INTERNATIONAL HYDROGRAPHIC ORGANIZATION

MANUAL ON HYDROGRAPHY

FINAL DRAFT

January 2005

PUBLISHED BY THE

INTERNATIONAL HYDROGRAPHIC BUREAU

M O N A C O

INTERNATIONAL HYDROGRAPHIC ORGANIZATION

MANUAL ON HYDROGRAPHY

FINAL DRAFT

January 2005

Published by the
International Hydrographic Bureau
4, Quai Antoine 1er
B.P. 445 - MC 98011 MONACO Cedex
Principauté de Monaco
Telefax: (377) 93 10 81 40
E-mail: info@ihb.mc
Web: www.iho.shom.fr

PREFACE

The IHO Manual on Hydrography general objective is to provide knowledge on the concepts involved in hydrography as well as guidance to plan and execute hydrographic surveys. The Manual is considered to be a professional guide for hydrographic surveyors and a tool for teachers and students involved in hydrographic courses or programs.

The preparation of this Manual started after the majority of IHO Member States (MS) responded in favour of proceeding with a project that could result in a Hydrographic Manual (1999). The IHB proposed then the establishment of a Working Group that met for the first time at the IHB premises, 20-22 June 2001, there the Table of Content was agreed; Team Leaders were identified to deal with specific subjects, being responsible for the compilation of experts' contributions, and a work program was defined. In 2004 a second meeting took place to review the result obtained and decide on a draft version of the Manual. After collecting comments from MS, the final version was prepared and the IHO Manual on Hydrography was published.

The Manual is considered to be a worthy product that contributes to the mission of the IHO, whose objectives are:

- The co-ordination of the activities of national hydrographic offices;
- The greatest possible uniformity in nautical charts and documents;
- The adoption of reliable and efficient methods of carrying out and exploiting hydrographic surveys;
- The development of the sciences in the field of hydrography and the techniques employed in descriptive oceanography.

It has to be acknowledged that several Hydrographic Offices (HO) have made great efforts in preparing and keeping up to date their version of a Hydrographic Manual, almost since their establishment, but the resources, both, in time and manpower required for this activity have precluded several HOs from continuing this practice, thus agreeing in the need to co-operate and co-ordinate the efforts for the preparation of an IHO Hydrographic Manual. A Manual that could be useful to everybody, containing specific aspects on hydrography and other matter only in general terms, as they are much more thoroughly covered in existing textbooks which refer to them in considerably greater detail.

The content of this Manual is divided into seven chapters:

- Chapter 1 refers to the principles of hydrographic surveying, including its specifications;
- Chapter 2 refers to positioning;
- Chapter 3 to refers to depth determination, including both the principles and techniques used;
- Chapter 4 provides information on sea floor classification and object detection;
- Chapter 5 refers in particular to water levels and flow;
- Chapter 6 is devoted to topographic surveying applied to hydrography;
- Chapter 7 provides, in a structured way, complete details on hydrographic practice;
- Annexes with Acronyms, Bibliography and other relevant information.

It is the IHB responsibility to keep this Manual up dated, following inputs from MS and other organisations who are encouraged to provide the IHB with relevant information in this regard. If needed, the IHB will request the advice of the FIG/IHO/ICA International Advisory Board on Standards of Competence for Hydrographic Surveyors and Nautical Cartographers, on the best way to include new material and/or organise the relevant chapter(s).

The IHB will keep the digital version of this Manual on the IHO web page and will print hard copies on demand. It is planed to update the Manual annually.

TABLE OF CONTENTS

INTRODUCTION

BRIEF HISTORY OF HYDROGRAPHY *by Admiral Ritchie (UK)*

The oldest navigational chart known today is the Carte Pisane, so named as it was bought in 1829 from a Pisan family by the Bibliothèque Nationale in Paris. It was drawn on an animal skin towards the end of the 13thC, probably in Genoa where a school of marine cartography had been established; there was a similar school in Venice, while a third school was developed on the isle of Majorca. Known as 'portolans' the charts produced by each of these schools were similar in style and content. The most striking feature was networks of interconnecting rhumb lines emanating from compass roses representing 32 winds directions, each one of which could be used with dividers to set a ship's course. The entire Mediterranean coastline was depicted, the coastal names shown on the land leaving the sea area clear for track plotting. There were a few symbols including the cross for submerged rock but no depth soundings.

By the 15thC Portuguese and Spanish portolans enabled Mediterranean seamen to sail to southern England and Flanders to load wool.

For generations the northern seamen had navigated from one headline to another using written directions and soundings handed down from their forebears, a method of pilotage known as 'caping the ship'. With the development of printing Pierre Garcie of Rouen was the first to publish caping information in his 'Routier de la Mer', which he illustrated with simple woodcut coastal views.

Cornelius Anthonisz, a draughtsman of Amsterdam realised that woodcut blocks could be used to print charts on paper, his first being his 'Karte van Ostland' of the Baltic and North Seas. Whilst adopting rhumb lines and other portolan features, he used Ptolomy's projection which had recently been rediscovered in Constantinople.

Anhonisz had shown the way but it was Lucas Janszoon Waghenaer of Enkhuizen in Holland who, forty years on, printed paper charts from copper plate engravings. For many years he had travelled widely as a sea pilot gathering hydrographic information and when he came ashore at the age of 49 he enlisted fellow mariners to supply him with such material for compiling his charts. In 1584 Waghenaer published his great atlas 'Spieghel der Zeevaerdt' (Mirror of the Sea) containing 45 charts covering the European coasts from Norway to the Strait of Gibraltar. He introduced many new features such as coastal recognition profiles behind the coastlines; reducing the distances between harbours so that their approaches could be shown on a larger scale; the introduction of symbols for buoys, beacons, church spires etc. and soundings reduced to their depth at half tide.

Waghenaer had made the great breakthrough in producing a paper chart designed by a seaman for seamen. He had a number of Dutch followers so that, for over 100 years, Dutch charts were widely available, even of British waters; eventually King Charles decided that the whole of Britain's coasts and harbours should be surveyed.

For this massive task he selected a naval officer named Greenville Collins, granted him the title 'Hydrographer to the King' and provided him with the yacht Merlin. The work began in 1681 and took eleven years.

There was no general topographic map of the Kingdom to which Collins could relate his charts, nor did he have any method of finding his longitude and only the quadrant to assess his latitude; his soundings reduced to low water were fixed by compass bearings of shore marks which in turn were fixed by compass and measuring chain. In 1693 the resulting charts were published in an atlas entitled 'Great Britain's Coasting Pilot', which contained 47 charts and 30 pages of tide tables, sailing directions and coastal views. Precisely engraved, the charts included soundings and leading lines for harbour entry etc. The Pilot appealed to British seamen, a further twenty editions being published during the next hundred years.

During the 16thC a school of hydrography was formed in Dieppe by the many sea pilots who sailed to distant shores. In 1661 John Baptiste Colbert became Chief Minister to Louis XIV and among his many tasks was that of revitalising the French Navy. He not only took over the Dieppe school but established similar hydrographic centres in a number of other French ports. This enabled him to have surveys made of the whole French coastline, every chart being directly connected to the national triangulation established by the Cassini dynasty.

Colbert's cadre of hydrographers were working in New France and the mass of material coming from Quebec led to the establishment in Paris of the 'Dépôt Général des Cartes et Plans', now recognised as the first national Hydrographic Office. Denmark was the next nation to establish a Hydrographic Office, followed closely by the British in 1795; a further twenty or so countries established such offices in 19thC.

About 1775 two British surveyors, Murdoch Mackenzie and his nephew of the same name were largely responsible for the invention of the station pointers, a device with which a vessel's position could be precisely plotted by the observation of two horizontal angles between three fixed marks onshore. This was a major technical advance which revolutionised sea surveying throughout the 19thC during which the demands for navigational charts both for war and peace increased dramatically.

Even before World War I a number of national Hydrographers were considering how international co-operation could lead to the exchange of and the standardisation in chart design. With the end of the War the British and French Hydrographers jointly called for an international Conference at which delegates from 22 countries gathered in London in June 1919. Many Resolutions were adopted by the Conference concerning chart standardisation and finally a Resolution to form an International Hydrographic Office with three Directors.

H.S.H. Prince Albert I of Monaco, who had been kept in touch with the proceedings of the Conference, generously agreed to provide a building in the Principality to house the Bureau where it remains.

The history of hydrography during the 20thC, during which there have been many technical developments, can be followed in the 75th Anniversary Commemorative Issue of the International Hydrographic Review dated March 1997.

IMPORTANCE OF HYDROGRAPHY

Firstly it is necessary to consider the IHO definition of Hydrography, which stands as follows:

> *That branch of applied sciences which deals with the measurement and description of the features of the seas and coastal areas for the primary purpose of navigation and all other marine purposes and activities, including –inter alia- offshore activities, research, protection of the environment, and prediction services.* (IHO Pub. S-32)

Therefore, the development of a National Maritime Policy requires a well developed capability to conduct all these activities which will allow the obtaining of basic knowledge of the geographical, geological and geophysical features of the seabed and coast, as well the currents, tides and certain physical properties of the sea water; all of this data must then be properly processed so that the nature of the sea bottom, its geographical relationship with the land and the characteristics and dynamics of the ocean can be accurately depicted in all zones of national shipping. In brief, Hydrography, as defined, is the key to progress on all maritime activities, normally of great national economic importance.

To adequately address areas of safe and efficient operation of maritime traffic control; coastal zone management; exploration and exploitation of marine resources; environmental protection and maritime defence, it is necessary to create a Hydrographic Service. The Hydrographic Service, through systematic data collection carried out on the coast and at sea, produces and disseminates information in support of maritime navigation safety and marine environment preservation, defence and exploitation.

To adequately address areas such as:

- Safe and efficient operation of maritime traffic control;
- Coastal Zone Management;
- Exploration and Exploitation of Marine Resources;
- Environmental Protection;
- Maritime Defence.

It is necessary to create a Hydrographic Service. The Hydrographic Service, through systematic data collection carried out on the coast and at sea, produces and disseminates information in support of maritime navigation safety and marine environment preservation, defence and exploitation.

FIELDS OF COMPETENCE ASSOCIATED WITH HYDROGRAPHY

Maritime Transport

More than 80% of international trade in the world is carried by sea. Maritime commerce is a basic element for a nation's economy. Many areas and ports in the world do not have accurate nor adequate nautical chart coverage. Modern nautical charts are required for safe navigation through a country's waters and along coasts and to enter its ports. A lack of adequate nautical charts prevents the development of maritime trade in the waters and ports of the concerned nations.

The shipping industry needs efficiency and safety. Poorly charted areas and the lack of information can cause voyages to be longer than necessary, and may prevent the optimum loading of ships, thus increasing costs. The saving of time and money resulting from the use of shorter and deeper routes and the possibility to use larger ships or load ships more deeply may produce important economies for national industry and commerce. It is also very important to note that the SOLAS Convention Chapter V considers a ship unseaworthy if it does not carry up-to-date charts necessary for the intended voyage.

A solution to these problems would not be possible without the quality maps and charts produced and continually updated and distributed by a Hydrographic Service. These charts, produced by means of modern hydrographic surveys, are required to enable the larger ships of today to navigate through national waters and enter ports the access to which was formerly insecure and therefore are essential tools for the creation of coastal nations' incomes.

Modern charts also provide information required to create the routeing systems established by international conventions and to meet the economic interests of the coastal state.

Coastal Zone Management

Adequate coastal zone management includes items such as construction of new ports and the maintenance and development of existing ones; dredging operations for the maintenance of charted depths and for the establishment, monitoring and improvement of channels; control of coastal erosion; land reclamation from the sea; establishment and monitoring of dumping grounds for industrial waste; extraction of mineral deposits; aquacultural activities; transportation and public works projects including construction of near shore infrastructure.

Precise large-scale surveys provide the primary data essential for projects involving all items mentioned above. Due to the rapid changes to which shorelines are subject, these surveys must be updated with the frequency dictated by the monitoring and analysis process. The information collected by Hydrographic Offices about the coastal zone provides essential input to coastal zone GIS (Geographic Information Systems) which are increasingly being used for better overall management and decision-making with regard to conflicting uses within the coastal region. The users of hydrographic information go beyond the traditional user group, mariners, to include government agencies, coastal managers, engineers, and scientists.

Exploration and exploitation of marine resources

Although intended primarily to support safety of navigation, the extensive data-bases amassed over the years by Hydrographic Offices, together with their various products and services, are of considerable economic value in assisting the management and exploitation of natural marine resources. In recent years, it has become more evident that inadequate hydrographic services not only restrict the growth of maritime trade but also lead to costly delays in resource exploration.

Coastal and offshore sedimentary areas may contain mineral deposits, in particular hydrocarbons, which require adequate surveys in order to be identified. If the existence of these hydrocarbons is confirmed, this will lead to the coastal nation's undertaking development of hydrocarbon production which implies interpretation of the sea floor morphology; navigation safety for the transportation of these hazardous cargoes; safety of offshore platforms and related sea floor transmission systems and the placement of production wells and the laying of pipelines. Bathymetric, tidal and meteorological data provided by a Hydrographic Service is a fundamental element in the development of a hydrocarbon industry.

The fishing industry is also a source of national wealth. Fishermen need marine information not only for the safe navigation of their vessels but also for safe deployment of their fishing gear, which will prevent costly losses. In addition, oceanographic charts, compiled and produced by Hydrographic Offices, are now being extensively used by the fishing industry.

Fishery activities need detailed charts in order to:

- avoid loss of fishing gear and fishing vessels on undetected or poorly charted obstructions;
- identify fishing areas;
- locate areas where fishing is limited or prohibited.

This kind of information is subject to frequent changes and therefore needs constant updating. Hydrographic surveying is essential to obtain timely and up-to-date information and should be periodically repeated.

The trend of modern fishery science is orientated towards habitat management; bathymetry and other ocean data will provide important input for proper species management and development.

Environment Protection and Management

An essential factor for the protection of the environment is safe and accurate navigation. Pollution caused by wrecks and oil spills are a major damage factor, the economic consequences of which are more devastating than is commonly imagined, but which, in some cases, have been estimated at US $ 3 billion for a single incident.

The value of navigation services for the protection of the marine environment has been internationally recognized. In this respect, it should be noted that Chapter 17 of Agenda 21 of the United Nations Conference on the Environment and Development (UNCED), held in 1992, recognized that **"Hydrographic charting is vitally important to navigational safety"**

Marine Science

Marine science depends largely on bathymetric information. Global tide and circulation models, local and regional models for a wide variety of scientific studies, marine geology/geophysics, the deployment/placement of scientific instrumentation and many other aspects of marine science depend on bathymetry provided by Hydrographic Services.

National Spatial Data Infrastructure

In the information age it is realised by governments that good quality and well managed spatial data are an essential ingredient to economic and commercial development, and to environmental protection. For this reason many nations are establishing national spatial data infrastructures, bringing together the services and data sets of major national spatial data providers, for example topography, geodesy, geophysics, meteorology, and bathymetry. The Hydrographic Service is an important part of the national spatial data infrastructure.

Maritime Boundary Delimitation

Good hydrographic data is essential to proper delimitation of the maritime boundaries as detailed in the United Nations Convention on the Law of the Sea.

Maritime Defense

Navies are major users of nautical chart products in that they must be prepared for deployment to many areas in the world and typically must maintain a large set of charts. The unique risks associated with the carriage of munitions and nuclear material make it important for such vessels to have up-to-date information. The marine data and information provided by national Hydrographic Offices support a variety of products used in naval operations. Surface, submarine, anti-submarine, mine-hunting and air-sea naval operations need nautical information products very different one from another. Hydrographic and oceanographic data necessary for the preparation of such products must be available if national investment in defence is to be optimised.

Tourism

Good charts are particularly important to the development of the economically important industry of tourism, especially involving cruise ships. The potential of the cruise ship industry is especially important to developing nations. Yet this important source of revenue cannot be properly developed if safe navigation to remote touristic landscapes is prevented or limited by a lack of adequate charts. Tourism is one of the major growth industries of the 21st Century.

Recreational boating

The recreational boating community represents a large percentage of mariners. It is generally not mandatory for leisure craft to carry charts and recreational mariners often do not update their charts; however, the advent of digital chart information is making it possible for the recreational user to have updated chart information readily available along with many types of value added information such as marina locations, etc. This development is likely to result in the recreational leisure sector becoming a significantly larger user of the hydrographic data as greater numbers of people become able to afford boat ownership. Again income from this sector is increasingly significant to many countries.

As it can be seen, it is extremely difficult to quantify the economic and commercial benefits which flow from a national hydrographic programme, but several studies by IHO Member States have suggested that the cost to benefit ratio is about 1:10 for major maritime nations. It is also true that volumes of maritime trade are growing continuously and, in the future, the exploitation and sustainable development of the national maritime zones will become a major pre-occupation of government and industry.

It should also be noted that, in economic parlance, the national hydrographic programme is regarded as a "Public Good". That is to say the necessary services required in the public interest will not be supplied at optimal levels by market forces alone. In every IHO Member State the provision of hydrographic services is a responsibility of central government, as an essential component of national economic development. This overall and important economic dimension of the work has sometimes been obscured by the emphasis on sector interests served by hydrographic services, and more recently by legislative or regulatory requirements. It is clear that the economic dimension of Hydrography deserves greater attention than it has received in the past.

CHAPTER 1
PRINCIPLES OF HYDROGRAPHIC SURVEYING

By Captain Muhammad ZAFARYAB (Pakistan)

1. INTRODUCTION

Hydrographic surveying deals with the configuration of the bottom and adjacent land areas of oceans, lakes, rivers, harbours, and other water forms on Earth. In strict sense, it is defined merely as the surveying of a water area; however, in modern usage it may include a wide variety of other objectives such as measurements of tides, current, gravity, earth magnetism, and determinations of the physical and chemical properties of water. The principal objective of most hydrographic surveys, is to obtain basic data for the compilation of nautical charts with emphasis on the features that may affect safe navigation. Other objectives include acquiring the information necessary for related marine navigational products and for coastal zone management, engineering, and science[1].

The purpose of hydrographic surveying is[2]:

- To collect, with systematic surveys at sea, along the coast and inland, georeferenced data related to:

    - Shoreline configuration, including man made infrastructure for maritime navigation i.e. all those features on shore that are of interest to mariners.
    - Depths in the area of interest (including all potential hazards to navigation and other marine activities).
    - Sea bottom composition.
    - Tides and Currents.
    - Physical properties of the water column.

- To process the information collected in order to create organized databases capable of feeding the production of thematic maps, nautical charts and other types of documentation for the following most common uses :

    - Maritime navigation and traffic management.
    - Naval operations.
    - Coastal zone management.
    - Marine environment preservation.
    - Exploitation of marine resources and laying of submarine cables/pipelines.
    - Maritime boundaries definition (Law of the Sea implementation).
    - Scientific studies.

Mariners have unquestioning faith in nautical charts and where no dangers are shown, they believe that none exist. Nautical chart is an end product of a hydrographic survey. Its accuracy and adequacy depend on the quality of the data collected during the surveys[3]. A nautical chart is a graphic portrayal of the

---

1 NOAA Hydrographic Manual Part-1, Edition July 4, 1976, P-1-3, www.thsoa.org/pdf/hm1976/part1ch123.pdf

2 International Hydrographic Organisation, Monaco, National Maritime Policies and Hydrographic Services (M-2), P-13.

3 NOAA Hydrographic manual Part-1, Edition July 4, 1976, P-1-3, www.thsoa.org/pdf/hm1976/part1ch123.pdf

marine environment; showing the nature and form of the coast, depths of the water and general character and configuration of the sea bottom, locations of dangers to navigation, rise and fall of the tides, cautions of manmade aids to navigation, and the characteristics of the Earth's magnetism. The actual form of a chart may vary from a traditional paper chart to an electronic chart.

An electronic chart is not simply a digital version of a paper chart; it introduces a new navigation methodology with capabilities and limitations very different from paper charts. The electronic chart has become the legal equivalent of the paper chart as approved by the International Maritime Organization. Divergences in purpose have led to the publication of various "new-generation" charts. Bathymetric charts developed from digital data or created from multi-beam sounding data allow the underwater relief to be visualised by means of varying blue tints and isobaths. Similarly, side-scan sonar mosaics have been published in the form of charts or atlases to characterise the large geomorphological structures. Such charts no longer have, as their object, the safety of navigation, but rather, the knowledge of the environment required for submarine navigation, oceanographic research or industrial applications, such as cable laying, seabed mining and oil exploitation.

Hydrographic surveying is undergoing fundamental changes in measurement technology. Multibeam acoustic and airborne laser systems now provide almost total seafloor coverage and measurement as compared to the earlier sampling by bathymetric profiles. The capability to position the data precisely in the horizontal plane has been increased enormously by the availability of satellite positioning systems, particularly when augmented by differential techniques. This advance in technology has been particularly significant since navigators are now able to position themselves with greater accuracy than that of the data on which older charts are based[4].

2. HYDROGRAPHIC SURVEYING

2.1 Survey Specifications

Requirements for hydrographic surveys arise as the result of policy decisions, product user reports or requests, national defence needs, and other demands. The inception of a specific hydrographic survey project follows an evaluation of all known requirements and the establishment of priorities. Among the many objective and subjective factors that influence the establishment of priorities are national and agency goal, quantitative and qualitative measures of shipping and boating, the adequacy of existing surveys, and the rate of change of the submarine topography in the area[5].

To accommodate in a systematic manner different accuracy requirements for areas to be surveyed, four orders of survey are defined by IHO in publication S-44 edition 98. These are described in subsequent paragraphs. Tables 1 and 2 summarize the overall, requirements and are in fact the essence of the complete standard[6].

2.1.1 Special Order hydrographic surveys approach engineering standards and their use is intended to be restricted to specific critical areas with minimum under keel clearance and where bottom characteristics are potentially hazardous to vessels. These areas have to be explicitly designated by the agency responsible for survey quality. Examples are harbours, berthing areas, and associated critical channels. All error sources must be minimized. Special Order requires the use of closely spaced lines in

---

[4] International Hydrographic Organisation, Monaco, National Maritime Policies and Hydrographic Services (M-2), P-19.

[5] NOAA Hydrographic manual Part-1, Edition July 4, 1976, P-2-1, www.thsoa.org/pdf/hm1976/part1ch123.pdf

[6] International Hydrographic Organisation, Monaco, IHO Standards for Hydrographic Surveys (S-44), P-14-8, fourth edition 1998.

conjunction with side scan sonar, multi-transducer arrays or high resolution multibeam echosounders to obtain 100% bottom search. It must be ensured that cubic features greater than 1m can be discerned by the sounding equipment. The use of side scan sonar in conjunction with a multibeam echosounder may be necessary in areas where thin and dangerous obstacles may be encountered.

2.1.2 Order 1 hydrographic surveys are intended for harbours, harbour approach channels, recommended tracks, inland navigation channels, and coastal areas of high commercial traffic density where under keel clearance is less critical and the geophysical properties of the seafloor are less hazardous to vessels (e.g. soft silt or sandy bottom). Order 1 surveys should be limited to areas with less than 100 m water depth. Although the requirement for seafloor search is less stringent than for Special Order, full bottom search is required in selected areas where the bottom characteristics and the risk of obstructions are potentially hazardous to vessels. For these areas searched, it must be ensured that cubic features greater than 2 m up to 40 m water depth or greater than 10% of the depth in areas deeper than 40 m can be discerned by the sounding equipment.

2.1.3 Order 2 hydrographic surveys are intended for areas with depths less than 200 m not covered by Special Order and Order 1 and where a general description of the bathymetry is sufficient to ensure there are no obstructions on the seafloor that will endanger the type of vessel expected to transit or work the area. It is the criteria for a variety of maritime uses for which higher order hydrographic surveys cannot be justified. Full bottom search may be required in selected areas where the bottom characteristics and the risk of obstructions may be potentially hazardous to vessels.

2.1.4 Order 3 hydrographic surveys are intended for all areas not covered by Special Order, and Orders 1 and 2 in water depths in excess of 200 m.

Notes:

For Special Order and Order 1 surveys the agency responsible for the survey quality may define a depth limit beyond which a detailed investigation of the seafloor is not required for safety of navigation purposes.

Side scan sonar should not be used for depth determination but to define areas requiring more detailed and accurate investigation.

TABLE 1.1
Summary of Minimum Standards for Hydrographic Surveys

ORDER	Special	1	2	3
Examples of Typical Areas	Harbours, berthing areas, and associated critical channels with minimum under keel clearances	Harbours, harbour approach channels, recommended tracks and some coastal areas with depths up to 100 m	Areas not described in Special Order and Order 1, or areas up to 200 m water depth	Offshore areas not described in Special Order, and Orders 1 and 2
Horizontal Accuracy (95% Confidence Level)	2 m	5 m + 5% of depth	20 m + 5% of depth	150 m + 5% of depth

ORDER	Special	1	2	3
Depth Accuracy for Reduced Depths (95% Confidence Level) (1)	a = 0.25 m b.= 0.0075	a = 0.5 m b.= 0.013	a = 0.1.0 m b.= 0.023	Same as Order 2
100% Bottom Search	Compulsory (2)	Required in selected areas (2)	May be required in selected areas	Not applicable
System Detection Capability	Cubic features > 1 m	Cubic features > 2 m in depths up to 40 m; 10% of depth beyond 40 m (3)	Same as Order 1	Not applicable
Maximum Line spacing (4)	Not applicable, as 100% search compulsory	3 x average depth or 25 m, whichever is greater	3-4 x average depth or 200 m, whichever is greater	4 x average depth

(1) To calculate the error limits for depth accuracy the corresponding values of 'a' and 'b' listed in Table 1 have to be introduced into the formula

$$\pm v\ [a^2 + (b*d)^2]$$

with

a constant depth error, i.e. the sum of all constant errors
b*d depth dependent error, i.e. the sum of all depth dependent errors
b factor of depth dependent error
d depth

(2) For safety of navigation purposes, the use of an accurately specified mechanical sweep to guarantee a minimum safe clearance depth throughout an area may be considered sufficient for Special Order and Order 1 surveys.

(3) The value of 40 m has been chosen considering the maximum expected draught of vessels.

(4) The line spacing can be expanded if procedures for ensuring an adequate sounding density are used

The rows of Table 1 are explained as follows:

- Row 1 "Examples of Typical Areas" gives examples of areas to which an order of survey might typically be applied.

- Row 2 "Horizontal Accuracy" lists positioning accuracies to be achieved to meet each order of survey.

- Row 3 "Depth Accuracy" specifies parameters to be used to calculate accuracies of reduced depths to be achieved to meet each order of survey.

- Row 4 "100% Bottom Search" specifies occasions when full bottom search should be conducted.

- Row 5 "System Detection Capability" specifies the detection capabilities of systems used for bottom search.

- Row 6 "Maximum Line Spacing" is to be interpreted as
  - spacing of sounding lines for single beam sounders, and
  - distance between the outer limits of swaths for swath sounding systems.

2.2 Survey Planning

Survey planning covers a wide range of activities from the development of an idea for a survey within the Hydrographic Office and its subsequent issue as Project Instructions / Hydrographic Instructions (HIs), to the detailed planning and organisation of a surveying ship to fulfil a practical task. It covers inter departmental liaison at Government level, diplomatic cooperation and the allocation of numerous expensive resources. It also covers prioritization of resources and day to day running of a survey ship employed on surveying task. Survey planning involves blending of these activities into a coherent pattern aimed at the achievement of a specific task.

A survey begins long before actual data collection starts. Some elements, which must be decided, are[7]:

- Exact area of the survey.
- Type of survey (reconnaissance or standard) and scale to meet standards of chart to be produced.
- Scope of the survey (short or long term).
- Platforms available (ships, launches, aircraft, leased vessels, cooperative agreements).
- Support work required (aerial or satellite photography, geodetics, tides).
- Limiting factors (budget, political or operational constraints, positioning systems limitations, logistics).

Once these issues are decided, all information available in the survey area is reviewed. This includes aerial photography, satellite data, topographic maps, existing nautical charts, geodetic information, tidal information, and anything else affecting the survey. HO will normally undertake this strategic planning of surveys in cooperation with other organisations and, from this, Projects Instructions / Hydrographic Instructions (HIs) will be compiled by the Hydrographer and issued for compliance. Details provided in Project Instructions / HIs will include some or all of the following, depending on the type of survey required[8]:

- Survey limits.
- Data requirement and resolution.
- Method of positional control, together with the accuracy expected.
- Use to be made of sonar.
- How the survey report is to be rendered and target date if appropriate.
- A general, and at times detailed, description of the reason for the survey priorities, methods to be employed, particular observations to be made and other relevant guidance or instruction.

In addition, appendices to HIs will give instruction or guidance on the following:

- Horizontal datum, projection and grid to be used.
- Wrecks in the area.

---

7 Bowditch -The American Practical Navigator, P-411, http://www.irbs.com/bowditch/

8 Admiralty, General Instructions for Hydrographic Surveys (GIHS), Sixth Edition, 1992, P- 5-3.

- Tidal datum and observations required.
- Particular instructions regarding the collection of data in respect of oceanography, geophysics, sailing directions, air photography etc.

On receipt of Project Instructions/HIs, the survey planners then compile sound velocity information, climatology, water clarity data, any past survey data, and information from lights lists, sailing directions, and notices to mariners. Tidal information is thoroughly reviewed and tide gauge locations chosen. Local vertical control data is reviewed to see if it meets the expected accuracy standards, so the tide gauges can be linked to the vertical datum used for the survey. Horizontal control is reviewed to check for accuracy and discrepancies and to determine sites for loc al positioning systems to be used in the survey.

Development of a general survey plan and subsequent site specific survey plans will create a more efficient survey. The general survey plan addresses the way that surveys are planned, performed, and processed. This plan must be well thought out and robust to account for as many contingencies as possible. This plan includes training, software, equipment maintenance and upgrades, logistics, all data requirements, schedule, safety, and weather. The site specific survey plan will address local notifications, survey lines, datum, data density, and specific equipment and personnel that will meet the general survey plan requirements. Few are described below:

- Training of surveyors should be catered during a survey operation in order to ensure appropriate competencies are maintained.

- Data logging and processing software are critical in a survey operation. These should be user friendly and personnel employed on these need to be well conversant with its all functions.

- Suitable survey platform and equipment should be selected. Some equipment will lend itself to particular types of surveys and others will be more general in use. It is paramount that a proper selection is made.

- The purpose for the survey will usually dictate the data requirement (density, coverage, and precision). However, if there are no impact to cost and schedule, then as many requirements should be addressed as possible.

- Schedule is often a critical element in a hydrographic survey. The data requirement usually has as a specific deliverable date assigned, such that the survey data collection and processing occur within a very specific time frame. This requires that the personnel and equipment resources be adequate to meet this need. In some cases, if the schedule cannot be met, then the survey simply will not be requested and other sources will be used. Considering this, it is important to plan and analyze all aspects of a general survey plan with the ability to meet schedule as a prime element.

- Safety is the primary consideration. It is incumbent on the person in charge in the field to evaluate every situation for possible hazards. If there is an identified hazard, then it needs to be addressed before continuing with the activity.

- Notifications to the local authorities / harbourmaster office should be made with enough time to allow them to notify the local mariners.

- Survey lines for multibeam surveys should follow the contours of the harbour bottom. This will reduce the changes in bottom coverage created by different water depths. However, when using a single beam survey system, the lines should run perpendicular to contours. This will help in determining changes in the bottom relief. Multibeam survey lines also need to be spaced so as to achieve the proper amount of overlap or data density to meet the survey standard.

- An integral part of the data of a survey is the reference datum. It is required, by good survey practice, to clearly indicate by note on the published survey the actual vertical and horizontal reference used, and the procedures used to establish the datum for the survey. WGS-84 is being used worldwide.
- Data density will vary based on method of survey, water depth, and need. The method of survey will be determined by equipment available for the survey, the personnel, and survey site conditions. If only a single beam survey system is available, then data density will be less. With a multibeam system, the greater the water depth the less dense the data will be, unless multiple passes are made. The type of survey will dictate the data redundancy or data overlap requirements.

- It is important to standardize the equipment as much as possible to limit training, maintenance and overheads.

2.3 Data Gathering

Data gathering is dependent upon various factors. The survey requirements, the platform and equipment available and the time specified for a particular task will determine the amount of data to be collected. A large amount of data can be collected using latest hydrographic software's and tools like multibeam echo sounders. In particular, the purpose of the survey will usually dictate the data requirement (data density, data coverage, and data precision). However, if there is no impact to cost and schedule, then as many data may be collected as possible during field survey. The data collection be made in methodical manner starting from one side of the area ending on other.

It should be noted that data redundancy and data density are not the same thing. Data density is the number of soundings per unit of area, while data redundancy refers to data overlap or data collected at a different time at the same location. The type of survey defines data redundancy or data overlap requirements. Full coverage surveys deal more with data density insuring that all bottom features/obstructions have been located. These need to be clearly understood by those requesting the survey and those doing the survey to insure compliance with the standards specified by IHO.

2.4 Data Processing

Data processing must be done under strict quality control criteria. Hydrographic data is either collected by automated systems or converted into an automated format. Final data processing and plotting are accomplished using onboard or office-based computer systems. A standard approach for a hydrographic survey is the collect-process-collect methodology[9]. The data collected is processed and subsequently gaps and areas with questionable data re-surveyed. Most of the hydrographic systems are capable of performing "field-finish" operations, wherein survey data is collected, processed, plotted and analyzed in the field. Comprehensive survey planning is required for an integrated approach that generates the base line for all real-time and post processing operation with the system. An example of such a model is given below[10]:

---

[9] B. Bourgeois, F. Petry, M. Harris & P. Alleman, "A GIS Integration Approach for Dynamically Reconfigurable Surveys", The Hydrographic Journal, January 1999, P 3-10.

[10] Pentti Junni & Ralf Lindgren, "The Hydrophic Information System – Co-operation, Concept and Future", Finish Maritime Administration, http://www.esri.com/library/userconf/proc97/proc97/pap619/p619.html

Data + Process = Job

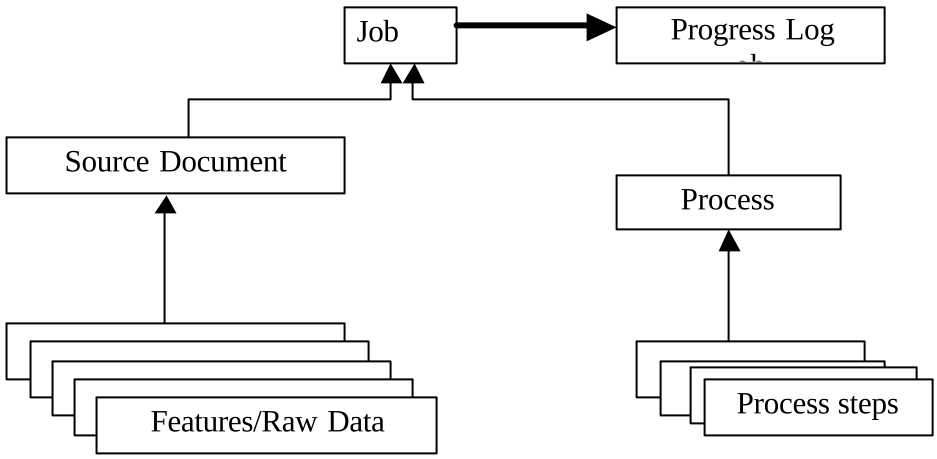


Fig. 1.1 "Data Processing model"

This model describes the different processes that can ideally handle the hydrographic information. The process contains several steps. The comments of each process step along with results, statistics should be recorded in a progress log. Further, the source and general quality information of any new data be described in source document which is stored in the database.

The core requirement of data processing is the generation of valid data; which has been sufficiently processed i.e. undergone through various procedures at various stages or represented so that evaluation can be made. These procedures/processing steps could be applied in real-time or during post processing but have to ensure that the final product meets the standards and specifications defined by IHO.

Care is to be exercised in processing the raw data. It is to be ensured that all errors have been eliminated and necessary corrections e.g. system calibration factors and sensor offsets, or variable values such as sound velocity profiles and tide values for the reduction of soundings, have been applied. The processing should strive to use all available sources of information to confirm the presence of navigationally significant soundings and quality data. Few processing steps outlined below are only to be interpreted as an indication, also with regard to their sequence, and are not necessarily exhaustive[11]:

- **Position:** Merging of positioning data from different sensors (if necessary), qualifying positioning data, and eliminating position jumps.

- **Depth corrections:** Corrections should be applied for water level changes, measurements of attitude sensors, and changes of the draught of the survey vessel (e. g. squat changing with speed; change over time caused by fuel consumption). It should be possible to re-process data for which corrections were applied in real-time.

- **Attitude corrections:** Attitude data (heading, pitch, roll) should be qualified and data jumps be eliminated.

---

[11] IHO, Monaco, Circular Letter 45/2001, "Guidelines for the Processing of High Volume Bathymetric Data", Para-3.2, dated 5 Oct 01.

- **Sound velocity:** Corrections due to refraction be calculated and applied; If these corrections have already been applied in real-time during the survey, it should be possible to override them by using another sound velocity profile with the advent of MBES, the application of S V has become critical.

- **Merging positions and depths:** The time offset (latency) and the geometric offset between sensors have to be taken into consideration.

2.5 Data Analysis

The accuracy of the results of survey measurement should always be quoted to show how good or reliable they are. Since no equipment is entirely free of errors, therefore, errors are introduced in all observations. In addition, errors are introduced in computations by approximations in formulae or by rounding. Observational techniques are designed to eliminate all but small random errors, which can then be analyzed by rigorous techniques to quantify the accuracy of the observations. Various errors, their sizes and procedures to eliminate are as under:

ERROR	SIZE	ELIMINATION
Blunder	Large	Training, care procedures.
Constant	Usually small, but fixed	Calibration or procedures
Periodic	Usually small, but variable	Procedure (repetition), even for large errors
Random	Usually small	Only reduced, by repetition

Constant, Systematic and Periodic errors are frequently considered together as 'Systematic Errors'. Constant and Systematic errors are accumulative, and thus cannot be reduced by repetition. Random errors are present in all observations; the result can never be 'exact'. These errors are as likely to be positive as negative and more likely to be of small size.

It is important that no method of adjustment can obtain an accurate solution from inaccurate observations. All errors other than small random ones must be eliminated prior to the adjustment. However, it may be possible to isolate a 'systematic' error by analysis, provided that sufficient data is available. It is clearly desirable to know when Constant and / or Systematic Errors are present in observations. Constant Errors are often difficult to detect, and may only become apparent during computation or in special checks, e.g., an incorrectly calibrated crystal of a Tellurometer can be detected if that Tellurometer is compared with another. Periodic and Random Errors however can often be detected by analyzing a series of observations. The algebraic difference between each observation and the mean of all the observations is called the Residual of that observation. If only random errors are present, then the Residuals will vary randomly in magnitude and sign. If systematic errors are present then the magnitudes and / or the signs of the Residuals will show systematic trends. To assist in data analysis, raw data attributes and metadata should be mentioned for subsequent evaluation.

2.6 Data Quality

Quality is about "fitness for the use". It has to do with the extent to which a data set, or map output satisfied the needs of the person judging it. Error is the difference between actual and true data. Error is a major issue in quality. It is often used as an umbrella term to describe all the types of the effects that cause data to depart from what they should be [12]. To allow a comprehensive assessment of the quality of survey data, it is necessary to record or document certain information together with the survey data. Such information is important to allow exploitation of survey data by a variety of users with different requirements, especially as requirements may not be known when survey data is collected. The process of documenting the data quality is called data attribution; the information on the data quality is called metadata. Metadata should comprise at least information on[13]:

- The survey in general as e.g. date, area, equipment used, name of survey platform.
- The geodetic reference system used, i.e. horizontal and vertical datum; including ties to WGS 84 if a local datum is used.
- Calibration procedures and results.
- Sound velocity.
- Tidal datum and reduction.
- Accuracies achieved and the respective confidence levels.

Metadata should preferably be in digital form and an integral part of the survey record. If this is not feasible similar information should be included in the documentation of a survey. Data quality can be achieved by effective quality control either by automatic or manual means[14].

- **Automatic (Non-interactive) Quality Control:** In this, the coordinates (i.e. positions and depths) obtained should be controlled automatically by a programme using suitable statistical algorithms which have been documented, tested and demonstrated to produce repeatable and accurate results.

- **Manual (Interactive) Quality Control:** In this, the use of 3-D visualisation tools is strongly recommended. These tools should allow viewing the data using a zoom facility. The interactive processing system should also offer different display modes for visualisation, e.g. depth plot, error plot, single profile, single beam, backscatter imagery etc. and should allow for the visualisation of the survey data in conjunction with other useful information as e.g. shoreline, wrecks, aids to navigation etc; editing of data should be possible in all modes and include an audit trail. If feasible, data displays should be geo-referenced. The flags set during the automatic stage, which correspond to depths shallower than the surrounding area, should require explicit operator action, at least, for Special Order and Order 1 surveys. If the operator overrules flags set during the automatic stage, this should be documented. If a flag is set by the operator, the type of flag used should indicate this.

2.7 Data Quality - Presentation

2.7.1 Chart Reliability Diagrams

Traditionally, the quality of bathymetric data has been a subjective procedure. For a user, the quality of the data which is presented is assessed through the chart reliability diagram. This diagram is displayed as an inset on a chart and indicates the areas surveyed together with some detail, e.g. scale, line spacing, year

---

[12] Patrick McGlamery, "Issues of Authenticity of Spatial data", University of Connecticut USA, 66th IFLA Council and General Conference, http://magic.lib.uconn.edu.

[13] International Hydrographic Organisation, Monaco, IHO Standards for Hydrographic Surveys (S-44), P-12, fourth edition 1998.

[14] IHO, Monaco, Circular Letter 45/200, "Guidelines for the Procesing of High Volume Bathymetric Data", Para-3.3, dated 5 Oct 01.

of survey. Unfortunately, the very nature of the information displayed on a reliability diagram, the ability to qualify data quality is severely limited. For instance, if the chart user is unaware what a pre 1970 sonar swept area means or what might be inferred from a line spacing of "n" metres, then the reliability diagram is of little real use in determining the quality of the depth data shown.

The original concept of the reliability diagram was to classify the quality of survey data and depict the different classifications on a diagram in terms of good, fair or poor quality. The diagram was intended to provide the mariner with the capacity to assess the danger of deviating from the recommended track. However, there has been growing concern over the complexity of the reliability diagram and the increasing difficulty of maintaining it in a form which is simple for the chart user. If they are too complicated; reliability diagrams become difficult to construct as a cartographic activity, prone to error in construction, and its use would be ignored by the mariners.

Reliability diagrams fall well short of achieving the fundamental aim of providing an indication of data quality to the mariner and in a very simple form. Furthermore, given the precise navigation capability which ENC and the ECDIS can facilitate, users require a far more definitive assessment of data quality to be available so that they can use the available information prudently. Thus, an alternative to the existing reliability diagram is required as the final quality indicator.

Source diagrams and similar variants shown on charts are all considered to present similar shortcomings.

2.7.2 Zones of Confidence (ZOC)

The ZOC concept was developed by the IHO to provide a mean of classifying bathymetric data. ZOC provide a simple and logical mean of displaying to the mariner the confidence that the national charting authority places on any particular selection of bathymetric data. It seeks to classify areas for navigation by identifying the various levels of confidence that can be placed in the underlying data using a combination of the following criteria:

* Depth and position accuracy,
* Thoroughness of seafloor search, and
* Conformance to an approved quality plan.

Under this concept six ZOCs were developed and subsequently approved for inclusion as a part of IHO S-57. ZOCs A1, A2, and B are generated from modern and future surveys with, critically, ZOCs A1 and A2 requiring a full area search. ZOCs C and D reflect low accuracy and poor quality data whilst ZOC U represents data which is un-assessed at the time of publication. ZOCs are designed to be depicted on paper charts, as an insert diagram in place of the current reliability diagram, and on electronic displays.

It must be emphasized that ZOCs are a charting standard and are not intended to be used for specifying standards for hydrographic surveys or for the management of data quality. The depth and position accuracy specified for each ZOC refer to the errors of the final depicted soundings and include not only survey errors but also any other errors introduced in the chart production process. The following paragraphs summaries individual ZOC specifications:

2.7.2.1 ZOC A1 – Position and depth data gathered in accordance with procedures and accuracies specified. Surveys conducted using recognized technology with a full area search undertaken with the aim of ensuring that all significant features are detected and depths measured. Typically, the survey would have been undertaken on WGS 84, using DGPS or a minimum three lines of position with multibeam, channel or mechanical sweep system. Due to the intensity of data gathering and the considerable time required to achieve this standard it can be expected that data with a ZOC A1 rating will most likely indicate critical channels, berthing areas, areas with minimum under keel clearances, navigation channels, recommended tracks, harbours and harbour approaches.

2.7.2.2 ZOC A2 – Position and depth data gathered in accordance with procedures and accuracies specified. Survey conducted using recognized technology with a full area search undertaken with the aim of ensuring that all significant features are detected and depths measured. Typically, the survey would have been conducted using a modern survey echosounder with sonar or mechanical sweep. Although position and depth accuracies not as high as ZOC A1, seafloor coverage is such that the mariner should have a high level of confidence in the quality of data.

2.7.2.3 ZOC B – Position and depth data gathered in accordance with procedures and accuracies specified. However, a full area search has not been achieved and uncharted features, hazardous to surface navigation, although not expected, may exist. This ZOC indicates to the mariner a reasonable level of confidence in the quality of data. ZOC B has the same position and depth accuracies as those required for ZOC A2 and would apply to (e.g.) modern surveys which have not achieved a full seafloor search and feature detection. The prudent mariner would require more under keel clearance in this ZOC than in ZOC A1 or A2.

2.7.2.4 ZOC C – Position and depth accuracy less than that achieved for ZOC B as described. Depth data may originate from sources other than a controlled, systematic hydrographic survey (e.g. passage sounding). A full area search has not been achieved and depth anomalies may be expected. ZOC C indicates that the mariner should navigate with special care and allow, with due regard to the depth of water in which they are navigating, greater safety margins to the charted information.

2.7.2.5 ZOC D – Position and depth data is of a very low quality or cannot be assessed due to a lack of supporting information. A full area search has not been achieved and large depth anomalies may be expected.

2.7.2.6 **ZOC U** – The quality of bathymetric data has yet to be assessed.

TABLE 1.2

Category of Zones of Confidence in Data – ZOC Table

1	2	3		4	5
ZOC[1]	**Position**[2] **Accuracy**	**Depth Accuracy**[3]		**Seafloor**[4] **Coverage**	**Typical Survey**[5] **Characteristics**
A1	± 5m	= 0.50 + 1%d Depth(m) 10 30 100 1000	 Accuracy (m) ± 0.6 ± 0.8 ± 1.5 ± 10.5	Full area search undertaken. All significant seafloor features[4] detected have had depths measured.	Controlled, systematic survey[6] high position and depth accuracy, achieved using DGPS or a minimum of three high quality lines of position (LOP) and a multibeam, channel or mechanical sweep system.
ZOC[1]	**Position**[2] **Accuracy**	**Depth Accuracy**[3]		**Seafloor**[4] **Coverage**	**Typical Survey**[5] **Characteristics**
A2	± 20m	= 1.00 + 2%d Depth(m) 10 30 100 1000	 Accuracy (m) ± 1.2 ± 1.6 ± 3.0 ± 21.0	Full area search undertaken. All significant seafloor features[4] detected have had depths measured.	Controlled, systematic survey[6] high position and depth accuracy less than ZOC A1, and using a modern survey eachosounder[7] and a sonar or mechanical sweep system.
B	± 50m	= 1.00 + 2%d Depth(m) 10 30 100 1000	 Accuracy (m) ± 1.2 ± 1.6 ± 3.0 ± 21.0	Full area search not achieved; uncharted features, hazardous to surface navigation, are not expected, but may exist.	Controlled, systematic survey[6] achieving similar depth but lesser position accuracies than ZOC A2, using a modern survey eachosounder[7] but no sonar or mechanical sweep system.
C	± 500m	= 2.00 + 5%d Depth(m) 10 30 100 1000	 Accuracy (m) ± 2.5 ± 3.5 ± 7.0 ± 52.0	Full AREA SEARCH NOT ACHIEVED; depth anomalies may be expected.	Low accuracy survey or data collected on an opportunity basis such as soundings on passage.
D	Worse than ZOC C	Worse Than ZOC C			Poor quality or data that cannot be assessed due to lack of information.
U		Un-assessed			

Notes:

To decide on a ZOC category, all conditions outlined in Columns 2 to 4 of the Table must be met, Explanatory note numbers quoted in the table have the following meanings:

1. *The allocation of a ZOC indicates that particular data meets minimum criteria for position and depth accuracy and seafloor coverage defined in the Table. ZOC categories reflect a charting standard and not just a hydrographic survey standard. Depth and position accuracies specified for each ZOC category refer to the errors of the final depicted soundings and include not only hydrographic survey errors but also other errors introduced in the chart production process. Data may be further qualified by Object Class "Quality of Data" (M_QUAL) sub-attributes as follows:*

   * *Positional Accuracy (POSACC) and Sounding Accuracy (SOUACC) may be used to indicate that a higher position or depth accuracy has been achieved than defined in this Table (e.g. a survey where full seafloor coverage was not achieved could not be classified higher that ZOC B; however, if the position accuracy was, for instance ± 15 metres, the sub-attribute POSACC could be used to indicate this).*

   * *Swept areas where the clearance depth is accurately known but the actual seabed depth is not accurately known may be accorded a "higher" ZOC (i.e. A1 or A2) providing position and depth accuracies of the swept depth meets the criteria in this Table. In this instance, Depth Range Value 1 (DRVAL1) may be used to specify the swept depth. The position accuracy criteria apply to the boundaries of swept areas.*

   * *SURSTA, SUREND and TECSOU may be used to indicate the start and end dates of the survey and the technique of sounding measurement.*

2. *Position accuracy criteria at 95% CI (2.45 sigma) with respect to the given datum. It is the cumulative error and includes survey, transformation and digitizing errors etc. Position accuracy need not be rigorously computed for ZOCs B, C and D but may be estimated based on type of equipment, calibration regime, historical accuracy etc.*

3. *Depth accuracy of depicted soundings for (e.g.) ZOC A1 = 0.50 metres + 1% d at 95% CI (2.00 sigma) where d = depth in metres at the critical depth. Depth accuracy need not be rigorously computed for ZOCs B, C and D but may be estimated based on type of equipment, calibration regime, historical accuracy etc.*

4. *Significant seafloor features are defined as those rising above depicted depths by more than:*

   Note: Mariners should have due regard to limitations of sounding equipment when assessing margins of safety to be applied.

Depth	***Significant Feature***
< 10 metres	*> 0.1 x depth*
10 to 30 metres	*> 1.0 metre*
> 30 metres	*> (0.1 x depth minus 2.0 metres*

5. *Typical Survey Characteristics – These descriptions should be seen as indicative examples only.*

6. *Controlled, systematic surveys (ZOCs A1, A2 and B) – surveys comprising planned survey lines, on a geodetic datum that can be transformed to WGS 84.*

7. *Modern survey echosounder – a high precision single beam depth measuring equipment, generally including all survey echosounders designed post 1970.*

2.8 Data Production

The final data production can both be in digital and analog form. Schematic diagram is given below.

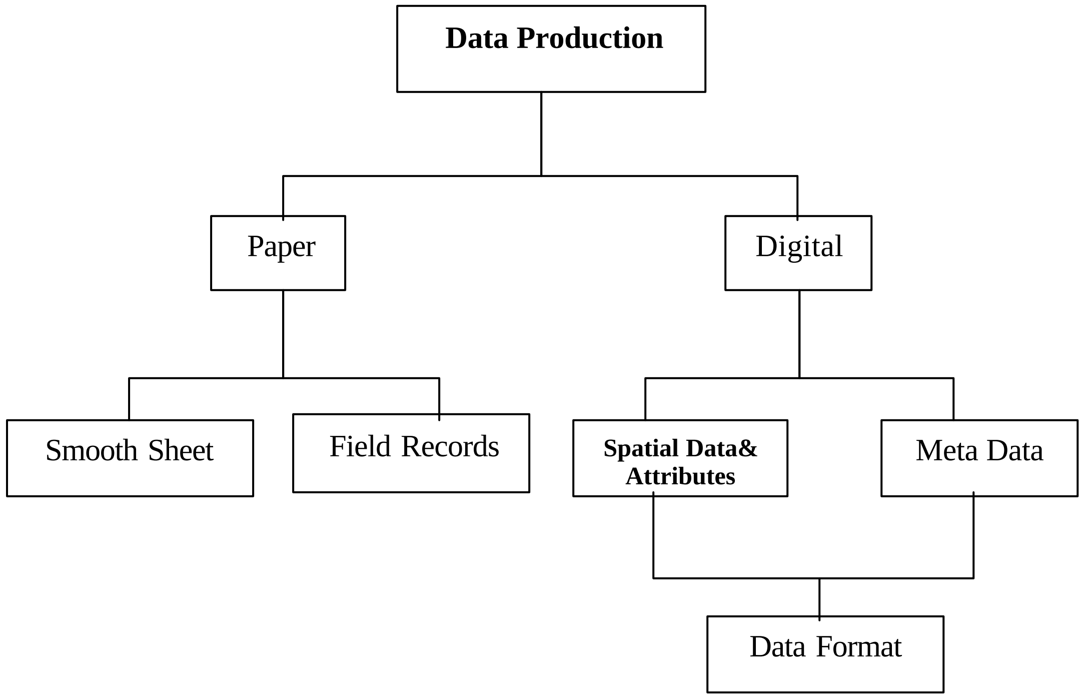


Fig. 1.2 "Digital Production schematic diagram"

Digital data should be in defined format to be directly imported into main database. As each survey typically includes numerous supporting documents and digital data files, deliverables must clearly be labelled in a manner that is both descriptive and intuitive to hydrographic office personnel. Ideally, standard operating procedures agreed by IHO and field units which covers such documents and digital data files, are enforced. **Manual** data should be clear, concise and in legible form which is properly labelled and marked[15].

After the data is collected, processed and plotted in the form of smooth sheets (manuscript in digital form), the inventory of the final deliverables is forwarded to Hydrographic office, the inventory of which should generally include[16]:

---

15 http://www.hydro.navy.gov.au/news/htf/htf.pdf.

16 Lieutenant Eric J. Sipos and Physical Scientist Castle Parker, "NOAA AHB Quality Assurance Inspections for Contract Hydrographic Surveys", NOAA Hydrographic Survey Division, USA

- Smooth Sheets.
- Digital files of smooth sheet with attributes.
- Raw and processed bathymetric data.
- Tide, sound velocity and vessel configuration files.
- Side scan data files.
- Descriptive report and supplemental reports.
- Field logs and documentation of processing.
- Calibration documentation.

2.9 Nautical Information System (NIS)

Nautical Information system is the combination of skilled persons, spatial and descriptive data, analytic methods and computer software and hardware - all organized to automate, manage and deliver information through presentation i.e. paper and digital charts. Previously, the main use of nautical chart databases was in the production of paper charts. Advances in navigation technology have set new demands on accuracy, reliability and the format of nautical charts. The positional accuracy of the chart should meet the increased accuracy of the positioning systems. To fully benefit from the dynamics of the modern positioning methods, the need of digital chart has arisen in parallel to the traditional printed charts. An international standard for digital hydrographic data has been developed by the International Hydrographic Organization (IHO). The valid version of the standard, S-57 edition 3.1 was adopted as the official IHO standard in November 2000 and is also specified in the International Maritime Organization (IMO) Performance Standards for Electronic Chart Display and Information Systems (ECDIS). S-57 describes the standard to be used for the exchange of digital hydrographic data between national Hydrographic Offices and for the distribution of digital data and products to manufacturers, mariners, and other data users. The most significant digital product being delivered in the S-57 format is the electronic navigational chart (ENC). The rapidly increased need for electronic navigational charts (ENC) has led to a situation for many hydrographic offices where there are two separate production lines for the two products, ENC cells and paper charts. It is essential for the safety of navigation that the products are not in conflict with one another. A typical NIS has four main functional subsystems[17] (Fig 1.3).

- **Data Input.** The data input subsystem allows the user to capture, collect, and transform spatial and thematic data into digital form. The data inputs are usually derived from a combination of hard copy maps, aerial photographs, remotely sensed images, reports, survey documents, etc.

- **Data Base - Storage and Retrieval.** Data storage and retrieval subsystem organizes the data, spatial and attribute, in a form which permits it to be quickly retrieved by the user for analysis, and permits rapid and accurate updates to be made to the database.

- **Data Base - Manipulation and Analysis.** The data manipulation and analysis subsystem allows the user to define and execute spatial and attribute procedures to generate derived information. This subsystem is commonly thought of as the heart of a GIS, and usually distinguishes it from other database information systems and computer-aided drafting (CAD) systems.

- **Data Output.** The data output subsystem allows the user to generate graphic displays, normally maps, and tabular reports representing derived information products.

---

[17] Dan Sherrill and Asa Carlsson, "The JANUS Solution for Hydrographic Information", T-Kartor AB Sweden- Box 5097 - S-291 05 Kristianstad – Sweden, ds@t-kartor.se & ac@t-kartor.se

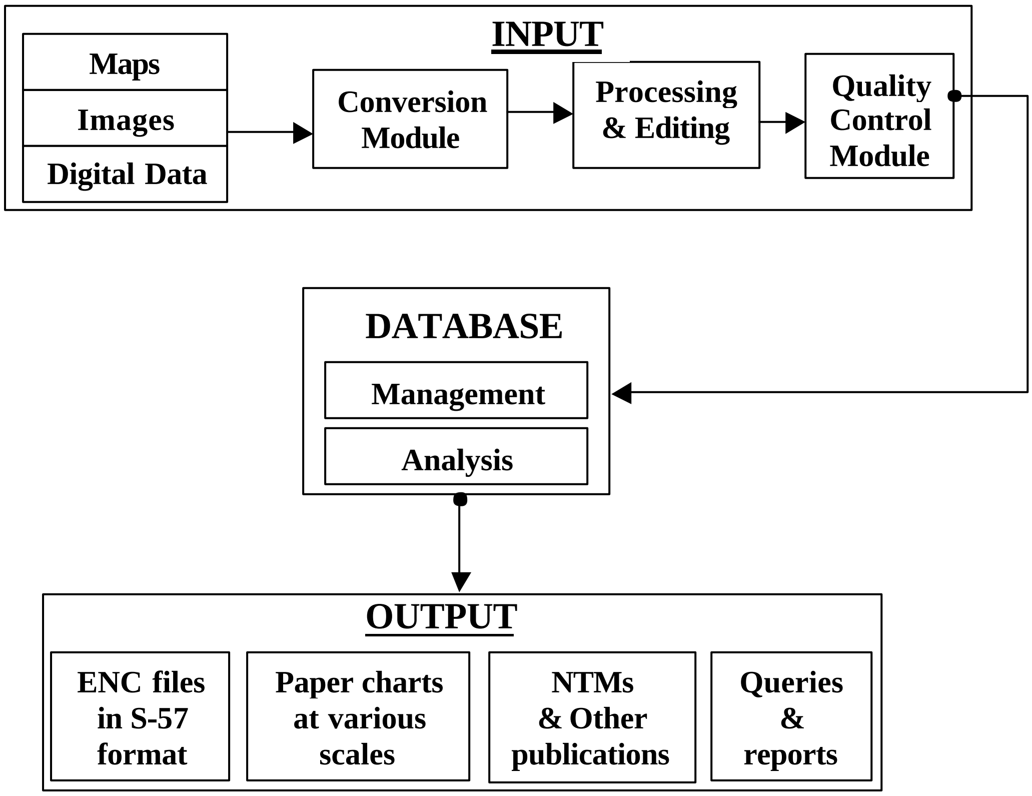


Fig. 1.3 "NIS Functional Subsystems"

There are four components of NIS; data, hardware, software, and users[18]. As shown in the Fig 1.4, the components must be integrated; they must be linked together and work in concert to support the management and analysis of spatial or mapped data.

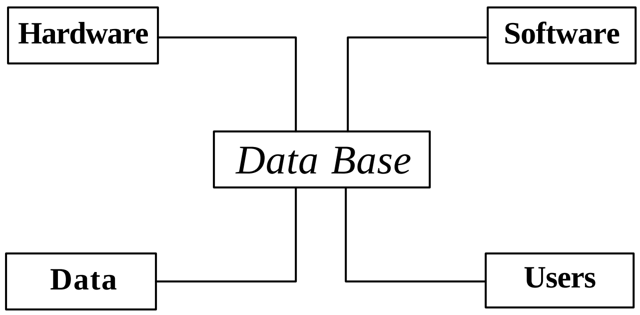


Fig. 1.4 "NIS Components "

[18] Lloyd P. Queen and Charles R. Blinn, "The Basics of geographic Information Systems", **lqueen@mercurry.forestry.umn.edu and cblinnlqueen@mercurry.forestry.umn.edu**

- **Data**. All data in a database are either spatial data or attribute data. Spatial data tells us where something occurs. Attribute data tells what occurs; it tells us the nature or characteristics of the spatial data.

- **Hardware.** Computer hardware must be able to support data input, output, storage, retrieval, display, and analysis.

- **Software.** Software used should be dynamic and have wide variety of functional capabilities.

- **Users.** The term "user" may refer to any individual who will use NIS to support project or program goals, or to an entire organization that will employ.

2.9.1 Compilation Process

Data compilation involves assembling all of the spatial and attribute data in NIS. Map data with common projections, scales, and coordinate systems must be pooled together in order to establish the centralized NIS database. Data must also be examined for compatibility in terms of content and time of data collection. Ultimately, the data will be stored in NIS according to the specific format requirements set by both the user and the chosen NIS software/hardware environment.

When all of the common data requirements are set by the user, a "base map" has been established. A base map is a set of standard requirements for data. It provides accurate standards for geographic control, and also defines a model or template that is used to shape all data into a compatible form. A base map is not necessarily a map rather, it is a comprehensive set of standards established and enacted to ensure quality control for the spatial and attribute data contained in the NIS.

Once the data are assembled and base map parameters are set the user must translate manuscript data into computer-compatible form. This process referred to as "conversion" or "digitizing," converts paper maps into numerical digits that can be stored in the computer. Digitizing can be performed using various techniques. Scanning is one technique. Another technique is line digitizing which uses a tablet and a tracing stylus. Digitizing simplifies map data into sets of points, lines, or cells that can be stored in the NIS computer. Each NIS software package will impose a specific form and design on the way that these sets of points, lines, and cells are stored as digital map files.

Following figure shows the various types of compilation processes.

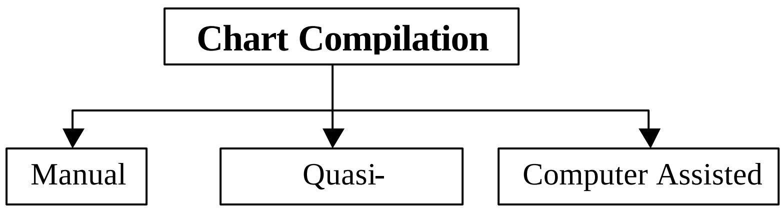


Fig. 1.5 "Chart Compilation Processes "

2.9.1.1 Manual: Traditional cartographic work, is based on colour separation and manual processes. The fair drawing is a manual method of preparing linework, symbols and topology (names) in accordance with the chart specifications. High quality linework is achieved by a process called "scribing" where the image is engraved on a coated film ensuring that cartographic specifications are carefully adhered to. Each colour used on a map is scribed on a separate film colour plate and symbols and names are combined photo-mechanically to produce colour separates for printing. Traditional cartography, defined as the manual techniques used for the production of a paper chart (before the advent of the computer), can be split in six components[19]:

- **Compilation Work.** The selection of the information collected for the production of the new edition of the paper chart.
- **Image generation.** The process of assigning symbol type, shape and structure to features on a map.
- **Image registration.** The technique to ensure that individual colour components fit each other in the map.
- **Contact copying at scale.** The operation used to produce same-size line, half-tone and continuous-tone positives and negatives by a direct contact process.
- **Image separation/combination.** The techniques used to produce multicolour maps by the sequential overprinting of a number of separate colour components.
- **Printing.** The charts are printed using offset lithographic process.

Quasi – automatic: Quasi-automatic cartography is the combination of manual and computer assisted techniques used for the production of a paper chart. Although, it contains the steps involved in manual but some of them are done by automatic means e.g. contouring is done by drafting machines instead of hands.

2.9.1.2 Computer Assisted Cartography: To improve services and to meet the growing demands for chart, computer assisted mapping systems are also used. The introduction of computer assisted mapping and geographic information systems have added new dimensions to cartographic techniques and usage of spatial data. The computer assisted cartographic steps are generally divided into the following six steps[20]:

- **Acquisition and Input.** Digital data is usually obtained from various sources e.g. digital files or scanning of old charts.
- **Verification.** All incoming data is verified for various and checked for formats, scale and feature coding etc.
- **Editing and attributing.** Main tasks involve ensuring features are topologically correct, attributed and symbolized according to Cartographic Digital Standards (CDS). Original manuscripts that were scanned require geo-referencing, and interactive editing and feature coding. All text or annotation on the map is added interactively.

---

[19] Lt Cdr Luis Pais, "Production and Distribution of ENC – The Portuguese Experience", Potugal (IHPT), hidrografia@hidrografico.pt, www.thsoa.org/pdf/h01/7_3.pdf.

[20] Vic Dohar and Dave Everett, "Geological Map Production for Dummies", Natural Resources Ottawa, Canada, vdohar@nrcan.gc.ca,http://pubs.usgs.gov/of/of00-325/dohar.html.

- **Quality Control.** A filtering process is used to create a report document that checks the project for completeness and correct feature attribution. Quality control officers ensure that the chart meets design specifications and that the digital data conforms to CDS. All maps are reviewed by the cartographers prior to publication.
- **Printing.** A final file is created for printing. Modern offset printing process may be single colour machines or may print multi colours in sequence.

2.9.2 Presentation

The real world is far too complex for a complete description to be practical, therefore a simplified, highly-specific, view of real world must be used. This is achieved by modelling the reality. The presentation of hydrographic information may vary to suit a particular use (e.g. it may be presented either graphically, using symbols or in a textual form). Therefore, the presentation of information should be independent of its storage. The concept of keeping information storage independent of presentation provides greater versatility and flexibility. It allows the same data to be used for many purposes without requiring any change to its structure or content. If the presentation style or medium changes, only the presentation model has to be changed. Therefore, the model described can be linked to many different presentation models. For example, ENC and paper charts present the same basic data in different ways via different presentation models[21].

2.9.2.1 Paper Charts. A Nautical Chart is a graphic portrayal that shows the nature and form of the coast, the depths of the water and general character and configuration of the sea bottom, locations of dangers to navigation, the rise and fall of the tides, locations of man-made aids to navigation, and the characteristics of the Earth's magnetism[22]. In addition to its basic elements, a chart is a working document used by the mariner both as a "road map" and worksheet and is essential for safe navigation. In conjunction with supplemental navigational aids, it is used to lay out courses and navigate ships by the shortest and most economical safe route.

Printed charts present all important information such as chart features with appropriate symbology and descriptive cartographic information texts and symbols. The volume of information is limited due to the size of the chart as well as the readability aspects of it. One of the most important aspects of the preparation work of the data to be published on the printed chart is cartographic generalization and cartographic editing of the data. These include e.g. displacement, aggregation, selection, rotation and text width, font and placement.

2.9.2.2 Digital Charts. Digital charts means a standardized database, as to content, structure and format, as shown in Fig 1.6.

---

[21] International Hydrographic Organisation, Monaco, Specifications for Chart Content and Display Aspects of ECDIS (S-52).

[22] http://chartmaker.ncd.noaa.gov/ncd/whatis.html.

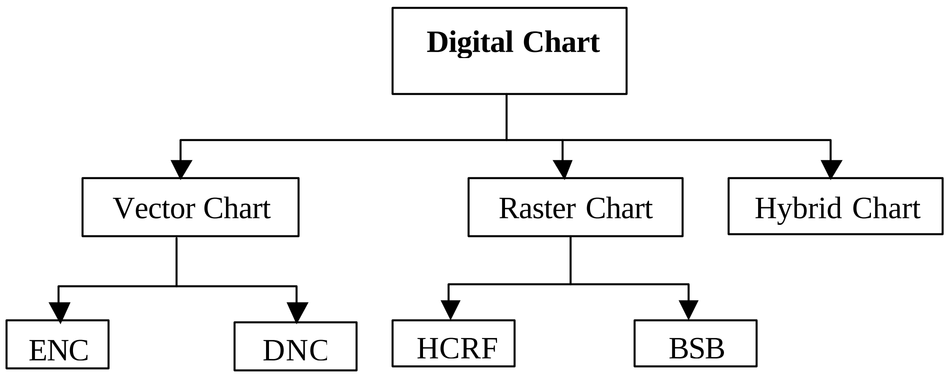


Fig. 1.6 "Digital Chart Types and Formats"

Digital charts[23] are a new navigation aid that can provide significant benefits to maritime navigation, safety, and commerce. More than simply a computer graphics display, digital chart systems combine both geographic and textual data into a readily useful operational tool. As an automated decision aid capable of continuously determining a vessel's position in relation to land, charted objects, aids to navigation, and unseen hazards, ENC are a real-time navigation system that integrates a variety of information that is displayed and interpreted by the Mariner. The most advanced form of digital chart systems represents an entirely new approach to maritime navigation.

2.9.2.3 Vector Charts

ENC: An Electronic Navigational Chart (ENC) is vector data conforming to IHO S-57 ENC product specification in terms of contents, structure and format. Issued for use with ECDIS on the authority of government authorized hydrographic offices, an ENC contains all the chart information necessary for safe navigation and may contain supplementary information in addition to that contained in the paper chart (e.g., sailing directions). In general, an S-57 ENC is an object-oriented, structurally layered data set designed for a range of hydrographic applications. As defined in IHO S-57 Edition 3, the data is comprised of a series of points, lines, features, and objects. The minimum size of a data is a “cell” which is a spherical rectangle (i.e. bordered by meridians and latitude circles). Adjacent cells do not overlap. The scale of the data contained in the cell is dependent upon the navigational purpose (e.g. general, coastal, approach, harbour). Other than a 5 Mb size limit to the amount of digital data contained in an ENC cell, there are no specifications regarding the dimensions of a cell as the smallest packaging.

DNC: The Digital Nautical Chart (DNC™) is a vector database of selected maritime features that can be used with shipboard integrated navigation systems (e.g. electronic chart systems), or other types of geographic information systems (GIS). Similar to IHO S-57 ENCs, the DNC database consists of points, lines and polygons that contain information regarding hydrography, aid-to-navigation, cultural land marks, land features, depths, obstructions, etc. Each theme (e.g. hydrography) is stored as thematic layer with geo-referenced properties. The DNC product is encapsulated using the Digest Annex C Vector

[23] Lee Alexander, Ph.D., “What is an ENC? It depends who you ask”, Offshore Systems Ltd., Vancouver, BC, Canada, http://www.osl.com/Support/what_is_enc.htm.

Relational Form (VRF) of encoding which consists of a set of relational database tables. Further the data is organised into series of “libraries” which are groupings of chart coverage which are commensurate with NIMA’s groupings of paper nautical charts scales (e.g. General, Coastal, Approach, and Harbour). In the ECDIS concept a DNC is a “system” electronic navigational chart (SENC) that contains specified data and display characteristics.

2.9.2.4 Raster Charts

Raster data formats are bitmaps with a geo-reference applied to them. A bitmap is a generic term for a computer image made up of a rectangular grid of very small (254 per inch is one standard) coloured squares or pixels. These bitmaps are usually generated by taking the original chart and scanning them to create a digital picture of the chart. Once this image has been acquired, a geo-reference is applied. This is the process of relating the grid positions of the bitmap pixels to their corresponding latitude and longitudes. In this way, a computer can relate pixel position to latitude and longitude. However, the system has no knowledge of the details of the features and details (such as the coast line) in the raster images it displays. Raster charts are produced by scanning at high resolution the original colour separates, which are used to print the paper charts. The digital files are carefully georeferenced to enable navigation software to map geographic positions to locations in the image. Metadata is added describing the chart, its datum, projection and other information about the chart and the digital file.

Hydrographic Chart Raster Format. (HCRF): This is the format developed by UKHO and used for its Admiralty Raster Chart Service (ARCS) and the Australian HO for its Seafarer Chart Service. Raster charts have the same standards of accuracy and reliability as paper charts. These are used with authorised compatible Electronic Charting Systems (ECS).

BSB Format: The (BSB) format is basically one or more raster images compressed in to an efficient package that is accompanied by the chart details within the package. These chart details include the geo referencing required for determining latitude and longitude as well as other particulars such as, scale, depth units, chart name, etc. The BSB format separates a chart into images depending on the number of compartments a chart contains where a 'compartment' is defined as main chart, chart inset, and chart continuation

2.9.2.5 Hybrid Charts

Ideally the master versions of all discrete digital cartographic product data would be held in vector form. The rise in capacity of computer systems over recent years has opened up the alternative of transferring quickly to digital methods by raster scanning the existing printing separates, and then using hybrid raster/vector techniques during a changeover period. Raster masters are replaced by vector masters in a sequence determined by costs and business priorities.

ANNEX A

ACRONYMS

ARCS	Admiralty Raster Chart Service
AHO	Australian Hydrographic Office
CAD	Computer aided drafting
CD	Compact Disk
CDS	Cartographic Digital Standards
DNC	Digital Nautical Chart
ECS	Electronic Charting System
ENC	Electronic Navigation Chart
ECDIS	Electronic Chart Display and Information System
GIS	Geographic Information System
HCRF	Hydrographic Chart Raster Format
HTF	Hydrographic Transfer Format
HI	Project Instruction/Hydrographic Instruction
IHO	International Hydrographic Organisation
IMO	International Maritime Organisation
NHO	National Hydrographic Office
NIMA	National Imagery and Mapping Agency
NIS	Nautical Information System
NTM	Notices to Mariners
QC	Quality Control
RNC	Raster Nautical Chart
SENC	System Electronic Navigational Chart
UKHO	United Kingdom Hydrographic Office
VRF	Vector Relational Form
WGS	World Geodetic System

REFERENCES

D. Russom & H. R. W. Halliwell	*"Some Basic Principles in the Compilation of Nautical Charts"*	International Hydrographic Review, Vol. LV No. 2, July 1978
William G. Swisher	*"National Ocean Survey Automated Information System"*	International Hydrographic Review, Vol. LVIII No. 2, July 1981
N. M. Anderson	*"Computer Assisted Cartography in the Canadian Hydrographic Service"*	International Hydrographic Review, Vol. LVIII No. 2, July 1981
Christer Palm	*"From Manuscript to Printed Chart"*	International Hydrographic Review, Vol. LX No. 2, July 1982
Martin Joseph	*"Assessing the Precision of Depth Data"*	International Hydrographic Review, Vol. LXVIII No. 2, July 1991
E. C. Bouwmeester and A. W. Heemink	*"Optimal Line Spacing in Hydrographic Survey"*	International Hydrographic Review, Vol. LXX No. 1, March 1993
H. Gorziglia	*"Computer Assisted Cartography at Hydrographic and Oceanographic Service of Chilean Navy"*	International Hydrographic Review, Vol. LXX N°, September 1993
Adam J. Kerr	*"Conceptual Model of a Regionally Integrated Data Base For ECDIS"*	International Hydrographic Review, Vol. LXXI No. 2, September 1994
Udo Laube	*"The Hydrographic and Wrecksearch Vessel "DENEB" as an Example for a Modern Survey Vessel"*	International Hydrographic Review, Vol. LXXII No. 2, March 1995
B. Bourgeois, F. Petry, M. Harris & P. Alleman	*"A GIS Integration Approach for Dynamically Reconfigurable Surveys"*	The Hydrographic Journal, January 1999.
Patrick McGlamery	*"Issues of Authenticity of Spatial Data"*	66th IFLA Council and General Conference 2000.
Neal G. Millet and Simon Evans	*"Hydrographic Management using GIS Technologies"*	U.S. Hydrographic Conference 2001.
Lieutenant Eric J. Sipos	*"NOAA AHB Assurance Inspections for Contract Hydrographic Surveys"*	U.S. Hydrographic Conference 2001.
Lt Cdr Luis Pais	*"Production and Distribution of ENC – The Portuguese experience"*	U.S. Hydrographic Conference 2001.

URL ADDRESSES

1.	Ames Remote	http://www.amesremote.com
2.	Caris Marine	http://www.caris.com
3.	Coastal Oceanographic, Inc	http://www.coastalo.comNational
4.	ESRI Software	http://www.esri.com/library/
5.	Federal Geographic Data Committee	http://fgdc.er.usgs.gov/fgdc.html
6.	Hydrographic Society of America	http://www.thsoa.org
7.	Imagery and Mapping Agency (NIMA)	http://www.nima.mil
8.	International Hydrographic Organisation	http://www.iho.shom.fr/iho.html
9.	JANUS Technologies	http://www.janus-tech.com
10.	NOAA National Coast Survey	http://chartmaker.ncd.noaa.gov
11.	Offshore Systems Ltd.	http://www.osl.com/corporate
12.	Primar Organisation	http://www.primar.org
12.	SeaBeam Instruments	http://www.seabeam.com/
13.	Reson, Inc	http://www.reson.com
14.	The Laser-scan Ltd.,	http://www.Laser-Scan.com/papers
15.	The GIS Primer	http://www.innovativesgis.com

BIBLIOGRAPHY

Admiralty	*"Manual of Hydrography"*	Vol I & Vol II
Bowditch	*"American Practical Navigator"*	U.S. Navy Hydrographic Office.
Admiralty	*"General Instructions for Hydrographic Surveyors"*	
Karl B. Jeffers	*"Hydrographic Manual"*	U.S. Department of Commerce
IHO Special Publication S-44	*"IHO Standards for Hydrographic Surveys"*	
IHO Special Publication S-52	*"Spécifications for Chart Content and Display Aspects of ECDIS"*	
IHO Special Publication S-57	*"IHO Transfer Standards for Digital Hydrographic Data"*	
IHO Special Publication S-61	*"Product Specification for Raster Navigational Charts (RNC)"*	
IHO Circular Letter 45/2001	*"Guidelines for the Processing of High Volume Bathymetric Data"*	Dated 5 Oct 2001.
	"Map Compilation Color Separation and revision"	Headquarters Department of Army, Washington.
Colonnel Sir Charles Close & Colonel H. St. J. L. Winterbotham	*"Text Book of Topographical and Geographical Surveying"*	Her Majesty's Stationery Office.

CHAPTER 2
POSITIONING

By Cdr. Lamberto LAMBERTI, Lt. Antonio DI LIETO (Italy) & Lt Cdr. Paul LAWRENCE RN

1. INTRODUCTION

Determination of position with relative reliability is the fundamental problem facing the reference frame of a Geographic Information System (GIS) and the principal purpose of the science of geodesy.

Determination of position for points on the earth's surface requires the establishment of appropriate co-ordinates in the selected geodetic reference system (DATUM).

The minimum information output, when the tool 'co-ordinates' is selected by the user, should be:

- The parameters that fully describe the reference system;
- The required co-ordinate details for the selected cartographic symbol or point.

In this way it is possible to unambiguously define the co-ordinates of a point or object with reference to the real world.

2. PRINCIPLES OF POSITIONING

2.1 The Earth

Calculation of position with repeatable accuracy is the central problem for the geographical reference of terrestrial information and the principal function of geodesy.

The geographical position of a point on the earth's surface can be defined in relation to a mathematically defined reference surface which is used in place of the true surface of the Earth (very close to an ellipsoid of rotation or bi-axial).

Reference surfaces should have two fundamental characteristics:

- mathematically defined;
- closely fitting the true surface in the desired location.

Reference surfaces used for limited areas are very often:

- the ellipsoid of rotation (or bi-axial);
- the local spheroid;
- the horizontal plane (or tangent plane);
- the geoid.

The first three have purely arithmetical definitions and they are used for horizontal positioning; the fourth surface has a physical definition and has a relationship with the others for height/separation value. A three-dimensional position is defined by 2 horizontal co-ordinates and a vertical component which is the height above the reference surface.

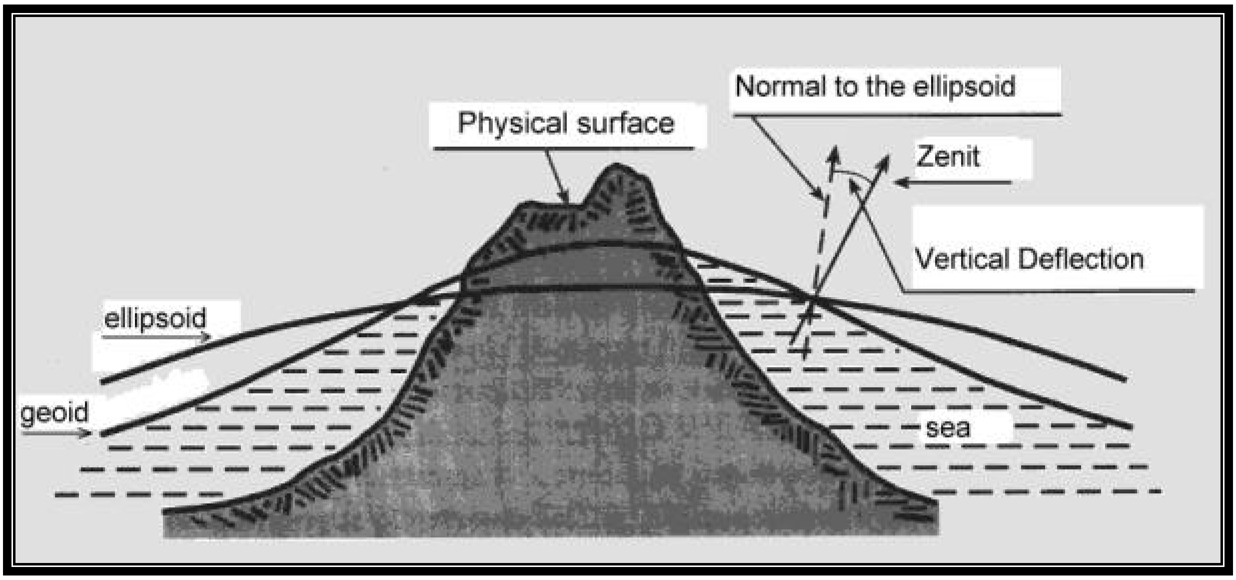


Fig. 2.1 "The Earth"

2.1.1 The Ellipsoid

The ellipsoid is a fourth order surface on which all curves of intersection with a plane are ellipses, which eventually degenerate into circles. For any selected point on the ellipsoid surface and for the normal to the tangent plan at this point, the ellipses produced by the intersection with such a surface and the normal form endless continuing planes, they are known as normal sections and have, at that point, a number of varying bending radii. This variation is a continuous function of the ellipsoid latitude of the selected point, of the ellipsoid shape parameters and the azimuth of the produced normal section. The two normal sections, which correspond to the minimum and maximum curving radii, are defined as the principal normal sections.

For geodetic purposes the ellipsoid of revolution, produced when an ellipse is rotated about its semi-minor axis, provides a well defined mathematical surface whose shape and size are defined by two parameters: lengths of *semi-minor axis* (***b***) and *semi-major axis* (***a***). The shape of a reference ellipsoid can also be described by either its *flattening:* $f = [(a - b) / a]$ or its *eccentricity:* $e = [(a^2 - b^2)^{1/2} / a]$.

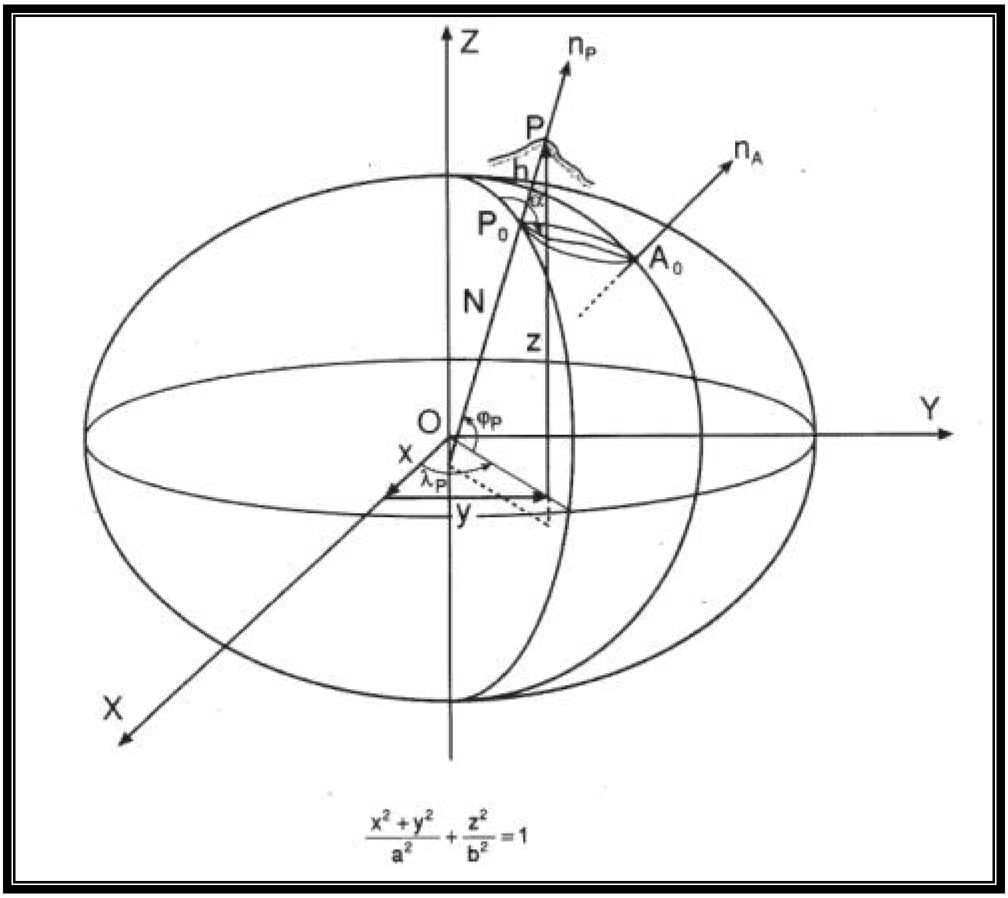


Fig. 2.2 "The Ellipsoid"

The ellipsoid surface is regular and derived mathematically; it is for these reasons that, as a reference surface, it is the widely used for horizontal co-ordinate systems. However it is of limited use as a reference for heighting as it is such a coarse approximation of the earth's shape.

2.1.2 The local Sphere

A local sphere is a reference surface which, at a selected latitude, has a radius equal to the geometric mean between the curving radii of the two principal normal sections of the ellipsoid being replaced at the point of interest on the surface.

Substitution is acceptable within a radius of approximately 100 km (in Geodetic Field) from the point of tangency between ellipsoid and sphere, it involves shifts in distances and angles of less than the sensitivity of the best survey tools (distances: 1cm +/- 1ppm; angles: 0.1").

Within a radius of 15 km (in Topographic Field) from the same point, it is acceptable to replace the sphere with a tangent plan, causing a shift in comparison with the ellipsoid surface less than the above stated accuracies.

2.1.3 The Geoid

The Geoid, defined as the equipotential surface of gravity strength field, is used as a reference surface for heights; Mean Sea Level (MSL) is the best approximation of such a surface. The physical meaning of gravity equipotential surfaces may easily be checked, as every point should be orthogonal to the direction indicated by a plumb line.

Unlike the ellipsoid, the Geoid can not be mathematically created or utilized in calculations because its shape depends on the irregular distribution of the mass inside the Earth.

2.2 Datum

A Datum is a Geodetic Reference System defined by the reference surface precisely positioned and held in the space; it is generated by a compensated net of points.

The SP-32 (IHO – Fifth Edition 1994) defines a geodetic Datum as *"a set of parameters specifying the reference surface or the reference co-ordinate system used for geodetic control in the calculation of co-ordinates for points on the Earth; commonly datums are defined separately as horizontal and vertical"*.

The determination of a unique reference surface for the whole Earth, essential for the use of satellite systems and associated survey and positioning techniques, has in the past been of little interest and difficult to achieve, due to the essentially local character of geodetic and topographic survey techniques. For this reason, there are many local geodetic systems worldwide, all defined with the sole purpose of obtaining a good approximation only for the area of interest.

Furthermore, for each nation it is normal to find two reference surfaces defined in different ways, because there has been a clear separation between the determination of positions for the horizontal and the vertical:

a. Local Ellipsoid;
b. Local Geoid.

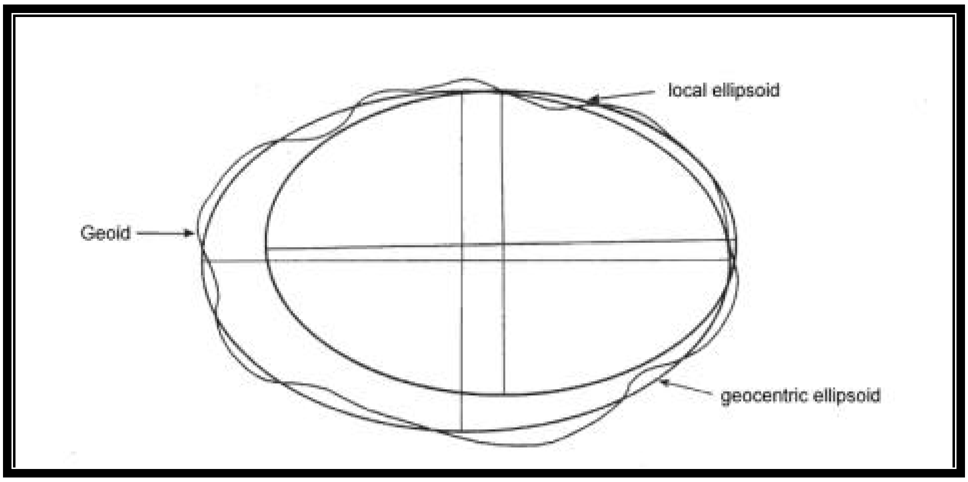


Fig. 2.3 "Datum orientation"

2.2.1 Horizontal Datum

Horizontal Datum is a mathematical model of the Earth which is used for calculating the geographical co-ordinates of points. A reference bi-axial ellipsoid in association with a local system is a horizontal geodetic reference system (that is bi-dimensional). It is defined from a set of 8 parameters: 2 for shape of the ellipsoid and 6 for position and orientation. Such a reference system is not geocentric, that is the ellipsoid centre is shifted from Earth's centre of mass by a quantity of about 100 metres; additionally the ellipsoid axes symmetry is not aligned to the mean terrestrial rotational axes, although angular shifts are very small, an order similar in quantity for the accuracy of the more sophisticated angular measurement capabilities.

The local ellipsoid must be positioned and orientated with regard to the Earth to enable translation from the measured geometric quantity (distances, angles, difference in elevations) to the calculation for the relative position associated with a point of known ellipsoid co-ordinates, conventionally selected in relation to local requirements. With satellite technology developments, it is now possible to directly obtain co-ordinates in comparison with a geocentric system which require no modification by the user and can be used internationally. In the past, when geocentric positioning was not possible, the only way for positioning and to directly reference systems was to establish an initial point (point of origin) and a connection with the local astronomic system (defined by the local vertical and by the terrestrial axis of rotation).

There are two parameters for shape which identify an ellipsoid, the other six (6 degrees of freedom of a rigid body in the space) which must be determined in the initial point, are:

a. ellipsoid or geodetic latitude;
b. ellipsoid or geodetic longitude;
c. geoid elevation (or orthometric height);
d. two components for the vertical deviation;
e. ellipsoid azimuth for a direction that has the origin as its point.

The policy continues that to connect the two fundamental surfaces, ellipsoid and geoid, selecting the point of origin for a known geodetic height, has to have an astronomically determined latitude and longitude. You therefore force the ellipsoid co-ordinates of the point of origin to coincide with the astronomical or celestial co-ordinates.

This condition produces two fundamental effects:

a. It binds a preset point on the ellipsoid to a direction in the space (eliminating two degrees of freedom);
b. It makes sure that the point is coincidence with the ellipsoid normal axis and with the geoid vertical axis (a further two fixed degrees of freedom removed).

Ascribing the point of origin ellipsoid height to be coincident with known geodetic height and aligning the ellipsoid rotational axis in the direction of the astronomical North, it is possible to fix the remaining two degrees of freedom of the ellipsoid relative to the geoid:

a. Sliding along the normal/vertical;
b. Rotating around to it.

On completion of these operations, the local ellipsoid of reference is focussed on the point of origin.

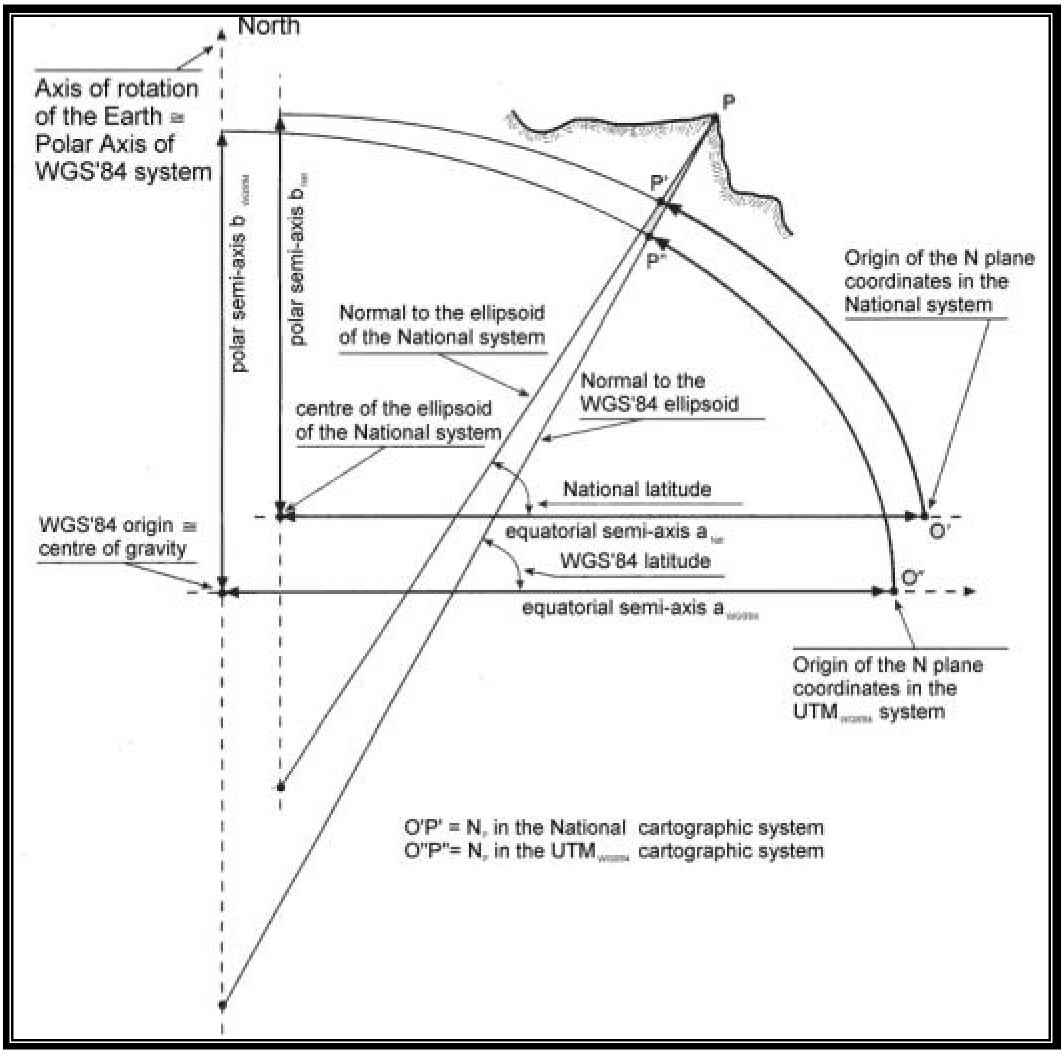


Fig. 2.4 "Horizontal Datum orientation"

2.2.2 Type of datum

Local geodetic systems employed in geodesy and cartography before the advent of satellite systems were based, as previously stated, on ellipsoids which approximately fitted the local geoid surface.

In this way, in operational applications, adjustments between the vertical and ellipsoidal normal are reduced and almost negligible, angular measurements on the ground can be quoted for ellipsoidal figures without corrections. This situation can be considered valid in cases for smaller nations with a limited surface area; it can also be acceptable, but with a degraded approximation, for wider zones, such as the whole of Europe or the United States.

The demand for wider reference systems has grown during recent decades in concomitance with the general globalization process.

For the past 50 years, it was recognised that there was a need to find a unique reference system for the whole globe, on which to present cartographic, geodetic and gravimetric products. The advent of satellite geodesy has made the adoption of single geocentric references essential and advanced the need to create a good middle approximation for every part of the globe.

The first systems with these characteristics were developed by the Department of the Defence of the United States: WGS60, WGS66 and WGS72 were increasingly reliable models of terrestrial physical reality, the culmination being the creation of WGS84.

WGS84 is the acronym for 'World Geodetic System 1984' and it defines the system as geodetic and universal in 1984. It is represent by an OXYZ Cartesian system with the origin at the centre of the Earth's conventional mass and Z axis directed towards the conventional earth North Pole (CTP. Conventional Terrestrial Pole), as defined by BIH (Bureau International Le Heure) in 1984, today named IERS (International Earth Rotation System). The X axis is at the intersection of the origin meridian plan passing through Greenwich, defined by IERS in 1984, and the CTP referred to the equatorial plane. The Y axis completes a clockwise orthogonal rotation and lies on the equatorial plane at 90° east to the X axis. The Cartesian terms match the Earth. The co-ordinate origin and axes are also at the centre of mass and the axes of the ellipsoid are coincident with the system (ellipsoid bi-axis, geocentric WGS84), the Z axis is the axis of symmetry.

EUREF, the IAG (International Association of Geodesy) sub-commission, which is responsible for the European Terrestrial Reference System realisation (ETRS), approved the European Terrestrial Reference Frame (ETRF) in 1989. The ETRF89 system is a realisation of the WGS84 system.

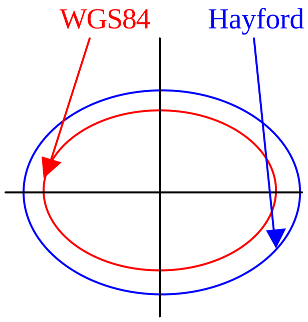
2.2.3 Datum transformation

With the development of a unique model, it became possible for all charts to be on only one reference system; however the transformation of charts from one datum to another is not a simple operation. For this reason many charts in circulation today are still referred to old systems.

Cartesian co-ordinates referred to a geocentric system or geographical co-ordinates referred to a geocentric ellipsoid are generated via satellite positioning techniques. To transform these co-ordinates into the local system related to the operational area, it is necessary to apply algorithms with parameters determined by means of probability computations in order to adapt the very precise satellite measured results to the net realised by the local system with its inevitable deformations.

Every ellipsoid, which is locally oriented, inevitably shifts from the geocentric one adopted in the WGS84 system, not only due to the different parameters but, importantly, also for centre position and axis orientation. Therefore, the geographical co-ordinates for the same point in the local datum and in the global one are different, the shifts translated into distances can be of hundreds metres.

The diagram shows the dimensional differences between the ellipsoid of Hayford and the corresponding WGS84:



System	Equatorial Semi-Axis [m]	Polar Semi-Axis [m]
WGS84	6 378 137	6 356 752.31
ED50	6 378 388	6 356 911.95

Note that the WGS84 ellipsoid is smaller both in the equatorial and polar dimensions.

The variations in dimension and origin are reflected in geodetic latitude (or ellipsoidal) and in the North horizontal co-ordinates (Gaussian) of a point on the earth's surface; the same occurs with ellipsoidal longitude and East co-ordinates.

The comparison with the geographical co-ordinates risks creating considerable confusion in the evaluation of horizontal co-ordinates definable by the adoption of the Gauss (UTM) representation. In fact, shifts in Gaussian co-ordinates are not the same as linear value shifts in ellipsoidal co-ordinates. This is because the length of the arc subtended by a degree of latitude or longitude depends on the dimension of the ellipsoid and also because it changes the point of origin. It is therefore vital to provide users with complete information and the necessary training to understand the problems.

To transform geographical and horizontal co-ordinates from one system to another it is necessary to apply to every point some variation in $\Delta\varphi$, $\Delta\lambda$, ΔN, ΔE, which are functions of the point; the shifts to be applied to every point alter with the position.

The transformation between two different local datums, in a same area, is often performed using empirical methods, based on the fact that the two reference surfaces, even though different, are very similar and the principal difference is one of orientation. In the case of the transformation between a global geocentric system, such as the WGS84, and a local geodetic system, the two surfaces are separated from each other and it is therefore necessary to apply more general algorithms of transformation.

Datum transformation has assumed considerable importance with the advent of GPS; in practice it is usually necessary that a GPS survey includes some points from the old geodetic system in which the survey must be structured; it is thus possible to calculate suitable transformation parameters which are valid for the immediate area of interest.

The simplest and most commonly used method consists of assuming the existence of a rotation and translation of the axes with a scale factor in the Cartesian systems connected to the aforementioned ellipsoids:

$$\begin{vmatrix} X_2 \\ Y_2 \\ Z_2 \end{vmatrix} = \begin{vmatrix} X_0 \\ Y_0 \\ Z_0 \end{vmatrix} + (1+K) \cdot \begin{vmatrix} 1 & E_Z & E_Y \\ E_Z & 1 & E_X \\ E_Y & E_X & 1 \end{vmatrix} \cdot \begin{vmatrix} X_1 \\ Y_1 \\ Z_1 \end{vmatrix} \qquad \textbf{(2.1)}$$

Where:

(X_1 Y_1 Z_1)	Cartesian co-ordinates of a point in the first system (S1);
(X_2 Y_2 Z_2)	Cartesian co-ordinates of the same point in the second system (S2);
(X_0 Y_0 Z_0)	co-ordinates, in S2, of S1 origin;
(1 + K)	scale factor;
(E_x, E_y, E_z)	rotations around S1 axes (expressed in radians and acting in anti-clockwise sense).

Such a model implies the perfect geometric congruence, except for scale factor, between all the points of the geodetic network, determined with GPS methods (for example in S2) and the same points, determined with the traditional techniques of triangulation and trilateration in S1. Naturally, this is not always the case in reality, mainly due to distortions induced in the classical geodetic networks from the propagation of errors which inevitably characterise the traditional procedures of measurement. The relationship (2.1) holds in most cases if it is applied within the limited extensions of the networks.

If together with it (2.1) the following formulae are used:

$$\begin{cases} X = (N+h) \cdot \cos f \cdot \cos ? \\ Y = (N+h) \cdot \cos f \cdot \text{sen}\, ? \\ Z = \left[(1-a)^2 \cdot N + h\right] \cdot \text{sen}\, f \end{cases} \quad \text{with} \quad N = \frac{a}{\sqrt{\cos^2 f + (1-a)^2 \cdot \text{sen} f}} \qquad \textbf{(2.2)}$$

They connect the geodetic co-ordinates j , l , & **h** related to the ellipsoid with semi-axis '**a**' and ellipticity (or compression or flattening) a , with the co-ordinates X, Y & Z related to the geocentric associate Cartesian system, the transformation formulae between the different systems are produced in geodetic co-ordinates.

The seven parameters, knowledge of which are necessary to apply (2.1), can be determined, in a local system, as the solution of a least squares adjustment, in which the observed quantities are the co-ordinates (Cartesian or geodetic) of a certain number (≥ 3) of points in the network, obtained via both GPS observations in S2 and classical terrestrial methods in S1.

2.2.4 Vertical datum

The first element necessary for the definition of height is the reference surface.

Once this is established, the orthogonal direction necessary for the measurement of elevation is specified, while the scale along that direction evolves from the reference system realisation.

As a result of the way these elements are selected, different systems of heights can be defined:

a. 'h' Ellipsoidal height: adopting as the reference surface a bi-axial ellipsoid;
b. 'H' Orthometric height (or elevation above Geoid surface): choosing as the reference an equipotential surface of gravity strength field, approximate to the MSL isolated from the periodic oscillations and shielded from the a-periodic ones (Geoid).

The second system enables the preservation of the physical meaning of height on the MSL. However, mathematical complications arise when determining the difference between the two surfaces (ellipsoid – geoid), known as geoidal undulation, knowledge of which is necessary to connect the two systems of heights.

The following figure shows the main relationship between ellipsoidal height **h** and orthometric **H**.

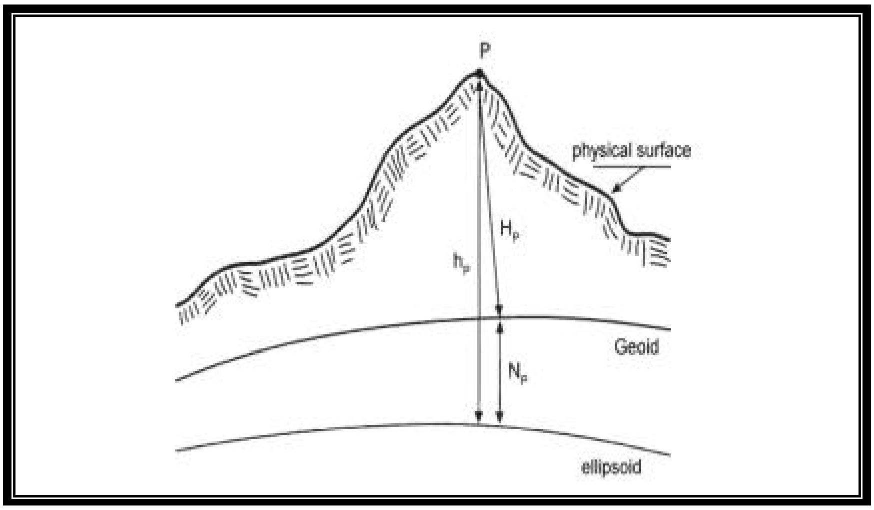


Fig. 2.5 "Vertical Datum"

In a first approximation, to within a few millimetres:

$$h_p = H_p + N_p \tag{2.3}$$

$\mathbf{h_p}$ is measurable with the GPS, while $\mathbf{H_p}$ is observable with levelling operations corrected for gravimetric observations; $\mathbf{N_p}$ (geoid height or undulation) is the elevation above the surface of the projected point P on the Geoid along the geoid vertical (plumb line). This formula is an approximation because it does not consider length differences between the normals or different scale factors which result from the different kinds of observations. For cartographic purposes the error produced by this approximation can normally be ignored.

For the traditional altimetry in cartographic work, MSL is conventionally assigned zero elevation (or level), since the surface of the sea is available from almost every where. The MSL is sufficiently determined from tide gauge observations over a long period to filter it from the short term affects of tide.

The SP-32 (IHO – Fifth Edition 1994) defines MSL as *"the average HEIGHT of the surface of the SEA at a TIDE STATION for all stage of the TIDE over a 19-year period, usually determined from hourly height readings measured from a fixed predetermined reference level (CHART DATUM)"*.

For a specific area of interest, the local Geoid is defined as the equipotential surface of the gravity strength field for a pre-determined point within the same area (usually a point on or near the coastline located at the conventionally defined MSL height).

Starting from this point, assumed as the fundamental zero level reference, using techniques of geometric levelling, it is possible to attribute a geoidal height to each point, known as benchmarks, in a network which extends over on the whole area, the elevation reference frame.

Depending on where we stand, the MSL can be nearer or farther away from the Earth's centre in comparison with another point; the elevations do not benefit from a global definition of the mean level of the sea and presently there is no global elevation reference system which enables unification or the direct comparison to be achieved between heights measured in various elevation systems.

2.3 Co-ordinates systems

The position is usually defined through curvilinear co-ordinates such as the latitude, the longitude and the height above the reference surface. In this case it is providing a position in (2+1) dimensions.

It is correct and necessary to distinguish between the following co-ordinate systems:

a. Plane rectangular (Grid);
b. Spherical;
c. Ellipsoidal (Geodetic);
d. Geoidal (Astronomical).

according to whether the plane, the sphere, the ellipsoid or the geoid is used as the surfaces of reference.

The ellipsoidal co-ordinates are also termed geodetic, while the geoidal co-ordinates are the astronomical ones.

According to this interpretation, the term 'geographical co-ordinate' is a general term which includes the types mentioned in c and d.

2.4 Principles of cartography

The representation of the ellipsoid on a plane (horizontal) surface is the fundamental problem and objective of Cartography.

Such a problem is made more complex by the ellipsoidal surface not being developable (or of the spherical surface in narrower field) on a plane surface. Thus it is not possible to transfer details from a three dimensional reference surface to a paper plan without the parameters which describe them (distances, areas, angles) suffering considerable deformations. Finding the best method of achieving this transfer will be focussed, therefore, towards the removal of some of them, or towards containing them within acceptable limits.

According to the selected method there are:

a. Charts in which distances are preserved (equidistant charts): this condition cannot be achieved for the whole paper, only along particular directions. It means along certain lines that the relationship (scale) is preserved between the measured distances on the paper and the measured distances on the reference surface;

b. Charts in which the areas are preserved (equivalent or equal area charts): this condition can be achieved over the whole paper. It means that the relationship is preserved between a measured area on the paper and a measured area on the reference surface. Linear and angular deformations are introduced, however, which create alterations of shape;
c. Charts in which the angles are preserved (conformal charts): this can also be achieved over the whole paper. It means that the measured angle between two geodetics transformed on the paper is equal to the angle between two corresponding directions on the reference surface (ellipsoid or sphere);
d. Charts in which the scale is the same in all directions at any point (orthomorphic charts): angles round a point are preserved and small shapes are not distorte d over the entire paper;
e. Charts in which none of the element above is rigorously preserved but where the relative deformations are contained within suitable tolerances (aphilatic or not orthomorphic charts).

Three indices allow the evaluation of the deformation entity, and therefore to calculate relative corrections. They are termed ‘forms of linear, superficial and angular deformation’ and they are respectively given from:

$$
\begin{aligned}
\mathrm{m}_{\mathrm{l}} &= \mathrm{dl}'/\mathrm{dl} \\
\mathrm{m}_{\mathrm{s}} &= \mathrm{dS}'/\mathrm{dS} \\
\mathrm{m}_{a} &= a' - a
\end{aligned}
\qquad \textbf{(2.4)}
$$

where with **dl’**, **dS’**& a**’** being the geometric elements belonging to the paper and with **dl**, **dS** & a the corresponding elements for the ellipsoid (a **’-** a is the angle by which the element **ds** has to rotate to get itself to **ds'**). The linear and superficial elements must be infinitesimally small to be able to quickly identify the size of the deformations.

The choice of the cartographic system depends on the purpose for which the chart is being produced. If the chart is to be used for navigation, it will have to be conformal. The angles on the paper (for example the angles between the courses marked on the paper and the meridians) will replicate, without variations, the direction of the vector angle.

The procedure, through which a connection is established between the points of the ellipsoid and the points of the cartographic plane, can be:

a. Geometric: which consists of establishing a projective relationship between them through appropriate geometric constructions, followed by relative analytical processes (trigonometric in general);
b. Analytical: consists of establishing a non-projective analytical connection between the points. It is necessary to write a system of equations which links the geographical co-ordinates of the various points on the ellipsoid to the orthogonal plane co-ordinates on the sheet referred to an appropriate system of axes.

The first method of chart construction is named ‘projection’, the second ‘representation’. The two methods are not incompatible, each system can be articulated through an arrangement of equations and appropriate projective systems can correspond to various analytical systems, even if they are sometimes approximate.

In modern cartography it is preferable to build charts through "representations".

Mixed systems exist in which selected elements of the network are transformed with one system and other elements with another. Such systems are termed 'projections or modified representations', they are commonly used in chart construction due to the particular characteristics they confer on the end product, which would not have been created in a pure projection or representation.

2.5 Projections

2.5.1 Perspective (or geometric) projections

To reproduce an ellipsoid determined section of a chart, it is necessary to study the centre of the area and to find the tangent plane to the ellipsoid at that point. It is then possible to project the ellipsoid geometric figures on such a plane from a suitable centre of projection.

Depending on the selected position for the point of projection, various transformations are produced, each with particular characteristics.

The centre of projection can be set:

a. at the ellipsoid centre (centre graphic or azimuthal projection): the charts produced with this system are useful for navigation, because the transformation of the arcs of maximum curvature of the single local spheres produces segments of straight lines on the plane of projection;
b. in relation to the point diametrically opposite to the zone to be represented (stereographic projection): it is the only conforming perspective projection and it is generally used for polar zones cartography;
c. at the extension of the diameter, but external to the ellipsoid ('scenographic' projection);
d. always on the same diameter but at infinite distance (orthographic projection).

2.5.2 Conic projections

The conic projection consists in taking a conic surface positioned according to the portion of ellipsoid for which the paper has to be created and projecting the ellipsoid on the conic surface from the centre of the ellipsoid. Subsequently, the conic surface will be turned into a plane and the chart so produced will not be deformed (equidistant) along the line of tangent; elsewhere it is aphilatic or not orthomorphic. The most common case is represented by the 'direct conic projection', which, in order to make it conformal, Lambert has maintained unchanged the projection principle for tracing the meridians but he has replaced an analytical representation system for the projection method for tracing the parallels. This is an orthomorphic modified projection.

2.5.3 Cylindrical projections

Cylindrical projections are obtained by taking a cylindrical surface, variously prepared, tangent to the ellipsoid and projecting above it the points belonging to the ellipsoid, from its centre.

Among the numerous possibilities of position for the projection cylinder, we are going to consider two which originate, after the development on the plane, the two cartographic systems most used: the direct cylindrical projection and the inverse one.

2.5.3.1 Direct cylindrical projection

The projection cylinder is a tangent to the equator and it has a coincident axis with the terrestrial ellipsoid smaller axis. The meridian and parallel grid (graticule) transforms itself, from that cylinder, in a series of straight lines orthogonal between them. The projection is aphilatic or 'not orthomorphic' in an equatorial band; it is conformal only for parallels and meridians, deformations are small in proximity of the equator but they grow approaching the poles.

The direct cylindrical projection can be made conformal and orthomorphic introducing an analytical connection between the parallels on the ellipsoid and the parallels on the chart; it remains the projection origin of transformed meridians.

The modified chart obtained in this way, termed Chart of Mercator (or Mercator projection), has the advantage of being conforming and presenting geographical grids transformed as straight lines orthogonal between them. In summary, this appears to be the ideal cartographic system for the equatorial area. For areas at the mean latitudes, a cylindrical surface intersecting the ellipsoid can be considered: there will no longer be an absence of deformations on the equator, but there will be on the two selected parallels, reductions in the band between and expansions in the external zones.

Additionally, the Chart of Mercator allows the navigation using 'loxodrome or rhumb line'. Though not representing the shortest distance between two points, which is the geodesic or orthodrome, the loxodromes are followed for short distances, because the route angle can easily be equated to the mean; for this reason, such charts are of usually employed for navigation.

2.5.3.2 Transverse cylindrical projection

The projection cylinder is tangent to a meridian with axis placed above the equatorial plan and the ellipsoid surface is projected above it from the centre of the ellipsoid itself. Deformations do not take place on the meridian of tangency; but they increase with increasing distance from it.

Meridian and parallel grid (graticule) are transformed into a net of curves that intersect at the same angles. The affect of the deformation is limited by reducing the zone to be projected, achieved by dividing the terrestrial surface into zones of limited width (generally 15° of longitude), and by projecting them above a cylinder tangent to their central meridian, along which deformations are avoid. To reduce deformations further, intersection of the cylinder, rather than a tangent, can be introduced. In such a method, the absence of deformation does not occur on the central meridian, but on two intersecting lines which are symmetrical to it: in the area enclosed between these lines there are contractions, while outside these zones there are increasing expansions.

2.5.4 Representations

The Gauss representation, which is the basis for official cartography of many countries, 'analytically' transforms the geographical grid (fig. 2.6), through very complex equations of correlation, in a network very similar to that obtained through the inverse cylindrical projection, by conferring on it the fundamental characteristic of conformity (in addition to those common to projections: rectilinearly between equator images and a meridian, and equidistance along a meridian).

The absence of equidistance (except for the selected central meridian) involves scale variation on the paper, in relation to the position of the measured element. The deformation increases with distance from the central meridian and equator. To reduce deformations the surface to be represented is carefully delineated; the ellipsoid is divided into zones with the central meridian (or zone meridian) chosen as the reference meridian on which the equidistance is achieved.

Through the correspondence formulae or Gauss equations, it is possible to obtain the cartographic co-ordinates, and therefore the plane, of the preset points on the ellipsoid (e.g. nodes of the geographical grid) on a plane representation X-Y (or N-E), remembering that the transformed meridian is shown by the X axis and that the Y one is shown in the parallel direction to the projection cylinder axis.

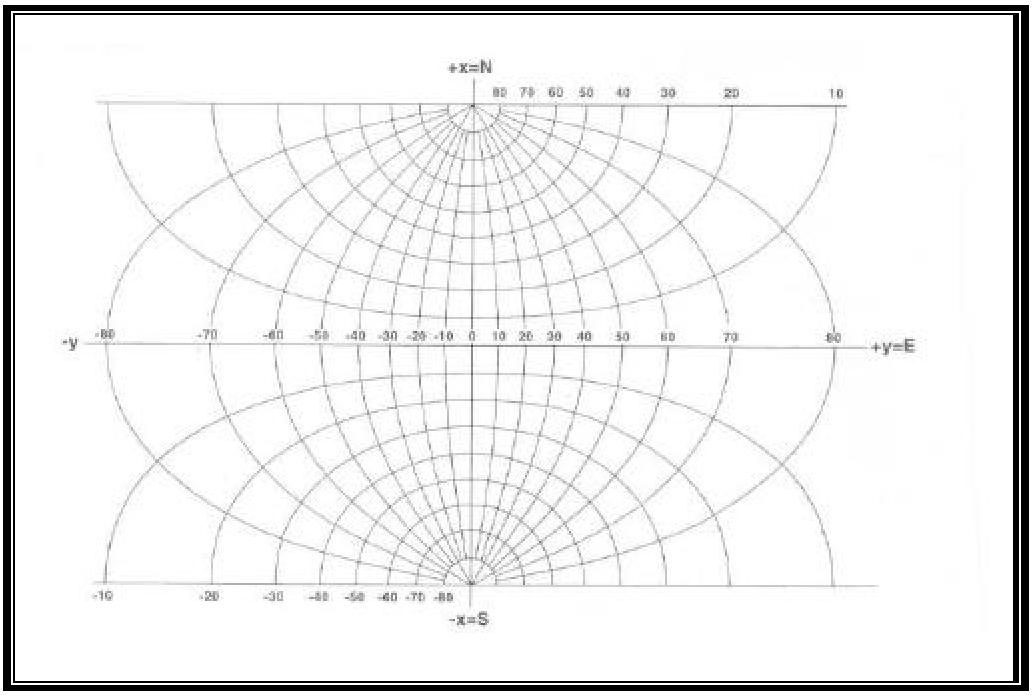


Fig. 2.6 "Geographic grid"

On paper, points with the same abscissa or ordinate are discreet straight lines parallel to the axis. Drawing onto the chart plane some of these straight lines (those corresponding to integer numbers of kilometres), creates a lattice network of squares, called a 'grid'.

In modern charts, on the sheets only the grids are shown, while the geographical grid (graticule) is shown only with traces of parallels and meridians on the sheet margin.

The presence of the grid allows operation in the horizontal field within the whole zone, with the only need for correction being the distances calculated through co-ordinates with the aid of linear deformation coefficient. Since the transformed curve of the geodetic is not a straight line segment corrections to the angles (through 'chord reduction') have to be introduced.

The cartographic system based on Gauss representation is internationally recognised as 'Universal Transversal Mercator Projection' or 'UTM' because of the analogy with the inverse cylindrical projection obtainable from the direct cylindrical projection (Mercator).

2.5.5 Universal Transverse Mercator projection

Universal Transverse Mercator (UTM) co-ordinates are used in surveying and mapping when the size of the project extends through several region plane zones or projections and are also utilised by the NATO Armies, Air Forces and Navies for mapping, charting and geodetic applications.

Differences between the UTM projection and the TM projection are in the scale at the central meridian, origin, and unit representation:

- The scale is 0.9996 at the central meridian of the UTM projection;
- The northing co-ordinate (NUTM) has an origin of zero at the equator in the Northern Hemisphere up to latitudes eighty four degrees north (84° N);
- The southing co-ordinate (SUTM) has an origin of ten million meters (10,000,000 m) in the Southern Hemisphere up to latitudes eighty degrees south (80° S).;
- The easting co-ordinate (EUTM) has an origin five hundred thousand meters (500,000 m) at the central meridian.
- The UTM system is divided into sixty (60) longitudinal zones. Each zone is six (6°) degrees in width extending three (3°) degrees on each side of the central meridian.

To compute the UTM co-ordinates of a point, the TM co-ordinates must be determined:

- The UTM northing or southing (NUTM, SUTM) co-ordinates are computed by multiplying the scale factor (0.9996) at the central meridian by the TM northing or southing (NTM, STM) co-ordinate values;
- In the Southern Hemisphere, a ten million meter (10,000,000 m) offset must be added to account for the origin;
- The UTM eastings (EUTM) are derived by multiplying the TM eastings (ETM) by the scale factor of the central meridian (0.9996) and adding a five-hundred thousand meter (500,000 m) offset to account for the origin;
- UTM co-ordinates are always expressed in meters.

UTM Northings, Southings, and Eastings

Northern Hemisphere:	$N_{UTM} = (0.9996)\ N_{TM}$
Southern Hemisphere:	$S_{UTM} = (0.9996)\ S_{UTM} + 10{,}000{,}000$ m
Easting co-ordinate:	$E_{UTM} = (0.9996)\ E_{TM} + 500{,}000$ m

The UTM zone (Z = UTM zone number) can be calculated from the geodetic longitude of the point (converted to decimal degrees):

- $Z = (180 + \lambda) / 6$ (east longitude)
- $Z = (180 - \lambda) / 6$ (west longitude)

If the computed zone value **Z** results in a decimal quantity, then the zone must be incremented by one whole zone number.

Example of UTM Zone Calculation:

λ = 15° 12' 33.5609'' E
Z = 195.20932247 / 6 = 32.53448
Z = 32 + 1
Z = 33

In the example above, **Z** is a decimal quantity, therefore, the zone equals seventeen (32) plus one (1).

3. HORIZONTAL CONTROL METHODS

3.1 Introduction

In the hydrographic field, the topographic survey, established to frame geographically the coastal territory or to create the land marks for hydrographic surveying, is carried out commencing from previously established topographic stations with co-ordinates already determined by geodetic survey operations.

Such points and the connecting network, termed the primary control, produce the adopted geodetic reference system (Datum).

Their horizontal determination can be obtained by:

a. classical methods of survey (astronomical observations and measurements of angles and distances);
b. mixed methods of survey;
c. photogrammetric methods of survey.

The first two methods accomplish the basic control networks, primary or of inferior order, via triangulation, trilateration and traverse operations. Afterwards, from the points of the primary control, the control can be extended as required for the particular survey needs with further measurements of angles and distances.

The development of the satellite technology has allowed the determination both of the stations of a primary basic control network and the points of the secondary control network to be derived without a geometric connection between them, until the topographic survey level of a particular site.

3.2 Classic method

3.2.1 Triangulation

3.2.1.1 Principles and specifications

In every country within their national boundaries some points are known, termed trigonometric stations, monumented in some fixed way and connected to each other in some form of a sequence of triangles, possibly of equilateral form.

The survey technique, called triangulation, creates, by primarily measuring angles, the determination of points of a triangular network, with every triangle having at least one common side.

The development, formed by triangles, can be made by continuing the networks (fig. 2.7a) or in a first phase 'made by chain' (fig. 2.7b). This last method has been successfully applied for the survey of states which are wide in latitude or in longitude (i.e. Argentina).

Additionally the chains can be related to themselves, in the case of a survey of a long and narrow zone; in such a case it is relevant to a more rigid scheme, such as quadrilateral with diagonals (fig. 2.7c).

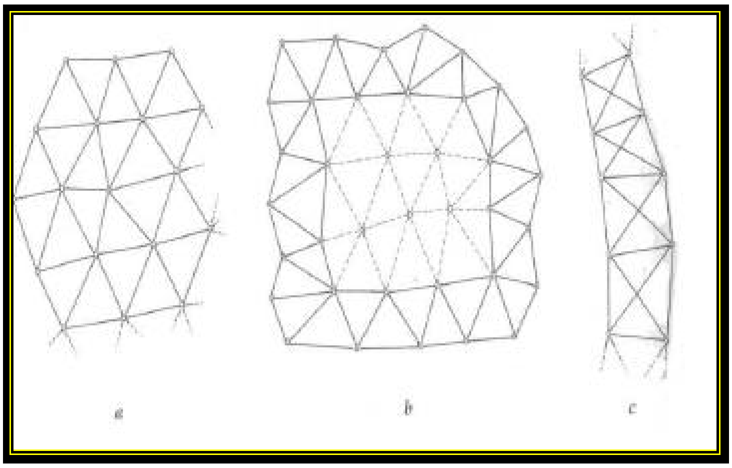


Fig. 2.7

Scale in a network can be achieved by the measurement of a single baseline, with all other measurements being angular. However errors of scale will accumulate through the network and this is best controlled/corrected by measuring other baselines. (NB. Before the advent of Electronic Distance Measurement the measurement of distance was a long and difficult task.)

Finally the orientation of the network has to be determined through measurement, by astronomic means, the azimuth of one side. As with scale, further azimuths should be determined throughout the network in order to correct/control the propagation of errors.

3.2.1.2 Base and angles measurements

To clarify how a triangulation survey is conducted, the aim is to determine the co-ordinates of points A, B, C, D, E and F (fig. 2.8); the points are connected so that they form a sequence of triangles. In general the AC side (normally named "base" in triangulation) and all the angles of the various triangles are measured: a_1, ß$_1$, ?$_1$ of the ABC triangle; a_2, ß$_2$, ?$_2$ of the ABD triangle, and so on.

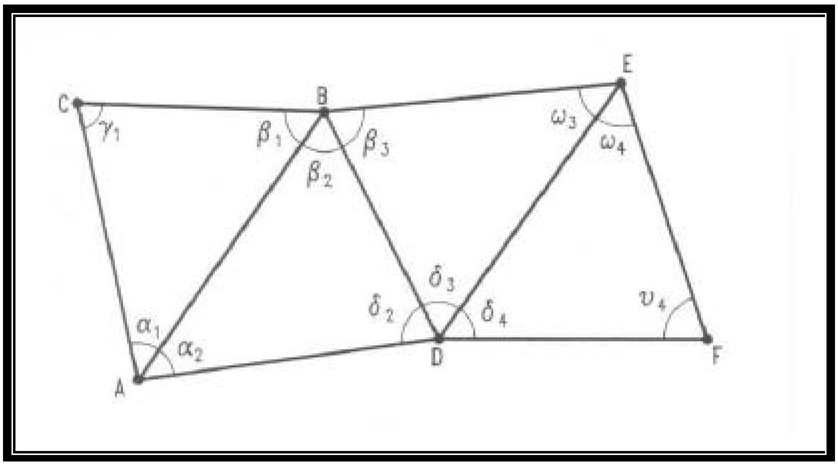


Fig. 2.8

The base length of the primary triangulation is in the order of about ten kilometres, and therefore, the measurement of the angles needs particular care; it is necessary to use theodolites reading to one or two tenth of a sexagesimal second, the purpose being to obtain, with suitable reiterations, the measurement of the directions with a root mean square error in the order of tenths of seconds.

To achieve measurements within these tolerances, particular importance should be attached to the targets, which need to be of conspicuously large structure and of suitable colouration. Diurnal or night time brightly lit targets can be used; the diurnal ones are produced by heliostats (or heliotropes) and at night by projectors. Both must allow collimation adjustments removed of any phase error and therefore require the presence of an operator at the point to be collimated.

Therefore in every triangle, having measured all three angles, the precision of each measurement needs to be verified; to calculate the error of angular closure (or angular closing error) of every triangle, verifying that the results are less than the pre-fixed tolerance:

$$e_a = \left|\sum a_i - 180°\right| \le t_a \tag{2.5}$$

where the summation $\mathbf{Sa_i}$ is the sum of the measured angles with the spherical excess removed; then to adjust the measured angles using a rigorous method or empirically adding to or subtracting from every angle a third of the angular closing error.

3.2.1.3 Computation and compensation

Once completed the verification of the tolerance, the first triangle ABC (in fig. 2.8) can be resolved, knowing a base and the three angles determining the other two in general through the application of the sine rule:

$$\overline{AB} = \overline{AC} \cdot \frac{\text{sen ?}_1}{\text{sen ß}_1} \quad \textbf{(2.6)}$$

$$\overline{BC} = \overline{AC} \cdot \frac{\text{sen a}_1}{\text{sen ß}_1} \quad \textbf{(2.7)}$$

We are now able to resolve the second triangle ABD, having determined its base, always applying the sine rule and so on.

If there is more than one measured base, it is necessary to use rigorous methods to conduct the compensation adjustment. The most frequently used method is of indirect observations:

The hyper-determination (over abundance of measurements) of the network allows the compensation adjustment calculation to be undertaken with a least squares approach.

Then, for instance, taking the ABD triangle (fig. 2.9), the unknown values are generated by the most probable values of the horizontal co-ordinates of points A, B, D (listed with X_A, X_B, X_D, Y_A, Y_B, Y_D,). Such co-ordinates are expressed as the sum of an initial approximate value and the relative corrections to apply to produce the resultant more probable final value by the use of the principle of the least squares.

Once the angular measurements are adjusted, the operations requiring completion are:

a. Formulation of a generating equation for every effected measurement. Particularly we impose the condition that an angle (i.e. $\mathbf{a_2}$), has to be equal to the difference of the two angles of direction measured on the AD base and on the AB base:

$$a_2 = (AD) - (AB) \quad \textbf{(2.8)}$$

from which:

$$a_2 - (AD) + (AB) = 0 \quad \textbf{(2.9)}$$

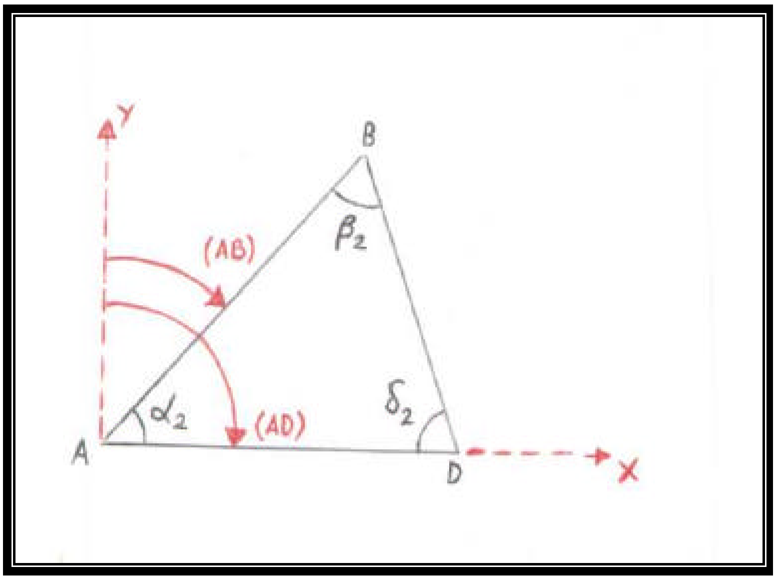


Fig. 2.9

The system of generating equations is an impossible task because the number of the equations (one for every measurement) is greater than the number of the unknowns (affect of the hyper-determination method).

The unknowns are contained in the measured angles of direction, they can be expressed in the following way:

$$(AD) = arctg\left[(X_D - X_A)/(Y_D - Y_A)\right] \qquad \textbf{(2.10)}$$

$$(AB) = arctg\left[(X_B - X_A)/(Y_B - Y_A)\right] \qquad \textbf{(2.11)}$$

Developing in Taylor's series, the function '**arctg**' of the two varying $\mathbf{X_i}$ and $\mathbf{Y_i}$ ($\mathbf{f[X_i,Y_i]}$) for a point whose co-ordinates $\mathbf{X_i}$° and $\mathbf{Y_i}$° represent the initial approximate co-ordinates of the points of the triangle, the increments $\mathbf{x_i}$ and $\mathbf{y_i}$ from these points constitute the corrections to be applied to calculate the adjusted final values (more probable value).

For present purposes, the development of the series terms greater than the first degree are considered negligible and are ignored:

$$f[X_i, Y_i] = f[X_i^0, Y_i^0] + \left|\partial f[X_i, Y_i]/\partial x\right|_{(X^0,Y^0)} x_i + \left|\partial f[X_i, Y_i]/\partial y\right|_{(X^0,Y^0)} y_i \qquad \textbf{(2.12)}$$

b. Formulation of the system of generated equations, imposing the existence of a observation residual ($\mathbf{v_i}$) resulting from the presence of inevitable accidental errors in the measurements; a generated equation of the type:

$$a_2 - (?D) + (??) = n_i \qquad \textbf{(2.13)}$$

Such a system results to being algebraically indeterminate since the number of equations is now less than the number of the unknowns (having inserted the observation residuals).

c. Formulation of the system of normal equations in the unknowns, corrections $\mathbf{x_i}$ and $\mathbf{y_i}$ introduced in the expression (2.12), resulted in imposing the condition that the sum of the squares of the observation residual, $\mathbf{v_i}$, to be a minimum. At this point the system is algebraically determinable with the number of equations equal to the number of the unknowns; it will allow the establishment of the adjusted values of the horizontal co-ordinates of the points of the triangulation.

3.2.2 Trilateration

3.2.2.1 Principles and specifications

This method of surveying is similar to triangulation, as the co-ordinates of a number of points are calculated by connecting the points in order to form a network of triangles with common sides, but where the principal measurements are distances not angles.

With the advent of the Electronic Distance Measuring (EDM) equipment and Electro-optic Distance Measuring (EODM) equipment, trilateration has wide applications and can totally replace triangulations; however the two methods normally coexist giving rise to mixed networks.

While in triangulations the controlled development of elements (triangles) can be achieved by measuring three angles for each triangle (control can be immediate through the sum of the three angles), in trilateration control has to be accomplished by examining adjacent triangles, after having calculated the angles in terms of the measured sides.

As for triangulations, for primary networks, the creation of a point of origin is always necessary and an azimuth by astronomical techniques for the control of orientation.

3.2.2.2 Angles and distance measurements

In comparison with triangulation, which can be undertaken by one operator with no requirement for the targets to be occupied except when using helioscopes or other lighting arrangements, trilateration always requires the occupation of the targets with prisms or some other form of reflector. This disadvantage is balanced by the advantage of being able to operate under less than perfect conditions of visibility, which allows more flexibility in planning and reduces working time.

3.2.2.3 Computation and compensation

By using the technique of indirect observations, calculation of compensation follows the same procedure of that related to triangulations. Generating equations are formulated in relationship to the measurements of sides and to satisfy the condition given, by Pitagora theorem, to the co-ordinates of points at the extreme of the measured side.

Taking the triangle in fig. 2.9, the equation relating to measured side AD will be:

$$(X_D - X_A)^2 + (Y_D - Y_A)^2 - \overline{AD}^2 = 0 \qquad \textbf{(2.14)}$$

As for triangulations, developing (2.14) in a Taylor's series, around an approximate value of the co-ordinates for points A and D (X_D°, X_A°, Y_D°, Y_A°), and considering only the first degree terms of such a development, the following expression can be produced:

$$\left(X_D^0 - X_A^0\right)^2 + 2\left(X_D^0 - X_A^0\right)\left(x_D - x_A\right) + \left(Y_D^0 - Y_A^0\right)^2 + 2\left(Y_D^0 - Y_A^0\right)\left(y_D - y_A\right) - \overline{AD}^2 = 0 \qquad \textbf{(2.15)}$$

where the increases ($x_D - x_A$) and ($y_D - y_A$) represent the corrections to apply to the initial approximate values of the co-ordinates, in order to create the adjusted most probable values.

The introduction of observation residuals and the application of the princip les of least squares enable the writing of the algebraically determined system of normal equations for the unknowns x_i and y_i.

3.3 Mixed method

The combination of angular, triangulation, and distance, trilateration, measurements requires care due to the different weights for the two methods of measurement. The weight of every observation is inversely proportional to the variance (**μ**) of the measurement.

Thus, assuming a root mean square error in angular measurements of ±1" (equivalent to $4.9 \cdot 10^{-6}$ radians) and a mean of relative error in distances of 10^{-5} m, the calculation of weights (applicable to P_a and P_d) emphasises that:

$$P_a \approx \left(10^{-6}\right)^2 \approx 10^{-12} \qquad \textbf{(2.16)}$$

$$P_d \approx \left(10^{-5}\right)^2 \approx 10^{-10} \qquad \textbf{(2.17)}$$

which indicates that angular measurements have an inferior weight 25 times to that of distances.

Thus, for the example, to combine observation equations, where residuals have the same precision of the associated measurements, resulting from measurements of distances and angles, it will require the angular equation terms to be multiplied by 100.

3.3.1 Traverse

3.3.1.1 Principles and specifications

The traverse surveys are frequently used in topography when undertaking more specific surveys over large areas or where lines of sight are obscured. These surveys are conducted by determining the co-ordinates of numerous points, connected to form a polygonal network. With the exception for the first and last points, the stations in a traverse have to be accessible and generally each station is visible from both the preceding and the following, marks for measurement of angles and distances.

Whether the first and last points of a polygonal network coincide or not, a traverse can be either closed or open. Whether absolute co-ordinates of some stations are known or not, it can be either oriented or not oriented.

In old topographical models, triangulation was the only available technique for creating a network of points over a wide area. Traverses were reserved for connecting points of the lowest order within a detailed survey. If the area was very small, a small network for closed traverse was surveyed; but if the area was large and the chart had to be at a large scale within the nearest known stations, the traverse connected the triangulation points and it was said to be open. Now days the use of EDM or EODM enables the survey of traverses over many kilometres and the programming of the surveys with more accurate traverses, which can directly connect to points of a national primary triangulation, completely replacing inferior order triangulation.

A significant defect with traverses is in the progressive increase of the error in the direction of progress, such error is the algebraic sum of all the errors created in the measurements of angles and distances from each mark, known as propagation of errors.

3.3.1.2 Base and angle measurements

In relationship to the measurements, of which there has to be at least one distance, the traverse can be:

a. Iso-determined: number of measurements equal to the number of the unknowns (co-ordinates of stations). If '**n**' is the number of marks, the number of measurements necessary is equal to (2n – 3);
b. Over-determined: number of redundant measurements in comparison to those necessary, thus there is a possibility to effect a control of the accidental errors, to adjust them and finally to obtain an evaluation of the precision of the final results. Furthermore, given the lower number of possible redundant measurements, the degree of over-determination can be at the most 3; empirical methods are applied for the adjustment of traverses rather than rigorous ones.

3.3.1.3 Computation and compensation

It is understood that the horizontal angles in association with the points of a traverse are those which are produced by making a clockwise rotation from the preceding direction towards the direction of advance. The calculation of the angles at a point in a traverse is therefore rigorous; knowing the angles of a direction it is possible to calculate the difference between the forward and back angles, if the difference is negative it is necessary to add 360°.

This is called the 'rule of transport'; a direction at a point A_i is given by the sum of the direction at the preceding point A_{i-1} and the angle to the point A_i, the measured angle between the two sides; if necessary 360° are added to or subtracted from the result to give a direction between 0° and 360°.

3.3.2 Not Oriented Open Traverse (iso-determined)

Reference fig. 10, the calculations to be developed in succession are:

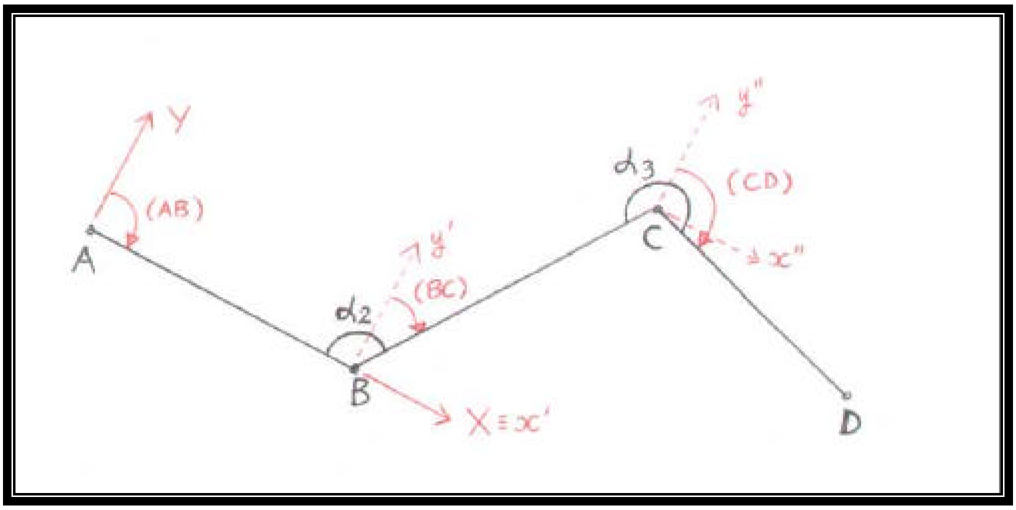


Fig. 2.10

a. Calculation of the angles of direction of the sides through the rule of the transport, remembering that the angle of initial direction (AB) it is obtained from the established local reference system (with direction of the x-axis on the first AB side and y-axis orthogonal to it). For example the angle of direction (BC) it is:

$$(BC) = (AB) + a_2 \pm 180^\circ \qquad \textbf{(2.18)}$$

b. Calculation of the initial co-ordinates, having defined as a partial reference system those centred on the preceding point to that being observed, with axes (indicated in the figure by x' y' , x'' y'') parallel to those initially described. For example, the co-ordinates of point C in comparison to point B are:

$$x_{C(B)} = \overline{BC} \cdot sen(BC) \qquad \textbf{(2.19)}$$

$$y_{C(B)} = \overline{BC} \cdot \cos(BC) \qquad \textbf{(2.20)}$$

c. Calculation of the final co-ordinates in comparison with the first local reference system centred on point A, which has the co-ordinates $X_A = 0$ and $Y_A = 0$. The final co-ordinates of point B are:

$$\begin{aligned} X_B &= X_A + x_{B(A)} \\ Y_B &= Y_A + y_{B(A)} \end{aligned} \qquad \textbf{(2.21)}$$

and so on for the following points.

It is important to notice that having the number of the measurements (angles a_A a_B and distances AB, BC, CD) equal in number to the unknowns (X_B X_C Y_C X_D Y_D final co-ordinates) the structure is iso-determined and it is therefore not possible to perform an adjustment or to appraise the precision of the final results.

3.3.3 Oriented Open Traverse (over-determined)

Reference fig. 2.11, the known elements of the problem are the absolute co-ordinates of the first and last stations of the traverse, A and D, relative to an external reference system (such as a national local Datum) and the co-ordinates, always in relation to the same reference system, of external points, P and Q, which serve to create the hyper-determination of the network. The measurements (angles a_A a_B a_C a_D and distances AB, BC, CD) are more related to the unknowns represented by the absolute co-ordinates of the intermediary points (X_B Y_B X_C Y_C), for every additional measurement there will be an equation of adjustment created.

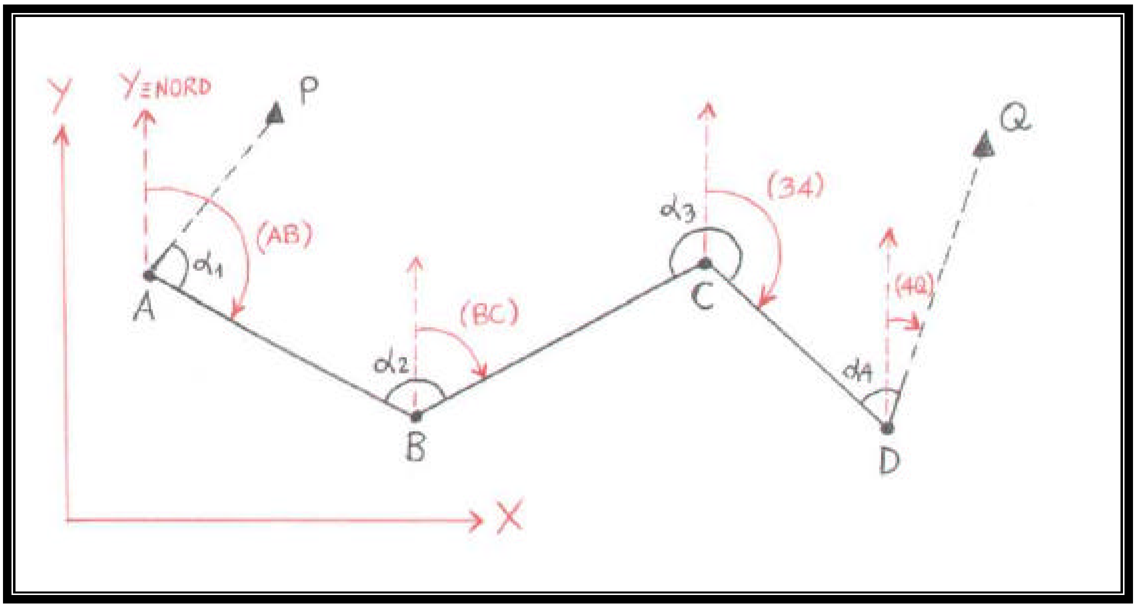


Fig. 2.11

The calculations to be developed are:

a. Calculation of the angles of direction, often known as the azimuths, unadjusted with the rule of the transport, starting from the first angle of direction (PA) already adjusted and calculated:

$$(PA) = arctg\left[(X_A - X_P)/(Y_A - Y_P)\right] \quad \textbf{(2.22)}$$

For example, the unadjusted angle of direction for the side DQ (equal to (DQ)') is:

$$(DQ)' = (CD) + a_D \pm 180° \quad \textbf{(2.23)}$$

b. Formulation of the first adjustment equation making use of the opportunity to calculate the final adjusted angle of direction (DQ):

$$(DQ) = arctg\left[(X_Q - X_D)/(Y_Q - Y_D)\right] \quad \textbf{(2.24)}$$

The condition, to be imposed at this point, is the equality among the already adjusted calculated value of (2.24) and the unadjusted in (2.23). The equation is:

$$(DQ)^{'} - (DQ) = 0 \quad \textbf{(2.25)}$$

With the unavoidable presence of accidental errors in the measurements of angles a_A ,a_B ,a_C ,a_D, which are present in the calculation of (DQ)', (2.25) will never be satisfied because of the presence of residuals called ‘error of angular closing’ and are termed ?a. The (2.25) then becomes:

$$(DQ)^{'} - (DQ) = \Delta a \quad \textbf{(2.26)}$$

remembering that ?a has to be smaller than an angular tolerance established for the project.

c. Calculation of the adjusted angles of direction:

$$\begin{aligned} (AB) &= (AB)^{'} - \mathrm{u}a \\ (BC) &= (BC)^{'} - 2\mathrm{u}a \\ (CD) &= (CD)^{'} - 3\mathrm{u}a \\ (DQ) &= (DQ)^{'} - 4\mathrm{u}a \end{aligned} \quad \textbf{(2.27)}$$

gives **ua**, where **u** represents ‘the unitary error of closing’ equal to the relationship between the error of angular closing and the number of the angles not adjusted on which to share it.

d. Calculation of the partially unadjusted co-ordinates, having defined the partial reference systems centred on the points and with parallel axes to those of the absolute system at the start. For example the partial unadjusted co-ordinates of point B relative to A are:

$$\begin{aligned} x_{B(A)} &= \overline{AB}^{'} sen(AB) \\ y_{B(A)} &= \overline{AB}^{'} \cos(AB) \end{aligned} \quad \textbf{(2.28)}$$

e. Formulation of the second and third equations by imposing the condition that the sum of all the partial co-ordinates is equal to the difference between the absolute co-ordinates of the last and the first points. There are two equations because one relates to the abscissas and the other to the ordinates:

$$\sum x' - (X_D - X_A) = 0$$
$$\sum y' - (Y_D - Y_A) = 0 \quad \textbf{(2.29)}$$

Likewise in the case of the angles, the equations will never be satisfied because the residuals, which are termed 'error of linear closing in abscissas' and 'error of linear closing in ordinates', are equal to:

$$\Delta x = \sum x' - (X_D - X_A)$$
$$\Delta y = \sum y' - (Y_D - Y_A) \quad \textbf{(2.30)}$$

Defining **?L** as:

$$\Delta L = \sqrt{\Delta x^2 + \Delta y^2} \quad \textbf{(2.31)}$$

?x and **?y** have to be such that **?L** is not greater than an established linear tolerance.

f. Calculation of the partially adjusted co-ordinates:

$$x_{2(1)} = x'_{2(1)} - u_x \qquad y_{2(1)} = y'_{2(1)} - u_y$$
$$x_{3(2)} = x'_{3(2)} - u_x \qquad y_{3(2)} = y'_{3(2)} - u_y \quad \textbf{(2.32)}$$
$$x_{4(3)} = x'_{4(3)} - u_x \qquad y_{4(3)} = y'_{4(3)} - u_y$$

where $\mathbf{u_x}$ and $\mathbf{u_y}$ represent the values of the unitary linear errors of closing and are equal to the relationship between the error of linear closing, related to the abscissas and to the ordinates and the number of partially unadjusted co-ordinates on which to share it in a uniform manner.

g. Calculation of the total co-ordinates (absolute) of the unknown intermediary points (B and C) departing from the known values of the initial point A and adding the values of the following partial co-ordinates.

$$X_B = X_A + x_{B(A)} \qquad Y_B = Y_A + y_{B(A)}$$
$$X_C = X_B + y_{C(B)} \qquad Y_C = Y_B + y_{C(B)} \quad \textbf{(2.33)}$$

3.3.4 Not Oriented Closed Traverse

Reference fig. 2.12, the known elements of the problem are represented by the co-ordinates of station A, in which the origin of the local reference Cartesian system has been settled with the x-axis in the direction of the first measured side AB, and from the ordinate, equal to 0 in the same local Cartesian system, of the second position B. The ten measured elements are all the inside angles and sides of the polygon, while the seven unknowns (X_B X_C Y_C X_D Y_D X_E Y_E) determine a hyper determination of a maximum possible Order 3.

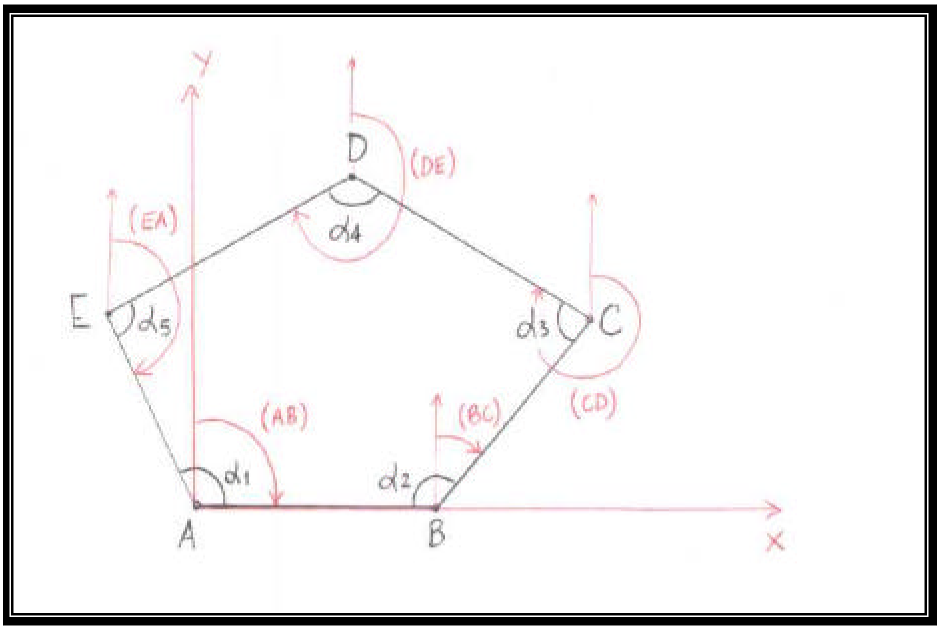


Fig. 2.12

The calculation has the following phases:

a. control and angular adjustment, imposing that the sum of the unadjusted measured angles is equal to the sum of the inside angles of a polygon with '**n**' sides ((n-2) 180°). Due to the inevitable accidental errors, the following observation residuals (error of angular closing **? a**) are generated:

$$\Delta a = \sum a' - (n-2) \cdot 180° \quad \textbf{(2.34)}$$

such that the result is smaller than a fixed tolerance. The unitary closing error '**ua**' (equal to angular closing error divided by the number of measured angles) has to be uniformly shared between all the measured angles.

$$\begin{aligned} a_1 &= a_1' - \text{u}a \\ a_2 &= a_2' - \text{u}a \end{aligned} \quad \textbf{(2.35)}$$

and so on. The calculated angles are now adjusted.

b. calculation of direction angles (in comparison to local system y-axis direction) using the rule of transport.

c. calculation of the unadjusted partial co-ordinates with (2.19) and (2.20).

d. control and side compensation, imposing that the sum of all the partial abscissas is zero (the same for ordinates). Taken into account in the calculation of the unadjusted partial co-ordinates, this condition will not be satisfied resulting in the residues **?x** and **?y** (side closing errors). Defined the quantity **?L** as:

$$\Delta L = \sqrt{\Delta x^2 + \Delta y^2} \qquad \textbf{(2.36)}$$

?x and **?y** have to be such that **?L** is not greater than an established linear tolerance. The unitary error of closing to be shared between the $\mathbf{u_x}$ unadjusted partial abscissas; it is equal to the relationship between the side closing error of the abscissas and the number of co-ordinates to be calculated. While the unitary error related to $\mathbf{u_y}$ ordinates must be calculated by dividing by the number of co-ordinates to be calculate less 1. This is because the co-ordinates of point B, in comparison to the origin at A ($\mathbf{y_{B(A)}}$), are unadjusted (fixed at 0 by axis choice), in order not to change the local Cartesian system orientation. It is now possible to proceed with the adjustment of partial co-ordinates, by subtracting $\mathbf{u_x}$ and $\mathbf{u_y}$ from the values of abscissas and of unadjusted ordinates, as detailed in (2.32), with the only exception for the value $\mathbf{y_{B(A)}}$, which is fixed at 0, as already stated.

e. calculation of the total co-ordinates with (2.21).

3.4 **Photogrammetric method** (also see Chapter 6)

Photogrammetry is a widespread technique for topographic surveying of the ground or objects through the use of photographs taken from different view points.

Conventional Photogrammetry is usually divided into two categories:

a. Terrestrial Photogrammetry, in which the photographs are taken from points on the ground;
b. Aerial Photogrammetry, in which photographs are taken from aircraft.

Such distinctions do not relate to the procedures of restitution, which are in principle the same, but to the methods and procedures used to obtain the images.

To ensure the topographical restitution of the photographed object it is necessary to have at least two images of the point of interest taken from two different positions. If the position of the cameras is known, the spatial co-ordinates of the points of interest on the two photographs can be calculated from the two straight lines intersecting the images with the relative optic centres. This is the fundamental principle of photogrammetry and it is common to all the techniques of photogrammetric survey.

During a photogrammetric survey there are three quantities, connected in different ways at various points of the survey. They are:

a. The three-dimensional co-ordinates (X, Y, Z) of the photographed objects;
b. The horizontal co-ordinates (x, y) of the images of the objects photographed on the plane of the film;
c. The entire parameters of the orientation, required to establish the position of the camera during the photograph.

At the moment of exposure two groups of quantities are assigned, although they may not be numerically known at the time: the co-ordinates of the photographed objects and the parameters of the orientation, i.e. the position and the optic characteristics of the camera. From knowledge of the real spatial co-ordinates and the horizontal co-ordinates on the film of some known points, the parameters of the orientation can be calculated. Finally in the restitution phase, with the parameters of orientation calculated, it is possible to determine the co-ordinates of all the observed points using the horizontal co-ordinates on the photogram.

One of the most important applications of photogrammetry is in cartographic production at variable scales from 1:500 to 1:50.000.

3.4.1 Aerophotogrammetry (Air photogrammetry)

Almost all charts are created by air photogrammetry. Due to this technique it is possible to generate topographical charts of large areas in relatively short times, instead of the many years required for traditional techniques.

Aerial photographs can be produced in different ways, depending on the kind of chart to be created and on the kind of camera to be used. Air photogrammetry generally employs cameras with nadir photographs (also called nadir point or plumb point), that is with the optic axis coincident with the vertical axis. This has the advantage of providing photograms with a constant scale if the ground is flat as well as allowing photogram stereoscopic observation.

Even if suitably enlarged, aerial photograms can not be used as maps of the photographed territory. The aerial photograph is a central perspective, while maps are produced with an orthogonal projection of the ground on the reference surface. Due to this difference, a vertical segment, which would be represented by a point in a map, is represented by a segment on a photograph.

Another difference between photography and cartographic representation is due to the fact that in the photogram the scale factor is definable only in the case when the object is perfectly horizontal and the axis of the camera strictly vertical. If in the observed area there are height differences, the scale of the photogram will vary from point to point and only an average scale can be defined; the choice of the average scale will determine the flight altitude.

To guarantee the fundamental principles of photogrammetry, each point of the area of survey has to be taken in separate photos, thus the two adjacent photograms have to result in an overlap of 50% of their length. To avoid the risk that some areas will not have this overlap due to variations in aircraft speed, a 60-70% overlap is normally adopted. The succession of photograms in a longitudinal direction is called a continuous-strip. Generally, it is necessary to take various continuous-strips, which are then placed transversally over each other to achieve an overlap of 15-30% of the photogram width to compensate inevitable aircraft drift.

3.4.1.1 Photogrammetric restitution

After having completed the survey, the two resulting photograms represent, from two different points, a perspective projection of the object. The photogram pairs are used for the restitution of the surveyed objects, through either complex equipment (stereoscopic plotting instruments) or a simple stereoscope, which allows the simultaneous observation of the objects via its binocular optical ability, allowing each eye to see only one photograph.

With stereoscopic photogrammetry the survey is not made on the plane, as with the traditional methods which obtain measurements from reality, but from a stereoscopic model (or stereomodel), observable through a pair of photographs, which dimensionally reconstruct it in an appropriate scale. In the traditional methods, a limited number of points are surveyed, while in photogrammetry the object is totally surveyed and subsequently the co-ordinates of the points of interest can be determined.

3.4.1.2 Analogue restitution

In analogue restitution the ground model is constructed by optic-mechanic means, from whose observation the paper can be drawn.

To be able to proceed to the restitution it is necessary to know, with great precision, the parameters of the interior orientation (or inner orientation):

a. The calibrated focal length of camera's objective lens;
b. The co-ordinates on the photogram of the calibrated Principal Point, which represents the footprint of the perpendicular from the interior perspective centre to the plane of the photograph (nodal point of the objective). These co-ordinates are calculated in the interior reference system of the photogram, defined by the intersection of the pairs of index marks engraved on the middle points of the sides of the photogram.

The procedure for analogue restitution consists of reconstructing the circumstances of the two photograms at exposure with a geometric similarity between the two configurations. The photograms are placed on two projectors which must be placed in such way as to show an interior orientation equal to that of the aerial continuous-strip camera. Then the parameters of the exterior orientation (or outer orientation) have to be determined, which allow the spatial position of the pair of photograms to precisely known and the ground model or the photographed object can be recreated. The exterior orientation is divided into:

a. Relative: it defines the position of the second photogram in relation to the first. Six parameters are necessary, i.e. the three relative co-ordinates of the second nodal point in relation to those of the first and from the rotations. The calculation of these parameters produces six pairs of homologous points, whilst manually eliminating the transversal parallax from each of them. In this way a stereoscopic model is defined, from which no metric information can be taken because its absolute orientation and the scale are not known;

b. Absolute: it defines the spatial position of the first photogram with reference to an earth fixed system through known points. Six other parameters are necessary because in space a body has six degrees of freedom. Generally these six parameters are the x_v y_v z_v spatial co-ordinates of the nodal point and the three φ_x φ_y τ rotations around three the Cartesian axes passing through the principal point (fig 2.13).

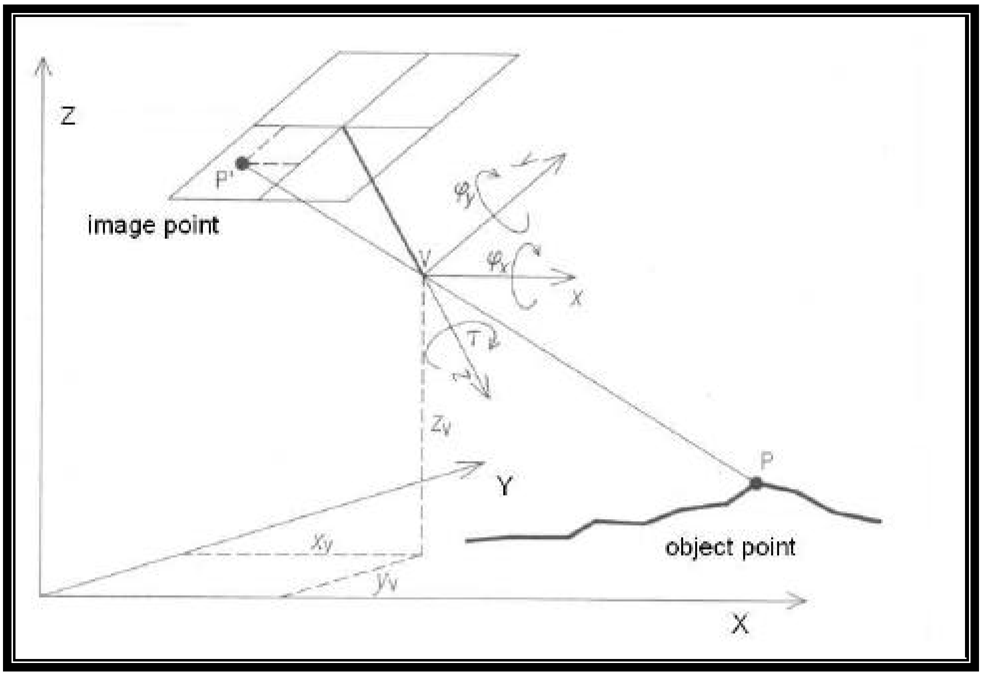


Fig. 2.13 "Twelve parameters determination for an Analogue Restitution"

The determination of the twelve parameters of the exterior orientation enables a return to the spatial position of the two photograms during exposure.

Normally the minimum number of ground control points it is five, of which four (known in the three co-ordinates) are distributed on the edges of the model and a fifth for vertical control, of which only the height is known, is positioned near the centre of the model. In this way the problem will be hyper-determined; there will be some residuals of observations, termed residuals of orientation, that allow the verification of the accuracy of the photogrammetric survey.

The differences of the control points, between the values of the ground co-ordinates and the model co-ordinates, should not be greater than certain limits.

3.4.1.3 The analytical restitution

Techniques of numerical photogrammetric restitution have been developed with the progress of the automatic numerical calculation; these methods make use of the computing power of modern computers to perform the photogrammetric compilation.

3.4.1.4 The digital photogrammetry

Traditional photogrammetry, that is stereoscopic or stereo-photogrammetry, can be achieved by analogue or analytical methods. In creation, the restitution in the analogue photogrammetry is achieved by optical systems; the co-ordinates of the observed points in analytical photogrammetry are mathematically determined.

The digital photogrammetry not only exploits the electronic calculators in the final phase, as in the analytical restitution, but also for the treatment of the images, which are recorded in digital form.

Traditional photographs can be also employed, initially modifying them through equipment which transforms the images into digital signals, such as a scanner.

The adoption of the digital images allows the automation of many operations, which must be performed by the operator such as the definition of the interior and exterior orientation in analytical photogrammetry.

3.4.1.5 Aerotriangulation (Aerial triangulation)

In the conduct of a photogrammetric survey, the co-ordinate determination of the ground control points is generally the phase which requires the greatest employment of time, at least 5 point for every model, which is for every pair. To reduce the number required, the co-ordinates of some can be also obtained through photogrammetric methods, through aerial triangulation.

The determination of the co-ordinates for the control points through the aerial triangulation can be achieved with the method of independent models. It consists of independently building the relative orientation of every model from the others; the models are linked through some points, known as tie points, which are common to the two models (the points common to the three photograms which have produced them) and are located in the marginal areas of the models themselves. In the end a single block of models is produced, of a length and width equal to that of the models linked between them. There would theoretically be only the five control points of the first model; in practice there are essential control points displaced to the edges and along the perimeter of the block of models and some altimetric points inside the block.

However, this technique is being overtaken by the employment of the GPS satellite positioning system, which allows the direct determination of the co-ordinates of the ground control points, whilst at the same time it offers the possibility of directly installing the GPS receivers in the aircraft.

The co-ordinates of points surveyed during exposure through the GPS receivers, using differential techniques with a fixed reference receiver on the ground, can be used during the aerial triangulation as additional data, adopting the method for independent models.

3.5 Inter-visibility of Geodetic Stations

3.5.1 Inter-visibility between two points must ALWAYS be checked in the field during the reconnaissance. However, many proposed lines can be checked during the office phase by plotting cross-sections from a map. A clearance of at least 5m, and preferably 10m, should be allowed on all grazing rays with particular care taken where buildings are shown near ends of lines.

3.5.2 For long lines, the earth's curvature needs to be taken into account when investigating inter-visibility. The formula in paragraph 3.5.3 must then be applied.

3.5.3 In fig 2.14, two stations ‘A’ and ‘B’ of heights ‘H_A’ and ‘H_B’, are a distance ‘D’ apart. The line of sight ‘AB’ will be tangential to a sphere concentric to the earth at a height ‘y’ and a distance ‘x’ from ‘A’. The problem is to determine what height of hill ‘h’, distance ‘d_A’ from ‘A’, will obstruct the line of sight.

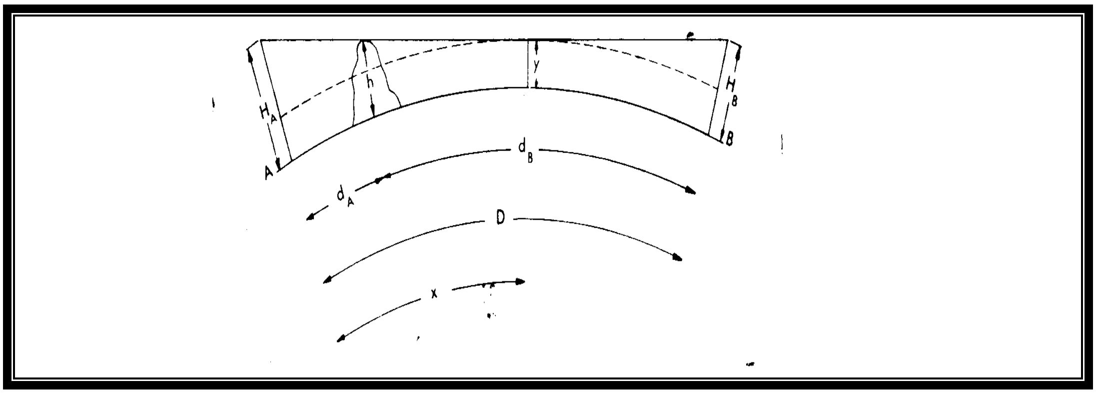


Fig. 2.14 "Intervisibility of Geodetic Station"

The height of an object distance 's' away, which appears on the horizon to an observer with the eye at sea-level, is:

Ks^2, where $K = \frac{\frac{1}{2} - k}{r}$ and k = co-efficient of refraction & r = radius of the earth

Therefore,

$$H_A - y = Kx^2$$
$$H_B - y = K(D-x)^2$$
$$\textit{Whence} \qquad \textbf{(2.37)}$$
$$x = \frac{D}{2} - \left(\frac{H_B - H_A}{2KD}\right) \quad and \quad y = H_A - Kx^2$$

$$h = y + K(d_A - x)^2$$
$$\textit{Therefore,} \quad h = \frac{d_A H_B}{D} + \frac{d_B H_A}{D} - Kd_A d_B \qquad \textbf{(2.38)}$$

Using this formula all inter-visibility problems can be solved. Care must be taken to use the correct units of measurement.

When heights are in metres and distances in kilometres, K = 0.0675.

Proof of formula:

$$h = y + K(d_A - x)^2 \qquad \textbf{(2.39)}$$

$$= H_A - Kx^2 + Kd_A^{\ 2} - 2Kd_A x + Kx^2$$

$$= H_A + Kd_A{}^2 - 2Kd_A \frac{D}{2} + \frac{2Kd_A H_B}{2KD} - \frac{2Kd_A H_A}{2KD}$$

$$= H_A + Kd_A{}^2 - (Kd_A d_A + Kd_A d_B) + \frac{d_A H_B - d_A H_A}{D}$$

$$= \left(\frac{d_A H_A + d_B H_A}{D} \right) + \left(\frac{d_A H_B - d_A H_A}{D} \right) - Kd_A d_B$$

$$= \frac{d_B H_A}{D} + \frac{d_A H_B}{D} - Kd_A d_B \qquad \textbf{(2.40)}$$

4. VERTICAL CONTROL METHODS

4.1 Geometric levelling (Spirit levelling method)

4.1.1 Principles and specifications

Levelling are operations which allow the measurement of difference orthometric heights (or Geoid elevations) between points or their difference in elevation.

The principle of the geometric levelling is: consider two points (A and B) to be a brief distance apart, not more than around 100 metres (fig. 2.15); two vertical stadia are set-up on them and at point M, equidistant from A and from B, an instrument which has its axis of horizontal collimation, or rather (for modest heights) parallel, to the tangent plane in M_0 to the Geoid. Two rounds of readings are taken from the stadia, l_A and l_B. The following expression can be immediately concluded from the figure, with the premise that the Geoid coincides, for the brief line being considered, with the local sphere at M_0:

$$Q_A + l_A = Q_B + l_B \qquad \textbf{(2.41)}$$

from which:

$$Q_B - Q_A = l_A - l_B \qquad \textbf{(2.42)}$$

where:

Q_A = Orthometric height (or Elevation) in A
Q_B = Orthometric height (or Elevation) in B

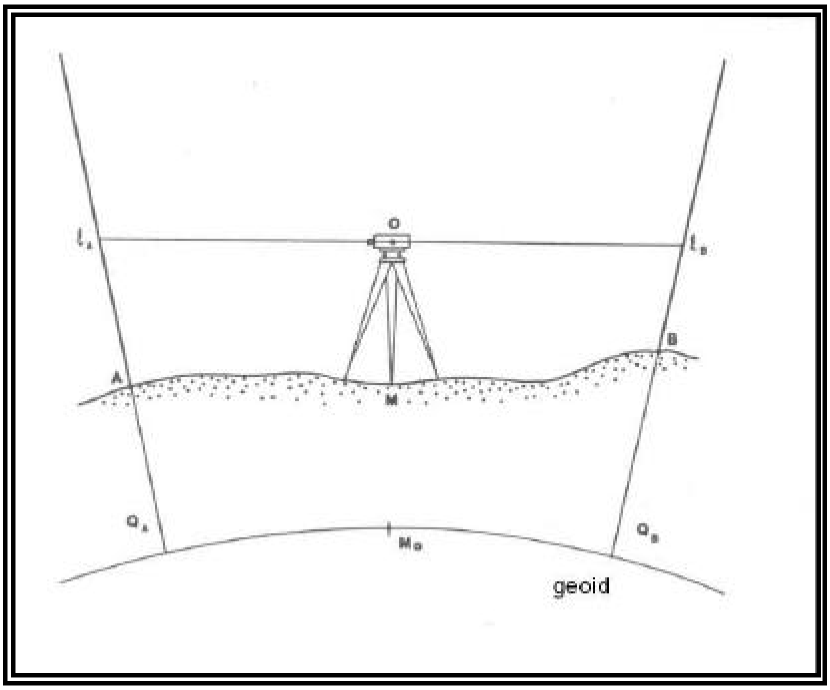


Fig. 2.15 “Geometric levelling”

Since the length of the observation is such as to make the influence of the terrestrial bending negligible, the tool which creates the collimation axis can theoretically be put in any intermediary position between A and B to reduce the influence of the atmospheric refraction.

When the aim is to calculate a difference in levels between points, a distant at which it is impossible to directly make a connection between them, it is necessary to undertake composite levelling. The distance between the start point A and the final point B of the levelling line, is divided into a number lines no greater than 100 metres with the stadia set at the division points.

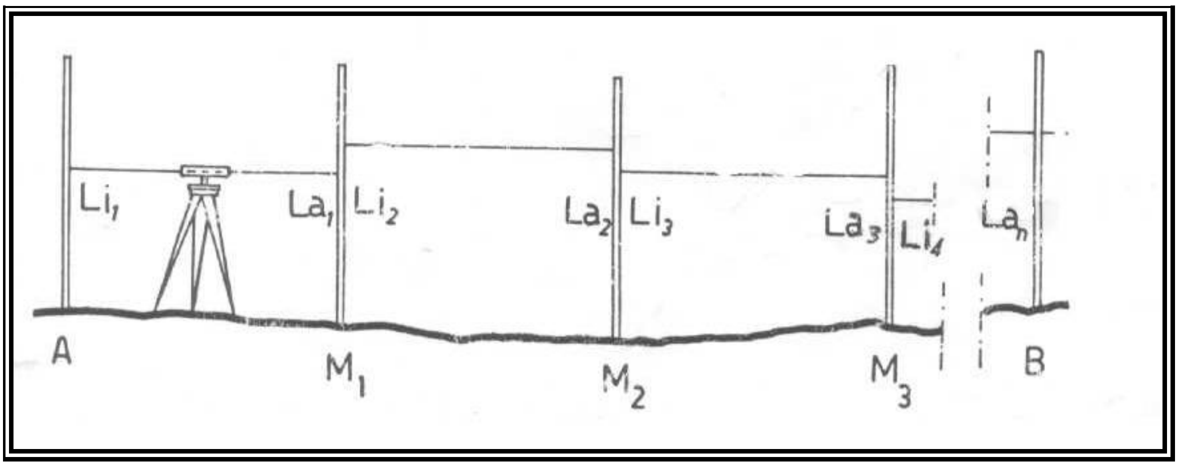


Fig. 2.16 “Level difference among several points”

Departing from A, the difference between M_1 and A is determined as detailed earlier. Thereafter the tool is transported to a point between M_1 and M_2, and the level difference ($l_{i2} - l_{a2}$) is determined between these points (after having rotated the stadia in M_1 on itself and transported that at A onto M_2); this process is repeated to the final point. The total difference in elevation will be:

$$Q_B - Q_A = \sum_n (l_{in} - l_{an}) \tag{2.43}$$

4.1.2 Measurements and quality control

An effective control of the measurements consists in making levelling runs in both directions, but returning by a different route of comparable length. The variation, between the values for the difference in elevation between the start and final points, has to be within established tolerances in relationship to the desired accuracy. The value to be used is the average between the two runs.

During the operation it is good practice to conduct an alignment control for the spirit level (spirit bubble or sensitive bubble) of the levelling instrument before every observation to the stadia.

Some levelling instruments are fitted with a circular level (or universal level or bull's eye level) and the modern ones are fitted with a self-aligning level.

4.1.3 Sources of error

Putting aside the possible inclination of the line of sight, the accidental errors of every observation can be separated in two parts:

a. error of collimation (or of reading the stadia): proportional to the square root of the distance of collimation;
b. error of aligning (or of reading) of the levels (in the auto-levels is replaced by the self-aligning level of the compensator): proportional to the same distance

The mean error of the entire levelling needs to be considered, supposing that the mean error of every single observation is constant and equal to s. Since the total difference in level is equal to the sum of the partial difference of elevations, subsequently determined, the mean error s_t of the whole levelling is:

$$s_t^2 = s_1^2 + s_2^2 + s_3^2 + \ldots\ldots + s_n^2 = ns^2 \tag{2.44}$$

from which:

$$s_t = s\sqrt{n} \tag{2.45}$$

4.1.4 Computation and compensation

As with other hyper-determinations, geometric levelling can be empirically adjusted or via rigorous methods, applying the theory of the least squares.

An elementary adjustment of a levelling line consists of assuming the average between the measurements conducted in both directions.

An empirical adjustment is applied in levels of limited precision, which are performed without the repetition of the measurements but close on the vertical datum point of departure (closed polygon) or on two vertical datum points of known elevations; in this case the closing error is distributed empirically between the differences in elevations.

With the assumption that the closing error is proportional to the distance over which the level is made then it is simply a case of dividing the closing error by the total distance levelled to give an error per metre of levelling. Then each intermediate point is corrected by the error per km of levelling multiplied by the length of observed level to that point.

The adjustment is more complicated when the lines of levelling constitute a network; in this case it is necessary to use to a rigorous network adjustment, preferably by the method of the indirect observations. The unknown quantities of the problem, resolved using the above method, are the corrections to be applied to the approximate values of elevation of the single points of the network, to consequentially obtain the most probable values for the structure.

The generating equations impose the condition that the difference between the measured difference of elevation and consequentially that from the approximation of the network, tends to be zero.

Due to the presence of the inevitable accidental residual errors in the measurements for the difference of elevations, these equations will not normally be satisfied, for the second constituent they will highlight the residues of adjustments. The equations in this form are termed generated equations.

With the distances between the vertical datum points being different, it is necessary to consider weights to assign to the measured differences of elevation; the weights are set to be equal to the inverse to the sum of the distances.

In order to reduce observations of different precision to the same weight (importance) it is necessary to scale their equations by the square root of the weight. We will now have a set of equations equal in number to the observations made. In order to obtain the most probable values for the unknowns (in this case corrections to the initial values of the elevations) it is necessary to reduce the observation equations to normal equations using the principle of least squares.

The subsequent solution of the normal equations will produce the unique and mathematically most probable values to correct the provisional elevations.

The knowledge of the mean error of the unity of weight, which is equal to:

$$m_0 = \pm\sqrt{\sum_i p_i ?_i^2} \Big/ (n-i) \qquad \textbf{(2.46)}$$

where:

p_i: the measure weights inversely proportional to the distances;
v_i: residual accidental errors on the measurements of the difference of elevations;
n: number of the generated equations;
i: number of the unknowns.

It is sufficient to consider the reliability of the task in terms of how much of it is assumed to be unitary weight 1/1km. If the known terms with the consequent residues of compensation are expressed in millimetres, m_0 represents the mean error in millimetres per kilometre; it is in this form that tolerance is normally expressed in geometric levelling (remembering that the tolerance or the maximum admitted error is considered to be equal to three times the mean quadratic error).

4.2 Trigonometric levelling (Trigonometrical heighting)

4.2.1 Principles and specifications

Trigonometric levelling is based on the use of a theodolite for the measurement of zenithal angles. It is employed for any distances, from a few meters to over 10 kms; it is often used for the determination of the elevations of positions in triangulation, it is also applied in other cases, such as when the distance between the points, for which the difference of elevation is required, is already known.

In every case for distances less than around 400 meters, the use of a plane surface of reference involves negligible errors and it results in simplified calculations with mean errors in the order of 5 cms.

Levelling in this case is termed 'eclimetric' and the difference in elevation between two points A and B (Δ_{AB}) is given by:

$$\Delta_{AB} = d \cdot \mathrm{cotg}\, \varphi_A + h - l \qquad \textbf{(2.47)}$$

where:

- d: is the horizontal distance between A and B (on the plane surface of reference);
- φ_A: is the zenithal angle to B measured by the theodolite at A;
- h: is the height of the theodolite related to the ground;
- l: is the height of the target at B related to the ground as measured from the theodolite.

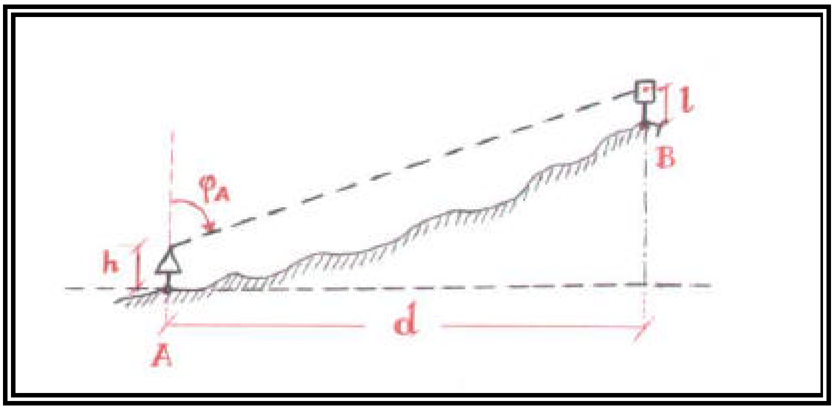


Fig. 2.17 "Trigonometric levelling"

The approximation of the plain surface of reference is not acceptable for distances greater than 400 metres. Thus three fundamental corrections must be considered, departing from the simplified calculations for the "eclimetric" levelling:

a. sphericity;
b. refraction;
c. height.

Taking account of these factors, the procedure for calculating the difference of elevation is termed trigonometric levelling. Since the distance between two points, between which the difference of elevation is being determined, is never greater than 20 kms and normally is less, the calculations can always be performed on the local sphere.

4.2.2 Correction for sphericity

This correction takes into account the bending of the local sphere relative to the plane adopted for the "eclimetric" levelling, with the assumption of negligible divergence between the normals (to the plane and the sphere at the point where the stadia is positioned) along which the difference of elevation is to be obtained.

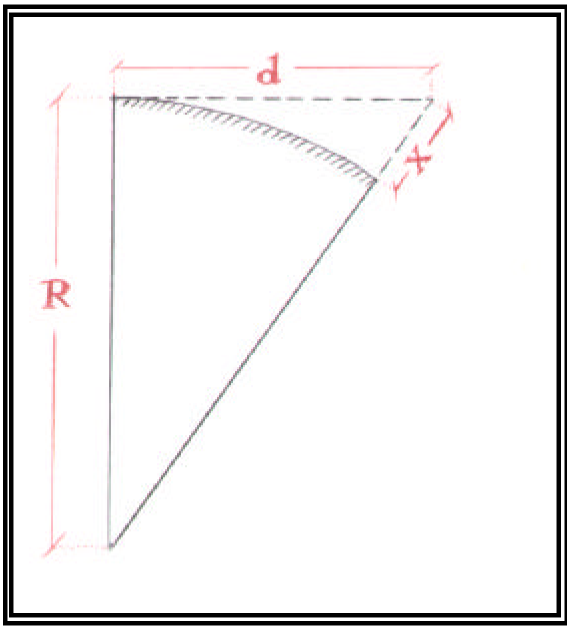


Fig. 2.18 "Correction for sphericity"

where:

X: is the correction for sphericity;
d: is the plain distance between the two points;
R: is the ray of the adopted local sphere.

Applying the theorem of Pythagoras to the triangle in figure 2.18:

$$d^2 + R^2 = (R + X)^2 \qquad \textbf{(2.48)}$$

developing and dividing both sides by **2R** and considering negligible the relationship $\mathbf{X^2/2R}$, the correction for sphericity is given by:

$$X = d^2/2R \quad \textbf{(2.49)}$$

4.2.3 Correction for refraction

This correction must be introduced to take into account the bending which the light ray experiences when passing through layers of the atmosphere of different density. Such bending tends always to result in bending downwards; therefore to a degree it accentuates the error sphericity.

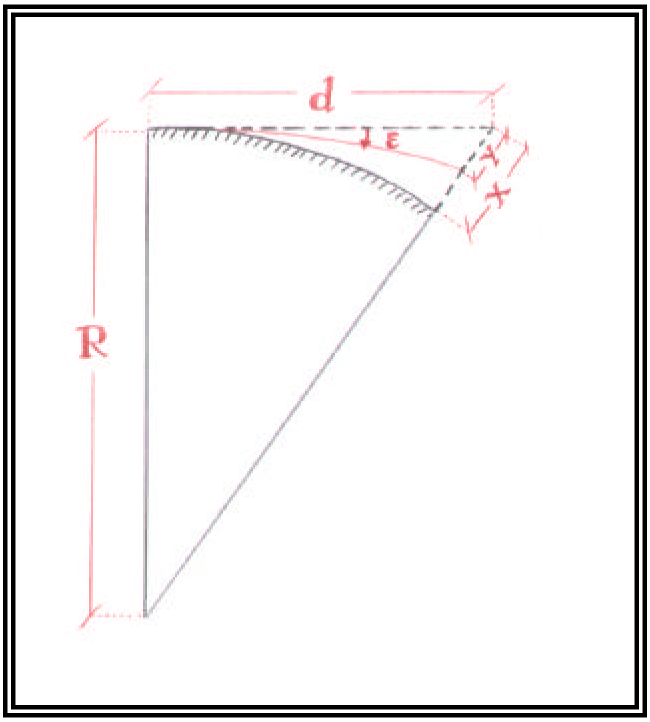


Fig. 2.19 "Correction of refraction"

where:

- X: is the correction for sphericity;
- Y: is the correction for refraction;
- R: is the ray of the adopted local sphere;
- e: is the dependent angle from the refraction coefficient K (˜ 0,14) [e = K d / 2R]

Assuming Y and e to be small, it is possible to write:

$$Y = d\,e \quad \textbf{(2.50)}$$

and therefore, replacing the expression of e in the (2.14), see 3.2.2.3 of Chapter 2, we can express it as:

$$Y = K\,d^2/2R \quad \textbf{(2.51)}$$

to this point the combination of the corrections of sphericity and refraction, identified in the qua ntity **(X – Y)**; it is possible to write as it follows:

$$(X-Y)=(1-K)d^2/2R \qquad \textbf{(2.52)}$$

4.2.4 Correction of height

The correction for height derives from the fact that the measured distance does not equate with the horizontal distance, which represents the quantity to use in the (2.47), see 4.2.1 of Chapter 2.

The relationship between '$\mathbf{d_{obl}}$' the oblique distance (measured) and '$\mathbf{d_{hor}}$' the horizontal distance is defined by:

$$d_{hor}=d_{obl}\cdot(1+Q_m/R) \qquad \textbf{(2.53)}$$

where $\mathbf{Q_m}$ represents the arithmetic average between the heights of the two points.

In summary, the formula to be adopted for trigonometric levelling from either end, taking into account of the three corrections described, is:

$$\Delta_{AB}=d_{obl}\cdot(1+Q_m/R)\cdot \mathrm{cotg}\, j_A+(1-K)\cdot d^2/2R+h-l \qquad \textbf{(2.54)}$$

adopting this approach, the weakness is the forecasting of the **K** coefficient of refraction, particularly for distances greater than 10 kms.

To remove this, the technique of simultaneous reciprocal trigonometric levelling can be employed, where two teams simultaneously measuring the two zenithal angles and the two oblique distances from the selected points. Two equations with two unknown are produced: $\mathbf{?_{AB}}$ and **K**. In this way it is no longer necessary to forecast **K**.

4.2.5 Sources of error

Because it is possible to consider the errors in the measurement of '**h**' and '**l**' as negligible, as well as the error for the mean height over of the distance (always less than the errors to those of the trigonometric levelling made over larger distances), for an analysis of the precision of this the simple formula can be used:

$$\Delta_{AB}=d_{or}\cdot \mathrm{cotg}\, j_A+(1-K)\cdot d^2/2R \qquad \textbf{(2.55)}$$

from the theory of errors, the $\mathbf{m_H}$ mean error (in this case a non linear function) of the $\mathbf{?_{AB}}$ difference of elevation will be:

$$m_H=\pm\sqrt{\left[\left(\frac{\partial\Delta_{AB}}{\partial d}\right)^2\cdot m_d^2+\left(\frac{\partial\Delta_{AB}}{\partial j_A}\right)^2\cdot m_{j_A}^2+\left(\frac{\partial\Delta_{AB}}{\partial K}\right)^2\cdot m_K^2\right]} \qquad \textbf{(2.56)}$$

in which $\mathbf{m_d}$, $\mathbf{m_{f\,A}}$ and $\mathbf{m_K}$ are respectively the mean errors of the distance, the zenithal angle and the coefficient of refraction. With the differentiation related to **d**, $\mathbf{f_A}$ and **K**, it is obtained that:

$$\frac{\partial \Delta_{AB}}{\partial d} = \cot g\, j_{A} + (1 - K) \cdot \frac{d}{R} \, (\text{with the second negligibleterm})$$

$$\frac{\partial \Delta_{AB}}{\partial j_{A}} = -\frac{d}{\mathrm{sen}^{2} j_{A}} \qquad \textbf{(2.57)}$$

$$\frac{\partial \Delta_{AB}}{\partial K} = -\frac{d^{2}}{2 \cdot R}$$

Analysing the three rooted terms in (2.56), it can be said:

a. in the first term, assuming as the mean error of distance the value of 1/50000 (2 cms for km), the error in the difference of elevation will depend on a , the angle of inclination, (a = 90° – f_A). With α=0°, the error removes itself. It is however always small (i.e. for α = ±10° and d = 5km, the error will be 1.6 cms).
b. in the second term, assigning to **a** a mean value of 10°, the error will depend on the mean error of **f_A,** the zenithal angle, and from **d**, the distance, (i.e. m·A = ±10° and d = 5 kms, the resultant error is 12.1 cms).
c. in the third term, the error is a function of the centre error of **K** and in this case, of the square of the distance (i.e. for m·K = ± 0,015 and d = 5 kms, the resultant error is 2.9 cms).

From such analysis it is evident that the greatest influence comes from the errors in the measurement of the zenithal angles. Thus, the angular measurements should always be undertaken from the two reciprocal faces of the instrument with the aim of compensating for the errors in instrument zenith. As a rule it is preferable to carry out the measurements when the **K** coefficient of refraction is more stable, which is around midday, even if at these times, due the sun's heat, the images appear less stable; this problem is overcome by taking the average of more measurements.

Nevertheless, for distances over some kilometres, the mean errors in the differences of elevation can be considered proportional to the distances themselves.

4.2.6 Computation and Compensation

In the theory of the errors, the weights of the measurements to be introduced in a calculation for adjustment are assumed proportional to the inverse of the squares of the mean errors of the measurements themselves. In this case being proportional to the distances, the weights to attribute to the several different compensating elevations are inversely assumed proportional to the squares of the distances. It is only worth considering for trigonometric levelling of medium and long distances, they are normally applied when undertaking the trigonometric networks of expansion. The trigonometric levelling over short distances involves detailed surveys and exploits the principle of the tacheometric (or tachymetric) levelling.

The procedures of adjustment are entirely comparable with those related to geometric levelling, with the only difference regarding the weights. It has to remember that, given the reliability of the trigonometric levelling for kilometric distances notably less than that of the geometric levelling, it is acceptable to conduct empirical adjustments.

4.3 Altimetry with GPS (GNSS Vertical Control Method)

GPS (exploiting the relative positioning) generates the base-line components between the surveyed positions, from which the XYZ geocentric co-ordinates are obtained in the WGS84 reference system. The φ, λ & **h** ellipsoidal co-ordinates are obtained with transformation formulae.

However in cartography the orthometric heights **H** are related to the surface of the Geoid and not the ellipsoid. Therefore it is important to know the Geoid undulation or its variation at a known points **H** and **h**. Only in small areas (< 10 km) and for cartographic purposes, can the Geoid be approximated to a horizontal plane.

For larger areas it is necessary to use global models of the Geoid; different global models (i.e. OSU91A, EGM96) are available in the processing software of GPS data and in the receivers. However these partially contain the effects of the distribution of local masses. Each national local estimate of the Geoid is performed by gravimetric measurements. The interpolations of these models produce values of the undulation **N**, necessary for orthometric height determination.

These local Geoids are gravimetric and independent from Geoid undulation values obtained from combining GPS and geometric levelling observations; they are estimated in a geocentric reference that does not coincide with WGS84 but introduces slight differences in origin of the geocentric axis term and of orientation of the axes of the reference system.

Therefore between the two reference systems it is necessary to conduct a transformation termed 'locating of the Geoid'.

To calculate this transformation, start from the orthometric height values **H** of some GPS positions, obtained via geometric levelling operations, and the $\mathbf{N_{WGS84}}$ experimental undulation is evaluate starting with the ellipsoidal height **h** derived by the GPS net compensation.
The effect of location d**N** is:

$$dN = N_{WGS84} - N_{localgeoid} \quad \textbf{(2.58)}$$

with $N_{WGS84} = h - H$.

The datum transformation in the strictest sense is a spatial rotation and translation with scale variation, but in small areas the altimetric part can be separated, estimating the parameters of the equation of a plane starting from d**N** values for at least of three points of which the height is known in both reference systems, with the following expression:

$$a_1 X_i + a_2 Y_i + a_3 = dN_i \quad \textbf{(2.59)}$$

with $\mathbf{X_i}$ and $\mathbf{Y_i}$ being the cartographic co-ordinates of the points for which the heights are twice measured and $\mathbf{a_1}$ $\mathbf{a_2}$ $\mathbf{a_3}$ parameters of the plane to be estimated. This plane describes the difference in Datum between N_{WGS84} and $N_{localgeoid}$. The three unknown parameters can be estimated to the minimum squares if the number of the points with double heights is greater than three.

5. INSTRUMENTS USED TO ESTABLISH HORIZONTAL AND VERTICAL CONTROL

5.1 GNSS Receiver (Global Positioning System)

GPS receivers can be classified according to the measurements they are able to acquire and the accuracy of the final positioning, as will be seen later in the paragraph 6.2:

a. Measures of code receivers: are able to acquire only the transmitted C/A component of the signal. They are often termed 'hand-held' due to the very small size of the receivers; some can receive a differential correction (in line with the standard protocol RTCM - 104) to improve the positioning precision. Their exclusive employment is for navigation.
b. Single frequency receivers: in addition to the code C/A, they can also acquire the L1 carrier phase. They perform positioning with measurements of the code or phase on L1 in absolute, relative or differential mode.
c. P-code double frequency receivers: are the most capable available in the market and can acquire all parts of the signal (L1, L2, C/A, P). They perform positioning with measurements of the code or phase on L1 and L2 (absolute, relative or differential). Thus they can be employed for all varieties of static and kinematic positioning. They are particularly suitable for the technique of dynamic initialisation 'On The Fly' (OTF).
d. Y-code double frequency receivers: identical to the P-code category, but they can also acquire the P-code using Anti-Spoofing (A/S).

5.2 Electronic instruments

The measurement of distances using electromagnetic wave distance measuring systems has undergone notable developments in the past few years; increasingly the producers of topographical instruments are including electromagnetic wave distance measuring devices in their theodolites. These systems, internationally termed EDM (Electronic Distance Measuring equipment) or DME (Distance Measuring Equipment)[1], operate in two different ways:

a. measurement of phase;
b. measurement of impulses.

5.2.1 Electronic Distance Measuring of phase

These instruments are based on the theory of the propagation of electromagnetic waves. They propagate using the sine rule, with speed equal to that of light in the air ($\mathbf{c_{air}}$), which is slightly inferior to that in a void, being equal to the relationship between the speed in the void ($\mathbf{c_0}$) and the index of refraction ($n_{\mathbf{air}}$) of the air which depends on temperature, pressure and humidity: $\{\mathbf{c_{air}=c_0}/n_{\mathbf{air}}\mathbf{(t,p,h)}\}$

These electronic distance measuring equipment are made up of three distinct parts: transmitter, reflector and receiver; the first and the last parts are contained together in the equipment set-up at the occupied station, the reflector is separate and is placed on the point the distance to which is to be determined.

The transmitter produces a signal at a previously established frequency; the reflector amplifies and reflects the signal, which is received by a phase discriminator capable of determining the phase difference between the transmitted and the received signals with an order of precision of a hundredth of radian.

[1] (IHO S-32 – fifth edition 1994, # 1406 and # 1576)

Since the signal has covered the distance between the two points twice, there and back, this double distance could easily be calculated if it were possible to determine the number of integer cycles which have passed between the transmission and the receipt of the signal. Being unable to determine this number of integer cycles, which is called the ambiguity, the electromagnetic distance measurements of phase use three different techniques to get round this problem:

a. the modulation for ten;
b. the method of the three frequencies;
c. the frequency modulation of the signal.

5.2.1.1 The modulation for ten

With this technique two or more signals are sent in sequence with different frequencies, varying multiples of 10 (hence the name), in order to measure the distance by the phase difference.

The first sent signal has a wavelength greater than double the range of the equipment. In this way the distance can be determined without ambiguity with the equation:

$$d = \left(\frac{\lambda}{2}\right) \cdot \left(\frac{\Delta\varphi}{2\pi}\right) \qquad \textbf{(2.60)}$$

where **d** represents the half double distance.

However, with this method, the distance is determined with low accuracy; if the range of the EDM from the target was 1 km, the signal would have a wavelength of at least 2 km, then, the distance would be measured with a precision equal to 1.59 m, applying (2.60) with the precision of the phase discriminator to 1/100 of radian. Such an error is obviously unsupportable in the measurement of the distances over a range of 1 km. To remove this problem, after the transmission of the first signal and the calculation of a first approximate value for the distance, a second signal is transmitted, with a wavelength equal to 1/100 of the previous signal. In this case, the determination of the distance requires the definition of the phase ambiguity, this is possible having already approximated the distance between the two points with sufficient precision to calculate it. In this way, the value of the distance is improved 100 times and the precision achieves, in the above case, a value of 1.6 cms, which could be considered acceptable. It is possible to transmit another signal of a wavelength equal to 1/100 of the second one, thus improving the precision to a few millimetres.

5.2.1.2 Method of the three frequencies

It comprises a variation on the previous method, by using two near equal frequencies with wavelengths of the order of the range of the equipment, which allows the determination of a first approximation of the distance. A third frequency with a wavelength very much smaller than the first two, enables the fine determination of the distance.

5.2.1.3 Variation of frequency

With this technique the frequency of the transmitted signal, starting at a set value, is increased (or decreased) until a zero phase difference is achieved between the transmitted and received signal. The determination of the distance could be calculated with an equation in which the number of cycles remains unknown, however, by continuing to increase the frequency (and therefore decreasing the wavelength); a zero phase difference will again be produced between the transmitted and received signals, when the number of integer cycles will be increased by a whole number. At this point, from the combination of the two equations (corresponding to the two values of wavelength) the phase ambiguity can be resolved.

In the first two techniques (modulation for ten and method of the three frequencies), the determination of the phase difference is necessary. This can be achieved through a phase discriminator composed of a transformer of sine waves, so that square waves (analogue-digital transformer) are transmitted and received, and using a counter of the time when the square waves are both positive and negative. This time is turned into a value of distance. Clearly, to increase the precision of the measure, this calculation is repeated thousands of times but it takes a few seconds to complete the measurement.

Recently some DME of phase have been produced without with discriminator. They use a mathematical correlation between the transmitted and received signals for determining the phase difference, enabling the achievement of greater precisions in the measurements of distance. According to the produced frequency, the phase DME can be classified as:

a. MDM (Microwave Distance Measurement);
b. EODM (Electro-Optical Distance Measurement) or geodimeters.

The first group use frequencies in the order of the 30 MHz (wavelengths centimetric), they are employed for determining long distances; in these instruments the reflector is active, that is it is capable of amplifying the received signal and reflecting it with greater power.

The requirement to alter the frequency of the transmitted signal involves some consideration of and allowance for the propagation of electromagnetic waves through the atmosphere. In fact only some ranges of frequency are capable passing through the atmosphere without large losses of power. Infrared rays (micrometric wavelengths), which require a limited consumption of energy power supply, are not overly influenced by the solar light, they are used for the determination of distances of 2-3 kilometres; the centimetric waves, termed Hertzian microwaves, which have wavelengths of a few centimetres, are also used for the determination of highly elevated distances, also in presence of fogs or precipitations, these require a significant power supply. If the signal has wavelengths in the visible range, wavelengths included between 0.3 and 1 micro-metre, the waves are created with specific optic systems and reflected with simple mirrors or prisms. For practical reasons, there is therefore the demand to emit very short waves from few centimetres in the MDMs to few tenth of micron in the geodimeters. This demand, however, is not reconciled with the need to emit waves with lengths in the order of metres to determine the fine value of the distance or waves of some kilometres to determine the first approximate value.

These two demands are satisfied by resorting to the frequency modulation in the MDMs or to the amplitude modulation in the geodimeters.

In the geodimeters, the wavelength of the carrier signal is constant and it assumes values of the order of a few microns (satisfying the first demand), while the wavelength modulated assumes varying values from a few metres to some kilometres (satisfying the second demand).

5.2.2 Electronic Distance Measuring of impulses

The operating principle of EDM, recently introduced into topographic surveying, is based on the measure of the time taken by a bright impulse to go from the distance meter to the reflector and back.

The same principle is used, for instance, by a particular system for measuring satellite altimetry, termed SLR (Satellite Laser Ranging), in which a Laser impulse is reflected back by an artificial reflecting satellite. The evolution of electronic systems has enabled the employment of these methods in topographical EDM, obtaining performances superior to those of phase EDM.

A diode light beam transmitting laser is excited for a short time interval. The exact measure of the time "**t**" between the transmission of the impulse and the following receipt would be enough to determine the distance:

$$d = \frac{v \cdot t}{2} \tag{2.61}$$

However the measurement of the time is made with certain errors. A time interval of 10^{-8} seconds (typical of a quartz clock) is enough for the bright impulse to cover 3 metres; this is not acceptable for an EDM. It is therefore necessary for a refinement in the measurement of the time, obtained by determining the fraction of the period of oscillation of the clock between the departure of the impulse and its receipt:

$$t = n \cdot T + t_A - t_B \tag{2.62}$$

where **T** is the period of the clock, **n** is the number of periods and therefore ***n*T** is the measurement of the time directly produced by the clock; $\mathbf{t_A}$ is the time between the transmission of the signal and the start of the clock oscillation and $\mathbf{t_B}$ is the time spent between the receipt of the signal to the completion of the final clock oscillation. To determine these two fractions of time, the voltage with which the laser diode is excited is gradually supplied in a linear manner; then, by determining the voltage $\mathbf{V_T}$ which would be used for a complete oscillation of the clock, the two fractions $\mathbf{t_A}$ and $\mathbf{t_B}$ can be calculated with a simple proportion:

$$t_A : V_A = t_B : V_A = T : V_T \tag{2.63}$$

where $\mathbf{V_A}$ and $\mathbf{V_B}$ are the voltages respectively supplied to the heads of the diode in the time $\mathbf{t_A}$ and $\mathbf{t_B}$.

In theory it would be enough for only one impulse to determine the distance; in practice thousands of impulses are transmitted to increasing the precision. Some EDM systems transmit up to 2000 impulses per second, employing 0.8 sec (1600 impulses) to achieve a standard error of 5mm + 1mm/km and 3 sec (6000 impulses) to obtain a standard error of 3mm + 1mm/km.

The many advantages of this method in comparison to that of the measurement of phase are evident:

a. It requires less time to take the measurements; after a few impulses (few milliseconds) a centimetric precision is obtained on the measurement of the distance, while the EDM of phase generally requires a few whole seconds. The ability to very quickly take measurements is useful when determining the distance of a moving point (and therefore in bathymetric surveys);
b. The signal can also be returned with weak power, because a small voltage is sufficient to stop the clock and complete the relevant time calculation. This allows notable increases in the range of the distance meter for equivalent intensities of the transmitted signal. In terms of power supply, the transmission of impulses is more economic than a continuous transmission of the carrier signal (greater battery life);

c. It is possible to obtain EDM which do not need reflectors to produce a signal return. These equipment have ranges strongly influenced by the quality and the colour of the reflecting surface, they do not operate over ranges of more 200-300 metres and they can achieve precisions of 5-10 mms. They are very useful for the measurement of distances of inaccessible points;
d. The quality of the measurement is not heavily influenced by environmental factors (temperature, pressure or humidity) as in EDM of phase measurement.

Besides these advantages, generally it is the higher cost of EDM of impulses which needs to be considered; probably justifiable only in the case where it is necessary to frequently measure distances of over 1 km.

5.2.3 Precision and range of EDM

Generally, EODMs, or geodimeters, use infrared waves, rarely waves included in the visible spectrum (with wavelengths in the order of 1-5 micrometres), or laser waves; in this equipment the reflector is passive, being made up of one or more three-squared prisms which reflect the signal parallel to the incidental ray. Increasing the number of prisms of the reflector increases the corresponding the range of the geodimeter, which can reach 4 or 5 kilometres.

The precision of EDM waves depends on numerous factors, presently it has reached comparable levels with that obtainable with wires of INVAR.

An important element of EDM is comprised by the oscillator, on whose stability the precision of the equipment depends. In fact, the frequency of the oscillator is a function of temperature; the law of variation of frequency as temperature varies must be memorised in the EDM, in order to be able to apply the appropriate corrections, which can reach 3-5 ppms for 20°C of temperature variation.

It is required to consider the atmospheric refraction which directly influences the wavelength of the transmitted signals. The effect of refraction depends on the values of temperature and atmospheric pressure which have to be inserted into the system, which then calculates, according to an empirical formula, the corrections to be applied in ppm to the measured distance. In other cases the builders provide some tables, through which the correction to apply to the distance can directly be determined, knowing the values of temperature and pressure. It is useful to remember that, in the first approximation, a correction of 1 ppm can derive from a variation of 1°C of temperature, of 3.5 hectopascals of atmospheric pressure or of 25 hectopascals of the partial pressure of the humidity of the air.

The aging of the equipment causes a variation of the nominal rated frequency of the oscillator, which can reach values of some ppms after 2-3 years of life. It is necessary, therefore, to have the system periodically re-calibration.

Finally, for determining the distance it is necessary to consider the instrument constant, termed the prism constant, because, generally the centre of reflecting surface of the prism is not coincident with the centre of the reflector. Such a constant is created by the reflectors and needs to be memorized in the EDM for every combination of prisms used.

As far as it affects the range of EDM, besides being a characteristic of this type of system, it also depends on atmospheric conditions and on the number of prisms being used. As previously stated, with the same power supply, the EDMs using impulses have greater ranges than those measuring by phase, they can achieve, under optimal atmospheric conditions, distances of 15 kilometres.

It should be noted that atmospheric conditions are considered:

a. unfavourable: a lot of haze or intense sun with strong refraction;
b. mean: light haze or veiled sun
c. good: no haze and cloudy sky

It is evident therefore, that the nominal precision declared by the builders of EDM is achievable only if all the factors which can influence the measurements are considered. In general phase EDM enables the achievement, without particular acumen, precisions of the order of s **= 5mm + 5ppm**

5.2.4 Total Stations

The collocating of an EDM and an electronic theodolite can be extremely productive, because it is possible to integrate the data coming from the distance meter with the angular measurements obtained with the theodolite. Thus, it is possible to immediately calculate other quantities, indirectly obtained, such as horizontal distances or rectangular co-ordinates, etc.

The collocated theodolite-EDM is called a Total Station or integrated station, as it enables singularly to obtain all the measurements for topographic surveying such as angles, distances, co-ordinates etc.

The surveyed data can be registered in a ‘field book’, but due to their digital nature, the data can be stored on magnetic media or the solid state memory. Thus the possible transcription errors of the operator are avoided and measurement operations are accelerated.

The inspirational principle of these systems is to automate the most repetitive operations of topographic surveying such as angular and distance readings, data recording, the input of station details, etc.

5.3 Optical instruments

5.3.1 Marine sextant (Circle to reflection)

The circle to reflection is a tool specifically built for the measure of horizontal angles between two objects. The precision of marine sextant in the measurement of angles varies from 20" to 10’.

It is a system of reflection and the measurement of the angles is based, as for the sextant, on the theory of the optics of the double reflection of a bright ray, with the difference that, in the circle, prisms are employed instead of mirrors.

The two prisms are set one sideways, the other higher and at the centre of a circular box provided with a handle. The prism at the centre is mobile and has two fins which limit the bright rays picked up by the prism to those which are reflected by the hypotenuse of the prism. The other prism is fixed and it is set at such a height from the plane of the box to cover only the interior half of the field of the telescope.

The telescope is fixed in such a way that the direct images of the objects appears in the upper part of its field and in the lower half for those reflected by the small prism. Inside the box a graduated circle is contained, fixed to the large prism, therefore rotating with respect to an index marked on the box.

To allow the equipment to work correctly, it is necessary that the two fundamental operation conditions of double reflection goniometers are respected, that is, in the case of prisms, they are exactly perpendicular to the plane of the box of the instrument and, when the two hypotenuse are parallel, the index marks 0° on the graduation vernier.

The large prism must not be able to move, except for rotation around its pivot; it is considered by construction perpendicular to the box. The perpendicularity of the small prism to the box can be adjusted by a screw. Set the vernier to 0°, if a distant object is observed through the telescope and the two parts of the image are seen to perfectly lined up vertically, the direct one above and the reflected one below, it indicates that the tool is perfectly rectified.

The parallelism between the hypotenuses of the prisms can be corrected with a special screw which makes the small prism rotate around a normal axis to the plane of the box.

5.3.2 Theodolites

The theodolite is an instrument which measures azimuth angles, via a graduated horizontal circle, and zenithal angles, via a graduated vertical circle.

Precision of theodolites in the measurement of angles varies from 0.1" to 10"; the tacheometris (or tachymetris) are differentiated from theodolites due to their lower obtainable precisions, from 10" to 10', in angular measurements.

In a theodolite three axes can be identified:

a. the primary axis, around which the alidade rotates;
b. the secondary axis, around which the telescope rotates;
c. the axis of collimation of the telescope.

The principal parts of a theodolite are:

a. the base (or tribrach), provided with a pedestal and three adjusting screws (levelling screws) on the basal plate which is the lowest part of the theodolite connected to the head of the tripod (or stand and legs), so as to be able, within certain limits, to centre the instrument over the reference mark. The lower spirit level or circular level (also termed universal level or bull's eye level) and the optic lead are anchored to it.

b. the alidade is a generally U-shaped frame, which can rotate around the vertical axis passing through the centre of the instrument (primary axis) and contains the engraved horizontal circle reading index. An upper spirit level (termed spirit bubble or sensitive bubble) is anchored to it to make the primary axis vertical and to set the origin of the zenith angles to the zenith, residual errors excepted.

c. the graduated horizontal circle, situated above the pedestal and under the alidade.

d. the telescope, hinged to the alidade so that its axis of collimation is perpendicular to its axis of rotation. The telescope has a magnification from 28 to 45 times, thus increasing the precision of the measurements.

e. the vertical circle, rigidly connected to the telescope, for reading of the zenithal angles

The theodolite can be of two types, depending on the system of lock used for the horizontal circle: repeating and reiterating.

a. The repeating theodolites (fig. 2.20) are those which allow the fixing of the horizontal circle both to the plinth and to the alidade through two separate screws. When both the locking screws are operating, the horizontal circle is fixed both to the plinth and to the alidade, so that the instrument cannot rotate around the primary axis.

b. In the reiterating theodolites (fig. 2.20), the horizontal circle remains independent from the plinth and from the alidade; it can rotate with the plinth through a special screw, usually protected against accidental manoeuvres. The alidade is locked to the plinth through a locking screw, together with a screw for minor movements.

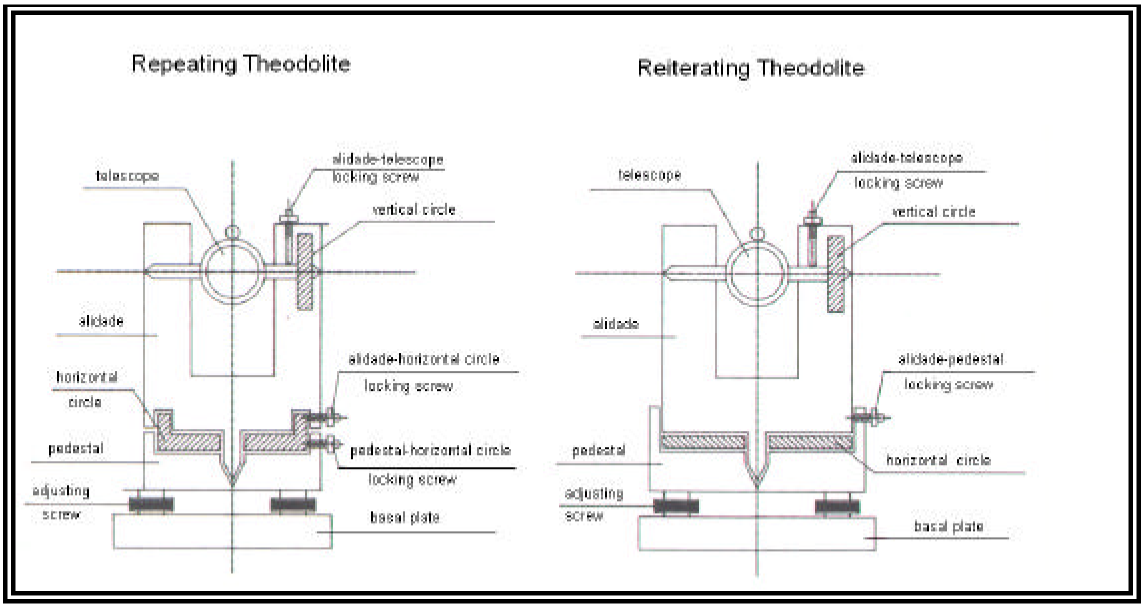


Fig. 2.20 "Theodolites"

Before obtaining any angular measurements, it is essential to verify that between the principal axes (primary, secondary and of coincidence) and other parts of the instrument some conditions of precision are achieved. Some of these are directly verified by the instrument builder, conditions of construction, and if the instrument is used with care, they can always be considered unaltered and therefore satisfied.

Some conditions, termed conditions of rectification, must be directly verified by the operator before beginning every measurement session. In particular one consists of establishing the verticality of the primary axis; this is achieved by using the spirit level, which is more sensitive than the circula r one used for precisely centring the principal axis of the instrument over the reference mark. To use the spirit level the alidade has to be rotated until the level is inline with the direction of two adjusting screws and, using them, the spirit bubble has to be centred. The level is correct, when by rotating the alidade 180° the spirit bubble remains centred; if not it will be necessary to use the rectification screw and the two adjusting screws. Final stage for correct levelling, the instrument is rotated 90° and using the third adjusting screw to centre the spirit bubble.

The other adjustment, normally only required when observers change, is to ensure that the telescope is correctly focused. This is achieved using the focusing ring on the telescope to ensure the reticle (or reticule) lines appear clear and sharp.

5.3.3 Levelling instruments (Levels) and Stadia

The levelling instrument (or level) is an instrument which allows the creation of an axis of horizontal collimation and it is used in geometric levelling. Modern levels are divided into:

a. Fixed Levels and Self-aligning Levels;
b. Digital Levels;
c. Laser Levels.

Having chosen the type of level, and thus defined the mechanism for reading, it is necessary to choose a stadia rod or a levelling rod or staff, whose principle of graduation connects it to the level. Levels with fixed or tilting telescopes have been made obsolete by modern digital and laser levels.

5.3.3.1 Fixed Telescope Level (Dumpy Level)

It consists of a telescope which forms a single unit with the pivot of rotation and with its base, similar to that of the theodolite. A spirit level is fixed to the telescope which allows the instrument to be levelled in position, in a similar fashion to a theodolite. Once the spirit level is centred in the two orthogonal directions, the level can be employed for determining the difference of elevation in any direction.

A condition periodically checked, is that the axis of the spirit level is parallel to the axis of collimation. To check the instrument, all that is required is to measure an already known difference of elevation between two points with the level in the middle and to move the reticle of the telescope with the special screw until the reading on the stadia is correct.

5.3.3.2 Fixed Telescope Level with Elevation Screw (Dumpy Level)

In these levels the telescope is not rigidly connected to the pivot of rotation but through a crossroad hinged at one end and connected at the other by a screw, called the elevation screw (or micrometer screw). The elevation screw allows the telescope to rotate, through a very small vertical angle; this enables a horizontal the line of sight to be achieved even if the primary axis is not vertical. These levels have a spherical level attached to the base, which, when centred, approximately makes the primary axis vertical. For each sight, it is necessary to use the elevation screw until the spirit level attached to the telescope is centred, thus making the axis of collimation horizontal.

5.3.3.3 Rotating Telescope Level (Y – Level)

In these levels the telescope can rotate through a vertical angle (180°) inside a muff connected solidly to the pivot of rotation. Attached to the telescope is a reversing spirit level with a double bend, which allows it to work even if it is turned upside-down. In these instruments there are therefore two axis of the level: the axis of rotation of the telescope (which coincides with the axis of the muff) and the axis of collimation. In the assumption that the two axes of the levels are parallel and that the axis of the muff coincides with the axis of collimation, two readings are made on the stadia, corresponding to the two extreme positions which the telescope can assume, each time centring the spirit level with the elevation screw.

Using the arithmetic average of the two readings, any error between the axis of the level and the axis of collimation is compensated, because the error is of opposite sign in the two readings obtained.

5.3.3.4 Self-aligning Level

In these instruments the axis of collimation is automatically made horizontal by an internal system, independently from the verticality of the primary axis. Since such systems, termed compensators, work within certain limits of rotation of the telescope, of the order of 10', the self-aligning levels are fitted with a circular level, which once centred, guarantees the correct operation of the instrument. Compensators, the actual design of which is different for each manufacturer, normally constitute a sensitive prism element suspended on a pendulum which uses the principle that the strength of gravity will create a horizontal line of sight.

5.3.3.5 Digital Level

These levels are similar to self-aligning levels but the reading on the levelling rod is made automatically, though it is possible for the traditional optic reading, in case of malfunction of the electronic parts or exhaustion of the batteries.

The stadia used with this type of levels are specific; on one side they have a graduation as with normal stadia, on the other side they have bar code graduations. The image of the bar code from the levelling rod is separately transmitted to the ocular sight to allow reading of the levelling rod and to an electronic survey system. The digital signal is decoded through a microprocessor which is able to produce, besides the difference of elevation, the horizontal distance between the two points.

The advantages introduced by these systems come from the ability to automatically record the survey data, with a considerable saving in time and with the total elimination of blunders during transcription. Correct operation is only guaranteed under good light conditions, that is measurements are performed in the open air. The precision of these levels is of 0.1 mm for the difference of elevations and of 1 cm for the distances.

5.3.3.6 Laser Level

These levels use the transmission of a laser beam which matches the line of sight of the telescope. Some of these instruments, which are normally self-aligning levels, do not require operator intervention. Once the equipment has been placed at the station with the aid of a circular level, a motor makes the laser beam continuously rotate, through a switch prism; in this way only one operator is required to perform levelling within a field of 200-300 m of the ray.

The levelling rods used for these levels have a sensor, decimetres in length, which can move on the stadia. When the laser beam hits the sensor, the value corresponding to the ray can be read and automatically recorded.

The precision of the measurements can be less than a millimetre, the system is ideal for the radial levelling.

6. POSITIONING METHODS (TECHNIQUES OF POSITIONING)

6.1 GNSS (GPS)

6.1.1 Description of Global Positioning System (GPS)

The GPS positioning system is based on the receipt of radio signals sent from an artificial satellite constellation in orbit around the earth, it is a real-time, all-weather, 24-hour, worldwide, 3-dimensional absolute satellite-based positioning system. The complete name of the system is NAVSTAR GPS which means NAVigation Satellite Timing And Ranging Global Positioning System. The system, created by the Department of the Defence in the United States, is currently managed in collaboration with the Department of the Commerce and has been projected to allow at every moment in every part of the world the three-dimensional positioning of objects, including whilst moving.

The system is divided into:

a. The spatial segment: is formed by 24 satellites in near circular orbit around the Earth at a height of about 20,200 kms. The satellites are distributed in groups of 4 about 6 circular orbits tilted 55° to the equatorial plane with a revolution period of about 12 hours. This constellation distribution ensures the visibility of at least 4 satellites (often 6 to 8) at all times and places with an elevation above 15° degrees from the horizon, which is fundamental for the positioning.

   The satellites have the followings functions:

   - To transmit information to users through a radio signal;
   - To maintain an accurate reference time due to the high degree of accuracy (from 10-12 to 10-14 sec) of the caesium and rubidium atomic clocks on board;
   - To receive and to store information from the control segment;
   - To make corrections to orbits.

   The satellites have been launched in different epochs, starting in 1978, in blocks which replaced earlier models with more advance ones.

b. Control segment: comprises 5 monitoring stations and an additional sixth at Sunnyvale, USA, where a copy of all the selected data and all the attached operations are preserved. Among the five stations, all of which are provided with meteorological stations for the evaluation of the troposphere affects on the radio signals sent by satellites, three stations (Ascension, Diego Garcia and Kwajalein) have the ability to send messages to the satellites and one (Colorado Springs, USA) is the Master station, where the necessary calculations for the determination of the new orbits are performed. In summary the tasks of the control segment are:

   - To continuously track the satellites and to process the received data for the calculation of the timing-space position (Ephemeris);
   - To check the general state of the system, in particular the satellite clocks;
   - To implement orbit corrections;
   - To upload new data to the satellites, including the forecast Ephemeris for the next 12 or 24 hours, which are then transmitted to users.

c. The user segment: is made up of users equipped with receivers with GPS antennas. These are passive systems in that they are able to acquire data without emitting some signal. Various types of receivers exist depending on the strategy used to analyze the received signal and the required positional accuracy.

d. The signal structure: Every satellite continually emits electromagnetic waves on carefully chosen frequencies to a very small sector on the earth's surface and is thus relatively sheltered from interference. These carrier waves transport the information to the user through code modulation. The onboard clocks produce a primary frequency f0 = 1.023 MHz; from this primary frequency the three fundamental parts of the GPS signal are precisely originated:

- *Carrying Component*
  It is made up of two sinusoidal waves called L1 and L2 respectively of frequency 154 x f0 = 1575.42MHz ($\lambda L1 \cong 19$ cm) and 120 x f0 = 1227.60 MHz ($\lambda L2 \cong 24$ cm).

- *Impulsive Component*
  It comprises two codes called Coarse Acquisition (C/A) and Precision (P), the former only modulates the L1 carrier frequency and the later both the L1 and the L2.

- Such codes are square waves formed by transitions of values +1 and –1 produced by a simple algorithm, which has as a characteristic the statistic balancing of positive and negative values; the codes are called "pseudo accidental" or PRN (Pseudo Random Noise). The C/A code frequency is 1.023 MHz (C/A≅300m), the P code has a frequency that is 10 x f0 = 10.23 MHz (P≅30 m). The C/A code is available for civil use while the P code is reserved to military use and other authorized users. The DoD USA have reserve the right to disguise the P code by encryption and using the so-called Anti-Spoofing (A–S) procedure. The encrypted P code is called Y-code.

- *Message Component*
  It is composed of the navigation D message which has a frequency f0/204800 = 50 Hz. It contains the ephemeris (or almanac) details of the satellites, information on their health and the onboard clocks.

6.1.2 Principles of positioning

GPS positioning uses the technique of Spatial Measurement Intersection. The geodetic reference system (Datum) exploited is called World Geodetic System 1984 (WGS84), which is created from a clockwise Cartesian axis rotation with the origin at the earth's centre of mass, with which the geocentric ellipsoid WGS84 is associated. If the satellite co-ordinates in this reference system are known, the unknown co-ordinates of a point are connected to the known co-ordinates of the observed satellites through the measurement of a sufficient number of distances between the satellites and the centre of phase of an antenna connected to the receiver at the required position. Essentially there are three principles of positioning:

a. Absolute positioning (or normal);
b. Relative positioning;
c. Differential positioning.

6.1.2.1 Absolute positioning

The aim of this method of positioning is the determination of positional co-ordinates in the WGS84 global reference system. This is achieved by using the signal's impulsive component (C/A code or P code if available) or to analyze the two carrier phases L1 and L2.

In the first case, the satellite-receiver distances are called 'pseudo-ranges' and they are calculated according to the flight time which is the time the signal takes to reach the receiver from the satellite. This time is measured by the receiver through correlations between the received signal and a copy produced by the receiver; the copy signal in the receiver is shifted in order to line it up with the satellite signal. The calculated time difference is influenced by the asynchronous error between the satellite and receiver clocks, in addition to the drift of the receiver clock, which is less accurate than the atomic clocks of the satellites.

These factors cannot be ignored in the measurement of flight time; it is for this reason that to the 3 clock unknown quantities of point position (transformable Cartesian x, y, z in φ , λ and height on the ellipsoid WGS84) it adds a fourth category, which identifies the receiver clock errors. From this it follows that there is a requirement to simultaneously observe a least four satellites to obtain an absolute position in real time.

In the second case the phase of the two carrier frequencies is analyzed and the satellite-receiver distance can be obtained by comparing the phase of the carrier signal at the moment of reception with the phase of the signal at the moment of transmission. In this case an additional unknown quantity for every observed satellite is introduced; it is the Initial Integer Ambiguity which is the integer of cycles the signal has traversed from the satellite to the receiver at the beginning measurement. Thus to every new observed satellite a corresponding new Ambiguity is created, due to the different distances. As a result, absolute positioning in real time with phase measurements is only possible if the Ambiguities of the satellites used for positioning are known; the procedure for this determination is called initialisation.

6.1.2.2 Relative positioning

The aim of relative positioning is the determination of the base-line vector or of the vector components which ties the two positions on which temporarily the two receivers are located. If the absolute co-ordinates of one of the two points are known, adding the components of the base-line vector, the absolute co-ordinates of the second position can be obtained. Such positioning can be achieved through measurements of code or phase, although in practise only phase measurement is used. A phase observation equation can be written for every receiver from which a satellite is observed at a given moment. Observing the same satellite at the same moment from two different receivers at the ends of the base-line and then subtracting one from the other produces two equations of phase, an equation to the simple differences. Inserting into the observation another satellite, and adding the difference between the two equations to the simple differences, an equation to the double differences is created. At the end of these two operations the result is the elimination of the clock errors of the two satellites. At this point the unknown quantities to be determined are the components of the base-line vector and the sum of the four initial ambiguities of the two satellites (considered as only an integer value). If the signal is interrupted the ambiguities change and a new initialisation is required. Finally the possible interruptions of the signal are separated through the difference between two equations of the double differences (termed equation to the triple differences) and establishing the continuity, the unknown Ambiguity quantity is eliminated.

6.1.2.3 Differential positioning

Differential positioning is similar to absolute positioning but has corrections for pseudo-range in real time or in delayed time, transmitted or stored by receivers set on points of known absolute co-ordinates. The remote receiver applies, in real time or in delayed time, the corrections to the measurements of pseudo-range or phase effected and then calculates the correct absolute position, improving the accuracy of the co-ordinates.

6.1.3 Performances of the system and sources of error

In relationship to the different positioning principles, they are classified by the different degrees of accuracy:

a.	Absolute (SPS) with measurements of code C/A on L1:	10 to 30 metres
b.	Absolute (PPS) with measurements of code P (Y) on L1/L2:	5 to 15 metres
c.	Relative with measurements of static phase:	10-8 to 10-6 of the base-line
d.	Relative with measurements of phase (RTK):	decimetre
e.	Differential with code phase measurements (DGPS):	few metres
f.	Differential with carrier phase measurements (RTK DGPS):	few centimetres

The elements (error sources) which have the most influence on system performance are:

a. *Clock errors* of the satellites and the receivers (off-set and drift);

b. Orbit errors (imperfections in the Ephemeris data);

c. *Delays* during atmosphere signal passage because of ionosphere and troposphere refraction, whose affects on the signal are considerable due to the use of double frequency receivers;

d. *Tropospheric error*. Humidity is included in this error. Humidity can delay a time signal by up to approximately 3 m. Satellites low on the horizon will be transmitting signals across the surface of the earth through the troposphere; whilst those directly overhead will transmit through much less of the troposphere. Masking the horizon angle to 15° can minimise the tropospheric error. If this blocks too many satellites, a compromise down to 10° may be necessary. Manufacturers model the tropospheric delay in software; tests have determined that these tropospheric models work reasonably well.

e. *Ionospheric error*. Sun-spots and other electromagnetic phenomenon cause errors in GPS range measurements of up to 30 m during the day and as high as 6 m at night. The errors are not predictable but can be estimated. The ionospheric error is assumed to be the same at the reference receiver as at the vessel receiver. This assumption is sound for GPS networks where the stations are separated by a few nautical miles. Ionospheric models have been implemented for dual frequency receivers.

f. *Multi-path*. Multi-path is a reception of a reflected signal in lieu of a direct signal. The reflection can occur below or above the antenna. Multi-path magnitude is less over water than over land, but it is still present and always changing. The placement of the GPS receiver antenna should avoid areas where multi-path is more likely to occur (i.e. rock outcrops, metal roofs, commercial roof-mounted heating/air conditioning, buildings, cars, ships, etc.). Increasing the height of the antenna is one method of reducing multi-path at a reference station. The multi-path occurrence on a satellite range can last several minutes. Masking out satellite signals from the horizon up to 15° will also reduce multi-path.

g. *Geometric configuration* of satellites used for positioning, given by GDOP (Geometrical Dilution of Precision). The GDOP is divided for this purpose into some indices (PDOP and TDOP) which have been introduced to establish a degree of quality control. The most general is called PDOP (Position Dilution of Precision), inversely proportional to the 'goodness' of the configuration, which is divided into two components for control purposes:

the vertical or VDOP (Vertical Dilution of Precision) and the commonly used horizontal or HDOP (Horizontal Dilution of Precision); occasionally the ratio HDOP/PDOP is considered (for horizontal control see Chapter 7).

h. *Voluntary errors*, for reducing the pseudo-range measurement performance on the position data of satellites, can be introduced at the discretion of the system managers. Such procedures, called Selective Availability (S/A), produced an uncertainty in the positioning through the calculation of pseudo-range in the order of a 100 metres, this was removed 1st May 2000. Differential operation could eliminate S/A. Even with S/A set to zero, DGPS is still required for most hydrographic surveying applications.

6.1.4 GPS tracking and signal acquisition techniques

6.1.4.1 Tracking Techniques

Two general modes are basically used to determine the distance, or range, between a NAVSTAR GPS satellite and a ground-based receiver antenna. These measurements are made by satellite signal phase comparison techniques. The carrier frequency phase or the phase of a digital code modulated on the carrier phase may be tracked to resolve the distance between the satellite and the receiver. The resultant positional accuracy is dependent on the tracking method used.

These two-phase tracking techniques are:

- Carrier phase tracking;
- Code phase tracking.

The GPS satellites actually broadcast on two carrier frequencies: L1 at 1575.42 MHz (19-cm wavelength) and L2 at 1227.60 MHz (24-cm wavelength). Modulated on these frequencies are the Coarse Acquisition (C/A) (300-m wavelength) and the Precise (P) codes (30-m wavelength). In addition, a 50-bps satellite navigation message containing the satellite ephemeris and health status of each satellite is transmitted. The C/A and P codes are both present on the L1 frequency. Only the P code is present on the L2 frequency. The higher frequency of the carrier signal (L-Band) has a wavelength of 19 and 24 cm from which a distance can be resolved through post-processing software to approximately 2 mm. The modulating code has a wavelength of 300 m and will only yield distances accurate to about 1 m. Both of these tracking methods have application in hydrographic and conventional surveying.

6.1.4.2 Signal Acquisition Techniques

The procedures of acquisition have characteristics and different accuracies; they relate to different approaches of signal handling. These are described as:

a. ***Stand-Alone***: single point absolute position with pseudo-range in the WGS84 geodetic reference system. The absolute accuracy, to the 95% of level of confidence, is between 10 and 30 meters for SPS (Standard Positioning Service) and between 5 and 15 for PPS (Precise Positioning Service). The applications are only navigational.

b. ***Differential (DGPS)***: the differential corrections, calculated in a reference station of known position, are applied to the absolute position generated by a stand-alone receiver. These corrections of code or phase, as previously described, can be transmitted by radio or cellular phone, meeting the RTCM protocol, and applied in real time, or stored in the reference station and applied during the post processing with suitable software. The ranges and the accuracies are described in the table below:

Range Correction tracking	Distance among Stations	Accuracy
Code phase	Some hundred km	Few metres
Carrier phase	About ten km	Few centimetres

c. ***Relative***: the co-ordinates of the base-line vector ends, which connect the positions occupied by the stations, have to be determined. The calculation is achieved by post processing using the method of the double differences, correcting the data acquired on the phase of GPS signal in the base and mobile receiver. The principal methods are:

Technique	Application
Static	Reference frame of high accuracy
Rapid Static	Reference frame with less accuracy
Stop and go Kinematic	Fiducial points, survey in detail
Continue Kinematic	Trajectories, continuous monitoring

The time of acquisition and the sampling interval (termed rate) are the discriminators in the relative methods. The rate has to be a good compromise between the demand of measurements and the size of the file to process them. For example, for static applications with long acquisition periods, it is common to sample with a time interval of 15 or 30 seconds; for kinematic applications it is necessary to reduce this interval, often down to 1 second. This value represents the sampling limit interval in many receivers; currently receivers are able to make measurements with a frequency of 20Hz. Methods, techniques of acquisition and fields of use are summarized in the table below:

Method	Time of measure	Length bases	Accuracy	Rate (per sec)	Notes
Static	>1/2 hour 1 hour from 3 to 4 hours varying	 10 kms 20/30 kms >100 kms	 10^{-6} to 10^{-8} (of baseline length)	 15-60	doubles frequency if with bases (20 kms
Rapid Static	20-30 min(L1s) 6-8 mins (L2)	<10-15 kms	10^{-6} (of baseline length)	5-15	necessity for good satellite configuration
Stop and go Kinematic	<1min	some kms <10 kms	centimetric	1-5	needs continuous contact with satellites Initialisation: - up to 30 mins: L1 - 5/6 mins: L1 + L2 - On the Fly (OTF):L1+L2
Continuous Kinematic	Continuous	Some kms	centimetric	1-5 (20Hz)	as above for the stop and go

6.1.5 DGPS

The GPS differential positioning (Differential GPS = DGPS) it is a technique in which two or more receivers are used; one on a station of a geodetic or topographic reference frame (Reference Station) and one (Rover Station) which occupies the new points to be determined in a survey vector (standing or in movement). The reference station calculates the Pseudo-Range Corrections (PRC) and their variations in time (RRC = Range Rate Correction). Both corrections can be transmitted in real time to the remote receiver of the rover station or they can be stored in the receiver of the reference station to be applied during the post-processing procedure.

When the procedure is performed in real time, a connection between the two stations (reference-rover) is created by radio modem or telephone modem.

In any case, the remote receiver (in real time) or the receiver/PC with post-processing software (in delayed time) apply the corrections to the measurements of pseudo-ranges and calculate the single point positions with these corrected observations.

The differential positioning can be applied to the range of code or phase.

6.1.5.1 DGPS with measures of code:

From a time series of PRC corrections its RRC variation in time can be quantified by numerical interpolation.

The range code correction, to an arbitrary epoch '**t**', can be approximated with the following:

$$PRC_t^{satel} = PRC_{t_0}^{satel} + RRC_{t_0}^{satel} \cdot (t - t_0) \qquad \textbf{(2.64)}$$

where the term **$(t - t_0)$**, called latency, is the determinant for the precision of positioning. This is nothing other than the time difference between the calculation of correction in the reference station receiver and its application (times of transmission, calculation etc) in the rover station receiver.

Applying such corrections of range, the satellite clock errors disappear from the range measurement equations. The possible disturbing effect, caused by a deliberate degrading of clocks and orbits data can be virtually eliminated. Similarly other troubling affects such as the ionospheric and tropospheric refraction.

Therefore the remote receiver position is calculated with the corrected pseudo-ranges of code. This correction can be transmitted or stored with a RTCM standard protocol and the technique is named RTCM Differential GPS.

The pseudo-range corrections can be transmitted to the GPS receiver by:

- Reference Station GPS receiver situated temporarily at a horizontal control point within the survey area or from a permanent station, with a modem by radiofrequencies (UHF/VHF/HF) or by telephone techniques (GSM/Satellite);
- Commercial fee-for-service Wide-Area Differential GPS system, using satellite broadcast techniques to deliver accurate GPS correctors, for instance Wide Area OmniSTAR system (FUGRO group) and LandStar (THALES group) systems;

- Free service by a DGPS MSK Radiobeacon Navigation Service (DGPS Beacon IALA System);
- Free service by world Wide Area Augmentation Systems (FAA WAAS, EGNOS, GPS/GLONASS, MSAS) Satellite Service.

Such techniques provided suitable results for the quick geo-referencing of significant details on the ground.

6.1.5.2 DGPS with measures of phase:

In this technique the satellite clock errors and the errors associated with the ionospheric and tropospheric refraction are eliminated. The correction of range of phase can be transmitted in real time by the reference station receiver to the rover station receiver through the RTCM protocol or through proper format of the receiver manufactures. DGPS with measurement of phase is used for kinematic applications of precision in real time: such techniques are termed RTK (Real Time Kinematic). The aim is for the time of latency to be removed or in practice much reduced (a few milliseconds).

6.1.6 RTK

The Real Time Kinematic (RTK) positioning is based on the use of at least of two GPS receivers, one as a reference station and one or more mobile receivers (rover stations). The receiver at the reference station performs measurements for the satellites in sight and transmits the corrections to the mobile receivers. At the same time the rover stations also perform measurements on the same satellites whilst processing the data received from the reference station; each rover then evaluates its position relatively to the reference station. Typically the reference and rover receivers acquire measurements every second, producing solutions of position with the same frequency.

Using receivers in the RTK mode, the measurements generated on the signal GPS carrier phase are utilised to reach centimetric accuracies.

The automatic initialisation, called OTF (On The Fly), is a common characteristic of the receivers capable of the RTK mode, for which both the reference and the rovers require at least five common satellites in sight simultaneously. Such a process consists of resolving the phase ambiguity, which is present in the measurement of range by phase and it removes the restrictions on the movement of the rover receivers during the process of initialisation, which lasts no more than few minutes. Initially the rover receiver produces a float solution or FLT with metric accuracy (the phase ambiguity is not fixed). When the initialisation is completed, the solution becomes a FIX type and the accuracy becomes centimetric.

The number of FIX type positions per second, produced by the RTK system (Update Rate), defines with what accuracy the route of a mobile receiver (rover) can be represented. The Update Rate is measured in Hertz and it can actually reach values of 20 Hz for some modern receivers.

Time of latency or Latency is the time period between the measurements affected by the receivers (reference and rovers) and the visualisation of the position in the rover receivers (including times of measurement, formatting and data transmission from the reference to the rover and FIX solution calculation); this parameter is very important for mobile vehicle guidance.

A vehicle which travels at 25 km/hr covers for example around 7 metres per second. For this the latency must be less than 1/7 (= 0.14) of a second to obtain positions with an accuracy of less than one metre.

The data transmission from the reference station, positioned within the survey area or from a permanent station, to the rover by radio modem or GSM modem, has been standardised in accordance with an international protocol named RTCM (Radio Technical Commission for Maritime service). Messages in this format need a transmission rate of at least 4800 bauds, other standards which support transmissions also exist at slower rates of 2400 bauds (Es. CMR, Compact Measurement Record).

6.1.6.1 RTK Positioning mode

The most common GPS receivers with RTK ability have four principal positioning modes:

a. Synchronised RTK (1Hz): is the technique often used for reaching centimetric accuracies between a reference station receiver and a mobile receiver. Typically the update rate is 1 Hz. The latency of the synchronised positions (FIX) is determined in large part by the data transmission, with a transmission at 4800 bauds it achieves around the one second. The RTK synchronised solution produces the highest possible accuracy for RTK modes and adapts itself well to dynamic applications.

b. Fast Synchronised RTK (5 or 10 Hz): has the same latency and accuracy of the above mode, but the positioning solutions are produced 5 to 10 times each second. Satisfactory results are obtained when it is connected at least to 9600 bauds.

c. Low Latency RTK: allows centimetric accuracies (a little inferior to the synchronised RTK positioning mode) almost instantly due to the reduction of the latency to about 20 milliseconds, which allows 20 FIX solutions each second. The technique, exploited for the drastically decreased latency, bases itself on the data phase forecast of the reference station, which generally have a continuous solution with variations independent of signal losses, satellite motion, clock running and atmospheric delay. Thus the errors in prediction of phase measurements of the reference station from the mobile station are influenced mainly by instability in the receiver clocks and from unexpected variations in satellite orbits.

d. Moving RTK Base-Line: different from the majority of the RTK applications, in which the reference station is fixed at a point of known co-ordinates, this technique uses pairs of receivers (reference and rover) both moving. This mode is dependant on the orientation determination of a mobile in which the two RTK receivers are positioned at the two extremities of the base-line (i.e. along the keel axis of a boat). The reference station receiver transmits the effected measurements to the rover, which calculates a RTK solution synchronised (base-line with orientation and length) at 1 or 5 or 10 Hz, with centimetric accuracy. The absolute positioning of the reference station, and therefore also of the rover station, has an accuracy equivalent to that of the absolute positioning with measurements of code (some about ten meters). The reference-rover distance should not be greater than 1 km to obtain good results.

6.1.7 Treatment of the data

6.1.7.1 Computational process in the Relative GPS positioning

The Relative GPS positioning is performed according to various phases in which all the differential quantities which have been analyzed are used. It usually starts from an approximate solution which is improved by the various processes.

In all the processing programs of GPS data are the basic phases of preliminary treatment to search for the cycle slips and to find anomalous data associated with coarse errors. A good preliminary treatment of the data is the basis of a good final solution. A GPS survey can be described in a number of ways; it can be performed with two or more receivers, accorded more sessions and days of measurement.

The most common approach (single base) involves single independent bases without considering their correlations. Such a strategy is exploited by the majority of the processing programs, because it produces good results aligned to a greater simplicity. As in all computational programs, with a linear approach to least squares, it is necessary to depart from approximate values. These are improved step by step by the processing. The principal phases of the treatment are:

a. *Solution single point with measures of code:*
The approximate solution is deduced with pseudo-ranges on the C/A code (Coarse/Acquisition or Clear/Access) or P code (Precise or Protected) if available.

b. *Net determination through single differences of phase:*
It is necessary to decide which independent base-lines are to be considered in the process. To achieve this it is necessary to create the single differences between the data files corresponding to the points of stations between which it is decided to calculate the base-lines.

c. *Treatment of data with the equations to the triple differences (solution TRP):*
Starting from the approximate co-ordinates previously produced, it is essential to determine the components $(\Delta x, \Delta y, \Delta z)_{TRP}$ of the base-line vector in the WGS84 geocentric reference system, without necessarily having knowledge of the phase ambiguity. The solutions have some disadvantages, such as a sequential propagation of errors from the three processes of differentiation. As an approximate value this result, which does not represent the optimal one, will be inserted into a further computational process to the double differences and essentially it is useful for appraising the cycle slips which, when present, cause discontinuity in the calculation of the base-line components.

d. *Expansion to the double differences and solution with no fixed ambiguity (solution FLT):*
Commencing with the station position, deduced with the third differences, the components $(\Delta x, \Delta y, \Delta z)_{FLT}$ of the base-line vector are determinate again via iterative process, together with the values of the phase ambiguities relative to the various combinations of two satellites and two receivers. The combinations of the phase ambiguities are the only unknowns.

e. *Fixing of the ambiguities to an integer value:*
The values of the phase ambiguities when determined are generally not integers, they must be fixed therefore to the nearest integer value. To do this the computational software inspects the standard deviations of the ambiguity parameters, verifying that they are equal to small fractions of a cycle. The correct fixing of the ambiguity is indicated by the RATIO quality factor. Its value has to be greater than certain limits in relationship to the length of the measured base-lines.

f. *Expansion to the double differences and solution with fixed ambiguity (FIX):*
The components $(\Delta x, \Delta y, \Delta z)_{FIX}$ of the base-line vector are determined again, with knowledge of the term containing the phase ambiguities, previously fixed to an integer value. Therefore the components Δx, Δy and Δz of the vector, connecting the positions on which the two receivers are set, are the only unknowns to be solved from the equation to the double differences.

This last passage normally represents the final result of the computational process; the resolution of the system of equations to the double differences gives the final solution of the base-line vector with ambiguity fixed to the integer value (FIX Solutions).

6.1.7.2 Statistic test on the quality of the elaboration

The correctness of the result of the calculation of a base-line can be valued according to statistic test; the principal ones are:

a. Test of the Ratio: it is the ratio between the two smaller values of Variance (σ^2), calculated from different groups of fixed integers; it is of value if the phase ambiguities have been fixed correctly. The calculation process generally separates more integer values of phase ambiguity, to be used in the FIX solution. All the solutions are calculated with the probable values of the ambiguities and the relative value of Variance of the unity of weight. Ratio is the ratio between the lowest second variance and the best (lower) in absolute terms. An elevated ratio means that between the two solutions there is considerable difference or perhaps improvement; decrease of the value of variance is an indication of correct fixing of the integer values. A value of Ratio >1.5 for static measurements and Ratio >3 for kinematic measurements is considered acceptable.

b. Test on the Variance of the unity of weight: The variance of the unity of weight, at the start fixed (also called variance of reference), has to be similar to the estimated value and, under normal conditions, to be equal to 1. The procedure consists of calculating the variance of threshold through a test with degrees of freedom equal to the redundancy. Elevated values of the estimated variance can highlight the presence of noise in the signalrelated to obstacles or satellites near to the horizon, local multiple reflections (multi-path), no calculation for tropospheric or ionospheric affects or incorrect calculation of the integer phase ambiguities.

6.2 Electromagnetic

The characteristics that establish the performances of a system of electronic navigation are:

a. The range which is the maximum distance from the stations at which it can usefully be employed. Being mainly tied to the radiated power and the sensitivity of the receiver, which compose a specific technical problem, faced by the manufacture.

b. Precision[2] and Accuracy[3] with which the system generates the position of the ship, which is related to factors which should be appreciated during the employment, with the aim of knowing the reliability of the positions.

The performances of a system, in relation to the accuracy, makes reference to two particular indices of output:

---

2 "The degree of refinement of a value" (IHO S32 – fifth edition 1994, #. 3987)

3 "The extent to which a measured or enumerated value agrees with the assumed or accepted value" or "the degree of conformance with the correct value" (IHO S32 – fifth edition 1994, # 21 and # 3987)

a. Repeatability or Repeatable Accuracy[4]: is a measure of the capability of the system to repeatedly return the mobile to the same position. It is influenced by the accidental errors of the measurement (due to the operators, to the instruments and to anomalies of propagation of the EM waves) and the geometry of the system (the angle of intersection between the individual LOPs).

b. Forecast ability: it is a measure of the capability of the electronic navigation system to minimise the size of the existing difference between the measurements and the estimate of positions produced from the base of calculations, having fixed a model of propagation and the geometry of the system. In the field of medium and high frequencies, the predictions for the electromagnetic propagation for the purpose of positioning are irrelevant; it is present with all of its complications in long range systems, and thus in low frequencies.

6.2.1 Accuracy in the position determination

When the accuracy of a navigation system is established, it is appropriate to specify the degree of reliability which can be assigned to such a value. Although the distribution of the errors is more often elliptic than circular, it is simpler to quote only one parameter, generated from the radius of a circle centred on the determined point.

The mariner has the percentage value (x percent) of probability of being in such a circle. With the objective of data exchangeability, it is important to clarify which statistical method was used in the determination of the performance and also include the degree of reliability (or level of confidence), expressed as the percentage of the tests which fell in a circle of determined radius.

For bi-dimensional measurements (in the horizontal co-ordinates x and y), the parameters generally have two values:

a. Circular Error Probable (**CEP**): radius of a circle within which there is about a 50% probability of finding the correct value[5];
b. Radial error or root mean square error in the distance (**1-sRMS** or **1DRMS**): with the assumption of equality of the standard deviations around two dimensions (σ_x σ_y)[6], of orthogonality between the axes x and y, of normal and not correlated distributions of error, the following relationship is valid:

$$\mathrm{DRMS} = \sqrt{s_x^2 + s_y^2} = \sqrt{2 \cdot s^2} = 1{,}414 \cdot s \qquad \mathbf{(2.65)}$$

Generally the measure of **2DRMS** is employed, which corresponds to the 98.5% level of confidence.

4 "In a navigation system, the measure of the accuracy with which the system permits the user to return to a position as defined only in terms of the co-ordinates peculiar to that system. The correlation between the geographical co-ordinates and the system co-ordinates may or not may be known" (IHO S32 – fifth edition 1994, # 4336)

5 (see also probable error: IHO S32 – fifth edition 1994 – # 1689)

6 (see also standard error: IHO S32 – fifth edition 1994 – # 1695)

6.2.2 Lines of Position (LOPs)

Limiting distances to less than 60 miles, in the study of radio wave navigation systems, it is valid to approximate to a horizontal terrestrial surface; for greater distances the line of position is considered as an arc of maximum circle.

The systems most often used for radio wave navigation produce circular and hyperbolic lines of position, which derive position from the measurement of a difference in time $\Delta \mathbf{t}$ or a difference in phase $\Delta\varphi$. Such measurements are translated into differences of distances (hyperbolic LOPs) or direct distances (circular LOPs) respectively with the relationships:

$$\Delta d = c \cdot \Delta t \tag{2.66}$$

$$\Delta d = \frac{c}{f} \cdot \left[\left(\frac{\Delta\varphi}{2\pi} \right) + n \right] \tag{2.67}$$

where:

- $\Delta \mathbf{d}$: is the difference of distance;
- **c**: is the speed of propagation of the electromagnetic waves;
- $\Delta \mathbf{t}$: is the difference of measured time;
- $\Delta\varphi$: is the difference of measured phase;
- **f**: is the frequency of the wave on which the measurement is effected $\Delta\phi$;
- **n**: is the number of integer cycles of the received wave.

An error in the measurement of $\Delta \mathbf{t}$ or $\Delta\varphi$ appears as an error in numeric line of position, while a deviation of **c** from its standard value, creates a distortion in the whole pattern of the lines.

6.2.3 Circular lines of position (C LOPs)

Measuring the distance from a point of known co-ordinates it is possible to determine a line of position, which is a circle having the observed position at the centre and the measured distance as the radius. The error in the measurement of a distance influences and modifies the relative line of position producing a band of uncertainty, whose proportions (standard deviation of the measurements of distance) are independent of the distance.

The intersection of two circular lines of position effected by error produce an area of uncertainty, inside which is the true position of the mobile. This area generally has the shape of a parallelogram. The circular systems are characterized by the fact that the angle of intersection between the LOPs varies in the area of coverage and, at a generic point P, is equal to that between the vector radii subtended by P to the stations.

In the case of σ equal for both the sets of circumferences, the radial error drawn by (2.65) then becomes :

$$d_{RMS} = \frac{\sqrt{2\sigma^2}}{\operatorname{sen}\alpha} = \frac{1{,}414 \cdot \sigma}{\operatorname{sen}\alpha} \tag{2.68}$$

where α is the angle of intersection between the two LOPs.

Considering therefore constant s in the area of coverage, it is immediate evident that the $\mathbf{d_{RMS}}$, in circular systems, is entirely dependent on the angle of intersection between the LOPs. The curves of equal distance are identified, therefore, by those of equal a . They are of the complete arcs with this angle and have their ends at the two stations.

6.2.4 Hyperbolic lines of position (H LOPs)

'A hyperbola is an open curve (line of points in the plan) with two part, all points of which have a constant difference in distances from two fixed points called foci' (IHO S32 – fifth edition 1994 - # 2353).

Referred to an orthogonal Cartesian system, they are a varying curved symmetric ray as related to the axis of the abscissas, the axis of the ordinates and to the origin.

In hyperbolic electronic navigation, the segment of the axis of the abscissas incorporated between the two foci A and B is called the base-line. Two fixed points in a plane can be the foci of endless hyperbolae, which will constitute a pattern of similarly focused hyperbolae.

In a system of similarly focused hyperbolae, representing geometric lines which differ one from the other by a constant quantity, it can be observed:

a. The hyperbolae cross the base line at regular intervals of distance;
b. The distance between two hyperbola increases with the increase of distance from the base-line.

In reality, the lines of position obtainable from hyperbolic radio navigation systems are hyperboloids. A pair of synchronized radio stations located at the hyperboloids foci, each can be paired with one or more other synchronized stations forming a hyperbolic chain. The observer is on one of the curved hyperboloids which are produced by two pairs of radio stations. With the measurements made onboard, the observer can determine his position by identifying the relevant spherical hyperbolas displayed on charts or from special tables built for such a purpose.

6.2.5 Determination methods of electromagnetic wave lines of position (EW LOPs)

An electromagnetic wave LOP can be produced by the direct or indirect measurement of:

a. Distance;
b. Difference between distances.

The measurements which dimensionally express a distance, are in reality obtained by the transformation of two possible and different kinds of measurement:

a. Difference of phase;
b. Difference of time.

6.2.6 Measurements of difference of phase

Differences of distances or distances can be determined from a measurement of phase difference.

a. Measures of distances:

Considering A, a point on the earth's surface for which co-ordinates are known in a stated reference system, with a station issuing a continuous electromagnetic wave with frequency **f** and at a generic point P, a suitable receiver able to measure the difference between the phases of the electromagnetic wave has, constantly, knowledge of the positions of A and P.

To make this possible, it requires the receiver to have an oscillatory wave of stable frequency which is synchronised with that of the issuing station.

In such a way, supposing known conditions of propagation in the medium which separates the station from the receiver, it is possible to know, constantly, the phase of the radio wave at station A and to make a comparison with the phase of the incoming radio wave at the receiver.

From the measurement of this phase difference it is possible to obtain the distance between the transmitting station and the receiver to less than multiples of 2π (or of 360°). The related line of position on the earth is represented by the circumference having A at the centre and the calculated distance for the radius.

Defining a lane as the space between two LOPs with a phase difference of 360°, in this case it is represented by the spherical circular area between the two circumferences. Therefore, starting from the transmitting station, every point equal to the wavelength corresponds to crossing a lane with a width equal to the wavelength.

The errors of measurement are expressed in cels (cents of lanes).

b. Measurement of difference of distances:

Sited at A and B are two radio wave transmitters of the same frequency, set at points on the earth's surface with known co-ordinates; at P a receiver able to receive separately the signals coming from the two stations and to calculate, at the same time, the phase difference of the two incoming radio waves. Except for multiples of 2π (or of 360°), such values allow the receiver to obtain the difference of the distances it is from the two stations A and B.

Being a hyperbola, the line of the points of difference of distances from two fixed points (called the foci) is constant, it results that from every point on a stated hyperbola the same phase difference is measured.

It is possible to conclude that a measurement of phase difference defines a hyperbolic line of position.

A receiver is able to measure in a lane only the absolute value of the phase difference, from 0° to 360°; this involves ambiguity, because such a difference is positive for one side of the hyperbola and negative for the other.

Appropriate techniques ensure the sides of hyperbolas are always positive. The phase difference is generally expressed in cents of lane (cels).

The identification, then, of the lane, to which the difference of measured phase refers, makes it essential to know the lane in which the receiver was positioned when it was set to work; the purpose is to regulate the special control switch which numerically records the passage across into a whole new lane.

6.2.7 Measurement of difference of time

Measurement of Difference of time involves both the measurement of a temporal interval delimited by two instants, which are recorded in succession, and the difference between two of these temporal intervals.

The two different ways of interpreting this quantity, allow the determination of two types of measurement: of distance and of difference of distances.

a. Measurement of distance:

   The distance is obtained by the measurement of the time which passes between the transmission of a transmitting station of known position at a known instant and the instant at which the signal reaches the receiver.

   What connects the measurement of this time interval to a distance is the speed of propagation of the electromagnetic waves. Therefore, the ability to forecast the anomalies of propagation defines the capability of the positioning system.

b. Measurement of difference of distances:

   Two transmitting stations at A and B are set at positions of known co-ordinates.

   The pulses from the two stations arrive sequentially at a receiver; with appropriate techniques, it is possible to measure the difference in time between the arrivals of the signals. It is clearly a function of the difference in distance of the receiver from the two stations.

   The measurement of difference of time is made on both sides of the hyperbola, causing an ambiguity, since the receiver is not able to establish which of the two pulses arrives first. To eliminate this, the transmission of pulses is not simultaneous but is made at intervals of a constant quantity (coding delay).

6.3 Acoustic Systems

Acoustic Positioning Systems were originally developed in the United States to support underwater research studies in the 1960s. Since then, such systems have played an important role in providing positioning for towed bodies, ROVs and in most phases of the offshore hydrocarbon industry, from initial exploration through to field development and maintenance. More recent developments and technical improvements have also seen it being used for military purposes.

Acoustic positioning is able to provide very high positional repeatability over a limited area, even at a great distance from the shore. For many users repeatability is more important than absolute accuracy, although the advent of GPS and integrated GPS/INS technology now makes it possible to achieve both high precision and accuracy.

Modern GPS developments such as DGPS, WADGPS and RTKGPS may have reduced the use of acoustic systems in areas such as seismic survey operations and seismic streamer tracking. However, in positioning rigs relative to wellheads (whether the rig is anchored or dynamically positioned), ROV tracking etc. acoustic positioning remains an important technique. Furthermore, in areas where sunspot activity (most pronounced around the magnetic equator and the Polar Regions) can cause interference to DGPS an acoustic system can provide a useful backup for GPS.

Acoustic Positioning Systems measure ranges and directions to beacons that are deployed on the seabed or fitted to ROVs and towed bodies. The accuracy achieved will depend on the technique used, range and environmental conditions. It can vary from a few metres to a few centimetres.

Acoustic positioning systems, produced by several manufacturers are generally available in the following 'standard' frequency bands

Classification	Frequency	Max Range
Low Frequency (LF)	8 – 16 kHz	>10km
Medium Frequency (MF)	18 – 36 kHz	2 – 3½ km
High Frequency (HF)	30 – 64 kHz	1500 m
Extra High Frequency (EHF)	50 – 110 kHz	<1000 m
Very High Frequency (VHF)	200 – 300 kHz	<100 m

6.3.1 Acoustic Techniques

There are 3 primary techniques used in acoustic positioning systems, Long Baseline, Short Baseline and Super or Ultra Short Baseline with some modern hybrid systems using a combination of these techniques.

6.3.1.1 Long Baseline Method (LBL)

LBL acoustic systems provide accurate fixing over a wide area by ranging from a vessel, towed sensor or mobile target, to three or more transponders located at known positions on the seabed. Transponders are interrogated by a transducer fitted to the surface vessel. The lines joining pairs of seabed transponders are termed **baselines** and can vary in length from 50m to over 6km depending on the water depth, seabed topography, the acoustic frequency used and the environmental conditions.

The LBL method provides accurate local control and high repeatability. If there is redundancy, i.e. 3 or more position lines, the quality of each position fix can also be estimated.

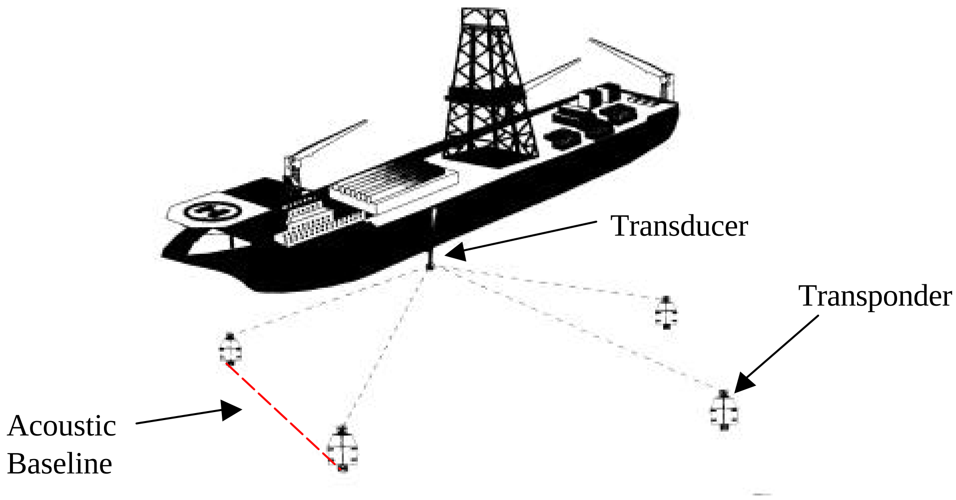


Fig. 2.21 "Long Baseline Meted"

Calibration of LBL Systems

Seabed transponders cannot be fixed or deployed as accurately as land based systems. Once laid, however, a pattern of transponders must be fixed relative to each other and then tied into the geodetic datum in use. The latter is usually achieved using GPS and the process of calibration generally follows three steps:

a. Relative Geometry:Relative positioning is achieved by nominating one of the transponders as the origin of the seabed array and defining its orientation by determining the direction to a second transponder. To achieve this, the ship steams at random throughout the area, aiming to cross each baseline at right angles at least once, gathering valid sets of slant ranges. These ranges can then be processed to fix the relative positions of the seabed transponders by trilateration and rigorous adjustment.

b. Orientation:The orientation process involves steaming with a constant heading along three legs at 90° to 120° intervals, taking two well-separated acoustic fixes on each leg. The effect of the tidal stream is cancelled by the course alterations and the network is aligned with north as defined from GPS positions or by the ship's gyro compass.

c. Absolute Positioning: This is achieved by matching fixes obtained from the deployed acoustic network with GPS positions.

6.3.1.2 Short Baseline Method (SBL)

SBL methods replace the large baselines formed between transponders on the seabed with baselines between reference points on the hull of a surface vessel, i.e. the co-ordinate frame is now fixed to the vessel instead of the seabed. Three or four transducers separated by distances of 10 to 100 metres are fitted to the hull of the vessel and connected to a ship-borne acoustic processor.

Underwater targets or seabed positions are marked by single acoustic beacons, the transmissions from which are received by the hull-mounted transducers. The returning signals – together with knowledge of the SV in the water column – are passed to a central processor, where the horizontal offset between the vessel and the beacon is computed. As with the LBL method, redundant observations are used to estimate the quality and accuracy of the position fix.

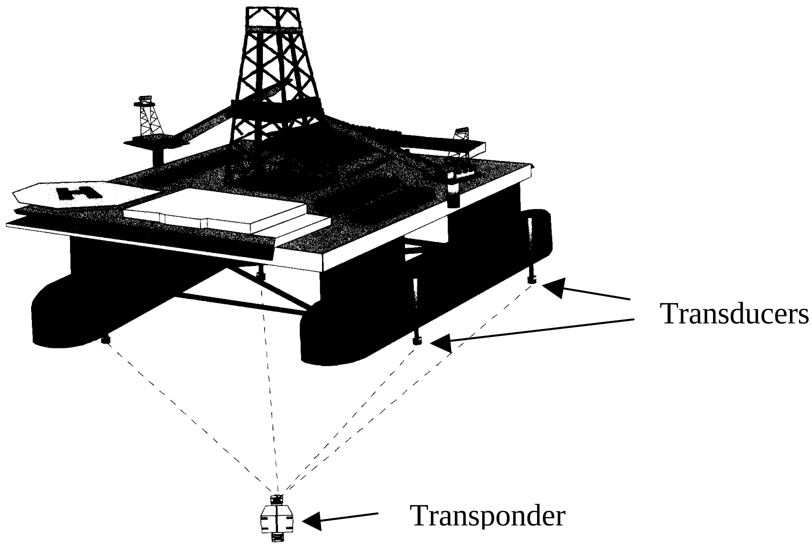


Fig. 2.22 "Short Baseline Meted"

The position of the transponders on board the vessel can be accurately determined during installation. Vessel heading and Roll and Pitch measurements have to be made during operation and as always a good knowledge of Sound Velocity is required.

6.3.1.3 Ultra or Super Short Baseline Method (USBL or SSBL)

In a USBL system the 3 or 4 hull mounted transponders of an SBL system are replaced with a single hull unit comprising an array of transducers. Phase comparison techniques are used to measure the angle of arrival of an acoustic signal in both the horizontal and vertical planes. Thus, a single beacon located either on the seabed or on a mobile target (e.g. a towed sonar body) can be fixed by measuring its range and bearing relative to the target.

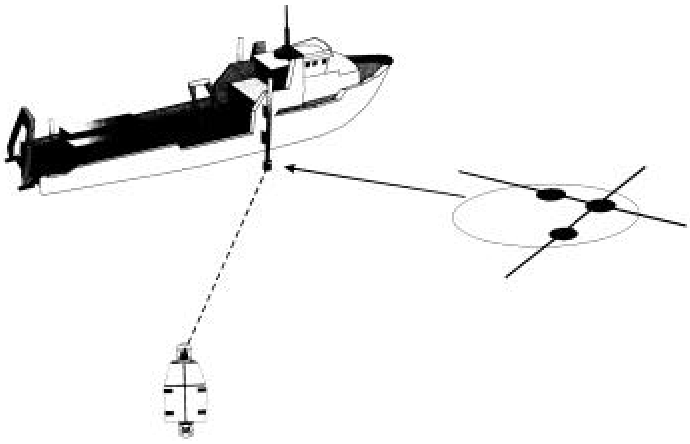


Fig. 2.23 "Ultra Short Baseline Meted"

The USBL method provides a simple positional reference input for dynamically positioned (DP) vessels and is also convenient for tracking towed bodies and ROVs.

Although more convenient to install, a USBL transducer requires careful adjustment and calibration. A compass reference is required and the bearing measurement must be compensated to allow for the pitch and roll of the vessel and for refraction effects in the water column. Unlike conventional LBL and SBL methods, there is no redundant information on standard USBL systems from which position accuracy can be estimated and accuracy is normally stated as between 0.5 to 1% of the maximum slant range measurement.

6.3.1.4 Combined Systems

These systems combine the benefits from all the above methods to provide a very reliable position with a good level of redundancy. The combined systems come in several varieties:

- Long and Ultra Short Baseline (LUSBL)
- Long and Short Baseline (LSBL)
- Short and Ultra Short Baseline (SUSBL)
- Long, Short and Ultra Short Baseline (LSUSBL)

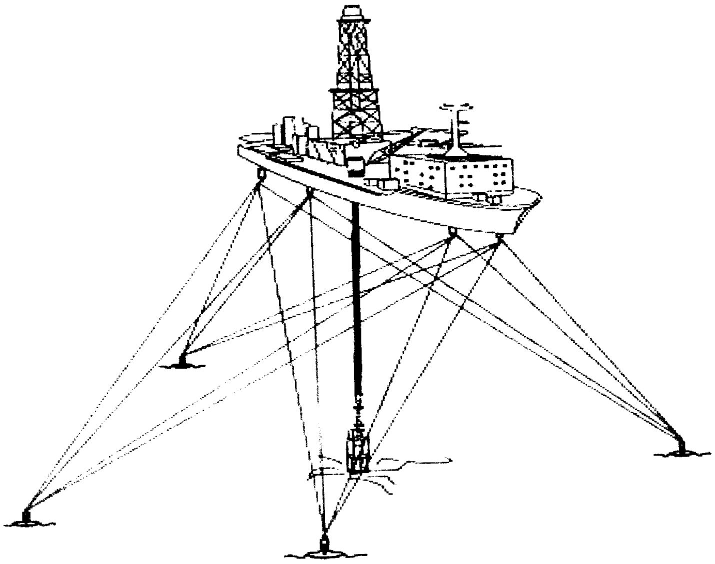


Fig. 2.24 “Combined System (LSUBL)”

6.3.1.5 Multi-User Systems

Multi-User systems are required when more than one vessel is working in close proximity and wishes to use the same acoustic system e.g. a drilling vessel in an oilfield might have a construction barge, a pipe lay barge and an ROV support vessel at the same location, all holding station by means of Dynamic Positioning (DP). This means that the potential for "acoustic pollution" is significant. The following solutions to this problem are either operational or under development (2004) are:

- Single "Master" seafloor beacon interrogation systems
- Master surface vessel with radio telemetry synchronisation to other vessels
- More channels within the same band through signal processing techniques
- The use of different frequency bands for different operations

6.3.2 Principles of Measurement

<u>Range Measurement</u>

a. If slant range (R) is determined by interrogating a transponder and θ is known then:

   $R = \dfrac{ct}{2}$ and the horizontal distance (Y) can be determined by: $Y = R\ sinq$

b. If transponder is replaced by an unintelligent 'pinger' beacon, direct slant range cannot be obtained and the depth must be known to calculate the horizontal distance: $Y = D\ tan\ q$

c. Knowledge of SV (c) allows θ to be determined by measuring differences in signal arrival times between hydrophones 1 & 2 (Figures 2.25 & 2.26). Therefore the angular measurement between a transducer/hydrophone and a beacon can be determined.

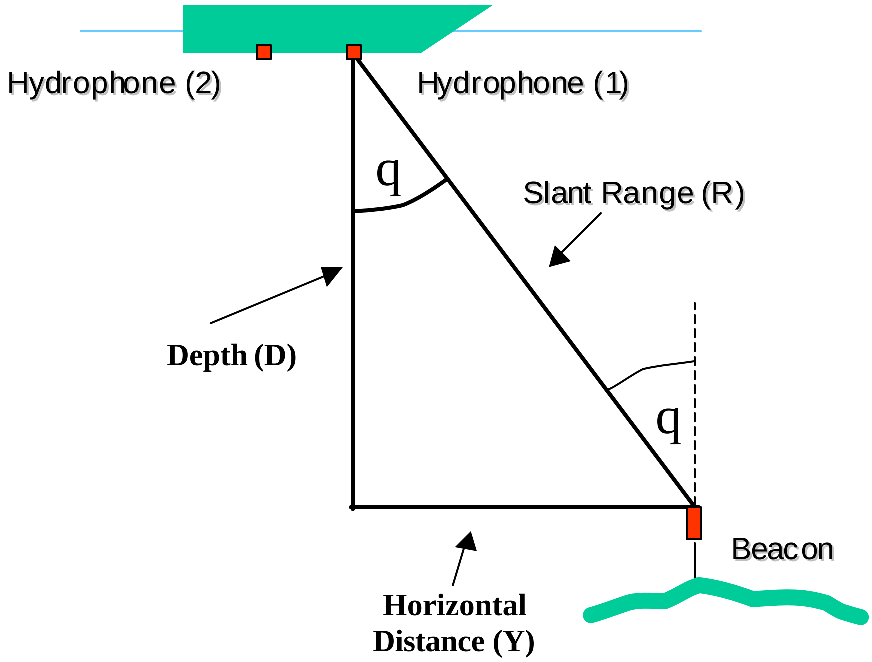


Fig. 2.25 "Range determination"

Angular Measurement

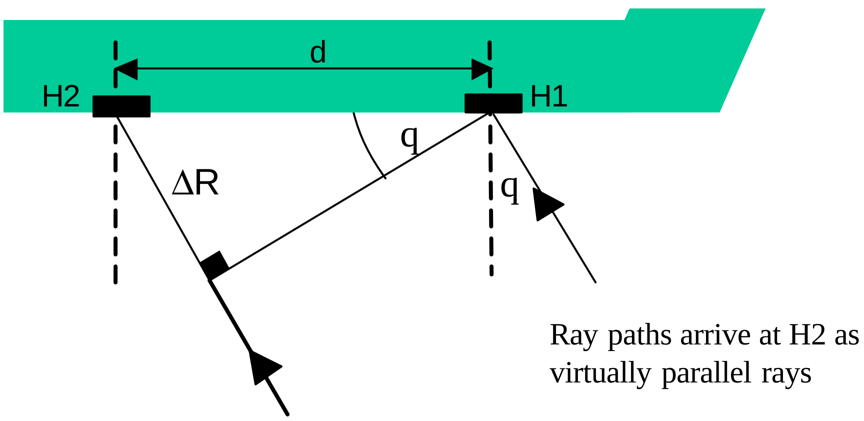


Fig. 2.26 "Angular Measurement"

- Provided Sound Velocity is known: $DR = cDt$ and $\sin q = \dfrac{c\Delta t}{d}$

Where:
c = Sound Velocity
Δt = Difference in arrival time of signal at H1 and H2
d = Distance between transducers/transducer elements/hydrophones

- A third transducer mounted at right angles to H1 and H2 enables the bearing of the beacon to be determined.

- When a vessel is directly over a transponder, two hydrophones in the same axis will receive signals in phase. This is a useful technique used in dynamic positioning, where any shift off station is sensed by signals arriving out of phase.

$$\sin q = \frac{c\Delta t}{d}$$

$$Y = R \sin q \quad = R\frac{c\Delta t}{d}$$

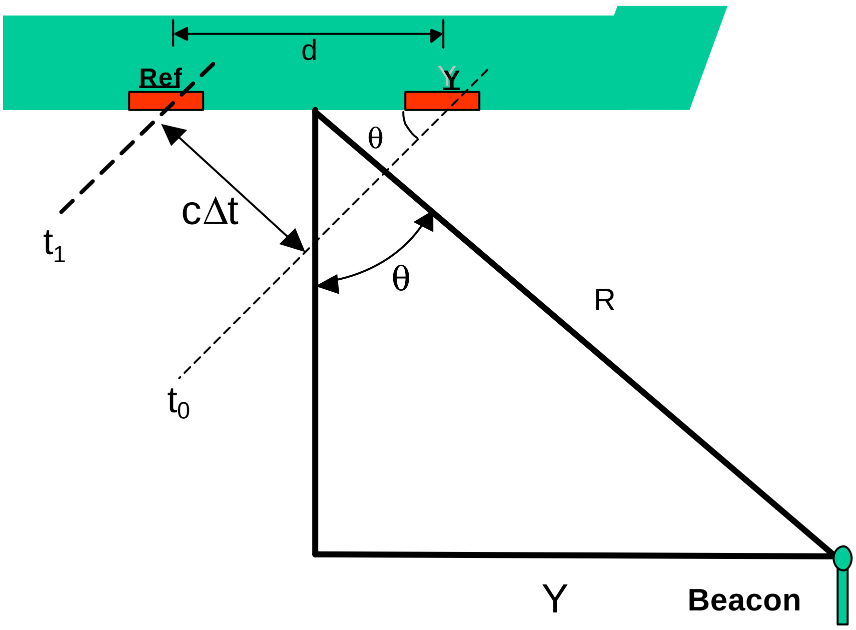


Fig. 2.27

<u>Calculating position in 2 planes</u>

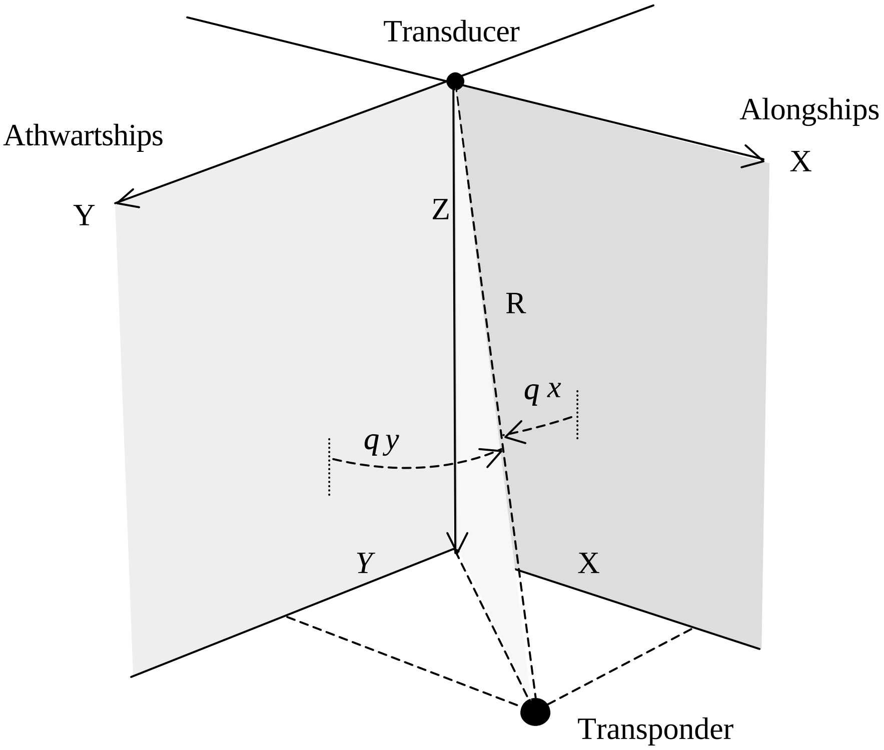


Fig. 2.28

$X = R \sin \theta x$ $\qquad$ $Y = R \sin \theta y$ and therefore $\qquad$ $z = R(1 - \sin^2 \theta y - \sin^2 \theta x)^{1/2}$

Note: Apparent position requires adjustment for:

a. Pitch and roll.
b. hydrophone alignment (at installation).
c. Hydrophone offset (fixed amount).
d. Transponder offset (fixed amount).

The Z co-ordinate is calculated from acoustic data, therefore depth information can be used to improve position accuracy under conditions of thermal gradients. Positioning accuracy is considered to be better than 1% of slant range.

6.3.3 Accuracy and errors sources

The overall accuracy of an acoustic fix will depend on:

a. The accuracy with which a transponder array is established relative to a geodetic datum.

b. Determination and supression of multipath effects (reflections). This is particularly noticeable in the region of fixed structures such as production platforms and is worse for SSBL and SBL systems than for LBL systems.

c. The accurate determination of sound velocity, velocity gradients and the amount of refraction.

d. The frequency used. Accuracy increases with increasing frequency but at the expense of range and the power required.

e. The fix geometry and, to a certain degree seabed topography i.e. whether or not there is a 'line of sight' between transponders.

f. The sophistication of the processing system and software being used.

g. Errors in time measurement owing to the presence of noise in the received signals. Noise may be:

- Ambient Noise (NA): Waves, wind, rain, marine life.
- Self Noise (NS): Propulsion, machinery, flow
- Reverberation Noise (NR): Volume reverberation, sea surface, seabed structures.

Signal to Noise ratio (SNR) = $E - N$

Where $E = SL - TL$
$N = 20Log_{10}NT$

And $NT = (NA^2 + NS^2 + NR^2)^{1/2}$

6.3.3.1 Sound Velocity Structure

Seawater is not a uniform, isotropic medium and therefore the velocity of sound in water is affected by changes in temperature (the dominant factor), salinity and depth. The mean value of SV in seawater will increase approximately as follows:

By 4½ m/s for every 1°C increase in temperature.
By 1.21 m/s for every part per thousand increase in salinity.
By 1 m/s for every 60 metre increase in depth.

All systems require a precise knowledge of the average sound velocity and preferably knowledge of the SV profile. This is usually obtained by using an independent TSD probe or velocity profiler.

6.4 Optical Techniques

The following paragraphs contain only a brief summary of the traditional methods employed in dredging, channel and harbour surveys. Most of them are no longer used due to the employment of Differential GPS techniques, however they are still valid. Chapter 7 contains more detailed explanation of these methods for hydrographic surveying.

6.4.1 Tag Line Positioning (Cable sounding)

The sounding survey with a cable is used for the lack of other positioning systems; it requires a tag line kept in tension from an operator who holds the end of the cable anchored on the beach.

On the vessel another operator will unwind the line from a winch, always keeping it in tension.

Then, at slow speed, the vessel begins the sounding line (generally perpendicular to the beach) steered by an operator who checks the direction followed by a planned fixed angle on circle to reflection (or sextant) or other visual method.

6.4.2 Sextant Resection Positioning (Inverse intersection)

This system needs two operators with circle to reflection (or sextants) in the vessel.

They measure the difference in azimuth of points selected during the planning. Every fix during the survey is the intersection between two LOPs; a sounding therefore is associated with the reading of two differences of azimuth.

6.4.3 Triangulation/Intersection Positioning (Direct intersection)

The direct intersection guarantees greater precisions, but it requires two operators on the ground and a reliable system of communication with the vessel.

The first operator, through a circle to reflection (or a theodolite), guides the vessel along the line, communicating by radio any required adjustments, while the second, using a total station, determines angles and distances of the vessel at established time intervals.

6.4.4 Range-Azimuth Positioning (Mixed system optic and electromagnetic)

It is a method which allows the generation of a position through the orthogonal intersection between two LOPs. An EDM system and a theodolite (or total station), which observes the vessel, are utilised for positioning.

REFERENCES

Luciano Surace	*"La georeferenziazione delle informazioni territoriali"1998*	Estratto dal "Bollettino di geodesia e scienze affini", 1998
A. Cina	*"GPS Principi Modalità e Tecniche di Posizionamento"*	Celid, Prima edizione – 2000
L. Costa	*"Topografia"*	Cooperativa Libraria Universitaria – Genova, Prima ristampa – 2001
IHO	*"Hydrographic Dictionary" S–32*	International Hydrographic Organization, Monaco, 5th edition – 1994
IHO	*"IHO Standards for Hydrographic Survey" S–44*	International Hydrographic Organization, Monaco, 4th edition – 1998
IHO	*"IHO Standards for Hydrographic Survey" Supplement to S–44*	International Hydrographic Organization, Monaco, Draft enclosed in letter IHB File N. S3/7198 – 29 April 1998
USACE	*EM 1110-2-1003* *"Hydrographic Surveying"*	U.S. Army Corps of Engineers, Department of the Army, Washington, 1 January 2002
USACE	*EM 1110-1-1004* *"Geodetic and Control Surveying"*	U.S. Army Corps of Engineers, Department of the Army, Washington, 1 June 2002
USACE	*EM 1110-1-1003* *"NAVSTAR Global Positioning System Surveying"*	U.S. Army Corps of Engineers, Department of the Army, Washington, 1 July 2003
USACE	*EM 1110-1-1005* *"Topographic Surveying"*	U.S. Army Corps of Engineers, Department of the Army, Washington, 31 August 1994
NOAA Melvin J. Umbach Rockville, Md.	*"Hydrographic Manual"*	U.S. Department of Commerce National Oceanic and Atmospheric Administration (NOAA) National Ocean Service (NOS), Fourth Edition 4th July 1976

Admiralty	"*Manual of Hydrographic Surveying*"	Hydrographic Department Admiralty (UKHO), Vol. I (1965) and Vol. II (1970)
Simo H. Laurila	"*Electronic Surveying in practice*"	John Wiley & Sons, Inc New York (USA), January 1983
Börje Forssell	"*Radio navigation system*"	Prentice Hall International (UK) Ltd, 1991

BIBLIOGRAPHY

Luciano Surace	*“La georeferenziazione delle informazioni territoriali” 1998*	Estratto dal "Bollettino di geodesia e scienze affini", 1998
A. Cina	*“GPS Principi Modalità e Tecniche di Posizionamento”*	Celid, Prima edizione – 2000
L. Costa	*“Topografia”*	Cooperativa Libraria Universitaria – Genova, Prima ristampa – 2001
Romagna Manoia G.	*“Manuale di Idrografia per la costruzione delle carte marine*	Accademia Navale di Livorno, terza edizione – 1949
II 3100.	*“Manuale dell’Ufficiale di Rotta”*	Istituto Idrografico della Marina, Genova, Quinta edizione – 1992 / Prima ristampa – 1998
NorMas FC 1028.	*“Norme di Massima per i Rilievi Idrografici”*	Istituto Idrografico della Marina, Genova, Seconda edizione – 1978
Admiralty	*“Manual of Hydrographic Surveying”*	Hydrographic Department Admiralty (UKHO), Vol. I (1965) and Vol. II (1970)
IHO	*“Hydrographic Dictionary” S–32*	International Hydrographic Organization, Monaco, 5th edition – 1994
IHO	*“IHO Standards for Hydrographic Survey” S–44*	International Hydrographic Organization, Monaco, 4th edition – 1998
IHO	*“IHO Standards for Hydrographic Survey” Supplement to S–44*	International Hydrographic Organization, Monaco, Draft enclosed in letter IHB File N. S3/7198 – 29 April 1998
USACE	*EM 1110-2-1003* *“Hydrographic Surveying”*	U.S. Army Corps of Engineers, Department of the Army, Washington, 1st January 2002
USACE	*EM 1110-1-1004* *“Geodetic and Control Surveying”* 1st June 2002	U.S. Army Corps of Engineers, Department of the Army, Washington.
USACE	*EM 1110-1-1003* *“NAVSTAR Global Positioning System Surveying”*	U.S. Army Corps of Engineers, Department of the Army, Washington, 1st July 2003

USACE	*EM 1110-1-1005* *"Topographic Surveying"*	U.S. Army Corps of Engineers, Department of the Army, Washington, 31 August 1994
NOAA Melvin J. Umbach Rockville, Md.	*"Hydrographic Manual"*	U.S. Department of Commerce National Oceanic and Atmospheric Administration (NOAA) National Ocean Service (NOS), Fourth Edition 4th July 1976
NOAA	NOS Hydrographic Surveys *"Specifications and Deliverables"*	U.S. Department of Commerce National Oceanic and Atmospheric Administration (NOAA) National Ocean Service (NOS), March 2003
Luigi Sinapi	*"Lezioni di Idrografia teorica ed operativa"*	Napoli, A.A. 2002/2003
Simo H. Laurila	*"Electronic Surveying and Navigation"*	John Wiley & Sons, Inc New York (USA), 1976
Simo H. Laurila	*"Electronic Surveying in practice"*	John Wiley & Sons, Inc New York (USA), January 1983
Börje Forssell	*"Radio navigation system"*	Prentice Hall International (UK) Ltd, 1991
Alam E. Ingham	*"Sea Surveying"*	John Wiley & Sons, Inc New York (USA), July 1974

CHAPTER 3
DEPTH DETERMINATION

by LCdr. Fernando ARTILHEIRO (Portugal)

1. INTRODUCTION

Depth determination is a fundamental task for a hydrographer, which requires specific knowledge of the medium, of underwater acoustics, of the plethora of devices available for depth measurement, of complementary sensors for attitude and heave measurement and proper procedures to achieve and meet the internationally recommended standards for accuracy and coverage as articulated in IHO publication S-44 4th Edition.

Lead line and sounding pole were the earliest methods used for directly measuring water depth. Their easy principles of operation ensured their continued use over many centuries.

Single beam echo sounders, derived from military sonars, were a major development and have been used in hydrographic surveying since the mid 1900s.

During the last decade, hydrographic surveying has experienced a conceptual change in depth measurement technology and methodology. Multibeam echo sounders (MBES) and airborne laser sounding systems (ALS) now provide almost total seafloor coverage and depth measurement. The high data density and high acquisition rates have led to huge bathymetric data sets and much ancillary data.

The state of the art of the depth measurement equipment was evaluated by the working group of the S-44 [IHO, 1998] as follows:

> *"**Single beam echo sounders** have reached a sub-decimetre accuracy in shallow water. The market offers a variety of equipment with different frequencies, pulse rates etc. and it is possible to satisfy most users' and, in particular, the hydrographers' needs. (…)*
>
> ***Multibeam echo sounder** technology is developing rapidly and offers great potential for accurate and total seafloor search if used with proper procedures and provided that the resolution of the system is adequate for proper detection of navigational hazards.*
>
> ***Airborne laser sounding** is a new technology which can offer substantial productivity gains for surveys in shallow, clear water. Airborne laser systems are capable of measuring depths to 50 m or more."*

Despite these new technologies, single beam echo sounders (SBES) still remain, for the present, the traditional equipment used on hydrographic surveys worldwide. These echo sounders have also evolved from analogue to digital recording, with greater precisions and higher accuracies and with specific features which allow a wider variety of purposes to be met. The use of digital echo sounders along with motion sensors, satellite positioning systems (such as GPS) and software for data acquisition have combined to optimize productivity with corresponding reductions in personnel for survey operations.

MBES have become a valuable tool for depth determination when full seafloor ensonification is required. An increasing number of National Hydrographic Offices (NHO) has adopted multibeam technology as the methodology of choice for the collection of bathymetric data for new chart production. The acceptance of multibeam data for use in published nautical charts is a sign of growing confidence in the technology. Notwithstanding their impressive capabilities, it is vital that planners, operators and checkers have indepth knowledge of MBES operating principles, as well as practice in data interpretation and validation.

Airborne laser sounding systems are being used by a few NHOs; these systems have, by far, the highest data acquisition rates and are particularly suited to near shore and shallow water areas. However, the high costs for the assets involved in data collection and their operation do not currently allow a more general use.

In this Chapter, Section 2 covers the broad acoustic fundamentals necessary for the understanding of sea water acoustic waves and physical characteristics, acoustic wave propagation and acoustic parameters. Section 3 deals with motion sensors. Section 4 covers transducer characteristics, their classification with regard to beam pattern, principles of operation and their installation. Section 5 describes the acoustic systems of single beam echo sounders and swath systems, both multibeam and interferometric sonars, with regard to their characteristics, principles of operation, installation and operational use. Finally, Section 6 explains non-acoustic systems, such as airborne laser and electromagnetic induction systems, remote sensing systems and classic mechanical devices.

The terminology used in this chapter follows, as far as possible, the Hydrographic Dictionary [IHO SP-32 5th Edition, 1994].

2. ACOUSTIC AND MOTION SENSORS FUNDAMENTALS

Sea water is the medium in which hydrographic measurements normally take place, therefore knowledge of sea water's physical properties and of acoustic waves propagation is important for full comprehension of the contents and aim of this chapter.

2.1 Sea water acoustic waves and physical characteristics

Despite electromagnetic waves having an excellent propagation in a vacuum and air, they hardly penetrate nor propagate through liquids. However, acoustic waves, either sonic or ultra-sonic, achieve good penetration and propagation through all elastic media once these media can be made to vibrate when exposed to pressure variations. The majority of the sensors used for depth determination use acoustic waves.

2.1.1 Acoustic field

The acoustic waves consist of subtle variations of the pressure field in the water. Sea water particles move longitudinally, back and forth, in the direction of the propagation of the wave, producing adjacent regions of compression and expansion, similar to those produced by longitudinal waves in a bar.

The intensity of the acoustic wave, **I**, is the amount of energy per second crossing a unit area. The acoustic intensity is given by:

$$I = \frac{p_e^2}{\rho c} \qquad \textbf{(3. 1)}$$

where $\boldsymbol{\rho}$ is the water density, **c** is the sound velocity in the water and $\mathbf{p_e}$ is the effective acoustic pressure[1], given by the root mean square of the peak pressure amplitude, **P**, i.e.:

$$p_e = \frac{P}{\sqrt{2}}$$

The acoustic wave intensity is computed using average acoustic pressure rather than instantaneous values. The acoustic pressure and intensity, due to their wide range variation, are usually expressed in logarithmic scales referred to pressure and intensity levels, the decibel scale being the most common logarithmic scale.

The acoustic intensity level, **IL**, is given by:

$$IL = 10\log_{10}\frac{I}{I_{Ref}} \qquad \textbf{(3. 2)}$$

where $\mathbf{I_{Ref}}$ is the reference intensity.

The acoustic intensity level is alternatively expressed by,

$$IL = 20\log_{10}\frac{p_e}{p_{Ref}} \qquad \textbf{(3. 3)}$$

where $\mathbf{p_{Ref}}$ is the reference pressure[2].

2.1.2 Sonar Equation

The sonar[3] equation is used to study and express the detection capability and performance of echo sounders as a function of operating conditions [Urick, 1975].

The sonar equation for echo sounders defines the signal or echo detection as the Echo Excess (**EE**),

$$EE = SL - 2\,TL - (NL-DI) + BS - DT \qquad \textbf{(3. 4)}$$

where **SL** = source level, **TL** = transmission loss, **NL** = noise level, **DI** = directivity index, **BS** = bottom backscattering strength and **DT** = detection threshold.

In this section each term of the sonar equation is presented and studied to enable a better understanding of the processes involved in acoustic signal propagation and echo detection.

The acoustic wave intensity $\mathbf{I_r}$ at a distance **r** from the transmitter is obtained by,

$$I_r = \frac{p_r^2}{\rho c} \quad W/m^2 \qquad \textbf{(3. 5)}$$

---

1 Pascal (Pa) is the unit of pressure in the International System (SI).

2 In underwater acoustics reference pressure is usually adopted as 1 μPa.

3 SOund NAvigation and Ranging.

Where $\mathbf{p_r}$ is the effective pressure at the distance $\mathbf{r}$ from the source and $\boldsymbol{\rho}\mathbf{c}$ is the acoustic impedance[4] (considering a sound velocity of 1500 m/s and sea water density of 1026 kg/m^3 the acoustic impedance is $\rho c = 1.54 \cdot 10^6\ kg/m^2s$).

The Source Level (**SL**) gives the acoustic signal intensity level referred to the intensity of a plane[5] wave with root mean square (rms) pressure 1 μPa, for a point located 1 metre away from the centre of the source (transmitter), i.e.:

$$SL = 10 \cdot \log_{10} \frac{I_1}{I_{Ref}} \qquad \textbf{(3. 6)}$$

The Transmission Loss (TL) takes into account the losses of acoustic intensity due to geometry, i.e. from spreading losses, proportional to $\mathbf{r^2}$ and the losses due to absorption, proportional to the coefficient of absorption, dependent on the physical and chemical sea water properties and on the acoustic frequency (see 2.3.1).

The spreading loss is caused by the geometry of the beam with its cone shape (Figure 3.1). The increase in area results in the decrease of power per unit of area.

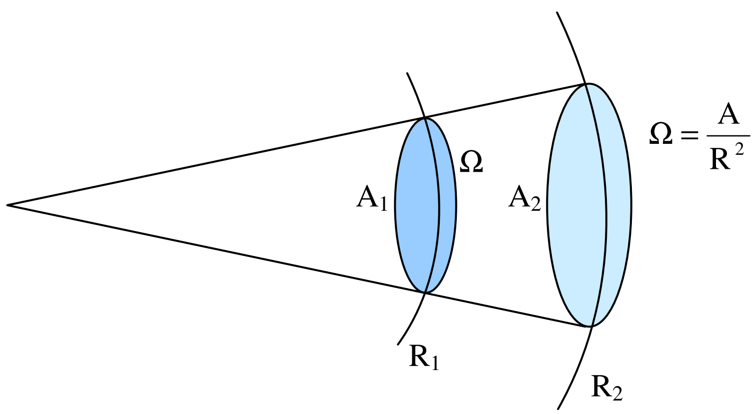


Fig. 3.1 "Spreading loss due to beam geometry"

The power, $\mathbf{\Pi}$, of the acoustic pulse is equal to Intensity x Area:

$$\Pi = I_1 \cdot A_1 = I_2 \cdot A_2$$

where $\mathbf{A_1} = \Omega \cdot R_1^2$ and $\mathbf{A_2} = \Omega \cdot R_2^2$, being $\boldsymbol{\Omega}$ the solid angle[6].

---

4 Acoustic impedance corresponds to the resistance of the medium to the wave propagating through it, i.e., a proportionality factor between the velocity and the acoustic pressure.

5 Plane waves occur in a small region away from the source where wave fronts (points where vibrations are in phase), are approximately plane and have negligible change in amplitude.

6 The solid angle, Ω, is the space enclosed by a conical surface. The value, expressed in steradian (sr), is obtained as $\Omega = S/R^2$, where S is the spherical surface with centre in the apex of the cone and radius R.

Therefore, the relation of intensities is given by:

$$\frac{I_1}{I_2} = \left(\frac{R_2}{R_1}\right)^2 \quad \textbf{(3. 7)}$$

If one considers the reference intensity at R_1 = 1 m, the distance at which the source level (SL) is determined, the logarithmic ratio of the intensities relates to the transmission loss due to spreading is:

$$10 \cdot \log_{10} \frac{I}{I_{Ref}} = 10 \cdot \log_{10} \frac{1}{R_2^2} = -20 \cdot \log_{10} R_2 \quad \textbf{(3. 8)}$$

Hence, the transmission loss is given by:

$$TL = 20 \log_{10} r + ar \quad \textbf{(3. 9)}$$

where **r** is the distance to the transducer and **a** the absorption coefficient.

The Noise Level (**NL**) is dependent on the environmental spectral noise level ($\mathbf{N_0}$) and on the transducer bandwidth during reception (**w**),

$$NL = N_0 + 10 \log_{10} w \quad \textbf{(3. 10)}$$

The noise in the ocean is generated by several sources [Urick, 1975] such as: waves, rain, seismic activity, thermal noise, living organisms and man-made.

Besides noise, it is also important to take into account the combined affect of the backscattering acoustic energy created by various marine bodies; these include surface waves, air bubbles, marine life, materials in suspension, etc. This contribution is known as the Reverberation Level (RL).

Transducers usually have the capacity to concentrate the energy within a conical shape (Figure 3.2). This property can be quantified, for the sonar equation, as the ratio from the intensity within the beam to the intensity of an omni directional point source with the same power.

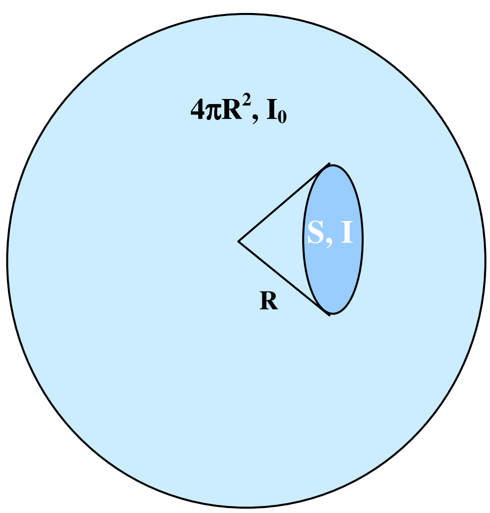


Fig. 3.2 "Ensonified surfaces from an omni directional source and directional source"

The power, **Π**, of the acoustic pulse is equal to (Intensity x Area). Therefore, considering the same power for the omni directional and directional sources:

$$\Pi = I_0 \cdot 4\pi R^2 = I \cdot S \qquad \textbf{(3. 11)}$$

The ratio of intensities is given by:

$$\frac{I}{I_0} = \frac{4\pi R^2}{S} \qquad \textbf{(3. 12)}$$

and the Directivity Index (DI) is obtained by:

$$DI = 10\log_{10}\frac{I}{I_{Ref}} = 10\log_{10}\frac{4\pi R^2}{S} \qquad \textbf{(3. 13)}$$

For an array with length, **L**, and wave length, **λ**, (with $L>>\lambda$) the Directivity Index is given by:

$$DI = 10\ \log_{10}(2L/\lambda) \qquad \textbf{(3. 14)}$$

The acoustic energy returned from the seafloor is the energy used by sonar systems, as well as the remote means by which to deduce some seafloor properties. Knowledge of the beam angle and the sound velocity profile in the water column allows one to obtain backscatter strength corrected for absorption and spherical spreading.

Each particle on the seafloor can be likened to a reflector and the seabed return as the sum of the energy contributions from the water-seabed interface and from the volume of sediments, due to some energy penetration into the sediments. However, the contribution from the volume of sediments is less significant when using high frequencies.

The seabed Backscattering Strength (**BS**) is usually described as the logarithmic sum of intrinsic backscattering strength per unit area or backscatter index (SB), which is dependent on the reflective properties of the seafloor and the effective instantaneous scattering area **A,** the area of the seafloor which contributes to the backscattered signal:

$$BS = SB + 10\ \log_{10} A\ \text{dB}. \qquad \textbf{(3. 15)}$$

The limits of the backscattering area are defined by the beam geometry, particularly by the beam width (of the transmit beam) in the along-track direction at normal incidence or nadir, $\boldsymbol{\phi_T}$, and by the beam width (of the receive beam) in the across-track direction at nadir, $\boldsymbol{\phi_R}$.

For the off-nadir directions, the backscattering area is bounded by the beam width, $\boldsymbol{\phi_T}$, and by the transmitted pulse length $\boldsymbol{\tau}$ (Figure 3.3). The seafloor backscattering strength may be given by:

$$BS = \begin{cases} SB + 10 \cdot \log_{10}\left(\phi_T \phi_R R^2\right) & \text{beam width constrict} \\ SB + 10 \cdot \log_{10}\left(\dfrac{c\tau}{2\sin\beta}\phi_T R\right) & \text{pulse length constrict} \end{cases} \quad \textbf{(3. 16)}$$

where **R** is the slant range from the transducer to the point on the seafloor, **c** is the velocity of the sound and **β** is the beam angle with reference to the vertical.

The backscatter coefficient, **SB**, is usually partially dependent upon the angle of incidence, the largest variation being near nadir and typically follows a Lambert's law [Urick, 1975 and de Moustier, 1993] dependence at larger angles of incidence. It is common to define:

$SB = BS_N$, for normal incidence ($\beta = 0°$)
$SB = BS_O \cdot \cos^2 \beta$, for oblique incidence ($\beta > 10\text{-}25°$)

Typically, BS_N will be about -15 dB and BS_O about -30 dB. These values may change within ±10 dB or even more depending on seabed type and roughness.

Looking at the beam footprint (ensonified area), Figure 3.3, the instantaneous ensonified area, **A**, is a function of the transmitted beam width, $\boldsymbol{\phi_T}$. The number of samples per beam depends on the sampling interval ($\boldsymbol{\tau_S}$).

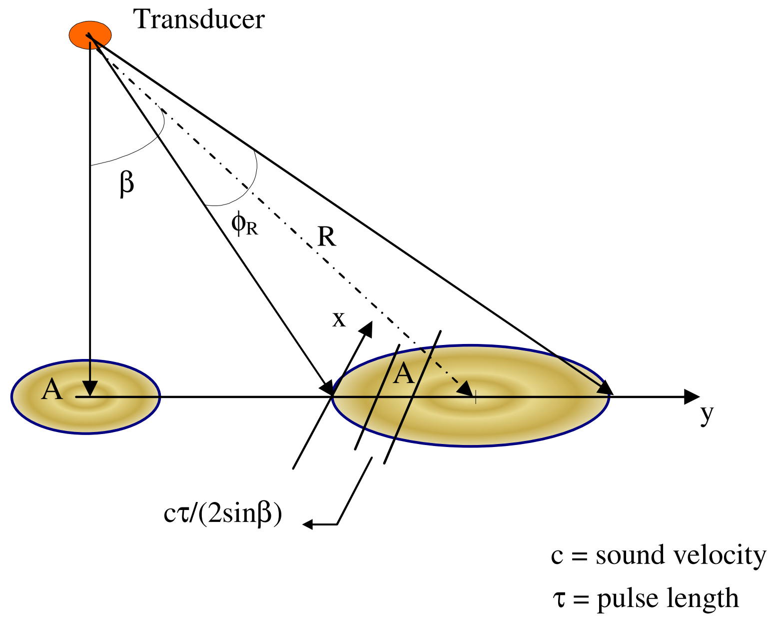


Fig. 3.3 "Backscatter samples"

The Detection Threshold (DT) is a system dependent parameter which establishes the lowest level above which the echo sounder can detect the returning echoes.

2.1.3 Temperature

The temperature at the sea surface varies with the geographic position on the earth, with the season of the year and the time of the day [Pickard and Emery, 1990]. The temperature field distribution is a complex one and can not be predicted with enough accuracy for hydrographic surveys; through the water column the behaviour of the temperature is also very complex. Such unpredictability necessitates a comprehensive distribution of sound velocity profile casts, both temporally and spatially, to maintain a representative currency of the sound velocity profiles for the survey area.

The depth measurement is quite sensitive to variations of the sound velocity profile; a variation of one degree Celsius in temperature translates to approximately 4.5 m/s in sound velocity variation.

The temperature variation is the dominant factor for sound velocity variation between the surface and the lower limit of the thermocline[7], thereafter pressure becomes the principal influence.

2.1.4 Salinity

The salinity is a measure of the quantity of dissolved salts and other minerals in sea water. It is normally defined as the total amount of dissolved solids in sea water in parts per thousand (ppt or ‰) by weight.

In practice, salinity is not determined directly but is computed from chlorinity, electrical conductivity, refractive index or some other property whose relationship to salinity is well established. As a result of the Law of Constancy of Proportions, the level of chlorinity in a sea water sample is used to establish the sample's salinity[8].

The average salinity of sea water is around 35 ‰. The rate of variation of sound velocity is approximately 1.3 m/s for a 1 ‰ alteration in salinity. Typically the salinity is measured with a CTD cast (Conductivity, Temperature and Depth) using the observable electrical conductivity, see 2.2.1.2.

2.1.5 Pressure

The pressure also impacts significantly on the sound velocity variation. Pressure is a function of depth and the rate of change of sound velocity is approximately 1.6 m/s for every alteration of 10 atmospheres, i.e. approximately 100 metres of water depth[9].

The pressure has a major influence on the sound velocity in deep water.

2.1.6 Density

Water density is dependent upon the previous parameters, i.e. temperature, salinity and pressure.

Fifty percent of the ocean waters have a density between 1027.7 and 1027.9 kg/m^3. The largest influence on density is compressibility with depth. Water with a density of 1028 kg/m^3 at the surface would have a density of 1051 kg/m^3 at a depth of 5000 metres.

---

7 The thermocline is also called discontinuity layer or thermal layer. The thermocline corresponds to a vertical negative temperature gradient in some layers of the water column, which is appreciably greater than the gradients above and below it. The main thermoclines in the ocean are either seasonal, due to heating of the surface water in summer, or permanent.

8 A joint committee (IAPO, UNESCO, ICES, and SCOR) proposed the universal adoption of the following equation for determining salinity from chlorinity: $S = 1.80655\ Cl$.

9 This is derived by the hydrostatic principle, i.e., $p(z) = p_0 + \rho g z$.

2.2 Salinity, Temperature, and Sound Velocity Determination

This subsection describes the instrumentation used for salinity, temperature and sound velocity determination as well as their operating principles and the calculation for mean sound velocity.

2.2.1 Instrumentation

2.2.1.1 Sound Velocity Profiler is the most common instrument used to measure the sound velocity profile through the water column. This instrument has one pressure sensor to measure depth, a transducer and a reflector a certain distance, **d**, apart. The sound velocity is calculated by the equation $c = 2d/\Delta t$, where $\mathbf{\Delta t}$ is the two-way travel time of the acoustic signal between the transducer and the reflector (similar to the depth measurement performed by echo sounders).

2.2.1.2 CTD is an electronic instrument with sensors for conductivity, temperature and depth. This instrument records the salinity by directly measuring the electrical conductivity of the sea water.

Sound velocity in the water varies with the medium's elasticity and density, which are dependent upon the salinity, temperature and pressure. With the information from the CTD (salinity, temperature and pressure) it is possible to calculate the sound velocity in the water based on empirical equations. One simple equation with adequate accuracy was presented by Coppens [Kinsler et al., 1982]:

$$C(Z, T, S) = 1449.05 + T[4.57 - T(0.0521 - 0.00023 \cdot T)] + [1.333 - T(0.0126 - 0.00009 \cdot T)](S - 35) + \Delta(Z) \qquad \mathbf{(3.\ 17)}$$

where **T** is the temperature in degrees Celsius (ºC), **S** is the salinity in parts per thousand (ppt), **Z** is the depth in km, and $\mathbf{\Delta(Z)} \approx 16.3 \cdot Z + 0.18 \cdot Z^2$.

This equation is valid for a latitude of 45º. For other latitudes, **Z** should be replaced by $Z[1 - 0.0026 \cdot \cos(2\varphi)]$, being $\boldsymbol{\varphi}$ the latitude.

2.2.1.3 Thermistors are elements whose electrical resistance depends on their temperature, which depends on the amount of heat radiation[10] falling on it from the sea. Thermistor chains are used to measure the water temperature at several depths through the water column. These chains, usually moored, consist of several thermistor elements, regularly spaced along a cable. A data logger samples each element sequentially and records the temperatures as a function of time.

2.2.2 Instrument operation

Important for successful operation of a sound velocity profiler, before deployment, the profiler should have the correct parameters entered with the required recording settings and be calibrated with the correct atmospheric offset in order to generate reliable depth measurements.

It should be stressed that, during the atmospheric offset calibration, a sound velocity profiler should not be in a pressurised compartment or the calibration will produce biased offsets and thus erroneous depth measurements.

[10] The heat radiation rate is given by Stefan's Law which states that the rate of emission of heat radiation from an object is proportional to the fourth power of its absolute temperature.

Before deployment, the profiler should be in the water for approximately 15 minutes for thermal stabilisation and during a sound velocity cast, it is recommended a constant deployment speed is maintained.

2.2.3 Data recording and processing

Sound velocity profiles should be edited and carefully checked for anomalous depths and sound velocity readings.

In general, velocity profilers record both depth and sound velocity, both downwards and upwards. The two profiles should be compared to confirm they are similar after which the profiles are often meaned to create the final profile, although this not necessarily required, in any event the readings should be compared and additional information removed to allow sorting into ascending or descending order.

2.2.4 Sound velocity computation

After the sound velocity profile has been validated, it can be applied to the survey file. The computation is used to correct depth measurements with sound velocity profile data.

For beams near the vertical, specific case of single beam echo sounders, it is accurate enough to use the average sound velocity in the water column. However, away from nadir, it is necessary to perform ray tracing to take account of the beam curvature due to any refraction phenomenon encountered; this is the procedure used in MBES (see 5.2.1.8.1).

For a signal transmitted vertically (i.e., $\theta_0 = 0°$), the harmonic mean sound velocity, $\mathbf{c_h}$ for a depth $\mathbf{z_n}$ is obtained by:

$$c_h(z_n) = \frac{z_n}{\sum_{i=1}^{n} \frac{1}{g_i} \ln\left(\frac{c_i}{c_{i-1}}\right)} \qquad \textbf{(3. 18)}$$

where $\boldsymbol{g_i}$ is the constant gradient at layer $\boldsymbol{i}$, given by: $g_i = \dfrac{c_i - c_{i-1}}{z_i - z_{i-1}}$.

2.3 Sea water sound propagation

This section focuses on sound propagation namely attenuation, reflection, and refraction.

2.3.1 Attenuation

Attenuation is the loss in energy of a propagating wave due to absorption, spherical spreading and scattering by particles in the water column.

The absorption is the result of dissociation and association of some molecules in the water column; magnesium sulphate ($MgSO_4$) is a major source of absorption in salt water. The rate of absorption is dependent on the physical and chemical properties of sea water and on the acoustic frequency being transmitted. From Figure 3.4 it can be seen that above a frequency of 100 kHz the absorption coefficient increases with the increase of temperature; thus, it can be expected that sonar range will vary with the water temperature.

The spherical spreading is dependent on the geometry; for a solid angle the acoustic energy spreads over a larger area as the distance from the source increases.

Both losses by absorption and spherical spreading are taken into account in the sonar equation (see 2.1.2). However, losses from scattering depend upon particles or bodies present in the water column; scattering is chiefly due to marine organisms, a major source of which is the deep scattering layer (DSL) which consists of a layer of plankton whose depth varies throughout the day.

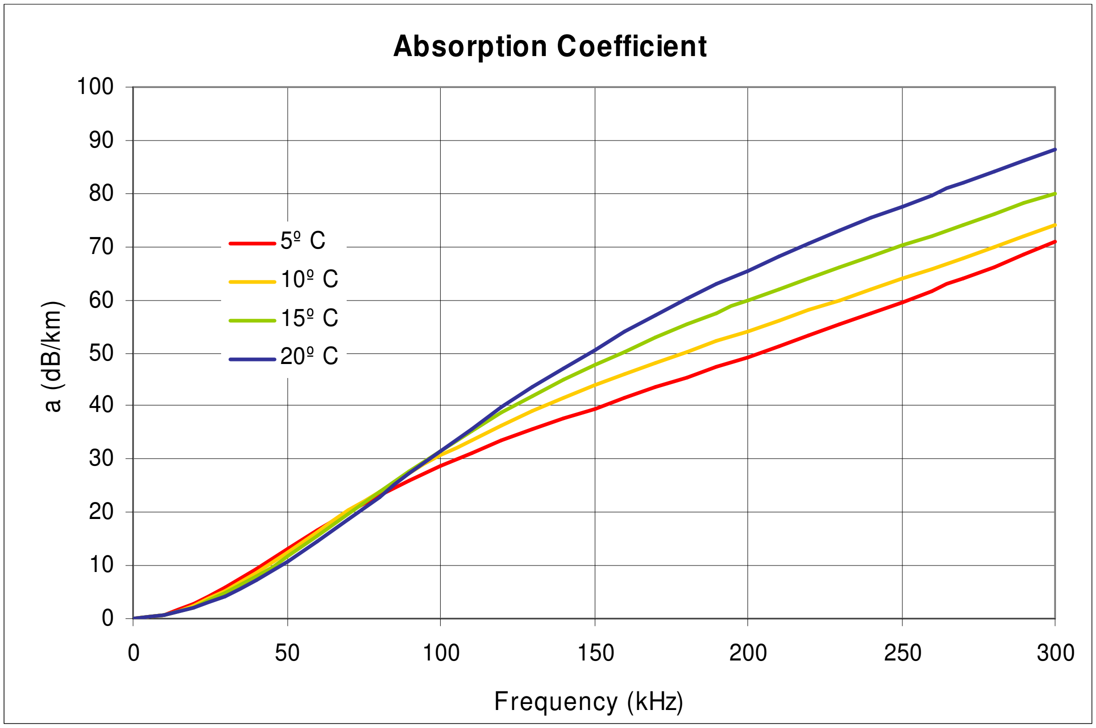


Fig. 3.4 "Absorption coefficient"

2.3.2 Refraction and reflection

Refraction is the process in which the direction of propagation of the acoustic wave is altered due to a change in sound velocity within the propagating medium or as the energy passes through an interface, representing a sound velocity discontinuity between two media.

Consider two media (Figure 3.5) with different sound velocities, $\mathbf{c_1}$ and $\mathbf{c_2}$; if $\mathbf{c_1}$ is greater than $\mathbf{c_2}$, the direction of the acoustic wave propagation is altered and the transmit angle will be smaller than the angle of incident. In contrast, if $\mathbf{c_1}$ is smaller than $\mathbf{c_2}$, the direction of the acoustic wave propagation is changed and the transmit angle will be greater than the angle of incident. For normal incidence no refraction occurs.

For normal incidence and smooth seabeds, the reflection coefficient[11] for the pressure, $\mathfrak{R}$, is obtained by the ratio of the amplitude pressure of the reflected wave by the amplitude pressure of the incident wave [Kinsler et al., 1982].

[11] It is also possible to define reflection coefficients for power and intensity. For normal incidence the reflection coefficient for power and intensity is the square of the reflection coefficient for the pressure.

$$\Re = \frac{P_R}{P_I} = \frac{\rho_2 c_2 - \rho_1 c_1}{\rho_2 c_2 + \rho_1 c_1} \qquad \textbf{(3. 19)}$$

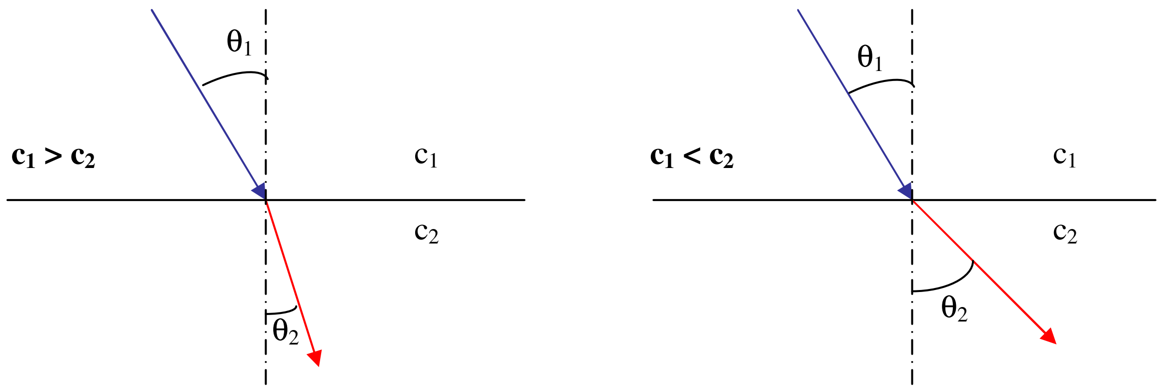


Fig. 3.5 "Refraction principle"

For general conditions, the ratio of acoustic intensity reflected and transmitted depends mainly on:

- Contrast between the acoustic impedances of the media;
- Seafloor roughness;
- Acoustical frequency.

2.4 Acoustic parameters

The characteristics of an echo sounder are determined by the transducers, namely the directivity, beam width, beam steering and side lobes level. In this subsection each of these parameters is analyzed.

2.4.1 Frequency

The echo sounder's acoustic frequency is the parameter which determines the range and the sound penetration of sediments. The attenuation of the acoustic signal in the water is proportional to the frequency. The higher the frequency is, the higher the attenuation will be and, consequently, the lower the range and the penetration into the seafloor.

The beam width is dependent on the acoustic wave length and on the size of the transducer. For the same beam width a lower frequency will require a larger transducer.

The frequencies of bathymetric echo sounders are typically:

- Waters shallower than 100 metres: frequencies higher than 200 kHz;
- Waters shallower than 1500 metres: frequencies from 50 to 200 kHz;
- Waters deeper than 1500 metres: frequencies from 12 to 50 kHz;

The frequencies for sediment echo sounders are below 8 kHz.

2.4.2 Band width

Taking $\mathbf{f_0}$ as the frequency of maximum power transmission (resonance frequency) and $\mathbf{f_1}$ and $\mathbf{f_2}$ as the frequencies corresponding to one half of that power, the band width is the frequency interval between these frequencies (Figure 3.6), i.e. $W = f_2 - f_1$

The transducer's quality factor, **Q**, is given by,

$$Q = \frac{f_0}{W} \qquad \textbf{(3. 20)}$$

From the definitions above one can see that **Q** and **W** have reciprocal variation. Hence, to optimize the transmission power, the transducer should transmit close to the resonance frequency and therefore have a small band width, i.e. a high quality factor value.

During reception it is necessary to have good echo discrimination from any other signal. Although, it must be well defined in the frequency range, the transducer band width should satisfy $W \geq 1/\tau$, where $\boldsymbol{\tau}$ is the pulse duration.

The optimized solution is to have a transmission transducer with high **Q** and a reception transducer operating in the same frequency of resonance but with a low **Q**.

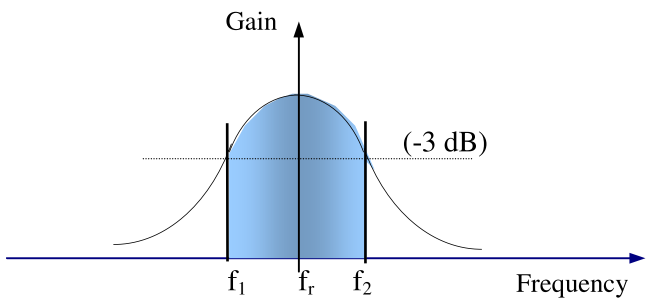


Fig. 3.6 "Transducer band width"

2.4.3 Pulse length

The pulse length determines the energy transmitted into the water; for the same power, the longer the pulse length, the higher the energy put into the water will be and thus the greater the range that can be achieved with the echo sounder.

To take advantage of the transducer resonant frequency, the pulse duration should be at least half its natural period. The drawback of longer pulses is the decrease in vertical resolution of two adjacent features (Figure 3.7).

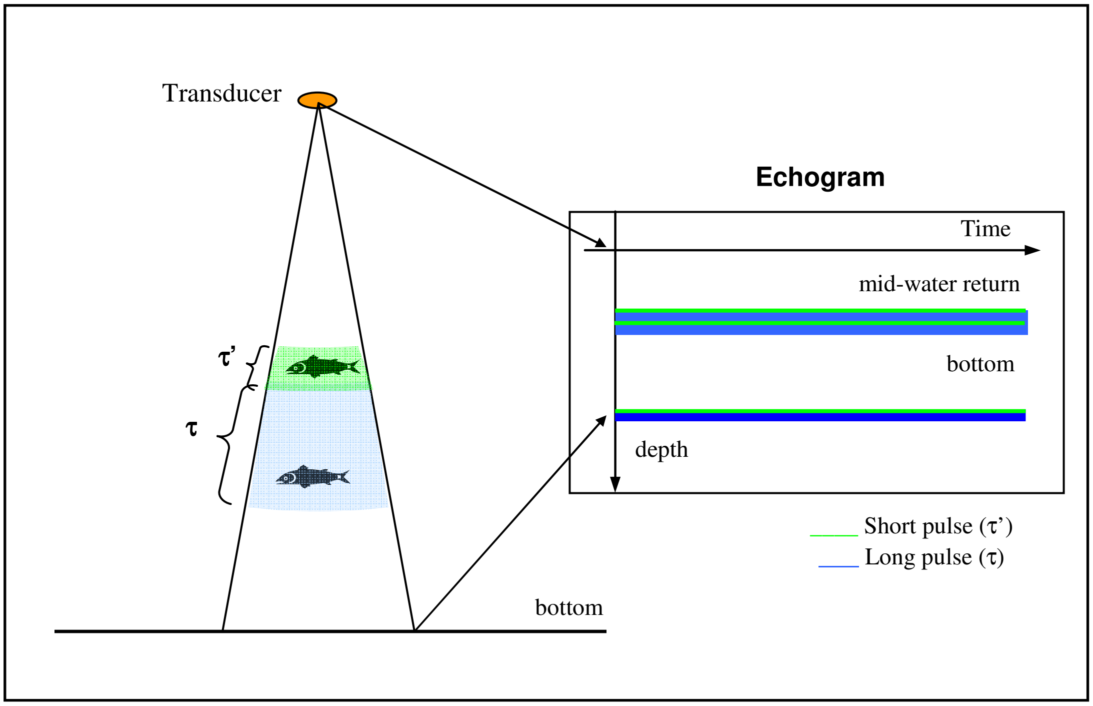


Fig. 3.7 "Resolution as a function of pulse length"

3. MOTION SENSORS

Being able to correct the observed depths and their positioning for vessel motion, i.e. attitude (roll, pitch, and heading) and heave was a considerable achievement and advance in hydrographic surveying quality and accuracy. For this purpose inertial sensors with a heading sensor (usually a gyro or fluxgate compass) or inertial sensors with the integration of GPS information are used to measure the attitude and heave of the survey vessel.

The attitude of the vessel consists of three rotations about the conventional three orthogonal axes defined for the vessel. Hereafter the vessel co-ordinate system is defined as a right-hand system with the x axis pointing towards the bow, the y axis pointing towards starboard and the z axis pointing downwards. In this reference system roll corresponds to a rotation about the x axis (the roll is positive when the starboard side is down), pitch corresponds to a rotation about the y axis (the pitch is positive when the bow is up), and yaw corresponds to a rotation about the z axis (the yaw is positive for a clockwise rotation).

To transform the collected data, referred to the vessel reference frame, to the local co-ordinate system it is necessary to perform rotations according to the sensed attitude. Hereafter, by convention, the local co-ordinate system is defined as a left-hand system with the x axis pointing to East, the y axis pointing to North, and the z axis pointing downwards.

This next section covers the fundamentals of motion sensing and measurement accuracy.

3.1 Principles of operation

3.1.1 Inertial sensors

Inertial sensors or Inertial Measurement Units (IMU) are the most common sensors used in hydrography for roll, pitch and heave measurements. These sensors apply Newton's laws for motion and consist of three accelerometers, mounted in tri-orthogonal axes, and three angular rate sensors placed in the same frame which thus experience the same angular motions as the vessel (strap down system). The output from this triad of accelerometers provides a good estimation of the gravity vector, the direction from which small angular displacements of the vessel are measured. The triad of angular rate sensors measures the angular displacements (i.e., roll, pitch, and yaw).

The heave is determined by the double integration of the linear acceleration sensed by the vertical accelerometer.

The data from the accelerometers are low-pass filtered to remove high frequency variations from the apparent vertical due to swell, quick turns or sudden velocity variations. On the other hand, the data from the angular rate sensors are high-pass filtered to remove the low frequency movements. Therefore the filter output is the platform attitude with frequency above the selected cut-off frequency (usually, cut-off frequencies of 5 to 20 seconds are acceptable).

When the vessel undergoes any systematic acceleration, whose duration exceeds the time constant of the low-pass filter applied to the accelerometers, such as prolonged turns or velocity variation, the centripetal or tangential acceleration are perceived as prolonged horizontal accelerations which can not be filtered out by the low-pass filter. The result is the deflection of the apparent vertical from the true vertical with the consequent errors on angular measurements (pitch and roll).

The combination from the two filters (low- and high-pass filters) and the relationship between the two pass bands establishes the sensor characteristics.

These inertial sensors, especially for heave, are very sensitive to the interval of time used for the integration. The equivalent cut-off frequency should be tuned to an adequate value, in order to detect the longer waves without dumping or attenuating the shorter ones.

3.1.2 Inertial sensors with the integration of GPS information

The integration of GPS information provides a means of determining the vessel's heading through the use of two GPS antennas in a base line, usually oriented longitudinally to the vessel's bow.

The velocity and rate of turn information provided by a GPS receiver and by the angular rate sensors can be used to compute the centripetal acceleration. Taking into account this information, roll and pitch measurements are compensated for the deflection of the apparent vertical. The output from this sensor is roll and pitch of higher accuracy which is not susceptible to any horizontal accelerations.

3.2 Roll, pitch, and heave measurement

Since the mid 1990s, affordable and accurate motion sensors have been utilised in hydrographic surveying. It is now considered an essential requirement not only for multibeam but also for single beam surveys when using automatic data acquisition systems. These sensors are used to compensate the vessel's motion for roll, pitch and heave.

The calculated depths must take into consideration the data from the motion sensor, i.e. the value from the oscillation of the survey vessel about the longitudinal axis (roll - θ_R), the value from the oscillation of the survey vessel about the transversal axis (pitch - θ_P), the vessel's heading (α) and the vertical oscillation (heave) see Figure 3.8 and Annex A.

3.3 Heading

The recording and application of vessel heading is essential for swath surveying systems. However, for single beam surveys the effect of heading variation (yaw) during a rotation is not significant if the positioning antenna and the transducer are located in the same vertical axis. When the positioning antenna and the transducer are not in the same vertical axis, for accurately positioned depths, it is necessary to take account of the vessel's heading.

For real-time heading measurement, several methods and equipment are available, such as: gyro compasses, fluxgate compasses and carrier-phase DGPS.

The heading measurement based on carrier-phase is conducted in the inertial sensors which integrates DGPS information. This solution allows for high accuracies.

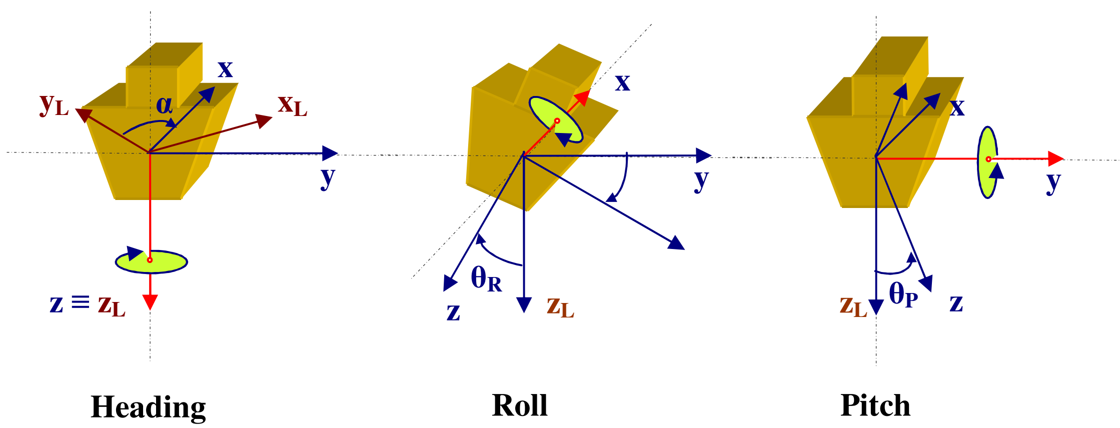


Fig. 3.8 "Vessel attitude"

3.4 Accuracy of measurement

The accuracy of roll, pitch, heave and heading should be as higher as possible. Presently available motion sensors are sufficiently accurate to be used in almost all survey orders. However, during horizontal accelerations of the survey vessel, either centripetal or tangential accelerations, inertial sensors, when used alone, have measurement biases due to the deviation of the apparent vertical.

For MBES it is recommended the inertial sensors are combined with the integration of DGPS information so that the effect of horizontal accelerations can be minimised. Usual systems accuracies, at 95% confidence level, are 0.05° for roll and pitch, 0.2° for heading and 10 centimetres or 10% of heave height whichever is greater.

During vessel turns the heave measurement is degraded due to the centripetal acceleration. Typically it is necessary to wait an interval of ten times the cut-off period to regain accurate measurements after the turn to allow stabilisation and settling of the measuring unit.

4. TRANSDUCERS

The transducers[12] are one of the echo sounders' components; it is transducer characteristics which dictate some of the operating features of an echo sounder. For this reason it is particularly important to study their operating principles, characteristics and related issues such as: beam width, directivity, beam steering, installation and coverage.

The transducers are the devices used for transmission and reception of the acoustic pulses. They operate by converting electrical energy into mechanical energy, i.e. transducers convert electrical pulses from a signal generator to longitudinal vibrations which propagate into the water column as a pressure wave [Seippel, 1983]. During the reception, reciprocally, the pressure waves are converted into electrical signals.

This section describes transducer classifications with regard to: operating principles, beam, beam width, and installation, at the end of the section the assessment of the ensonification is presented.

4.1 Classification with Regard to Operation

Transducers are classified with regard to their operating principle, i.e. magnetostrictive, piezoelectric, and electrostrictive.

4.1.1 Magnetostrictive

These transducers have an axis of iron with a coil of nickel. A D.C. (direct current) current or pulse through the axis generates a magnetic field in the coil which produces a contraction and consequently a reduction of its diameter. When the electric current along the axis stops the coil returns to its original size.

The application of an A.C. (alternating current) signal generates contractions and expansions according to the characteristics of the applied signal. The amplitude of the induced vibration will be a maximum if the frequency is equal or harmonically related to the transducer material's natural frequency or frequency of resonance[13].

This type of transducer is, however, less efficient than transducers which operate with the piezoelectric effect.

4.1.2 Piezoelectric

These transducers are made from two plates with a layer of quartz crystals between. The application of an electric potential across the plates produces a variation in the thickness of the quartz layer (piezoelectric effect). Alteration of the electric potential causes vibration of the quartz and consequently vibration of the entire unit. Reciprocally, the mechanical compression of the crystal produces a potential

[12] The transducer is the device, underwater antenna, used to transmit acoustic pulses and to receive them back. In particular, if the device is used only for transmission it is called a projector and if it is exclusively used for reception, operated in passive mode, it is called a hydrophone.

[13] This phenomenon corresponds to the re-enforcement or prolongation of any wave motion, such as acoustic waves. The frequency of resonance is the frequency at which a transducer vibrates more readily.

difference between opposite crystal faces. The amplitude of the vibration will be a maximum if the frequency of the electric potential matches the quartz natural frequency.

4.1.3 Electrostrictive

These transducers are based on the same principle of the piezoelectric transducers. However, the materials used (usually polycrystalline ceramics or certain synthetic polymers) do not have naturally piezoelectric characteristics, therefore during the manufacturing processes they need to be polarized.

Electrostrictive transducers are used almost exclusively these days. These transducers are lighter, reversible and can be arranged in arrays. These arrays with a set of small elements, when properly arranged, allow, according to the product theorem (see 4.2, equation 3.26), similar characteristics to a single piece transducer.

4.2 Beam width

The pressure amplitude generated by a transducer, in polar co-ordinates, is given by the product:

$$P(r,\theta) = P_{ax}(r) \cdot h(\theta) \qquad \textbf{(3. 21)}$$

Where $\boldsymbol{\theta}$ is the angle referred to the beam axis, line of maximum pressure amplitude and intensity, **r** is the range from a particular point to the transducer, $\mathbf{P_{ax}(r)}$ is the pressure amplitude in the acoustic beam axis, and $\mathbf{h(\theta)}$ is the directional factor which corresponds to the relative signal strength. The directional factor is normalized for $\theta = 0$, i.e. $h(0) = 1$, hence $P(r,0) = P_{ax}(r)$.

The transducer directivity is usually represented by a beam pattern diagram $B(\theta) = h^2(\theta)$ or in a logarithmic scale as: $b(\theta) = 10 \cdot \log_{10}(B(\theta)) = 20 \cdot \log_{10}(h(\theta))$.

The transducer can be characterized by its beam width $\mathbf{b_w}$; this is commonly defined by the angle at the -3 dB level, that is to say, the angular aperture corresponding to half power referred to the beam axis $b_w = 2\theta_{-3dB}$, see Figure 3.9.

The depth measurement is performed in any direction within the cone defined by the beam width.

The beam width is related to the physical dimensions of the transducer and to the frequency of the acoustic pulses. For instance, the beam width at the -3 dB level from a circular piston transducer, with a **D** diameter, can be approximated by:

$$b_w = 60\ \lambda / D \text{ (degrees)}, \qquad \textbf{(3. 22)}$$

and for a rectangular face transducer, with length **L** and width **W**, the beam widths at the -3 dB level in the two dimensions can be respectively approximated by:

$$b_w = 50\ \lambda / L \text{ and } b_w = 50\ \lambda / W \text{ (degrees)}, \qquad \textbf{(3. 23)}$$

For a linear array of **N** omni directional transducer elements, with a distance **d** apart, the sum of the signals from the elements generates a directional beam pattern diagram (figures 3.10 and 3.11).

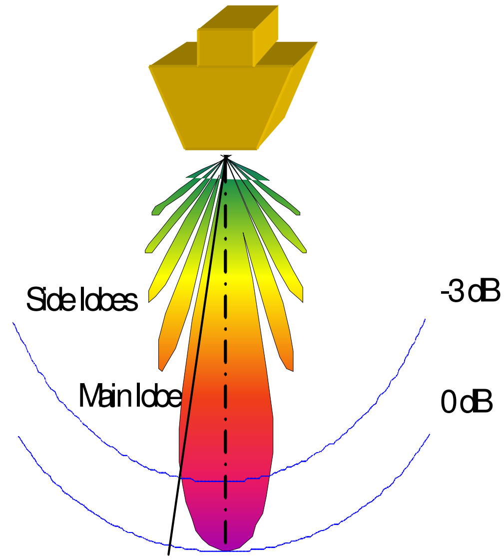


Fig. 3.9 "Beam width defined by the angle at a –3dB level"

The acoustic axis direction is perpendicular to the face of the transducer array. The beam width, at the -3 dB level, is given approximately by:

$$b_w = 50\, \lambda / ((N\text{-}1)d) \text{ (degrees)} \qquad \textbf{(3. 24)}$$

where **λ** is the acoustic wavelength.

The directional factor of the array of transducer elements is given by [Kinsler et al., 1982]:

$$h_{array}(\theta) = \left| \frac{\sin(N\pi \frac{d}{\lambda} \sin\theta)}{N \cdot \sin(\pi \frac{d}{\lambda} \sin\theta)} \right| \qquad \textbf{(3. 25)}$$

Product theorem – is a law of acoustics which defines the directional factor of an array of **N** transducer elements as the product of the directional factor of the transducer element by the directional factor of the array, i.e.

$$h(\theta) = h_e(\theta) \cdot h_{array}(\theta) \qquad \textbf{(3. 26)}$$

and the pressure amplitude is given by:

$$P(r, \theta, \phi) = P_{ax}(r) \cdot h_e(\theta, \phi) \cdot h_{array}(\theta, \phi)\,. \qquad \textbf{(3. 27)}$$

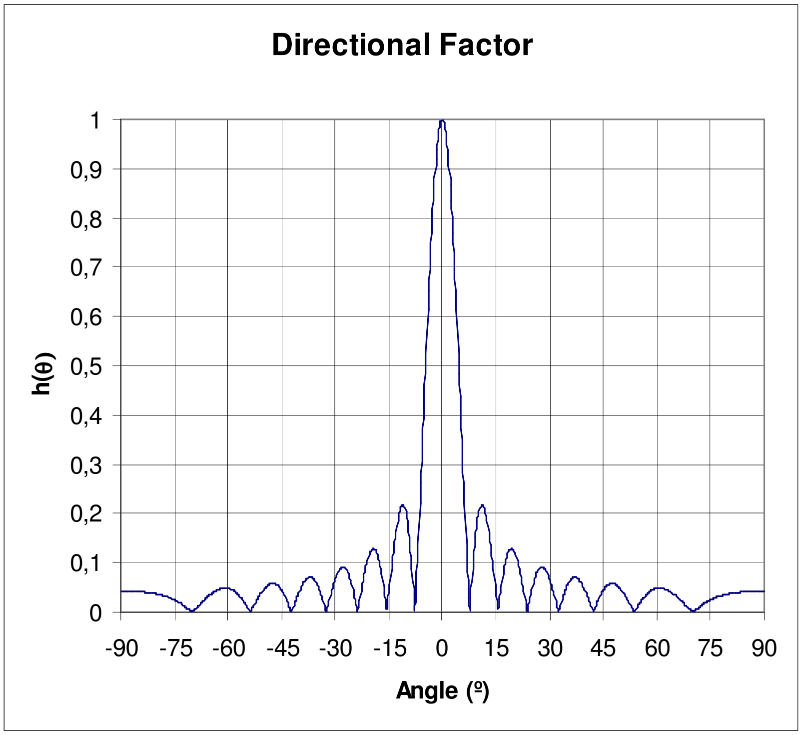


Fig. 3.10 "Directional factor"

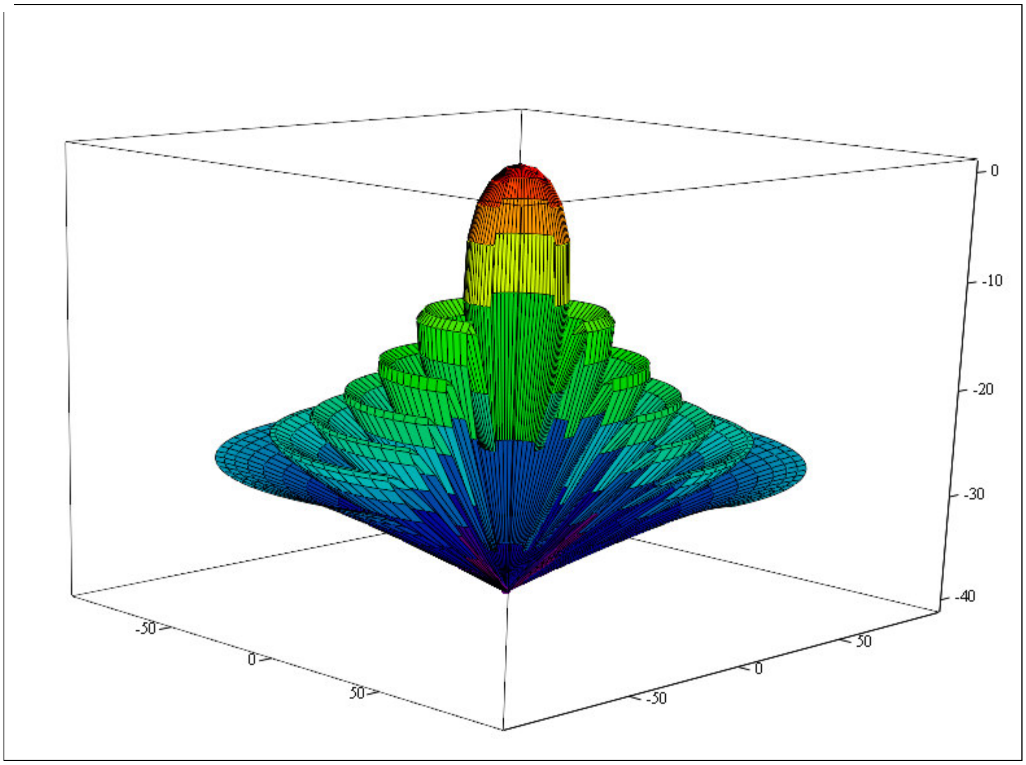


Fig. 3.11 "Beam pattern of a beam perpendicular to the transducer face"

For a linear element, the directional factor is:

$$h_{linear}(\theta) = \left| \frac{\sin(\pi \frac{L}{\lambda} \sin\theta)}{\pi \frac{L}{\lambda} \sin\theta} \right| \qquad \textbf{(3. 28)}$$

The directional factor for an array is only valid for the far field, i.e. in the area where two waves generated by a central and an outer element from the same array have a phase difference smaller than 180 degrees.

$$k\sqrt{R^2 + \left(\frac{L}{2}\right)^2} - kR \le \pi \qquad \textbf{(3. 29)}$$

where **k** is the wave number, i.e. $k = 2\pi/\lambda$.

For instance, for a frequency of 100 kHz and a linear transducer array with L = 0.5 m, the far field corresponds to distances greater than 4.0 metres. This is usually the restriction for the minimum depth measurement.

In the near field, interfering processes create a more complex representation of the acoustic pressure.

The transducer or transducer array beam axis is normal to the transducer face. To form beams not normal to the face of the array, it is necessary to steer the beam. This process is achieved through beam steering techniques.

One array with **N** omni directional elements may steer a beam by introducing a phase or time delay in each element. The corresponding directional factor is (Figure 3.12):

$$h_{array}(\theta) = \left| \frac{\sin\left[N\pi \frac{d}{\lambda}\left(\sin\theta - \sin\theta_{ax}\right)\right]}{N \cdot \sin\left[\pi \frac{d}{\lambda}\left(\sin\theta - \sin\theta_{ax}\right)\right]} \right| \qquad \textbf{(3. 30)}$$

The result of this equation is the beam axis steered to the direction $\boldsymbol{\theta_{ax}}$ (Figure 3.13)

Beam steering can be accomplished by introducing a time delay or phase difference to the elements of the array (equation 3.31).

Beam steering has two aims: beam stabilization and beam forming during the reception phase.

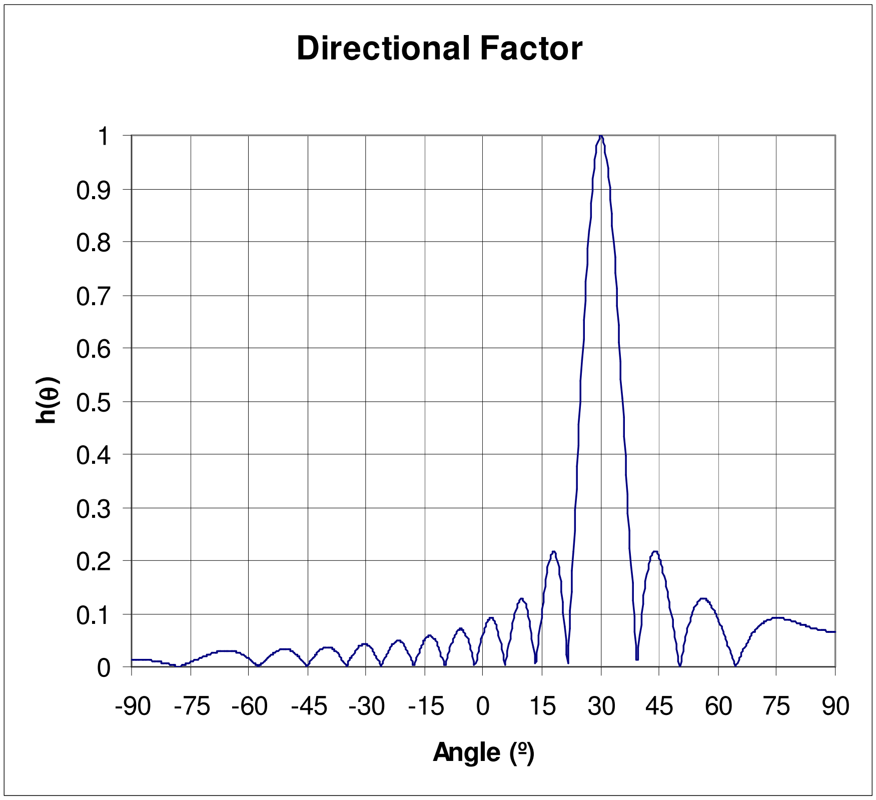


Fig. 3.12 "Directional factor of a beam steered 30 degrees"

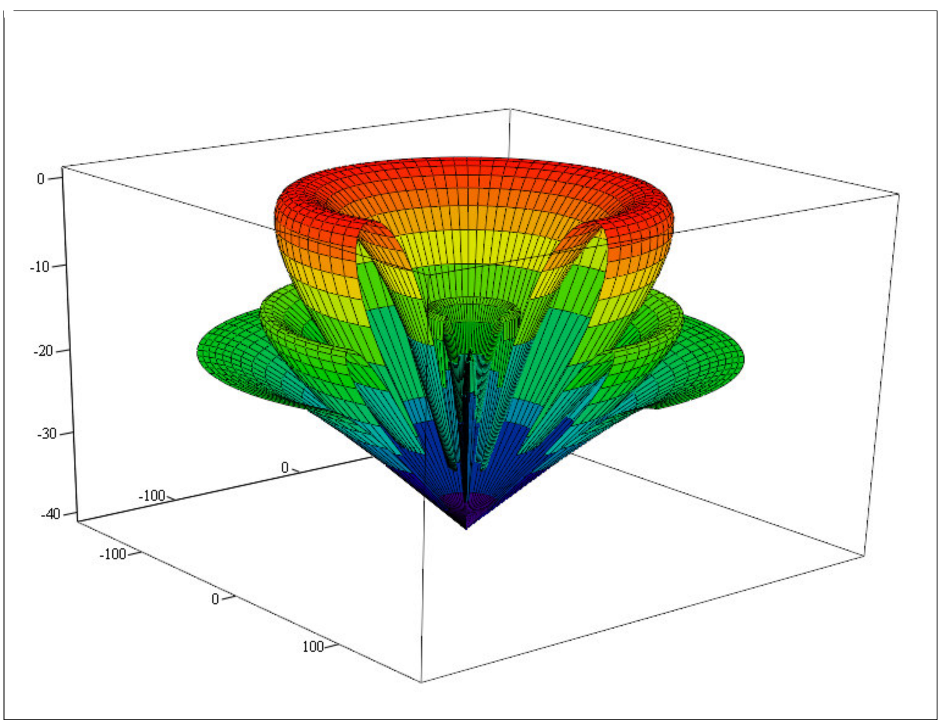


Fig. 3.13 "Beam pattern of a beam steered 30 degrees"

For beam stabilization, it is necessary to measure the angle referred to the normal direction to the array, being the time delay obtained from:

$$\Delta t_n = \frac{nd}{c}\sin(\theta_{ax}) \tag{3. 31}$$

During beam forming, the signals from each element of the array are copied for each beam, the time delay applied to one element from a particular channel or beam is:

$$\Delta t_{n,i} = \frac{nd}{c}\sin(\theta_{ax_i}) \tag{3. 32}$$

where **i** is the order of the beam and **n** is the number of the transducer element.

Considering two dimensionless elements transmitting a pulse with the same frequency but with a time delay, the acoustical axis is steered to the direction where the wave fronts, coming from the two elements, arrive at the same time (Figure 3.14).

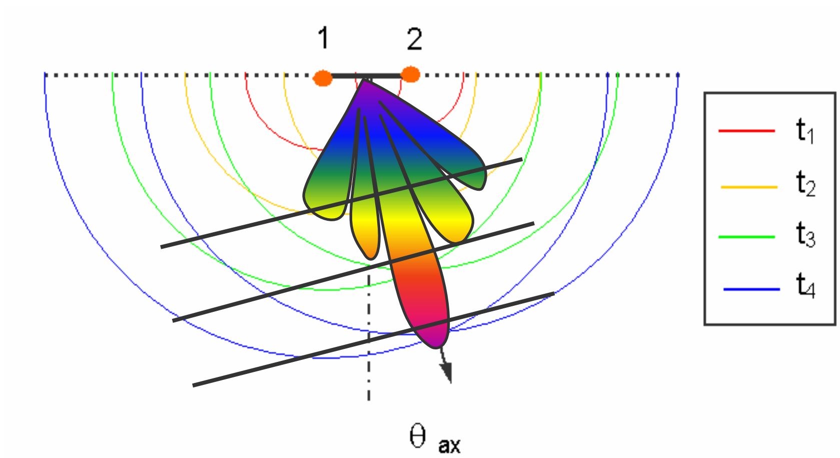


Fig. 3.14 "Illustration of a beam steered θ_{ax} with two transducer elements."

The beam width, usually defined at the -3 dB level, increases with the steering angle, i.e.

$$b_{w_i} = 50\frac{\lambda}{(N-1)d\cdot\cos(\theta_{ax_i})} \text{ (degrees)} \tag{3. 33}$$

Due to the beam's conical shape, when steered, its intersection with the plane of an assumed flat seabed results in a hyperbolic footprint (Figure 3.15).

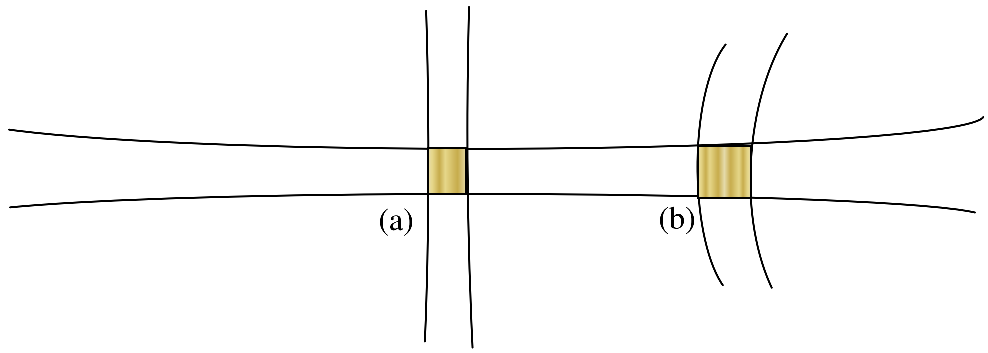


Fig. 3.15 "Linear (a) and hyperbolic (b) footprint."

Side lobes have undesirable effects, such as the detection of echoes corresponding to those lobes. This is the general case for MBES using large beam angles where the erroneous detection is made at the nadir or in the case of rocky spots (Figure 3.16). This effect results in wavy bathymetric contours which sometimes can be identified as an “omega” shape. Lobe reduction is vital for the successful operation of MBES; it is achieved using shading functions, applied during echo reception, corresponding to variable gain applied to the transducer elements.

Considering all transducer elements with the same amplitude, the side lobes will have approximately -13 dB. The technique used to reduce the side lobes consists of superimposing a window, which amplifies the signals from different elements with different gains. These windows are usually symmetric to the centre of the array.

The Dolph-Chebychev window is quite often used; this window has the advantage of optimising the level of the side lobes for particular beam widths; it produces the same amplitude level for all side lobes.

The disadvantage of windows superimposition is the resultant reduction of the directivity.

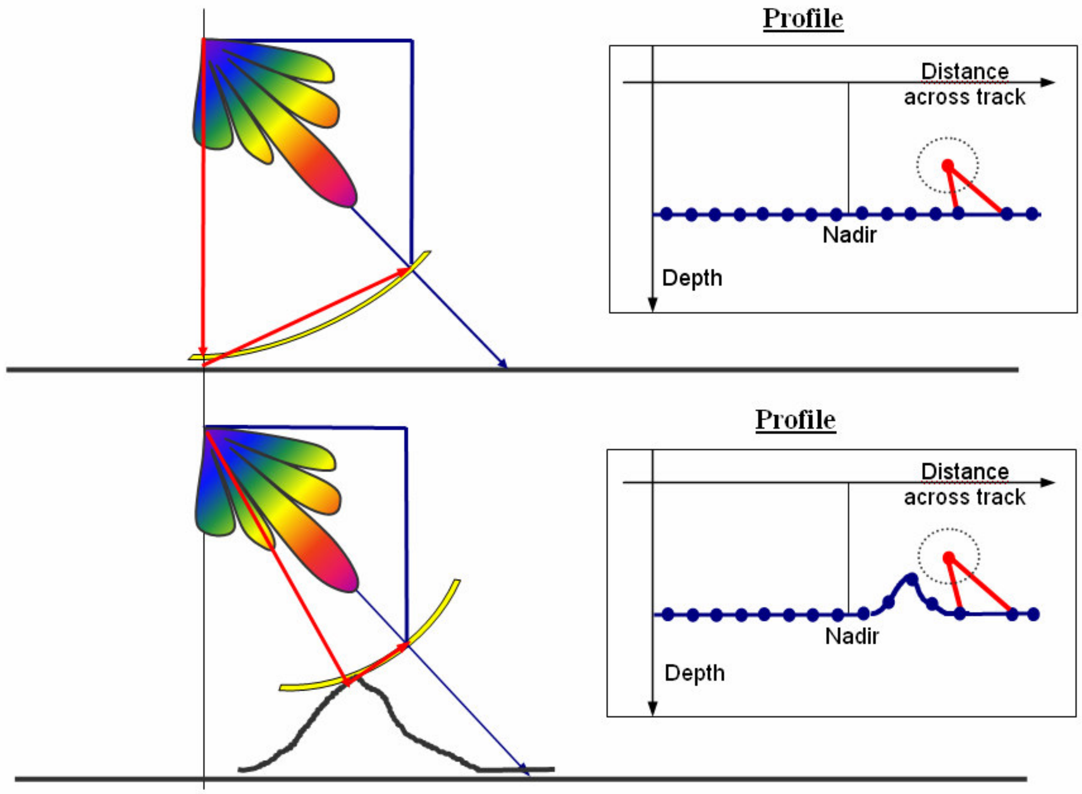


Fig. 3.16 "Depth measurement corresponding to side lobes producing both incorrect depth determination and positioning."

4.3 Classification with Regard to Beam

Echo sounders can be divided into single beam and multibeam. SBES may have transducers either with a single transducer piece or an array. MBES have transducer arrays built up from several elements. As mentioned before, this is a result of the need for beam forming in multiple directions and, sometimes, beam steering to compensate for platform attitude.

4.3.1 Single beam

Single beams require only a transducer, for both transmission and reception, but a transducer array may be used particularly when stabilization is required. Knowledge of roll and pitch angles are needed for beam stabilization.

Beam width is a function of the transducer dimensions and acoustic wave length. The higher the frequency and the larger the transducer is, the narrower the beam will be. Thus to have a narrow beam in low frequencies, a large transducer is required.

The transducer selected for SBES may have a narrow beam when high directivity is required or a wide beam when directivity is not the main concern but the detection of minimum depths or obstacles on the seafloor is the priority.

Wide beams have the capacity to detect echoes within a large solid angle, which is useful for the detection of hazards to navigation requiring further investigation. These beams are usually not stabilized, for common sea conditions the attitude of the transducer does not impact on the measurements.

On the other hand, narrow beams, typically 2º to 5º, are usually required for high resolution mapping (Figure 3.17). These beams might be stabilized in order to measure the depth vertically below the transducer.

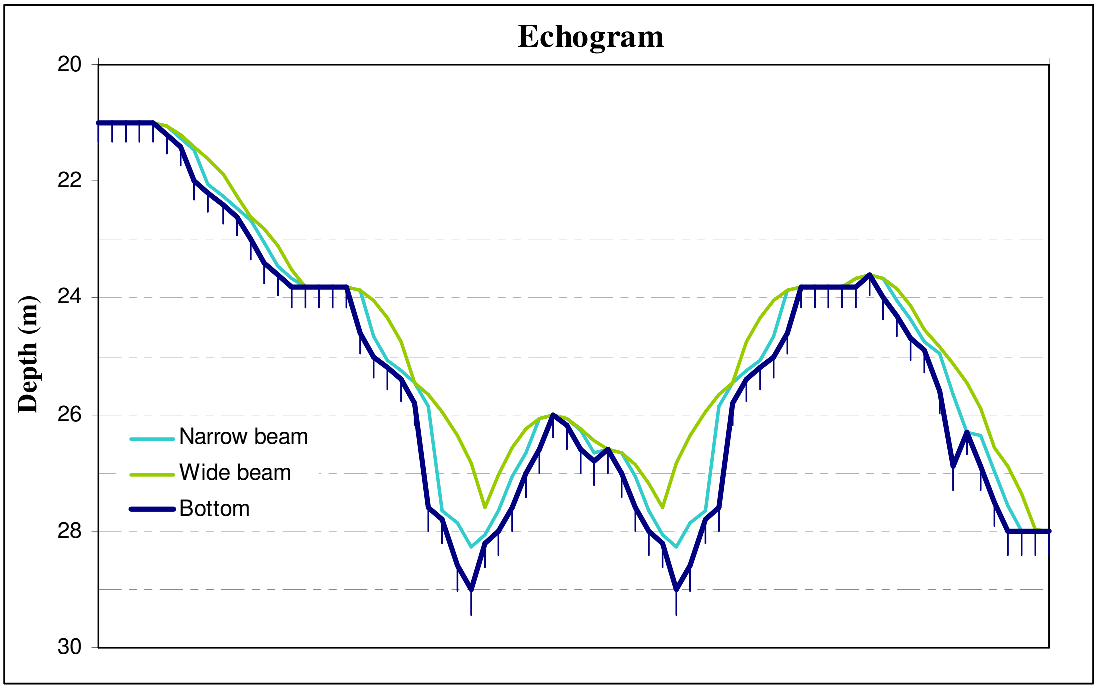


Fig. 3.17 "Illustration of depth measurement using a single narrow- and wide-beam."

4.3.2 Multibeam

MBES usually have separated transducer arrays for transmission and reception, i.e. one projector and one hydrophone, where the first is oriented longitudinally and the second is oriented transversally to the vessel's bow. The most common is to have only one transmitted beam with a fan shape, narrow along track and broad across track.

The reception transducer forms several beams, in predefined directions, narrow across track and broad along track, guaranteeing, regardless of the attitude of the surveying platform, intersection between the transmission and the reception beams.

The reduction of side lobes is essential for correct depth measurement and positioning of MBES. It is common to have side lobes below -20 dB.

4.4 Classification with Regard to Installation

The installation of the transducers on board the survey vessel can be undertaken in several ways. The decision on the way in which the transducer should be installed depends on system portability, keeping it free from vessel noise sources, including turbulent water flow under the keel, and the need to lower it close to the sea floor. The transducer installation can be keel mounted, towed or portable. Each of these is described in the following paragraphs.

4.4.1 Keel mounted

This is the common installation for single beam and multibeam transducers in large vessels, particularly for those designed for deep water surveys.

The installation on the keel can be optionally chosen from:

4.4.1.1 Flush mounted – the transducer is mounted with the face in the hull plane. This option is used either in single beam or multibeam transducers. The advantage is that it does not require a special structure for the installation; the disadvantage may be the vessel noise.

4.4.1.2 Blister – the transducer is mounted in a structure with a small hull shape. This option is used for both single and multibeam transducers. The advantage is the reduction in hull water flow effect at the transducers face; the disadvantage is the need for a special structure for the installation.

4.4.1.3 Gondola - the transducer is mounted in a special gondola shaped structure (Figure 3.18). This option is used for multibeam transducers, particularly for deep water operation. The advantages are the reduction of vessel noise and the elimination of hull water flow noise at the transducer face as it passes in between the hull and the gondola; the disadvantages are the need for a special structure for the installation and consequently an increase in the vessel's draught of the order of a metre.

4.4.2 Towed

The transducer installation in a towed fish is used for side scan sonars when it is essential to have good stability of the transducer, reduction of vessel noise and the ability to lower the transducer close to the seabed.

4.4.3 Portable

This method of installation is commonly used either on single beam and multibeam transducers in small vessels, specifically aimed at shallow water surveys. This installation can be achieved either on the side or over the bow (Figure 3.19) of the vessel. The support structure for the transducer should be rigid and resistant to torque.

Fig. 3.18 "Gondola installation"

Fig. 3.19 "Transducer installation over the bow"

4.5 Coverage

The seafloor coverage, i.e. the ensonified area by SBES, is the area within the beam, where the footprint size is given by (Figure 3.20):

$$a = 2 \cdot z \cdot \tan\left(\frac{\phi}{2}\right) \qquad \textbf{(3. 34)}$$

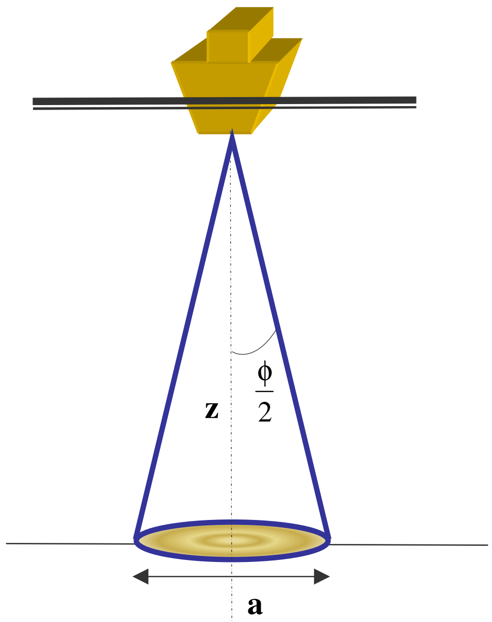


Fig. 3.20 "Single beam coverage"

For multibeam echo sounders, the ensonified area is the result of the intersection of the transmitted and received beam patterns and is dependent upon beam pointing angle, beam width, depth and mean slope of the seafloor. The ensonified area for each beam can be approximated by an ellipse. Taking a flat and levelled seafloor, the length of this ellipse in the athwart ships direction is approximately given by $\mathbf{a_y}$,

$$a_y = \frac{2z}{\cos^2(\beta)} \tan\left(\frac{\phi_R}{2}\right) \qquad \textbf{(3. 35)}$$

where $\mathbf{z}$ is the mean depth, $\boldsymbol{\beta}$ is the beam pointing angle and $\boldsymbol{\phi_R}$ is the width of the reception beam in the athwart ships direction. In the presence of a slope, defined by an angle $\boldsymbol{\zeta}$, the length of the acoustic footprint is approximately:

$$a_y = \frac{2z}{\cos(\beta)\cos(\beta - \zeta)} \tan\left(\frac{\phi_R}{2}\right) \qquad \textbf{(3. 36)}$$

The width or dimension of the footprint ellipse in the fore-aft direction, for a flat seafloor, is approximately given by $\mathbf{a_x}$,

$$a_x = \frac{2z}{\cos(\beta)} \tan\left(\frac{\phi_T}{2}\right) \tag{3. 37}$$

where ϕ_T is the width of the transmitted beam.

The coverage of the seafloor is a function of the dimension of the ensonified areas, beam spacing across-track, ping rate, ship's speed, yaw, pitch and roll. In order to achieve full coverage of the seafloor, the ensonified areas from consecutive pings must overlap one another, so that every single point on the seafloor is ensonified, at least, by one acoustic pulse.

The width of the swath for a flat seafloor is given by,

$$S_W = 2z \cdot \tan\left(\frac{\Delta\theta}{2}\right) \tag{3. 38}$$

where $\Delta\theta$ is the angular coverage between the outer beams of the MBES, effectively used for hydrographic purposes.

5. ACOUSTIC SYSTEMS

In this section, the acoustic systems applied in hydrographic surveying are described. These systems are divided according to their ability to cover the seafloor, i.e. SBES and swath systems, either multibeam or interferometric sonar[14].

5.1 Single beam echo sounders

These echo sounders are devices for depth determination by measuring the time interval between the emission of a sonic or ultrasonic pulse and the return of its echo from the seabed.

Traditionally, the main purpose of the echo sounder is to produce a consistent and high resolution vertical seabed profile on echo trace. The echo trace, after a cautious interpretation, is sampled and digitized manually to produce soundings.

During the last decade, the technology applied in SBES has progressively improved with automatic digitisers, recorders without moving parts and annotation of positions on the echo trace. Recently, built-in computers and signal processors have allowed more sophisticated real time signal processing and data presentation on graphical colour displays, rather than a paper recorder.

5.1.1 Principles of operation

An echo sounder works by converting electrical energy, from the pulse generator, into acoustic energy. As the transducers do not transmit in all directions, the acoustic energy is projected into the water in the form of a vertically oriented beam.

---

[14] A system of determining distance of an underwater object by measuring the interval of time, between transmission of an underwater sonic or ultrasonic signal and the echo return. The name sonar is derived from the words SOund NAvigation and Ranging.

The acoustic pulse travels through the water column and hits the seabed. The interaction with the sea floor results in reflection, transmission and scattering.

The reflected energy which returns to the transducer, the echo, is sensed by the transducer. The strength of the echo decreases rapidly with time, for that reason the echo is automatically adjusted in accordance with its energy level using the Automatic Gain Control (AGC) set in the factory and the Time Variable Gain (TVG) to compensate for the echo's decrease as a function of time. After amplification the electric signal is passed to an envelope detector and compared to the threshold setting to filter the noise from the signal. The output signal is then visualized or recorded.

The resultant observable is the time interval between pulse transmission and echo reception, **t**, being the measured depth given by:

$$z_m = \frac{1}{2} \cdot t \cdot \bar{c} \qquad \textbf{(3. 39)}$$

where $\bar{c}$ is the mean sound velocity in the water column.

5.1.1.1 Echo sounder parameters need to be set correctly in order to obtain high accuracy and a clear record of the seabed. The most important parameters are:

a) Power – The operating range of the echo sounder depends on pulse length, frequency and transmitted power. To optimize the use of the echo sounder, the transmitted power should be kept at the lowest values consummate with adequate detection. Increases in power will result in high levels of echoes but also in higher reverberation levels, creating a poor record. The power is limited by cavitation[15] phenomenon and by the braking stress of the transducer material.

b) Gain – The gain entails signal amplification. The amplification of the signal also amplifies the noise and consequently the data record may be confused. It is recommended that the gain is adjusted according to the seabed type and to the transmission power.

c) Register intensity – This parameter is used in analogue echo sounders to adjust the recording intensity.

d) Pulse length – The pulse length is usually selected automatically as a function of the operating range. The pulse length is responsible for the vertical resolution of the echo sounder, short pulses are necessary for a better resolution. It may be necessary to increase the pulse length in areas with poor reflectivity or with steep slopes.

In shallow waters, where resolution is more important, short pulses must be used. This will reduce the probability of false echoes due to strong reverberation.

e) Scale – Corresponds to the depth scale of the echo sounder recording window. The width of the recording paper is fixed; therefore at small scales one will have a low vertical resolution.

---

[15] This corresponds to the production of small voids in the water. This phenomenon occurs when the acoustic pressure exceeds the hydrostatic pressure.

f) Phase scale – The phase scale is one way to overcome limitations of the recording resolution imposed by the echo trace scale. The phase scale consists of recording just one depth window which should be changed, either manually or automatically, to maintain the seabed recording with a satisfactory vertical resolution no matter the water depth (Figure 3.21).

g) Draught – This parameter corresponds to the immersion of the transducer; in order to record the depth with reference to the instantaneous water level, the draught should be set and verified before survey operations commence and regularly thereafter.

h) Paper's speed – This speed is particularly important and should be selected to ensure good horizontal resolution from depth measurements.

i) Sound velocity – This is the nominal value of sound velocity that should correspond to the mean sound velocity in the area of operation. In surveys with more demanding accuracies, the sound velocity may be set to the sound velocity at the transducer face or to 1500 m/s and then during data processing, the depth must be corrected by applying the actual sound velocity profile.

In classical analogue echo sounders, this parameter does not correspond to the sound velocity but to the value that calibrates the mechanical and electrical echo sounder components to measure the correct water depth.

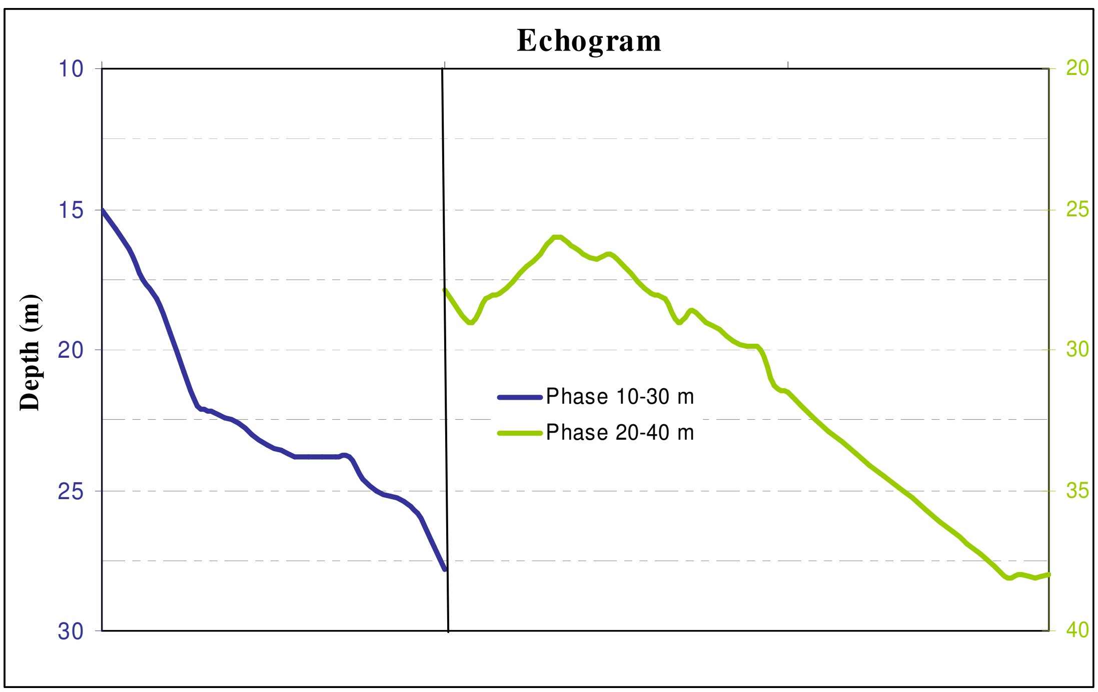


Fig. 3.21 "Recording scales"

The general working principles of SBES were referred to above. However, it is possible to identify two types of echo sounders, digital or analogue.

The traditional analogue echo sounder, whose diagram is presented in Figure 3.22, begins the cycle with the generation of an electrical pulse and the transmission of a burst of energy into the water. After the echo reception and conversion into electrical energy, the low voltage signal is pre-amplified and passed to a recording amplifier, in order to be recorded on an echo trace, which is a graphic record of depth measurements obtained by an echo sounder with adequate vertical and horizontal resolutions. After the recording phase is completed, it is possible to initiate another cycle.

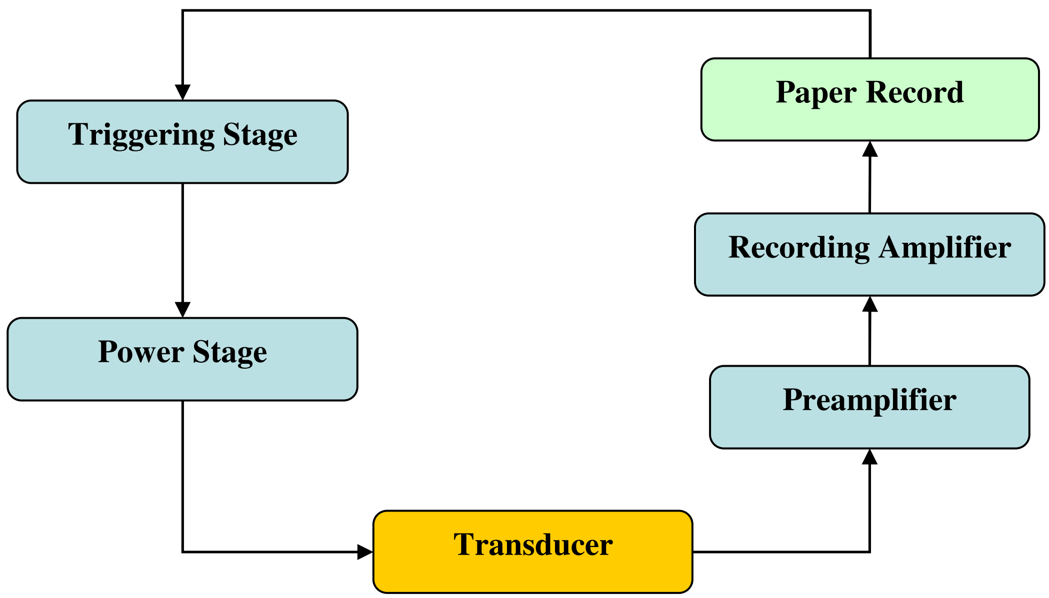


Fig. 3.22 "Analogue echo sounder – block diagram"

Hydrographic echo sounders for shallow waters are usually built with two channels (low and high frequency). The simultaneous recording of two frequencies allows the separation of the seabed return from the soft surface sediments and the underlying rock due to their different acoustic impedances.

The digital echo sounder, see Figure 3.23, works in a similar manner to the analogue echo sounder for the signal transmission. However, during the echo reception, the received signal is amplified as a function of time (time varying gain) and passed through an envelope detector where it is finally converted to digital format, which is the signal that is processed to determine the depth. This allows the information to be stored and displayed in several formats.

5.1.12 <u>Accuracy</u> in the depth measurement is a function of several factors, the echo sounder itself and the medium. Usually, it is necessary to calculate the error budget based on those factors (see 5.1.4).

5.1.1.3 <u>Resolution</u> is the ability to separate returns from two or more objects close together; it is generally expressed as the minimum distance between two objects that can be separated. In depth measurement, a major concern is the vertical resolution of the echo sounder which is dependent on:

a) pulse length - larger pulses have smaller resolution (see 2.4.3). Two objects inside a narrow beam will be recorded as a signal target if they are less than half a pulse length apart; they will be resolved as two separate echoes if they are more than half a pulse length apart;
b) sensitivity and resolution of the recording medium;
c) transmit beam width.

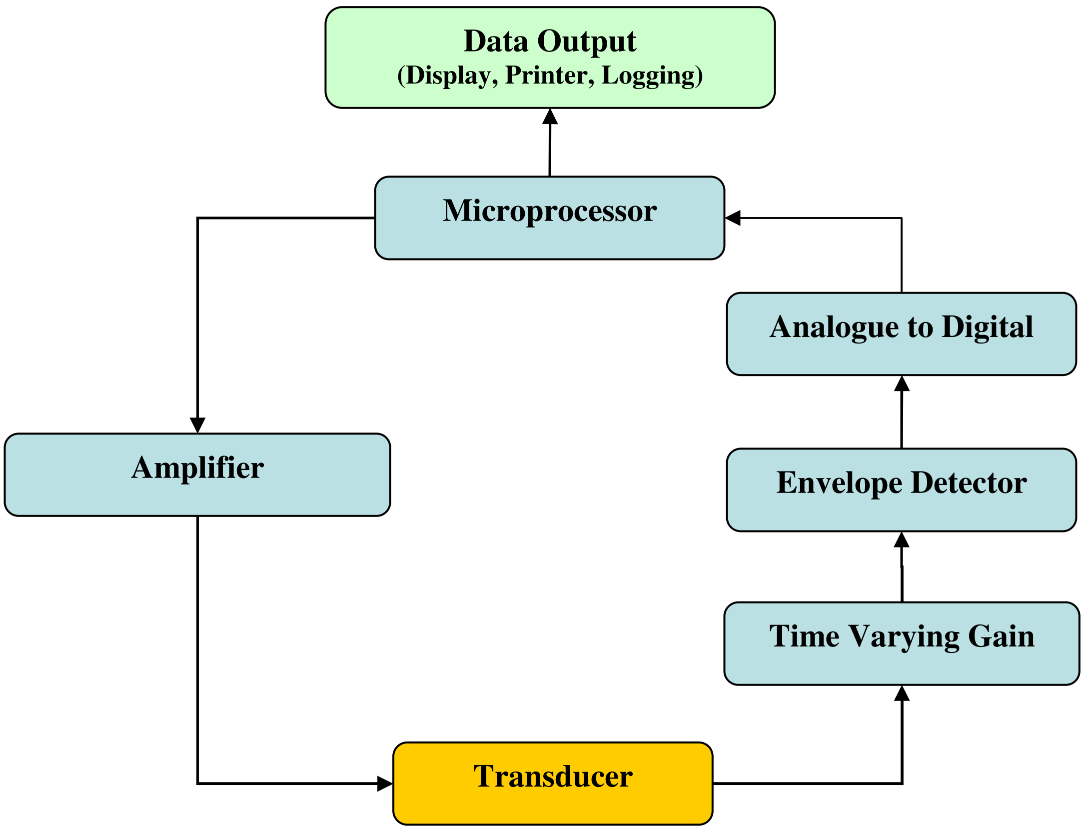


Fig. 3.23 "Digital echo sounder – block diagram"

5.1.1.4 Frequency of an echo sounder is selected based on the intended use of the equipment, i.e. the depth ranges. In some cases it is desirable to use the same device in several depths, for that purpose echo sounders may have more than one frequency and transducer in order to improve the data acquisition and data quality.

The frequencies are often allocated to channels. The echo sounders with two channels are mainly used in shallow and coastal waters; for deep waters, it is usual to use a single low frequency.

5.1.2 Installation and calibration

The transducer may be fixed under the hull or mounted on the side or over the bow. The relevant considerations are that the transducer should be placed, as far as possible, away from the vessel's own sources of noise, deep enough to avoid the surface noise and to stay submerged even in rough seas. It is also very important that the transducer is securely fixed and vertically orientated. It is desirable for the heave compensator and the positioning antenna to be located in the same vertical axis as the transducer.

Echo sounder calibration is a routine task which consists of adjusting the equipment to ensure correct depth measurement. The calibration can be conducted with a bar check or with a special transducer. The purpose is to set the sound velocity parameter so as to adjust the mechanical and electrical components. It may also be possible to correct the measured depths during post processing with the application of the sound velocity profile.

In shallow waters, echo sounder calibration for the average sound velocity in the water column may be performed in the following ways:

a) Bar check consists of lowering a bar or plate underneath the transducer at several depths (for instance, every two metres) either recording the depth error to apply afterwards during the data processing or forcing the echo sounder to record the correct depth from the bar or plate through the adjustment of the sound velocity parameter (Figure 3.24). In such cases the value adopted for calibration is the mean value of the observations. This method should be used down to 20-30 metres.

b) Calibration transducer is an apparatus designed to perform the calibration knowing an exact path length. The calibration procedure consists of making the echo sounder record the correct two-way path inside the calibration transducer by the adjustment of the sound velocity parameter. The calibration transducer is lowered to several depths, each adjustment of the echo sounder, due to the performed measurement, is only valid for the corresponding depth. The calibration value used should be the mean of all the observations. This method should be used down to 20-30 metres.

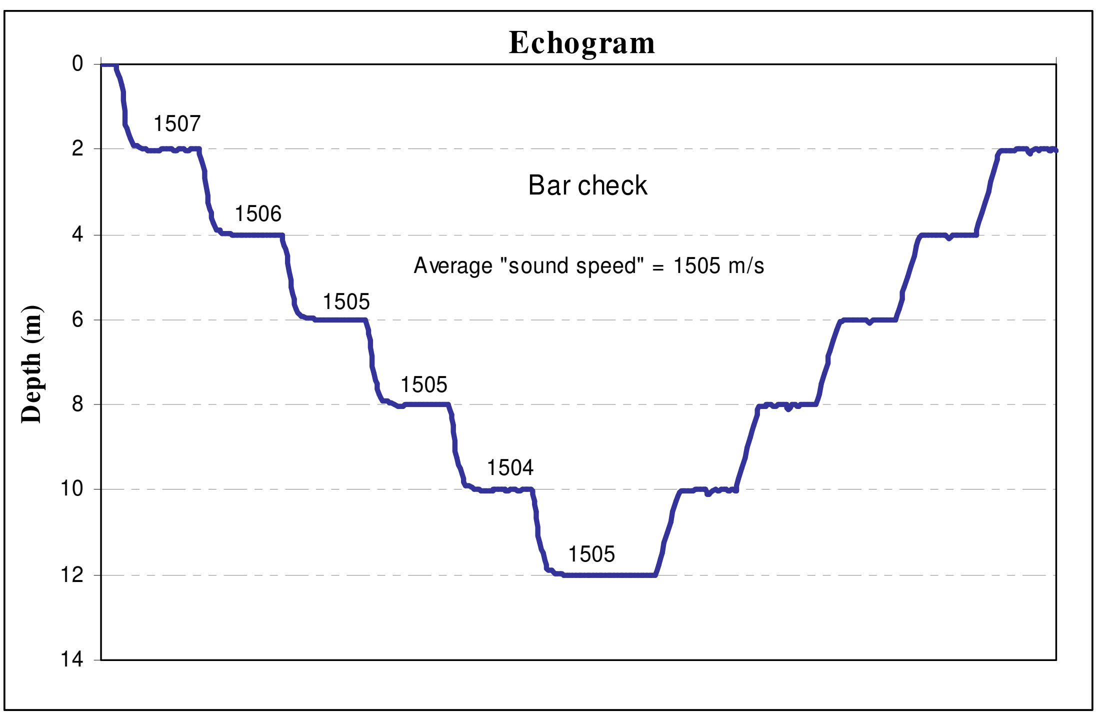


Fig. 3.24 "Bar check illustration"

c) A combined method may also be used with a calibration transducer and a sound velocity profiler. This method is usually used for depths greater than those detailed above. With a sound velocity profile and the adjustment of the true sound velocity at the transducer draft, a similar procedure to that described in b) above is followed. In more modern echo sounders, the sound velocity parameter is set to the actual sound velocity.

   The depth correction is computed during data processing, with the assumption that data was collected using the true sound velocity at the transducer draft. The depth correction is based on the difference between the sound velocity used during data collection and the harmonic mean sound velocity computed from the sound velocity profile.

For depths greater than 200 metres, it is not required to correct the measured depths for sound velocity, a standard sound velocity of 1500 m/s is usually used or values may be selected from the Mathews Tables (N.P. 139).

5.1.3 Operation and data recording

Operation of individual echo sounders should be referred to their operator's manual. Nevertheless, it is important to stress the following aspects:

- Prior to the start of the survey it is necessary to calibrate the echo sounder for the actual sound velocity;
- A general scale, adequate for the expected depths, should be selected;
- The frequency channel should be chosen according to their range capability;
- When using an analogue echo sounder, it is essential to set the gain and the recording intensity to produce a legible trace.

5.1.4 Sources of errors and quality control techniques

Errors in depth determination can be divided into three categories: large errors (blunders), systematic errors and random errors.

Blunder is the terminology used to define errors made by machines; these are caused by defective mechanical or electronic components.

Systematic errors are mainly the result of the offsets (fixed errors) or biases (errors that vary under different conditions) in motion sensing of the survey vessel, misalignment of the transducer and other sensor mounting angles. These errors can be easily corrected if the sign and size of the systematic error can be identified. This category of errors can be determined and removed during calibration of the system.

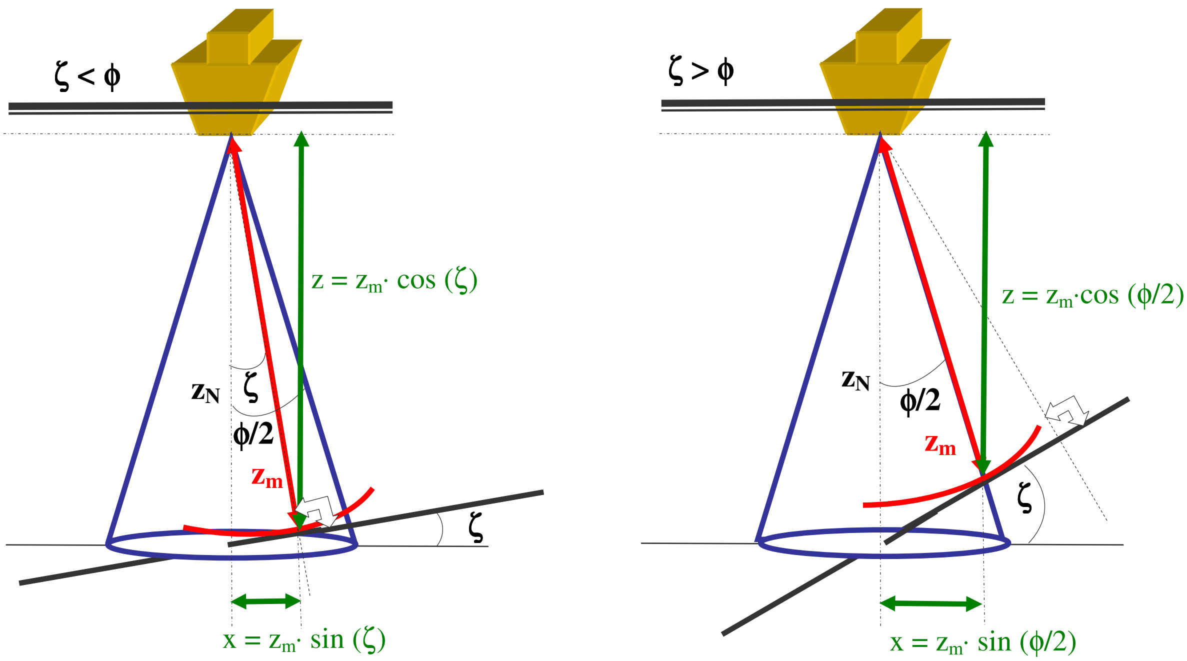
After removing blunders and systematic errors in the depth data, random errors will remain and these can be analysed using statistical techniques.

Hydrographers should be aware of the sources which contribute to the depth error and their individual impact. This section identifies several sources of errors and presents the usual techniques used for quality control.

5.1.4.1 Due to bottom slope - Taking into consideration the different seafloor slopes, in Figure 3.25, the error on the depth measurement, **dz**, depends on both beam width and slope. If no correction is applied, the error in depth will be given by:

$$dz = \begin{cases} z_m\left(\sec(\zeta) - 1\right) & \text{if} \quad \zeta < \dfrac{\phi}{2} \\ z_m\left(\sec\left(\dfrac{\phi}{2}\right) - 1\right) & \text{if} \quad \zeta > \dfrac{\phi}{2} \end{cases} \tag{3.40}$$

where $\frac{\phi}{2}$ is half beam width and $\boldsymbol{\zeta}$ is the slope of the seafloor.



a) slope less than one half the beam width **b)** slope greater than one half the beam width

Fig. 3.25 "Effect of beam width and seafloor slope on depth measurement and positioning"

5.1.4.2 Due to sound velocity. Sound velocity variation is difficult to monitor and produces, in single beam echo sounders, errors in the depth measurement (**dz$_m$**), proportional to the mean sound velocity error or variation (**dc**) and to the depth

$$dz_c = \frac{1}{2} \cdot t \cdot dc \tag{3.41}$$

or

$$dz_c = z \cdot \frac{dc}{c} \tag{3.42}$$

The magnitude of the sound velocity error varies with:

a) accuracy of sound velocity determination;
b) temporal variation of sound velocity;
c) spatial variation of sound velocity.

Note that the depth variance, σ_{zc}^2, due to sound velocity measurement error and to sound velocity variation is written as,

$$\sigma_{zc}^2 = \left(\frac{z}{c}\right)^2 \left(\sigma_{cm}^2 + \sigma_c^2\right) \qquad \textbf{(3. 43)}$$

where σ_{cm}^2 is the sound velocity measurement variance and σ_c^2 is the sound velocity variance due to spatial and temporal variations.

Sound velocity variation, temporal and spatial, is a major external contribution to depth measurement errors. It is important, that during survey planning or at the beginning of the survey, to carry out a number of sound velocity measurements or sound velocity profiles across the survey area distributed throughout the day to assist the hydrographer in deciding on the frequency and location of profiles to be conducted within the survey area during data gathering operations.

5.1.4.3 Due to time measurement. An echo sounder effectively measures time, converting the measurement to depth. The error in time measurement, **dt**, relates directly to depth error, $\mathbf{dz_t}$. In modern echo sounders, time measurement error is usually small and constant. This error is also taken into consideration during calibration.

$$dz_t = \frac{1}{2} c \cdot dt \qquad \textbf{(3. 44)}$$

The major error is a function of identification of the measurement point within the echo, i.e. on the algorithms used for signal detection.

Note that the depth variance, $\sigma_{t_m}^2$, due to time measurement error is written as,

$$\sigma_{zt}^2 = \left(\frac{1}{2} c\right)^2 \sigma_{tm}^2 \qquad \textbf{(3. 45)}$$

where σ_{tm}^2 is the time measurement variance.

5.1.4.4 Due to roll, pitch, and heave. Roll and pitch contribute to the error in depth measurement when the magnitude of those angles is greater than one half the beam widths, $\frac{\phi}{2}$. Figure 3.26 depicts the error in depth measurement and in positioning due to roll, $\boldsymbol{\theta_R}$; this figure can be adapted for pitch changing $\boldsymbol{\theta_R}$ by $\boldsymbol{\theta_P}$.

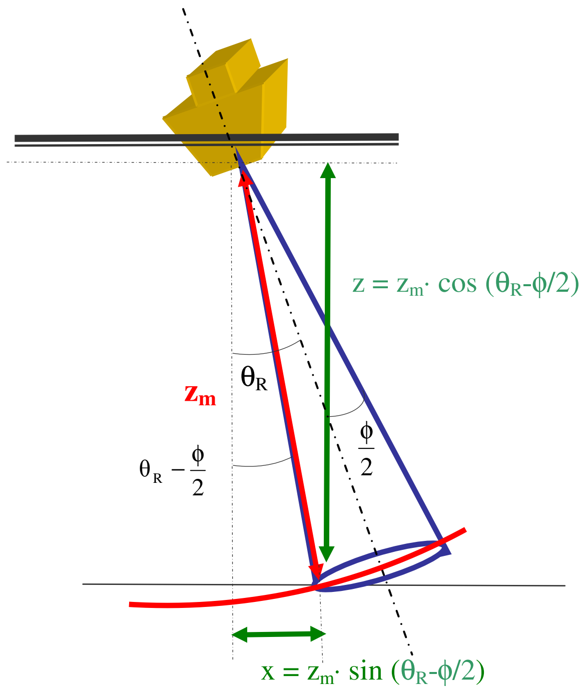


Fig. 3.26 "Effect of beam width and seafloor slope on depth measurement and positioning"

Wide beam echo sounders are usually immune to the roll and pitch of the survey vessel.

For narrow beam echo sounders, this effect may be compensated with beam stabilisation, i.e. keeping the beam vertical regardless of the vessel's attitude or correcting the measured depth and position as follows:

$$dz_{roll} = \begin{cases} z_m\left(1-\sec\left(\theta_R - \frac{\phi}{2}\right)\right) & \text{if} \quad \theta_R > \frac{\phi}{2} \\ 0 & \text{if} \quad \theta_R < \frac{\phi}{2} \end{cases} \qquad \textbf{(3. 46)}$$

The heave (**h**), effect caused by the action of sea and swell on the survey vessel, is measured with inertial sensors or heave compensators. The heave compensator should be placed over the transducer to measure the heave in the same vertical axis.

When using inertial sensors, installation should be close to the centre of gravity of the survey vessel with the known lever arms from the centre of gravity to the transducer; with the roll and pitch instantaneous angles, the measured heave, $\mathbf{h_m}$, can be transferred to the transducer position, $\mathbf{h_t}$, through the application of the induced heave, $\mathbf{h_i}$, (Figure 3.27).

$$h_t = h_m + h_i \qquad \textbf{(3. 47)}$$

To calculate the induced heave, consider the vessel to be a rigid body which is free to rotate around the three axes (x, y & z) as mentioned in 3. The rotation about the centre of gravity (roll and pitch), near which heave is usually measured, corresponds to a transducer depth variation, from the vessel reference frame (identified with the script **V**) to a local co-ordinate system (identified with the script **L**). This difference is called induced heave.

The induced heave, adapted from Hare [1995] for the reference frames defined in 3 and Annex A, is given by:

$$h_i = z_t^L - z_t^V = -x_t^V \sin(\theta_P) + y_t^V \cos(\theta_P)\sin(\theta_R) + z_t^V(\cos(\theta_P)\cos(\theta_R) - 1) \qquad \textbf{(3. 48)}$$

where $\boldsymbol{\theta_R}$ is the roll angle, $\boldsymbol{\theta_P}$ is the pitch angle and ($\mathbf{x_t}$, **yt** & $\mathbf{z_t}$) are the transducer co-ordinates.

The total error on the depth measurement due to heave is therefore,

$$dh = dh_m + dh_i \qquad \textbf{(3. 49)}$$

where $\mathbf{dh_m}$ is the error in heave measurement and $\mathbf{dh_i}$ is the error in the induced heave determination.

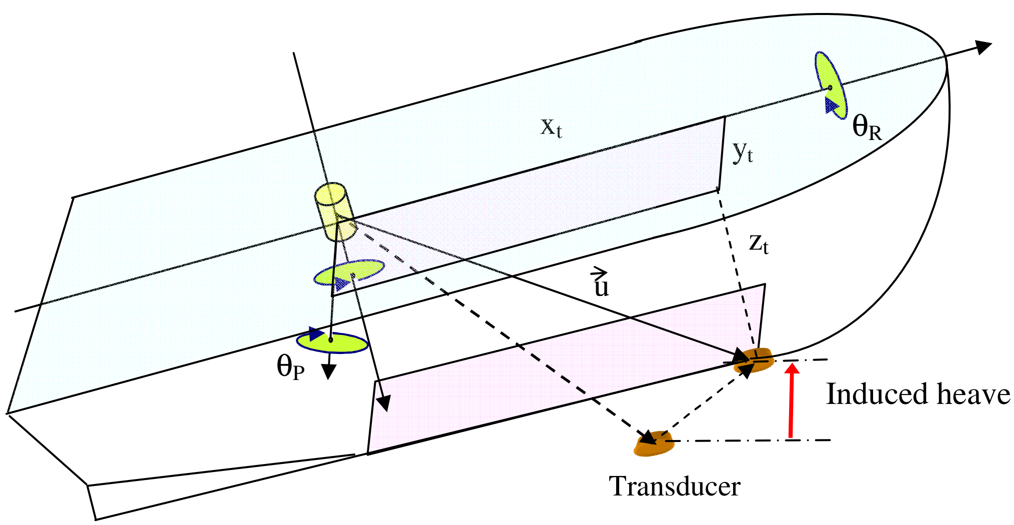


Fig. 3.27 "Induced heave"

Note that the induced heave variance is dependent upon the accuracy of the transducer offsets from the motion sensor and from the accuracy of the roll and pitch angles [Hare, 1995].
The total heave variance corresponds to the depth variance, σ_h^2, i.e.

$$\sigma_h^2 = \sigma_{hm}^2 + \sigma_{hi}^2 \qquad \textbf{(3. 50)}$$

where σ_{hm}^2 is the heave measurement variance and σ_{hi}^2 is the induced heave variance. This later error is typically negligible when compared to the heave measurement error.

When no heave compensator is available, it is possible manually to smooth the data in the echo trace. This task requires considerable experience in the interpretation of the echo trace in order to preserve seafloor features. The general procedure, in non significant roll conditions, is that the echo trace should be smoothed half way between crests and troughs (Figure 3.28).

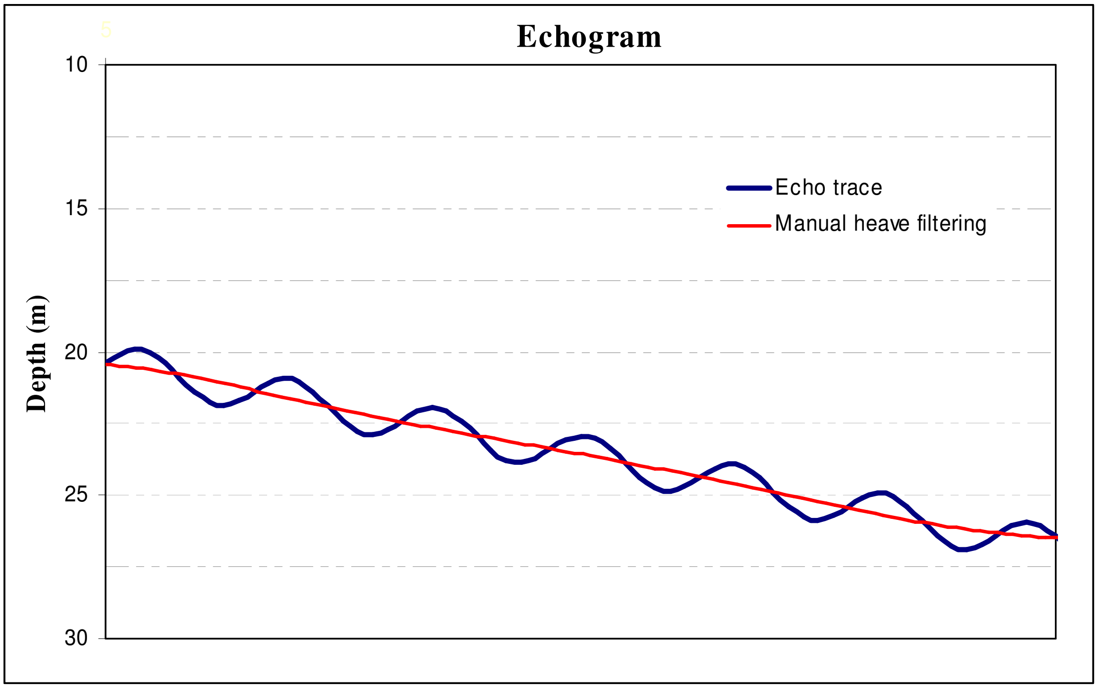


Fig. 3.28 "Manual heave filtering"

5.1.4.5 Due to draught, settlement, squat and relative position of transducer. The accurate measurement of the transducer draught is fundamental to the accuracy of the total depth. Even so it is generally necessary to update that value during the survey. The reasons for the draught variation are mainly due to the fuel and water consumption. The draught variation, for the same displacement, increases with the decrease of the float area at the sea surface. The draught error propagates directly as a depth error, $\mathbf{dz_{draught}}$.

Settlement is the general lowering in level of a moving vessel, relative to its motionless state level. This effect, particularly noticeable in shallow waters, is due to the regional depression of the surface of the water in which the vessel moves. The depth error due to settlement is referred to $\mathbf{dz_{settlement}}$.

Squat is another effect that takes place under dynamic conditions, the change in level of the bow and stern from the rest condition in response to the elevation and depression of the water level about the hull resulting from the bow and stern wave systems. In surveying vessels where the change in squat is significant, it is usually computed in a table of squat versus vessel speed. The depth error due to squat is referred to $\mathbf{dz_{squat}}$.

The relative position of the transducer from the motion sensor or heave compensator needs to be taken into consideration to correct the measured depth for induced heave, see 5.1.4.4.

The total error due to transducer position to the water line, $\mathbf{dz_i}$ is:

$$dz_i = \sqrt{dz^2_{draught} + dz^2_{settlement} + dz^2_{squat}} \qquad \textbf{(3. 51)}$$

Note that the total depth variance due to transducer immersion is written as:

$$\sigma_i^2 = \sigma^2_{draught} + \sigma^2_{settlement} + \sigma^2_{squat} \qquad \textbf{(3. 52)}$$

where $\sigma^2_{draught}$ is the draught variance, $\sigma^2_{settlement}$ is the settlement variance and σ^2_{squat} is the squat variance.

5.1.4.6 Record reading and resolution. The record reading and resolution for depth measurement is dependent on the operating principles of the echo sounder. In the case of analogue recording, the operator should select appropriate echo sounder parameters during survey operations in order to have, as far as possible, a clean echo trace and adequate resolution. On the other hand, the digital record no longer has such a degree of dependence on the operator during the survey but supervision is required during data acquisition.

When the data is recorded on paper, it is necessary to select the gain and intensity for a legible record; it is also necessary to have a vertical scale with sufficient discrimination. It is common to use scale phases for that reason (see 5.1.1.1).

The echo trace should be prepared for reading; this task consists of identifying the points on the seabed that will be selected for depth reading. This is usually performed with the aid of a digitising table.

The error associated with record reading depends on the experience and care of the hydrographer. Consider a paper record width of 20 centimetres and scale 0-200 metres, an error reading of 0.5 mm will produce an error in the depth of 0.5 metres. Therefore, this scale is not adequate for depth recording in shallow waters. The reading error is referred to $\mathbf{dz_{read}}$, with variance σ_r^2.

5.1.4.7 Interpretation of the echo trace is a hydrographer's responsibility. The interpretation requires experience to identify particular shapes, multiple echoes and false echoes.

a) False echoes – are caused by foreign matter such as kelp or fish in the water column (Figure 3.29), by layers of water separated by sudden changes of temperature or salinity or both.

False echoes are occasionally recorded by echo sounders and might be interpreted erroneously as correct depths. In cases of doubt over the validity of measured depths, an investigation of that sounding should be performed and the particular part of the survey line re-run if necessary.

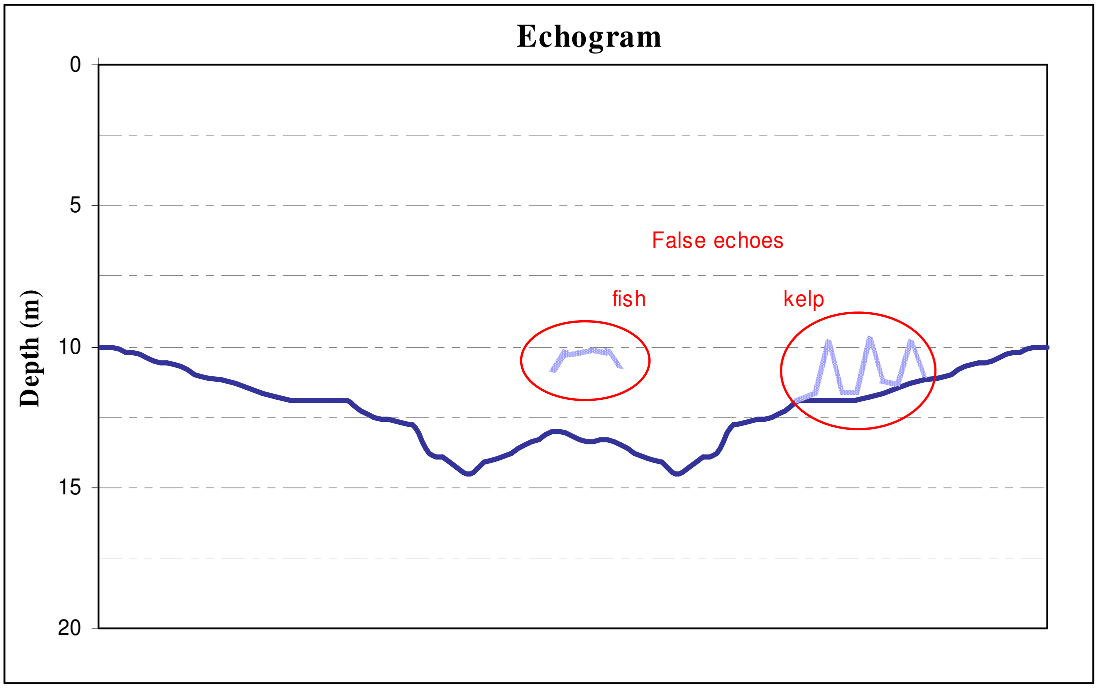


Fig. 3.29 "False echoes"

b) Multiple echoes – are echoes received subsequent to the very first one due to a multiplicity of reflections back and forth between the seafloor and the surface. These reflections are often recorded as multiples of the first depth (Figure 3.30).

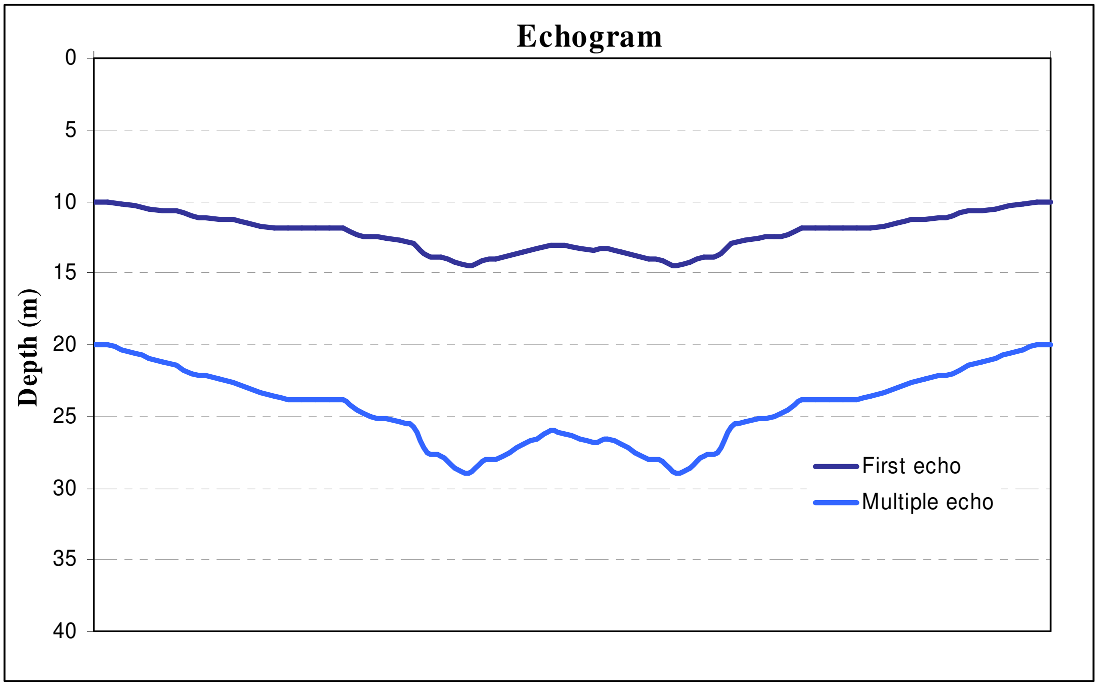


Fig. 3.30 "Multiple echoes record"

c) **Heave** - the oscillatory rise and fall of a vessel due to the entire hull being lifted by the force of the sea, may be compensated during acquisition by a heave sensor or may be filtered manually afterwards. The hydrographer's experience is the tool used for that purpose, even though; it is sometimes difficult to differentiate the heave in an irregular seabed.

d) **Side echoes** – are also false echoes but they are the result of echo detection in the side lobes which results in errors in depth measurement and positioning (see 4.2).

e) **Unconsolidated sediments** – are usually detected by high frequency echo sounders. In shallow waters, it is recommended that two frequencies are used at the same time to differentiate soft sediments from the bed rock (Figure 3.31).

5.1.4.8 Depth reduction. The measured depths, corrected for the attitude of the surveying vessel, are reduced to the vertical datum by the application of the tide. The depth error due to tide error measurement is referred to $\mathbf{dz_{tide}}$.

In addition to the error in tide measurement, sometimes a more significant error is the co-tidal correction which results from the difference of the actual tide in the survey area and in the tide gauge. The depth error from co-tidal error is $\mathbf{dz_{co\text{-}tidal}}$. This may be quite significant several miles away from the tide gauge (see Chapter 4). A co-tidal model or weighted averages from two or more tide gauges may be required.

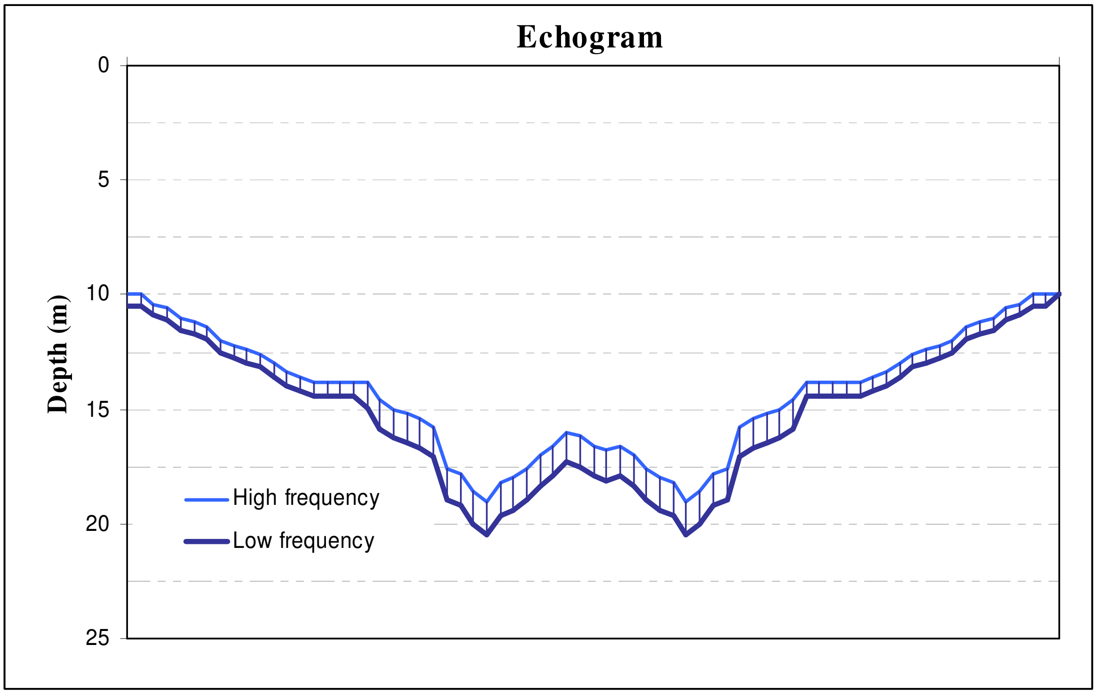


Fig. 3.31 "Dual channel recording"

Note that the tide variance, σ^2_{tide} , due to measurement error and co-tidal variation is written as,

$$\sigma^2_{tide} = \left(\sigma_{tide\,m}^{\;2} + \sigma^2_{co-tidal}\right) \tag{3. 53}$$

where $\sigma_{tide\,m}^{\;2}$ is the tide measurement variance and $\sigma^2_{co-tidal}$ is the co-tidal component variance.

The tide determination using GPS-RTK (Real Time Kinematic) will provide accurate local tide determination, however, the tidal computation requires a model of the differences between the reference ellipsoid, the WGS84, and the vertical datum.

The quality control is performed by statistical calculations based on the comparison of soundings from the check lines against neighbouring soundings from the survey lines. The statistical results of the comparison should comply with the accuracy recommendations in the S-44 (Figure 3.32).

According to the errors presented above, the estimated reduced depth variance is written as follows,

$$\sigma^2_z = \sigma_{z\,c}^{\;2} + \sigma_{z\,t}^{\;2} + \sigma^2_h + \sigma^2_i + \sigma^2_r + \sigma^2_{tide}. \tag{3. 54}$$

The estimated error on the reduced depth, at the 68 percent (or 1σ) confidence level, is obtained by the square-root of equation 3.54. Assuming that the component errors follow approximately a normal distribution, the estimated error on the reduced depth, at the 95 percent (or 2σ) confidence level, is obtained by substituting each variance σ^2 with $(2\sigma)^2$.

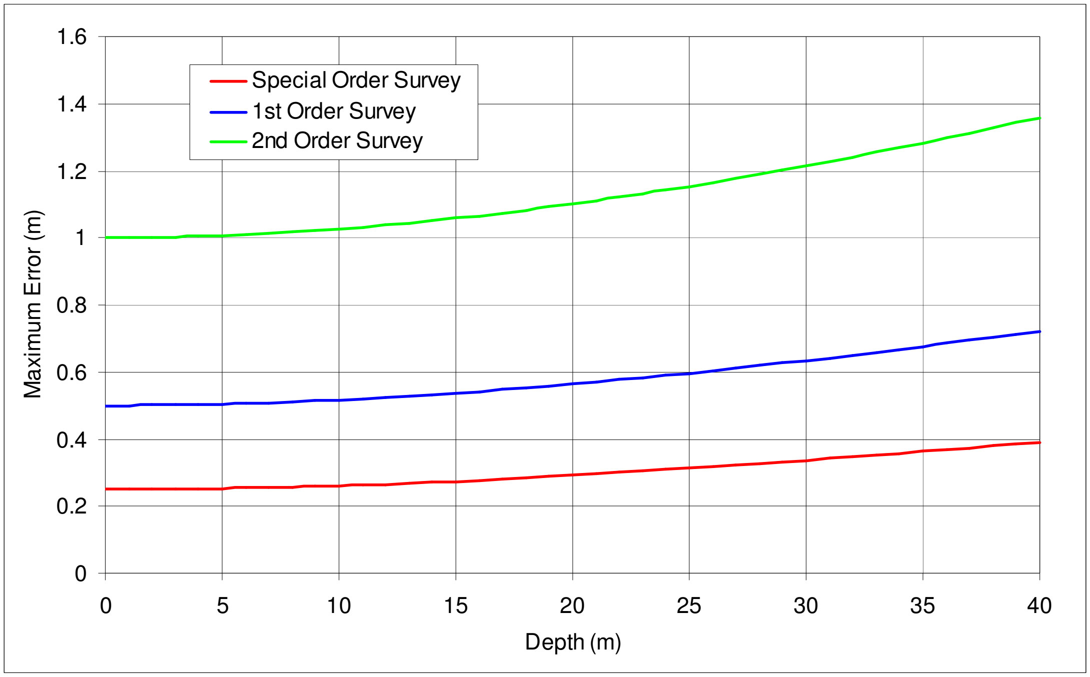


Fig. 3.32 "Minimum depth accuracy requirements for Special Order, Order 1 and Order 2 (S-44)"

For each surveying system it is recommended an error budget is created to assess compliance with the S-44 requirements. Figure 3.33 represents the error estimation for a particular echo sounder and operating conditions.

Echo sounder Characteristics:
Frequency = 200 kHz
Beam width = 20º (β = 10º)
Pulse duration = 0.1 ms

Operating Conditions:
Mean sound speed = 1500 m/s
Heave = 1 m
Roll = 5º
Vessel speed = 8 knots
Settlement = N/A
Squat = 0.05 m

Estimated or standard errors (1 σ):

Draft error = 0.05 m
Sound speed variation = 5 m/s
<u>Depth errors due to</u>:
sound speed variation = (5/1500)z m
time measurement = 0.02 m
squat = 0.05 m
heave = 0.10 m
tide error = 0.05 m
co-tidal error = 0.05 m

Estimated errors (2 σ):
(95% confidence level)

Draft error = 0.10 m
Sound speed variation = 10 m/s
<u>Depth errors due to</u>:
sound speed variation = (1/150)z m
time measurement = 0.04 m
squat = 0.10 m
heave = 0.10 m
tide error = 0.10 m
co-tidal error = 0.10 m

$$\sigma_z = \sqrt{\sigma_{zc}^2 + \sigma_{zt}^2 + \sigma_i^2 + \sigma_h^2 + \sigma_r^2 + \sigma_{tide}^2}\,.$$

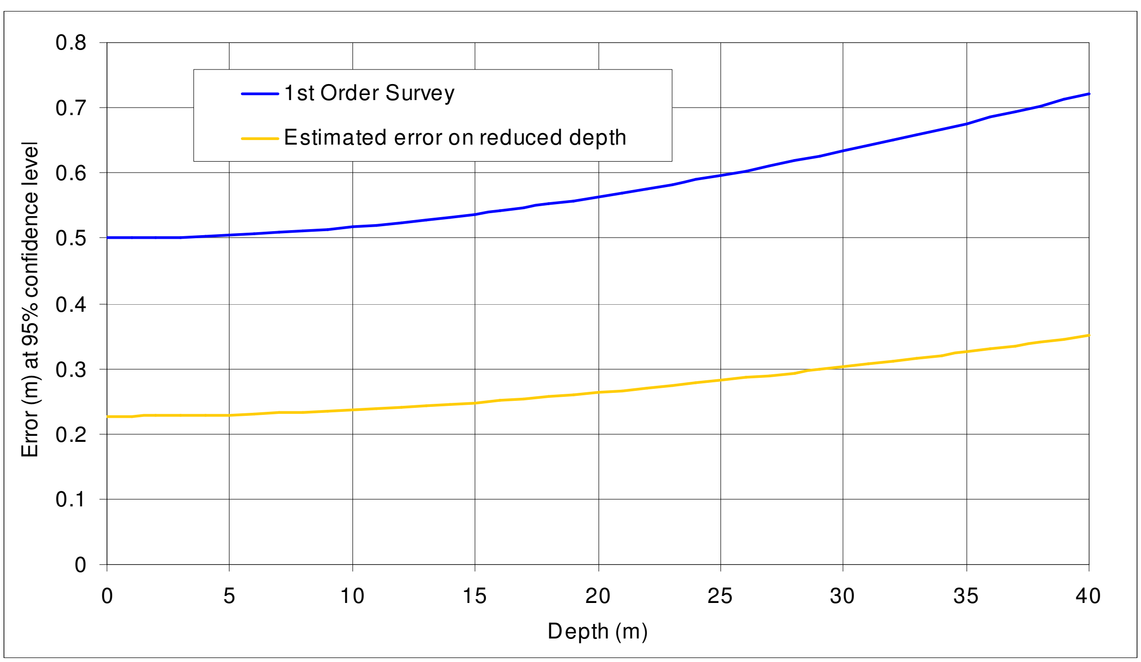


Fig. 3.33 "Depth error estimation"

5.2 Swath Systems

Swath systems measure the depth in a strip of seafloor extending outwards from the sonar transducer. These systems are arranged in a way that the profile, line where measurements are performed, lies at right angles to the direction of survey vessel motion. As the vessel moves, the profiles sweep out a band on the seafloor surface, known as a swath.

This section describes the multibeam systems and interferometric sonars.

5.2.1 Multibeam systems

MBES are a valuable tool for depth determination when full seafloor coverage is required. These systems may allow complete seafloor ensonification with the consequent increase in resolution and detection capability.

5.2.1.1 Principles of operation. Multibeam principle of operation is, in general, based on a fan shaped transmission pulse directed towards the seafloor and, after the reflection of the acoustic energy by the seabed; several beams are electronically formed, using signal processing techniques, with known beam angles. The two-way travel time between transmission and reception is computed by seabed detection algorithms. With the application of ray tracing (see 5.2.1.8.1), it is possible to determine the depth and the transversal distance to the centre of the ensonified area.

The transmitted beam is wide across-track and narrow along-track; conversely, the beams formed during reception are narrow across-track and wide along-track. The intersections of those beams in the seafloor plan are the footprints (ensonified areas) for which the depths are measured.

Since depths are measured from a floating platform, with six possible degrees of freedom (three translations and three rotations), for accurate computation of depth measurement and its associated positioning, accurate measurements of latitude, longitude, heave, roll, pitch and heading are required.

A. **Bottom detection** is the process used in MBES to determine the time of arrival and the amplitude of the acoustic signal, representing a reflection from the seabed. The reliability of this process affects the quality of the measurements. Depth blunders can be, among other factors, related to a poor performance of the algorithms used for seabed detection. Seabed detection algorithms can be categorized into two main divisions: amplitude detection and phase detection algorithms.

a) **Amplitude detection**. The transducer array emits an acoustic pulse towards the seabed and then starts the listening period. In this phase, the returned signal is sampled in time for each beam angle. The travel time of the signal for the correspondent depth point is defined by the detected amplitude of the reflected signal (Figure 3.34).

The most common methods of amplitude detection are as follows:

i) Leading Edge of the Reflected Signal. This method is commonly used when the angle of incidence of the acoustic signal to the seafloor is approximately zero degrees. The bottom detection time is defined for the first arrival inside the beam angle.

With the increasing angle of incidence, the returned signal loses its sharp form (short rise time) and the leading edge method no longer performs well. Two other methods can be employed, which take into account the variation of the reflected signal strength samples across the beam footprint.

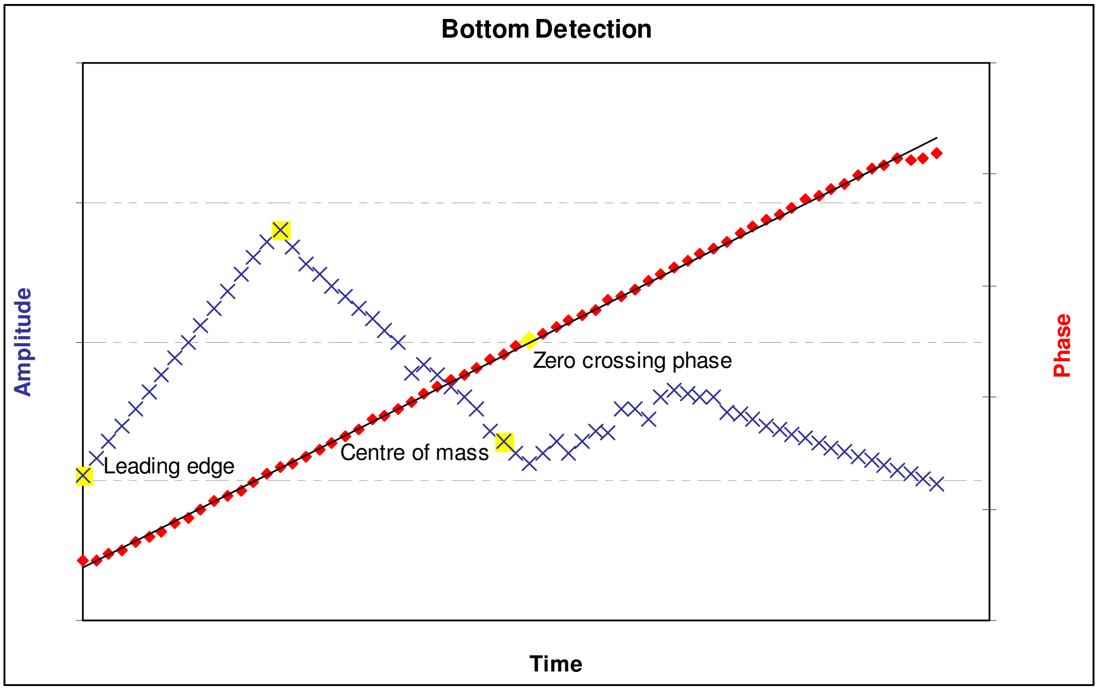


Fig. 3.34 "Bottom detection methods (signal detection)"

ii) <u>Maximum Amplitude of the Reflected Signal</u>. The bottom detection is defined by the time of maximum backscatter amplitude.

iii) <u>Centre of Mass of the Reflected Signal</u>. This method consists of determining the time corresponding to the centre of gravity of the amplitude signal.

b) Phase detection. Amplitude detection is the technique used for the inner beams (close to nadir), where the backscatter amplitude has higher values and a smaller number of samples. For the outer beams, the backscatter amplitude decreases and the number of samples becomes very large. Consequently, the echo is so smeared out in time that amplitude detection methods obtain poor results. Against a sloping seabed in the across-track direction away from the source, the smear of the echo is also enhanced. Hence, the phase detection method is commonly used for large angles of incident.

In this technique, the transducer array for each beam is divided into two sub-arrays, often overlapping, with the centres of the sub-arrays a number of wavelengths apart. The angular directions are pre-determined and each sub-array forms a beam in that direction, the advantage being that in the case of the arrival of simultaneous echoes from different directions, the MBES system resolves only the echoes in the direction of the formed beam. The sequence of phase difference estimates are then used to estimate the time of arrival of the echo in the pre-determined direction by finding the zero crossing of the phase sequence [de Moustier, 1993]. Figure 3.35 depicts an example of the phase detection method. The equivalent of the centre of the two sub-arrays are represented by ***A*** and ***B*** at a distance ℓ from each other, where $\boldsymbol{\theta}$ is the angle of the received signal measured from the acoustical axis. A second order polynomial may be fitted to a limited sequence of differential phase estimates to refine the detection of the zero crossing phase.

When the signal return is from the acoustic axis direction, i.e. $\theta = 0$, the signals at the two sub-arrays are in phase, and this corresponds to the acoustic travel time.

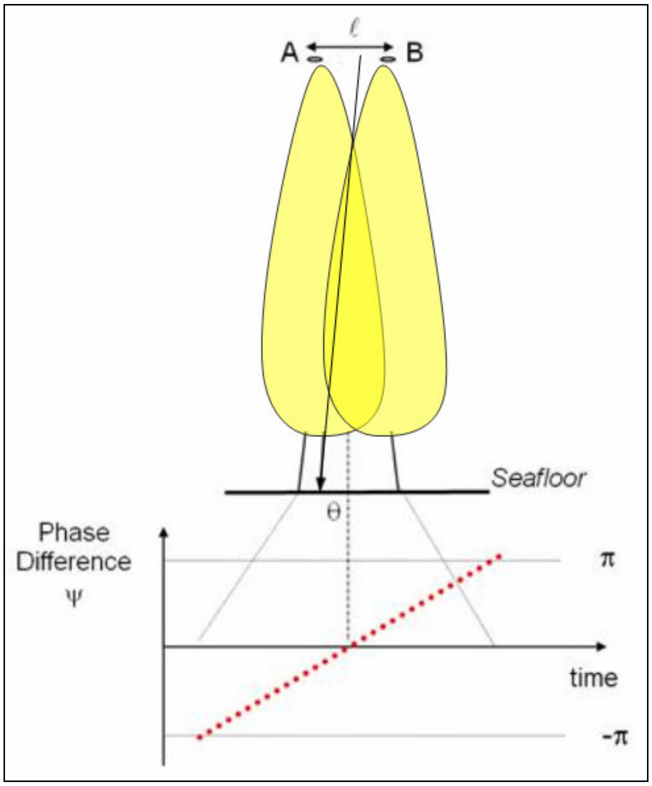


Fig. 3.35 "Bottom detection methods by phase difference (zero crossing phase)"

Across the swath, a combination of amplitude and phase detection is usually required for robust bottom detection. Near nadir, the amplitude detection should be used due to the fact that the time series for these beams is too short for a robust phase detection. Amplitude detection is also used in the case of steep slopes occurring well off-nadir, associated with bathymetric heights, except for the extreme case of a seafloor that is steeply sloping away from the transducer. Phase detection for the nadir beams is more likely to be the result of a gross error (blunder) due to mid-water column returns or due to higher returns from the side lobes. Off-nadir detections are better by phase, but amplitude detection may be chosen when a higher return is caused by the difference of the reflective properties of the target, by a near specula reflection or by a large variance of the curve fit. These conditions may occur due to features such as wrecks and boulders.

B) Fast Fourier Transform (FFT)

According to 4.1, during beam forming the signals from each element of the array are copied for each beam. The sum of the amplitude of the **N** transducer elements is itself a Fourier transform of the vector, with **N** elements, corresponding to the radiation pattern of the linear array. If **N** is a power of 2 the computation is less demanding and the Fourier transform is called Fast Fourier Transform[16]. This method has the advantage to speed up the beam forming process.

Details on this method are addressed in de Moustier [1993].

5.2.1.1 Accuracy. Measurement of range and beam angle for multibeam systems is more complex than for single beam echo sounders. Consequently, there are a number of factors that contribute to the error in those measurements, including: beam angle, incident angle on the seafloor, transmit and receive beam widths, attitude and heave accuracy, bottom detection algorithms and sound speed profile variation.

It is usually necessary to calculate the error budget based on those factors (see 5.2.1.8.).

5.2.1.2 Resolution. Multibeam systems with their capability of full seafloor ensonification contribute to a better seafloor representation and, when compared to SBES, to higher mapping resolution. However, as far as the depth measurements are concerned, resolution will depend on the acoustic frequency, transmit and receive beam widths and on the algorithm used to perform seabed detection.

Resolution in depth measurement is a function of pulse length and of the dimensions of the ensonified area. The ensonified area of MBES echo sounders, near normal incidence, is relatively small, thus the resolution is higher than the single beam echo sounder.

5.2.1.3 Frequency of a MBES is selected based on the intended use of the equipment, basically depth ranges and resolution.

The frequencies of the MBES are typically:

a.	Waters shallower than 100 metres:	frequencies higher than 200 kHz;
b.	Waters shallower than 1500 metres:	frequencies from 50 to 200 kHz;
c.	Waters deeper than 1500 metres:	frequencies from 12 to 50 kHz.

5.2.1.4 System associated sensors and integrity

Multibeam systems, besides the echo sounder itself, include:

a) motion sensor – for attitude (roll, pitch and heading) and heave measurement. Presently, these sensors comprise an inertial measurement unit (IMU) and a pair of GPS receivers with respective antennas. As a result of the technology involved this sensor is also capable of providing positioning with high accuracy;
b) sound velocity profiler – to measure the sound velocity through the water column;
c) sound velocity probe – to measure the sound velocity at the transducer face. This should be considered compulsory for flat arrays where beam steering is required;

---

[16] It is still possible to apply the FFT even when N it is not a power of 2 by adding virtual elements with zero amplitude (padded zeros), see de Moustier [1993].

d) Positioning system – as stated above, positioning in new technology systems is integrated with motion sensors. GPS either in pseudo-differential mode or in Real Time Kinematic mode (RTK) is the system commonly used worldwide;
e) Heading sensor – also integrated in motion sensors, the optimal and more accurate solution is the heading obtained by GPS dual receivers.

5.2.1.5 Installation and calibration (patch test)

The installation of multibeam transducers can be either fixed to the hull, on the side or over the bow. The hull installation is used in large vessels or when it is a permanent fixture, the other installations are used for temporary short term purposes in small vessels.

The calibration or patch test is an essential procedure which consists of determining the composite offset angles (roll, pitch and azimuth) for both transducer and motion sensor and the latency from the positioning system. Detailed analysis and procedures may be found in Godin [1996].

The calibration must be performed after the installation and either after long periods of inoperability or after appreciable changes in the installation.

Before calibration it is necessary to check the installation parameters and to perform a sound velocity profile to update the calculation of the refraction solution.

a) Positioning time delay

The time delay or latency is the time offset between depth measurement and positioning. The procedure to determine the time offset consists of running two pairs of survey lines at different speeds over a sloping seabed, the steeper the gradient the higher the resolution of this parameter will be. The slope should be regular and with enough extension to guarantee adequate sampling. Figure 3.36 illustrates the time delay calibration.

The time delay is obtained by measuring the longitudinal displacement of the soundings along the slope due to the different speeds of the vessel. To avoid any influence from the pitch offset the lines should be run on the same course.

The time delay, $\boldsymbol{\delta t}$, may be obtained by the equation:

$$\delta t = \frac{\Delta x}{v_2 - v_1} \qquad \textbf{(3. 55)}$$

where $\boldsymbol{\Delta x}$ is the horizontal separation between the two sounding profiles near nadir, and $\mathbf{v_1}$ and $\mathbf{v_2}$ are the speeds of the vessel for line 1 and 2 respectively.

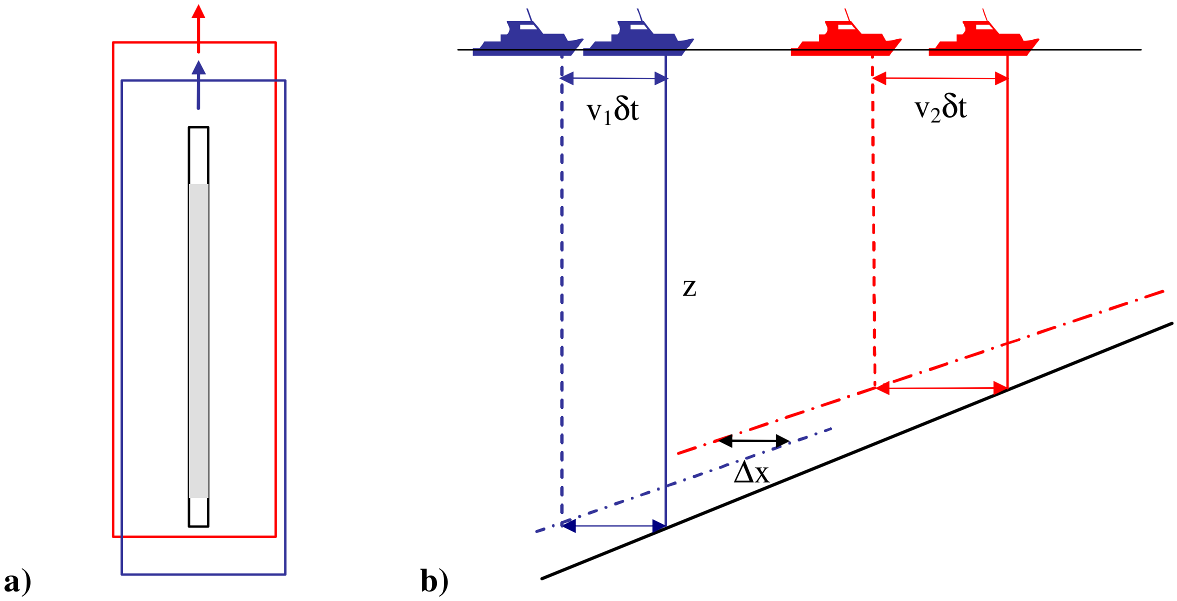


Fig. 3.36 "Time delay calibration.
a) Top view of the two lines;
b) Longitudinal section where the separation of the two sounding profiles from the actual seafloor is visible. "

b) Pitch offset

The pitch offset is the composite angle offset from the inertial measurement unit and from the transducer alignment with the local vertical in the longitudinal plane of the vessel. The procedure to determine the pitch offset consists of running two pairs of reciprocal survey lines at the same speed over a sloping seafloor, as mentioned in the time delay calibration, the steeper the gradient the higher the resolution of this parameter will be. The slope should be regular and with enough extension to guarantee adequate sampling. Figure 3.37 depicts the pitch offset calibration.

After the correct determination of time delay, the pitch offset is obtained by measuring the longitudinal displacement of the soundings along the slope due to the pitch offset. To avoid any influence from time delay the system should already be compensated for positioning latency.

The pitch offset, $\boldsymbol{\delta\theta_p}$, may be obtained by the equation:

$$\delta\theta_p = \tan^{-1}\left(\frac{\Delta x}{2 \cdot z}\right) \qquad \textbf{(3. 56)}$$

where $\boldsymbol{\Delta x}$ is the horizontal separation between the two sounding profiles near nadir, and $\mathbf{v_1}$ and $\mathbf{v_2}$ are the speeds of the vessel for line 1 and 2 respectively.

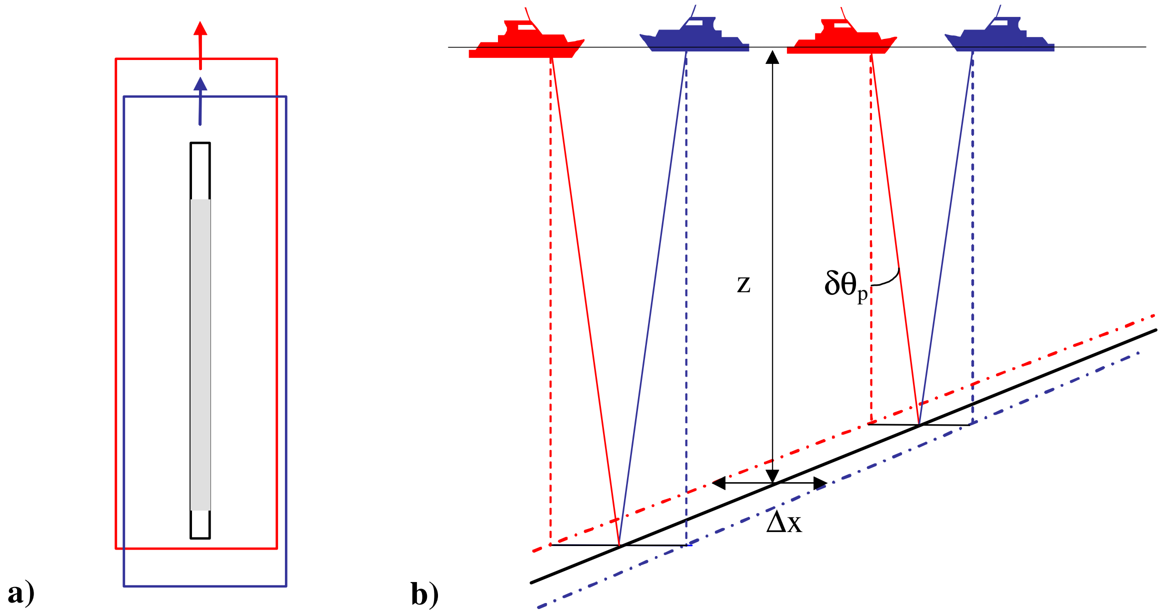


Fig. 3.37 "Pitch offset calibration.
a) Top view of the two lines;
b) Longitudinal section where the separation of the two sounding profiles from the actual seafloor due to pitch offset are visible."

c) <u>Azimuthal offset</u>

The azimuthal offset is the composite angle offset from the heading sensor and from the transducer alignment perpendicularly to the longitudinal axis of the vessel. The procedure to determine the azimuthal offset consists of running two adjacent pairs of reciprocal lines, at the same speed, in an area with a well defined bathymetric feature such as a shoal. The adjacent lines should overlap (not more than 20% the swath width) in the outer beams in the location of the bathymetric feature. Figure 3.38 depicts the azimuthal offset calibration.

After the correct determination of time delay and pitch offset, the azimuthal offset is obtained by measuring the longitudinal displacement of the bathymetric feature from adjacent lines. To avoid any influence from time delay and pitch offset the system should already be compensated for these offsets.

The azimuthal offset, $\delta\alpha$, may be obtained by the equation:

$$\delta\alpha = \tan^{-1}\left(\frac{\Delta x}{\Delta L}\right) \qquad \textbf{(3. 57)}$$

where Δx is the horizontal separation from the bathymetric feature from the adjacent reciprocal lines and ΔL is the distance between lines.

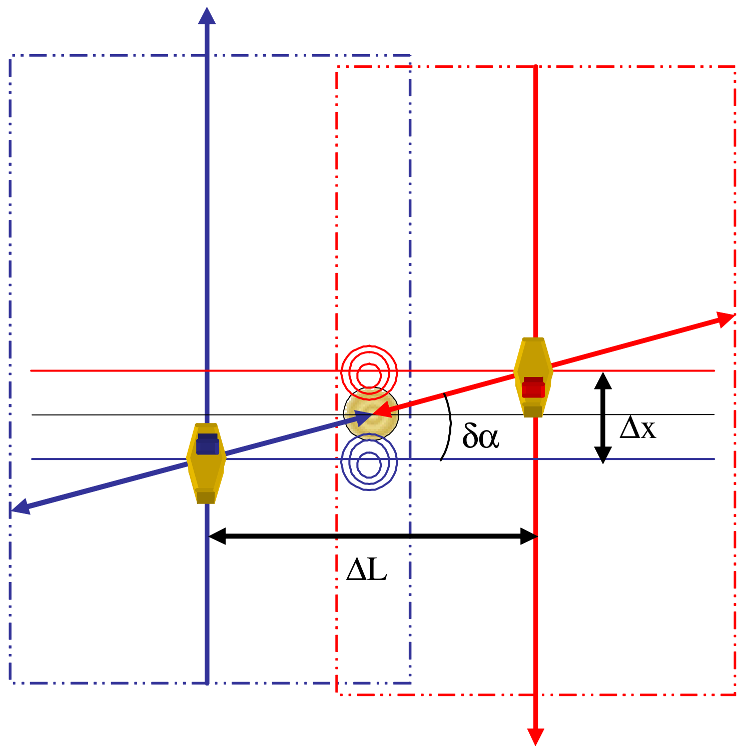


Fig. 3.38 "Azimuthal offset calibration"

d) <u>Roll offset</u>

The roll offset is the composite angle offset from the inertial measurement unit and from the transducer alignment with the local vertical in the transversal plane of the vessel. The procedure to determine the roll offset consists of running a pair of reciprocal survey lines, at the same speed, in a regular and flat seafloor. The lines should overlap each other. Figure 3.39 depicts the roll offset calibration.

After the correct determination of time delay, pitch and azimuthal offsets, the roll offset is obtained by measuring the vertical displacement of the outer beams from the reciprocal lines. To avoid any influence from time delay, pitch and azimuthal offsets the system should already be compensated for these offsets.

The roll offset, $\boldsymbol{\delta\theta_R}$, may be obtained by the equation:

$$\delta\theta_R = \tan^{-1}\left(\frac{\Delta z}{2 \cdot \Delta y}\right) \qquad \textbf{(3. 58)}$$

where $\boldsymbol{\Delta z}$ is the vertical displacement between outer beams from reciprocal lines and $\boldsymbol{\Delta y}$ is half swath width or the distance from nadir to the point where vertical displacement is measured.

Calibration is usually performed by interactive tools. The adjustment or offset calculation should be done in several sampling sections in order to obtain an average. The offsets may have an uncertainty of the order of the motion sensor repeatability.

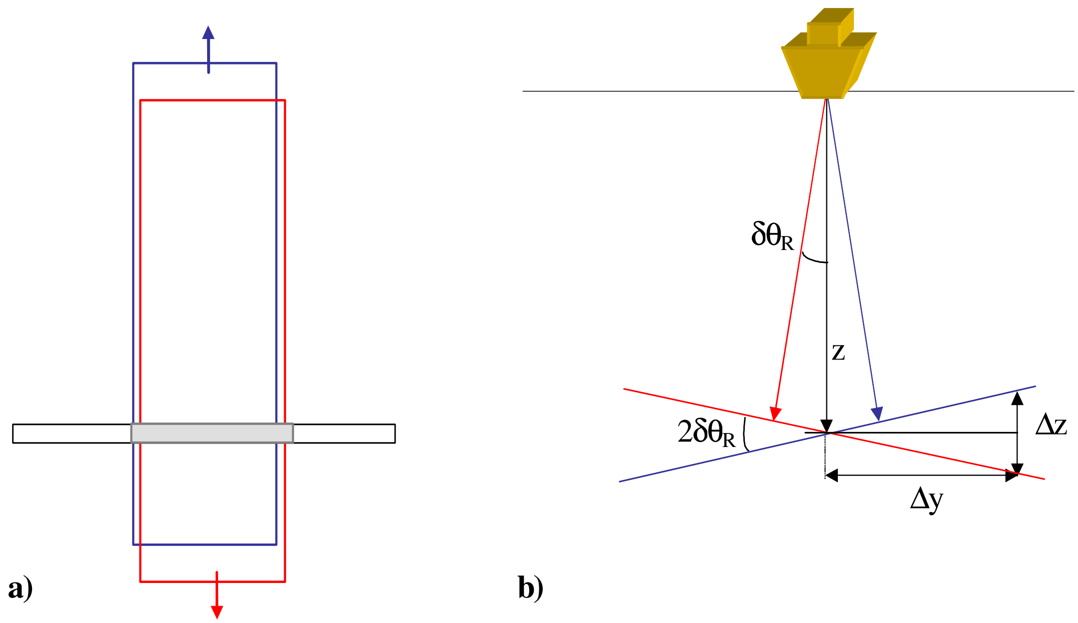


Fig. 3.39 "Roll offset calibration.
a) Top view of the two reciprocal lines;
b) Cross section where the pronounced separation of the sounding profiles from nadir to the outer beams due to roll offset are visible."

5.2.1.7 <u>Operation and data recording</u>

Vessel configuration and calibration parameters should be checked at the beginning of the survey. Some system parameters should also be checked. These are mainly the parameters used during data acquisition and may vary with each survey's location (e.g. maximum operating depth, expected depth, maximum ping rate, etc.).

At the beginning of the survey a sound velocity profile should be performed and transferred to the echo sounder to be used, generally, in real time. Sound velocity at the transducer face should be compared with the value given from the sound velocity probe. During the survey session several sound profile casts should be performed according to the pre-analysis of the temporal and spatial variation of sound velocity.

Whilst surveying, the systems are almost completely automatic, the hydrographer should, however, monitor the data acquisition and the data integrity. Full seafloor ensonification and overlap with the adjacent swaths must be guaranteed and monitored. It is most important to compare the overlapping outer beams of adjacent swaths and to check for any trend of curvature upward or downward of each ping.

At the end of every survey session it is strongly recommended that a back up of the collected data is created.

5.2.1.8 Sources of errors and quality control techniques

Sources of errors were discussed in 5.1.4 but these were for SBES. Hereafter the sources of errors are analyzed for MBES, some of the errors are common to both systems, i.e. they do not vary with beam angle. For this reason some of the errors are referred back to 5.1.4. Further details on multibeam uncertainty can be found in Hare [1995] and Lurton [2002].

5.2.1.8.1 Due to sound velocity. Errors in sound velocity or in its variation result in incorrect refraction solutions and, consequently, to errors in depth measurement and positioning.

The ray[17] tracing is based on Snell's law which states the relation between the ray direction and the acoustic wave velocity:

$$\frac{\operatorname{sen}\theta_0}{c_0} = \cdots = \frac{\operatorname{sen}\theta_i}{c_i} = \kappa \qquad \textbf{(3. 59)}$$

where $\mathbf{c_i}$ is the sound velocity, $\boldsymbol{\theta_i}$ is the incidence angle referred to the vertical at the depth $\mathbf{z_i}$, and $\boldsymbol{\kappa}$ is the ray parameter or Snell constant.

Assuming that the sound velocity profile is discrete (Figure 3.40), it is reasonable to assume that the sound velocity gradient in a layer, between two measurements, is constant. Though, the sound velocity is represented as follows:

$$c^i(z) = c_{i-1} + g_i(z - z_{i-1}), \quad \text{for} \quad z_{i-1} \le z \le z_i \qquad \textbf{(3. 60)}$$

where $\mathbf{g_i}$ is the constant gradient at layer **i**, given by:

$$g_i = \frac{c_i - c_{i-1}}{z_i - z_{i-1}} \qquad \textbf{(3. 61)}$$

In each layer the acoustic pulse travels a path with constant radius of curvature, $\boldsymbol{\rho_i}$, given by:

$$\rho_i = -\frac{1}{\kappa g_i} \qquad \textbf{(3. 62)}$$

[17] The acoustic ray corresponds to a line drawn from the source, in each point is perpendicular to the wave front.

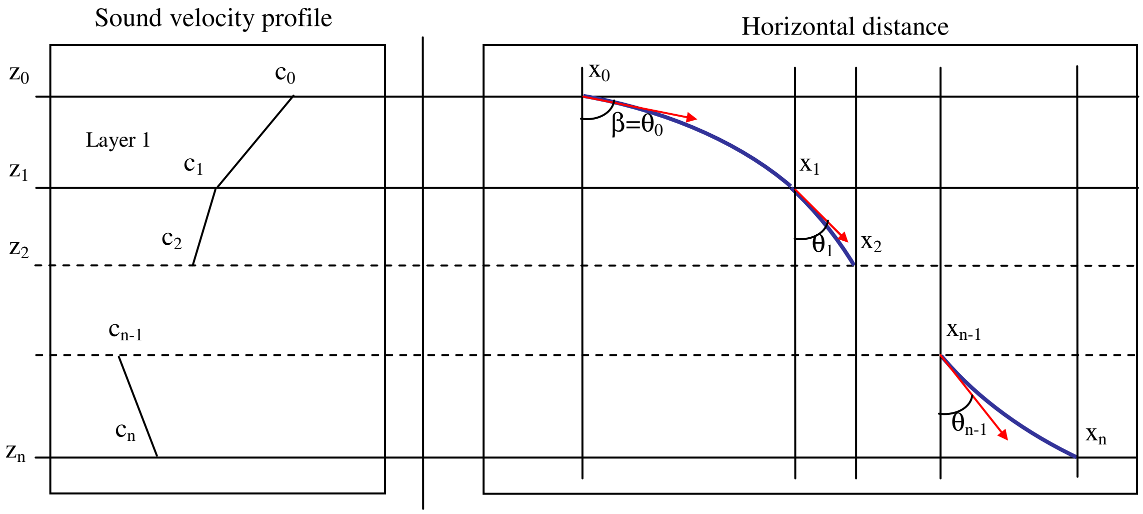


Fig. 3.40 "Ray tracing"

Considering the launching angle $\boldsymbol{\theta_0}$ (or $\boldsymbol{\beta}$) in a depth with sound velocity $\mathbf{c_0}$, the horizontal distance travelled by the acoustic pulse to cross the layer **i** is:

$$\Delta x_i = \rho_i(\cos\theta_i - \cos\theta_{i-1}) = \frac{\cos\theta_{i-1} - \cos\theta_i}{\kappa g_i} \qquad \textbf{(3. 63)}$$

The replacement of $\mathbf{cos(\boldsymbol{\theta_i})}$ by $\sqrt{1-(\kappa c_i)^2}$ produces:

$$\Delta x_i = \frac{\sqrt{1-(\kappa c_{i-1})^2} - \sqrt{1-(\kappa c_i)^2}}{\kappa g_i} \qquad \textbf{(3. 64)}$$

The travel time of the acoustic pulse in layer ***i***, is obtained by:

$$\Delta t_i = \frac{1}{g_i}\int_{c_{i-1}}^{c_i} \frac{dc}{c \cdot \cos(\theta)} \qquad \textbf{(3. 65)}$$

which can be written:

$$\Delta t_i = \frac{1}{g_i}\ln\left(\frac{c_i}{c_{i-1}}\frac{1+\sqrt{1-(\kappa c_{i-1})^2}}{1+\sqrt{1-(\kappa c_i)^2}}\right) \qquad \textbf{(3. 66)}$$

To obtain the total horizontal distance travelled by the acoustic signal and the travel time, it is necessary to sum the distances $\boldsymbol{\Delta x_i}$ and the times $\boldsymbol{\Delta t_i}$ from the transducer to the seabed,

$$x = \sum_{i=1}^{n} \frac{\left[1-(\kappa c_{i-1})^2\right]^{1/2} - \left[1-(\kappa c_i)^2\right]^{1/2}}{\kappa g_i} \quad \textbf{(3. 67)}$$

$$t = \sum_{i=1}^{n} \frac{1}{g_i} \ln\left(\frac{c_i}{c_{i-1}} \frac{1+\sqrt{1-(\kappa c_{i-1})^2}}{1+\sqrt{1-(\kappa c_i)^2}} \right) \quad \textbf{(3. 68)}$$

Depth determination and sounding positioning are the result of the integration of the echo along each direction, fixed by the beam pointing angle, using the updated sound velocity profile between the transmission and one-way travel time (**$\Delta t/2$**).

Taking one sound velocity profile with constant gradient, **g**, the depth is obtained as:

$$z = \int_0^{\Delta t/2} (c_0 + g \cdot z) \cdot \cos(\theta) dt \quad \textbf{(3. 69)}$$

The depth error, **dz_c**, due to the gradient variation, **dg**, and surface sound velocity variation, **dc_0**, by differentiation of equation 3.69 can be approximated by:

$$dz_c = \frac{z^2}{2c_0}\left(1-\tan^2(\beta)\right)dg + \frac{z}{c_0} dc_0 \quad \textbf{(3. 70)}$$

where **β** and **c_0** are respectively the beam pointing angle and the sound velocity from the sound velocity profile at the transducer face. In equation 3.70 the first term corresponds to both the range and ray bending errors due to the variation of the profile gradient, whereas the second term corresponds to depth error due to the sound velocity profile offset at the transducer depth. Assuming there is no correlation of these errors, the depth variance due to sound velocity errors is written as:

$$\sigma^2_{z\,c\_profile} = \left(\frac{z^2}{2c_0}\right)^2 \left(1-\tan^2(\beta)\right)^2 \sigma_g^2 + \left(\frac{z}{c_0}\right)^2 \sigma_{c_0}^2 \quad \textbf{(3. 71)}$$

where **σ_g** corresponds to the variance of the gradient of the sound velocity profile and **σc_0** corresponds to the variance of the initial value of the sound velocity profile used for depth calculation.

Sound velocity errors are, in practice, difficult to quantify and the problems with temporal and spatial variations may sometimes be so significant that the only practical solution is to limit multibeam angular coverage.

There is another error component due to sound velocity error or variation at the transducer face; this component introduces an error on the beam pointing angle which also introduces errors in depth measurement and positioning.

For beam steering or stabilization it is necessary to introduce time delays in the transducer elements (4.2). To compute the delays it is necessary to know the sound velocity at the transducer draught, this is usually achieved by a sound velocity sensor installed close to the transducer. Any error in the sound velocity at the transducer face will be propagated as an error in the beam pointing angle (Figure 3.41).

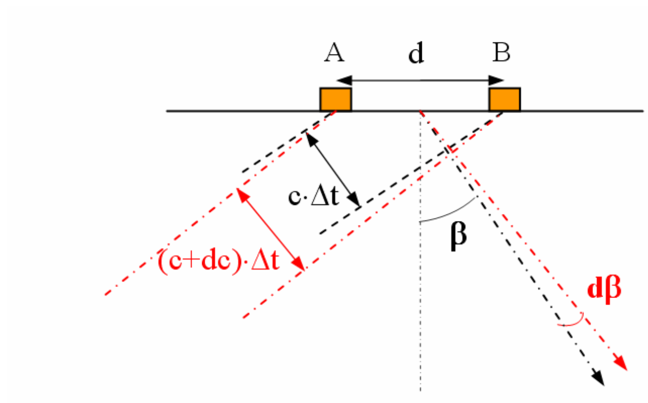


Fig. 3.41 "Beam steering error due to sound velocity variation"

The time delay to steer the beam the angle **β**, is obtained by:

$$\Delta t = \frac{d}{c_0}\sin(\beta) \qquad \textbf{(3. 72)}$$

hence:

$$\beta = \operatorname{asin}\left(\frac{c_0 \cdot \Delta t}{d}\right) \qquad \textbf{(3. 73)}$$

by differentiation and the appropriate simplification gives:

$$d\beta = \frac{\tan(\beta)}{c_0} dc_0 \qquad \textbf{(3. 74)}$$

The error in beam steering propagates to an error in depth, given by:

$$dz_\beta = -\frac{z}{c_0} \cdot \tan^2(\beta) dc_0 \qquad \textbf{(3. 75)}$$

The depth variance due to beam steering is therefore:

$$\sigma_{z\beta}^2 = \left(\frac{z}{c_0}\right)^2 \tan^4(\beta)\sigma_{c0}^2 \qquad \textbf{(3. 76)}$$

where $\boldsymbol{\sigma c_0}$ corresponds to the variance of the sound velocity used for beam steering, usually obtained by a sound velocity sensor.

Note that the estimated total depth variance due to sound velocity errors is written as:

$$\sigma_{z\,c}^{2} = \sigma_{z\,c\_profile}^{2} + \sigma_{z\beta}^{2} \qquad \textbf{(3. 77)}$$

The errors mentioned above may be detected by visual inspection of the data by trying to detect abnormal curvature of the profiles (set of beams).

5.2.1.8.2 Due to motion sensing. The depth measurement is dependent on pitch error and roll error, the contribution being given as follows:

$$dz_{\theta_R} = R \cdot \cos(\theta_P) \cdot \sin(\beta - \theta_R)d\theta_R \qquad \textbf{(3. 78)}$$

and:

$$dz_{\theta_P} = R \cdot \sin(\theta_P) \cdot \cos(\beta - \theta_R)d\theta_P \qquad \textbf{(3. 79)}$$

The depth variances are respectively:

$$\sigma_{z\theta_R}^{2} = (z \cdot \cos(\theta_P) \cdot \tan(\beta - \theta_R))^2 \sigma_{\theta R}^{2} \qquad \textbf{(3. 80)}$$

and:

$$\sigma_{z\theta_P}^{2} = (z \cdot \sin(\theta_P))^2 \sigma_{\theta P}^{2} \qquad \textbf{(3. 81)}$$

The total depth variance due to vessel attitude and heave is:

$$\sigma_{z\,motion}^{2} = \sigma_{z\theta_R}^{2} + \sigma_{z\theta_P}^{2} + \sigma_h^2 \qquad \textbf{(3. 82)}$$

where σ_h^2 is the heave variance.

5.2.1.8.3 Due to draught, settlement, squat and relative position of transducer. The accurate measurement of transducer draught and knowledge of vessel behaviour in dynamic conditions, settlement and squat, are fundamental to the accuracy of the measured depths. These errors will contribute to the depth error independent of the beam angle.

The total depth variance due to transducer immersion, see 5.1.4.4, is written as:

$$\sigma_i^2 = \sigma_{draught}^2 + \sigma_{settlement}^2 + \sigma_{squat}^2 \qquad \textbf{(3. 83)}$$

where $\sigma_{draught}^2$ is the draught variance, $\sigma_{settlement}^2$ is the settlement variance, and σ_{squat}^2 is the squat variance.

5.2.1.8.4 Depth reduction. This issue was previously analysed in 5.1.4.8.

The QC may be performed by statistical calculations based on the comparison of soundings from the check lines against the bathymetric surface generated from the main survey lines. The statistics generated by the comparison should comply with the accuracy recommendations in the S-44.

According to the errors presented above, the estimated reduced depth variance is written as follows:

$$\sigma_z^2 = \sigma_{z\,c}^{\;2} + \sigma_{h\,\text{motion}}^2 + \sigma_i^2 + \sigma_{\text{tide}}^2 + \sigma_{z\,\text{det ection}}^2 \qquad \textbf{(3. 84)}$$

where $\sigma_{z\,\text{det ection}}$ corresponds to the depth variance due to the seabed detection algorithm implemented in the MBES system [Lurton, 2002].

The estimated error on the reduced depth, at the 68 percent (or 1σ) confidence level, is obtained by the square-root of equation 3.84. Assuming that the component errors follow approximately a normal distribution, the estimated error on the reduced depth, at the 95 percent (or 2σ) confidence level, is obtained by substituting each variance σ^2 by $(2\sigma)^2$.

5.2.2 Interferometric sonars

5.2.2.1 Interferometric sonar systems apply the phase content of the sonar signal to measure the angle of a wave front returned from the seafloor or from a target. This principle differs from MBES which forms a set of receive beams and performs seabed detection for each beam, either by amplitude or phase, to detect the returning signal across the swath [Hughes Clarke, 2000].

These sonars have two or more horizontal arrays, each array produces a beam that is narrow along track and wide across track. One of these arrays is used for transmission, ensonifying a patch of the seafloor, scattering acoustic energy in all directions. Part of the scattered energy will return back towards the transducers, which measure the angle made with the transducers. The range is also calculated from the observed two-way travel time.

The method used for angle measurement can be diverse. The simplest method is to add the signal copied from the two arrays together, being the resultant amplitude "fringes", corresponding to variations in signal strength. If the transducer arrays are separated by half a wave length, there will only be one fringe, being the zero phase direction perpendicular to the transducer array axis and the direction can be unambiguously determined. If the transducer arrays are separated by several wave lengths, the angle of the detected wave front may be derived from directions where maxima (or minima) of the received signal occur (Figure 3.42). However, this method, when used alone, produces only a few measurements. Using the gradient of the fringes to produce more measurements extends this method.

5.2.2.2 Forward looking sonars. Horizontal aperture sonars are used to detect obstructions ahead of the vessel by mechanical or electronic scanning in the horizontal plane. These systems are especially appropriate for the detection of obstructions in unsurveyed or poorly surveyed areas.

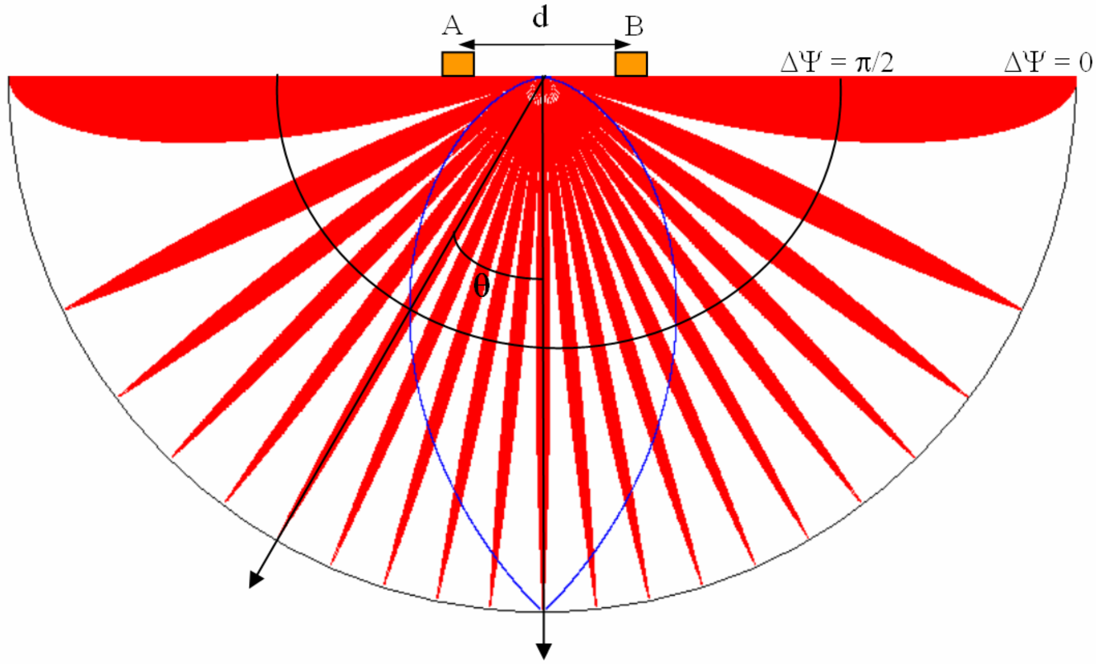


Fig. 3.42 "Pattern resulting from the interference, either constructive or destructive, from signals received at two arrays, separated by a distance (d) 10 times the acoustic wave length (red) and half wave length (blue)"

6. NON ACOUSTIC SYSTEMS

In addition to the acoustic systems, presented in previous sections, there are some electromagnetic systems which may be used for depth determination, such as the airborne laser and electromagnetic induction systems, as well as depth determination derived from satellite altimetry. These systems and the traditional mechanical methods for depth measurement and sweeping are described below.

6.1 Airborne Laser Systems

Laser[18] systems offer both an alternative and a complement to surveys with acoustic systems in shallow waters [Guenther et al., 1996].

A laser system is composed of a laser scanning system, global positioning system (GPS) and an inertial measurement unit (IMU).

6.1.1 Principles of operation

Hydrographic airborne laser sounding, LIDAR (LIght Detection And Ranging), is a system for measuring the water depth. This system emits laser pulses, at two frequencies (blue-green and infrared), in an arc pattern across the flight path of the airborne platform; it records both signal arrivals from the light pulse

[18] Laser is the acronym for Light Amplification by Stimulated Emission of Radiation. The laser basically consists of an emitting diode that produces a light source at a specific frequency.

reflected by the sea surface and by the seafloor (Figure 3.43). The measured time difference, between the two returns, is converted to distance.

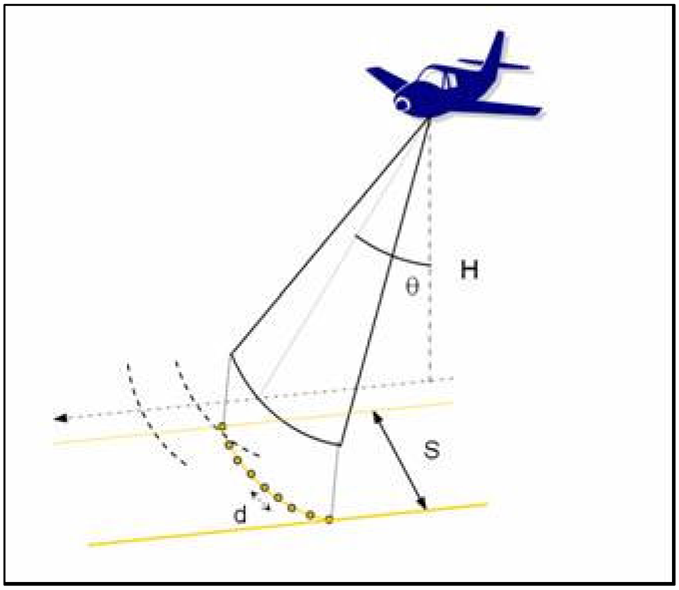


Fig. 3.43 "Geometry of Lidar measurements"

Propagation of light through sea water, like the propagation of acoustic energy, depends on the temperature, pressure and salinity. The sea water is, to some extent, transparent to light. In ideal conditions, no material in suspension, the attenuation is a function of the absorption and scattering.

The transparency of sea water to the infrared and optical regions of the electromagnetic spectrum is a function of the quantity of material suspended in the water. Therefore, water transparency[19] is a constraint to the use of laser sounding systems. Lidar maximum depth operation is about 2 to 3 times observed Secchi disk[20] depth.

A pulse of light of two different frequencies is transmitted in the direction of the sea; part of the energy from the infrared beam is reflected by the sea surface back towards the aircraft, the blue-green laser beam is transmitted to the water and partly reflected by the seafloor, also to be detected by the receiver. Using accurate timing, the distance to the seafloor can be measured knowing the speed of light in the water. The depth calculation requires, in addition, knowledge of the geometry of measurements according to Snell's law (Figure 3.44).

$$\frac{\operatorname{sen}\theta_a}{c_a} = \frac{\operatorname{sen}\theta_w}{c_w} \qquad \textbf{(3.85)}$$

where $\boldsymbol{\theta_a}$ and $\boldsymbol{\theta_w}$ are the incidence angles in the air and into the water and $\mathbf{c_a}$ and $\mathbf{c_w}$ the respective speeds of light in the air and the water.

---

19 The transparency of the sea water, i.e. the transmission of visible light through the water, can be measured quantitatively by determining the Secchi disk depth.

20 The Secchi disk is a simple device to measure the water transparency. The disk is a white plate, approximately 30 cm in diameter, fastened horizontally on the end of a rope marked in metres. The disk is lowered into the sea water and the depth at which it is lost to sight is noted.

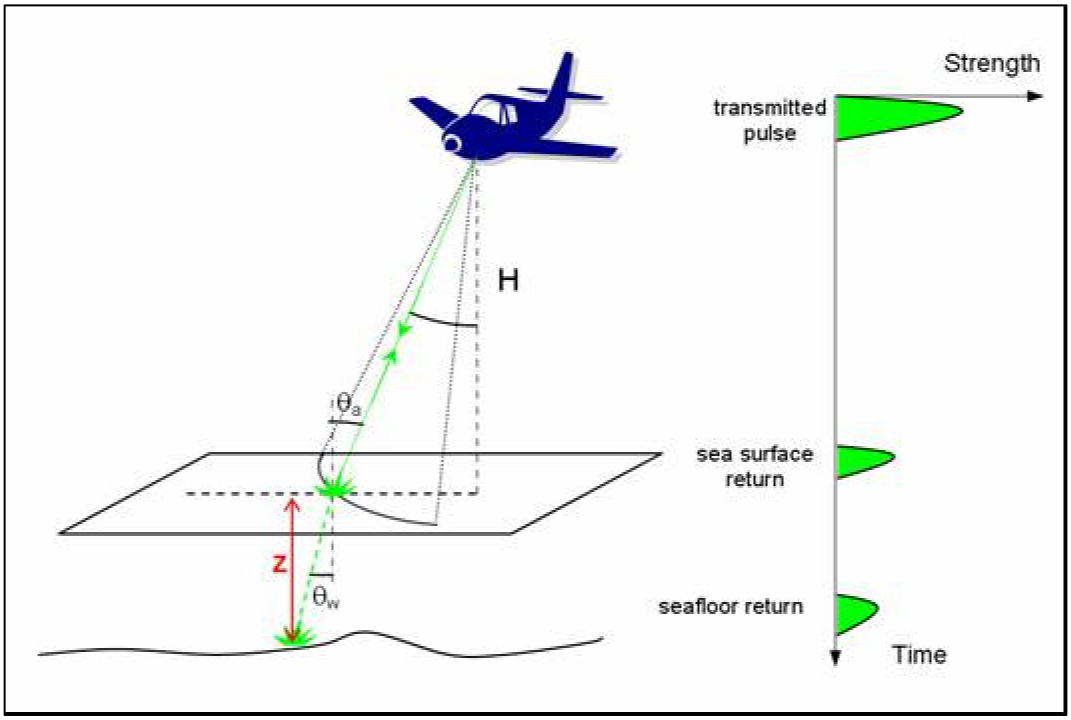


Fig. 3.44 "Lidar working principle"

6.1.2 Capabilities and limitations

Laser systems are efficient in shallow waters due to their outstanding productivity [Axelsson and Alfredsson, 1999]. This productivity results from the high surveying speed and the swath width, which is independent of water depth. In contrast, multibeam systems are operated at low surveying speed and the swath width is proportional to the water depth (usually 3 times the water depth).

Laser systems give good coverage, close to full coverage, in extreme conditions of salinity and temperature, where acoustic systems may produce poor quality data.

Undoubtedly, safety is a major advantage of laser system operation, particularly where under water hazards may be risky for surface navigation.

Despite the capabilities mentioned above, laser systems are very sensitive to suspended material and turbidity in the water column. The maximum operating depths, in optimal operation conditions, i.e. in very clear waters, is about 50-70 metres.

6.2 Airborne Electromagnetic Systems

Airborne electromagnetic induction systems have been used for over 40 years to detect highly conductive metallic mineral deposits. Advances in this technology have allowed the use of electromagnetic induction principles for mapping seafloor formations in shallow water. Detailed information on airborne electromagnetic induction systems is referred to Zollinger et al. [1987] and Smith and Keating [1996].

6.2.1 Principles of operation

The operating principle of these systems is based on a geophysical survey technique for measuring the electrical conductivity of bedrock or the thickness of a conductive layer.

A magnetic dipole transmitter, placed on a helicopter or a fixed wing aircraft, produces a magnetic field, the primary field, and a towed receiver is used to detect the secondary magnetic field induced in the ground.

Assuming horizontal layers, signal processing in time or frequency domain can be used to determine the conductivity, σ_w, and thickness of the seawater column, i.e. the water depth, and conductivity, σ_s, of the seafloor (Figure 3.45).

6.2.2 Capabilities and limitations

This no-acoustic system, due to the low frequencies involved, has the capability of operating over thick ice. However, this system is limited to water depths of less than 100 metres and is, at present, for reconnaissance purposes only.

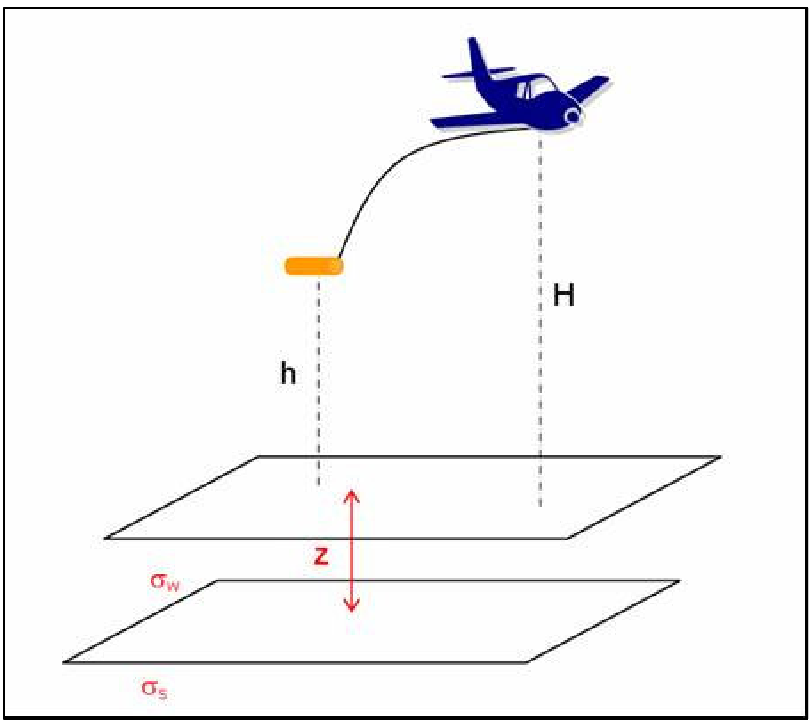


Fig. 3.45 "Electromagnetic airborne working principle"

6.3 Remote Sensing

This section presents depth estimation derived from aerial photography and from satellite altimetry, as an additional method for the coverage of extensive areas.

6.3.1 Photo-bathymetry

It is a common practise for aerial photography to be used to delineate the coast line and, sometimes, very helpful in reconnaissance, planning of hydrographic surveys, location of shoals and the creation of a qualitative description of the seafloor rather than a means by which to determine the water depth.

6.3.1.1 Principles of operation

Digital image processors have the ability to correlate light intensity with depth. However, variation in light intensity is also dependent on material in suspension and on the reflective properties of the seafloor. Thus a local calibration should be undertaken to account for these variations.

6.3.1.2 Capabilities and limitations

The application of photo-bathymetry, within the present limits of this technology, remains mainly a tool for reconnaissance and planning in areas where there is insufficient or no depth information.

6.3.2 Others

Satellite images in the visible band may be used in a similar way to photo-bathymetry. However, satellites may be equipped with high resolution altimeters for mapping the oceans' surface and, with the appropriate data processing, it is possible to estimate the depths all over the globe.

The ocean surface has an irregular shape which replicates, to some extent, the topography of the ocean floor. Seafloor features, such as seamounts, contribute to a local modification of the earth's gravity field, inducing a deflection of the vertical, which causes a slope in the sea surface and consequently the seawater is be pulled thus generating a bulge on the sea surface. The ocean surface can be mapped with an accurate satellite altimeter and the anomaly, i.e. the difference between the observed ocean surface and a theoretical surface, such as that created from WGS84 ellipsoid, can be determined and the water depth estimated (Figure 3.46).

The integration of satellite altimetry with bathymetric measurements may produce a more reliable data set which contributes to the knowledge of the seafloor topography in areas where hydrographic surveys are sparse [Smith and Sandwell, 1997].

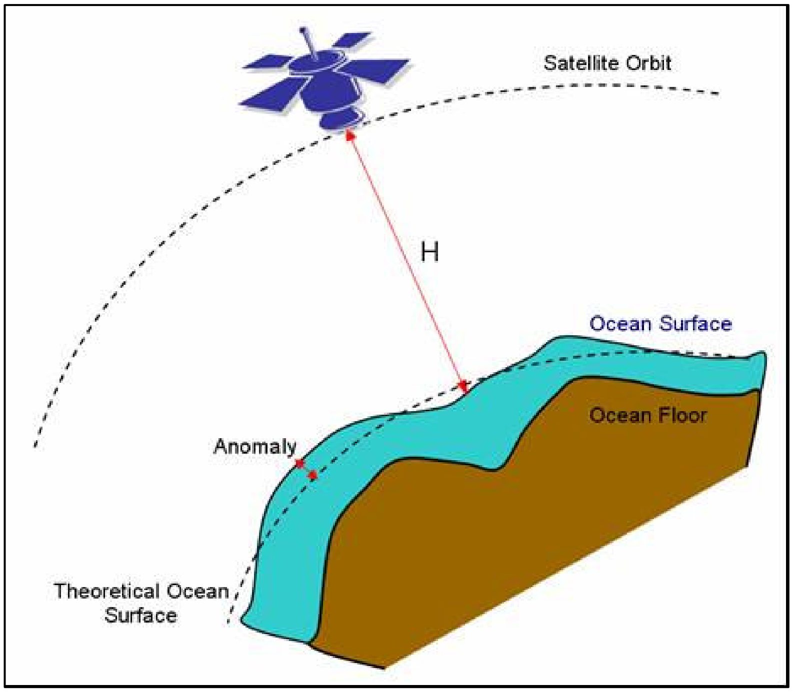


Fig. 3.46 "Satellite derived bathymetry"

6.4 Mechanic systems

Mechanic systems are the earliest tools used for depth measurement. Nonetheless, these systems still and their place and remain in use to this day.

The systems studied up until now perform indirect measurements and are sensitive to seawater characteristics. In typical conditions, gross errors in the depth measurements are likely to occur, these are generated by echoes from within the water column and therefore they do not related to the seafloor, for instance they can be caused by: kelp, schools of fish, deep scattering layer, thermal plumes and sediments in suspension. Additionally errors may occur near piers, where echo detection occurs from the returns of side lobes from the pier itself.

Mechanical methods (lead line or sounding pole) are not sensitive to these particular environmental conditions and may provide an alternatively method.

Bar or wire sweep methods are an unambiguous way to detect minimum depths over wrecks or over obstructions and to guarantee minimum depths throughout a navigation channel.

6.4.1 Lead line and sounding pole

The lead line aids the hydrographer in resolving echo sounder misinterpretation caused by spurious returns.

When the bottom is visible, a lead line or sounding pole can consistently be placed on the high points and the depth measured. In other areas, detection and measurement may be more difficult and sweeping methods may be preferred.

6.4.1.1 Description

A lead line is a graduated line with attached marks and fastened to a sounding lead. The line is used for determining the depth of water when sounding manually, generally, in depths of less than 50 metres.

A correction, to compensate for the shrinking and stretching of the line, may be applied to the depths obtained; this error source has, however, been overcome by inserting a wire heart inside the rope.

A sounding pole is a pole graduated with marks which is also used for determining the depth of water when manual sounding, generally it is used in depths of less than 4 metres.

As previously mentioned, at present, these tools are often used to check anomalous soundings gathered with acoustical systems which occur in shallow waters.

6.4.1.2 Sources of error

The sources of depth measurement error with lead line are mainly due to:

a) Line curvature is induced by current and produces a depth error. The correction may be problematical and, for this reason, it is recommended only undertaking the measurements when the impact of the current will be negligible, the only remaining the affect being the residual speed of the vessel.

b) Heave will contribute to the error in the depth measurement. Heave leads to difficulty in reading the depth; this is overcome by taking an average of the reading between wave crests and troughs.

6.4.1.3 Operation, data recording, and processing

The direct depth measurement should be performed with the vessel dead in the water and, if possible, avoiding periods of strong currents and tidal flow. It is normal, between successive sounding positions, to keep the lead line in the water to check for any prominent seafloor features.

6.4.2 Bar sweep

Bathymetric coverage with SBES only measures the depth along the survey lines, leaving the seafloor between lines without coverage or detailed information, although side scan sonar is often used to search and locate any prominent seafloor features between the SBES lines. For rock pinnacles or wrecks, SBES may not detect the minimum depth when the echo is possibly too weak to be sensed by the receiver, this is particularly the case for masts or sharp pieces of metal.

For navigation safety purposes, the use of an accurate mechanical sweep, either bar sweep or wire sweep, is an adequate means to guarantee a minimum safe clearance depth throughout an area and, according to S-44, it may be considered sufficient for Special Order and Order 1 surveys.

6.4.2.1 Description

The sweep is made of a bar about 5-6 metres in length. Each end of the bar may be packed with lead, or other heavy material, to provide more weight and reduce lift when underway. The bar is suspended below the vessel by graduated lines.

This instrument is very easy to manufacture. Trial and error tests may be used to obtain the best solution.

This is often more effective and easier to handle than a wire sweep.

6.4.2.2 Operation methodology

The bar or rod should be suspended horizontally under a vessel. The sweep may be equipped with rockers or other sensors to record contact with the seabed.

The depth of the bar should be referred to the vertical datum, the tide height should be recorded during the sweep operation and depths reduced as appropriate.

A complete coverage of the navigation area at a safe clearance depth should be performed; in the event of an obstruction being detected, full coverage around the obstruction is recommended to confirm that the minimum depth is detected.

6.4.3 Wire sweep

As an alternative to the bar sweep, a wire sweep may be used to determine the least depth over a bathymetric feature when, from the general nature of the visible terrain, the existence of a rock pinnacle or obstruction is suspected.

Detailed information on wire sweeping may be found in NOAA [1976].

6.4.3.1 Description

The sweep is constructed from two small trawl boards or doors (identical to those used by fish trawlers). The trawl boards are connected by 40 to 60 metres of oval link chain. The sweep is bridled and towed so that the connecting chain is dragged along the seabed approximately 60 metres astern of the towing vessel NOAA [1976].

REFERENCES

Artilheiro, F. (1996)	*"Analysis and Procedures of Multibeam Data Cleaning for Bathymetric Charting"*	Master's Report, Department of Geodesy and Geomatics Engineering, University of New Brunswick, Fredericton New Brunswick, Canada.
Axelsson, R. and M. Alfredsson (1999)	*"Capacity and Capability for Hydrographic Missions"*	US Hydrographic Conference 1999.
Clay, C. e H. Medwin (1977)	*"Acoustical Oceanography"*	Wiley and Sons, Toronto.
de Moustier, C. (1988)	*"State of the Art in Swath Bathymetry Survey Systems"*	International Hydrographic Review, LXV(2), p. 25.
de Moustier, C. (1993).	*"Signal Processing for Swath Bathymetry and Concurrent Seafloor Acoustic Imaging"*	Acoustic Signal Processing for Ocean Exploration, J.M.F. Moura and I.M.G. Lourtie Eds., pp. 329-354.
Godin, A. (1996).	*"The Calibration of Shallow Water Multibeam Echo-Sounding Systems"*	Proceedings of the Canadian Hydrographic Conference '96, Halifax, NS, Canada, pp. 25-31.
Guenther, G., R. Thomas, and P. LaRocque (1996).	*"Design Considerations for Achieving High Accuracy with the SHOALS Bathymetric Lidar System"*	SPIE: Laser Remote Sensing of Natural Waters: From Theory to Practice. 15, pp. 54-71.
Hare, R. (1995).	*"Depth and Position Error Budgets for Multibeam Echosounding"*	International Hydrographic Review (LXXII), Monaco, pp 37-69.
Hughes Clarke, J. (2000).	*"Present-day Methods of Depth Measurement"*	In: P. Cook and C. Carlton (eds) Continental Shelf Limits - The Scientific and Legal Interface. Oxford University Press, New York.
IHO (1994).	*"Hydrographic Dictionary. Special publication No. 32, 5th edition"*	International Hydrographic Organization, Monaco.
IHO (1998).	*"IHO Standards for Hydrographic Surveys. Special publication No. 44, 4th edition"*	International Hydrographic Organization, Monaco.

Kinsler, L., A. Frey, A. Coppens, and J. Sanders (1982).	*"Fundamentals of Acoustics".*	Wiley and Sons, Toronto.
Lurton, X. (2002).	*"Acoustical Measurement Accuracy Modelling for Bathymetric Sonars"*	Canadian Hydrographic Conference 2002.
NOAA (1976).	*"Hydrographic Manual. 4th edition"*	National Oceanic and Atmospheric Administration. US Department of Commerce.
Pφhner, F. (1993).	*"Model for Calculation of Uncertainty in Multibeam Depth Soundings"*	Report from Simrad Subsea AS, Horten, Norway, FEMME 93, 16 p.
Pickard, G. and W. Emery (1990).	*"Descriptive Physical Oceanography – An Introduction, 5th edition"*	Pergamon Press, Oxford.
Seippel, R. (1983).	*"Transducers, Sensors and Detectors"*	Prentice-Hall.
Smith, R. and P. Keating (1996).	*"The usefulness of multicomponent, time-domain airborne electromagnetic measurements"*	Geophysics, Vol. 61, No. 1, pp. 74–81.
Smith, W. and D. Sandwell (1997).	*"Global Seafloor Topography from Satellite Altimetry and Ship Depth Sounding"*	Science 277. pp. 1956-1962.
OMG (1996).	*"Multibeam Sonar Surveying Training Course. Ocean Mapping Group"*	University of New Brunswick.
Urick, R. (1975).	*Principles of Underwater Acoustics.*	McGraw-Hill, Toronto.
Zollinger, R., H. Morrinson, P. Lazenby, and A. Becker (1987).	*"Airborne Electromagnetic Bathymetry"*	Geophysics, Vol. 52 no. 8, pp. 1172-1137.

BIBLIOGRAPHY

Andersen, O. and P. Knudsen (1998).	*"Global Gravity Field from ERS1 and GEOSAT Geodetic Mission Altimetry"*	Journal Geophysics Research 103(C4), pp. 8129-8137.
Arabelos, D. (1997).	*"On the Possibility to Estimate the Bottom Topography from Marine Gravity and Satellite Altimetry Data Using Collocation"*	In: R. Forsberg, M. Feissel, R. Dietrich (eds) Geodesy on the Move Gravity, Geoid, Geodynamics, and Antarctica IAG Symposia 119, Springer – Verlag Berlin Heidelberg, pp. 105-112.
Burtch, R. (2002).	*"Lidar Principles and Applications"*	IMAGIN Conference 2002, Traverse City.
Calmant, S. and N. Baudry (1996).	*"Modelling Bathymetry by Inverting Satellite Altimetry Data: A Review"*	Marine Geophysics Research 18, pp. 23-134.
Collet, C., J. Provost, P. Rostaing, P. Pérez, and P. Bouthemy (2000).	*"SPOT Satellite Data Analysis for Bathymetric Mapping"*	IEEE, pp. 964-967.
Denbigh, P. (1989).	*"Swath bathymetry: Principles of operation and an analysis of errors"*	IEEE Journal of Oceanic Engineering 14, pp. 289–298.
Dixon, T., M. Naraghi, M. McNutt, and S. Smith (1983).	*"Bathymetric Prediction from SEASAT Altimeter Data"*	Journal Geophysics Research 88, pp. 1563-1571.
Durey, L., G. Terrie, R. Arnone, and A. Martinez (1997).	*"Bottom Reflectance Maps from Hyperspectral Sensors: An Application to AAHIS Data"*	In Proceedings, ERIM Fourth International Conference on Remote Sensing for Marine and Coastal Environments, Orlando, pp. 17-19.
Geng, X. and A. Zielinski (1999).	*"Precise Multibeam Acoustic Bathymetry"*	Marine Geodesy, 22, pp. 157-167.
Guenther, G. (1985).	*"Airborne Laser Hydrography: System Design and Performance Factors"*	NOAA Professional Paper Series, National Ocean Service.
Guenther, G., A. Cunningham, P. LaRocque, and D. Reid (2000).	*"Meeting the Accuracy Challenge in Airborne Lidar Bathymetry"*	*Proceedings of EARSeL Symposium 2000*. Dresden, Germany.

Guenther, G., M. Brooks, and P. LaRocque (1998).	*"New Capabilities of the SHOALS Airborne Lidar Bathymeter"*	Proceedings 5th International Conference on Remote Sensing for Marine and Coastal Environments, ERIM International, October 5-7, San Diego, CA, Vol. I, 47-55.
Guenther, G., P. LaRocque, and W. Lillycrop (1994).	*"Multiple Surface Channels in SHOALS Airborne Lidar"*	SPIE: Ocean Optics XII, Vol. 2258, pp. 422-430.
Guenther, G., R. Thomas, and P. LaRocque (1996).	*"Design Considerations for Achieving High Accuracy with the SHOALS Bathymetric Lidar System"*	SPIE: Laser Remote Sensing of Natural Waters from Theory to Practice, Vol. 2964, pp. 54-71.
Hammerstad E. (1995).	*"Simrad EM 950/1000 - Error Model for Australian Navy"*	Extract of Report, Simrad Subsea AS, Horten, Norway, 4 p.
Hare, R. and A. Godin (1996).	*"Estimating Depth and Positioning Errors for the Creed/ EM 1000 Swath Sounding System"*	Proceedings of the Canadian Hydrographic Conference '96. Halifax, NS, Canada, pp. 9-15.
Herlihy, D., B. Hillard, and T. Rulon (1989).	*"National Oceanic and Atmospheric Administration Sea Beam System - Patch Test"*	International Hydrographic Review, Monaco, LXVI(2), pp. 119-139.
Hughes Clarke, J. (1995).	*"Reference Frame and Integration."*	Lecture IV-1 in Coastal Multibeam Hydrographic Surveys. United States / Canada Hydrographic Commission Multibeam Working Group, St. Andrews, New Brunswick, Canada.
Hughes Clarke, J. (1995a).	*"Interactive Bathymetry Data Cleaning"*	Lecture X-4 from Coastal Multibeam Hydrographic Surveys. United States / Canada Hydrographic Commission Multibeam Working Group, St. Andrews, New Brunswick, Canada.
Ingham, A. (1992).	*"Hydrography for the Surveyor and Engineer"*	3rd edition, BSP, Oxford.
Irish, J. and W. Lillycrop (1999).	*"Scanning Laser Mapping of the Coastal Zone: The SHOALS System"*	ISPRS Journal of Photogrammetry and Remote Sensing, 54. pp. 123-129.
Irish, J., J. McClung, and W. Lillycrop (2000).	*"Airborne Lidar Bathymetry: the SHOALS System"*	PIANC Bulletin. 2000 (103), pp. 43-53.
Jung, W. and P. Vogt (1992).	*"Predicting Bathymetry from Geosat ERM and Ship Borne Profiles in the South Atlantic Ocean"*	Tectonophysics 210, pp. 235-253.

Lillycrop W., L. Parson, and J. Irish (1996). *"Development and Operation of the SHOALS Airborne Lidar Hydrographic Survey System"* SPIE: Laser Remote Sensing of Natural Waters from Theory to Practice, Vol. 2964, pp. 26-37.

Lillycrop, W. and J. Banic, (1993). *"Advancements in the US Army Corps of Engineers Hydrographic Survey Capabilities: The SHOALS System"* Marine Geodesy, Vol. 15, pp. 177-185.

Lillycrop, W., J. Irish, and L. Parson (1997). *"SHOALS System: Three Years of Operation with Airborne Lidar Bathymetry - Experiences, Capability and Technology Advancements"* Sea Technology, Vol. 38, No. 6, pp. 17-25.

Lillycrop, W., L. Parson, L. Estep, P. LaRocque, G. Guenther, M. Reed, and C. Truitt (1994). *"Field Test Results of the U.S. Army Corps of Engineers Airborne Lidar Hydrographic Survey System"* Proceedings of the 6th Biennial National Ocean Service International Hydrographic Conference, Norfolk, VA, pp. 144-151.

Parson, L., W. Lillycrop, C. Klein, R. Ives, and S. Orlando (1996). *"Use of LIDAR Technology for Collecting Shallow Bathymetry of Florida Bay"* Journal of Coastal Research, Vol. 13, No. 4.

Pope, R., B. Reed, G West, and W. Lillycrop. (1997). *"Use of an Airborne Laser Depth Sounding System in a Complex Shallow-water Environment"* Proceedings of Hydrographic Symposium XVth International Hydro Conference. Monaco.

Quinn, R., (1992), *"Coastal Base Mapping with the LARSEN Scanning Lidar System and Other Sensors"* Proceedings, 5th Biennial National Ocean Service International Hydrographic Conference, Baltimore.

Riley, J. (1995). *"Evaluating SHOALS Bathymetry Using NOAA Hydrographic Survey Data"* Proceedings 24th Joint Meeting of UNIR Sea Bottom Surveys Panel, Tokyo, Japan.

Sinclair, M. (1998). *"Australians Get on Board with New Laser Airborne Depth Sounder"* Sea Technology, June 1998, pp. 19-25.

Sinclair, M. (1999). *"Laser Hydrography - Commercial Survey Operations"* Hydro 99.

Smith, R. and M. Smith (2000). *"Airborne Lidar and Airborne Hyperspectral Imagery: A Fusion of Two Proven Sensors for Improved Hydrographic Surveying"* Proceedings Canadian Hydrographic Conference 2000.

Thomas, R. and G. Guenther (1990). *"Water Surface Detection Strategy for an Airborne Laser Bathymeter"* SPIE: Ocean Optics X, Vol. 1302, pp. 597-611.

USACE (2002).	*"Hydrographic Surveying Manual"*	U.S. Army Corps of Engineers, Department of the Army, Washington.
Vergos, G. and M. Sideris (1998).	*"On Improving the Determination of the Gravity Field by Estimating the Bottom Ocean Topography with Satellite Altimetry and Shipborne Gravity Data"*	Department of Geomatics Engineering, University of Calgary.
Whitman, E. (1996).	*"Laser Airborne Bathymetry - Lifting the Littoral"*	Sea Technology, August 1996, pp. 95-98.
Wozencraft, J. (2001).	*"The Coastal Zone Revealed Through Shoals Lidar Data"*	Proceedings US Hydrographic Conference 2001.
Wright, C. and J. Brock (2002).	*"EAARL: A LIDAR for Mapping Shallow Coral Reefs and Other Coastal Environments"*	Seventh International Conference on Remote Sensing for Marine and Coastal Environments Proceedings 2002.
Yakima, W., Wilt M., H. Morrison, K. Lee, and N. Goldstein (1989).	*"Electromagnetic Sounding in the Columbia Basin"*	Geophysics, Vol. 54, No. 8, pp. 952-961.

CHAPTER 3 – ANNEX A
REFERENCE AND COORDINATE SYSTEMS

A. Reference and Coordinate Systems

Depth determination is performed in a vessel in dynamic conditions. Usually, a reference system (vessel coordinate system), three orthogonal axes, is used on board to locate the hydrographic sensors and to measure the vessel's attitude and heave.

The vessel's attitude consists of angular displacements about those axes, roll (transversally) about the x axis, pitch (longitudinally) about the y axis, and yaw (horizontally) about the z axis. Considering an orthogonal right-hand reference system with the z axis pointing downward; with the usual convention for most attitude sensors roll is positive when starboard side is down, pitch is positive when bow is up, and yaw is positive when rotating clockwise.

Considering Figure A.1, the rotation θ_1 in the yz plane, i.e., rotation about the x axis, can be expressed by the rotation matrix,

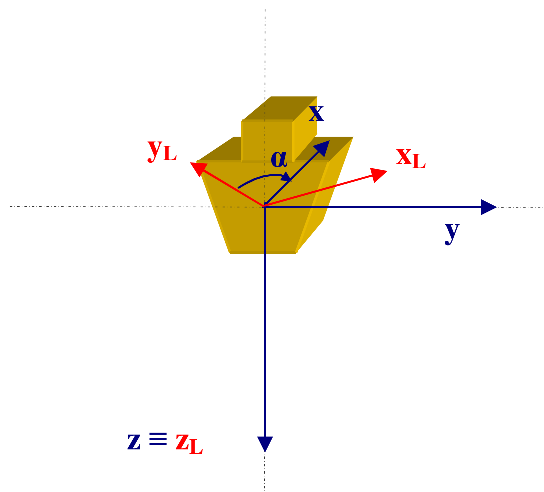


Fig. A.1. "Vessel reference system"

$$R_1(\theta) = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos(\theta) & \sin(\theta) \\ 0 & -\sin(\theta) & \cos(\theta) \end{bmatrix} \quad \textbf{(A. 1)}$$

and the rotations about the *y* and *z* axes are respectively:

$$R_2(\theta) = \begin{bmatrix} \cos(\theta) & 0 & -\sin(\theta) \\ 0 & 1 & 0 \\ \sin(\theta) & 0 & \cos(\theta) \end{bmatrix}, \quad \textbf{(A. 2)}$$

$$R_3(\theta) = \begin{bmatrix} \cos(\theta) & \sin(\theta) & 0 \\ -\sin(\theta) & \cos(\theta) & 0 \\ 0 & 0 & 1 \end{bmatrix}. \qquad \textbf{(A. 3)}$$

The transformation which results from three sequential rotations is represented by the product of the rotation matrices. Successive rotations are applied to the left of this product.

Considering the successive rotations (θ_1, θ_2, θ_3) about the x, y and z axes, the transformation matrix is given by,

$$R_3(\theta_3) \cdot R_2(\theta_2) \cdot R_1(\theta_1) =$$

$$= \begin{bmatrix} \cos(\theta_3)\cos(\theta_2) & \sin(\theta_3)\cos(\theta_1)+\cos(\theta_3)\sin(\theta_2)\sin(\theta_1) & \sin(\theta_3)\sin(\theta_1)-\cos(\theta_3)\sin(\theta_2)\cos(\theta_1) \\ -\sin(\theta_3)\cos(\theta_2) & \cos(\theta_3)\cos(\theta_1)-\sin(\theta_3)\sin(\theta_2)\sin(\theta_1) & \cos(\theta_3)\sin(\theta_1)+\sin(\theta_3)\sin(\theta_2)\cos(\theta_1) \\ \sin(\theta_2) & -\cos(\theta_2)\sin(\theta_1) & \cos(\theta_2)\cos(\theta_1) \end{bmatrix}. \qquad \textbf{(A. 4)}$$

The measured depths, initially referred to the vessel's frame, need to be positioned in a local coordinate system. Considering an orthogonal left-hand local coordinate system; with the x axis pointing to East, y axis pointing to the geographic North, and the z axis pointing downward.

In a survey vessel with roll, pitch, and heading respectively: θ_R, θ_P, and α; a beam with slant range R and angle β (Figure A.2), will be transferred from the tri-orthogonal, right-hand, vessel coordinate system $(x, y, z)_V$ to the tri-orthogonal, left-hand, local coordinate system $(x, y, z)_L$, with the rotation about the x axis the reciprocal of the roll angle ($-\theta_R$), the rotation about the y axis the reciprocal of the pitch angle ($-\theta_p$), and the rotation about the z axis the reciprocal of the heading angle ($-\alpha$) and since the two z axes are both positive downward, but the vessel coordinate system is a right-hand system and the local coordinate system is a left-hand system, it is necessary to swap the x and y coordinates during the transformation from vessel to local coordinate systems. This is performed by applying the matrix R_{xy}.

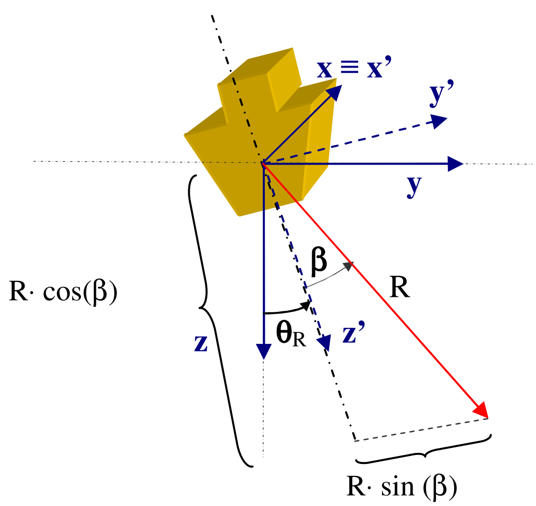


Fig. A.2 "Vessel and local level reference system"

$$\begin{bmatrix} x \\ y \\ z \end{bmatrix}_L = Rxy \cdot R_3(-\alpha) \cdot R_2(-\theta_P) \cdot R_1(-\theta_R) \begin{bmatrix} x \\ y \\ z \end{bmatrix}_V = T(\alpha, \theta_P, \theta_R) \begin{bmatrix} 0 \\ R\sin(\theta) \\ R\cos(\theta) \end{bmatrix}_V \qquad \textbf{(A. 5)}$$

where R_{xy} is given by,

$$Rxy = \begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

$$T(\alpha, \theta_P, \theta_R) = \begin{bmatrix} \sin(\alpha)\cos(\theta_P) & \cos(\alpha)\cos(\theta_R) + \sin(\alpha)\sin(\theta_P)\sin(\theta_R) & -\cos(\alpha)\sin(\theta_R) + \sin(\alpha)\sin(\theta_P)\cos(\theta_R) \\ \cos(\alpha)\cos(\theta_P) & -\sin(\alpha)\cos(\theta_R) + \cos(\alpha)\sin(\theta_P)\sin(\theta_R) & \sin(\alpha)\sin(\theta_R) + \cos(\alpha)\sin(\theta_P)\cos(\theta_R) \\ -\sin(\theta_P) & \cos(\theta_P)\sin(\theta_R) & \cos(\theta_P)\cos(\theta_R) \end{bmatrix}$$

where $T(\alpha, \theta_P, \theta_R)$ is the transformation matrix from reference frame measurements into the local coordinate system.

$$\begin{bmatrix} x \\ y \\ z \end{bmatrix}_L = \begin{bmatrix} (\cos(\alpha)\cos(\theta_R) + \sin(\alpha)\sin(\theta_P)\sin(\theta_R)) \cdot R\sin(\beta) + (-\cos(\alpha)\sin(\theta_R) + \sin(\alpha)\sin(\theta_P)\cos(\theta_R)) \cdot R\cos(\beta) \\ (-\sin(\alpha)\cos(\theta_R) + \cos(\alpha)\sin(\theta_P)\sin(\theta_R)) \cdot R\sin(\beta) + (\sin(\alpha)\sin(\theta_R) + \cos(\alpha)\sin(\theta_P)\cos(\theta_R)) \cdot R\cos(\beta) \\ \cos(\theta_P)\sin(\theta_R) \cdot R\sin(\beta) + \cos(\theta_P)\cos(\theta_R) \cdot R\cos(\beta) \end{bmatrix}$$

Note that the beam angle is positive to starboard side and negative to port.

CHAPTER 4
SEAFLOOR CLASSIFICATION AND FEATURE DETECTION

by Lieutenant Commander Peter JOHNSON (Australia)

1. INTRODUCTION

1.1 Hydrography includes the description of the features of the seas for a number of purposes not restricted to navigation. The advent of sonar and swath echo sounders now enables a more complete and detailed description to the benefit of safer navigation and other uses. Obviously, it is impracticable to find every feature in every depth so the IHO have determined the minimum size of feature which should be searched for and measured in any particular area. Classification of the seafloor has been employed for minewarfare operations for many years but the advent of automated classification software has enabled wider usage, particularly in fishery and resource industries.

1.2 In this chapter, the phrases seafloor classification and seafloor characterisation, and feature detection and object detection are synonymous

2. SEAFLOOR FEATURE DETECTION

2.1 Background

2.1.1 To ensure safe navigation it is necessary to detect features on the seafloor which may be a hazard to navigation, whether natural or man made. A feature is defined as any item on the seafloor which is distinctly different from the surrounding area; it can be anything from an isolated rock on a flat sand seafloor to a wreck or obstruction. This activity is called seafloor feature detection. Feature detection can also be used to detect and identify features which are of interest to other seafarers, such as wellheads and mine-like features. The latter may not be of navigational significance but are, nonetheless, of importance to those concerned.

2.1.2 A traditional survey will develop the bathymetry of an area by running a regular series of sounding lines throughout the area. Multibeam echo sounder (MBES) or side scan sonar (SSS) coverage is utilised for feature detection and to provide information regarding seafloor classification. In some instances the detection of features is more important than the acquisition of bathymetry. Specific features which have been identified on the MBES or SSS image will usually require a more positive check of its position and the least depth.

2.2 Standards

2.2.1 There are a number of feature detection standards the most relevant being those contained in IHO S-44 and IHO S-57.

2.2.2 IHO S-44 - Minimum Standards for Hydrographic Surveys

2.2.1.1 S-44 Table 1, summarised at Tables 4.1 and 4.2 below, specifies where a feature search is to be undertaken and system detection capabilities for each Order of survey:

2.2.1.2 Once detected any features considered significant should have its position and the least depth over it determined to the standards detailed in S-44 Table 1.

IHO S-44 Order and example areas		Search Requirement
Special Order	harbours, berthing areas and associated critical channels with minimum under keel clearances.	100% search compulsory.
Order 1	harbours, harbour approach channels, recommended tracks and some coastal areas with depths up to 100 m.	100% search required in selected areas.
Order 2	areas not described in Special Order and Order 1 or areas up to 200 m water depth.	100% search may be required in selected areas.
Order 3	offshore areas not described in Special Order and Orders 1 & 2.	no search required.

Table 4.1 "IHO S-44 Search Requirements"

<table>
<tr><th>IHO S-44 Order</th><th>System Detection Capabilities</th></tr>
<tr><td>Special Order</td><td>cubic features >1.0 m detectable.</td></tr>
<tr><td>Order 1</td><td rowspan="2">cubic features >2.0 m in depths down to 40 m detectable or 10% of depth beyond 40 m (this depth chosen with regard to the maximum expected draught of vessels).</td></tr>
<tr><td>Order 2</td></tr>
<tr><td>Order 3</td><td>not applicable.</td></tr>
</table>

Table 4.2 "IHO S-44 System Detection Capabilities"

2.2.3 IHO S-57 - Transfer Standards for Digital Hydrographic Data

2.2.3.1 S-57 specifies "Zones of Confidence" (ZOC) as the method of encoding data quality information. ZOC were adopted to provide a simple and logical means of classifying all bathymetric data and displaying to the mariner the confidence the national charting authority places in it. Areas are classified by identifying various levels of confidence that can be placed in underlying data using a combination of depth and position accuracy, thoroughness of seafloor search and conformance to an approved quality plan.

2.2.3.2 ZOC A1, A2 and B are generated from modern and future surveys with, significantly, ZOC A1 and A2 requiring a full seafloor search, i.e. full feature detection. ZOC C and D reflect low accuracy and poor quality data, whilst ZOC U represents data which is unassessed, but not unsurveyed, at the time of publication. ZOC are designed to be depicted on paper charts, as an insert diagram in place of the current reliability diagram, and on electronic displays.

2.2.3.3 It must be emphasised that ZOC are a charting standard and are not intended to be used for specifying standards for hydrographic surveys or for management of data quality by individual hydrographic authorities. Depth and position accuracy specified for each ZOC refer to errors of final depicted soundings and include not only survey errors but other errors introduced in the chart production process.

2.2.3.4 S-57 ZOC Feature Detection criteria are at Table 4.3:

S-57 ZOC	Search Requirement
ZOC A1	full area search undertaken, all significant seafloor features detected and have had their depths measured. (see Note)
ZOC A2	
ZOC B	full area search not achieved, uncharted features hazardous to navigation may exist.
ZOC C	full area search not achieved, depth anomalies may be expected.
ZOC D	full area search not achieved, large depth anomalies may be expected.
ZOC U	quality of bathymetric data yet to be assessed.

Table 4.3 "ZOC Feature Detection Criteria"

Note: Significant seafloor features are defined in S-57 as those rising above depicted depths by more than:

- 0.1 x depth, in depths <10 m;
- 1.0 m in depths of 10-30 m;
- (0.1 x depth) minus 2.0 m in >30 m.

2.2.3.5 S-57 also details the relevant position and depth accuracy required of measured features.

2.2.4 Detection of Hazardous Features

2.2.4.1 The surveyor must remain cognisant of the fact that many features which are potentially hazardous to navigation do not fit the S-44 "cubic feature" criteria; for example the masts of wrecks and wellheads. However, ZOC criteria do take such features into account if they rise above depicted depths by the prescribed amount. The ability to detect such features is a critical issue when considering the type of system to be used to undertake feature detection. For instance, these types of features will normally be detected by SSS but may not be detected by multibeam, lidar and other such systems due, for example, to the beam footprint or "filtering" algorithms.

2.2.4.2 As far as the surveyor is concerned the purpose of a sonar sweep is to ensonify the area between adjacent lines of soundings in order to detect any feature of significance to the mariner. Although no hard and fast definition of the minimum length of a wreck can be given, features less than three metres in length are unlikely to be sufficiently proud of the seafloor to cause concern. There will of course be occasions when this is not so (i.e. in coral areas or when searching for masts) and the Surveyor must examine all sources of data available to him before deciding on the minimum length feature he wishes to detect.

2.2.4.3 Note that in all calculations that follow, involving speeds over the ground that must not be exceeded, the feature length is used and no account is taken of feature height. What is used for calculations is the maximum length of feature that just fails to receive five 'pings', this being considered the minimum to achieve feature detection. How much of the energy in the five pings on the feature that returns to the transducer is dependent upon:

- feature shape, extent, composition and aspect;
- sonar conditions;
- nature of the seafloor and other factors.

2.2.4.4 The amount of energy returned from the feature will control the intensity of the printed mark.

2.2.5 Military Requirements

2.2.5.1 Military forces often require detection of features smaller or deeper than those required for the safety of navigation, for example some strive to detect features with a volumetric size of 0.5 m on the continental shelf in depths to 200 m. Minewarfare forces, using specialised sensors, aim to detect and classify even smaller features. Whilst these reflect particular capabilities not normally required of the surveyor employed in nautical charting, there is a knock on effect in the development of systems capable of achieving them becoming available on the commercial market.

2.2.6 Reporting Features

2.2.6.1 Whilst it is desirable to investigate every feature which meets the above criteria in complex areas this will not be possible. Surveyors may need to use their own judgement as to which features warrant investigation after considering the available resources, the likely use of the area (draught of vessels etc) and the likely significance of the feature noting the general depths in the area. For example, a shoal of 26 m in general depths of 28 m may not warrant further investigation if the draught of vessels using the area is only 12 m. This will particularly be the case if a ship transiting the area must at some point pass through general depths of, say, 20 m. In such cases it may only be necessary to ensure that there is no indication of much shoaler water (i.e. by interlining, sonar etc.).

2.2.6.2 The above criteria should also be used to ascertain whether or not a feature should be included in any Report of Survey. In complex areas this list can become unwieldy; therefore the Report need only include those features which are truly significant in terms of general depths and likely usage.

2.2.6.3 At the end of each survey the surveyor, being the only person with all the facts at his disposal, must give a firm opinion as to the status of each feature located, i.e. wreck, sea floor type, unexamined etc., with findings included in the Report. Newly discovered features, which may be dangerous to surface or submarine navigation, and charted features, which are found to be significantly changed, are to be reported to the responsible National Hydrographic Office (NHO) immediately. Uncharted features in depths less than 750 m would normally be considered for Notice to Mariners action.

2.3 Methods of feature detection

2.3.1 Overview

2.3.1.1 There are a number of methods with which to achieve feature detection. SSS has a well proven feature detection capability and can still be considered the most reliable means. However, SSS is subject to operational limitations in that it is generally towed behind the survey vessel, which introduces positional errors for features. These errors can be reduced by use of transponders in the towfish and/or running past the feature in the opposite direction to obtain an average position. SSS operations are also subject to the nadir gap which requires lines to be run with sufficient overlap to detect features under adjacent tracks.

2.3.1.2 One of the main limitations of SSS is the speed of advance required to achieve sufficient pings on a particular feature. With few exceptions this limits side scan operations to about six knots, which impacts rate of effort. The advent of MBES offers the chance of meeting feature detection requirements at higher speeds and therefore increased rate of effort. To date, however, MBES detection of features of the size that meet IHO Special Order and ZOC A1/A2 requirements or other small and potentially hazardous features, cannot be guaranteed unless certain precautions are taken, such as limiting the useable swath width and calculating an appropriate speed of advance for 'ping' rate.

2.3.1.2.1 The geometry of a SSS transducer in relation to a feature is the key factor which makes it such a successful tool for feature detection. The shadows cast behind a feature, proud of the seafloor, are the telltale sign that a feature has been ensonified. The geometry of the MBES transducer in relation to seafloor features results in the loss of almost all shadow-casting capability. A surveyor wishing to use MBES for feature detection must then rely on the multibeam's other characteristics in order to look for any features. These characteristics are high resolution bathymetry and amplitude backscatter coupled with a positioning capability allowing for very accurate repeatability. In addition, whilst features are normally capable of being detected by an operator during SSS data acquisition, detection using MBES is far more uncertain at this stage and post processing is usually required to allow results to be seen.

2.3.1.3 Other sensors which can be used for feature detection include single beam echo sounder, forward looking sonar, magnetometer and remote methods such as airborne lidar bathymetry (ALB) and airborne electromagnetic bathymetry (AEMB). Mechanical feature detection methods, less used these days, include wire sweep, drag and diver.

2.3.2 Side Scan Sonar (SSS)

2.3.2.1 Dual-channel SSS is now accepted as an essential aid to modern surveying and it remains the case that no survey on the continental shelf can be considered complete unless a comprehensive sonar sweep has been carried out and all contacts investigated.

2.3.2.2 In addition to locating wrecks and obstructions between survey lines, SSS also provides a considerable amount of other seafloor information. These data, when combined with seafloor samples and depth contours to produce seafloor classification, are of great value to those involved with amphibious, minewarfare and submarine operations. The importance of this information has grown over the years to such an extent that, in many surveys, sonar rather than bathymetric considerations govern the selection of line direction and spacing. However, great care is needed in the preparation and checking of these data if their full potential is to be realised.

2.3.2.3 When used in hydrographic surveying, SSS has four main functions:

- The detection of wrecks and obstructions between sounding lines. Although precise position and least depth cannot be determined by SSS, a properly tuned and operated sonar will detect nearly all significant features between lines.
- The detection of other seafloor features. Correctly used, SSS can detect very small seafloor features. Whilst not hazardous to navigation the positions of such features, or groups of features, are of considerable importance in both submarine and minewarfare operations.
- The gathering of seafloor classification data. Knowledge of the texture of the seafloor, combined with samples, is of great importance for submarine bottoming and minewarfare operations, and for fisheries and resource development.
- The identification of mobile areas of seafloor. The presence of sand-waves and ripples are indications that the seafloor in a particular area is mobile. On major shipping routes such areas may require periodic re-survey to ensure safety of navigation.

2.3.3 Theoretical Considerations

2.3.3.1 The strength of the signal returned by a given feature is governed by several factors linked by an expression known as the "sonar equation" which may be used to determine whether a particular type of feature will or will not be detected. A good explanation of the terms involved in this equation is given in the 1981 FIG/IHO "Report on the Detection of Depth Anomalies". The standard textbook that should be consulted if a further study of this subject is required is "Principles of Underwater Sound" by R.J. Urick. It must be stressed that this equation can form only the starting point for a consideration of SSS performance. It ignores signal losses and other acoustic parameters, as well as the limitations of the towfish and the recorder.

2.3.3.2 Short range coverage. There is a region close to the towfish where gaps in the sonar cover may occur. These gaps need to be considered in two planes (see Figure 4.1):

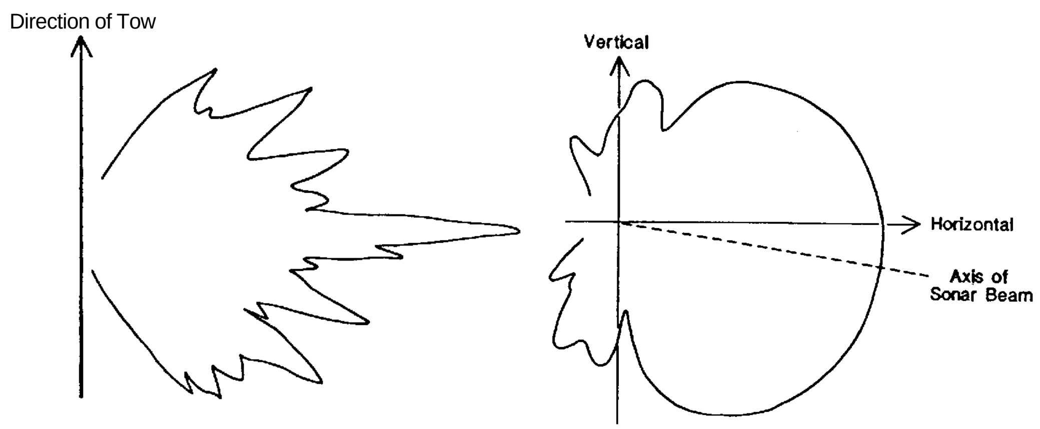


Fig. 4.1 "SSS Horizontal and Vertical Beam Coverage"

- The vertical plane. The main beam of the sonar has a width in the vertical plane of about 50°, with the beam axis tilted 10° downwards. There is, therefore, a region under the towfish which lies outside the main beam; the size of this region is governed by the height of the transducers off the seafloor. The original concept of this area not being ensonified at all is incorrect. Unless the fish is a long way off the seafloor this zone is covered by side lobes from the transducers, and parts will receive some sound energy from the fringes of the main beam. (The "edge" of a beam is usually taken as the half-power line, but this is not an absolute cut-off point and some energy exists outside it). Whilst a gap in the record under the fish does occur, it is considerably smaller than originally thought and may only be a few metres in extent. Nevertheless, this gap must be covered by sonar from the adjacent lines.

- The horizontal plane. There is an area close to the towfish (the "near field") where the sound pulses have parallel edges. As a result, gaps may occur between individual pulses of sound. The gap between pulses in the near field is a function of ship speed and pulse repetition rate. Beyond this area, the spreading of the beams closes the gaps to give total coverage. Small contacts are therefore likely to be missed close to the fish rather than further away from it.

2.3.3.3 Planning Area Searches. Two different methods of planning area searches can be used:

- Detecting contacts close to the towfish. The search is planned so that the smallest required contact can be detected close to the towfish. The limiting case requires such contacts in the near-field of the sonar beam to receive five pulses; outside this area, beam expansion ensures they will receive at least five pulses.

- Detecting contacts further away from towfish. The zone where small contacts may not be detected can be calculated for a given range scale in use and speed over the ground. Line spacing can then be adjusted so that sweeps from adjacent lines at least cover the gap. Alternatively, line spacing can be fixed and speed adjusted to ensure that full coverage is achieved. Thus with a range scale of 150 m in use and at a speed at which small contacts may not be detected within the first 25 m, line spacing must not be more than 125 m.

2.3.3.4 The second of the above methods is usually employed on area searches as it allows a faster speed of advance. For a line-spacing of 125 m using the 150 m range-scale, one metre contacts will be detected in the near field at a speed of 3.6 knots. Relying on detecting them from adjacent lines allows a speed increase to seven knots. Details of the calculation follow (see 'Target Detection' and 'Calculation of Speed of Advance').

2.3.3.5 Confirming SSS Performance. Whilst these calculations will provide theoretical capabilities it is essential that a SSS's performance is confirmed in the field prior to use. This is achieved by selecting a suitable feature, reflecting the type and size of feature required to be detected during the survey, and towing the SSS past it. Both sonar channels, i.e. both sides, and each range scale should be tested to determine the maximum detection range.

2.3.3.6 Position of the Side Scan Towfish. Towing the sonar transducers astern of the vessel has several advantages including removing the sensor from the effects of vessel motion and operating it at a height above the seafloor which will enable the optimum shadow. However, there is a disadvantage in that it also introduces uncertainty as to the position of the towfish. This error has three components:

- an along-track component, caused by uncertainty in how far the fish is astern of the vessel; this depends on the length of cable out, depth of towfish and vertical catenary of the cable (the last two also vary with the ship's speed);

- an across-track component, caused by deflection of the fish by tidal stream or current, and by ship manoeuvres;

- errors in the position of the ship or boat, which will be transferred to the towfish.

2.3.3.7 Towfish position can be determined using an ultra short baseline (USBL) positioning system which requires transducers/receivers to be fitted in the vessel and towfish; however the accuracy of this system deteriorates rapidly depending on the length of tow. An alternative method, under development in Australia (2004), utilises the direction and angle of depression of the tow cable over the stern of the vessel, together with a model of the catenary of the tow cable to predict, quite accurately, the towfish position.

2.3.3.8 In addition, the attitude of the towfish may vary both longitudinally and about its axis and thus the direction of the transducer beams may fluctuate. This is especially true if the ship's course or speed are frequently changing and emphasises the need for generous overlaps during sonar sweeping. Planning to theoretical limits of performance is almost certain to lead to gaps in the sweep in reality.

2.3.3.9 Hull Mounting. SSS can be mounted in the hull of a surface vessel. The advantages of this are that its position, and hence orientation, are accurately known and therefore the positioning of detected features is relatively easy. Hull mounting also enables freedom of manoeuvre for the vessel which is no longer required to tow the sensor. However there are a number of disadvantages to hull mounting including the effect of vessel motion on side scan ensonification and performance, possible mutual interference with other hull mounted sensors, i.e. MBES, and the fact that it is unlikely that the SSS will be operated at the optimum height above the sea floor. Hull mounting is often the best method when operating in shallow water or in areas where the seafloor topography is potential hazardous, reef strewn for example. Otherwise, the disadvantages of hull mounting would normally outweigh the advantages.

2.3.4 Operational Constraints

2.3.4.1 Hydrodynamic Stability of the Towfish. Under most conditions the towfish is largely decoupled from the effects of ship's motion by the flexibility of the tow-cable. The assumption is usually made that the fish is completely stable in roll, pitch and yaw, although some motion in all these planes undoubtedly occurs. Roll probably has relatively little effect on the sonar picture, being compensated for by the wide beam angle in the vertical plane. A permanent list in one direction, which may be caused by a distorted fin or a twist in the cable, can however markedly decrease performance. This should be suspected if one channel gives a different quality of picture to the other.

2.3.4.2 In extreme cases it may be necessary to rely only on the "good" channel and allow for this in planning survey lines. Pitch and yaw are more significant; with such a narrow beam-width in the horizontal plane, these motions could decrease detection probabilities of small features. A feature that would receive at least five pulses with a stable fish may only receive three or four if the fish is oscillating in either of these directions.

2.3.4.3 The problem of towfish stability is believed to be less important than that of fish position. In rough weather the effects of towfish oscillation can usually be clearly seen on the trace. Under these conditions the reduction in the probability of detecting small features must be considered. With the increasing use of heave compensators and motion sensors for echo sounders and the greater importance attached to detecting small contacts, sonar conditions rather than echo sounder performance may be the limiting factor for effective surveying.

2.3.4.4 Height of Towfish. For most work the optimum height of the towfish above the seafloor is 10% of the range-scale in use, i.e. on the 150 m scale the fish should be 15 m above the seafloor. Side scan transducers are directed slightly downwards; flying the fish too close to the seafloor may reduce the range from which returns can be received. If the fish is too high acoustic shadows may not be formed behind obstructions making them more difficult to detect; this is especially true in deep water when a compromise has to be made between the need for getting the fish down to a useful depth and maintaining a reasonable speed of advance.

2.3.4.5 In areas of very high seafloor relief it may be prudent to tow the sonar higher than normal; in this event the reduction in acoustic shadow on features standing proud of the seafloor must be borne in mind. This effect is worst close in to the towfish where detection of small contacts is already at its most difficult.

2.3.4.6 In shallow water it may not be possible to get the fish as high off the seafloor as desirable. Although the recorder will be giving a background trace across the entire width of the paper, the sonar beam may not be ensonifying the entire range. Under these conditions the only solution is to reduce both the range scale and the line spacing.

2.3.4.7 As a further limitation in shallow water the transducers may be very close to the surface with little tow-cable streamed. This will introduce the problem of surface noise (such as waves and ships wake) degrading performance and may also lead to the towfish being adversely affected by the motion of the ship. The effects of water layers and thermoclines on side scan can usually be ignored; they have very little effect on the range at the frequencies used.

2.3.4.8 When investigating contacts with sonar the towfish should always be sufficiently high above the seafloor to allow it to pass over the obstruction in the event of an accidental "on top". The least depth over a feature can usually be estimated initially from the shadow length obtained during the area search.

2.3.4.9 If it becomes necessary to tow the fish at a height other than the optimum, a confidence check should always be carried out to confirm the system continues to meet detection and other requirements. Towfish height can easily be controlled by a combination of wire out and ship's speed. Quickly heaving in a length of cable will "snatch" the fish upwards rapidly, after which it will settle back down more slowly. This technique can be very useful in lifting the fish over unexpected dangers. As the length of wire streamed increases this method becomes less effective.

2.3.4.10 Depressors. Some side scan fish can be equipped with depressors which will drive the fish deeper for any given length of tow cable or speed of advance. Whilst this can reduce the length of tow required there are a number of disadvantages to using depressors:

- they increase strain on the cable resulting in the requirement for a more powerful winch if scope is to be adjusted underway; manual operations can become impracticable;

- the shorter scope of cable results in the transmission of ship movement down to the fish;

- they can reduce the effect of an increase in speed and/or reduction in scope of tow cable on the fish height, thus negating the use of this technique to overcome unexpected dangers.

2.3.4.11 When operating in close proximity to the sea floor it is prudent to ensure the towfish is fitted with a trip mechanism that enables it to flip over and still be retrieved after a strike. In this case it is possible the fins will be lost but at least the towfish itself is recovered.

2.3.4.12 Direction of tow. In normal circumstances SSS should be towed into and out-of the predominant tidal stream/current in order to minimise their effect on the towfish in the form of across track positional errors. Where tidal stream/current effects are not an issue the side scan should be towed parallel to the bathymetric contours. This minimises the requirement to have to continually adjust the scope of tow when steaming into and out-of shallow water.

2.3.4.13 However, there are exceptions to these rules. In sandwave areas, in particular, it may be necessary to tow the SSS at right angles to the axis of the sandwaves. This ensures that the side scan looks along the sandwave crests/troughs to avoid the possibility of shadow areas where features will not be detected, see Figure 4.2.

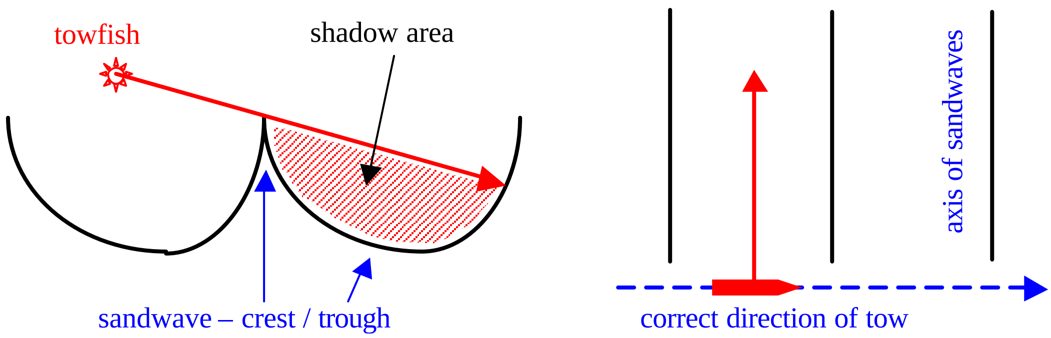


Fig. 4.2 "SSS – Potential Shadow Areas in Sandwaves and Correct Direction of Tow"

2.3.4.14 Effective Sonar Range. The presence of marks on the sonar trace does not necessarily indicate that returning echoes are being received. Transmission losses, interference from other sources of noise, water conditions and recorder limitations all restrict the useful range of SSS. A maximum range of 270 m is about all that can be expected for even large wrecks, with small contacts (1-2 m) unlikely to be detected beyond about 120-150 m. Detection range varies between different SSS models and frequencies - the higher the frequency the less the detection range, although the resulting picture may be better. The best results will usually be achieved by restricting the range scale to 150 m to take advantage of the higher pulse rates and greater definition. A short test using a suitable seafloor contact at varying ranges will usually provide information on sonar conditions in the survey area.

2.3.5 Distortions of Sonar Records

2.3.5.1 Sonographs never represent isometric maps of the seafloor. Various distorting factors have to be recognised when interpreting sonograph mosaics in map form, unless the distortions have been eliminated digitally before the mosaic has been compiled. The main causes of distortion are:

- compression of sonograph picture with speed increase - a distortion will occur parallel to the course made good due to variable ship speeds and constant paper feed speed, resulting usually in a compression of the record in this direction;

- the height of the fish above the seafloor will introduce a lateral distortion perpendicular to the direction of travel;

- a sloping seafloor will introduce distortions perpendicular to the direction of travel which are different on the up-slope and down-slope sides.

2.3.5.2 For a given ship's speed, range scale, paper speed and towfish height, the distortions can be calculated. During area sweeps these effects generally only need to be considered when plotting contacts; during investigations they need to be considered in detail. Speed during investigations should be adjusted to give as little distortion as possible, about three knots is usually ideal.

2.3.5.3 Lloyd Mirror Effect. During sonar operations in very calm conditions reflection of some of the sonar energy can occur from the sea-surface, as shown in Figure 4.3. This is known as the Lloyd Mirror effect and results in a series of maxima and minima in the sonar image. This effect normally occurs only when the fish is close to the surface and can be minimised by towing the fish deeper.

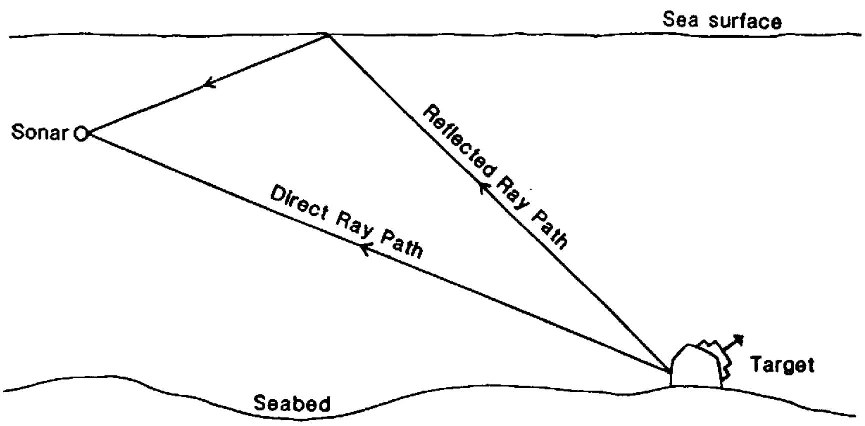


Fig. 4.3 "Lloyd Mirror Effect"

2.3.5.4 Cross Talk. Cross talk between two SSS channels can result in a mirror image of sea floor features from one channel being displayed on the opposite channel, albeit usually fainter. Cross talk can result in the true image on the effected side being obscured. This may prevent detection of features or to the erroneous 'detection' of what are, in effect, copies of real features from the opposite side. This can be a particular problem in areas where there are numerous features in which case it can be difficult to verify what is real and what is not.

2.3.5.5 Tilt Effect. If the side scan fish is not being towed level, in other words it is tilted to one side, the channel that is facing downwards towards the sea floor will result in a stronger return signal and therefore a darker image; on the other hand the channel that is facing upwards will result in a lighter image. Seafloor classification is based on interpreting the image shading, a result of the relative strength of the return signal from different seafloor types. The tilt effect can therefore result in difficult or even erroneous interpretation.

2.3.5.6 Automatic Gain Control (AGC). AGC was introduced as a means of ensuring the SSS image was optimised for feature detection. In other words in areas of strong return, such as rock, the gain was automatically decreased to enable features to be detected against a 'light' background. However, as with the tilt effect, altering the gain and hence the resulting image shading, also renders seafloor classification difficult, if not impossible. For this reason AGC should be turned off if the sonar image is to be used for seafloor classification.

2.3.5.7 Wash and Wake. If the SSS is towed too close to the surface the image can be affected by returns from the wash or wake of other vessels or even the towing vessel itself if it has recently made a turn. Again, such interference can seriously impact seafloor classification and it is important that a sonar log is maintained so that such incidents can be recorded to assist subsequent image interpretation.

2.3.5.8 Thermocline. As with any sonar, side scan transmissions are subject to the effects of their passing through water with changing properties and which may result in distortion of the image. Whilst software can be used to 'mould' the image back into shape, it is the important for the surveyor to know, and hence the degree of sonar ensonification which is used to overcome this problem. For instance, in areas significant to navigation, a higher level of ensonification redundancy may be required with adjacent lines run in opposite direction and possibly additional lines at right angles, with a short range scale selected. In less important areas the range scale employed may be greater and the degree of overlap and redundancy less and therefore distortion can become more of a problem.

2.3.5.9 "Sound Underwater Images - A Guide to the Generation and Interpretation of Side Scan Sonar Data" (Fish JP & Carr HA, 1990) is an example of a reference text that may be used to assist sonar interpretation.

2.3.6 Feature Detection

2.3.6.1 The following assumptions are made:

- feature size is defined as the length presented normal to the sonar beam;
- the minimum number of returns to make a discernible mark on the trace is taken as five;
- sound velocity is assumed to be 1500 m/sec;
- beam angle of the sonar is 1.5°.

2.3.6.2 Terms and Units:

pulse interval -	t	seconds
pulse repetition interval -	F	pulses per second
ship's speed (over ground) -	V	metres per second
feature length -	L	metres
velocity of sound in seawater -	C	metres per second
recorder range scale -	Rm	metres
beam width -	Bw	metres
slant range to contact -	Rs	metres
length of array -	l	metres
distance travelled between pulses -	d	metres

2.3.6.3 Basic Equations:

$F = \frac{C}{2Rm}$ pulses per second; or, $t = \frac{1}{F}$ seconds

Because ϕ is a very small angle, beam width at a given range (Bw) = $Rs \cdot \phi$

2.3.6.4 It can be seen from Figure 4.4 that Feature A is the largest feature that CANNOT receive five pings; it can receive a maximum of four (i.e. pings 2, 3 and 4 and either ping 1 or 5). However, theoretically, a small increase in Feature A's length would mean that it received five pings; in general, for N pulses its length is given by:

$$L = V \cdot t \cdot (N - 1) - Bw \qquad \textbf{(4.1)}$$

2.3.6.5 Feature B is the smallest feature that MUST (theoretically) receive five pings; it is caught by the first and just missed by the sixth. Its length is given by:

$$L = V \cdot t \cdot N - Bw \qquad \textbf{(4.2)}$$

Essentially this is the same equation as used to determine speed whilst echo sounding. Both formulae assume that the sonar beam is divergent.

2.3.6.6 In general, equation (4.1) is used when determining either:

- the length of feature that will receive five pings at a given speed over the ground;
- the speed over the ground that cannot be exceeded if a feature of a given length is to receive five pings.

2.3.6.7 There may be occasions when the surveyor feels it more prudent to use equation (4.2) giving a greater probability of detection.

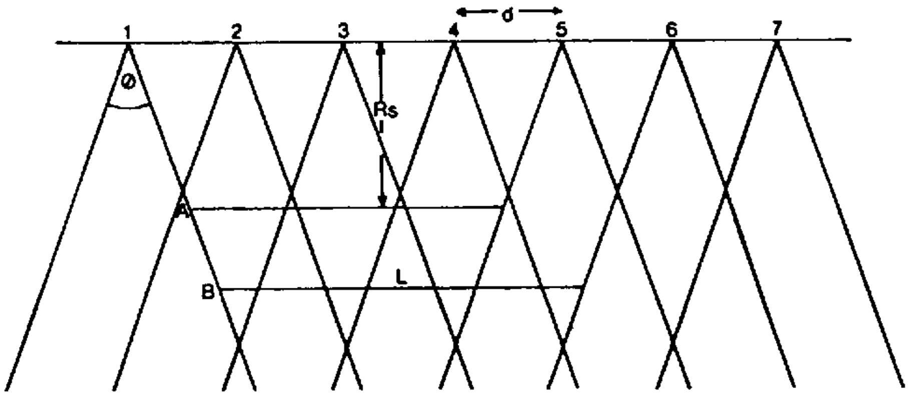


Fig. 4.4 "Diagram showing Feature Detection"

2.3.7 Calculation of Speed of Advance (SoA)

2.3.7.1 A typical survey scale is 1:25,000 in which case the usual spacing of lines is 125 m. In general, it is advantageous if bathymetry and sonar sweeping can be carried out at the same time. With lines 125 m apart a swathe 25 m either side of adjacent lines is ensonified, although this may be reduced with wayward line-keeping.

2.3.7.2 To recognise a feature on the SSS trace it is necessary to ensure it receives five pings. To identify it as a significant feature requires a confirmatory detection from another line. This does not mean that contacts not detected on adjacent lines may be discarded as spurious but that a small wreck at the outer edge of the side scan trace may easily be overlooked.

2.3.7.3 In an area sweep it is then necessary to determine the speed over the ground which must not be exceeded in order that a feature of one metre in length should receive five pings from two adjacent lines. This gives the Speed over the Ground (SoG) which should not be exceeded.

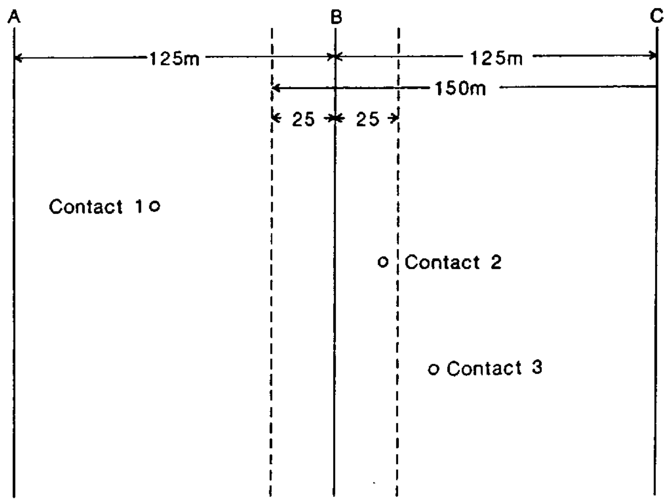


Fig. 4.5 "Calculating Speed of Advance"

2.3.7.4 In Figure 4.5 A, B and C are three lines spaced 125 m apart. A survey vessel is operating its side scan on the 0-150 m range scale. What criteria must be satisfied?

2.3.7.5 Near Field. The near field limit is usually within 20 m. Therefore with a 25 m overlap from adjacent lines a feature which would not have received five pings at a given range in the near field on line B will get five pings from both lines A and C. In this case the near field detection speed of 3.6 knots is not a limiting factor.

2.3.7.6 Far Field. Contact 1 should be detected from lines A and B, Contact 2 will get five pings from lines A and C, Contact 3 from lines B and C. It is necessary to calculate the speed over the ground that must not be exceeded if a contact of length L m is to get five pings at 25 m.

If L = 3.0 m then:

From equation (4.1) the maximum length of feature that will not get five pings is:

$$L = V \cdot t \cdot (N - 1) - Bw$$

where $Bw = 25.0 \cdot \phi$

$N = 5$

$t = 0.2$ sec

$L = 2.999$ m (see Note)

Note: because theoretically a slightly longer feature, i.e. 3.0 m, should get five pings.

rearranging: $$V = \frac{L + Bw}{(N - 1) \cdot t}$$

$$= \frac{2.999 + 0.6545}{4 \cdot (0.2)}$$

$$= 4.57 \text{ m/sec or } 8.9 \text{ knots}$$

2.3.7.7 In fact for practical reasons the fish should not be towed at speeds over the ground in excess of eight knots, or small features will be missed, or 10 knots through the water since above this speed the towfish is liable to yaw. Note also that if five pings to a feature are to be "guaranteed" then equation (4.2) should be used giving a V of 3.65 m/sec or 7.1 knots.

2.3.7.8 If the requirement is to detect features 1.0 m in length from two lines then:

$$V = \frac{0.999 + 0.6545}{4 \cdot (0.2)}$$

$$= 2.067 \text{ m/sec or } 4.0 \text{ knots}$$

2.3.7.9 However if five pings into a one metre feature from one line only are to be required then:

$$V = \frac{0.999 + (72.5 \cdot Bw)}{4 \cdot (0.2)}$$

$$= 3.623 \text{ m/sec or } 7.0 \text{ knots}$$

2.3.7.10 The danger with using the last of the above equations is that the probability of detection of a small feature in a "one chance only" situation is low.

2.3.7.11 "Fast" SSS. As technology evolves some SSS are able to be operated at faster speeds over the ground than was previously possible. An example is the Klein 5000 series, which employs beam steering and focussing techniques simultaneously generating several adjacent, parallel beams per side. This "multibeam" design permits higher towing speeds whilst providing high resolution imagery. Other SSS developments include the use of interferometric, multi-pulse and synthetic aperture techniques. However, as with all such sensors, it is essential that its performance is validated against known features, which represent features required to be detected. Validation should be followed up by initial and regular repeat confidence checks in the survey area.

2.3.8 Track-Keeping Errors

2.3.8.1 A question that needs to be addressed is how far off track can the survey vessel go before a gap in coverage is created? Assuming only one detection (five pings) is required to a 1.0 m feature, a standard 1:25,000 survey is being undertaken with lines 125 m apart and range scale 0-150 m selected, then overlap is 25 m. The sum of any errors must be contained within this figure. For example:

fish position	e1	10 m
vessel navigation	e2	5 m
slope effect	e3	1 m
sound velocity variations	e4	1.5 m
therefore	? e^2 =	128.25 m
total error RMS	E =	11.3 m

2.3.8.2 Overlap is 25 m, however only 24 m is useable (the contact has to paint) therefore total allowable track error $= \sqrt{[24^2 - \Sigma e^2]} = 21$ m

2.3.8.3 This assumes that a feature is detectable at 149 m where it will paint as a black dot 0.8 mm by 0.8 mm with a 1 mm shadow (that is if the shadow is not obliterated by the 150 m range line). A more prudent off track allowance would be 15 m; this plots as 0.6 mm at a scale of 1:25 000.

2.3.9 Practical use of Side Scan Sonar

2.3.9.1 Area Sweep is the name given to the standard hydrographic sonar search method. The categories of sonar sweep required for any given survey will be specified in the survey instructions. An example of categories of SSS search is as follows:

Category A and B. Search in one direction and/or its reciprocal using SSS. Where practicable, adjacent lines are to be run in opposite directions. Searches for all listed wrecks are to be conducted. Examples of sonar line spacing, range scale, overlap to be achieved and maximum speed over the ground to be used are given at Table 4.4.

Category A sweeps are intended to be the standard sweeps for coastal and inshore areas not subject to routine re-survey. These sweeps are designed to achieve a theoretical seafloor ensonification of 240%, i.e. [2 x effective sonar range/line spacing] x 100 = % ensonification. Category B sweeps achieve a theoretical seafloor ensonification of 133% and may be used for routine re-surveys and in depths greater than 100 m where detection of all features is less critical.

Category C. Only searches for listed wrecks are to be conducted.

Category D. Special searches as ordered. This includes special instructions for use of particular side scan and hull mounted sonars etc.

Category	Type of Survey	Sonar Line Spacing	Sonar Range Scale	Max SoG	Adjacent Line Overlap[1]
A1	Special	125 m	150 m	6 kn	25 m
A2	inshore & coastal surveys at >1:25,000 in depths <15 m	62.5 m	75 m	8 kn[2]	12.5 m
	inshore & coastal surveys at >1:25,000 in depths <50 m	125 m	150 m		25 m
	shelf surveys in depths >50 m and/or scale <1:25,000	250 m	300 m		50 m
B1	routine re-surveys	250 m	150 m		50 m
B2	shelf survey scale <1:25 000 in depths >100 m	500 m	300 m		100 m

Table 4.4 - SSS Search - Categories A and B - Example Criteria

Notes:

1. The overlap under adjacent lines is to allow for limited wayward line-keeping and positional inaccuracies. If the surveyor considers positional inaccuracies and/or wayward line-keeping exceed this figure then he should adjust the line spacing or range scale, with subsequent speed adjustments, as necessary.

2. See previous comments with regard to use of “fast” SSS which may enable these speeds to be increased.

2.3.9.2 It is emphasised that these reflect minimum standards; if in doubt over sonar performance, line spacing should be tightened or speed reduced.

2.3.9.3 The use of a regular series of parallel straight lines remains the most efficient way of covering a survey area. The line direction will be close to the direction of the tidal stream to minimise towfish offset. The line spacing for the sonar lines is determined by the range scale in use and the overlap required. It is recommended that the overlap between adjacent swaths is 125%.

2.3.9.4 For military surveys on the continental shelf in water depths less than 200 m, the requirement is often to detect all contacts larger than one metre in extent. With existing equipment this cannot easily be achieved and a compromise between the requirements of sonar and bathymetry must be reached. A sonar sweep which will detect one metre contacts in depths less than 140 m provides this compromise. For the normal scale of l:25,000, this means a line spacing of 125 m, sonar range scale of 150 m and a speed over the ground no faster than 7 knots. Existing equipment cannot effectively be deployed deeper than 150 m and, in water between 150 and 200 m depth, the search will be restricted to locating large wrecks and obstructions.

2.3.9.5 Autonomous Underwater Vehicles (AUV). The employment of AUV equipped with SSS and MBES is becoming increasingly common. These platforms enable sensors to be operated at great depth and at the appropriate altitude above the seafloor. Thus it is likely that small features will be capable of detection at greater depths than is currently possible when employing surface vessel mounted or towed sensors.

2.3.9.6 Sonar sweeps should always be undertaken with lines orientated as closely as possible parallel to the main tidal flow in the survey area. The cross-track errors in the position of the towfish are invariably greater than those along the track and every effort should be made to minimise them. At a speed of 6 knots with 400 m of wire out and a tidal stream of 2 knots, a difference of 10° between tidal flow and line direction can offset the fish 17 m from the line.

2.3.9.7 The running of an extra sonar line immediately outside each edge of the survey area is necessary to ensure that the ordered category of sweep continues to the limit of the area. Similarly, care must be taken to ensure that the sonar fish has cleared the edge of the survey area before a survey line is ended.

2.3.9.8 It must be remembered that speed and feature detection probabilities calculated here are theoretical and take no account of adverse sonar conditions and equipment failings.

2.3.9.9 Plotting of Contacts. The detection of seafloor contacts between survey lines is one of the main reasons for using SSS. The ultimate use of the information must always be considered when deciding which contacts to plot; for example, submarines will not take the ground in areas of rough seafloor and minewarfare operations will usually be selected to avoid them. In areas of smooth seafloor the aim must always be to detect and plot every contact; in more rugged areas this standard will have to be relaxed. All such contacts must be plotted and allocated a contact number which will ultimately be included in the seafloor classification model.

2.3.9.10 Various techniques have been developed to plot contacts from manuscript side scan records; all attempt to reduce the errors in the contact position caused by errors in towfish position and orientation. Different techniques are to be used for contacts plotted from area searches, investigations and examinations:

- Contacts from area searches are usually plotted from two directions 180° apart. The standard "layback and offset" method should be used, with the mean of the two positions adopted as the most likely position.

- Investigations should produce a minimum of two pairs of passes for each contact at right-angles to each other, orientated in such a way as to fix the extremities.

- When a contact is examined by echo-sounder, the best "on top" position is to be used in preference to any side scan derived one, where possible an echo sounder line should pass the length of the long axis of the contact.

2.3.9.11 Measurements by Sonar. A good "beam-on" side scan picture of a wreck or obstruction can usually be used to estimate its height above the seafloor using the sonar "shadow". Although not accurate enough for charting purposes, this height is very useful for the safety of both ship and towfish when planning investigations. Estimates of the beam and length of a wreck can also be obtained from the sonar trace. The following points should always be considered:

- when estimating heights from sonar shadows the presence of higher parts of the wreck (such as masts), which do not throw a detectable shadow, should always be borne in mind;

- shadow heights must be measured from both sides of the wreck and the results meaned, which helps to correct for errors introduced by seafloor slope (it should be noted that heights obtained in the near nadir area by this method may be overestimated by up to 20%);

- measurements for length and breadth should always be taken perpendicular to the towfish track and must always be corrected for slant range distortions.

2.3.9.12 Conduct of Investigations. Investigations (or examinations) are conducted to improve the classification of a contact located during an area search. The following technique is recommended:

- relocate the contact by SSS, aiming to pass 50-100 m from it; this will normally be sufficient to eliminate ephemeral contacts;

- verify and/or improve its position;

- conduct the examination.

2.3.9.13 The 150 m scale is usually best (use of the 75 m scale may result in the shadow from a large contact extending off the trace). Speed should be kept to about 3 knots, to reduce distortions in the record, with the fish about 15 m clear of the seafloor. Providing good pictures are obtained, four runs (comprising two perpendicular pairs) should be sufficient. In the case of wrecks, one pair of tracks should be parallel to the long axis of the wreck and one pair perpendicular to it.

2.3.9.14 The above procedure will usually give sufficient data to determine whether an echo sounder examination is required and also will allow measurements of length, beam and height to be made. The side scan should always be recovered before close sounding. If several contacts which need sonar examination exist in the same general area, time can usually be saved by examining the whole group with sonar before recovering the sonar and obtaining a least depth by echo sounder.

2.3.9.15 Disproving Searches. Charted wrecks, obstructions or other dangerous features which have not been located and examined during a survey must be disproved if possible. They will not be removed from the chart without a positive statement from the surveyor in charge that this is justified and why. The procedure for conducting a disproving search is outlined below:

- Features whose positions have been previously established but which cannot be found during the survey need a very detailed investigation to disprove them. Such searches are to include a sonar sweep in two directions at right angles to each other and a close echo sounder search over a radius of between 0.5 and 2.5 NM from the charted position. Consideration might also be given to undertaking a wire sweep.

- When searching for an feature whose position is only known approximately [usually a (PA) wreck], the sonar search should also be undertaken in two directions at right angles and consideration should be given to extending the search over a radius of at least 2.5 NM, a distance based on the statistical probability of such a search being successful. However, if the surveyor is confident that the initial area search in one direction was entirely thorough, and that the sonar equipment was operating satisfactorily, he may consider that a second search in another direction is not necessary, having regard to the size and history of the wreck concerned and the position in which it is alleged to lie. If, during the initial sonar sweep, a magnetometer was also deployed and no marked magnetic anomaly was detected within 2.5 NM of the charted position, this may be accepted as additional evidence that a wreck with a predominantly ferrous content does not exist in the area.

- Searches for wrecks not within a regular survey area must be extended to a radius of at least 2.5 NM. Whether there is need to carry out a second sweep at right angles to the first will depend on the same considerations as above.

2.3.9.16 Whatever the outcome of such searches, whether as part of a larger survey or as individual examinations, the surveyor must report the findings in full, in an appropriate manner and with supporting records as necessary, together with a positive recommendation as to future charting action.

2.3.10 Positions Errors of Sonar Contacts

2.3.10.1 During normal area surveys the surveyor's primary concern is to attempt to ensonify the entire seafloor in order to detect any significant feature. Any features of significant size will then usually be accurately fixed by echo sounder.

2.3.10.2 However in some special surveys it is essential that as precise a position as possible is given for each contact, particularly for small seafloor contacts. These will not necessarily be fixed by echo sounder. It is thus necessary to consider all the errors accruing in the plotting of a contact from SSS trace.

2.3.10.3 Uncertainties in the position of a contact will derive from the following (i.e. 1 sigma)(±):

uncertainty in vessel position -	5 m
uncertainty in towfish position (see Note) -	10 m
variations due to assumed SV (1500 m/sec) -	1.5 m
resolution of paper trace. (0.75% range scale) -	0.75 m

errors due to seafloor slope -	1 m
therefore, total error (RMS) (1 sigma) =	11.4 m

Note: This can be an unknown quantity depending on use of a precision towfish tracking system. Evidence suggests that the fish can oscillate 20 m about the towin g vessels track. The value is also dependant on the depth and length of tow cable. An estimate of ±10 m is therefore assumed.

2.3.10.4 The values given above are examples only and the list is not exhaustive. The surveyor should consider the table of errors for each part of his survey and comment on them in the Report of Survey, as is the case with echo sounder errors.

2.3.10.5 Uncertainty in the position of the towfish is the greatest potential source of error. Unless a method of accurately positioning the towfish is employed surveyors should make every effort to minimise the offsets by planning tracks parallel to the prevailing tidal stream or current. If this is not possible every opportunity must be taken to quantify the offset of the towfish to the track by reference to seafloor features whose positions are known. If there is any risk that full ensonification is not being achieved, the simplest solution is to close up the sonar lines, accepting that this will result in a reduction in rate of effort.

2.3.11 Plotting and Measurements from Sonar Records

2.3.11.1 Layback. Layback is the distance astern of the navaid position that the towfish is assumed to be (see Figure 4.6). In the normal course it can be computed as follows:

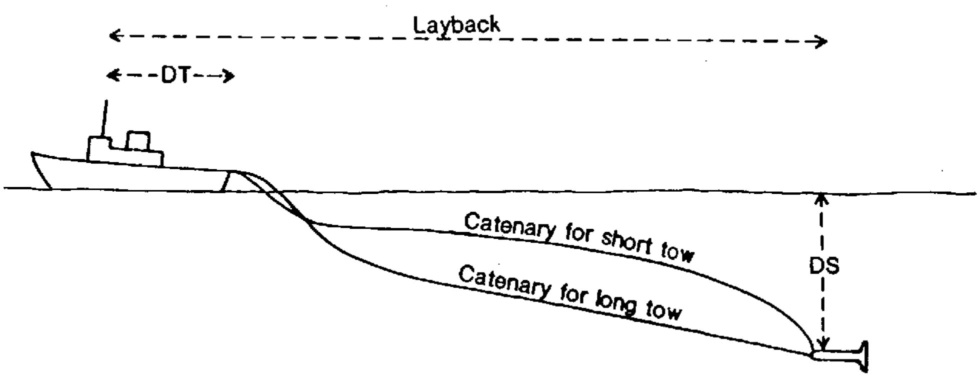


Fig. 4.6 "Side scan Layback"

Note: When the wire out exceeds 100 m, the bight of wire has a greater effect on the tow than the hydrodynamic properties of the towfish.

Layback = $DT + \sqrt{[WO^2 - DS^2]}$

where:	DT =	horizontal distance from fix point to tow point;
	WO =	amount of wire out from tow point;
	DS =	depth of towfish below surface.

2.3.11.2 This assumes that the wire takes a straight line path from the tow point to the towfish. Obviously this is a simplification; the wire is actually in an irregular catenary in both horizontal and vertical planes.

2.3.11.3 Correction for Slant Range. Slant range may be corrected to horizontal range simply by use of Pythagoras' theorem. If the seafloor is sloping then a correction factor will have to be applied.

2.3.11.4 Geometry of Heighting from SSS. One of the most important capabilities of SSS is its ability to enable the height of a feature to be measured from the length of its shadow on the sonar trace. However, this capability depends on the SSS being operated at the correct height above the seafloor and selection of the optimum range scale. The geometry of heighting from SSS is shown at Figure 4.7.

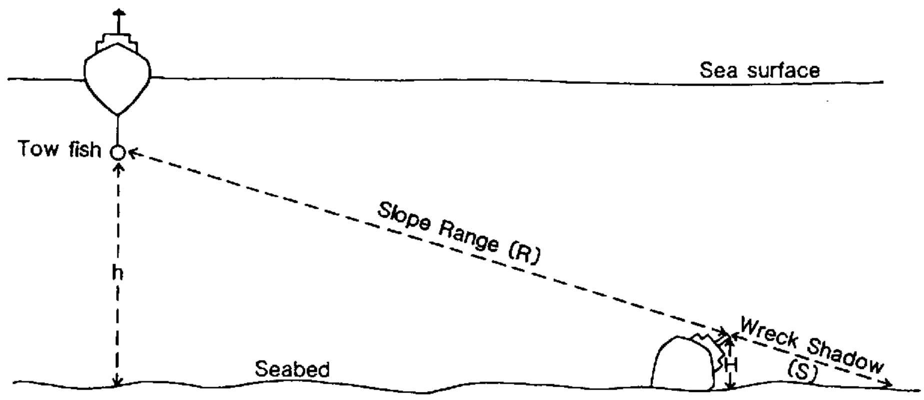


Fig. 4.7 "Heighting from SSS"

Therefore, by similar triangles - $$H = \frac{S \cdot h}{R + S}$$

Where:	H =	height of the feature
	S =	length of feature shadow
	R =	slope range
	h =	height of towfish above seafloor

2.3.12 Multibeam Echo Sounders (MBES)

2.3.12.1 For bathymetry the MBES has quickly proven its superior capabilities allowing it to provide (in theory) 100% ensonification of the seafloor whilst meeting IHO specifications for bathymetry. The fact than a MBES transducer is rigidly mounted to the hull of the survey vessel means that its position may be calculated as accurately as that of the positioning system in use. Coupled with the capability of forming discrete beams, the MBES is the tool of choice for bathymetric surveys.

2.3.12.2 Given a MBES's positional capabilities, subsequent passes over the same stationary feature should yield exactly the same geo-referenced position. The small difference, if any, in the contact's position is of great advantage when looking for features which may be revisited for purposes of in-situ identification either by ROV or diver. Unfortunately, however, the fixed transducer results in broad grazing angles which are not conducive to real time feature detection using the same shadow-casting principles of the SSS. Detection, therefore, must focus on variations in the resultant bathymetry caused by a feature on the seafloor.

2.3.12.3 Survey Methods. The requirements for a multibeam sonar survey where SSS is towed simultaneously are similar to the requirements for a traditional SBES. The use of a regular series of parallel straight lines remains the most efficient way of covering a survey area. The line direction will probably be determined by the side scan requirement in that the direction is close to the direction of the tidal stream. One difference with the multibeam is that since the system collects data in a matrix that is as dense along the line as athwartships, there is no requirement to cross the contours at right angles to determine their position accurately.

2.3.12.4 Line spacing for the sonar lines is determined as usual by the range scale in use and the overlap required. The difference here is that almost certainly 100% coverage will be specified for bathymetry as well. In shallow depths (i.e. under 30 m), the line spacing required to achieve 100% bathymetric coverage with the multibeam may be less than that required for side scan. It will be for the surveyor to determine if it is more efficient to complete the side scan coverage as normal, and then to run interlines using multibeam alone where required or to complete the multibeam coverage on the first pass.

2.3.12.5 Where multibeam determines the line spacing, the required spacing will depend on the average and minimum depths in an area. The multibeam swath width is depth dependant. Where the depth varies significantly over the survey area, it may be more efficient to split the complete area into subsections and to run each subsection at a line spacing appropriate to its depth. Current recommendations are to achieve an average overlap between adjacent swaths of 25% with a minimum overlap of 10%.

2.3.12.6 Where multibeam alone determines the line direction for a survey, and where the sound velocity profile throughout an area is similar, then the most efficient line direction is parallel to the depth contour lines. In this way, the swath width and the overlap between adjacent swaths will be more even and the line spacing can be wider.

2.3.13 Considerations whenusing MBES

2.3.13.1 Despite early predictions and manufacturer's claims, the detection of small and potentially hazardous features by MBES cannot be taken for granted. For instance, even if, say, the mast of a wreck is "pinged" by MBES, built in noise reduction algorithms will likely eliminate the feature; whilst turning such filters down or off would introduce so much noise as to make the data unusable.

2.3.13.2 Another fundamental factor is MBES beam geometry. The various makes and models are of different design and, in some instances, leave relatively large gaps that are not ensonified between beams. Interferometric MBES, for example, can suffer from poor feature detection in the nadir area due, simply, to the physics of that type of system.

2.3.13.3 Surveyors must verify the performance of a MBES before it is employed for feature detection; including determination of an appropriate swath width, ping rate, speed over ground etc. Many agencies responsible for nautical charting still require the use of SSS for feature detection, with MBES providing bathymetry and as a check on SSS feature detection. MBES beam geometry and feature detection potential is discussed in detail at "How Effectively Have You Covered

Your Bottom?" - Miller JE, Hughes Clarke JE, & Paterson J - The Hydrographic Journal No. 83 January 1997.

2.3.14 Magnetometer

2.3.14.1 This instrument can prove very useful in differentiating wreck from rock if the wreck is ferrous. A brief outline of the theory of operation of magnetometers can be found in the 1981 FIG/IHO "Report on the Detection of Depth Anomalies".

2.3.14.2 Whenever possible, a magnetometer should be used during the basic sonar sweep, as this will provide additional evidence of the existence of ferrous material on or below the seafloor, although it cannot locate it precisely.

2.3.14.3 The intensity of the magnetic field from a ferrous feature falls off proportionally with the cube of the distance from the feature. A general formula for computing the change in field in nanoteslas (nT) to be expected as the magnetometer is displaced from the feature is:

$$M = \frac{50{,}000 \cdot W}{D^3}$$

where: M = change in field intensity in nT,

W = weight of ferrous metal in tonnes, and

D = distance of feature from detector in metres.

2.3.14.4 Generally, 5 nT is the smallest change of magnetic field intensity that can be reliably detected. Then, for a change in intensity of 5 nT, the equation above can be written to give:

$$D = \sqrt[3]{10\,000 \cdot W}$$

or for a series of features:

Feature	**Detection Range**
100 kg anchor -	10 m
1 tonne mine like feature -	22 m
2 tonne cannon -	27 m
10 tonne wreck -	46 m
100 tonne wreck -	100 m
1000 tonne wreck -	200 m

2.3.14.5 For example, during an area sweep with lines 125 m apart in a water depth of 50 m and with the magnetometer towing 3 m below the surface, from the above table it can be seen that:

- a 100 tonne ferrous wreck will probably be detected from at least one of a pair of adjacent lines and anything larger than 1000 tonnes should be detected on several lines;

- a 10 tonne ferrous wreck may just be detected directly below the magnetometer;

- anything smaller than 10 tonne is unlikely to be detected;

- a ship of about 1,000 tonne (ferrous metal) must tow the magnetometer 200 m astern or else tabulated detection ranges will be seriously degraded.

2.3.14.6 Many magnetometers are designed to be towed very close to the seafloor. This will increase the probability of detection of small ferrous features. However, care will have to be taken to prevent fouling the SSS cable, a danger less evident with a surface towed magnetometer.

2.3.15 Other Methods of Feature Detection

2.3.15.1 Other sensors with potential for feature detection include:

Singlebeam Echo Sounder (SBES). Not normally employed for feature detection in shallow water due to its relatively narrow beam width, which makes a full area search impracticable. SBES can be used as a check on MBES which have poor nadir feature detection performance and in deep water beyond the range of shallow water MBES. However, in all these instances use of SSS for feature detection should be considered.

Airborne Lidar Bathymetry. Contemporary ALB systems (i.e. LADS Mk. 2 and SHOALS) are capable of a full area search and of detecting features two metres square. This means they can meet IHO standards in clear waters suitable for ALB operations. Future development to further decrease spot size to enable detection of smaller features is expected.

Airborne Electromagnetic (AEM) Bathymetry. Originally designed for geophysical survey, AEM methods offer the potential for feature detection but this capability has yet to be demonstrated to IHO standards.

Forward Looking Sonars (FLS). Originally designed purely for navigation and collision avoidance, some recent FLS developments offer bathymetric and feature detection capabilities. To date, however, these capabilities have not been demonstrated as meeting IHO feature detection but they may achieve low order bathymetry standards. They cannot, therefore, currently be considered a stand-alone hydrographic survey sensor.

2.3.16 Obtaining Definitive Least Depth over a Feature

2.3.16.1 The surveyor must establish the least depth over wrecks and obstructions and the following guidance may assist in deciding upon the method of examination, i.e. obtaining the least depth. Whichever method is employed, the opinion of the surveyor as to the accuracy of the least depth obtained is of vital importance and must be stated in the Report of Survey. If a least depth is not achieved, the examination must still result in positive recommendations regarding the likely accuracy of the depth obtained and future charting action.

2.3.16.2 The horizontal and vertical accuracy of a least depth must reflect the accuracy criteria detailed for the survey as a whole and, in turn, those standards in IHO S-44 and/or S-57.

2.3.17 Echo sounder Least Depth

2.3.17.1 The least depth may be obtained by saturation SBES sounding. The required line spacing is to be calculated from knowledge of the echo sounder beam width and general depths in the area, allowing an overlap of at least 25% between lines. Attention is drawn to chapter 3, paragraph 3.5, with regard to calculating the area ensonified by a SBES.

2.3.17.2 Alternatively, MBES may enable the least depth to be obtained. However, as noted previously, if MBES is employed the surveyor must be certain that the system's capabilities are such that the definitive least depth is able to be determined. This is particularly the case if the least depth is over a mast or similar feature. Considerations here include the beam width and spacing, speed over ground, optimum part of the swath (i.e. nadir, inner or mid swath) to be placed over the feature, number and direction of passes required. It may be, however, that MBES is best employed to identify the boundary of a feature to enable a first-pass or, at least, a less extensive SBES examination to determine the least depth.

2.3.18 Use of Divers

2.3.18.1 An alternative is the use of divers, assuming visibility, strength of tidal stream and depth of the feature allow their employment. Where divers can be employed, ships should plan to allow sufficient time for the task to be completed safely and accurately. If depth gauges are used to determine depth, the accuracy of the gauges should be determined. The least depth over a feature can usually be obtained by divers in less than an hour, whereas a wire drift sweep can often take four hours or more.

2.3.18.2 In certain circumstances, the surveyor will be directed to use divers. If the least depth is likely to be less than 30 m, the use of a diver must be considered. If a wreck has been wire swept or investigated by diver within the last five years, its position is unchanged and echo sounder depths over it show no significant alteration, the use of divers should not be necessary.

2.3.18.3 Where general depths around the wreck are markedly different from those charted or when it is known that salvage/dispersal work has taken place since the last survey, the use of divers may be necessary.

2.3.18.4 If SSS traces indicate the vessel to be lying on its side or with its keel uppermost and several consistent echo sounder depths have been obtained, further investigation should not be necessary. However, if there is any possibility that there are projecting structures which may not have been revealed on sonar or echo sounder, then divers should be used.

2.3.18.5 Areas charted as 'foul', especially in an anchorage, need special consideration as seafloor movement may expose debris not previously considered hazardous; a diver's report is especially useful in these circumstances.

2.3.18.6 In areas of strong tidal stream and mobile seafloor, wreckage may shift and it is possible for the least depth over it to become markedly less. Wrecks in such areas should always be viewed with suspicion and, where other evidence suggests it to be necessary, diving should be carried out.

2.3.19 Other Methods

2.3.19.1 Other methods of obtaining the least depth over a feature include wire sweeping (see next paragraph) and the use of autonomous and remote vehicles equipped with suitable sensors. These, if nothing else, can be used to identify the shoalest point on a feature for subsequent measurement. These methods are not described in detail here.

2.3.20 Methods of Wire Sweeping Wrecks

2.3.20.1 In many cases the only positive means of establishing the least depth over a rock pinnacle or wreck is by use of a wire drift sweep. There are several methods:

2.3.20.2 Single Vessel Drift Sweep. This is a slow but accurate method which is, nevertheless, impossible if wind and tide are at right angles and difficult if opposed. Wire angles must be minimal and there must be no ahead or astern movement during drift. Surveyors using this method should beware of the gentle foul, of leaving gaps in swept path and of excessive wire angles.

2.3.20.3 The optimum situation for a single ship sweep:

- the wreck should be properly examined by echo sounder first;
- a marker buoy should be laid approximately one sweep width up tide of the wreck;
- angle of sweep to be less than 20°;
- no engines used, i.e. drifting;
- constant tension maintained on the sweep.

2.3.20.4 Two Vessel Drift Sweep. The procedure is similar to single vessel sweep. Considerations are:

- greater swept path than single vessel sweep (100-120 m maximum);
- need to know position of wing vessel;
- good vessel handling required;
- vessels to be stopped and drifting;
- sag (wire out) and lift (wire tension);
- greater tendency for vessels to roll;
- vessels will slowly pull together.

2.3.20.5 Accuracy factors include:

- sweep angle is caused by movement through the water and tension placed on wire sweep and must be kept to a minimum;

- wire sag is affected by weight of the wire and the width of the sweep;

- greater tendency for vessels to roll, hence less accuracy than single ship drift sweep.

2.3.20.6 Underway or Drag Sweep.

2.3.20.7 Accuracy factors are:

- the sag tends to disappear due to wire lifting on movement through the water;

- variable tension of wire and drag speed means uncertain angle of sweep.

2.3.20.8 Drift and drag sweeping are discussed in detail in the "Admiralty Manual of Hydrographic Surveying", Volume 2, UK Hydrographic Office, 1969.

2.4 Side Scan Sonar records

2.4.1.1 This section outlines records associated with SSS. The surveyor is to be scrupulous in confirming that there are no inconsistencies between any of the records.

2.4.1.2 Bridge records will vary from ship to ship, depending on the type of data logging equipment in use and preferences of the surveyor. However, it is recommended the following information should be available to the sonar interpreter:

- date and time;

- speed over ground;

- base course and course over ground;

- ship's head;

- wire out;

- remarks, including sea state.

2.4.1.3 Sonar Contact Book. This is the master record for all sonar contacts. Where applicable, it should contain the following for each record evaluated:

- sonar roll number and associated echo roll (or digital equivalents);

- dates and times;

- contact number;

- position details;

- port/starboard;

- slope range;

- layback;
- height of fish above seafloor;
- contact assessment, i.e. shadow, cross-talk, intensity, initial classification;
- further action required, i.e. investigate, interline, quick look, no further action (NFA) etc.;
- action complete with final classification and reference to associated wreck records if appropriate.

2.4.1.4 The sonograph (if applicable) must be marked up simultaneously with the echo sounder trace and should carry a comprehensive title. It should be remembered that the deck book and sonograph may become separated and there is merit in including sufficient information in the latter to enable it to stand alone for analysis and checking purposes.

2.4.2 Wreck Records

2.4.2.1 The accurate processing of wreck records is a time consuming task. The establishment of a fool-proof procedure at the outset will often save confusion and errors later. The position and details of individual wrecks may appear on several documents and great care is needed to ensure that these records are both consistent and correct.

2.4.2.2 The surveyor must ensure that the following activities take place:

- working records are logged and systematically stored;
- all contacts are investigated and examined in an orderly way;
- wreck reports are completed where needed;
- all wrecks are plotted on both working and fair records;
- all positions and details are consistent.

2.4.2.3 Wreck data may appear in the following fair records:

- air sheet (or digital equivalent) on completion;
- sonar track plot;
- seafloor texture tracing;
- annotated side scan and echo sounder traces (or digital equivalents, i.e. side scan contact thumbnails);
- the Report of Survey.

2.4.2.4 Positional accuracy of wrecks. The position of a wreck in all records must be consistent. The following procedure is recommended:

- select the best echo sounder “on top”; determine the navaid readings for that position, either from an “on top” fix or from the wreck investigation plot and convert this to latitude and longitude to provide the master position;

- record the position taken during the best echo sounder “on top”;

- plot the master position on the track plot, sonar contact plot, seafloor texture tracing and sounding tracing (as appropriate);

- record the master position in the Report of Survey.

2.4.2.5 The Fair Sheet (or equivalent) should show the position and least depth of each wreck located. If it has not been possible to examine it fully, a danger circle in red should be inserted with the legend “Wk(NFS)” - indication “not fully surveyed’. It is important that no depth should be inserted in the circle as this may be mistakenly treated as the least depth during subsequent processing.

2.4.2.6 The sonar tracing is to show the position of each wreck using the appropriate symbols contained in chart INT 1.

2.4.2.7 Each listed wreck or obstruction is to be accompanied by representative examples of echo sounder and SSS traces illustrating the feature (screen images, if the echo sounder does not have paper trace). Traces are to be annotated with the date/time of fixes bracketing the feature, the ship’s course and speed made good over the ground and, in the case of SSS traces, the ship’s true course and the distance of the sonar fish from the point of fix. The least depth obtained or calculated should also be inserted.

2.4.2.8 As much detail as possible is to be shown and should include the following:

- position in which the wreck was located, together with the horizontal datum of the survey;

- fix obtained - this is to indicate which corrections were applied;

- the least depth recorded, how it was obtained and whether the surveyor considers it to be definitive; if the charted depth is different, the surveyor should express his view as to the reason for the difference; if the height of the wreck has been calculated from SSS traces, it should be stated whether it is a mean of heights obtained from opposite directions;

- approximate dimensions and orientation, together with any evidence (i.e. diver’s report) about the wreck’s identity and condition;

- details of the tidal reduction used;

- general remarks, especially any correlation with other wrecks in the vicinity or listed; existence and depth of scour; general depths and nature of seafloor.

2.4.3 Sonar Coverage Records

2.4.3.1 Whenever sonar is used during a survey, a tracing at the same scale as the Fair Sheet (or digital equivalent) is to be prepared to show the following data:

- vessel's track whilst carrying out the sonar search;
- limits of the area searched by sonar;
- limits of areas closely examined (examination tracks need not be shown);
- positions and identifying numbers of all wrecks and features listed in the Report of Survey;
- positions and identifying numbers of all wrecks and obstructions located during the survey.

2.4.3.2 When a searchlight sonar has been used in conjunction with SSS, the tracing is also to include:

- areas of numerous echoes;
- all firm contacts and the direction in which they were obtained (ephemeral contacts should not be shown);
- classification and quality of these contacts and whether examined.

2.4.3.3 All positions of contacts and wrecks are to be carefully cross-checked with other tracings, forms and reports. The following symbols are to be used on sonar tracings:

wreck -	Wk
wreck, not fully surveyed -	Wk(NFS)
possible wreck -	Wk(U) (see Note)
bottom -	B
good sea floor contact -	g
fair sea floor contact -	f
swept wreck -	Wk

Note: where it has not been possible to confirm the identity of a contact as a wreck, but it is sufficiently strong to merit its classification as a 'possible wreck', the additional qualification of "(U)" (unexamined) should be used to indicate an inconclusive examination. "(U)" should also be used when a contact has not been examined at all. The classification of "Wk(U)" should result in a wreck report.

2.4.3.4 Ship's track and fixes. Where the ship's track for sonar operations differs from those of main sounding, sufficient fixes are to be identified and annotated on the tracing and should be abbreviated except for the ends of line.

2.4.3.5 Limits of area searched. Green line for SSS; red line for searchlight sonar and blue outline for areas of intensive search (with result in manuscript or reference to other record).

2.4.3.6 Listed wrecks. Non-dangerous wreck symbol in black with Wreck List number.

2.4.3.7 Located wrecks. Black circle 5 mm in diameter.

2.4.3.8 When searchlight sonar, alone, has been used, the tracing is to encompass the entire survey area (ideally an overlay of the largest scale chart or topographic map covering the area). It is to depict the limits of the area swept by searchlight sonar and may be combined with any other tracing, providing clarity can be maintained. This information is used by the charting authority in assigning data quality attributes.

2.4.3.9 Sonar tracings are to carry a clear and comprehensive key to the symbols used. In addition, SSS tracings are to carry a table showing the operating specifications, including range scale, mode (survey or search), beam depression and average towfish height.

2.4.3.10 Some of the data required above may be combined with other tracings provided their inclusion does not interfere with the clarity of existing tracing.

3. SEAFLOOR CLASSIFICATION

3.1 Background

3.1.1 There are three requirements for seafloor classification, i.e. nautical charting, commercial/ environmental and military.

3.1.1.1 Nautical Charting. A relatively simple classification method is used for nautical charting and navigational purposes; it is defined as determining the composition of the seafloor. A list of the classifications is contained in Chart INT 1. The mariner needs this information:

- to decide where to anchor;
- to determine the type of holding ground and how much cable to use;
- to help assess the safety of an anchorage;
- to provide an additional check on navigation.

3.1.1.2 Commercial/Environmental. A more detailed classification, usually obtained using commercial processing software and used for:

- offshore engineering (siting oil platforms, beacons and sea walls);
- mineral exploration;
- fishing etc.

3.1.1.3 Military. A combination of four basic seafloor types with detailed and specific additional data and attributes. Military users rely upon this information for:

- amphibious operations;
- mine countermeasures, i.e. selecting operating areas in order to avoid those of unfavourable seafloor topography;
- submarine and anti-submarine operations, i.e. selection of safe areas for submarines to take the seafloor;

- sonar acoustic performance.

3.1.1.4 In future, military seafloor classification information is likely to be distributed to headquarters and operational units in the form of Additional Military Layers (AML). These are able to be read in embedded geographic information systems and command tactical decision making systems.

3.1.2 Seafloor Classification Models

3.1.2.1 Information is normally presented as a seafloor classification model, examples of which are at Figure 4.8. Data may be obtained by single and MBES, SSS and actual sampling, and is presented as a mixture of symbols and words. Like all fair records the information must be accurately and clearly plotted.

3.1.2.2 The following information is to be shown in seafloor classification models:

- natures of the seafloor from samples;
- texture of the seafloor from echo sounder, SSS etc.;
- seafloor contacts and features (i.e. wrecks, sand waves, trawl scours);
- depth contours.

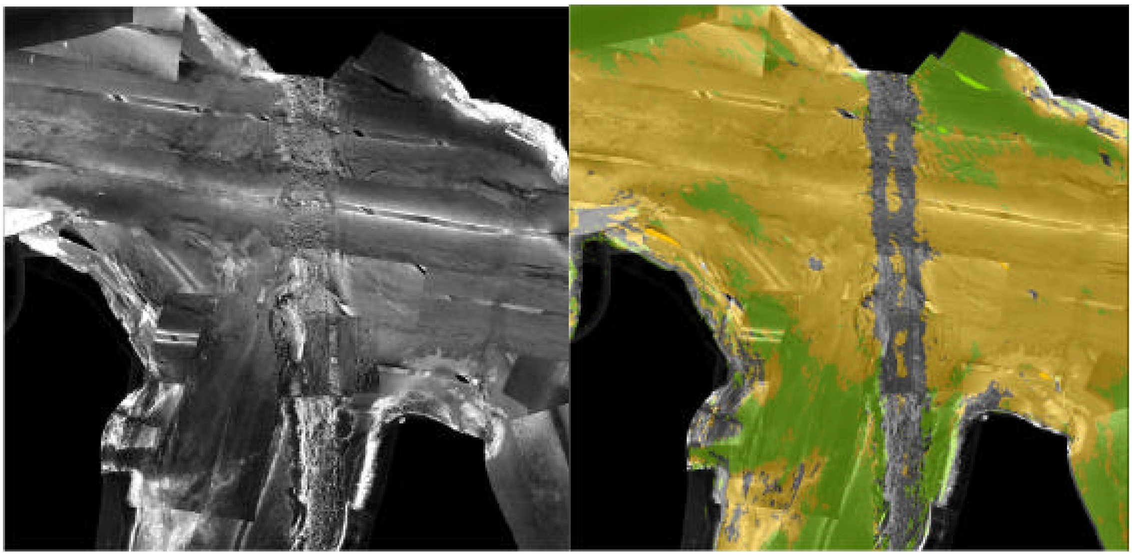


Fig. 4.8 "Example of SSS Mosaic and Classification Models"
(using QinetiQ "Classiphi" software)

3.1.2.3 Examples of Sonar Records. The problems met in identifying wrecks on sonar records are well known to surveyors and need no further amplification. Examples of sonar records for seafloor classification comparisons can be found in "Sonagraphs of the Seafloor" by Belderson, Kenyon, Stride and Stubbs.

3.1.3 Seafloor Samples

3.1.3.1 The nature of the seafloor is to be obtained in depths less than 200 m as follows:

- as required to assist with the interpretation of any SSS records;
- as required to provide ground truth and confirmation of seafloor classification models;
- in all likely anchorages;
- on all banks, shoals and seamounts, particularly when these are likely to be unstable and in the channels between them;
- on the summit and at the base of seamounts, in depths greater than 200 m, when depths are not extreme and appropriate sampling methods are available.

3.1.3.2 In addition, the nature of the seafloor is to be obtained at regular intervals throughout the survey ground. The frequency of sampling will vary, depending on the depth and the extent to which it is homogeneous, with samples obtained at intervals of between 1.0 and 1.7 km in depths less than 200 m.

3.1.3.3 The nature of the seafloor obtained from samples is to be included in the classification model. The correlation between samples and the texture derived from the sonar record is very important; it provides the only real confidence check on the interpretation. It follows that seafloor samples must fulfil three conditions, i.e. they must be:

- a complete sample - underway samplers are known to lose much of the finer portions of the sample as they are recovered;
- from an individual spot - underway samplers may be dragged for several hundred metres, and cannot provide a "spot" sample;
- accurately positioned - samples must be fixed to the same accuracy as any other item of survey information, with the fix taken as the sampler hits the seafloor.

3.1.3.4 To fulfil the above requirements samples must be taken by grab or corer with the ship stopped and the fix obtained by the main survey navigation aid (or one of comparable accuracy). Their position on the classification model is shown by a small dot surrounded by a circle, with the classification positioned next to it.

3.1.4 Nature of the Seafloor

3.1.4.1 The seafloor is formed of rock of various types overlaid in most places by unconsolidated sediments from two main sources:

- materials washed from adjacent land masses or from erosion of the seafloor itself;
- biologically produced sediments which are formed from decaying animal and vegetable products within the ocean basins.

3.1.5 Classifying Samples

3.1.5.1 Classification entails describing a sample under two main headings:

- a descriptive adjective, such as ‘coarse’, ‘small’, etc.;
- a general description, such as ‘Rock’, ‘Mud’, etc.

3.1.5.2 Mixed Samples. Most natural sediments are rarely composed of only one type of sediment, they are often a mixture. When this occurs, classification should follow the principle of listing the most predominant material first, for example *“fSbkSh”* indicates that there is more sand in the sample than there is shell.

3.1.5.3 Grain Size and Grading. Sediments are graded according to grain size at Table 4.5.

General Description	Name		Limits (mm)	Remarks
Mud	*M*	Clay	< 0.002	when dried on hand, will not rub off easily.
		Silt	0.002 – 0.063	when dried on hand, will rub off easily.
Sand	*fS*	very fine Sand	0.063 – 0.125	
		fine Sand	0.125 – 0.250	
	mS	medium Sand	0.250 – 0.50	
	cS	coarse Sand	0.5 – 1.0	
		very course sand	1.0 – 2.0	
Gravel	*smG*	Granules	2.0 – 4.0	from thickness of standard pencil lead to size of small pea.
	P	Pebbles	4.0 – 64.0	small pea to clenched fist size.
	lG	Cobbles	64.0 – 256.0	clenched fist to man’s head size.
Rock	*R*	Boulders	> 256.0	larger than a man’s head size.
		Rock		

Table 4.5 “Sediment Grain Size”
(taken from the UKHO Hydrographic Quality Assurance Instructions for Admiralty Surveys)

3.1.5.4 The size of grain can be determined by eye or by comparison with standard samples illustrated in a “comparator disk”, if held. The finer sediments are the hardest to classify. If size cannot be classified with the naked eye or by comparison, the sample may be placed between the teeth. If it feels gritty then it is silt; if it feels smooth and buttery in texture then it is clay. It is extremely difficult to estimate the relative percentages when samples contain sand, silt and clay.

3.1.5.5 Rock. A sample should only be classified as ‘rock’ if positive evidence is available. If the only evidence held by the collector is a score or dent or damaged sampler, the abbreviation *“h”* (hard) should be used.

3.1.5.6 Other Descriptions. Where additional qualities can be identified or the seafloor type can be positively classified as comprising another distinct material, the various references should be consulted for guidance.

3.1.6 Methods of Obtaining Seafloor Samples

3.1.6.1 Samples of the seafloor can be obtained by a variety of means, the most common are:

- lead lines;
- grabs;
- snappers and scoops;
- corers;
- dredges;
- divers;
- remotely operated vehicles (ROV) and submersibles;
- opportunity based sampling (i.e. from anchors).

3.1.6.2 Selection and use of the appropriate device will depend on the nature of the investigation, the character of the seafloor, the depth of water and the shipboard equipment available for lowering and retrieving sampling equipment.

3.1.6.3 Sounding Leads. The armed lead line is a traditional method of obtaining and indicating the nature of the thin surface layer of the seafloor. It can give no idea of the depth of a surface layer or what is underneath. Leads are armed with tallow or a similarly sticky substance such as petroleum jelly or grease to which particles of sediment adhere. When the seafloor is strewn with larger features such as pebbles or rock, an impression of the sea floor material may be obtained but this cannot be guaranteed.

3.1.6.4 Advantages of the lead line are that it is cheap and simple to operate. Disadvantages are:

- larger material may not be detected (for example boulders);
- only the surface layer is sampled;
- sampling becomes unreliable as depths increase;
- the sample is contaminated by the material used for arming;
- the sample is disturbed when collected.

3.1.6.5 Grabs, Snappers and Scoops. These are supplied for the purpose of collecting medium size samples of the surface and immediate sub-surface layer of the seafloor. They usually comprise a bucket or scoop, which is activated on hitting the seafloor. Some are spring-loaded; others close when raised off the seafloor. Grabs are rarely suitable for sampling soft or liquid mud as the sample is often washed out of the bucket before it reaches the surface.

3.1.6.6 Shipek Grab. The Shipek Grab consists of two concentric half cylinders; the inner half cylinder or sampling bucket is held open against a pair of powerful axial springs by a pawl. On striking the seafloor a sliding weight trips the pawl and allows the bucket to rotate through 180° under the torque of the springs. During this rotation the bucket scoops a sample from the seafloor. The bucket then remains closed whilst the grab is hauled to the surface. The Shipek Grab is most effective on soft and unconsolidated sediment. It is liable to bounce on a compacted seafloor and the closing action of the bucket can lift the grab off the seafloor giving only a superficial sample or none at all. In these conditions improved results can sometimes be obtained by reducing the speed of impact of the grab on the seafloor.

3.1.6.7 Corers. These are used to obtain an undisturbed vertical sample of the seafloor. They often penetrate a considerable distance below the seafloor surface. Corers usually comprise a tube or box shaped cutting mechanism similar to an apple corer or pastry cutter. They are driven into the seafloor and when withdrawn they retain an undisturbed sample of the sediment layers.

3.1.6.8 Retaining mechanisms vary from creating a vacuum on the back of the sample to cover plates or shutters. Often there is a combination of methods to hold the sample in place. Corers may be driven into the seafloor by a number of means: their own weight, explosives, pneumatics or mechanical vibration.

3.1.6.9 Dredges. Dredges are designed to be dragged along the seafloor collecting loose material and sediment. They often incorporate a filter that allows smaller sediments to pass through. Samples are always disturbed but do reflect the seafloor materials over a reasonably large area. Dredges can be deployed in all depths of water.

3.1.6.10 Other Sampling Methods

Divers. An inspection by divers allows a positive identification of the seafloor. Large as well as small features can be identified. Divers are limited by the depth to which they can work but, for shallow water and with time permitting, this is a good method of obtaining samples.

Autonomous and Untethered Underwater and Remotely Operated Vehicles (AUV, UUV & ROV) (see Note). AUV, UUV and ROV can assist in classifying the seafloor either by collecting samples (usually scoop or grab) or by obtaining video images for later inspection.

Note: AUV are programmed to operate independent of a support platform; UUV are untethered but under remote control from a support platform; whilst ROV are tethered.

Opportunity Basis Sampling. Useful samples can also be obtained from ship’s cables, anchors or buoy moorings. These samples must be used with some caution since only samples of a clinging nature are likely to survive the washing action of equipment on its way to the surface.

3.1.7 Seafloor Sample Records

3.1.7.1 Seafloor Sample Log. Data should be formatted to assist in the archiving of relevant data and such that it will be readily available for interested authorities. The Report of Survey is to contain full details of the methods of sampling employed during a survey together with any problems that may have been experienced.

3.1.7.2 The location and classification of seafloor samples obtained is to be shown on a tracing or digital model accompanying the bathymetric data.

3.2 Classification Sensors

3.2.1 This section describes the various sensors used for seafloor classification.

3.2.1.1 SSS. In addition to locating wrecks and obstructions between survey lines, SSS also provides a considerable amount of other seafloor information. These data, when combined with seafloor samples and depth contours to produce seafloor classification models, are of great value. The importance of this information has grown over the years to such an extent that, in many surveys, sonar rather than bathymetric considerations govern the selection of line direction and spacing. Great care is needed in the preparation and checking of these tracings if their full potential is to be realised.

3.2.1.2 MBES. The introduction of MBES systems in hydrographic surveying has meant not only the ability to determine bathymetry more accurately and with greater coverage than before, but also the ability to determine seafloor boundaries and sediment types relatively quickly and effectively. With this in mind, the surveyor is now able to interpret the backscatter imagery from swath systems as well as side scan imagery. The added benefit of obtaining backscatter information from MBES systems, while collecting bathymetric data, allows a more cost (and time) effective survey to be conducted.

3.2.1.3 SBES. Commercial seafloor classification software that is capable of being fitted to SBES has been available for some years. Used particularly in the fishing industry, a typical system is described below.

3.2.1.4 Other Methods. Other sensors with potential for seafloor classification include:

- Airborne Lidar Bathymetry (ALB). Research is continuing into the extraction of information other than bathymetry from the laser return waveform including turbidity and seafloor classification.

- Airborne Electromagnetic (AEM) Bathymetry. AEM methods offer the potential to obtain seafloor classification information but this capability has yet to be developed.

- Remote Sensing. Seafloor classification information can be obtained from satellite and aerial imagery in shallow water but still requires ground truth data.

- Forward Looking Sonars (FLS). Originally designed purely for navigation and collision avoidance, some recent FLS developments offer bathymetric and seafloor classification capabilities. For example, the Thales Underwater Systems “Petrel” FLS matches the energy of acoustic returns to the ambient noise level and beam angle of incidence on the seafloor to provide a seafloor reverberation figure of merit which will be unique for varying seafloor densities, materials and porosity. By ground-truthing these figures of merit a real-time swath seafloor classification capability is available in parallel to bathymetry.

3.3 Classification - Theory

3.3.1 This section introduces the collection and interpretation of backscatter information and compares the methods used by SSS and MBES. The advantages and disadvantages of each are discussed. It also covers the methods that the MBES system uses to remove the distorting effects due to the angle at which the signal hits the seafloor and other causes.

3.3.1.1 SSS, and most MBES systems, can display a representation of the seafloor using the principle of acoustic imaging. Most SSS pictures show relatively unsophisticated representations of the returning ping in the sense that the image is only corrected for a limited range of measurable parameters. For example, modern side scan receivers often have the ability to measure the forward velocity of the vessel and adjust the along track axis of the image so that the scale in this direction equates to the scale across track. Also, they can measure the height of the fish above the seafloor, and remove this portion from the image so that the image starts at the seafloor underneath the towfish and covers the seafloor out to the maximum range of the set. The image can be corrected so that the distance on the image equates to the distance on the seafloor, however this is normally achieved by making the assumption that the seafloor is level. Since this is in fact not the case, there will be distortions on the side scan image.

3.3.1.2 On the other hand, the provision of backscatter information is a by product of the bathymetric data collection for a MBES system. It is akin to the output of SSS and produces a representation of the seafloor in terms of the intensity of the returning echo. The significant difference between the two is that the MBES is measuring the depth concurrently with the backscatter information and this allows for a more sophisticated level of display. The data on depth, when combined with beam angle, effectively gives the position on the seafloor to which the backscatter information relates and therefore provides a true geometric correction of the backscatter image.

3.3.2 Backscatter Imagery

3.3.2.1 The result of the MBES side scan imaging based on backscatter information is a mosaic covering the seafloor which displays the backscatter intensity equating to each point on the seafloor. There is normally an ability to combine backscatter and depth information so that they are co-registered by position. Assuming the lines have been run appropriately, the imaging should provide 100% coverage and it may be that the backscatter information covers more than the bathymetry if beams have been invalidated for accuracy reasons. It is likely that the extra backscatter information will not be used since it does not have depth information associated with it, but it remains available just the same.

3.3.2.2 A certain amount of post-processing will have been carried out to normalise the backscatter image to remove the distorting effects on the original signal return. The corrections will depend on range (to correct for attenuation and beam spreading), source power (which should be recorded with the echo information) and beam directivity; both transmit and receive, if this varies over time. Additionally, there will be corrections to be applied that depend on the signal path and the area that is ensonified. These are corrections for beam angle, ray path and local seafloor slope, which can all be combined into a grazing angle at which the signal hits the seafloor. Figure 4.9 shows examples of scattering strength for different seafloor types at different grazing angles.

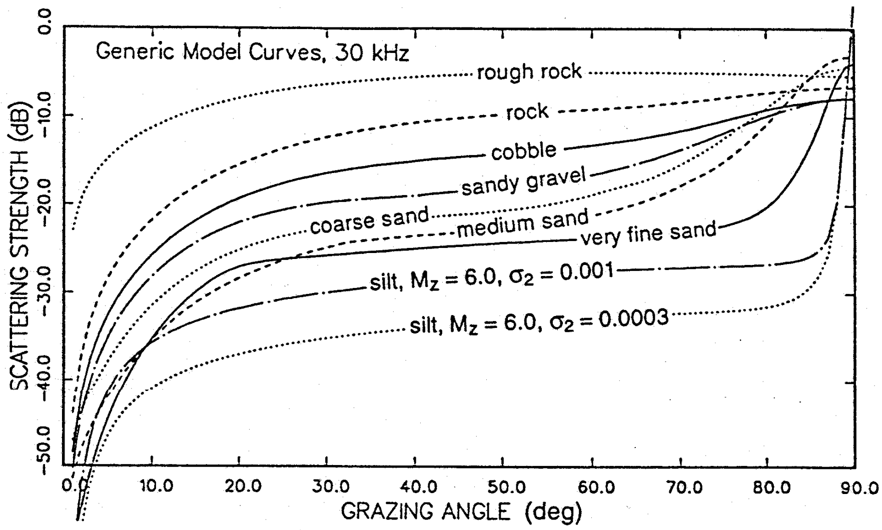


Fig. 4.9 "Examples of Scattering Strength"
(from "High Frequency Ocean Environmental Acoustic Models Handbook", October 1994)

3.3.3 Side Scan Registration

3.3.3.1 The correction of the image for position is termed side scan registration (as the term side scan is often used with a MBES system to refer to the backscatter intensity image). The required correction translates between the slant range given by the time of travel and the true seafloor position, or at least a true distance from the point underneath the transducer to the feature patch of seafloor.

3.3.3.2 As mentioned before, the method used with SSS images tends to be quite simplistic, but, using the extra depth information available in MBES systems, knowledge of the sound velocity profile and the attitude of the vessel at the point of transmission, the registration can be made more accurate. A large part of the calculation has already been carried out to produce the corrected depths in the bathymetry application of the MBES and sometimes that information can be made available to the side scan image.

3.3.4 Mosaicing

3.3.4.1 The transformation of the MBES side scan image into a regular raster image is called mosaicing. The image will be positionally corrected for the movement of the vessel; however there may still be some problems with the mosaicing procedure. In some MBES, the small footprint size in the central beams may leave small gaps between each individual footprint. The aim with the side scan image is to produce a regular raster image that allows direct comparison of one point with another and gaps in data may make this difficult. It may be possible to fill in the gaps by interpretation.

3.3.4.2 If coverage is in excess of 100% and there is overlapping of data, it is likely that the data will have been collected at different angles and directions of ensonification. Rather than attempt to combine the data, the data from the preferred beam is accepted whilst the other beam is suppressed. There will inevitably be a discontinuity where the two swaths meet but the above method minimises the distortion this will cause. There are various methods available that can automatically choose the preferred beam, for example giving preference to the mid-beam over the nadir and the far range.

3.3.4.3 The interpretation of the backscatter image will therefore depend on knowledge of the information that the system retains and its method of presenting the data. Some systems have the ability to retain information on the distribution of data within the beam, so detail that is smaller than the beam footprint can still be seen. Other methods use a reduced data set, retaining (for example) only the average or the peak intensity for each beam, which provides less detail. Figure 4.10 shows that the bathymetry alone does not provide the same information on the change in seafloor type as the raster backscatter mosaiced image.

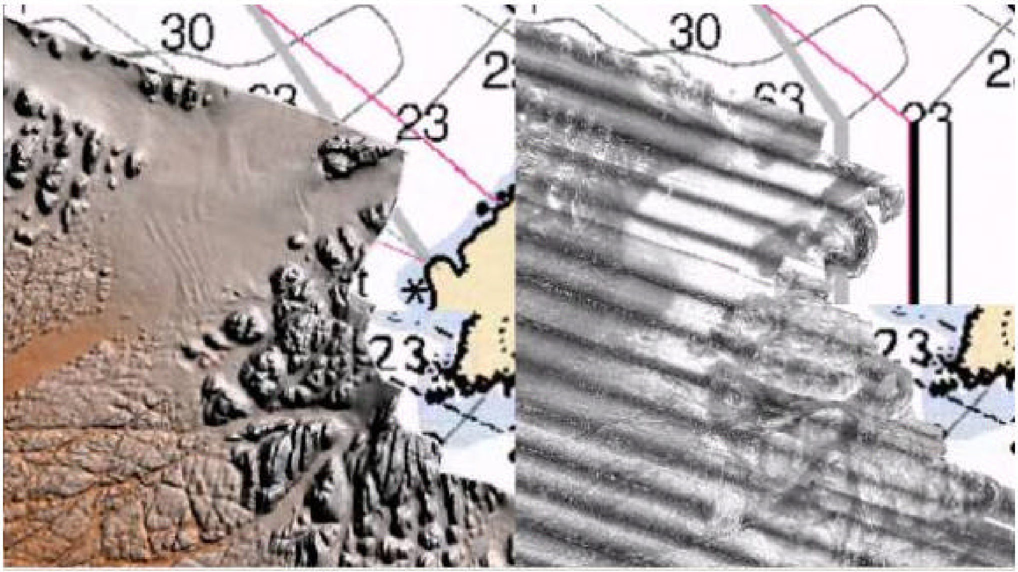


Fig. 4.10 "Seafloor Imagery - Bathymetry (left) v. Raster Backscatter Mosaiced Image (right)"

3.3.5 Classification - General

3.3.5.1 There are further complications when attempting to classify the type of seafloor. Different acoustic impedance characteristics of the seafloor will affect the shape and the characteristics of the return. If the seafloor is rough, but with detail smaller than the beam footprint, then this will have an influence on the intensity of the return.

3.3.5.2 The only way to truly allow for these different effects is to have full knowledge of the seafloor in advance and this is only possible where actual ground truthing (i.e. seafloor sampling) has happened. However, certain types of seafloor will have different general characteristics; hence the backscatter may be used to conduct general classification. If particular returns are matched by ground truthing, then a 'library' of backscatter classes can be built up, enabling automatic classification. This library can be as complex as required with different areas of the roughness/hardness graph assigned unique classifications. There are a number of different software tools for this purpose although each will likely have a different procedure and requirements to perform its task.

3.3.5.3 Classification of the seafloor using the acoustic image is a rapidly developing field. Initial advances were made with the use of vertical incidence systems (SBES), where the method was to study all the parameters of the returning echo, including the variation in intensity over time and the frequency scatter graph, to provide an indication of the sea floor type.

3.3.5.4 The requirement for seafloor classification depends on the final use of the information. In return, the particular parameters that are used to identify a particular seafloor type may depend on the classification requirement. Typical characteristics that may be measured are the type of seafloor in traditional hydrographic terms, which would classify the seafloor in terms of the grain size, texture and type. Other characteristics may be physical properties of the seafloor that may be relevant for, say, a pipeline survey, or acoustic properties that may be of interest to minewarfare, anti-submarine warfare, and oceanographers. These include:

- sediment type, i.e.:
  - grain size, texture, i.e. sand, silt, clay, gravel;
  - mineralogy, i.e. ash, clay, silica, carbonate;
  - genetic, i.e. biogenous, terrigenous;
- physical properties, i.e. grain size, density, and porosity;
- acoustic properties, i.e. velocities, attenuation;
- geotechnical properties, i.e. shear strength, elastic moduli;
- morphology, i.e. texture and relief.

3.3.5.5 Various approaches have been taken to the problem of seafloor classification, focusing on different properties of the returning signal and with different methodologies to achieve the result. In order to achieve this remote classification we look at systems and models developed for the interaction of sound with the seafloor and the effect that this interaction should have on the pulse shape. One such system adopted for seafloor classification using SBES is RoxAnn, developed by Marine Microsystems Ltd.

3.3.6 RoxAnn

3.3.6.1 RoxAnn is one of a number of commercial seafloor classification systems that is connected, in this instance, to existing echo sounders (typically vertical incidence systems) by means of a "head amplifier" which matches the impedance of the system to that of the echo sounder. The design was based on observations of echo sounder performance in known areas of different seafloor types. Sediment classification is accomplished by the identification of two parameters (see Figure 4.11):

- E1 - the integrated energy under the tail of the first return, i.e. roughness;

- E2 - the integrated energy under the second (multiple) return, i.e. hardness

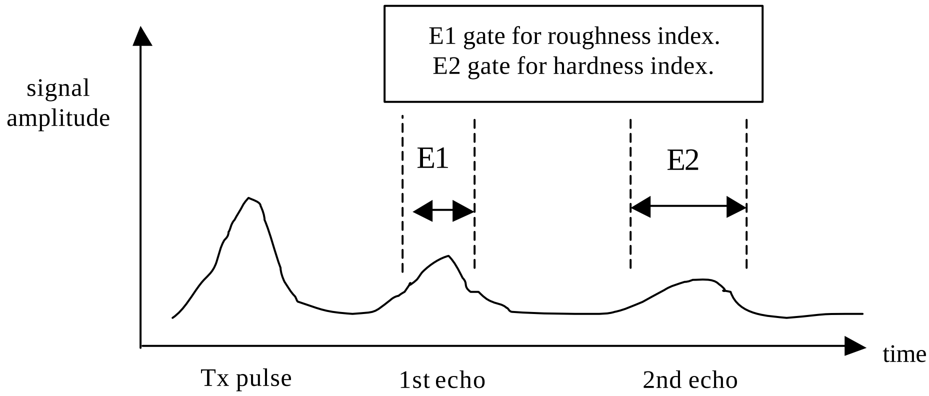


Fig. 4.11 "RoxAnn - Quantification of Roughness (E1) & Hardness (E2)"

3.3.6.2 Then, by use of a look-up table which will graph hardness against roughness, you can enter an observed value that has been ground-truthed and calibrate the system for automatic classification in that locality. The system requires periodic recalibration and will also require recalibration when moving to a new area. Figure 4.12 shows values of E1 and E2 plotted, and then a 'known' seafloor type allocated.

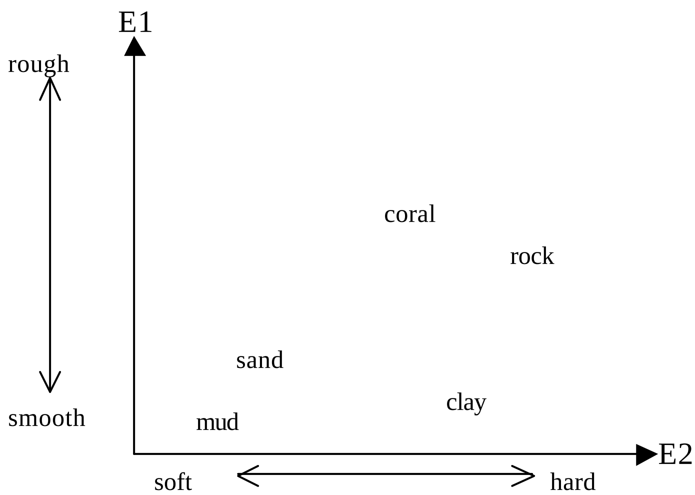


Fig. 4.12 "RoxAnn - Values of E1 and E2 (example only)"

3.3.6.3 The advantages of this system are that it is relatively simple and inexpensive. The disadvantages are that the system is not quantitative, it must be calibrated and it depends on multiple returns which raises the question about variability as a function of sea state.

3.3.7 Classification using MBES

3.3.7.1 MBES provide us with both geo-referenced measurements of the instantaneous back-scatter intensity, and spot estimates of seafloor elevation (soundings). Both of these can be used, either together or separately, to attempt seafloor classification, usually in conjunction with commercial software packages designed for that purpose. For those systems which are calibrated, or for which at least a relative calibration can be performed, the backscattered intensity measurements can be reduced according to:

- the range to feature (attenuation and spherical spreading);
- source power, beam directivity (transmit and receive);
- area ensonified (beam angle, refracted ray path, local seafloor slope).

3.3.7.2 There are three main methods employed in swath sonar seafloor classification, based on the variability in the echo structure to infer information on the nature of the seafloor:

- texture mapping and spectral estimation;
- echo amplitude peak probability density function;
- acoustic backscatter angular dependence functions.

3.3.8 Textural Mapping

3.3.8.1 This method looks at the variation of backscattered intensity as a function of 2D space (horizontal dimensions). It is based on the identification of significant changes in the characteristics of the echoes both within a ping and over a number of consecutive pings. In essence, it is the estimation of the 2D spatial statistics of an acoustic backscatter amplitude image of the seafloor.

3.3.8.2 Even in the absence of a calibrated sonar system, it is easy to see that the textural characteristics of side scan imagery contain information about the seafloor. Most early SSS were developed for the purpose of feature recognition, in which the aim was to use the full dynamic range of the display device, generally a wet paper recorder or graphical monitor, to maximise the contrast in the returned echo. For this purpose automatic gain controls were developed. The detrimental side of this development was that in most cases, the absolute level of the backscattered intensity was not preserved. Nevertheless, such a signal processing technique was ideal for bringing out textural information in the imagery. This has been achieved by the introduction of two methods:

- power spectra;
- grey level co-occurrence matrices.

3.3.9 Power Spectra

3.3.9.1 The seafloor acoustic backscatter changes roughly as the $\cos^2$ of the angle of incidence (Lambert's Law) out to low grazing angles. Therefore, it can be assumed that the variations in the amplitude of seafloor echoes received by the sonar over this angular sector are expressions of the inherent roughness of the backscattered surface. This would indicate the possibility of classifying these returns, and hence infer seafloor types, based on their spectral shape.

3.3.9.2 When applied to multibeam sonars, this method must be limited to the outer segment of the swath where the angular dependence of seafloor acoustic backscatter levels off and where the length of the instantaneous ensonified area is relatively constant across-track. In the near vertical incidence region a combination of the high aspect ratio of the hull-mounted sonar, the rapidly changing size of the ensonified area and the regular angular dependence function of acoustic backscatter put severe limitations on the assumption that spectral shape directly relates to seafloor type.

3.3.9.3 In addition, because the time series of backscatter strength obtained with a multibeam configuration is actually a composite of several beam traces, there is the possibility of introducing energy into the power spectrum at spatial wavelengths equivalent to the beam spacing.

3.3.9.4 As you move between shallower and deeper water, the pulse length of many shallow water swath systems is varied. This changes the instantaneously ensonified area and the length scales that can be observed with the power spectra.

3.3.10 Grey Level Co-occurrence Matrices

3.3.10.1 To identify boundaries of like texture patterns in the side scan image, the classic image processing techniques of Grey Level Co-occurrence Matrices (GLCM) is used. This technique characterises the 2D spatial inter-relationships of the grey levels (where the darkness of the grey refers to the intensity of backscatter) in an image with a scale ranging from fine texture, corresponding to frequent level changes over short distances, to coarse texture corresponding to few level changes over long distances. Co-occurrence matrices are computed for a set of distances and angular spatial relationships. Each GLCM will correspond with a different texture, which can then be interpreted as a seafloor type.

3.3.10.2 One drawback of the GLCM method is that it must be implemented on a side scan mosaic, which is a raster product. As discussed, the mosaicing process is a compromise between preserving the across track resolution of the backscatter amplitude data and honouring the along-track resolution. Thus the side scan mosaics commonly are averaged (or median filtered) versions of the raw intensity data. As such, they cannot exhibit the same statistical characteristics as the original raw data. Therefore, the characteristics used for classification are only applicable to data that has undergone exactly the same transformation from a raw side scan time series to a raster product. In addition, some form of ground truthing is required to identify the sea floor type because there are no models that link GLCM to specific seafloor physical properties and different lithobgy's (seafloor characteristics) can exhibit the same textural characteristics.

3.3.11 Echo Amplitude Peak Probability Density Function

3.3.11.1 Echo Amplitude Peak Probability Density Function (PDF) seafloor acoustic backscatter is a reverberation process whose stochastic (statistical) behaviour can be described by the Gaussian distributed instantaneous quadrature samples, with an envelope (echo amplitude) distributed according to a Rice probability density function and a phase uniformly distribution. Remembering that the end members of the Rice PDF are a Gaussian shape when the returned signal is mostly coherent and a Rayleigh shape when it is mostly scattered, it is possible to derive a measurement of coherence from the statistics of the envelope. Figure 4.13 shows a comparison of Rayleigh and Gaussian statistical curves, measuring the probability of an echo's amplitude.

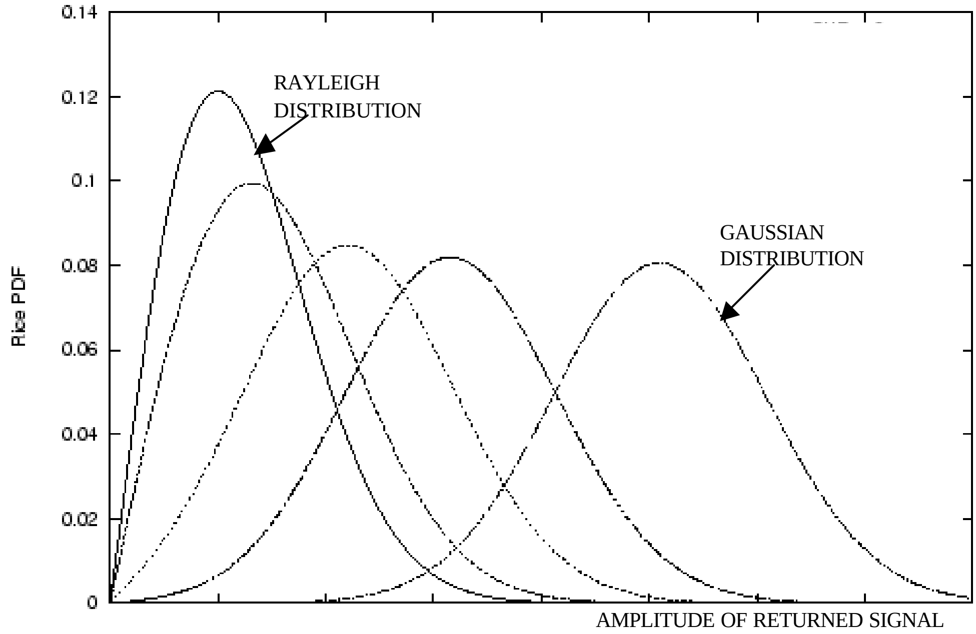


Fig. 4.13 “Comparison of Rayleigh and Gaussian Curves”

3.3.11.2 The mean and the variance of instantaneous amplitude values are dimensional quantities and thus imply that the sonar system must retain at least the relative changes in backscattered amplitudes of the echoes. Any changes in source power and/or receiver gain settings can be taken into account (compensation for automatic gain control). Swath backscatter amplitude data (side scan) presented as mean amplitude is easy to understand as a classification tool. Approximately constant mean amplitude over a region suggests a homogenous seafloor type and gross changes in mean amplitude suggest changes in the seafloor. However, presentations of regional changes in seafloor backscatter amplitude generally ignore, or try to empirically normalise for, changes in the ensonification geometry (grazing angles) across the swath.

3.3.11.3 This fitting of the observed PDF against the standard models is performed on normalised echo PDFs, thus the absolute mean amplitude of backscatter is ignored in the process. The method is to attempt to separate out the ratio of coherent and incoherent components in the data. With data from a calibrated sonar system, the normalised step can be skipped and the absolute amplitudes used instead.

3.3.12 Angular Dependence

3.3.12.1 There are a number of models that predict the angular dependence of seafloor acoustic backscatter based on several factors (generating parameters). These include the impedance contrast of the sediment-water interface, the statistics of the roughness of that interface and any possible contributions from volume inhomogeneities within the sediment layers. The quantity of interest in this method is the backscattering strength per unit solid (3D) angle. This is obtained from the measurement and compared to model predictions to estimate the generating parameters.

3.3.12.2 High aspect ratio systems such as multibeam sonars provide measurements of the backscattered amplitude at grazing angles ranging from vertical (90°) through grazing angles as low as 15°. This is in contrast to the distribution of grazing angles recovered by deep-towed, low aspect ratio SSS, which tend to be biased to very low angles.

3.3.12.3 Commensurate with this method is the requirement to know the ray path of the acoustic energy as it strikes the seafloor and the 3D slope of the seafloor interface that it strikes. This leads to assigning an instantaneous measure of backscatter amplitude to that angle. In order to arrive at a good estimate of the mean backscatter strength at that grazing angle, a large number (>10) of instantaneous measurements are used. This obviously assumes the seafloor under investigation is not changing over the length of the multibeam swath (i.e. the seafloor type is the same from nadir to the far range).

3.3.13 Acoustic Backscatter Data Interpretation

3.3.13.1 In the first instance, interpretation of a digital side scan image is often difficult. At the very limit of resolution is the single instantaneous sample of backscatter intensity. This is derived from a complex sum of all the individual scattered contributions from within the ensonified area and also the scattered contributions from the volume of sediment below the ensonified area. Notwithstanding the derived solution, there are three main effects that are noticeable in any side scan mosaic:

- variations in backscatter strength due to changing seafloor type;
- variations in backscatter strength due to changes in seafloor slope;
- true cast shadows.

3.3.13.2 For conventional side scan the first two are ambiguous. There is no way to unambiguously tell whether fluctuations are due to slope or texture. In reality it is rare to see a significant change in the seafloor slope without a change in texture. In contrast to conventional side scan, swath sonar systems can resolve the ambiguity for those cases where the topographic wavelength is greater than the beam spacing, although roughness at shorter wavelengths cannot be resolved.

3.3.13.3 True cast shadows can be recognised by both systems as long as the signal-to-noise ration is high enough - thus the drop in signal strength is greater than that expected for any real sediment type. Interestingly, swath sonars cannot predict the presence of a shadow from the topographic information alone. This is because a shadow implies slopes steeper than the ray path and thus swath sonars cannot see behind the shadowing feature. This is important to remember when using side angular sectors. Steep topography facing away from the sonar is not adequately resolved and thus the resulting derived terrain model will be distorted. Even with swath sonars, it is often not clear whether short wavelength variations in seafloor backscatter are a result of either of the above effects. The only way to resolve this is to image the seafloor from multiple near-orthogonal directions.

3.3.13.4 First and foremost, the surveyor is concerned about verifying potential hydrographic hazards on the seafloor. Any confirmation or denial of the validity of an anomalous sounding represents an aid in the interpretation of the data. This ultimately allows greater confidence in the quality of the sounding data which will appear on a navigation chart. As we have seen there are resolution limitations to high-speed multibeam imagery, which means that you cannot always resolve the discrete hydrographic anomalies that are of interest. This lends to the discussion of deploying conventional, towed SSS in conjunction with a swath sonar system.

3.3.13.5 When the beam reaches the seafloor some of the beam is reflected back in the form of an echo, but much of the energy is scattered in all directions, and some is even absorbed into the seafloor. The vertical incidence case is concerned mainly with the reflection properties of the seafloor, and again, different characteristics of the echo sounder beam have an effect on the amount of the signal that is reflected. The frequency of the signal is one of the most important attributes in this respect. The multibeam case is more complicated and the scattering properties of the seafloor take on a larger importance.

3.3.13.6 Returns from a smooth hard seafloor. As the sound wave travels through the water, it moves by displacing water particles, causing them to vibrate and so allow the passage of the wave. The water has a low acoustic impedance, or a low resistance to the movement of the wave. When the wave reaches the seafloor however, the seafloor has a high acoustic impedance and does not allow the sound wave to continue into the seafloor. The particles are densely packed and are not able to move easily. Since the total energy must be maintained and the energy cannot pass into the seafloor in the form of a sound wave, it must go somewhere and the result is that it is radiated back into the water. Some, probably a small percentage, will be reflected back in the direction of the incoming wave and will travel back to be received at the sonar transducer as an echo.

3.3.13.7 Effects of different types of seafloor and differing angles of incidence. Different types of seafloor will have different levels of acoustic impedance. If the level is low then some of the sound energy is absorbed into the seafloor and the returning echo will be weaker. If the level is high then more is reflected. Similarly, the intensity of the reflected signal is also dependant on the angle of incidence. If the angle is high, approaching 90°, then a large part of the sound wave will be reflected back towards the sonar. If the angle is low then the major part of the sound wave will be scattered in a direction away from the transducer, however some will still return as a weak echo.

3.3.13.8 The type of seafloor will have an effect on the returning signal as well. The relationship between the angle of incidence, the type of seafloor and the level of the returning signal is not a straightforward one. For the beam arriving at a low level of incidence, if the seafloor is rough then there will be more faces that are near to a right angle to the incoming sound wave and therefore give a stronger reflection. A smooth seafloor will in general result in more of the signal being scattered in other directions and not back in the direction of the sonar receiver. For a high angle of incidence, however, the situation is likely to be reversed and a smooth seafloor may give a better return. This will however depend on a number of factors such as particle size and the composition of the seafloor.

3.3.14 Military Classification Models

3.3.14.1 In preparing a military classification (or texture) model from the sonar records the first task for the surveyor is to decide whether the texture of the seafloor is mud, sand, gravel or rock. It is appreciated that the seafloor contains a wide variety of combinations of the four basic categories, but more detailed analysis is best undertaken by written descriptions. Clearly defined boundaries between different types of seafloor should be shown as firm lines and ill-defined limits should be depicted as pecked lines. Figure 4.14 shows an example of a military classification model.

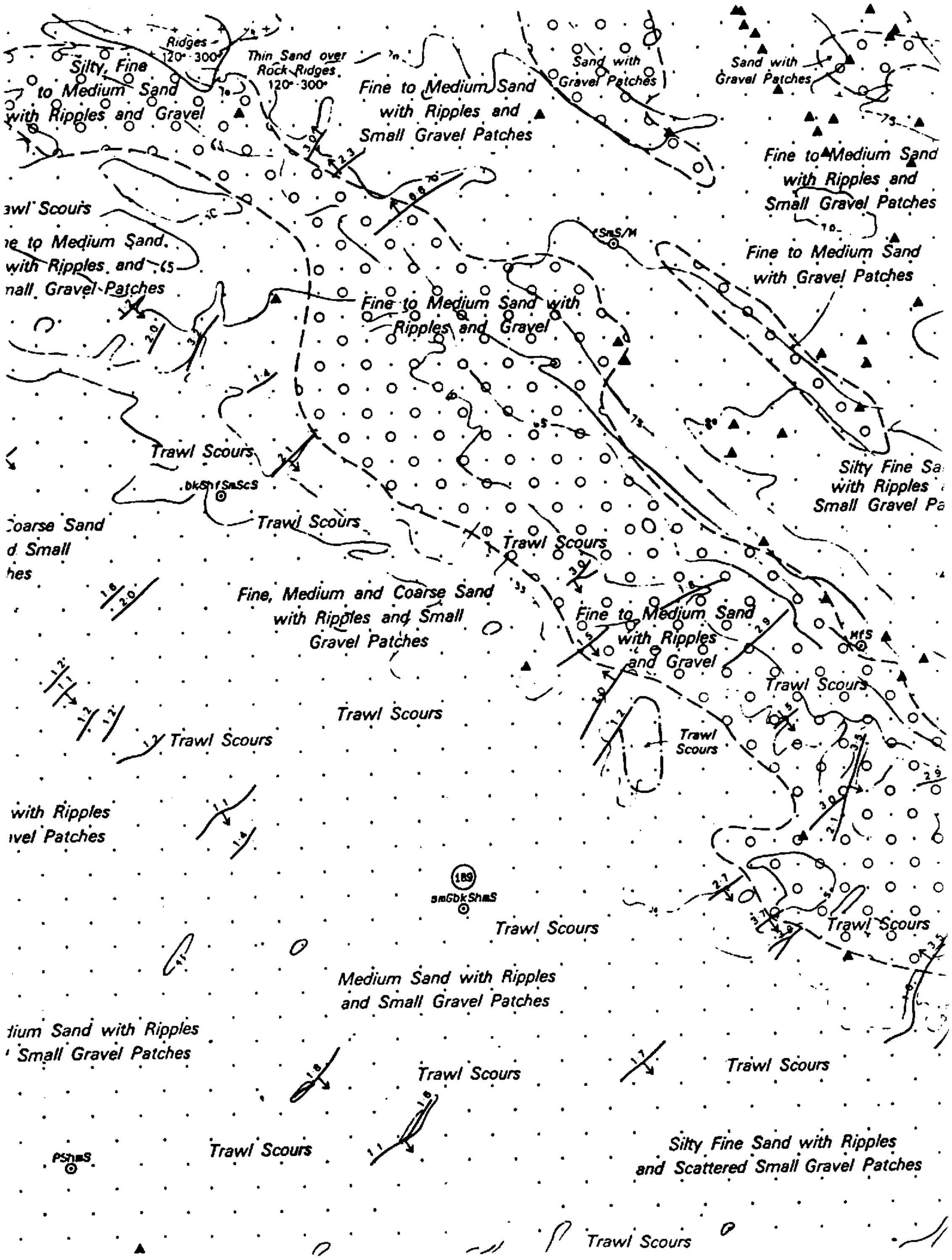


Fig. 4.14 "Example Military Seafloor Texture Model"

3.3.14.2 The graphic representation of the seafloor texture should be amplified by use of written descriptions. Examples of the terms to be used, together with their definitions follow. It is stressed that these are not exhaustive; other words may be used providing the meaning is clear to all who may use the information. Descriptions of "negative" features (for example "flat featureless sand") are as useful as information on more prominent characteristics. Written descriptions should be kept brief.

3.3.14.3 Features such as wrecks, sand-waves, trawl scours and pipelines also form an important part of the description of the seafloor. These features are invariably more important than written descriptions and in congested areas their inclusion should take priority.

3.3.14.4 Sandwaves are a common feature of seafloor topography and may occur either as isolated features or in fields. Different symbols are used for each type:

- Isolated sandwaves. To ensure accuracy, the position of the crest of the wave must be plotted from the echo-trace and not from the sonar record. The symbol for an isolated sandwave is then to be positioned along the line of the crest. If the wave is asymmetric, a small arrow is to be inserted pointing down the steeper side of the feature, the arrow should be omitted if the wave is symmetrical. Details of the height of the crest above the trough should be included.

- Sandwave fields. Many sandwaves occur in groups with similar height and orientation. Under these conditions individual waves need not be plotted. The extent of the field should be delineated, again referring to the echo-trace for accuracy and the symbol for a sandwave field inserted. The orientation of the crests should be indicated, as should the wavelength, height, symmetry and the steeper side.

3.3.14.5 For plotting purposes, a sandwave is defined as having a height greater than one metre. Features smaller than this should be classed as ripples. The wavelength is defined as the distance between two adjacent crests; the height is the difference in depth between a crest and its adjacent trough. As a rule of thumb, crests which plot closer together than one centimetre on paper and are similar in orientation, height and wavelength should be considered a field. Ripples are often superimposed on waves and may have a different orientation a brief written description such as "Ripples 120/300" should be placed next to the sand-wave symbol.

3.3.14.6 Small sea floor contacts. All non-ephemeral contacts larger than one metre must be plotted. Where more than five such contacts exist per square centimetre the area may be delineated and a notation made. Wherever possible the number of contacts in each area should be stated, written descriptions may be included where useful.

3.3.14.7 Wrecks and obstructions. All wrecks and obstructions located during the survey must be included in the classification model. Wrecks are to be shown by the "non dangerous" wreck symbol, oriented in the same direction as the wreck. The extent and direction of any scour is to be noted, i.e. "Scour 155/50m". Other obstructions are to be shown using the "foul" symbol, with a written description if possible, i.e. "wellhead".

3.3.14.8 Small depressions. Certain areas of seafloor may contain small depressions, distinguishable on the sonar trace by the "shadow" being in front of the contact. Some may show a pronounced lip and include "pock-marks". Unless their origin is known (for example if an oil-rig is moved during a survey) classification should not be attempted.

3.3.14.9 Trawl scours. In many areas, trawl scours are a frequent and distinctive part of the seafloor. Their importance is increased by the fact that they are most often met in otherwise flat areas. Isolated trawl scours are to be shown individually; concentrations of them may be delineated and the wording "numerous trawl scours" inserted.

3.3.14.10 Pipelines. All pipelines detected during a survey are to be plotted. Areas of buried pipe should not be interpolated unless they are visible on the side scan trace, in this event the word “buried” is to be inserted as required. Pipes which stand proud of the seafloor are to have their height in metres noted at intervals.

3.3.14.11 Depth contours. Contours are to be included with the normal vertical interval being five metres. In areas where a large range of depth occurs this may be expanded at the discretion of the surveyor, providing the presentation of the “form” of the texture is maintained. The purpose of drawing the depth contours is to assist the surveyor in his interpretation of the sonograph.

3.3.14.12 Descriptions for use on Military Seafloor Classification Models:

Sandwaves. Straight or sinuous ridges of sand commonly aligned across the dominant tidal stream or current. Minimum height is one metre. Crest separation (wavelength) can be up to 1000 m with heights reaching 20 m. May be symmetrical or asymmetrical, and may have ripples on them.

Ripples. Small ridges of sand, similar in shape to sandwaves but with a height of less than one metre. Usually orientated transverse to the tidal or current flow with wavelength of less than 15 m. May not be detectable with an echo sounder.

Furrows & Ridges. Longitudinal bed-forms in gravel, sand or mud, some of which can be 9 km long and up to 14 m wide. They may be solitary, but more usually occur in groups. They are generally parallel to the prevailing currents.

Sand Ribbons. Normally apparent overlying a coarser type of seafloor. Most are straight and parallel with currents. Can be up to 15 km long, 200 m wide and are generally only a few centimetres thick. Typically have a "laddered” appearance due to the presence of ripples.

Gravel/ Sand/ Mud Patches. Thinly-spread patches of gravel, sand or mud no more than 100 m across and commonly less than 2 m thick. May be depositional and subject to movement. Shape may be determined by the relief of the underlying seafloor.

Rock Outcrop. A patch of rock covering a small area. Refers to a cohesive group, not a collection of boulders.

Pinnacle. A rock of limited horizontal extent with height considerably greater than surrounding rocks.

Ledge. Rock outcrop with length in excess of 300 m and relatively narrow in comparison. Often found in groups, with similar direction and extent.

Bank. Usually of sand or gravel, but may be of rock. A rise in the seafloor over a relatively small area, but fairly prominent in relation to its surroundings. When formed of sediment it is often oriented along the tidal flow.

Large/ Small. Preferred to big, great, high/little, slight, mini, etc.

Broad/ Narrow. Used to express width when qualifying such features as sand ribbons. Broad should only be used for ribbons over 150 m wide, narrow for those less than 10 m wide.

Smooth. Preferred to even or level, and may refer to a seafloor that is either flat or sloping. Will usually refer only to mud.

Flat. Must only be used to describe level surfaces (i.e. no significant gradient).

Sloping. Refers to any area where there is a general trend in the depth of the seafloor, i.e. a sea floor gradient. A sloping seafloor may be smooth but cannot be flat.

Gentle. Gradual, slowly changing.

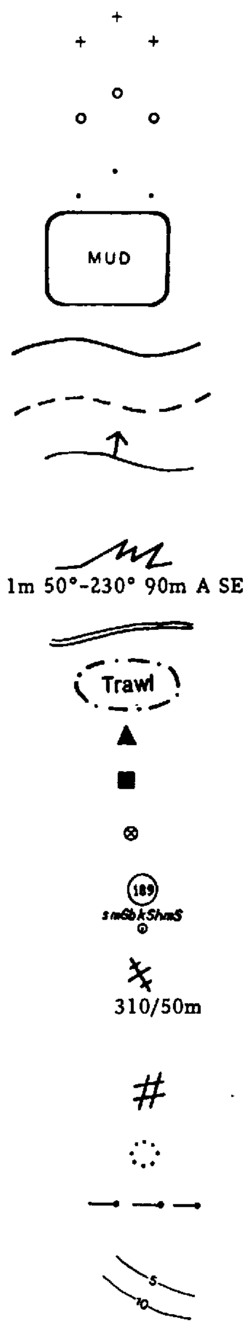
Regular. Used to qualify a series of features which are uniform in amplitude and wavelength, i.e. sandwaves, ridges.

Irregular. Used to qualify features which are not uniform but do have a specific entity, i.e. sandwaves. Can also be used to describe an area of rock where no regular structure is evident.

Prominent. Used to describe a feature or series of features which is or are very obvious in relation to their general surroundings.

Featureless. Applied normally to either a flat or smooth seafloor where the featureless aspect is either unusual or of considerable extent.

3.3.14.13 Symbols for use on Military Seafloor Classification Models are at Figure 4.15:



- Rock exposure
- Gravel
- Sand
- Mud
- Clearly defined seafloor texture boundary
- Ill-defined seafloor texture boundary
- Sand-wave crest with crest-to-trough height in metres, symbol denotes direction of steepest side
- Sand-wave areas (with height, orientation, wavelength and symmetry; A = Asymmetrical; S = Symmetrical)
- Prominent trawl scour
- Area of concentrated trawl scours
- Significant sonar contact
- Patch of highly reflective seafloor without significant height or extent
- Pockmark
- Position of seafloor sample with abbreviated description and identifying number if retained
- Wreck, orientated as appropriate, with direction and extent of scour
- Foul or obstruction (with classification if applicable)
- Submerged well
- Pipeline
- Seafloor contour at 5 m vertical intervals.

Fig. 4.15 "Symbols for use on Military Seafloor Classification Models"

REFERENCES

1987	*"The Use of Side Scan Sonar for Hydrographic Surveying and the Gathering of Bottom Texture Information"*	UK Hydrographic Department Professional Paper No. 24
1990	*"Sound Underwater Images" – A Guide to the Generation and Interpretation of Side Scan Sonar Data.*	Fish JP & Carr HA
1994	*"High Frequency Ocean Environmental Acoustic Models Handbook"*	Washington University Seattle Applied Physics Lab
1997	*"Admiralty Manual of Hydrographic Surveying"*	UK Hydrographic Office
1998	*"Side Scan Versus MBES Object Detection - A Comparative Analysis"*	Brissette MB & Hughes Clarke JE
2001	*"LEEUWIN Class Operating System"*	Australian Hydrographic Service
2004	*"Hydrographic Quality Assurance Instructions for Admiralty Surveys"*	UK Hydrographic Office

CHAPTER 5
WATER LEVELS AND FLOW

by Cdr. Jerry MILLS and Mr. Stephen GILL (USA)

1. INTRODUCTION

The purpose of this chapter is to provide the hydrographer and technical reader the fundamental information required to understand and apply water levels, derived water level products and datums, and water currents to carry out field operations in support of hydrographic surveying and mapping activities. The hydrographer is concerned not only with the elevation of the sea surface, which is affected significantly by tides, but also with the elevation of lake and river surfaces, where tidal phenomena may have little effect. The term 'tide' is traditionally accepted and widely used by hydrographers in connection with the instrumentation used to measure the elevation of the water surface, though the term 'water level' would be more technically correct. The term 'current' similarly is accepted in many areas in connection with tidal currents; however water currents are greatly affected by much more than the tide producing forces. The term 'flow' is often used instead of currents.

Tidal forces play such a significant role in completing most hydrographic surveys that tide producing forces and fundamental tidal variations are only described in general with appropriate technical references in this chapter. It is important for the hydrographer to understand why tide, water level and water current characteristics vary both over time and spatially so that they are taken fully into account for survey planning and operations which will lead to successful production of accurate surveys and charts.

Because procedures and approaches to measuring and applying water levels, tides and currents vary depending upon the country, this chapter covers general principles using documented examples as appropriate for illustration.

2. TIDES AND WATER LEVELS

2.1 Principles of Tides and Water Levels

The observed tides at any given port are the result of many factors, including the response of the ocean basin to the tide producing forces, to the modifications of the tide due to shallow water effects of local embayments and rivers, to the regional and local effects of weather on water levels.

2.1.1 Astronomical Tide Producing Forces

At the surface of the Earth, the Earth's gravitational attraction acts in a direction inward toward its centre of mass and thus holds the ocean waters confined to this surface. However, the gravitational forces of the Moon and Sun, and centrifugal force of the Sun-Earth-Moon system, act externally upon the Earth's ocean waters. These external forces are exerted as tide-producing, or tractive forces. Their affects are superimposed upon the Earth's gravitational force and act to draw the ocean waters horizontally to various points on the Earth's surface.

A high tide is produced in ocean waters by the 'heaping' action resulting from the horizontal flow of water toward the region of maximum attraction of the combined lunar and solar gravitational forces. An additional high tide is produced at a position on the opposite side of the Earth, where the centrifugal force of the orbiting system overpowers the gravitational attraction of the Sun and Moon. Low tides are created by a compensating withdrawal of water from regions around the Earth midway between these two tidal bulges. The alternation of high and low tides is caused by the daily (or *diurnal*) rotation of the solid body of the Earth with respect to these two tidal bulges and the tidal depression. The changing arrival times of any two successive high or low tides at any one location are the result of numerous factors. Fundamental tide producing forces have two components due to the Sun (solar) and the Moon (lunar).

2.1.1.1 Origin of Tide -Producing Forces

It appears to an observer on the Earth that the Moon revolves around the Earth, but in reality the Moon and the Earth each revolve around their common centre of mass known as the *barycentre*. The two astronomical bodies tend to be pulled together by gravitational attraction and simultaneously thrown apart by centrifugal force produced as they revolve around the barycentre. The gravitational attraction and centrifugal force are equal in magnitude and opposite in direction; thus, the Earth and Moon are neither pulled toward each other, nor are they further separated from each other. There is a similar effect for the Earth and Sun system, but they are separate and distinct from those of the Earth and Moon (thus the lunar and solar components).

These gravitational and centrifugal forces are balanced only at the centres of mass of the individual bodies. At points on, above or within the bodies, the two forces are not in equilibrium, resulting in tides of the ocean, atmosphere and lithosphere. On the side of the Earth turned toward the Moon or Sun, a net (or *differential*) tide-producing force acts in the direction of the Moon or Sun's gravitational attraction or toward the Moon or Sun. On the side of the Earth directly opposite the Moon or Sun, the net tide-producing force acts in the direction of the greater centrifugal force or away from the Moon or Sun.

2.1.1.2 Centrifugal Force

The barycentre of the Earth/Moon system lies at a point approximately 1,700 km beneath the Earth's surface, on the side toward the Moon, and along a line connecting the individual centres of mass of the Earth and Moon (Figure 5.1). The centre of mass of the Earth describes an orbit (E_1, E_2, E_3...) around the barycentre (G) just as the centre of mass of the Moon describes its own monthly orbit (M_1, M_2, M_3...) around this same point.

As the Earth revolves around the barycentre, the centrifugal force produced at the Earth's centre of mass is directed away from the barycentre in the same manner in which an object whirled on a string around one's head exerts a tug upon the restraining hand. Because the centre of mass of the Earth is on the opposite side of the barycentre to the Moon, the centrifugal force produced at the Earth's centre of mass is directed away from the Moon. All points in or on the surface of the Earth experience the same magnitude and direction of this centrifugal force. This fact is indicated by the common direction and length of the arrows representing the centrifugal force (F_c) at points A, B and C in Figure 5.1 and the thin arrows at these same points in Figure 5.2. In a similar fashion, the barycentre of the Earth/Sun system lies at a point well within the Sun because the Sun is so massive relative to the Earth; however the same theory applies.

It is important to note that the centrifugal force produced by the daily rotation of the Earth on its own axis is of no consequence in tidal theory. This element plays no part in the creation of the differential tide-producing forces because the force at any particular location remains constant with time, so the water surface is always in equilibrium with respect to it.

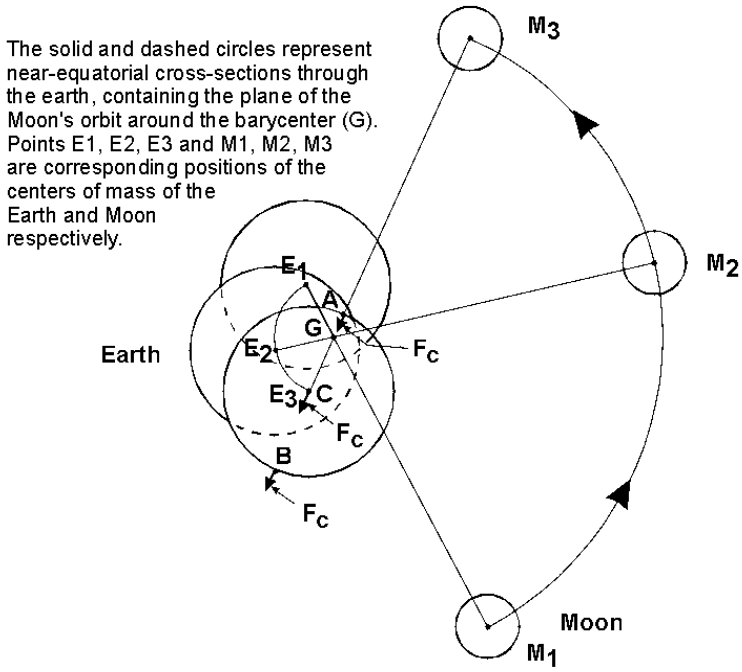
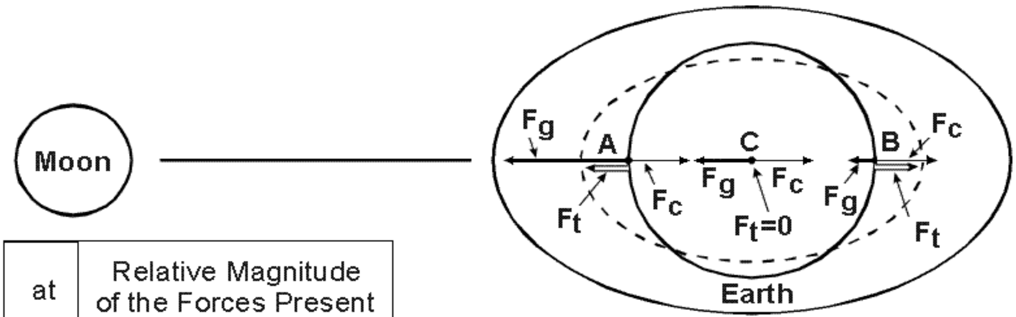


Fig. 5.1

Type of Force	Designation
F_c = centrifugal force due to Earth's revolution around the barycenter	thin arrow
F_g = gravitational force due to the Moon	heavy arrow
F_t = the resultant tide-raising force due to the Moon	double shafted arrow



at	Relative Magnitude of the Forces Present
A	$F_g > F_c > F_t$
	$\vee \quad \parallel \quad \vee$
C	$F_g = F_c > 0$
	$\vee \quad \parallel \quad \wedge$
B	$F_g < F_c > F_t$

A north-south cross-section through the Earth's center in the plane of the Moon's hour angle; the dashed ellipse represents a profile through the spheroid composing the tidal force envelope; the solid ellipse shows the resulting effect on the Earth's waters.

Fig. 5.2

2.1.1.3 Gravitational Force

While the affect of the external centrifugal force is constant at all points on the Earth, the affect of gravitational force produced by another astronomical body changes from place to place. This is because the magnitude of the gravitational force exerted varies with the distance of the attracting body. Thus, in the theory of tides, another variable influence is introduced, based upon the different distances of various points on the Earth's surface from the Moon's centre of mass. The relative gravitational attraction exerted by the Moon at various positions on the Earth is indicated in Figure 5.2 by arrows labelled F_g which are heavier than those representing the centrifugal force components.

Analogous to the fact that the Earth's own centrifugal force plays no part in producing tides, the effects of the Earth's own gravitational force plays no direct part in the origin of tides. Again, this is because the Earth's gravitational force at any particular location remains constant with time.

2.1.1.4 Differential Tide-Producing Forces

The centrifugal force acting on the centre of the Earth as the result of its revolving around the barycentre is equal and opposite to the gravitational force exerted by the Moon on the centre of the Earth. This is indicated at point C in Figure 5.2 by the thin and heavy arrows of equal length, pointing in opposite directions. The net result of this circumstance is that the tide-producing force (F_t) at the Earth's centre of mass is zero.

The *sublunar point*, point A in Figure 5.2, is approximately 6,400 km nearer to the Moon than point C. Here, the force produced by the Moon's gravitational pull is larger than the gravitational force at C due to the Moon; because the centrifugal force is everywhere equal and opposite to the gravitational pull of the Moon at the Earth's centre of mass, the larger gravitational pull at point A overpowers the centrifugal force, for a net force in the direction of the Moon. This is indicated in Figure 5.2 by the double-shafted arrow. The resulting tide produced on the side of the Earth toward the Moon is known as the *direct tide*.

On the opposite side of the Earth, the *antipodal point*, point B, is about 6,400 km farther from the Moon than point C, the Moon's gravitational force is less than at C; because the centrifugal force at point B is greater than the gravitational attraction of the Moon at point C, the resultant tide-producing force at this point is again directed away from the Earth's centre. This force is indicated by the double-shafted arrow at point B. The tide produced in the antipodal point is known as the *opposite tide*

There is also a similar set up of differential forces in the Earth/Sun system.

2.1.1.5 Tractive Force

Tide-producing forces have a magnitude of only about one 9-millionth of the Earth's own gravitational attraction. The tide-producing forces, therefore, are wholly insufficient to noticeably lift water against the pull of the Earth's gravity. Instead, the tides are generated by the horizontal component of the tide-producing forces. At any point on the Earth's surface, the tide-producing force may be resolved into two components – one vertical, or perpendicular to the Earth's surface, and the other horizontal, or tangential to the Earth's surface; because the horizontal component is not opposed in any way to gravity, it can act to draw particles of water freely over the Earth's surface toward the sublunar and antipodal points.

The horizontal component, known as the tractive ('drawing') component of force, is the actual mechanism for producing the tides. The tractive force is zero at the sublunar and antipodal points, because the tide-producing force is entirely vertical at these points; thus, there is no horizontal component. Any water accumulated in these locations by tractive flow from other points on the Earth's surface tends to remain in a stable configuration, or tidal 'bulge'. Thus, there exists a constant tendency for water to be drawn from other points on the Earth's surface toward the sublunar point (A in Figure 5.2) and its antipodal point (B in Figure 5.2), and to be heaped at these points in two tidal bulges. As a special case of Newton's Universal Law of Gravitation, the tide-producing force varies inversely as the third power of the distance of the centre of mass of the attracting body from a given point on the surface of the Earth.

Within a band around the Earth roughly midway between the sublunar and antipodal points, the tractive force is also zero, because the tide-producing force is directed vertically. There is, therefore, a tendency for the formation of a stable depression in this region.

2.1.1.6 The Tidal Force Envelope

If the ocean waters were to respond exclusively to tractive forces and the Earth were covered with water without continents, the water surface would approximate the shape of a prolate spheroid. The major axis of the spheroid would lie along a line connecting the centres of mass of the Earth and Moon, the minor axis would be centred on, and at right angles to, the major axis. The two tidal humps and the tidal depression are represented in this force envelope by the directions of the major axis and rotated minor axis of the spheroid, respectively. From a purely theoretical point of view, the daily rotation of the solid Earth with respect to these two tidal bulges and the depression may be conceived to be the cause of the lunar tides. With respect to the Sun, the resulting bulges and depressions may be conceived to be the cause of the solar tides.

As the Earth rotates, one would ideally expect to find a high tide followed by a low tide at the same place 6 hours later, then a second high tide after 12 hours, and so forth. This would nearly be the case if a smooth, continent-free Earth were covered to a uniform depth with water, if the tidal force envelope of the Moon alone could be considered, the position of the Moon were fixed and invariable in distance and relative orientation with respect to the Earth and Sun and if there were no other accelerating or retarding influences affecting the motions of the waters of the Earth; however this is far from the actual situation.

First, the tidal force envelope produced by the affect of the Moon is accompanied by, and interacts with, a tidal force envelope produced by the Sun. The tidal force exerted by the Sun is a composite of the Sun's gravitational attraction and the centrifugal force created by the revolution of the Earth around the centre of mass of the Earth-Sun system. The position of this force envelope shifts with the relative orbital position of the Earth with respect to the Sun; due of the great difference between the average distances of the Moon and Sun from the Earth (384,400 km and 150,000,000 km, respectively), the tide-producing force of the Moon is approximately 2.5 times that of the Sun, even though the Sun is so much more massive than the Moon.

Second, there exists a wide range of astronomical variables in the production of the tides. Some of these are the changing distances of the Moon from the Earth and the Earth from the Sun, the angle which the Moon in its orbit makes with the Earth's equator, the angle which the Sun appears in the yearly orbit of the Earth about the Sun and the variable phase relationships of the Moon with respect to the Sun and Earth. Some of the principal types of tides resulting from these purely astronomical influences are described below.

Third, other affects come into play, causing the water level to differ from the astronomically induced tide. These include restrictions to the flow of water caused by the continents and meteorological affects, among others.

2.1.2 Tidal Characteristics

The actual characteristics of the tide at locations around the Earth differ significantly from the idealised tidal envelope discussed previously. First of all, water is a somewhat viscous fluid, which lags in its response to the tide-generating forces. More significantly, the Earth is not a smooth sphere with uniformly deep water covering its entire surface. Tidal movements are affected by friction with the ocean floor and with other ocean currents; the continents interrupt, restrict and reflect tidal movements; the shape and size of the ocean basins accentuate or dampen various components of the tide producing forces.

The rise and fall of the tide does not occur at a uniform rate. From low water, the tide begins rising very slowly at first, but at a constantly increasing rate for about 3 hours when the rate of rise is at a maximum. The rise continues for about 3 more hours, but at a constantly decreasing rate until high water. The falling tide follows a similar pattern of increasing then decreasing rate. When the rise and fall of the tide is represented graphically, it can be seen to approximate the form of a sine curve. At any location, however, the rise and fall of the tide, and consequently the shape of the curve, will be characterised by a variety of features. These features will vary considerably from location to location. Of these features, three may be considered as constituting the principle characteristics of the tide. These three are the time of tide, the range of tide and the type of tide. The hydrographer must understand and consider each of these three characteristics in order to compute and apply tidal reductions to soundings.

2.1.2.1 Time of Tide

A stationary moon would appear to cross the meridian at any given place once every day; but, because the Moon revolves around the Earth in the same direction in which the Earth is rotating, any point on the Earth must actually rotate approximately 12.5° extra each day to catch up with the Moon. This 12.5° requires about 50 minutes, resulting in a 'tidal day' of 24 hours and 50 minutes.

The time of tide refers to the time of occurrence of high or low water with respect to the Moon's meridian passage. This characteristic of the tide at a specific place is described by the high and low water *lunitidal* intervals. The lunitidal interval is the elapsed time between a meridian passage of the Moon and the high or low water. Lunitidal intervals are not constant along any given meridian. The intervals vary, based upon the interruption of the tidal wave by land masses and by the resistance of the ocean floor as the wave moves into shallow water.

Even at any given place, the intervals are not constant, but they do vary periodically within relatively narrow limits. This limited variation in lunitidal interval results from the interaction between the tidal forces of the Moon and the Sun. Between new Moon and first quarter, and between full Moon and third quarter, this interaction can cause an acceleration in tidal arrival times. Between first quarter and full Moon, and between third quarter and new Moon, the interaction can result in a lagging of the tidal arrival.

Lunitidal intervals are defined both in terms of the Moon's meridian passage over Greenwich and the Moon's meridian passage over the local longitude. They are known respectively as the Greenwich lunitidal interval and the local lunitidal interval. Greenwich intervals are more useful in that they can be used to relate the times of tide at one place to those at any other place. The time of tide is important in analysis and prediction of the tides and in the computation of tidal zoning correctors.

2.1.2.2 Tidal Range

The difference in height between consecutive high and low tides occurring at a given place is known as the range. In the open ocean, the actual height of the tidal wave crest is relatively small (usually 1m or less) and uniform. It is only when the tidal crests and troughs move into shallow water, against land masses, and into confining channels that large tidal ranges and noticeable variations in range are apparent.

The range of tide at a particular location is not constant, but varies from day to day. Part of this variation is caused by the effects of wind and weather, but mostly it is a periodic phenomenon related to the positions of the Sun and the Moon relative to the Earth. In this day to day change, the tide responds to three variations, each associated with a particular movement of the Moon.

Lunar Phase Effects: Spring and Neap Tides – At most places, the Moon's phase has the greatest affect on the range of the tide. It has been noted that the tides originate from the combined effects of tractive forces generated by both the Sun and the Moon; because of the Moon's changing position with respect to the Earth and Sun (Figure 5.3) during its monthly cycle of phases, tractive forces generated by the Moon and Sun act variously along a common line and at changing angles relative to each other.

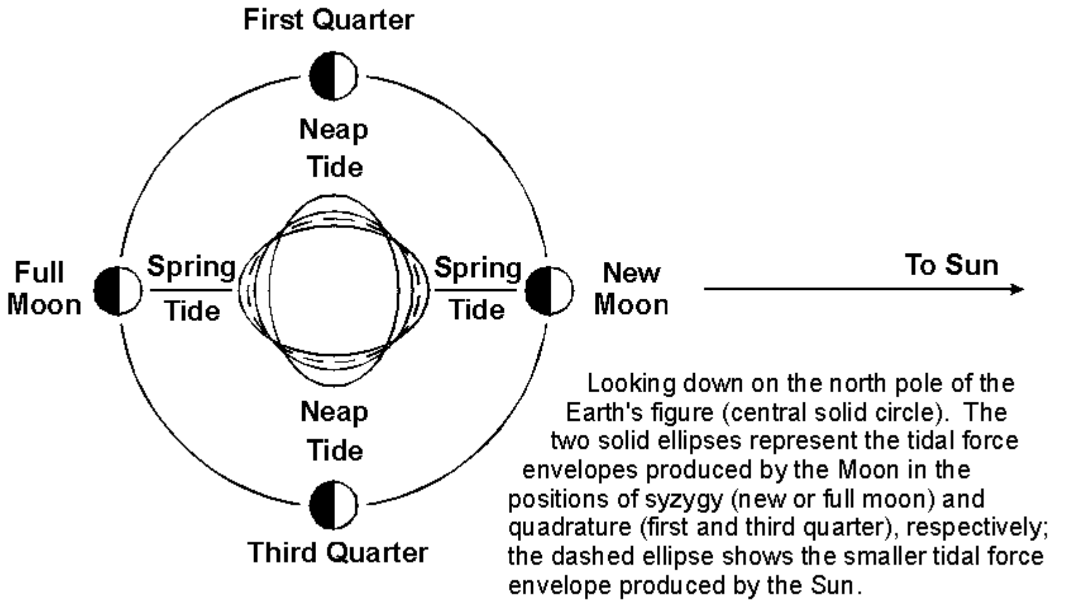


Fig. 5.3

When the Moon is at new phase and full phase (both positions being called *syzygy*), the gravitational attractions of Moon and Sun act to reinforce each other; as the resultant or combined tidal force is also increased, the high tides are higher than average and low tides are lower. This means the tidal range is greater at all locations which display a consecutive high and low water. Such greater-than-average tides resulting at the syzygy positions of the Moon are known as *spring tides* – a term which merely implies a 'welling-up' of the water and bears no relationship to the season of the year.

At first- and third-quarter phases (*quadrature*) of the Moon, the gravitational attractions of the Moon and Sun upon the waters of the Earth are exerted at right angles to each other. Each force tends in part to counteract the other. In the tidal force envelope representing these combined forces, both the maximum and minimum force values are reduced. High tides are lower than average and low tides are higher. Such tides of diminished range are called *neap* tides, from a Greek word meaning 'scanty'.

Parallax Effects (Moon and Sun) – Since the Moon follows an elliptical path (Figure 5.4), the distance between the Earth and Moon will vary throughout the month by about 50,000 km. The Moon's gravitational attraction for the Earth's waters will change in inverse proportion to the third power of the distance between Earth and Moon, in accordance with the previously mentioned variation of Newton's Law of Gravitation. Once each month, when the Moon is closest to the Earth (*perigee*), the tide-generating forces will be higher than usual, thus producing above-average ranges in the tides. Approximately 2 weeks later, when the Moon (at *apogee*) is farthest from the Earth, the lunar tide-producing force will be smaller and the tidal ranges will be less than average. Similarly, in the Sun-Earth system, when the Earth is closest to the Sun (*perihelion*), about January 2 of each year, the tidal ranges will be enhanced, and when the Earth is farthest from the Sun (*aphelion*), around July 2, the tidal ranges will be reduced.

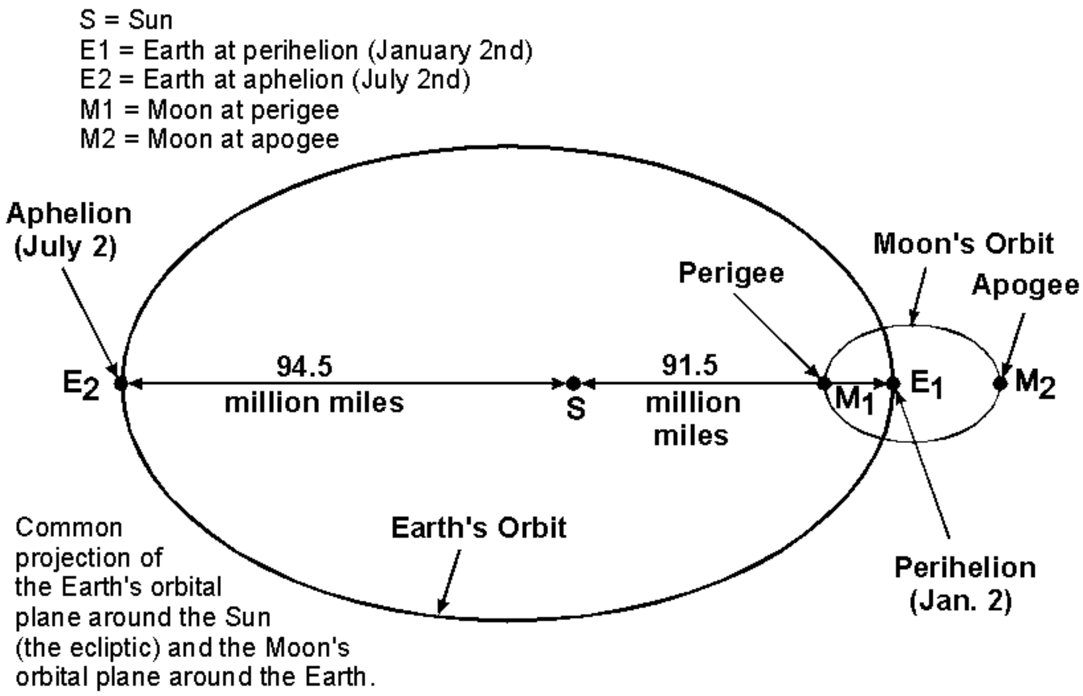


Fig. 5.4

When perigee, perihelion, and either the new or full Moon occur at approximately the same time, considerably increased tidal ranges result. When apogee, aphelion, and the first- or third-quarter Moon coincide at approximately the same time, considerably reduced tidal ranges will normally occur.

Lunar Declination Effects: The Diurnal Inequality – The plane of the Moon's orbit is inclined only about 5° to the plane of the Earth's orbit (the *ecliptic*) and thus the Moon in its monthly revolution around the Earth remains very close to the ecliptic.

The ecliptic is inclined 23.5° to the Earth's equator, north and south of which the Sun appears to move once each half year to produce the seasons. In a similar fashion, the Moon, in making a revolution around the Earth once each month, passes from a position of maximum angular distance north of the equator to a position of maximum angular distance south of the equator during each half month. This is *declination*.

Twice each month, the Moon crosses the equator. In Figure 5.5, this condition is shown by the dashed position of the Moon. The corresponding tidal force envelope due to the Moon is depicted, in profile, by the dashed ellipse. Tides occurring when the Moon is near the equator are known as *equatorial tides*, while those occurring when the Moon is near its maximum northern or southern declination are known as *tropic tides*.

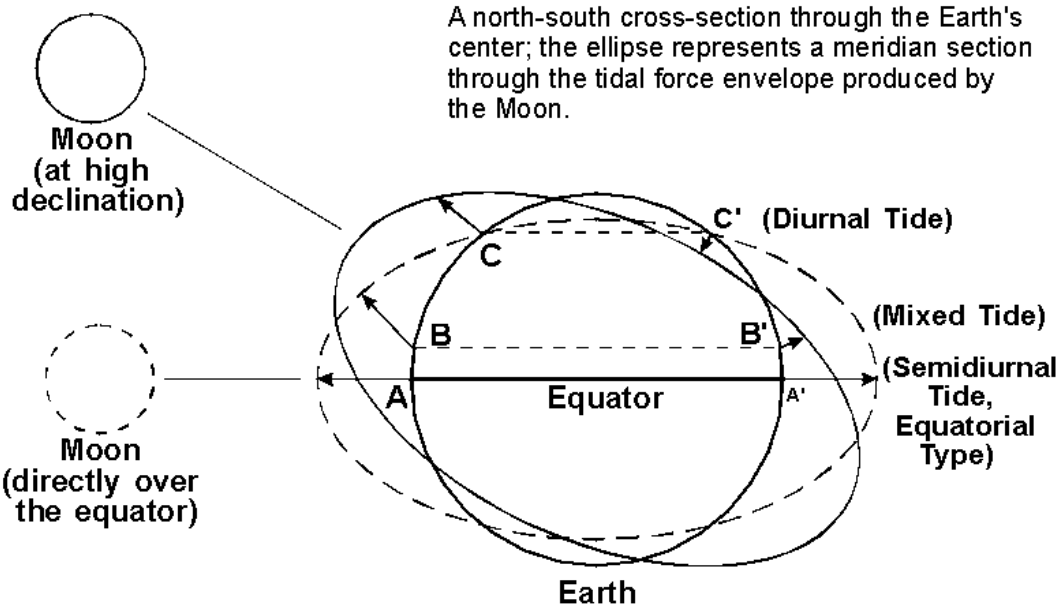


Fig. 5.5

Variability – The effects of phase, parallax and declination are not exhibited everywhere in equal measure, though all three do occur on all parts of the Earth. Phase inequalities are most commonly greatest, but in any particular area any one of the three variations may exert the predominant influence on the range variation of the tide. The month of the Moon's phases, the *synodic month*, is approximately 29.5 days; the month of the Moon' s distance, the *anomalistic month*, is approximately 27.5 days; and the month of the Moon's declination, the *tropic month* is approximately 29.3 days. It follows, therefore, that considerable overall variation in the range of the tide occurs at any place as a result of the progressively changing relations of the three variations to each other. At Seattle, for example, the mean range of tide is about 2.3 m, but individual ranges within a single day may vary from less than 1.5 m to more than 4.5 m.

The range of tide is subject to other periodic variations (for example, the solar parallax differences noted earlier), but the three discussed above are the principal variations. The fact all these major variations cycle completely in 29.5 days or less is the main reason why the hydrographer must operate key tide stations for a minimum of 30 days. Although the variation in range from one 30-day period to another will change somewhat, any 30 consecutive days obtained in combination with long-term stations will usually suffice for preparation of hydrographic tide reductions. One important long-term deviation in range of tide is due to a slowly varying change in orientation of the Moon's orbit called the regression of the Moon's nodes. This variation results in a measurable corresponding slow difference in the range of tide. This deviation gives ride to the need to use nodal factors or modal corrections when performing harmonic analyses or tide predictions and is important in the determination of various tidal datums (see section 2.1.4).

2.1.2.3 Types of Tide

Of the three principal tidal characteristics, type of tide is the most fundamental. If the tides at two places are of the same type, but differ in time or in range, the tide at one place can be related simply and accurately to the tide at the other location. This similarity underlies the hydrographer's ability to extend sounding datums and compute accurate water-level reducers in areas where a relatively short series of tidal observations have been obtained. On the other hand, if the type of tide at the two places differs, the fact the time or range may be the same does not necessarily indicate a simple relationship between the two places. Differences in time and range of tide are merely differences in degree, but differences in type of tide are differences in the basic nature of the tide.

The type of tide refers to the characteristic form of the rise and fall of the tide as revealed by the tide curve. Although the tide curve for any particular place will differ in some respects from that of any other place, tide curves may be grouped into three large classes or types. These types are semidiurnal, diurnal, and mixed tides.

Referring to Figure 5.5, as the points A and A' lie along the major axis of this ellipse, the height of the high tide represented at A is the same as that which occurs as this point rotates to position A' some 12 hours later. When the Moon is over the equator – or at certain other force-equalizing declinations – the two high tides and two low tides on a given day are similar in height at any location. Successive high tides and low tides are then also nearly equally spaced in time and occur uniformly twice daily. This is known as the *semidiurnal* type of tide. A semidiurnal curve of tidal height versus time is shown in the top diagram of Figure 5.6. The semidiurnal type of tide is one in which the full cycle of high and low water is completed in half a day. There are two high and two low waters in each lunar day of 24 hours 50 minutes. To be classified as a semidiurnal tide, the two daily tidal cycles must resemble each other such that, although they are not identical, the two highs do not differ much and the two lows do not differ much.

However, with the changing angular distance of the Moon above or below the equator (as shown in Figure 5.5), the tidal force envelope produced by the Moon is canted, and differences between the heights of two daily tides of the same phase begin to occur. Variations in the heights of the tides resulting from the changes in the declination angle of the Moon and in the corresponding lines of gravitational force action give rise to a phenomenon known as the *diurnal inequality.*

In Figure 5.5, point B is beneath a bulge in the tidal envelope. One-half day later, at point B', it is again beneath the bulge, but the height of the tide is not as great as at B. This situation gives rise to a twice-daily tide displaying unequal heights in successive high or low waters, or in both pairs of tides. This type of tide, exhibiting a strong diurnal inequality, is known as a *mixed tide*. (See the middle diagram in Figure 5.6.) The mixed type of tide is one in which two highs and two lows occur each day, but in which there are marked differences between the two high waters or between the two low waters of the day. This type of tide is named a mixed tide because it has the properties of a mixture of semidiurnal and diurnal tides.

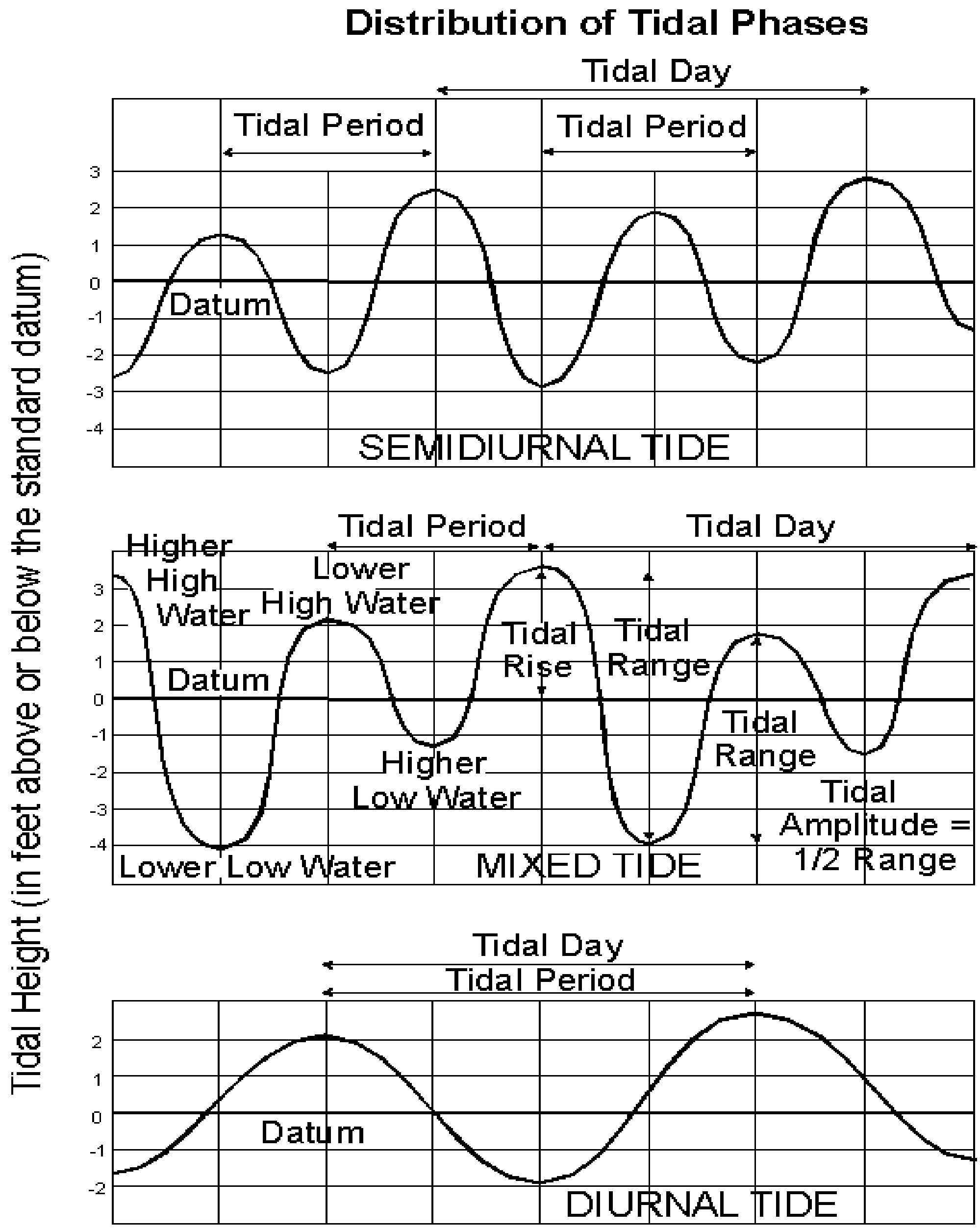


Fig. 5.6

Finally, as depicted in Figure 5.5, the point C is seen to lie beneath a portion of the tidal force envelope. One-half day later, however, as this point rotates to position C', it is seen to lie above the force envelope. At this location, therefore, the tidal forces present produce only one high water and one low water each day. The resultant diurnal type of tide is shown in the bottom diagram of Figure 5.6. The diurnal type of tide describes those tides in which one high and one low water occur in a lunar day. In this type of tide, the period of the rise, and also of the fall, of tide is approximately 12 hours as opposed to the 6 hour periods of semidiurnal tides.

Examples of each of the three types of tide are shown in Figure 5.6, using three days of tidal records from Hampton Roads, Virginia; Pensacola, Florida and San Francisco, California. The horizontal line through each curve represents the mean level of the sea and the magnitude of the rise and fall of the tide above and below mean sea level is indicated by the scale on the left.

The upper curve, for Hampton Roads, illustrates the semidiurnal type of tide. Two high and two low waters occurred each day, with the morning and evening tides differing relatively little. The bottom curve, for Pensacola, illustrates the diurnal type of tide, one high and one low occurring each day. The curve for San Francisco illustrates one form of the mixed type of tide. Two high and two low waters occurred each day, but the morning tides differ considerably from the afternoon tides. In this particular case, the difference is seen in both the high and low waters.

The difference between corresponding morning and afternoon tides, or diurnal inequality, arises primarily from the fact that the Moon's orbit is inclined to the plane of the equator. This inclination results in the existence of both diurnal and semidiurnal tide producing forces. These forces affect the actual rise and fall of water level to differing degrees in different places, mostly as a result of local and basin response to the forces and thus result in different magnitudes of diurnal inequality. In fact, the distinction between mixed tides and semidiurnal tides is based entirely on this difference in magnitude.

Refer to the tide curve for San Francisco in Figure 5.6, which illustrates a mixed tide. While there is considerable inequality in both the high and low waters, the inequality of the low waters is greater. In Hampton Roads, the inequality, though not very large is exhibited primarily in the high waters. As Figure 5.7 illustrates, the inequality may be featured primarily in the low waters, primarily in the high waters or it may appear equally in the highs and the lows. It is also significant that the diurnal inequality is a feature of the time of tide as well as of the height of tide. Just as the inequality of the height varies from place to place and day to day, the duration of the rise and fall and the lunitidal interval also vary.

To distinguish the two tides of the day, definite names have been given to each of the tides. Of the two highs, the higher is called 'higher high water' (HHW) and the lower, 'lower high water' (LHW). Similarly, the two lows are called 'lower low water' (LLW) and 'higher low water' (HLW). (See Figure 5.6.) As a measure of the inequality, the terms 'diurnal high water inequality' (DHQ) and 'diurnal low water inequality' (DLQ) are used. The DHQ is defined as half the difference between mean higher high water and mean lower high water and the DLQ is defined as half the difference between the means of lower low water and higher low water.

These may be more meaningfully understood as the difference between the mean high water and the mean higher high water, and the difference between the mean low water and the mean lower low water, respectively.

Examination of a month's series of tides, as in the curves shown in Figure 5.7, will show that the diurnal inequality also varies in magnitude in relation to the Moon's declination, the inequality being the least when the Moon is near the equator, as it was in this month from the 3rd to the 5th and the 18th to the 20th, and being the greatest when the Moon is near maximum north or south declination, as it was from the 11th to the 13th and the 25th to the 27th.

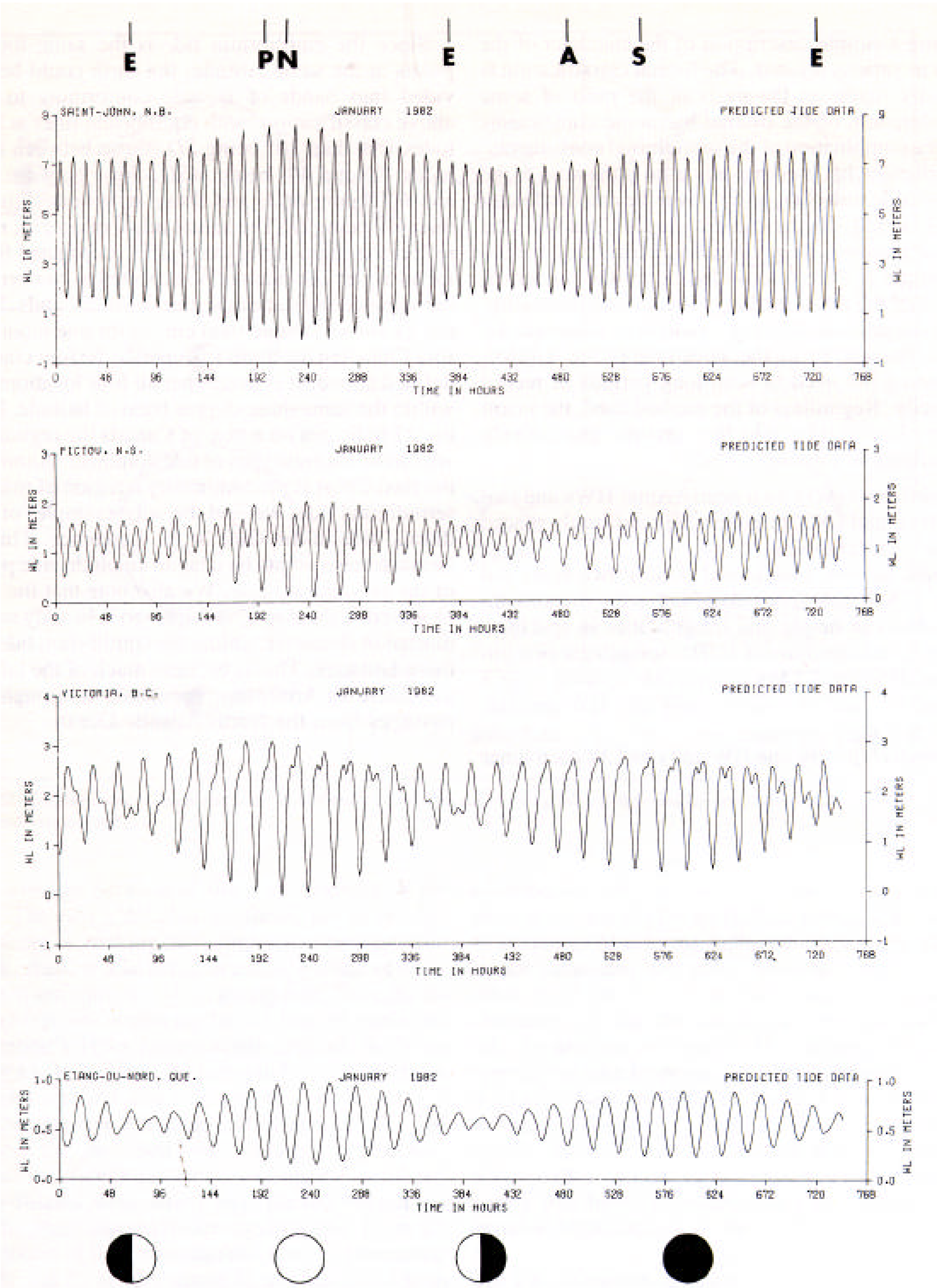

E
PN
E
A
S
E
SAINT-JOHN, N.B.
JANUARY 1982
PREDICTED TIDE DATA
WL IN METERS
TIME IN HOURS
PICTOU, N.S.
JANUARY 1982
PREDICTED TIDE DATA
VICTORIA, B.C.
JANUARY 1982
PREDICTED TIDE DATA
ETANG-DU-NORD, QUE.
JANUARY 1982
PREDICTED TIDE DATA

Fig. 5.7

2.1.2.4 Basin and Coastal Effects

Although it is the combined forces of the Sun and Moon which set the tidal wave in motion, it is often the size and shape of the ocean basin which controls the characteristics of the tide. For example, Pensacola, Florida is certainly not near the pole – the region to which Figure 5.5 would restrict diurnal tides – but it nonetheless has an unmistakable diurnal tide. Similarly, San Francisco and Hampton Roads are at about the same latitude, but have distinctly different tidal characteristics. Much like the water in a bathtub can be set to surging resonantly from end to end, the tidal oscillations in an ocean basin or a restricted sea can be accentuated by the basin's natural period of resonance. The Pacific Ocean basin accentuates the diurnal component of the tides, resulting in diurnal or strongly mixed tides. The Atlantic, on the other hand, accentuates the semidiurnal component of the tides. As a more localized example, much of the northern Gulf of Mexico responds primarily to the diurnal components of the tide.

The predominate type of tide can change over relatively short distances. For example, the east coast of Florida exhibits semidiurnal tides, most of the west coast of Florida exhibits mixed tides and most of the panhandle of Florida has diurnal tides.

Certain coastal and seafloor configurations can greatly increase the range of tide. Much like a wind wave crests and breaks as it approaches a beach, a tidal wave also increases in height as it encounters shallow water and constricting shoreline. Examination of tidal ranges at coastal stations will almost always show increasing range as the tide moves up a bay or inlet. Cook Inlet, Alaska, a funnel-shaped and shelving body of water, is a particularly good example. As the tide moves in from the Gulf of Alaska, its range increases from about 3 m at the entrance to about 10 m near the head at Anchorage. Above Anchorage, in Turnagain Arm, the inlet narrows and shoals even more. At the times of highest tides, the rising tide flows up the arm against the last of the ebb and rises into a vertical wall of water reaching nearly 2 m high. Such a tidally generated wall of water is known as a *tidal bore*. Tidal bores occur in several rivers or estuaries around the world where the range of tide is great and the configuration of the shore and seabed is ideal.

2.1.3 Non-Tidal Water Level Variations

Changes in the observed water level along the coasts are due not only to tidal forces but are also driven by variety of other forces over a wide-range of time scales. At the highest frequencies, water levels can be affected by tsunamis, seiche and storm surge. Local wind and barometric pressure changes can have a large affect, especially in shallow water. Wind set-up from onshore winds and low barometric pressure will generally cause water levels to above those predicted while offshore winds and high barometric pressure tend to have the opposite effect. Strong seasonal meteorological patterns will have affects on monthly sea levels. ENSO (El Niño Southerly Oscillation) affects on monthly mean sea levels in the Pacific Ocean are particularly noticeable. Short-term and seasonal affects are also found in tidal estuaries with strong river flows and are driven by run-off characteristics of individual watersheds and controlling dams upstream. The Great Lakes and other large lakes are sensitive to the annual evapo-transpiration cycles and the net gain or loss of water volume. Seasonal variations in oceanic circulation patterns and deviations in ocean eddies also may affect coastal levels. Depending on the spatial scale of the meteorological event, the effects can be seen basin-wide, regionally or only locally. The hydrographer needs to be generally aware of these dependencies in planning and conducting survey operations and to distinguish any anomalies in water level measurements due to weather or natural causes versus gauge malfunction.

2.1.4 Tide and Water Level Datums

The hydrographer must be able to relate all measured depths, regardless of the stage of tide or level of water at the time of sounding, to a common plane or datum. The datum used to reckon heights or depths for marine applications is a vertical datum called a 'water level datum'. For tidally derived datums, most datums are computed over, or referred to, specific 19 year periods or tidal datum epochs. The 19 year period is important as discussed in section 2.1.2.2, because of the 19 year modulation of lunar constituents by the long-term variation in the plane of the Moon's orbit called the regression of the Moon's nodes.

The water level datum to which the soundings on any particular survey are referred is known as the 'sounding datum'. The datum to which depths on a chart are referred is known as the 'chart datum'. A water level datum is called a 'tidal datum' when defined in terms of a certain phase of tide. In United States coastal waters, Mean Lower Low Water (MLLW) is used for both sounding and chart datums. MLLW is computed from tabulation of the observations of the tide, in this case the average of the lower low waters each tidal day over a 19 year period. The United States presently refers all tidal datums computed from tide observations to the 1983-2001 National Tidal Datum Epoch (NTDE) and updates to new NTDEs only after analysis of relative mean sea level change. In contrast, some Chart Datums are derived from harmonic analyses of observations and constructing time series of tide predictions over 19 year periods. The Canadian Chart Datum is the surface of Lower Low Water, Large Tide or LLWLT which encompasses the previously used datum of Lowest Normal Tide (LNT). British Charts now use a Chart Datum of Lowest Astronomical Tide (LAT) based on the lowest predicted tide expected to occur in a 19 year period. LAT is determined at a particular location by performing a harmonic analysis of the observations, then using the resulting harmonic constituents in a prediction equation to predict the elevation of the lowest predicted tide to occur over a 19 year period. Use of LAT has been adopted for international use by the International Hydrographic Organisation (IHO). Harmonic analyses have also been used to determine other Chart Datums. Chart Datums used on some older British Admiralty Charts were Mean Low Water Springs (MLWS) and Indian Spring Low Water (ISLW). MLWS and ISLW are derived from summations of the amplitudes of various major harmonic constituents below local Mean Sea Level.

In areas where there is very little or no tide, other water level datums are used. For the Black Sea, Mean Sea Level or Mean Water level is used. In the Great Lakes, both Canada and the United States use a separate fixed Low Water Datum (LWD) for each lake based on analyses of monthly means during low water stages. In non-tidal lagoons and bays in the coastal United States where the area transitions from tidal to non-tidal, a LWD is used which is determined by subtracting 0.2 m from the local Mean Sea Level derived from observations and adjusted to a 19 year period.

There are a variety of local Chart Datums employed in tidal rivers as well. In the United States, Chart Datums have been derived from analyses of measurements during the low river stages over a period of time and then are held fixed for charting purposes. Examples are the Hudson River Datum and Columbia River Datum derived from MLLW based on observations during the lowest river stages during the year.

The water level datum is a local plane of elevation which applies only in the specific area in which water level measurements have been made. Whether tidal or non-tidal, it is permanently referred to the land by levelling from the water level gauge to a local network of bench marks. Computational procedures for determining tidal datums are addressed later in this Chapter.

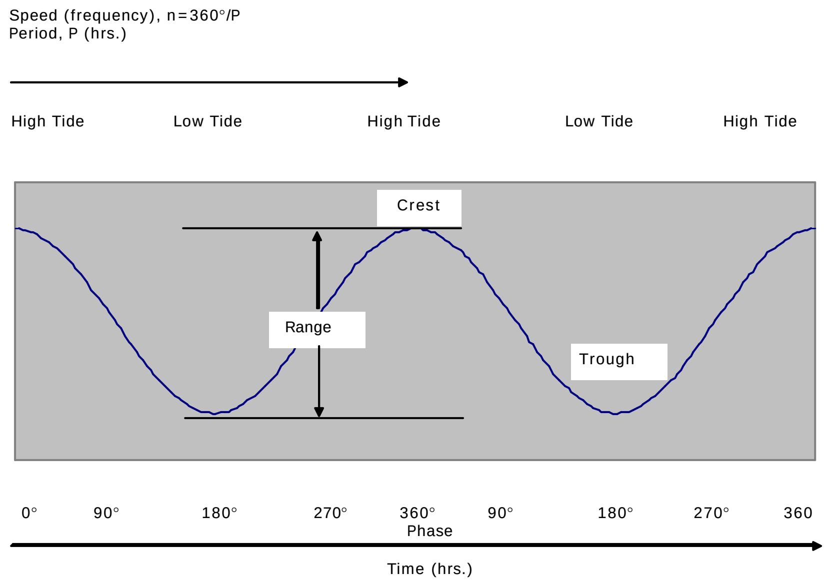
Water level datums are completely distinct from geodetic vertical datums. For instance, the United States and Canada use the North American Vertical Datum of 1988 (NAVD 88) and the International Great Lakes Datum of 1985 (IGLD 85) as the vertical datums for geodetic purposes. The relationship between NAVD 88 (or IGLD 85) and local mean sea level or mean water level varies considerably from place to place. In fact, it is impossible to transfer a tidal datum from one place to another with geodetic levelling, without considering local tidal conditions. The geodetic network does, however, establish relationships between the many tide stations and their tidal datum elevations around the North American continent, and it could be used to recover a connected local tidal datum if the tidal bench marks are destroyed. This requires level connections or GPS connections between geodetic and tidal bench mark networks.

2.1.5 Harmonic Analysis and Tide Prediction

Each of the tide-generating motions described in the above sections can be represented by a simple cosine curve as illustrated in Figure 5.8. The horizontal axis represents time and the vertical represents the magnitude of the tide-generating force. The crests give the times of the maximums in the tide-generating force and the troughs, the minimums. For example in Figure 5.8, in the Sun-Earth system, noon, with the Sun overhead, is at the first crest. Six hours later a minimum occurs at the trough. The second maximum is at midnight with the second crest. Another trough is for dawn and then back to the original noon crest.

Each one of the tide-generating motions, represented by a simple harmonic cosine curve, is known as a tidal component, tidal constituent, or harmonic constituent. A letter or letters and usually a subscript are used to designate each constituent. The tidal constituent described above, for example, called the Principal Solar semidiurnal constituent, is designated S_2. The Principal Lunar semidiurnal constituent is designated M_2. S is for sun and M is for moon and the subscript $_2$ means that there are two complete tidal cycles for each astronomic cycle. Thus, these are said to be semidiurnal constituents. Constituents are described by their tidal period (the time from maximum to maximum), **P**. The period for the S_2 is 12.00 solar hours (hr.) and the period for the M_2 is 12.42 solar hours:

$$\text{Speed (frequency), } n = 360°/P$$



From S . Hicks (2004)

Fig. 5.8 "The tide curve visualized as a wave-form"

In tidal work, each constituent (cosine curve) is more often described by its speed (or frequency in degrees per hour). The cosine curve is divided into 360° (from crest to crest). The speed **n** of the constituent is 360°/P. Thus, for S_2, n = 360°/12.00 = 30°/hr; for M_2, n = 360°/12.42 = 28.984°/hr.

There are an infinite number of constituents to describe almost all of the perturbations in the relative motions of the Sun, Moon, and Earth (including the distance and declinational aspects). However, after about 37, the effects of these motions in representing the actual tides are extremely small in most locations in the United States. For tidally complex areas inside estuaries, such as Anchorage, Alaska and Philadelphia, Pennsylvania, it takes over one hundred constituents to adequately describe the tide curve. These additional constituents are artefacts which combine the fundamental diurnal and semidiurnal constituents to produce high frequency (from 3 to 13 cycles per day) constituents which attempt to describe the complex non-linear affects of seabed friction and shallow water.

The representations of the various astronomical events and the development of their periods and speeds are essential in understanding the harmonic analysis techniques. The development of frictional, shallow water and compound tidal constituents are outside of the scope of this chapter.

The Principal Solar semidiurnal constituent, S_2, represents the Earth spinning relative to the Sun. The Earth rotates once in 24 mean solar hours or, since around the world is 360°; it is going at the rate of 360°/24 = 15°/hr. However, there is a maximum in the solar tide producing force under the Sun and again on the opposite side (midnight). So, the period (maximum to maximum) of the constituent is 12 mean solar hours and the speed is: S_2 360°/12 = 30°/hr.

The Principal Lunar semidiurnal constituent, M_2, represents the Earth spinning relative to the Moon. Since the Moon is moving eastward, it takes 24.8412 mean solar hours to bring the Moon back overhead. Again, there are two maximums in this lunar day, so the period is only 12.4206 mean solar hours and its speed is: M_2 360°/12.4206 = 28.984°/hr.

S_2 and M_2 get into phase (maxima lined up) and out of phase (maximum of one lined up with the minimum of the other) to produce spring and neap tides, respectively (Figure 5.3). Spring tides occur at the times of full moon and new moon while neap tides occur at the times of the first and third quarter moons. The revolution of the Moon around the Earth relative to the Sun takes 29.5306 days (called the synodic month or one lunation). Since there are two maximums, spring tides occur every 29.5306/2 = 14.765 days and neap tides 7.383 days later than the springs.

The Larger Lunar Elliptic semidiurnal constituent, N_2, and the Smaller Lunar Elliptic semidiurnal constituent, L_2, are two constituents designed to simulate the cycle of perigee to perigee. These are completely artificial constituents in contrast with S_2 and M_2 which have realistic relationships to the solar and lunar envelopes of the tide-generating forces. Perigee to perigee occurs every 27.5546 days (the anomalistic month) or 661.31 mean solar hours. The speed of perigee to perigee is thus 360°/661.31 = 0.544°/hr. This is a lunar event and the speed of M_2 is 28.984°/hr. The constituent speeds are, therefore:

$$N_2 \quad 28.984 - 0.544 = 28.440°/\text{hr.}$$
$$L_2 \quad 28.984 + 0.544 = 29.528°/\text{hr.}$$

Thus, when N_2 and L_2 are in phase every 27.5546 days (the anomalistic month) they add to M_2 to simulate the near approach to the Moon (perigee). Also, 13.7773 days later they are out of phase simulating apogee (the Moon farthest away).

The Luni-solar Declinational diurnal constituent, K_1, and the Principal Lunar Declinational diurnal constituent, O_1, are also artificial constituents designed to simulate the cycle of maximum declination to maximum declination of the moon. Maximum north to maximum north occurs every 27.3216 days (the tropical month) or 655.72 mean solar hours. However, both north and south declinations produce the same results. The north to south (and south to north) cycle is 655.72/2 = 327.86 hours. The speed is 360°/327.86 = 1.098°/hr. The speeds of the constituents, as they modify M_2, will be the speed of M_2 plus and minus the speed of the north to south cycle. Since the maximum is only felt once per day as the earth spins, the constituent speeds are half the sum and difference:

$$K_1 \quad (28.984 + 1.098)/2 = 15.041°/\text{hr.}$$
$$O_1 \quad (28.984 - 1.098)/2 = 13.943°/\text{hr.}$$

Thus, when K_1 and O_1 are in phase, every 13.6608 days (one half of the tropical month, i.e. the month with respect to the vernal equinox), they add to M_2 to simulate the maximum declination of the Moon north or south. They account for the diurnal inequality due to the Moon (the two high tides and/or the two low tides being unequal in height each tidal day) and, at extremes, diurnal tides (one high tide and one low tide each tidal day).

The Luni-solar Declinational diurnal constituent, K_1, and the Principal Solar Declinational diurnal constituent, P_1, are designed to simulate the cycle of maximum declination to maximum declination of the Sun. Maximum north to maximum north occurs every 365.2422 days (the tropical year) or 8765.81 mean solar hours. However, both north and south declinations produce the same results. The north to south (and south to north) cycle is 8765.81/2 = 4382.91 hrs. The speed is 360°/4382.91 = 0.082°/hr. The speeds of the constituents, as they modify S_2, will be the speed of S_2 plus and minus the speed of the north to south cycle. Since the maximum is only felt once per day as the earth spins, the constituent speeds are half of the sum and difference:

$$K_1 \quad (30.000 + 0.082)/2 = 15.041°/\text{hr.}$$
$$P_1 \quad (30.000 - 0.082)/2 = 14.959°/\text{hr.}$$

Thus, when K_1 and P_1 are in phase every 182.62 days (one half of the tropical year, i.e. the year with respect to the vernal equinox), they add to S_2 to simulate the maximum declination of the Sun north or south. These constituents also contribute to diurnal inequality.

The theoretical relative magnitudes of the various constituents are also of interest. It must be remembered, however, that they are computed from the tide-generating forces and are not necessarily the values in the observed tide. They are based on the value, one, for M_2, since M_2 is usually the dominant constituent. The relative magnitude values, together with the periods of the constituents (360°/speed), are:

M_2	1.00	12.42 hrs.
S_2	0.46	12.00 hrs.
O_1	0.41	25.82 hrs.
K_1	0.40	23.93 hrs.
N_2	0.20	12.66 hrs.
P_1	0.19	24.07 hrs.
L_2	0.03	12.19 hrs.

2.1.5.1 Harmonic Analysis

The mathematical process of looking at one constituent at a time from an observed time series is called harmonic analysis. By knowing the periods of the constituents, it is possible to remove them, providing there is a series which is long enough. Generally, one year is desirable but one month can provide adequate results with dominant semidiurnal tides. Standard analyses are made for 37 constituents by the U.S., although several of them may be quite small at many of the stations.

From a harmonic analysis of the observed water level series, two values are obtained for each tidal constituent. Amplitude, the vertical distance between mean tide level and the level of the crest (when plotted as a cosine curve) is one of the values. The other value is the phase lag (Epoch). The phase lag is the amount of time elapsed from the maximum astronomic event to the first maximum of its corresponding constituent tide. It is usually expressed in degrees of one complete cosine curve (360°) of that constituent. These two values are known as harmonic constants and are illustrated in Figure 5.9. It must be remembered that they are unique to the particular station location from which they were derived. Also, the harmonic constants are treated as a constant even though in the strictest sense they are not because the computed values are affected by noise in the signal, the length of the series analysed, etc. The accepted constants which are used are considered the best estimates of the actual (unknown) values. When any natural event or engineering project occurs, such as erosion, deposition, dredging or breakwater construction, which has the potential to cause major alterations in the adjacent topography new measurements and a new harmonic analysis must be undertaken.

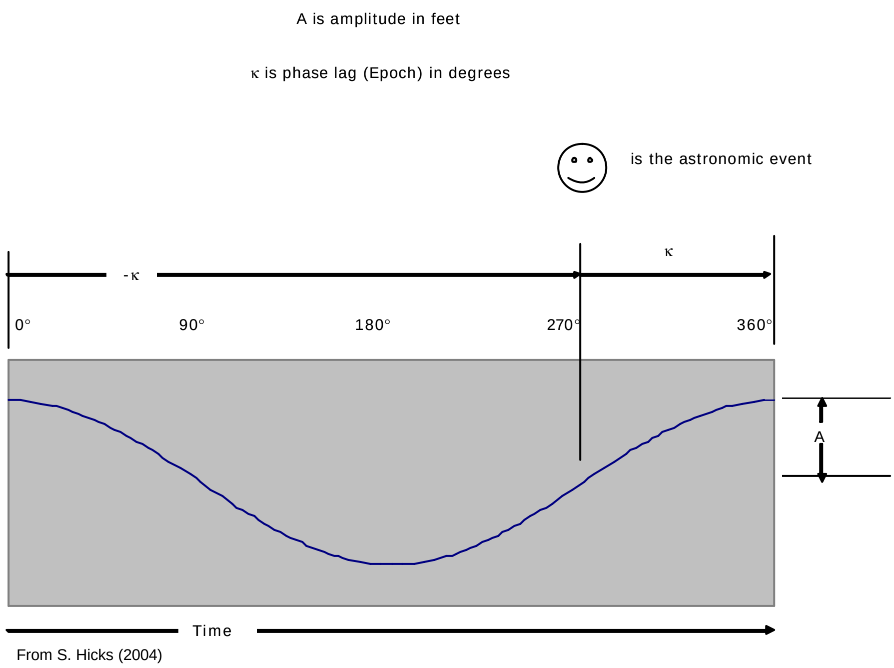


Fig. 5.9 "The amplitude and phase lag of a harmonic constituent"

2.1.5.2 Prediction of Tides

To predict the tide, say for a calendar year, it is necessary to know the harmonic constants (amplitudes and phase lags) for the constituents at each location for which predictions are desired. These are obtained from a harmonic analysis of the observed tide at each station as described above. Adjustments are made for the astronomic configurations for the beginning of the year. Knowing the phase lag of each constituent from the harmonic analysis, the first maximum of each cosine curve occurs after the event by the amount of its phase lag. The amplitude of each cosine curve is that found from the harmonic analysis.

Finally, at each hour of the year, the heights of all the cosine curves are added. When plotted, the resulting curve is normally very similar (in shape and size) to the original observed curve.

The times and heights of the high tides and low tides are placed as predictions for the forthcoming year; the vast number of predictions is possible by applying corrections to those stations for which harmonic constants have been determined - the Primary Control Stations (Reference Stations). Subordinate Stations (those without harmonic constants) are referred to their nearby Reference Stations by empirical constants. Thus, predictions are also obtained for these Subordinate Stations.

The type of tide at a given location is largely a function of the declinations of the Sun and Moon. The declinations are constantly varying such that the type of tide changes throughout the month and year at many of the locations. A more rigorous classification system is available using the amplitudes of the major constituents at each location. Quantitatively, where the ratio of (K_1 + O_1) to (M_2 + S_2) is less than 0.25, the tide is classified as semidiurnal; where the ratio is from 0.25 to 1.5, the tide is mixed mainly semidiurnal; where the ratio is from 1.6 to 3.0, the tide is mixed mainly diurnal; and where the ratio is greater than 3.0, it is diurnal.

The characteristics of diurnal inequality and its fortnightly variation can be explained by considering the combination of diurnal and semidiurnal constituents resulting from the diurnal and semidiurnal tide producing forces. As represented in Figure 5.10, where the semidiurnal constituent is represented by a dotted line and the diurnal constituent by the dashed line, the resultant tide, shown by a solid line, is clearly the sum of the two constituents.

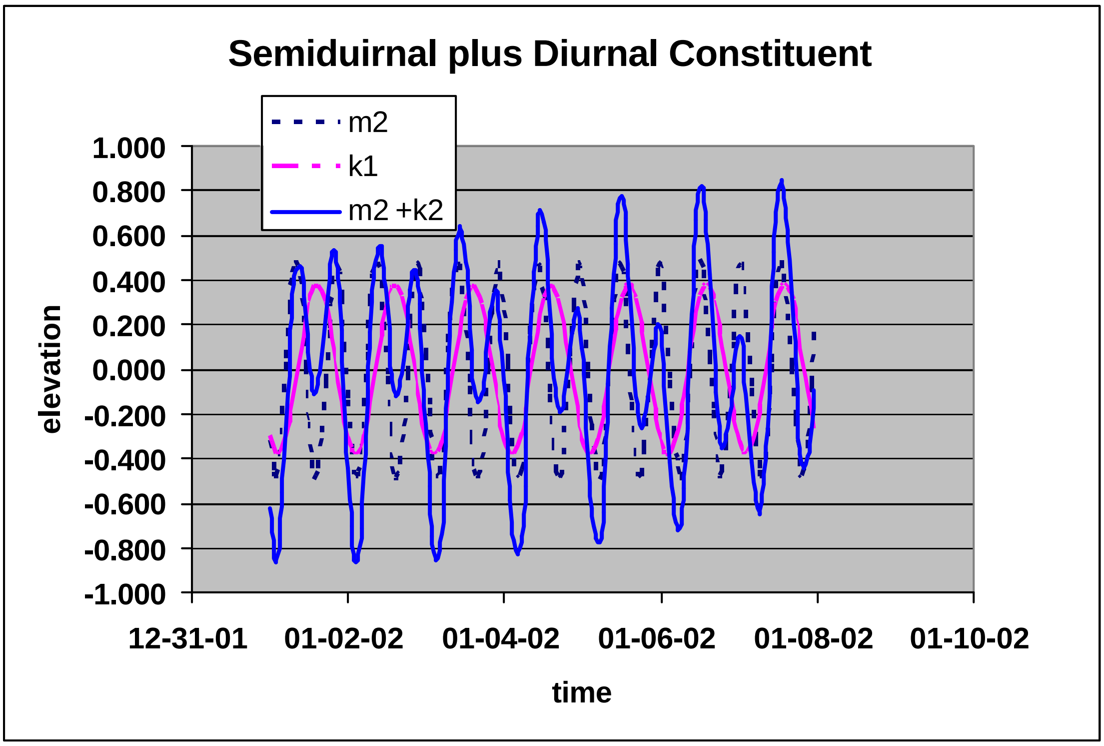


Fig. 5.10

The relative ranges of the constituents at any location, as well as the relative times of the two constituents, depend not only on the relative magnitude and phase of the tide producing forces, but also on the hydrographic characteristics of the tidal basin and the local area. For this reason, the same tide producing forces can result in different relative times and ranges of the diurnal and semidiurnal constituents at different places. Figure 5.10 portrays the simple case where the ranges of the two constituents are equal, but the relative time of the highs and lows varies. In each case, there is considerable diurnal inequality, but there are profound differences in the phase of the tide which exhibits the inequality. In the upper diagram, where the low waters occur at the same time, the diurnal inequality is exhibited in the low waters. In the middle diagram, where the high waters occur simultaneously, the inequality is exhibited in the high waters. And, in the lower diagram, where the two constituents are at mean sea level at the same time, the inequality is featured equally in the highs and the lows. These three diagrams depict the three general classes into which diurnal inequality of tidal heights are grouped.

In actually occurring tides, not only do the times of the constituents have different relations, but the ranges of the two constituents also differ. Refer to the lower diagram in Figure 5.10. If the range of the semidiurnal constituent (dotted line) remains as shown, but the range of the diurnal constituent (dashed line) becomes greater, it can be seen that the lower high water will become lower, and the higher low water will become higher. When the range of the diurnal constituent becomes twice that of the semidiurnal constituent, the lower high water and the higher low water will be equal in height, resulting in a 'vanishing tide'. As the range of the diurnal constituent increases further, there will be but one high and one low water in a day, a diurnal tide. Combining the affects of time and range, it turns out that if the range of the diurnal constituent is less than 2 times that of the semidiurnal constituent, there will be two high and two low waters daily; if the diurnal range is between 2 and 4 times the semidiurnal there may be two high and two low waters or there may be only one high and one low water in a day; and if the diurnal range exceeds 4 times the semidiurnal, only one high and one low water in the day will occur.

It should be noted that the magnitudes of both diurnal and semidiurnal forces vary during the month, the former being greatest at maximum north and south declination, the latter peaking when the Moon is over the equator. The tide at a given place, therefore, exhibits varying degrees of inequality within any two week period.

In reality, there are over 70 tidal constituents which combine to produce the resultant tide. Of these, there are four major semidiurnal constituents and three major diurnal constituents which are combined into the semidiurnal and diurnal constituents pictured in Figure 5.11.

Each constituent is based on some motion of the Earth, Moon or Sun, or combination thereof. The most important of these constituents complete their cycle within a month and all but the most insignificant complete their cycle within about 18.6 years. The 19 year period of operation required for designation as a primary tide station is based on this timetable. The whole-year period of 19 years is used rather than the 18.6 year cycle, because seasonal variations are often much greater than some of the minor astronomic constituents.

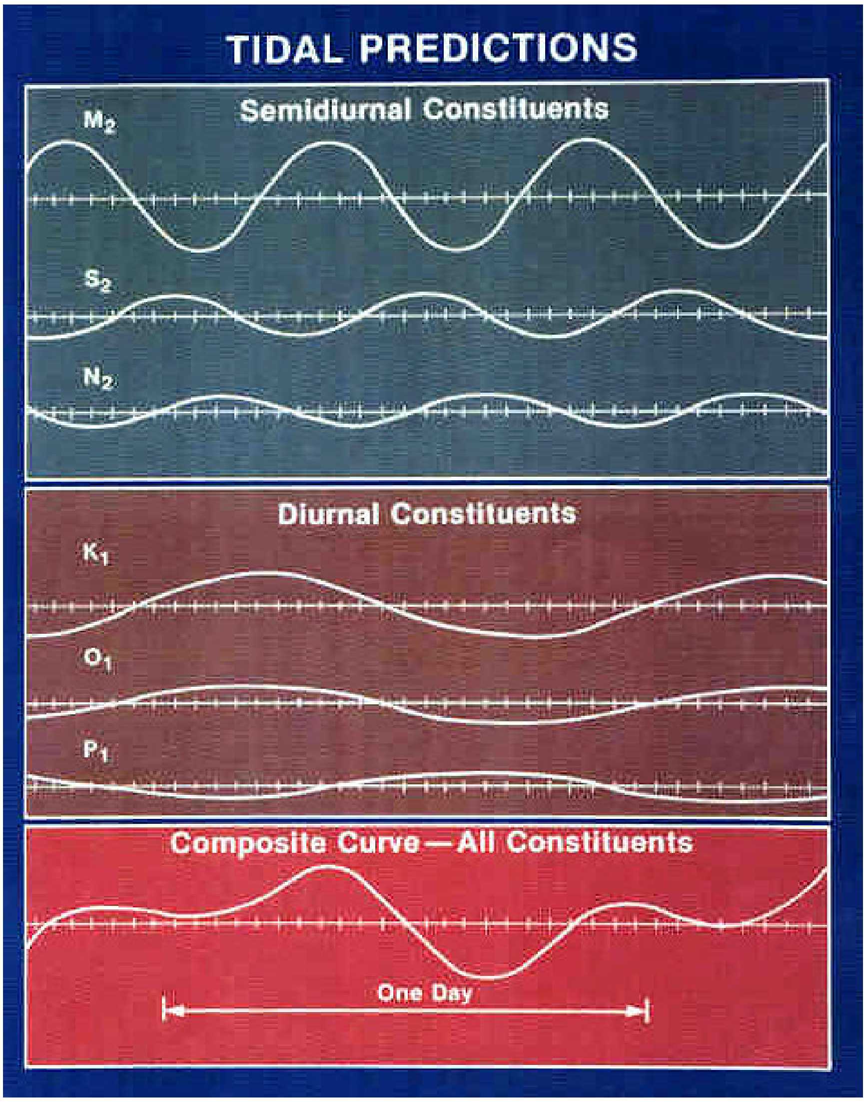


Fig. 5.11

2.2 Operational Support Functions

This section covers the water level and vertical datum requirements for operational support of hydrographic surveys. The scope of this support is comprised of the following functional areas:

a. tide and water level requirement planning;
b. preliminary tide and water level zoning development;
c. control water level station operation;
d. supplemental water level station installation, operation and removal;
e. data quality control, processing, and tabulation;
f. tide and water level datum computation and datum recovery;
g. generation of water level reducers and final tidal zoning.

2.2.1 Error Budget Considerations

The water level reducers can be a significant corrector to soundings to reduce them relative to chart datum particularly in shallow water areas with relatively high ranges of tide. The errors associated with water level reducers are generally not depth dependent, however. The portion of the error of the water level reducers must be balanced against all other sounding errors to ensure that the total sounding error budget is not exceeded. The allowable contribution of the error for tides and water levels to the total survey error budget typically falls between 0.20 m and 0.45 m depending on the complexity of the tides.

The total error of the tides and water levels can be considered to have component errors of:

a. the measurement error of the gauge/sensor and processing error to refer the measurements to station datum. The measurement error, including the dynamic effects, should not exceed 0.10 m at the 95% confidence level (see IHO Standards for Hydrographic Surveys, S-44, April 1998, section 4.2). The processing error also includes interpolation error of the water level at the exact time of the soundings. A estimate for a typical processing error is 0.10 m at the 95% confidence level.

b. the error in computation of first reduction tidal datums and for the adjustment to 19 year periods for short term stations. The shorter the time series, the less accurate the datum, i.e. bigger the error. An inappropriate control station also decreases accuracy. NOAA has determined that the estimated error of an adjusted tidal datum based on one month of data is 0.08 m for the Atlantic and Pacific coasts and 0.11 m for the coast in the Gulf of Mexico (at the 95% confidence level).

c. the error in application of tidal zoning. Tidal zoning is the extrapolation and/or interpolation of tidal characteristics from a known shore point(s) to a desired survey area using time differences and range ratios. The greater the extrapolation/interpolation, the greater the uncertainty and error. Estimates for typical errors associated with tidal zoning are 0.20 m at the 95% confidence level. However, errors for this component can easily exceed 0.20 m if tidal characteristics are very complex, or not well-defined, and if there are pronounced differential effects of meteorology on the water levels across the survey area.

2.2.2 Tide and Water Level Requirement Planning

The planning of tide and water level support for hydrographic surveys requires attention to each of the six functional areas listed above. In the context of the complete survey operation and generation of output product, the planning involves:

a. determination of overall error budget;
b. study of the tide and water level characteristics and the meteorological and oceanographic environment;
c. determination of which control stations to use and what existing vertical control is in the area, placement, logistics and time period of short-term water level stations and equipment, including GPS and geodetic datum connections;
d. construction of zoning schemes;
e. development of operational data collection, quality control and data processing and analyses functions;
f. development of final zoning and datum determination procedures, application of water level reducers to the hydrographic sheets and estimation of final error budget.

Project planning attempts to minimise and balance these potential sources of errors through the use and specification of accurate reliable water level gauges, optimisation of the mix of zoning required, the number of station locations required and the length of observations required within practical limits of the survey area and survey duration. The practical limits depend upon the tidal characteristics of the area and suitability of the coastline for the installation and operation of appropriate water level gauges.

The hydrographer should plan operations to ensure the collection of continuous and valid data series. Any break in the water level measurement series affects the accuracy of datum computations. Breaks in data also result in increased error in the tide reducers when interpolation is required to provide data at the time of soundings. At a critical measurement site where the water level measurement data cannot be transmitted or monitored during hydrographic operations, an independent backup sensor or a complete redundant water level collection system should be installed and operated during the project.

The locations of tide stations are selected to meet two sets of criteria. First, for adequate coverage, the stations must be sufficient in number and appropriately distributed to accurately portray the tidal or water level regime for the survey area. Second, the specific sites must be suitable for accurate measurement of the full range of water levels experienced.

The density and distribution of tide gauges depends on the changes in water level (usually tidal) characteristics of the survey area. The measurement of tide is generally planned to identify every 0.1 m change in range for areas with 3 m or less range, every 0.2 m change in range for areas with more than 3 m of range and to identify every 0.3 hour change in Greenwich interval.

In determining coverage requirements, tidal characteristics are first evaluated in a general geographic sense. The type of tide and changes in the type (semidiurnal, diurnal, or mixed) are analysed. The source from which the tide advances into the area is determined and the strength of the tide is evaluated relative to seasonal and localised meteorological influences. The areas of transition from tidal to non-tidal regimes are particularly important, since non-tidal areas receive different treatment for low water datum determination.

Next, the tidal characteristics are evaluated in a localised geographical context. Complex changes occur to the tide across shallow inlets, extensive marshes and narrow constrictions. Lagoons may cut off tidal flow at low water and constant river flow affects the tide at all stages. In large bays of comparatively shallow depth having a small range of tide, the wind has considerable affect on the time and height of tide. This is also true in broad stretches of rivers or along shores where the water is shoal. Man-made influences such as bulkheads, dredging, dams, levees, hydroelectric intakes and water level management practices can have significant impacts.

After this analysis, approximate station locations are identified. Stations are usually required on both sides of any significant impediments to tidal flow; at frequent intervals in very shoal areas and in the narrow upper reaches of tidal rivers; at the head of navigation or limit of survey of all rivers and streams; and on both sides of transitions from tidal to non-tidal or between diurnal, mixed, and semidiurnal tides. The survey area is usually bracketed with stations so that extrapolation of water level reducers is not required. When surveying exposed channel approaches where depths are not much greater than the draft of vessels, the water level data provided by an inshore gauge alone may not be accurate enough for the reduction of soundings. In such surveys, a temporary station on an offshore structure may be very desirable. Also, overlapping coverage is normally planned, so that at least two stations are operating for any given portion of the survey area. This overlap aids interpolation for zoning purposes and provides some backup data should one gauge malfunction.

In many cases, historical information is available to assist in planning water level coverage. Primary and secondary station information, as well as tide and water level data from previous hydrographic surveys, provide good indications as to how many and approximately where tide stations may be required for a new survey. Where historical information is not available, the planner must estimate the requirements by analysing data for nearby areas with similar physiographic characteristics. In these situations, it is prudent to err on the side of too many stations rather than being unable to provide satisfactory control for the entire survey area. Soundings acquired with insufficient tide control cannot be corrected with data from gauges installed after the survey.

2.2.3 Preliminary Tide and Water Level Zoning

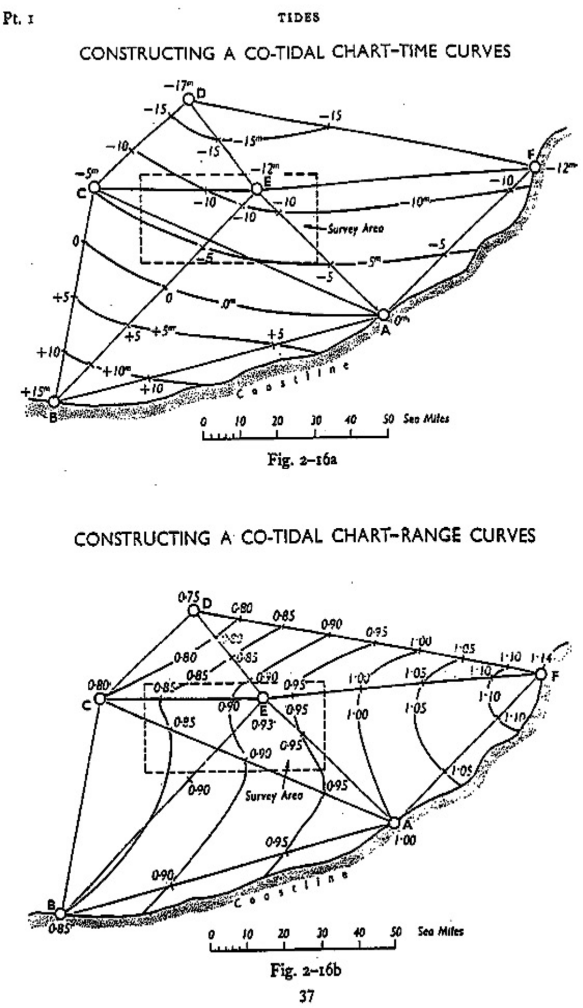
Tide and water level zoning is a tool used to extrapolate or interpolate the tide or water level variations from the closest water level station to the survey area. In many instances, interpolation or extrapolation is not necessary and water level reducers are provided directly from the water level gauge referenced to Chart Datum. In most cases, existing stations are not near the survey area or not enough water level stations can be installed in a practical sense to provide direct control everywhere. The estimated errors in the extrapolation and interpolation of water levels must be balanced with the total error budget. The more stations which can be established throughout a survey area, the less the zoning error. The more stations required, the higher the cost and logistical complexity of operations.

Any zoning scheme requires an oceanographic study of the water level variations in the survey area. For tidal areas, co-tidal maps of the time and range of tide are constructed based on historical data, hydrodynamic models and other information sources. Based on how fast the time and range of tide progress through a given survey area, the co-tidal lines are used to delineate discrete geospatial zones of equal time and range of tide. Once this is constructed, time and range correctors to appropriate operational stations or tide prediction stations can be calculated.

The techniques described above will provide correctors in the immediate vicinity of a tide station. In many instances, the survey area will fall between two or more tide stations, each of which has a different range of tide. In such situations, the correctors for the intermediate area must be interpolated into correction zones from the surrounding stations. In most cases, the zoning provided with predicted tides will be adequate for this purpose. However, should predicted zoning be unavailable or proved incorrect, the hydrographer can prepare co-tidal and co-range charts in the field from preliminary observed water levels.

A co-tidal chart portrays lines of equal Greenwich lunitidal intervals. For zoning in the field, co-tidal charts are usually drawn to show lines of equal time of high or low water before or after the relevant time at a reference tide station.

Co-range charts portray lines of equal tidal range. For field use, the lines are usually labelled with ratios relative to the reference station. These relationships to the reference gauge facilitate the preparation of reducers in the manner. Figures 5.12 and 5.13 are examples of co-tidal and co-range charts of a hypothetical bay in which a survey is conducted.



254

Fig. 5.12 and 5.13

Co-tidal chart – Co-tidal charts are generally constructed using GIS drawing tools. The following is a simple manual example to illustrate some of the fundamentals. To construct a co-tidal chart, the hydrographer should plot the reference station and all secondary, tertiary and short-term tide stations in the survey area. For best results, the survey area should be within a nearly equilateral triangle or a quadrilateral formed by tide stations. For each station, the time of arrival of high or low water before or after the time of arrival at the reference station is annotated. In some cases, the time differences are the same for high and low water.

For simplicity, such a case is depicted in Figure 5.12. In many cases, however, separate co-tidal charts for high and low waters are required. Adjacent and opposite stations are connected by straight lines. Periodic intervals along each line are then interpolated and marked. The time segments used depend upon the range of tide and the precision desired for reducers. For most areas, 10 min is a suitable interval to select. The corresponding interval marks along each line are connected by a smooth curve as shown in Figure 5.12. When two interpolated points conflict, precedence is given to the mark along the shorter line and to the marks on lines which the curves intersect closest to perpendicularity. In many instances, the survey areas are so complex that drawing of interpolation lines connecting stations is not practical and the co-tidal lines are placed by the oceanographer using GIS tools.

Co-range Charts – As shown in Figure 5.13, the chart is laid out as for the co-tidal chart. Instead of times, tidal ranges or range ratios to the reference station are annotated. Each connecting line is interpolated by increments, usually 0.1 m of range or the equivalent ratio increment. Smooth co-range lines are then drawn through the corresponding points on each line, giving precedence in the same manner as on co-tidal lines.

Zoning Charts – The zoning chart is constructed by overlaying the co-tidal chart on the co-range chart. The hydrographer can then select regions in which to apply range and time correctors to the reference station height and time. Examination of Figures 5.12 and 5.13 will reveal that the co-tidal lines and the co-range lines are not parallel. This difference in orientation is typical of most areas and often results in irregular shaped corrector zones which may not be operationally efficient. For the purposes of simplifying preliminary field correctors, however, the hydrographer can adjust the size and shape of zones to accommodate the operational situation. For example, if a sounding system of east-west lines were planned, it might be most efficient to alter the zones into east-west bands across the bay. It becomes a matter of judgment in balancing operational considerations against the need for accuracy and precision. Regardless of the zoning selected by the hydrographer in the field, however, final zoning will be based on a complete analysis of the observed water levels and will be designed for maximum accuracy.

Offshore Zoning – When it is impossible, as it is for offshore soundings, to bracket a survey area with tide stations, water-level zones must be selected on more theoretical considerations. Where the continental shelf is broad and the tidal wave approaches parallel to the shore, as it does on much of the east coast of the United States, the tide will arrive offshore earlier than inshore. On other coasts, such as the west coast of the United States, the tide wave is nearly perpendicular to shore with minimal time and range differences offshore. For offshore sounding reducers, estimates of time and range corrections to be applied to coastal tide stations may be made from existing co-tidal charts or from existing ocean tide models.

2.2.4 Control Water Level Station Operation

Control water level stations are those which already have accepted datums computed for them and which are typically in operation during the survey. They may be operated by the agency or country performing the survey or maintained by another entity. These control stations are typically used as references for tidal prediction, as direct sources of water level reducers during survey operations, as a control data source to which zoning correctors are applied and control for simultaneous comparison with short-term stations for datum recovery or datum determination. These long-term control stations are usually part of each nation's national network of tide and water level stations.

2.2.5 Supplemental Water Level Station Requirements

These stations are used to provide time series data during survey operations, tidal datum references and tidal zoning which all factor into the production of final water level reducers for specific survey areas. Station locations and requirements may be modified after station reconnaissance or as survey operations progress.

The duration of continuous data acquisition should be a 30 day minimum except for zoning gauges. Data acquisition is from at least 4 hours before the beginning of the hydrographic survey operations to 4 hours after the ending of hydrographic survey operations and/or shoreline verification in the applicable areas. Stations identified as '30 day' stations are the 'main' subordinate stations for datum establishment, providing tide reducers for a given project and for harmonic analysis from which harmonic constants for tide prediction can be derived. At these stations, data must be collected throughout the entire survey period in specified areas for which they are required and not less than 30 continuous days are required for accurate tidal datum determination. Additionally, supplemental and/or back-up gauges may also be necessary based upon the complexity of the hydrodynamics and/or the severity of environmental conditions of the project area.

A complete supplemental water level measurement station installation shall consist of the following:

a. the installation of the water level measurement system [water level sensor(s), ancillary measurements sensors (if required), a Data Collection Platform (DCP) or data logger and satellite transmitter (if installed)] and supporting structure for the DCP and sensor, and a tide staff (if required).

b. the recovery and/or installation of a minimum number of bench marks and a level connection between the bench marks and the water level sensor(s) and tide staff as appropriate on installation and removal of the gauges. Static GPS measurements should also be made to a subset of the bench marks.

2.2.5.1 Water Level Measurement Systems

2.2.5.1.1 Water Level Sensor and Data Collection Platform

Various types of water level sensors and station configurations are possible. There are several types of water level sensors which are used by various countries for support of hydrographic surveys. The U.S. uses air acoustic and pressure (vented) digital bubbler systems for control and supplemental station in tidal areas and float driven shaft angle encoders for the Great Lakes control stations, see Figure 5.14. Many other types of float driven and internal non-vented pressure systems are deployed around the globe.

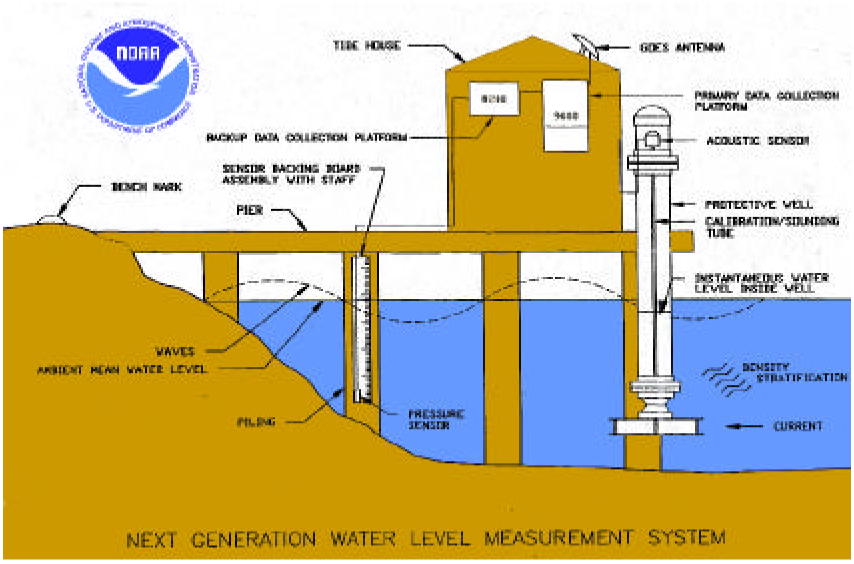


Fig. 5.14

The sensor measurement range should be greater than the expected range of water level. Gauge/sensor systems should be calibrated prior to deployment and the calibration should be checked after removal from operations. The calibration standard's accuracy must be traceable to some national or international standard. The required water level sensor resolution is a function of the tidal range of the area in which hydrographic surveys are planned. For tidal range less than or equal to 5 m, the required water level sensor resolution should be 1 mm or better; for tidal range between 5 m and 10 m, the required water level sensor resolution should be 3 mm or better; and for tidal range greater than 10 m, the required water level sensor resolution should be 5 mm or better.

The data acquisition systems should acquire and store water level measurements at time steps required for tabulation of the significant variations in the water levels. For tides, the U.S. uses a 6 minutes interval to ensure tabulation of high and low tides to the nearest tenth of an hour. Other longer sampling intervals may be appropriate for lakes and non-tidal areas, although the sampling interval should be short enough to measure any seiche action. Many sensors employ burst sampling at high rates to provide a data point at the sampling or reporting interval. NOAA systems use a 3 minute average of higher rate samples from the sensors to derive the 6 minute interval data points. Sample statistical outliers and standard deviations are then used as quality control parameters. Water level data loggers should have a clock accuracy of within one minute per month. Known error sources for each sensor shall be handled appropriately through ancillary measurements and/or correction algorithms. Examples of such errors are water density variations for pressure gauges, barometric pressure correction for non-vented systems, sound path air temperature differences for acoustic systems and high frequency wave action and high velocity currents for all sensor types.

For tidal datum applications, it is important for gauges and sensors to be carefully maintained with either frequent calibration checks or cycled swaps of calibrated sensors for long-term installations. The sensor 'zero' must be precisely related to either a tide staff and/or the bench marks through staff/gauge comparisons or direct levelling between the sensor and the bench marks. Vertical stability of the sensor 'zero', both physically and internally, must be monitored and any movement taken into account in the data reduction and datum computation.

The hydrographer should install a tide staff at a station if the reference measurement point of a sensor (zero of a gauge) cannot be directly levelled to the local bench marks, i.e. orifice is laid over sea floor in case of pressure based bubbler gauges. Even if a pressure gauge can be levelled directly, staff readings are still required for assessment of variations in gauge performance due to density variations in the water column over time. The tide staff should be mounted independent of the water level sensor so that stability of the staff or sensor is maintained. Staff should not be mounted to the same pile on which the water level sensor is located. The staff should be plumbed. When two or more staff scales are joined to form a long staff, the hydrographer should take extra care to ensure the accuracy of the staff throughout its length. The distance between staff zero and the rod stop should be measured before the staff is installed and after it is removed and the rod stop above staff zero height should be reported on the documentation forms.

2.2.5.1.2 Tide Staffs

In areas of large tidal range and long sloping beaches (i.e. Cook Inlet and the Gulf of Maine), the installation and maintenance of tide staffs can be extremely difficult and costly. In these cases, the physical installation of a tide staff(s) may be substituted by systematic levelling to the water's edge from the closest bench mark. The bench mark becomes the "staff stop" and the elevation difference to the water's edge becomes the "staff reading".

When using pressure sensors, for instance, a series of gauge/staff comparisons through a significant portion of a tidal cycle should be required at the start, at frequent intervals during deployment and at the end of a deployment. The staff to gauge observations at the start and end of deployment should be at least each three hours long and the periodic observations during the deployment should be 1 hour long.

In general, the gauge and staff should be read simultaneously and recorded once a day (minimum of three days in each seven day period) for the duration of the water level measurements. The average staff-to-gauge difference should be applied to water level measurements to relate the data to staff zero. Frequent gauge/staff comparisons (at least three times per week or minimum eight times per month) during deployment should be required to assist in assuring measurement stability and minimising processing type errors. A higher number of independent staff readings decrease the uncertainty in transferring the measurements to station datum and the bench marks. If logistically, it is not practical to have a local tide observer or for the field party to visit the station because the survey area is a long distance from the station, then whenever visits are made, a 'burst' sample of several staff readings should be made over a few hours time period instead of one discrete reading.

If the old staff is found destroyed by elements during the deployment, then a new staff should be installed for the remainder period of the deployment and a new staff to gauge constant needs to be derived by new sets of staff to gauge observations.

2.2.5.1.3 Bench Marks and Levelling

A network of bench marks is an integral part of every water level measurement station. A bench mark is a fixed physic al object or marker (monumentation) set for stability and used as a reference to the vertical and/or horizontal datums. Bench marks in the vicinity of a water level measurement station are used as the reference for the local tidal datums derived from the water level data. The relationship between the bench marks and the water level sensor or tide staff are established by differential levelling. Since gauge measurements are referenced to the bench marks, it follows that the overall quality of the datums is partly dependent on both the quality of the bench mark installation and the quality of the levelling between the bench marks and the gauge.

2.2.5.1.4 Number and Type of Bench Marks

The number and type of bench marks required depends on the duration of the water level measurements. Each station typically has one bench mark designated as the primary bench mark (PBM), which should be levelled to on every run. The PBM is typically the most stable mark in close proximity to the water level measurement station. The most desirable bench mark for GPS observations will have from 10° above the horizon 360° of horizontal clearance around the mark. If the PBM is determined to be unstable, another mark should be designated as PBM. The date of change and the elevation difference between the old and new PBM should be documented. For stations installed for longer than one-month, form 3 to 5 bench marks should be established or recovered and levelled to for each station.

2.2.5.1.5 Levelling

At least third-order levels should be run at short-term subordinate stations operated for less than one-year. Levels should be run between the water level sensor(s) or tide staff and the required number of bench marks when the water level measurement station is installed, modified (i.e. water level sensor serviced or replaced), for bracketing purposes or prior to removal. In any case, levels are required at a maximum interval of six months during the station's operation and are recommended after severe storms, hurricanes, earthquakes to document stability (see stability discussed below).

Bracketing levels to appropriate number of marks (five for 30-day minimum stations) are required if smooth tides are required 30 days or more prior to the planned removal of a applicable gauge(s) or after 6 months for stations collecting data for long term hydrographic projects.

2.2.5.1.6 Stability

If there is an unresolved movement of the water level sensor or tide staff zero relative to the PBM, from one levelling to the next, of greater than 0.010 m, the hydrographer should verify the apparent movement by re-running the levels between the sensor zero or tide staff to the PBM. This threshold of 0.010 m should not be confused with the closure tolerances used for the order and class of levelling.

2.2.5.1.7 GPS Observations at Bench Marks

Static GPS surveys should be conducted on a minimum of one bench mark, preferably two marks if time and resources permit, at each subordinate water level station installed/occupied for hydrography. Static GPS surveys should be conducted at water level stations concurrently with the occupation of NAVD 88 marks, if possible, to accomplish water level datum transfers using GPS-derived orthometric heights.

High accuracy static differential GPS surveys require a geodetic quality, dual frequency, full-wavelength GPS receiver with a minimum of 10 channels for tracking GPS satellites. A choke ring antenna is preferred, however, any geodetic quality ground plane antenna may be used. More important than antenna type, i.e. choke ring or ground plane, is that the same antennas or identical antennas should be used during the entire observing sessions. If not, a correction for the difference in antenna phase patterns (modelled phase patterns) must be applied. This is extremely critical for obtaining precise vertical results. The antenna cable length between the antenna and receiver should be kept to a minimum when possible; 10 meters is the typical antenna cable length. If a longer antenna cable is required, the cable must be fabricated from low loss coaxial cable (RG233 for up to 30 meters and RG214 over 30 meters).

The most desirable bench mark for GPS observations will have from 10° above the horizon 360° of horizontal clearance around the mark. Newly established marks should be set in locations which have the required clearances, if at all possible.

Meteorological data (air temperature, barometric pressure and relative humidity) need to be collected, if available, during the GPS observations. Meteorological data should be collected at or near the antenna phase centre. All equipment should be checked for proper calibration periodically.

2.2.5.2 Station Documentation

The documentation package:

a. installation of a station;
b. performance of bracketing levels;
c. gauge maintenance and repair;
d. removal of the station.

The station documentation generally includes, but is not limited to the following:

a. calibration test documentation from an independent source other than the manufacturer for each sensor used to collect water level or ancillary data;

b. a station report documenting the station configuration information and related metadata;

c. new or updated nautical chart section or equivalent map indicating the exact location of the station, with chart number or map name and scale shown;

d. large-scale sketch of the station site and digital GIS compatible file provided on diskette showing the relative location of the water level gauge, staff (if any), bench marks and major reference objects found in the bench mark descriptions. The sketch should include an arrow indicating north direction, a title block, latitude and longitude (derived from handheld GPS) of the gauge and all bench marks;

e. new or updated description of how to reach the station from a major geographical landmark;

f. photographs of station components and bench marks. Digital photographs are preferred. As a minimum, photographs should show a view of the water level measurement system as installed, including sensors and gauge housing; a front view of the staff (if any); multiple views of the surroundings and other views necessary to document the location; photographs of each bench mark, including a location view and a close-up showing the bench mark stamping. All photographs should be annotated and referenced with the station name, number, location and date of the photograph;

g. description/recovery notes of bench mark;

h. level records and level abstract, including level equipment information;

i. datum offset computation worksheet or Staff/Gauge difference work sheet as appropriate showing how sensor 'zero' is referenced to the bench marks.

2.2.6 Data Processing and Tabulation

2.2.6.1 Data Quality Control

The required output product used in generation of tide reducers and for tidal datum determination is a continuous time series of discrete interval water level data for the desired time period of hydrography and for a specified minimum time period from which to derive tidal datums. (Note: this discrete time interval is typically 6 to 10 minutes but for discussion purposes, 6 minutes will be used.) The 6 minute interval water level data from the water level gauges should be quality controlled for invalid and suspect data as a final review prior to product generation and application. This includes checking for data gaps, data discontinuities, datum shifts, anomalous data points, data points outside of expected tolerances such as expected maximum and minimum values and for anomalous trends in the elevations due to sensor drift or vertical movement of the tide station components and bench marks.

Quality control should include comparisons with simultaneous data from backup gauges, predicted tides or data from nearby stations, as appropriate. Data editing and gap filling should use documented mathematically sound algorithms and procedures and an audit trail should be used to track all changes and edits to observed data. All inferred data should be appropriately flagged. Water level measurements from each station should be related to a single, common datum, referred to as Station Datum. Station Datum is an arbitrary datum and should not be confused with a tidal datum such as MLLW. All discontinuities, jumps or other changes in the gauge record (refer to the specific gauge user's guide) which may be due to vertical movement of any one of the gauge, staff, or bench marks should be fully documented. To avoid confusion all data should be recorded on UTC (Co-ordinated Universal Time – also known as Greenwich

Mean Time - GMT) and the units of measurement should be properly denoted on all hard-copy output and digital files.

2.2.6.2 Data Processing and Tabulation of the Tide

The continuous 6 minute interval water level data are used to generate the standard tabulation output products. These products include the times and heights of the high and low waters, hourly heights, maximum and minimum monthly water levels and monthly mean values for the desired parameters. Examples of these tabulation products are found in Figures 5.14 and 5.15 for tide stations. The times and heights of the high and low waters should be derived from appropriate curve-fitting of the 6 minute interval data. For purposes of tabulation of the high and low tides and not non-tidal high frequency noise, successive high and low tides should not be tabulated unless they are appropriately derived. Hourly heights should be derived from every 6 minute value observed on the hour. Monthly mean sea level and monthly mean water level should be computed from the average of the hourly heights over each calendar month of data. Data should be tabulated relative to a documented consistent station datum such as tide staff zero, arbitrary station datum, MLLW, etc. over the duration of the data observations. Descriptions of general procedures used in tabulation are also found in the Tide and Current Glossary, Manual of Tide Observations and Tidal Datum Planes.

2.2.6.3 Data Editing and Gap Filling Specifications

When backup sensor data are not available, data gaps in 6 minute data should not be filled if the gaps are greater than three consecutive days in length. Data gap filling should use documented mathematically and scientifically sound algorithms and procedures and an audit trail should be used to track all gap-fills in observed data. Data gaps of less than 3 hours can be inferred using interpolation and curve-fitting techniques. Data gaps of longer than 3 hours should use external data sources such as data from a nearby station. All data derived through gap-filling procedures should be marked as inferred. Individual hourly heights, high and low waters, and daily means derived from inferred data should also be designated as inferred.

2.2.6.4 Computation of Monthly Means

When tabulation of the tides covers monthly time periods, monthly means of the various tidal parameters are computed for subsequent use in tidal datum determination and for quality control of long term data sets. Monthly mean sea level, for instance, is an important parameter for understanding long term sea level trends and seasonal variations in water levels. For purposes of monthly mean computation, monthly means should not be computed if gaps in data are greater than three consecutive days.

Fig. 5.15 "Example of a Monthly Tabulation of the Tide"

Jan 28 2003 08:24 HIGH/LOW WATER LEVEL DATA October, 2002
National Ocean Service (NOAA)
Station: 8454049 T.M.: 0 W
Name: QUONSET POINT, RI Units: Meters
Type: Mixed Datum: Station Datum
Note: > Higher-High/Lower-Low [] Inferred Tide Quality: Verified

High			Low		High			Low	
Day	Time	Height	Time	Height	Day	Time	Height	Time	Height
1	7.5	8.037	2.4	7.326	16	<9.7	[8.292]	2.6	7.394
	<20.2	8.071	<12.9	7.197		<21.3	8.782	14.6	7.563
2	8.8	8.000	2.6	7.173	17	10.6	8.345	<6.0	7.470
	<21.4	8.176	<14.3	7.066		<22.8	8.323	<15.4	7.245
3	9.5	8.233	3.2	7.157	18	10.7	8.257	4.0	7.248
	<22.3	8.314	<15.6	7.049		23.3	8.230	16.7	7.196
4	10.5	8.525	4.1	7.163	19	<11.8	8.296	<4.3	7.140
	<23.1	8.599	<16.3	7.057		<23.4	8.292	17.1	7.204
5	<11.5	8.632	4.4	7.109	20			<5.0	7.066
	23.8	8.466	<17.1	6.873		12.4	8.209	<17.5	6.994
6			<5.8	6.670	21	0.4	[8.128]	5.8	7.036
	12.2	8.477	18.2	6.832		<12.8	8.297	18.1	7.090
7	<0.5	8.582	<6.4	6.961	22	0.9	8.142	<6.5	6.999
	<13.3	8.819	19.2	6.969		<13.4	8.216	19.0	7.040
8	1.3	8.457	6.9	6.888	23	1.4	[8.075]	<6.9	7.013
	<14.0	8.644	<20.1	6.877		<13.7	[8.180]	<19.1	6.915
9	2.3	8.355	<7.9	6.852	24	2.1	7.934	7.3	6.969
	<14.9	8.631	20.9	6.986		<14.7	8.164	19.9	7.093
10	3.4	8.316	<8.2	6.969	25	2.9	[7.993]	<8.0	7.047
	<15.8	8.497	21.2	7.086		<15.4	8.156	<20.3	7.136
11	4.3	8.240	<9.4	7.129	26	3.8	[8.061]	8.3	7.204
	<16.7	8.455	22.1	7.305		<16.2	8.607	23.5	7.389
12	5.2	8.295	<10.3	7.380	27	4.6	7.974	<9.1	7.090
	<17.7	8.462				<17.1	8.216	21.9	7.348
13	5.9	8.266	0.5	7.481	28	5.4	7.860	<10.5	7.064
	<18.7	8.344	11.8	7.461		<17.9	8.008		
14	6.8	8.077	<2.2	7.401	29	6.2	7.949	1.5	7.243
	<20.1	8.161	<12.7	7.190		<18.6	8.042	<11.6	7.109
15	8.3	8.156	2.0	7.349	30	7.3	[8.052]	<1.5	7.197
	20.9	8.273	<14.1	7.344		<20.0	[8.154]	13.0	7.211
					31	8.3	8.215	2.1	7.239
						<20.7	8.290	<14.1	7.222

Highest Tide: 8.819 13.3 Hrs Oct 7 2002
Lowest Tide: 6.670 5.8 Hrs Oct 6 2002

Monthly Means:	MHHW	8.357					
	MHW	8.272	DHQ	0.085			
	MTL	7.707	GT	1.266	HWI	0.42	Hrs
	DTL	7.724	MN	1.131	LWI	6.13	Hrs
	MSL	7.668					
	MLW	7.141	DLQ	0.050			
	MLLW	7.091					

2.2.7 Computation of Tidal Datums

A vertical datum is called a tidal datum when it is defined by a certain phase of the tide. Tidal datums are local datums and should not be extended into areas which have differing hydrographic characteristics without substantiating measurements. In order that they may be recovered when needed, such datums are referenced to fixed points known as bench marks.

Basic Procedures:

a. Make Observations – Tidal datums are computed from continuous observations of the water level over specified lengths of time. Observations are made at specific locations called tide stations. Each tide station consists of a water level gauge or sensor(s), a data collection platform or data logger and data transmission system and a set of tidal bench marks established in the vicinity of the tide station. NOS collects water level data at 6 minute intervals.

b. Tabulate the Tide – Once water level observations are quality controlled and any small gaps filled, the data are processed by tabulating the high and low tides and hourly heights for each day. Tidal parameters from these daily tabulations of the tide are then reduced to mean values, typically on a calendar month basis for longer period records or over a few days or weeks for shorter-term records.

c. Compute Tidal Datums – First reduction tidal datums are determined directly by meaning values of the tidal parameters over a 19 year NTDE (National Tidal Datum Epoch). Equivalent NTDE tidal datums are computed from tide stations operating for shorter time periods through comparison of simultaneous data between the short term station and a long term station.

d. Compute Bench Mark Elevations – Once the tidal datums are computed from the tabulations, the elevations are transferred to the bench marks established on the land through the elevation differences established by differential levelling between the tide gauge sensor ‘zero’ and the bench marks during the station operation. The bench mark elevations and descriptions are disseminated by NOS through a published bench mark sheet for each station. Connections between tidal datum elevations and geodetic elevations are obtained after levelling between tidal bench marks and geodetic network benchmarks. Traditionally, this has been accomplished using differential levelling, however GPS surveying techniques can also be used (NGS, 1997).

The locations of tide stations are organized into a hierarchy:

a. <u>Control tide stations</u> are generally those which have been operated for 19 or more years, are expected to continuously operate in the future and are used to obtain a continuous record of the water levels in a locality. Control tide stations are sited to provide datum control for national applications and located in as many places as needed for datum control.

b. <u>Secondary water level stations</u> are those which are operated for less than 19 years but more than 1 year and have a planned finite lifetime. Secondary stations provide control in bays and estuaries where localized tidal effects are not realized at the nearest control station. Observations at a secondary station are not usually sufficient for a precise independent determination of tidal datums, but when reduced by comparison with simultaneous observations at a suitable control tide station very satisfactory results may be obtained.

c. Tertiary water level stations are those which are operated for more than a month but less than 1 year. Short-term water level measurement stations (secondary and tertiary) may have their data reduced to equivalent 19 year tidal datums through mathematical simultaneous comparison with a nearby control station.

Control (or primary) tide stations, secondary stations and tertiary stations are located at strategic locations for network coverage. The site selection criteria include spatial coverage of significant changes in tidal characteristics such as: changes in tide type, changes in range of tide, changes in time of tide, changes in daily mean sea level and changes in long term mean sea level trends. Other criteria include coverage of critical navigation areas and transitional zones, historical sites, proximity to the geodetic network and the availability of existing structures, such as piers suitable for the location of the scientific equipment.

Procedure for Simultaneous Comparison:

Conceptually, the following steps need to be completed in order to compute equivalent NTDE tidal datums at short term stations using the method of comparison of simultaneous observations:

a. select the time period over which the simultaneous comparison will be made;

b. select the appropriate control tide station for the subordinate station of interest;

c. obtain the simultaneous data from subordinate and control stations and obtain or tabulate the tides and compute monthly means, as appropriate;

d. obtain the accepted (relative to the NTDE in the U.S. for example) values of the tidal datums at the control station;

e. compute the mean differences and/or ratios (as appropriate) in the tidal parameters between the subordinate and control station over the period of simultaneous comparison;

f. apply the mean differences and ratios computed in step e, above, to the accepted values at the control station to obtain equivalent or corrected NTDE values for the subordinate station.

Compute Bench Mark Elevations.

Once the tidal datums are computed from the tabulations, the elevations are transferred to the bench marks established on the land through the elevation differences established by differential levelling between the tide gauge sensor ‘zero’ and the bench marks during the station operation (NOS Specifications and Deliverables, 2000). Connections between tidal datum elevations and geodetic elevations are obtained after levelling between tidal bench marks and geodetic network benchmarks. Traditionally, this has been accomplished using differential levelling, however GPS surveying techniques can also be used (NGS, 1997).

2.2.7.1 Tidal Datum Recovery

Whenever tide stations are installed at historical sites, measures should be taken to ‘recover’ the established tidal datums through levelling which should be accomplished by referencing the gauge or tide staff ‘zero’ to more than one existing bench mark with a published tidal elevation. Through this process, the published MLLW elevation is transferred by level differences to the ‘new’ gauge or tide staff and compared to the MLLW elevation computed from the new data on the same ‘zero’. Factors affecting the datum recovery (i.e. differences between old and newly computed datums) include the length of each data series used to compute the datums, the geographical location, the tidal characteristics in the region, the length of time between re-occupations, the sea level trends in the region and the control station used. Based on all of these factors, the datum recovery can be expected to vary from +/- 0.03 m to +/- 0.08 m. Hence, this process also serves as a very useful quality control procedure. After a successful datum recovery is performed and benchmark stability is established, the historical value of Mean Lower Low Water (MLLW) should be used as the operational datum reference for data from the gauge during hydrographic survey operations.

2.2.7.2 Quality Control of Datums

It is essential for tidal datum quality control to have data processing and levelling procedures carried out to the fullest extent. Caution must also be used in computing tidal datums in riverine systems or in regions of unknown tidal regimes. Tide-by-tide comparisons between subordinate and control station data will often detect anomalous differences which should be investigated for possible gauge malfunction or sensor movement. Datums should be established from more than one bench mark. Differences in elevations between bench marks based on new levelling must agree with previously established differences from the published bench mark sheets. Any changes in the elevation differences must be reconciled before using in any datum recovery procedure. Datum accuracy at a subordinate station depends on various factors, but availability and choice of an adequate control station of similar tidal characteristics, similar daily mean sea level and seasonal mean sea level variations, and similar sea level trends are the most important. The length of series will also determine accuracy. The longer the series, the more accurate the datum and the greater quality control and confidence gained from analyzing numerous monthly mean differences between the subordinate and control station. At re-occupied historical stations for which datum recoveries are made, updated datums should be computed from the new time series and compared with the historical datums as the survey progresses.

2.2.7.3 Geodetic Datum Relationships

Tidal datums are local vertical datums which may change considerably within a geographical area. A geodetic datum is a fixed plane of reference for vertical control of land elevations. The North American Vertical Datum of 1988 (NAVD 88) is the accepted geodetic reference datum of the U.S. National Geodetic Spatial Reference System and is officially supported by the National Geodetic Survey (NGS) through a network of GPS continuously operating reference stations. The relationship of tidal datums to NAVD has many hydrographic, coastal mapping and engineering applications including monitoring sea level change and the deployment of GPS electronic chart display and information systems, etc. In some countries, the local datum of Mean Sea Level (MSL) has been confused over time with the national geodetic reference datum because the geodetic datums were originally derived from MSL measurements at tide gauges. However as relative sea level has changed with vertical land movement and global sea level rise, the geodetic datums became de-coupled from local oceanographic MSL. NAVD88, for instance, used only one tide station as a starting reference and is not considered a MSL correlated datum.

Existing geodetic marks in the vicinity of a subordinate tidal station should be searched for and recovered. A search routine is available at **http://www.ngs.noaa.gov**. An orthometric level connection and ellipsoidal GPS tie is required at a subordinate tide station which has geodetic bench marks located nearby. NAVD 88 height elevations for published bench marks are given in Helmert orthometric height units by NGS. The GPS ellipsoid network height accuracies are classified as conforming to 2 cm or 5 cm standards accuracies (Refer to NOAA Technical Memorandum NOS NGS-58). At the present time, GPS ellipsoid heights conforming to the 2 cm accuracy standards are required for contract hydrographic surveying projects. Refer to Section 4.2.8 GPS Observations and User's Guide for GPS Observations, NOAA/NOS, updated January 2003.

An orthometric level connection is preferred over ellipsoidal GPS tie, where applicable, for deriving NAVD 88 heights. An orthometric level connection is required if any geodetic marks (up to five marks) are located within a radius of 0.8 km from the subordinate tide station location. If suitable marks are found in the NGS database, and are farther than 0.8 km but less than 10 km from a subordinate tide station, then a GPS tie is required to derive the ellipsoid heights. If a minimum of five existing tidal bench marks within 1 km of a subordinate tide station location are not found, or suitable geodetic marks are not found in the NGS database within 10 km of a subordinate tide station, then five new bench marks should be installed, described, connected by levels, and GPS observations should be undertaken on at least one of the five marks. (Refer to User's Guide for Writing Bench Mark Descriptions, NOAA/NOS, Updated January 2002, User's Guide for GPS Observations, NOAA/NOS, Updated January 2003, and Section 4.2.8 GPS Observations.) At least two geodetic bench marks should be used to validate the levelling or GPS ellipsoid height connection for quality control purposes.

2.2.8 Final Zoning and Tide Reducers

Data relative to MLLW from subordinate stations installed specifically for the survey, or from existing primary control stations, as appropriate, should be applied to reduce sounding data to chart datum, either directly or indirectly through a correction technique referred to as tidal zoning. Whether corrected or direct, time series data relative to MLLW or other applicable LWD applied to reference hydrographic soundings to chart datum are referred to as 'tide reducers' or 'water level reducers'.

2.2.8.1 Construction of Final Tidal Zoning Schemes

As tidal characteristics vary spatially, data from deployed water level gauges may not be representative of water levels across a survey area. Tidal zoning should be implemented to facilitate the provision of time series water level data relative to chart datum for any point within the survey area such that prescribed accuracy requirements are maintained for the water level measurement component of the hydrographic survey. NOS currently utilises the 'discrete tidal zoning' method for operations, where survey areas are broken up into a scheme of cells bounding areas of common tidal characteristics. The minimum requirement is for a new cell for every 0.06 m change in mean range of tide and every 0.3 hour progression in time of tide (Greenwich high and low water intervals). Phase and amplitude corrections for appropriate tide station data should be assigned to each cell.

Preliminary zoning, which is based on available historical tide station data and estuarine and global tide models, is referenced to an applicable predictions reference station for utilisation during field work. For final processing, preliminary zoning should be superseded by 'final zoning' which is a refinement based on new data collected at subordinate stations during the survey. With the final zoning scheme, correctors for each zone should be derived from a subordinate station specifically installed for the survey rather than the reference station used with preliminary zoning. Zoning errors should be minimized such that when combined with errors from actual water level measurement at the gauge and errors in reduction to chart datum, the total error of the tide reducers is within specified tolerances. The final zoning scheme and all data utilized in its development should be documented and submitted.

2.2.8.2 Tide Reducer Files and Final Tide Note

Verified time series data collected at appropriate subordinate stations are referenced to the NTDE Mean Lower Low Water (Chart Datum) through datum computation procedures. Time series data collected in six minute intervals and reduced to chart datum as specified, both from subordinate gauges operated during the survey should be used either directly or corrected through use of a zoning scheme such that tide reducers are within specified tolerances. A Final Tide Note should be submitted for each hydrographic sheet with information as to what final tidal zoning should be applied to which stations to obtain the final tide reducers. An example Final tide Note and final tidal zoning graphic is found in Figures 5.16 and 5.17.

Fig. 5.16 "FINAL TIDE NOTE and FINAL TIDAL ZONING CHART"

DATE: December 22, 1999

HYDROGRAPHIC BRANCH: Pacific
HYDROGRAPHIC PROJECT: OPR-P342-RA-99
HYDROGRAPHIC SHEET: H-10910

LOCALITY: 6 NM Northwest of Cape Kasilof, AK

TIME PERIOD: July 22 - August 20, 1999

TIDE STATION USED: 945-5711 Cape Kasilof, AK
Lat. 60° 20.2'N Lon. 151° 22.8'W
PLANE OF REFERENCE (MEAN LOWER LOW WATER): 0.000 meters
HEIGHT OF HIGH WATER ABOVE PLANE OF REFERENCE: 5.850 meters

REMARKS: RECOMMENDED ZONING
Use zone(s) identified as: CK394, CK395, CK399, CK400, CK401, CK407, CK408, CK409, CK434, CK435, CK441, CK442, CK443, CK467, CK468, CK469, CK470, CK477, CK480, CK481, CK482, CK483, CK493 & CK494.

Refer to attachments for zoning information.

Note 1: Provided time series data are tabulated in metric units (Meters), relative to MLLW and on Greenwich Mean Time.

Note 2: Nikiski, AK served as datum control for subordinate tide stations and for tidal zoning in this hydrographic survey. Accepted datums for this station have been updated recently and have changed significantly from previous values.

The current National Tidal Datum Epoch (NTDE) used to compute tidal datums at tide stations is the 1960-78 NTDE. Traditionally, NTDEs have been adjusted when significant changes in mean sea level (MSL) trends were found through analyses amongst the National Water Level Observation Network (NWLON) stations. Epochs are updated to ensure that tidal datums are the most accurate and practical for navigation, surveying and engineering applications and reflect the existing local sea level conditions. For instance, analyses of sea level trends show that a new NTDE is necessary and efforts are underway to update the 1960-1978 NTDE to a more recent 19 year time period.

Note: This example of Field Tide Note and Final Tidal Zoning Chart was written in December 1999, at that time NTDE was 1960-1978, now the new NTDE is 1983-2001.

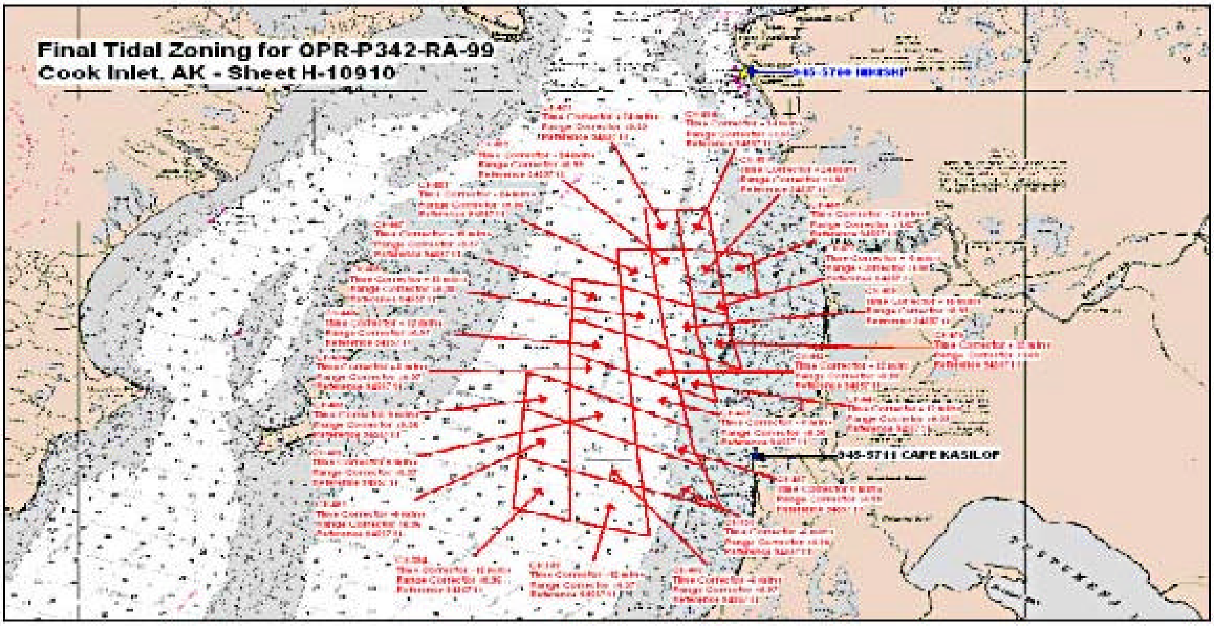


Fig. 5.17

The final observed water level measurements should be reported as heights in meters to three decimal places (i.e. 0.001 m). All heights should be referenced to station datum and should be referenced to UTC. The final tide reducer time series data should be referenced to MLLW and should be referenced to UTC.

The original raw water level data and also the correctors used to convert the data to chart datum should be retained until notified in writing or at least two years after the survey is completed. All algorithms and conversions used to provide correctors should be fully supported by the calibrations, maintenance documentation, levelling records and sound engineering/oceanographic practices. Sensors for measurements used to convert data (i.e. pressure to heights) should be calibrated and maintained for the entire water level collection period.

2.2.9 Using Kinematic GPS for Vertical Control

The technology of using Kinematic GPS for vertical control in hydrographic surveying is becoming much more commonplace after being in a research-to-operations mode for several years. Kinematic GPS is a form of centimetre level-differential positioning which uses primarily carrier phase observables in which the differential corrections are formulated in conjunction with a mobile GPS receiver (i.e. a ship or launch) and at least one static base station.

Kinematic GPS requires an accurate horizontal and vertical reference frame in order to determine an accurate position on each sounding, relative to NAD83 (for instance), and to determine an accurate depth of each sounding such as MLLW, LAT or other appropriate local chart datum. The issue of determining the separation between kinematic differential GPS vertical datum and the local chart datum is important to resolve for each survey area. This separation is not constant and may be quite complex. They are usually not well known and may require additional measurements to understand the complexity of the geodesy in the area as well as the tidal characteristics. Constant relationships may be adequate in small survey areas, simple interpolation in others; or complex interpolation models and continuous zoning schemes may be required. Tidal datums, bathymetry and geodesy must be put into the same vertical reference frame prior to the survey operations.

The availability of control for use of kinematic GPS for surveying must be evaluated during planning and if required, geodetic and tidal datum control must be established before operational collection of soundings to establish the relationship of the tidal datum and GPS reference surfaces throughout the survey area. The amount of field work required is dependent on the adequacy of existing tide and geodetic control (NOS, 2000).

3. WATER LEVEL FLOW AND TIDAL CURRENTS

3.1 Introduction

The hydrographer is required to have working knowledge of predicted and observed oceanographic and meteorological conditions to be able to conduct successful field data collection surveys and to conduct safe and efficient navigation necessary for the surveys. Besides the rise and fall of the tides, the tidal currents are often a predominant variable that affects field operations. Often the hydrographer is not only required to take soundings for nautical chart production, but must assess the tidal characteristics and the movements of the tidal currents and be able to describe them for such applications as the coast pilot volumes and tidal prediction products. In addition, the hydrographer is often required to deploy and retrieve current meter instrumentation and moorings.

3.2 Principles of Tidal Currents

Water current is horizontal movement of water. Currents may be classified as tidal and non-tidal. Tidal currents are caused by gravitational interactions between the Sun, Moon, and Earth and are part of the same general movement of the oceans which results in the vertical rise and fall of the tide. Non-tidal currents include the permanent currents in the general circulatory systems of the sea as well as temporary currents arising from more pronounced meteorological variability.

Much like tidal datums and heights, various countries use various terminology to describe the same phenomena. The United Kingdom uses the term tidal streams instead of tidal current and the term tidal flow to describe the actual flow or total current flow, which is a combination of tidal and non-tidal components.

Residuals are sometimes referred to as the difference between observed total and predicted tidal currents or the difference between tidal streams and tidal flow. Although tidal currents are derived from the same tide producing forces as the tide, tidal currents are much more variable and complex to predict than the tidal heights. The rise and fall of the tide is termed a scalar (varying heights only) quantity while tidal currents are vector (both speed and associated direction vary) quantities. Speed and direction of currents at a given location not only vary over time, but with depth as well. And the characteristics of the current at any given location cannot be extended very far, especially in areas of complex shallow water bathymetry and complex topography (shoreline configurations). Current patterns in complex areas also may exhibit eddies and gyres of various sizes set up by bathymetry and channel configuration on shallow waters. It is not uncommon to find current shear patterns in which there are significant changes in direction and amplitude. Because of this spatial variability, tidal predictions derived from a current meter measurement are typically only valid for a small area at a given depth, and are not necessarily transferable within a region or throughout a water column.

Types of non-tidal currents include:

- Oceanic circulation currents;
- Gyres, Western and Eastern Boundary currents, Equatorial counter-current, etc.;
- Thermohaline circulation;

- Wind driven currents (Ekman to about 100M);
- Seiches;
- River flow and hydraulic currents.

In the open ocean, tidal currents or streams tend to be rotary in nature (Figure 5.18). In theory if the earth were covered completely by water, at the time when the earth or sun is aligned with the Equator, the tidal currents at the equator would move back and forth (east and west) in a reversing manner in response to the tides. With latitude, the currents would show elliptical patterns increasing with latitude to a circular pattern at the poles. The pattern at any given latitude would vary depending upon declination of the Moon or Sun. The effects of coriolis force also reinforce the rotary nature of the currents on the ocean such that they rotate clockwise in the northern hemisphere and counter(anti)-clockwise in the southern hemisphere.

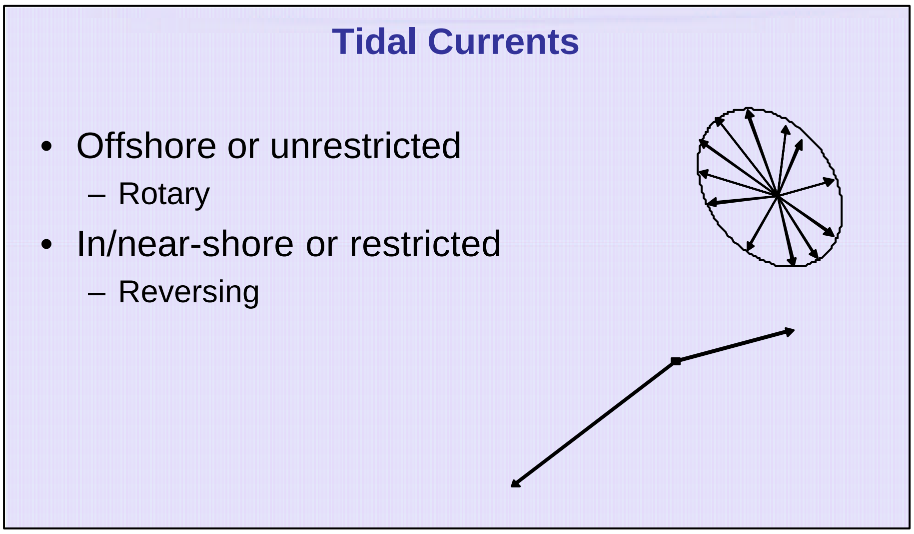


Fig. 5.18

In the near shore areas, tidal currents and streams tend to be more reversing in nature responding to the bathymetry and topography of the estuaries and bays (Figure 5.19). The phases of reversing currents are described as having slack periods, maximum floods and maximum ebbs. Slack water is the short time period between the reversal from flood to ebb. Typically, flood currents are those which are incoming or towards the shore or upstream. Ebb currents are those which are outgoing, offshore or downstream. These tides will show many of the characteristics of the tide described for tidal heights. Floods and ebbs will display semi-diurnal, mixed and diurnal tide characteristics very similar to the corresponding tidal height characteristics in a given area. Their strengths and velocities will exhibit variations in response to changing declinations of the Moon and Sun and to the perigee/apogee and perihelion/aphelion cycles (Figure 5.4). Tidal currents in mixed tide regimes display inequalities in the floods and ebbs each day, just like the tidal heights.

The current direction is sometimes referred to as the set and the speed is sometimes referred to as the drift. The current direction, by convention, is the compass direction toward which current flows (the opposite of the convention for winds). Speeds are defined in terms of knots (navigation) or meters/sec (scientific) (1 knot = 0.51444 m/s).

Hydraulic currents are currents due to the height difference in water level of two interconnected basins (Hell's Gate NY, Cape Cod Canal & the Chesapeake and Delaware Canal). The height differences in tidal waters are caused by the phase difference in the tide at each end of a strait or a canal. Non-tidal hydraulic currents occur in the interconnecting water ways of the Great Lakes for instance and typically are in a downstream direction only.

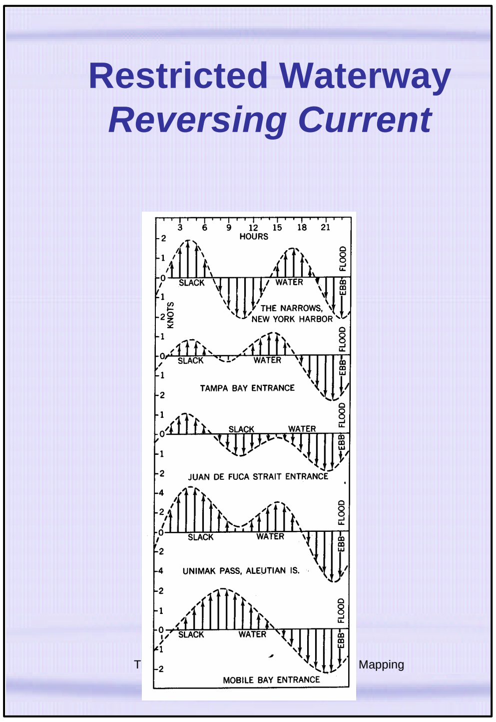


Fig. 5.19

In theory, tidal currents should have consistent relationships between the times and strengths of the flood and ebbs and the times and heights of the high and low tides because they are related and forced by the same tide producing forces. However the response of basins and estuaries to the tide producing forces and the resulting relationships of tidal current to tidal heights is complex and varies by location. In some locations, maximum currents occur at mid-tide and at others, maximum currents occur nearer high and low tides.

3.3 Measurement of Currents

There are two distinct methods for measuring currents: Lagrangian use of floats, dyes, drift cards, drifters, or current drogues and Eulerian use of a current meter at a single location(s). Both types have their advantages and disadvantages depending upon the application. Lagrangian devices require tracking of concentrations or changes in position of drifters over time, they are useful for trajectory modelling and forecasting for application to HAZMAT spills and oil spills or for studies of estuarine circulation patterns. Sub-surface drifters can also be deployed to track bottom currents. Eulerian devices provide good time series information of currents at specific locations and depths used in traditional tidal current prediction applications for recreational and commercial navigation and fisheries vessel operations. Both types of measurements are useful for complete understanding of current regimes and for the development and calibration of hydrodynamic circulation models. Hydrographic survey vessels may be required to deploy a variety of current measurement devices depending upon the survey area and the information required.

The earliest current measurement systems were Lagrangian in nature, using drifting ship tracks or using drift poles deployed from ships. For near-shore work, these were replaced by moored current meter systems of various mechanical and electro-mechanical designs. These systems are installed in sub-surface mooring designs with several meters deployed along the vertical mooring line depending upon depth with the top current meter deployed as close to the surface as possible. Mechanical current meters use combinations of vanes, rotors and propellers to measure speed and direction. The meters are generally internal recording with the data collected upon retrieval. Deployment periods are generally short-term (a few months maximum). Modern current meter systems use acoustic doppler current profiler (ADCP) technology to measure current profiles in the water column over time from a bottom mounted current meter. ADCPs can also be deployed horizontally to measure across channel currents over time at fixed depths and can be towed to measure currents with depth over cross channel transects. These current meters can also be deployed from surface buoys in a downward looking direction and can be configured to provide data in real-time using acoustic modem technology or direct cables depending upon the deployment. The ADCPs provide profiles of current speed and direction by providing information for fixed vertical bins in the water column. Figure 5.20 shows some typical deployment configurations for current meters.

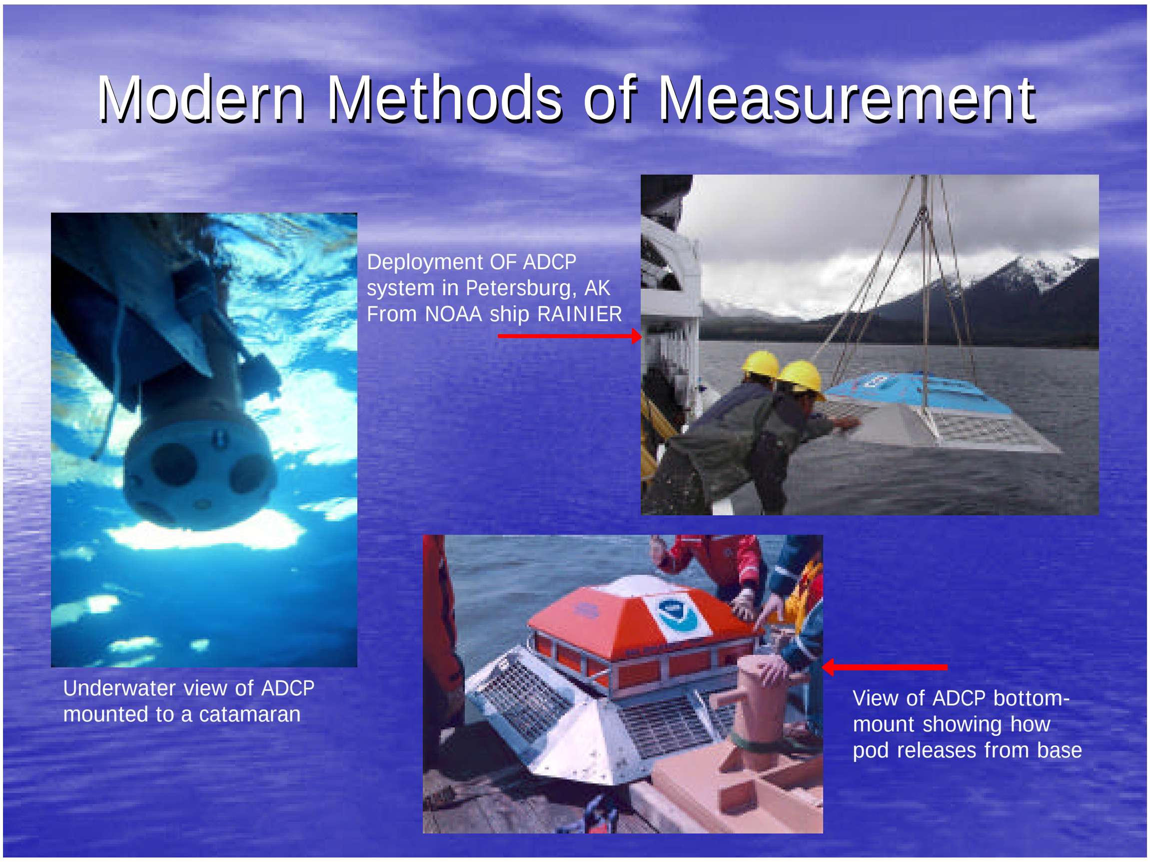


Fig. 5.20

New high frequency radar systems are being developed that provide surface current maps over wide-areas which should also be beneficial to conducting hydrographic survey operations. These shore-based systems use a transponder and receiver antennae to provide current vectors for fixed surface area bins in near-real-time (see Figure 5.21).

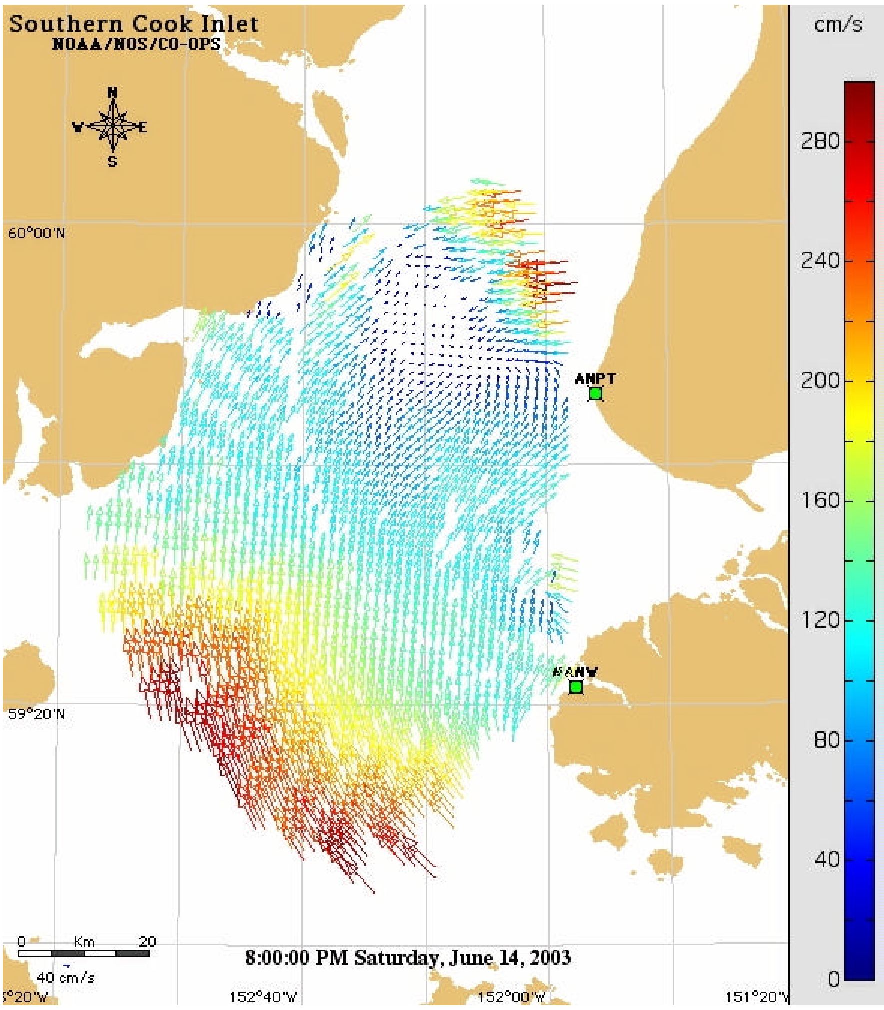


Fig. 5.21

3.4 Tidal Current Prediction

Tidal currents, like tidal heights, can be predicted because they are caused by the interaction of the well-known Earth-Moon-Sun system. Also, like tidal height predictions, tidal currents are predicted by performing harmonic analysis of measurements obtained from preferably 29 days of data to span an entire lunar month. A minimum of 15 days of data can be used for harmonic analysis currents, simply because it is historically and logistically hard to routinely obtain more than that in a typical deployment. Although the approach and theory are the same for harmonic analysis of tides and tidal currents, analyses of tidal currents are more complex. For example, for reversing currents, two sets of constituents are obtained for the major and minimum axes with the major axis being the principal current direction. In addition, the analysis must try to handle the presence of any non-tidal permanent current found in the analysis of the observations.

Mariners are generally interested in the timing and strength of four phases of the tidal current cycle. NOAA (U.S.) tide prediction tables include predictions for slack before flood (SBF), maximum flood current (MFC), slack before ebb (SBE), and maximum ebb current (MEC). In areas where the currents are never at true slack (zero speed), slack flood current (SFC) and slack ebb current (SEC) values are predicted as well. Tidal current prediction stations also utilise the same concept of reference stations and secondary stations in the tide table products such that speed ratios are used to correct daily reference station predictions to the desired location(s).

Predictions of tidal currents have similar limitations to those for tidal height predictions. Extreme care must be taken to extrapolate a tide or tidal current prediction beyond the location of the measurement. This is especially the case for tidal currents due to the nature of the spatial variation in speed and direction in shallow water estuaries and rivers, the significant affects of non-tidal forcing due to river flow, wind speed and direction and natural non-tidal circulation patterns. Just as for tidal heights, tidal current predictions are much less useful in areas with low signal-to-noise ratio (low tidal forcing relative to non-tidal forcing).

REFERENCES

June 2000.	*"Tidal Datums and Their Applications"*	Special Publication No. CO-OPS 1, NOAA/NOS.
Reprinted 1965.	*"Manual of Tide Observations"*	U.S. Department of Commerce, Publication 30-1.
1951.	*"Tidal Datum Planes"*	U.S. Department of Commerce, Special Publication No.135, Marmer.
October 1989.	*"Tide and Current Glossary"*	U.S. Department of Commerce, NOAA, NOS.
1974.	*"Variability of Tidal Datums and Accuracy in Determining Datums from Short Series of Observations"*	NOAA Technical Report NOS 64, Swanson.
December 2003.	*"Computational Techniques for Tidal Datums"*	NOAA Technical Report NOS CO-OPS 2, U.S. Department of Commerce, NOAA, NOS.
September 1984.	*"Standards and Specifications for Geodetic Control Networks"*	Federal Geodetic Control Committee.
November 1997.	*"Guidelines for Establishing GPS-Derived Ellipsoid Heights (Standards: 2CM and 5CM)"*	Version 4.3, NOAA Technical Memorandum NOS NGS-58.
August 2000.	*"NOS RTK Team Final Report"*	NOAA/NOS Team Final Report.
December 2004.	*"Understanding Tides"*	Steacy Dopp Hicks.
1983.	*"Canadian Tidal Manua "*	Warren D. Forrester, Ph. D. under contract to Department of Fisheries and Oceans, Ottawa.
1969.	*"Chapter 2, Tides and Tidal Streams"*	Admiralty Manual of Hydrographic Surveying, Vol 2.
March 2003.	*"NOS Hydrographic Survey Specifications and Deliverables"*	NOAA, National Ocean Service, Office of Coast Survey, Silver Spring, MD, USA.

CHAPTER 6
TOPOGRAPHIC SURVEYING

by Mr. Federico MAYER and Mr. Hector SALGADO (Argentina)

1. INTRODUCTION

From the hydrographic point of view a Topographic Survey consists of a series of tasks carried out with the aim of determining the composition of those parts of the earth's surface which emerge from the water. It includes the coastal relief and the location of permanent natural or artificial objects and features.

Such information is partly obtained by determining the position of points on the ground, which allows their shape as well as details of the features to be depicted, enabling their location and description to be charted. Other sources of data include remote sensing processes from aerial photogrammetric information, other airborne sensors or satellite imagery products. In these cases it is necessary to create ground control points in order to adjust the information to the reference frame in use.

The term topography often has other applications, for example in oceanography it is used to depict seafloor surfaces or the boundaries of certain water mass characteristics. All these meanings share a common external description of surfaces covering a physical body.

This chapter deals with the methods applicable to the description of coastal features as part of hydrographic surveys, particularly with regards to the appearance of the ground and the location of detail. It includes coastlining and location fixes, generally related to the high water line for marine surveys, the information on these areas ranges from this line to the low water line, as well as conspicuous coastal features which allow the mariner to position himself relative to near shore dangers.

Except in harbours or coastal areas, where operations or projects are planned or expected to be undertaken, it is necessary to make detailed observations of coastal formations by topographic survey methods.

In some cases, much of the topographic surveying may be undertaken via photogrammetric processes. In these surveys, control is achieved by positioning details on the ground which may be identified in images. Additionally it is necessary to add information which may provide a proper interpretation of the structure of coastal features.

In coastal topographic surveys it is also essential to locate all aids to navigation within the surveying area; if required, the horizontal and vertical geodetic control network should be made denser. In all these cases, it is essential that the reference system for the topographic survey co-ordinates, the geodetic control and aids to navigation (reference stations, lights, beacons, etc.) is consistent with the reference system used for the rest of the hydrographic survey. This precaution is fundamental for the mariner, who positions himself with the use of the aids to navigation and other coastal details, to be able to rely on the charted depths at every fix.

This chapter will deal first with the methods applied to land surveying, then it will deal with remote sensing ranging from photogrammetric processes to satellite imagery.

Except for the restatement of some basic principles, which are deemed essential, it is assumed the reader has previously examined Chapter 2 (Positioning) where subjects related to co-ordinates on the spheroid and the plane, horizontal/vertical control methods and positioning equipment and methods are covered in more depth.

2. TOPOGRAPHY, COASTLINE DELINEATION AND AIDS TO NAVIGATION POSITIONING

2.1 Specifications

2.1.1 All tasks shall assume, as a minimum, the specifications stated in publication S-44 (IHO Standards for Hydrographic Surveying), particularly those relating to Chapter 2.

2.1.2 In S-44 2.2, a relative accuracy of 1:100000 (10 ppm) for Primary Horizontal Control, or an error below 10 cm for a 95% confidence level, is expected. For secondary stations (primary control densification), the relative error limit expected for such standard is 1:10000 (100 ppm), or an error below 0.50m.

2.1.3 In S-44 2.4, errors for a 95% confidence level with regard to positions for other important details and coastal features are expected to be below the following limits:

TABLE 6.1 (TABLE 2 ON S-44)			
	SPECIAL ORDER	ORDER 1	ORDERS 2 and 3
Fixed aids and conspicuous objects for navigation	2 m	2 m	5 m
Coastline	10 m	20 m	20 m
Floating aids to navigation	10 m	10 m	20 m
Topographic features	10 m	20 m	20 m

2.1.4 Thorough checks must be conducted to confirm that the reference system used to show all the control point co-ordinates is the same. Verification should include an analysis of records and whenever doubts arise, field checking should be included.

2.1.5 To check positioning accuracies, a strict routine for cross checking between physically obtained control point details and the supplied co-ordinates should be instituted. This will avoid the situation of co-ordinates from measured closed circuits returning to the same control point being exclusively used; instead, other ways of ensuring the expected consistency should be included. Therefore, at least one connection that ensures the transfer of co-ordinates from one control point to another should be included in the applied measurements.

2.1.6 When satellite services (GNSS) are used for altimetric purposes, it should be ensured that, besides the accuracy of the process being undertaken, corrections between heights above the reference spheroid used and the mean sea level are accurate enough for level calculations to be made with errors below those in S-44 4.2, that is ± 5 cm for Special Order, and ± 10 cm for Orders 1, 2 and 3. The main purpose of this precaution is to meet the requirements directly associated with sea levels, water intakes or artificial outlets, surveying for coastal projects, ground control for photogrammetry, harbour surveys, etc.

Exceptions to these requirements are surveys intended to show the coastline from the sea, the sea-level positioning for conspicuous objects or the heighting of lights, signals and beacons where errors up ± 0.3 m are allowed for groups of signals (leading lines) and up to ± 0.5 m for an isolated signal or object. In the case of ground control points intended to define the coastline shape, the error tolerance can be ± 0.5 m for Special Order and ± 1 m for Orders 1 or 2, when the ground slope is below 10%. On steeper slopes error tolerance can be up to ± 1 m ±·0.8 iH, where **H** is the horizontal error, which is shown in Table 6.1 and **i** is the slope (elevation angle tangent).

2.1.7 The principle methods of coastlining are:

a. Real Time Kinematic with GNSS (RTK trough GPS, etc.);
b. Resection fixes (EDOM, sextant, theodolite, etc.);
c. Traverses (EODM, Total Stations, level and tachstaff, tachymetry or sextant and 10' pole)*;
d. Intersection (EODM, theodolite or sextant);
e. Air photography;
f. Existing maps.

(*) *In traverses with sextant and 10' pole, the horizontal angles are measured by sextant (see 5.3.1 at Chapter 2) as well as the distances with a special rod, where an angle is converted into a distance (parallactic method, through the measure between two separate marks of a known distance apart).*

2.1.8 The methods used will depend upon the scale of the survey, the time and the equipment available; i.e. existing maps, where small details can be shown, could well be used for scales of 1:50000 or smaller (1:100000). Similarly air photography can be used, but it is likely that such images will be reduced and interpreted as necessary at the National Hydrographic Office (NHO).

Photogrammetric restitution is also a suitable method (derived from aerial information), but it is advisable to complement this process with ground data collected during the field reconnaissance.

2.2 Positioning methods and Accuracies

2.2.1 GNSS (See 6.1 at Chapter 2)

Methods using single navigation systems are only applicable for cases for which, as shown in Table 6.1, errors of ± 20 m are acceptable. Using particular caution, including an experimental calculation for corrections to points known before and after surveying for periods over 2 hours between sunrise and sunset, it could be applied to cases which, according to the Table above, require ± 10 m accuracies as long as the calculation of such corrections are consistent with the given limits.

Methods using the observable codes in differential mode (DGPS, etc.) with reference stations at geodetic control points may be used for cases requiring ± 5 m for the highest accuracies. In cases with more accurate requirements (i.e. ± 2 m Table 6.1), the processes used should be phase measurement of carrier waves L_1, L_1/L_2, etc.

In these cases, the following possible vector errors should be considered:

TABLE 6.2

VECTOR LENGTH	L_1	L_1/L_2
Up to 10 Km	± 1 cm ± 1 ppm	± 1 cm ± 1 ppm
10 to 40 Km	± 1 cm ± 2 ppm	
40 to 200 Km	NOT APPLICABLE	
Above 200 Km		± 2 cm ± 2 ppm (*)

(*) *With suitable time periods, special equipment and software, errors may be below ± 1 cm ± 1 ppm.*

Regarding Table 6.2, it should be noted that upon the expected GNSS development from 2005, consideration should be given for updating to allow provision for the additional band L5 and the full operational reception compatible of GPS, GLONASS and GALILEO.

Likewise, the increasing potential of operating using real-time kinematic (RTK) mode suggests that its use may exceed present surveying capabilities and its use for some ground control positioning may be expected. For the present (2004), such techniques may be considered to have errors of ± 5 cm ±5 ppm.

Moreover, within GNSS development, not withstanding the above, new differential services, in addition to those already existing, are planned to come into operation:

- Ground - Based Augmentation Systems (GBASs) with transmissions from earth stations near airports as well as other intensively used sites;

- Satellite - Based Augmentation Systems (SBASs) with satellites receiving differential correction signals from different stations and then transmitting adjusted corrections. One of the most complete networks scheduled to become fully operational by 2005 is the so-called WAAS (Wide Area Augmentation System) sponsored by USFAA (US Federal Aviation Association).

Some of these services are operating with different characteristics, though they are expected to increase in number and introduce greater capabilities. This exploitation has generated the possibility of conducting more surveys without the need of establishing reference stations. Nevertheless, it is not advisable to be over optimistic with its application if there is no relatively near earth station contributing to the system. Another method is to implement active station networks, the reception of which is centralized and transmits precise ephemeris calculations which are applicable to a particular region.

Returning to coded differential equipment with base stations operating on control points, there are some which, by means of a so-called "sub-metric" treatment, may achieve errors in the order of 10 cm ± 10 ppm without strictly using the L phase of carriers and allow base-rover distances as much as 10 Km.

There is a wide variety of equipment on offer but very few fulfil such error limits. It is, therefore, advisable to check the procedures with a test by stationing them at several distances on the existing control points in order to obtain a reliable assessment.

For the remainder of this chapter, it is assumed that the equipment in use is measuring the phases of carrier wave(s) (L1 or L1/2) within the limitations stated in Table 2 and RTK mode error (± 5 cm ± 5 ppm) as mentioned.

Ideally, to carry out a topographic survey, all the points should be surveyed from base stations with control marks. Wherever existing control points are not sufficient it should be expedient to increase the density of them. Fig. 1 illustrates such a plan, i.e. from existing network marks, new control points are generated by GNSS vectors using geodetic receivers in a static relative mode. To make corrections to geodetic heights (above the spheroid), in order to obtain heights above mean sea level or others associated with it (see Chapter 2), it is necessary to tie in altimetric control points.

It is desirable that photogrammetric ground control points and aid to navigation signals are calculated, as a minimum, from two control points. Faster methods such as stop and go or real-time kinematic (RTK) modes may be applied both to these types of control points and to surveyed ground features, as long as they meet the requirements in Table 6.1.

If, while surveying, the need arises for the generation of additional control points, they should be obtained from two previously determined control points.

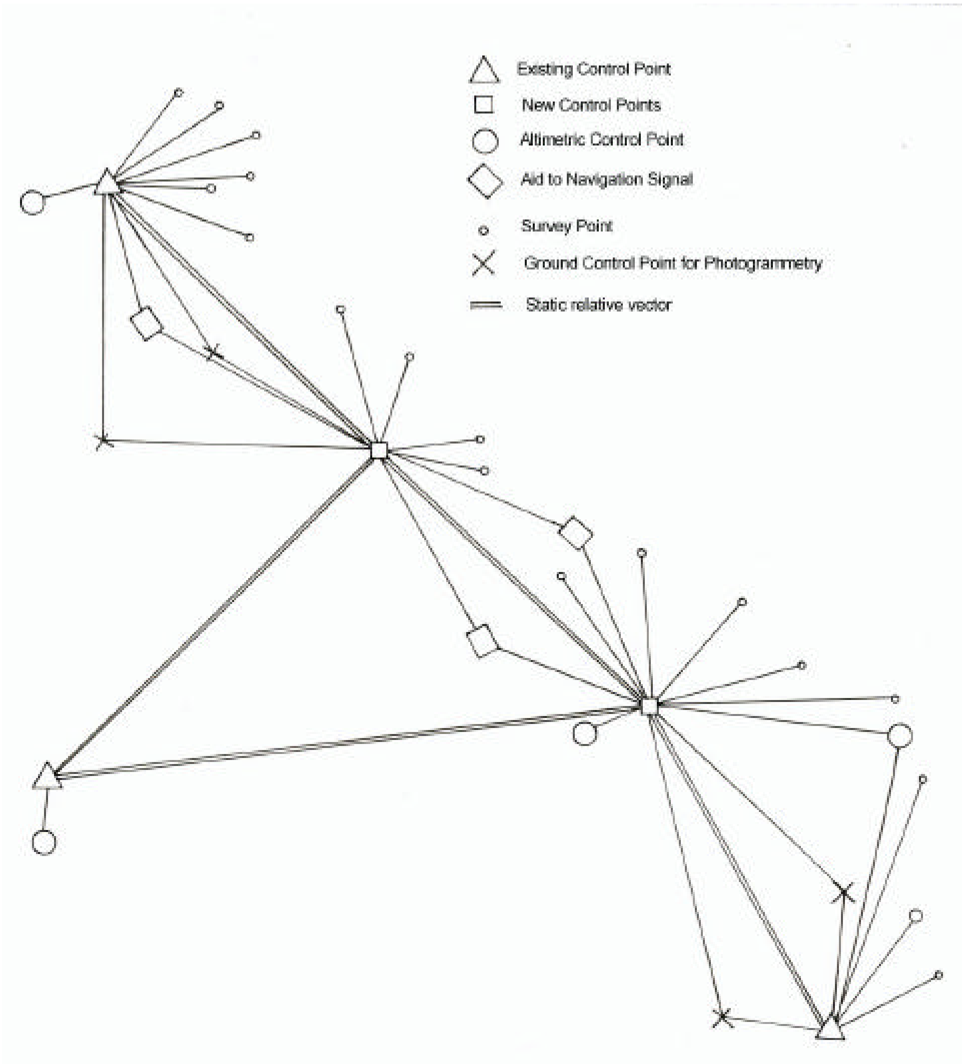


Fig. 6.1

The ease of creating new control points as well as the cost of building and preserving bench marks, or other marks, is setting the trend for minimum monumentation. In such cases, illustrations such as Fig. 6.2 may be chosen.

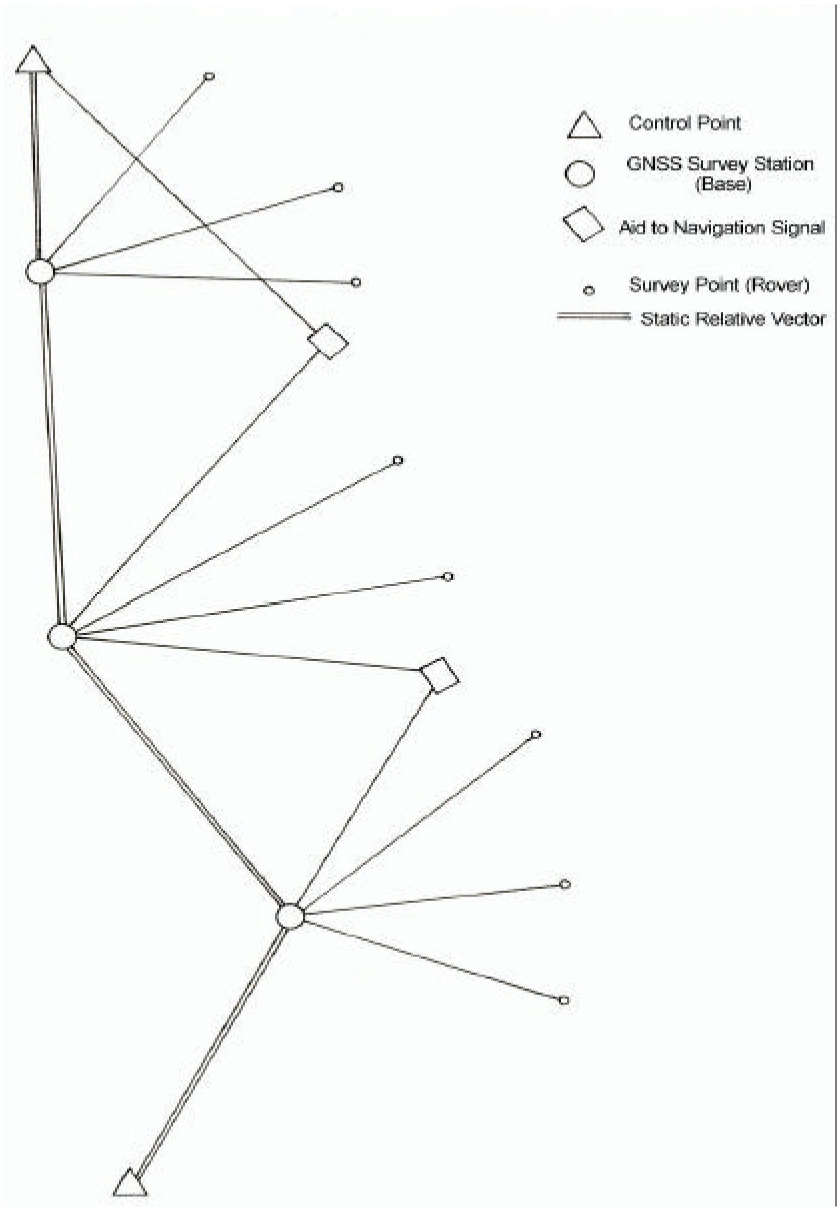


Fig. 6.2

Besides serving as a base for rover receiver reference, GNSS survey stations are connected to each other by static relative vector measurements thus forming, as a minimum, a traverse between control points without generating additional monumentation. In most cases, these traverses will have been measured with the same instruments as used to survey ground features.

2.2.2 Triangulation (See 3.2.1 at Chapter 2)

It is a technique based on principal angular measurements. Before the middle of the 20^{th} century, it was the most common method for establishing geodetic control networks and for sole calculation of conspicuous points, marks and other aids to navigation or photogrammetric ground control points. Since the 1960s Electronic Distance Measuring equipment (EDM) or Electro Optical Distance Measurement (EODM) has superseded the above methods. More recently they have been replaced by satellite methods, particularly since a permanent global coverage was established in the 1990´s

The earliest form of triangulation for hydrographic purposes consisted of a series of observations as in Fig. 3, with a relatively small number of measured sides (baselines) and a large number of angular measurements, which are showed here with the observed directions. Such a diagram provides a great deal of redundancy; each double-diagonal quadrilateral has three angular verifications created by adding or subtracting values. Nevertheless the network scale is still determined by the baselines.

In old unconnected geodetic controls, position and orientation were established from astronomic observations of latitude, longitude and azimuth in a datum. Now days, if marks are use from these kinds of networks, it is usually necessary to re-observe and recalculate via GNSS in order to convert co-ordinates into a universal system like WGS 84 (see 2.2.3 at Chapter 2).

In general it should be noted that distances from the baselines could be measured within accuracies ranging from 1ppm to 3ppm, directions from ± 0.5" to ± 2", and transition from a base to another (that is, the contrast between the base transfer by triangle resolution and the other measured base) could normally be checked within 20 ppm and 40 ppm.

These limitations should be taken into account when trying to adjust an old triangulation network to a present framework via GNSS observations, with distances of 200 or 300 km there can be differences of several meters (2 or 3). Besides tolerating differences of these orders, it is necessary to have a sufficient number of well-distributed connections to common datums and of datum conversion algorithms in order to absorb the distortions typical of the old networks (see Chapter 2)

Not withstanding the above statement, densification by GNSS of datums with fixed co-ordinates computed from old triangulations should be avoided; such cases often lead to distortions and inaccuracies in the final results. If unavoidable due to the need to keep the co-ordinates of an old datum, it will be necessary to adopt very particular computation strategies and the limitations of the values obtained must be stated at an early stage.

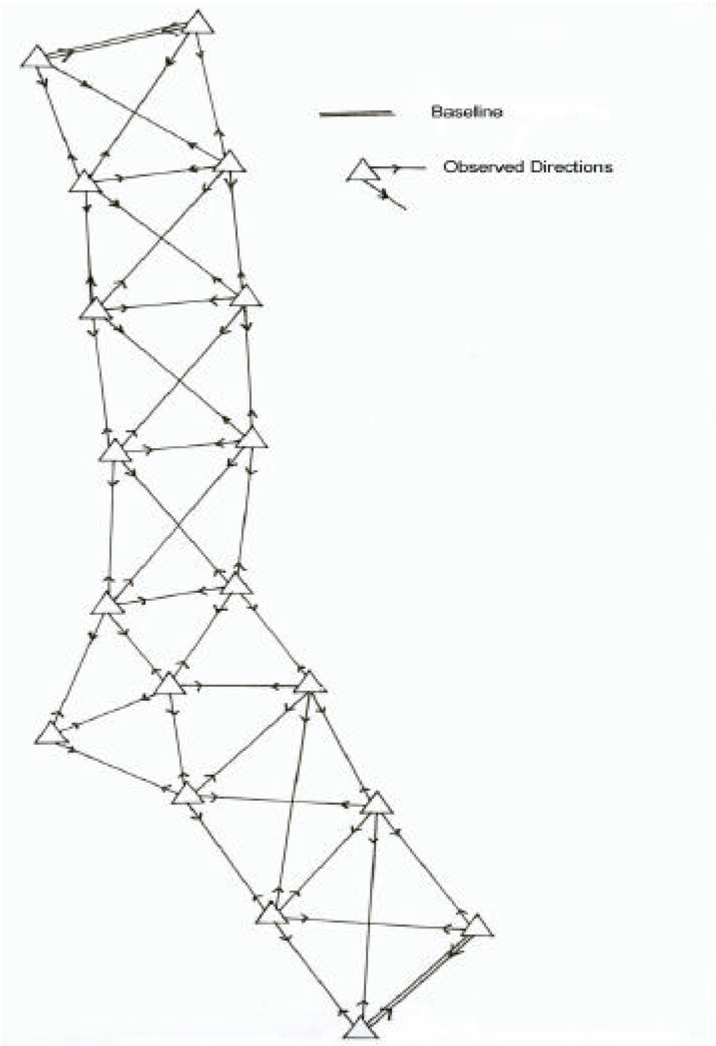


Fig. 6.3

A control network with the characteristics as in Fig. 6.3 had, in general, sides with lengths ranging from 15 to 25 km, 18 km on average, with triangle closure errors from ± 1" to ± 2"; this was termed first-order triangulation. The following densifications had shorter sides (10 to 15 km) with closure errors from ± 2" to ± 4"; they were designated second-order triangulation. There were also third-order and fourth-order triangulations with shorter sides and higher tolerances, ± 5", for third-order triangulation, and ± 10", for fourth-order triangulation. Table 6.3 details typical values and aspects of these orders.

Table 6.3

TRIANGULATION MEASURES CHARACTERISICS					
ORDER	SIDES LENGHTS (Km)	TYPICAL THEODOLITE DIRECT READING ERROR (“) (*)	TYPICAL NUMBER OF REITERA TIONS (*)	TYPICAL DIRECTIO N ERROR (“)	TRIANGLE CLOUSURE ERROR TOLERANCE (“)
1st.	15 to 25	0.1 to 0.2	9 to 18	0.1 to 0.5	1 to 2
2nd.	10 to 15	1”	6 to 9	1 to 2	2 to 4
3rd.	5 to 10	1” to 10”	4 to 6	2 to 3	5
4th	2 to 10	10”	2 to 4	5	10

(*) *See 5.3.2 at Chapter 2*

For each order of work, the co-ordinates of the higher orders were taken as fixed co-ordinates and generally the baselines and astronomical stations were exclusively for the highest two orders.

In lower order work, it was normal to select a few higher order points at a time, as in Fig. 6.4 left; though in some cases for control densification networks selection of a larger number of points with shorter side observations (Fig. 6.4, right) was carried out, particularly whenever triangulation towers had been removed. These towers were used to elevate the line of sight over trees, topographic features and other obstacles interfering with the observations. Obviously their removal prevented long sights from being conducted which led to this type of solution.

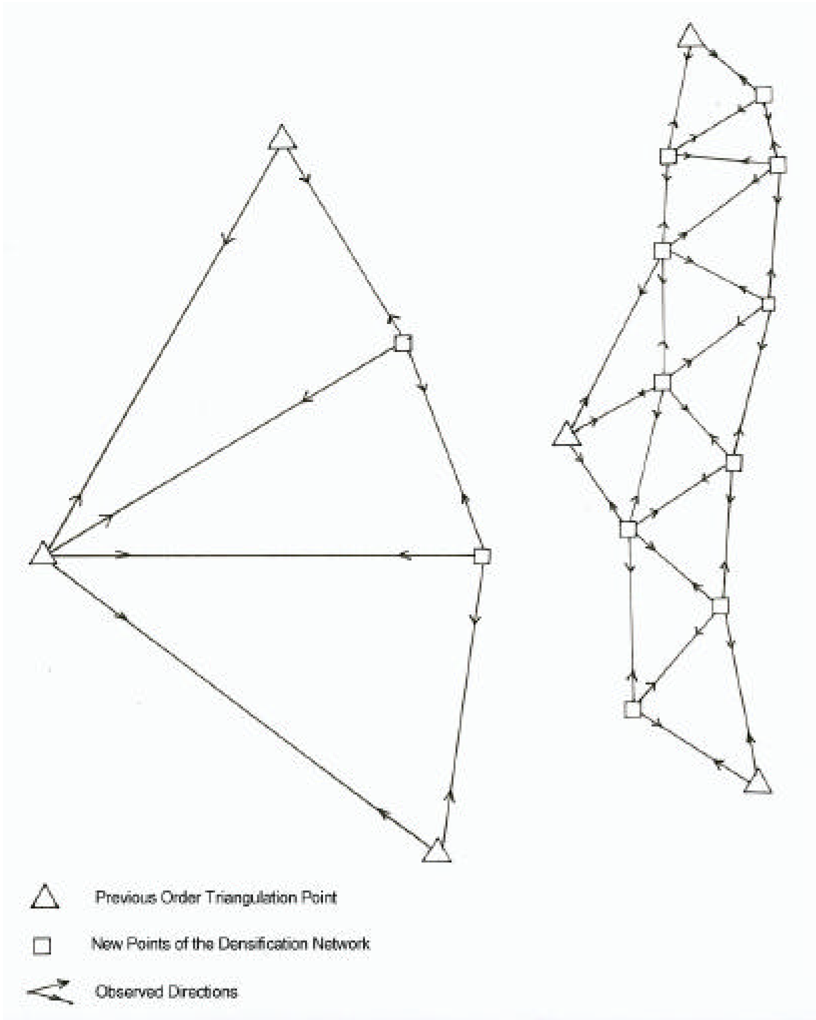


Fig. 6.4

In some cases, in hydrographic surveying, the term *triangulation* has been used to describe survey points with flare triangulation by intersection (see 2.2.4). These flares were dropped with parachutes vertically over the point to be located and, whilst burning, directions from control points would be simultaneously observed; this was conducted towards many coastal points requiring to be surveyed, as many were not visible. Balloons, luminous shots or mobile, elevated signals were also used for the same purposes.

The term triangulation has also been used when measuring angles to control points with a sextant, sometimes in combination with observations from such points. The exclusive use of observations from points to be calculated is treated as a resection in 2.2.4.

Although these survey techniques are becoming obsolete due to intensive use of other more responsive modern methods, they still are effective.

One of the typical problems of triangulation is the error propagation dependence on the figure shapes, on which the error results (positions) rely not only on the measurement error but also on the network geometry. This problem is dealt with for particular cases in 2.2.4 though it admits more complex generalisations. For example, a single chain of equilateral triangles is more rigid than a chain with unequal angles. Also, a two-diagonal square chain is more rigid than a chain with rectangles or trapeziums with similar diagonals.

2.2.3 **Traverse** (See 3.3.1 to 3.3.4 at Chapter 2)

Before the 1950-1960 decade, use of combined distances and directions was restricted to small areas but later, with the development of EDM and EODM equipment, larger networks with longer sides (5, 10, 15, 20, ... km) could be created. As stated at the beginning of 2.2.2, these techniques superseded triangulation.

For some time (about 1960) a new technique based on exclusive side measurement (trilateration) (See 3.2.2 at Chapter 2) was considered but it was quickly rejected, mainly due to a lack of internal checks. To clarify this concept, a single triangle has an angle closure condition while a trilateral of the same shape has no way of being checked; a quadrilateral with two diagonals and all its directions measured, as stated in 2.2.2, has four closure conditions whilst the same trilateration geometrical figure with its 6 sides measured has only one verification. This advantage with triangulation is limited since the method requires some sides being measured (baseline); however, trilateration can be conducted without observing any angles.

A combination of both techniques resulted in a suitable solution, although sometimes termed triangulateration, here it will be termed traverse, although often a traverse may be a simple succession of measured angles and distances.

One of the most important properties of traverses is that error propagation is independent from configuration; that is there is no requirement for complex network design involving suitable geometries or erecting towers to facilitate certain lines of sight. From the practical point of view with this kind of network, uniformity of control points with survey stations or aid to navigation requirements was possible.

In general, it is advisable to maintain a reasonable balance between the accuracies of both types of measurements (directions and distances) in order to improve the geometry independence in relation to the accuracy of results. One of the applicable rules is

$$\frac{\sigma DIST}{DIST} = \frac{\sigma ANG}{200000}$$

where $\sigma \boldsymbol{DIST}$ is the distance standard deviation stated in the same unit as **DIST**, and $\sigma \boldsymbol{ANG}$ is the standard deviation of a measured direction stated in sexagesimal seconds. Then, for $\sigma ANG = \pm 1''$ the distances required are 5 ppm (1/200000) and for ± 4", 25 ppm (1/40000) is enough.

The required angular or distance errors must never be confused with the instrument reading or resolution capabilities. The observer limitations, the environmental conditions, the correction accuracies, the time when measurement were made, etc., must also be considered.

For example, for an inclined distance measurement with an angle of elevation of 20º and 5 km in length, with an elevation difference error of ± 0.5 m, the error in its horizontal projection is expected to be

$$0.5 \text{ m} \tan 20° = 0.18 \text{ m}$$

Thus, in spite of being measured with EODM equipment, whose error may be in the order of ± 1 cm ± 2 ppm, and with the inclined distance error of ± 2 cm, if used for transferring a horizontal position, the error is ± 18 cm.

A distance measured with EDM equipment must be corrected for environmental conditions (pressure, temperature, humidity).

Humidity is calculated according to pressure and temperature with dry and wet bulb observations, it is very important for measurements taken by microwaves. No measurement should be taken with EDM in an oversaturated atmosphere (rain, drizzle, fog); with EODM measurements humidity is not so important, although the luminous wavelength used should be considered. LASER radiations have an advantage since they are basically monochromatic, it is generally sufficient to obtain pressure and temperature data. For long distances (more than 5 km) it is recommended that the environmental parameters at both ends of the distances to be measured are obtained and then averaged.

Manufacturers usually provide the instructions for making the necessary corrections to their equipment. In EODM, the reflector prisms should be used with the equipment with which a calibration was conducted to avoid errors in measured distances, sometimes above 1 cm.

In distances above 5 km corrections for earth and ray curvature must be made. Such correction is:

$$+\frac{(1-k)^2}{24R^2}D^3$$

Where **k** is the refraction coefficient (rate between earth and ray radius). In mean conditions it is 0.25 for microwaves and 0.13 for luminous waves. It is sufficient to introduce its approximate mean value as earth radius.

$$R = 6371000\text{m}$$

Figure 6.5 illustrates **D** (measured distance) and **S** (reduced distance to reference surface) meanings. This is necessary for the above correction and the correction for point elevations, which is detailed below.

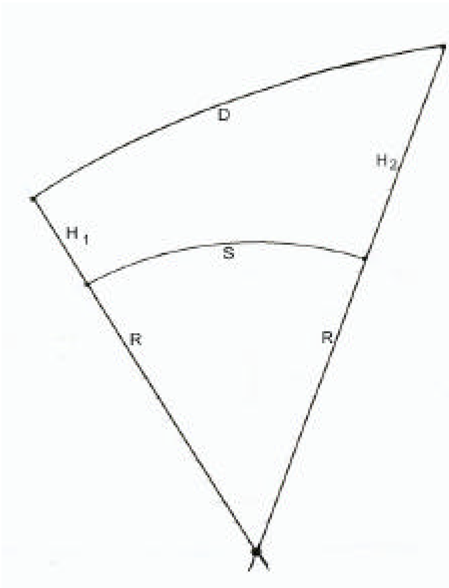


Fig. 6.5

It is important to note that the above correction for curvatures takes into account the geometric effect of both arcs as well as the physical influence produced as a consequence of ray propagation at a slightly lower level than that for mean environmental conditions at both ends.

Correction for ray elevation and inclination is more significant. The general expression is:

$$S = \sqrt{\frac{D^2 - (H_2 - H_1)^2}{(1+\frac{H_1}{R})(1+\frac{H_2}{R})}}$$

The way in which such elevations are obtained, especially their difference ΔH = (H2 - H1), affect the correction error. By simply considering the numerator:

$$S \approx \sqrt{D^2 - \Delta H^2}$$

we can deduce the influence:

$$dS \approx \frac{\Delta h}{\sqrt{D^2 - \Delta H^2}} d\Delta h = id\Delta H$$

previously mentioned. Therefore, the elevation difference error must be multiplied by the slope,
i = tan a , in order to obtain the influence produced on the corrected distance error.

In topographic surveys it is usual to make calculations in plane co-ordinates; for this purpose it is necessary to have previously applied corrections to the projection plane. The way in which this type of corrections may be calculated is detailed in 2.2.5.

The general and most correct way of calculating a traverse network on the representation plane is via a previous calculation of the co-ordinates for every new point commencing with the values of the known points and the uncorrected observations. It is necessary to average some results taken from different points starting with the additional redundant observations. When the provisional co-ordinates have been accepted, the above corrections should be applied and then observation equations are to be obtained, the unknown quantities of which are the corrections to the co-ordinates, in order to conduct a least squares adjustment.

If any observation exceeds tolerance levels (maximum admissible error) the original records should be checked, if no apparent cause is found regarding the error source re-measurement should be considered. If there is sufficient redundancy, the erroneous observation may be removed and a new adjustment conducted.

In some basic traverse circuits an approximate adjustment may be achieved by distributing the angle closure error first and then the co-ordinate closure error proportional to side length or some other logical criteria.

Angle closure errors in traverses must be below:

$$\pm\left(5''+2''\sqrt{n}\right)$$

where **n** is the number of angular stations making up the circuit. In secondary traverses, intended to increase the density of the control points, the error may extend to:

$$\pm\,(10''+10''\sqrt{n})$$

When the purpose is limited to fixing co-ordinates of coastal details, larger tolerances may be permitted.

Co-ordinates closure errors should be no more than the values stated in Table 1 depending on the use of the network, noting that adjusted co-ordinates for intermediate points will have errors in the order of half the closure error. Nevertheless, for control networks, closure errors should not be greater then ±(0.2 m + 10 ppm) for primary horizontal control and ±(0.5 m + 100ppm) for secondary stations to meet the requirements in 2.1.2.

When errors are greater than a traverse tolerance, there are methods available to assist in locating the error source. For example, when an angle closure error is detected the grid bearing of the suspect side is calculated from all components of the co-ordinate closure error. However, if there is a large angle error, the angles should not be adjusted, on calculating the traverse in both directions, only at the affected point will the co-ordinate values roughly match.

When a network is accurately adjusted by least squares from provisional co-ordinates, the process enables, from the variance-covariance matrix, the calculation of adjusted co-ordinate errors. A similar calculation in a traverse may not be so clear since co-ordinate closure errors are more general. In such cases, the mid points may be permitted to have an error in the order of half the closure error, decreasing towards each end.

Traverse calculations in plane co-ordinates are very simple. The initial grid bearing is obtained from increases in ΔE, ΔN. Two control points whose co-ordinates are known in advance are represented as **P** and **Q** in Fig. 6. Then:

$$\operatorname{tg} B_{PQ} = \frac{E_Q - E_P}{N_Q - N_P} = \frac{\Delta E_{PQ}}{\Delta N_{PQ}}$$

where $\Delta \mathbf{E_{PQ}}$ and $\Delta \mathbf{N_{PQ}}$ signs (+/-) also define the quadrant.

If true azimuth referred to true north, rather than grid north as the referred available orientation, with grid declination γ (the definition of which is given in Annex A), this be should be taken into account. From this point forward only plane orientations (grid bearing) will be considered. Moreover, if a Transverse Mercator projection is used, it is assumed that corrections to observations (distances and directions) for the plane of representation have been made according to specifications in 2.2.5.

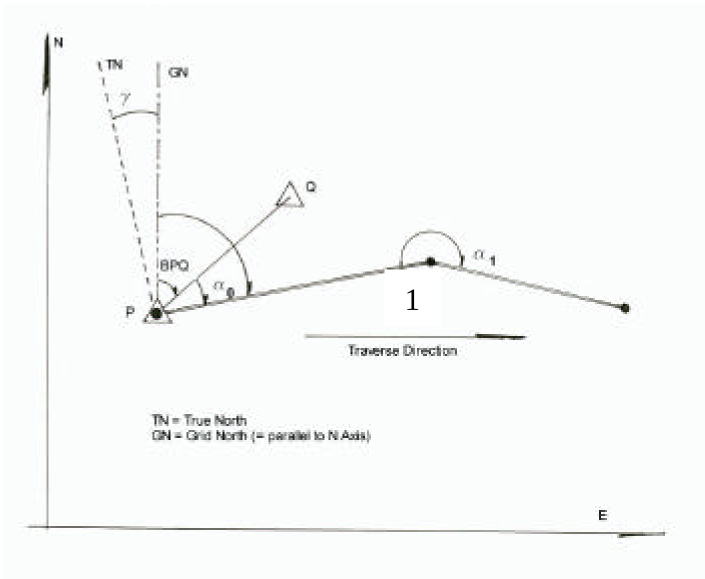


Fig. 6.6

Returning to computing a traverse, the grid azimuth of the first side is obtained by simply adding the first angle:

$$B_{P1} = B_{PQ} + \alpha_0$$

And using the following general form of bearing transfer:

$$B_{12} = B_{PQ} + \alpha_1 \pm 180^\circ$$

The sign **+** is used in case the previous addition ($\mathbf{B_{P1}}$ + α_1) is below 180º and the sign **-** when it is above. The latter is the most common case.

Increases in co-ordinates are obtained with the expressions for converting polar co-ordinates into plane co-ordinates:

$$\Delta N = S \cos B$$
$$\Delta E = S \operatorname{sen} B$$

It should be remembered that, in the cases of simple traverses, before making such conversions, it is normal to adjust an angle by distributing the closure error if it is below the given tolerance level. In more complex traverses, network calculations may be supplemented with the algorithms related to intersections or resections according to descriptions in 2.2.4 and 2.2.5. Adjustment requirements mentioned above should also be considered.

As regards adjustments, their respective methods will not be developed further, since such processes are expected to be developed at the NHO where appropriate software is available. It should be remembered, however, that good results may be achieved only if the data is checked in the field to ensure that closure errors or the calculation of point co-ordinates carried out by different methods show an acceptable consistency with the above specifications.

A simple traverse is deemed to be fully closed if it starts from a pair of control points and ends at another pair. There are then three possible closure errors available: one angle closure error and two co-ordinates closure errors. This case is illustrated at the top of Fig. 6.7; it allows an initial angular adjustment and a subsequent distribution of the differences in co-ordinates. There is a special case of a simple closed traverse which makes a circuit, starting and ending at the same point. Even though it may be properly checked as detailed above, it is not advisable to conduct such methods for the reasons set out in 2.1.5

A simple traverse is termed half closed when a direction to another control point has not been measured from the final point; this means that no known angle checking or its corresponding adjustment is permissible. Nevertheless, if co-ordinate closure errors are acceptable, a similar distribution as in the previous case may be carried out as illustrated in the second case of Fig. 6.7.

A simple traverse is deemed to be precariously closed when, although it starts and ends at control points, there is no final measured direction with an orientation. The only check is to confirm that the measured distance between the control points P and R generated from the traverse is fairly consistent with the distance calculated from their known co-ordinates; this illustrated in the third case in Fig. 6.7. The simplest way of calculating the distance is by giving it an arbitrary or approximate orientation for the initial calculation and by then rotating the orientation and adjusting the length according to the differences to the end point.

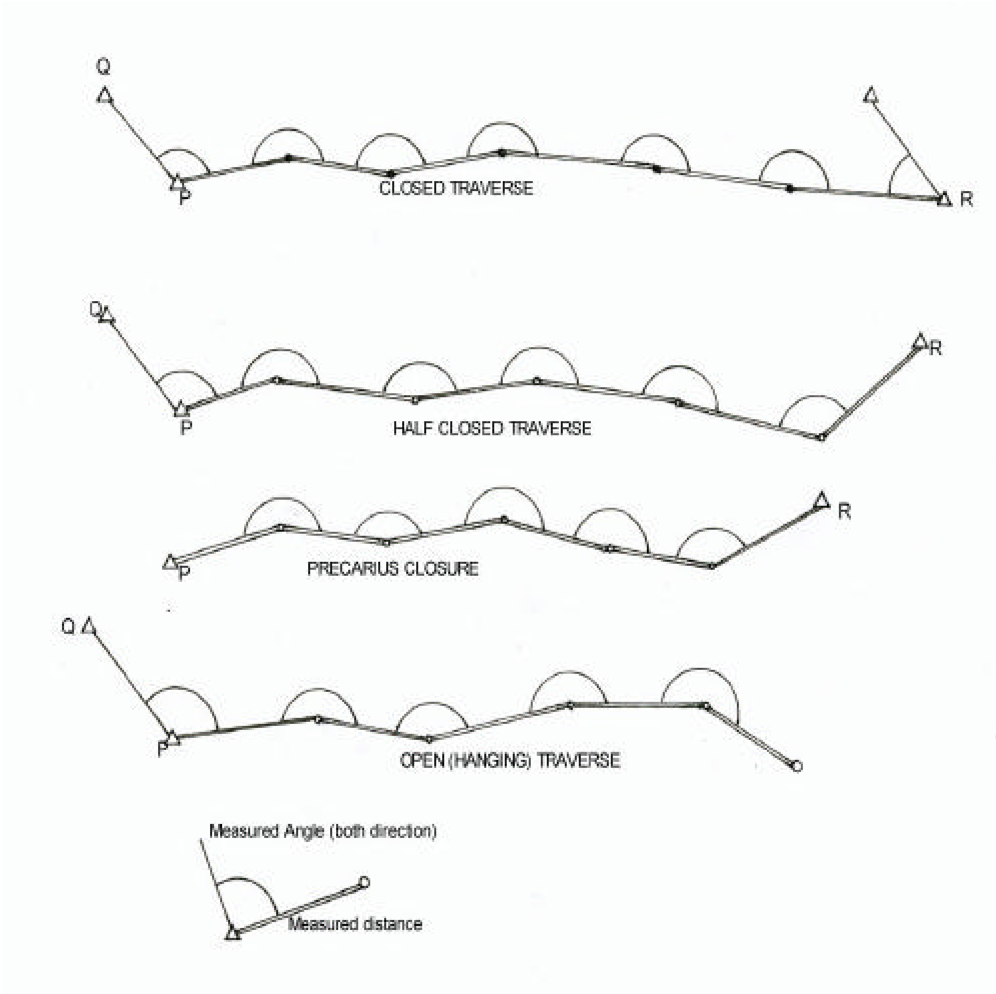


Fig. 6.7

A simple traverse is deemed to be an open, not closed, or hanging traverse only if it starts from known points but ends at new unknown marks, thus no closure verification or adjustment can be made; this is not a recommended configuration. When it is the only choice, extreme caution should be exercised and the temporary nature of subsequent results should be clearly stated.

Traverses are closely associated with trigonometric levelling operations. These consist of a series of measurements taken to determine differences in elevations by vertical angles. (See 4.2 at Chapter 2)

The most precise way of obtaining a trigonometric difference of elevation consists in measuring the direct distance between the points and the vertical angles reciprocally and simultaneously from both stations:

$$\Delta H_{12} = \frac{i_1 + s_1}{2} - \frac{i_2 + s_2}{2} + D\,\mathrm{sen}\left(\frac{a_1 - a_2}{2}\right)$$

where (see Fig. 6.8):

i_1	= theodolite height above bench mark in point 1;
s_1	= signal (target) height above bench mark in point 1;
i_2, s_2	= theodolite and signal heights above bench mark in point 2;
D	= slant and elevated distance (see Fig. 6.5);
ΔH_{12}	= difference of elevations between benchmarks 1 and 2.

The elevation angles (a) are positive when they are above the horizon and they are negative when below the horizon. In Fig. 6.8 a_1 is the positive angle and a_2 is the negative angle. It is necessary for both to be simultaneously measured in order for a correct adjustment of ray curvature, which changes throughout the day.

A trigonometric difference of elevation obtained under these conditions may have an error of

$$\pm 0.01 \text{ m} \cdot K$$

where **K** is the distance expressed in kilometres, which is an error of 1 cm/Km.

If slant distance (**D**) has not been measured and ground distance reduced to the reference level, commonly the mean sea level, is available, which is the case of triangulation or intersection (see Fig. 5), the formula to be applied is:

$$\Delta H_{12} = \frac{i_1 + s_1}{2} - \frac{i_2 + s_2}{2} + S\left(1 + \frac{Hm}{R}\right)\mathrm{tg}\left(\frac{a_1 - a_2}{2}\right)$$

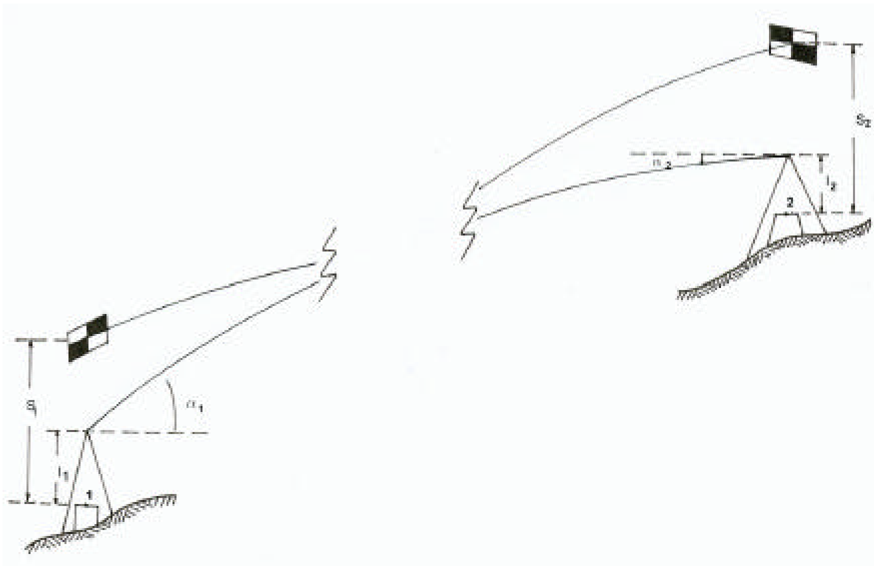


Fig. 6.8

If the elevation angle is known only at point 1, the formulae to be applied are:

$$\Delta H_{12} = i_1 - s_2 + D \operatorname{sen} a_1 + \frac{(1-k)}{2R} D^2$$

$$\Delta H_{12} = i_1 - s_2 + S\left(1 + \frac{Hm}{R}\right) \operatorname{tg} a_1 + \frac{(1-k)}{2R} S^2$$

In the last three formulae, **R** is the terrestrial mean radius, in principle 6371 Km, but a more correct value relative to latitude and azimuth may be used for the adopted spheroid. The same is valid for the formula above to transfer **D** to **S** (See Fig.6.5)

$$S = \sqrt{\frac{D^2 - \Delta H^2}{\left(1 + \frac{H_1}{R}\right)\left(1 + \frac{H_2}{R}\right)}}$$

$$Hm = \frac{H_1 + H_2}{2}$$

Hm is the mean elevation

If only $\mathbf{H_1}$ is available, it may be calculated as:

$$H_m = H_1 + \frac{\Delta H_{12}}{2}$$

Where $\Delta \mathbf{H_{12}}$ is obtained by an iterative process which improves the value $\mathbf{H_2}$.

Coefficient **k** has the above stated meaning and it can be considered to have a value of:

$$k = 0.13 \pm 0.05$$

then the error of a non-reciprocal trigonometric difference of elevation may be:

$$\pm (0.01 \text{ m K} + 0.004 \text{ m K}^2)$$

The use of trigonometric levelling is ideal both for reducing the sides due to differences of elevations and height and for other altimetric requirements to overcome possible accuracies.

2.2.4 Intersection and Resection

The most general form of intersection is when directions are observed from two control points into a mark, whose co-ordinates are required. Directions in orientation mean that directions are measured from the same stations to other known points, it then being possible to obtain the grid bearings of both directions. In some very special cases these are astronomic or gyroscopic orientations; in such cases it is required to go from the true azimuth to the grid bearing by applying the grid declination shown as g in Fig. 6.6.

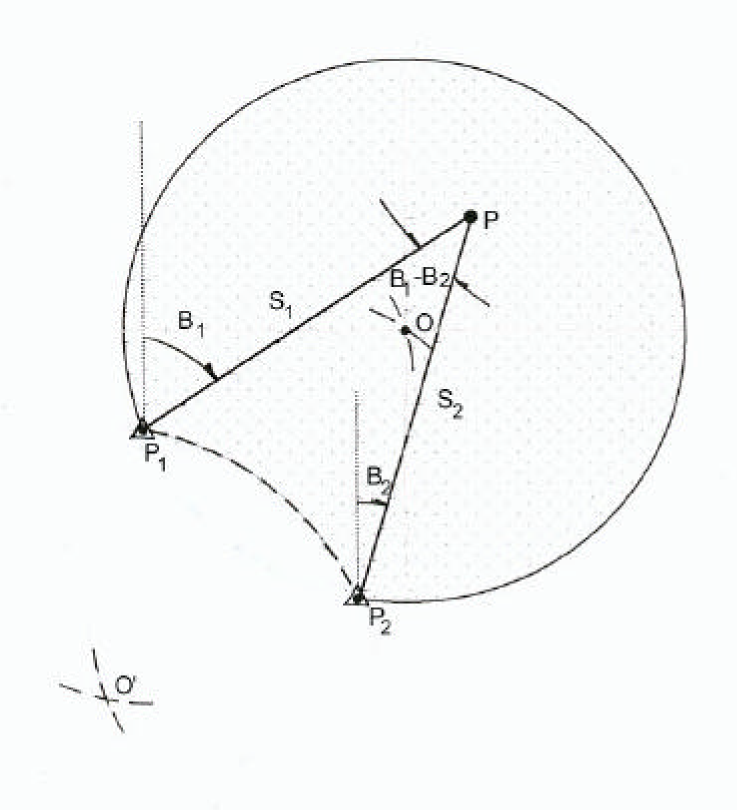


Fig. 6.9

Fig. 6.9 shows a typical intersection example. It should be made clear that in some cases, especially in short distances, reciprocal directions between known points (P_1 - P_2; P_2 - P_1) are used as the origin of plane orientations (B_1; B_2) to the point intended to be determined (P).

Besides grid bearing errors arising mainly from angular errors, distances (P_1 - P ; P_2 - P) and the angle between these directions, which is equal to the difference (B_1 - B_2), contribute to the errors in the co-ordinates of P. The simplest rule is that the angle should range between 30º and 150º. The area for this condition is shaded in Fig. 6.9 and corresponds to the limits of two circles centred at O and O' which are obtained as the vertices of two equilateral triangles with a common side P_1P_2.

Outside this area, errors largely increase to reach indetermination for B_1 - B_2 when equal to O° or 180°.

Another intersection case is shown when distances are measured from P_1 and P_2 to the point to be determined (P). These distances (S_1; S_2) define two symmetrical solutions as regards axis P_1 - P_2. To solve this ambiguity it must be known if P is on the left side of P_1 to P_2 (this is the case in the figure), or on the right side (a symmetrical case). An alternative solution is to note, when seen from P, which is the known point on the right or on the left (in the case of Figure 6.9, P_1, is on the right and P_2 on the left).

Algorithms to make corrections to the plane and obtain the co-ordinates of P, taking into account the cases stated, are shown in 2.2.5.

In cases of intersection, directions (straight lines) or distances (arcs), the best solutions are obtained when the crossing angle (B_1 - B_2) tends to 90°. In these cases the error ellipse tends to be a circle. Strictly speaking, taking into account that errors in measured directions and distances increase their influence with distance and such ideal solutions slightly differ from the 90º rule, its use, however, is a good way to quickly examining the suitability of the set up.

The most common resection case is when three known control points are observed from a new point, as in Fig. 6.10. This case is usually known as Pothenot-Snellius resection.

In this case, indetermination occurs when the circumference of a circle passes through the three known points. The same angles (a , b) to the control points can be measured to any point located on that line. It is relatively easy to avoid this situation by plotting on a chart the known control points and seeing if they lie on a circle centred on the unknown point. Another method is to check the addition:

$$\alpha + \beta + \omega$$

If it is near 180° such situation must be avoided.

The algorithm to solve this case, including corrections to compute on the projection plane, is shown in 2.2.5.

Resections have been very frequently used by hydrographers, both in topographic surveying by theodolite and hydrographic surveying by sextant. The advantage being that it was only necessary to put signals on the control points, the surveyor then being free to carry out his tasks without the assistance from ashore.

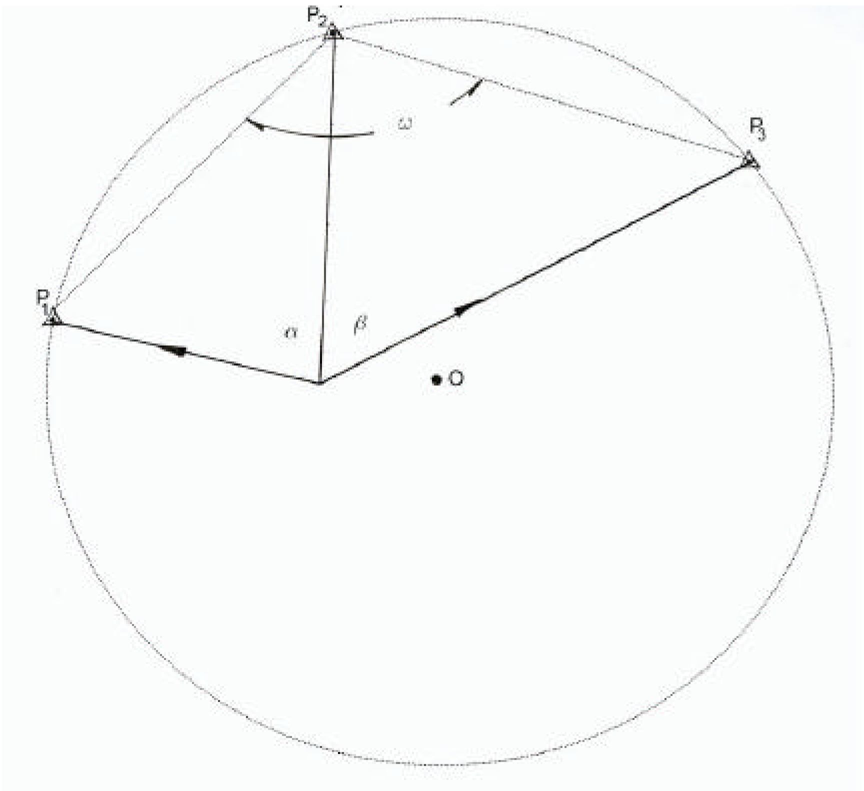


Fig. 6.10

It is possible to present multiple resections as generally given in Fig. 6.11.

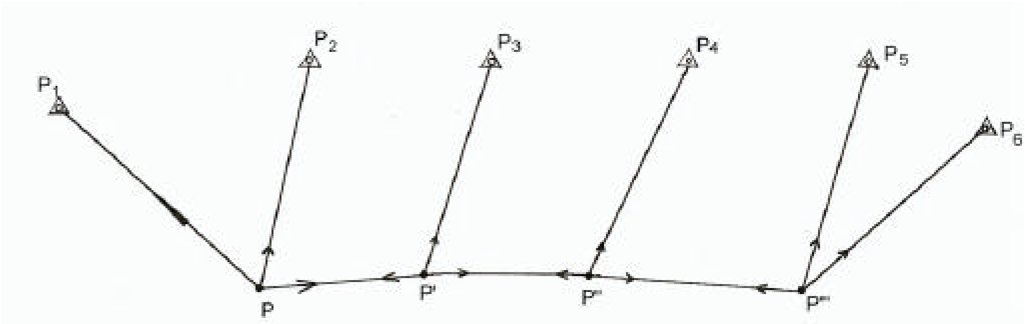


Fig. 6.11

That is from new points P, P', P'', P''', ... directions to known points P1, P2 ... P6 are seen. In such cases it should be noted that in the first and last points, two known control points are seen; at the intermediate points, besides reciprocal directions, a sight to one of those known points is sufficient.

Where there are only 2 new points and 4 control points are seen, it is known as Marek solution. If only the two control points seen from two new points are used, it is called Hansen solution. These particular cases are illustrated in Fig. 6.12.

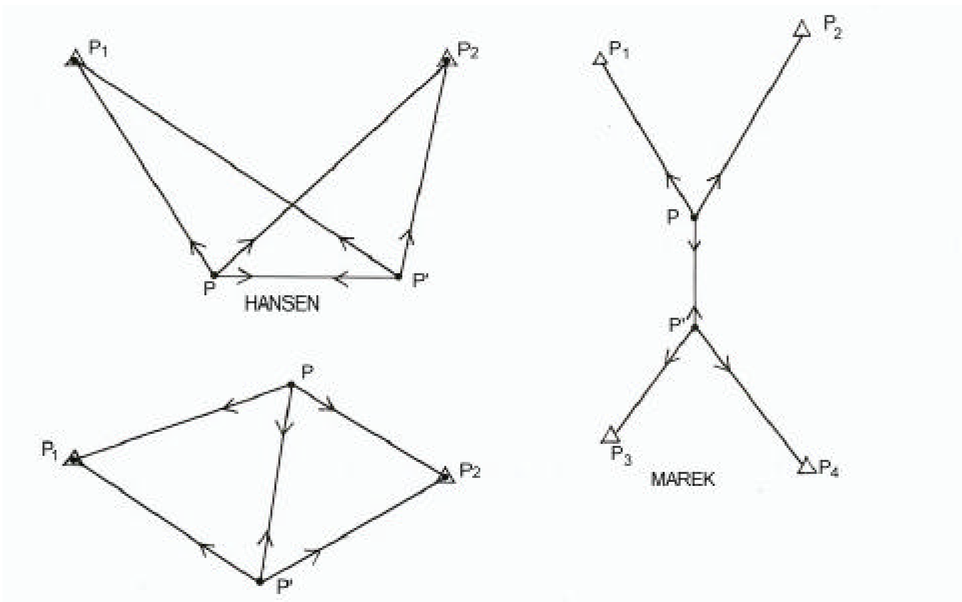


Fig. 6.12

Even though these multiple resection cases may be used whenever required, they are not recommended due to their limited opportunities for checking. A simple solution to apply is by incorporating additional sights to provide redundancy and the opportunity to check.

Therefore more than three directions to known points should be seen from every new point, or that new points will be interconnected by reciprocal lines of sight, as shown in Fig. 13; even though every new point is determined by directions to three known points, reciprocal lines of sight between new points include them in calculations of adjacent points.

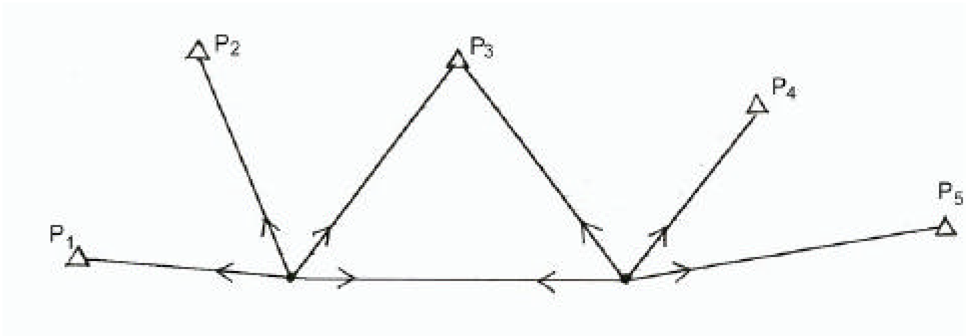


Fig. 6.13

Configurations originally close to indetermination can be improved in this way.

Solutions of this kind require some type of adjustment, whether rigorously by least squares or by iterative average of several positive solutions by trying to give more strength to cases further away from the situation of indetermination.

2.2.5 Usual Algorithms

a) Corrections to projection plane (See ANNEX A)

One of the processes to be conducted for calculations with rectilinear figures on the representation plane to be correct is related to the corrections that must be applied to measured observations (distances and directions). In this section we shall deal with the Gauss Krüger projection, also known as Transverse Mercator, which is often used for topographic calculations.

ANNEX A deals with the nature of this projection for cases of "tangent cylinder", that is those in which distance deformation starts on the central meridian:

$$m = \frac{ds'}{ds} = 1 + \frac{x^2}{2R^2} + ...$$

where **x** is the east co-ordinate referred to the central meridian:

$$x = E - X_0$$

when a false easting value $\mathbf{X_0}$ is used.

If this coefficient is applied between two points 1 and 2 (not infinitely near) a relationship:

$$\frac{S'}{S} = 1 + \frac{x_1^2 + x_1 x_2 + x_2^2}{6R^2}$$

is obtained. It should be noted that if a point is on one side of the central meridian and the other point on the other side, the product $\mathbf{x_1} \cdot \mathbf{x_2}$ will be negative.

Also that **R** (mean terrestrial radius) must be calculated for mean latitude of the working area and the representation system includes a coefficient (**K**) in order to contract distances over the central meridian, as in the case of UTM representation (where K = 0.9996, see ANNEX A). The coefficient for reducing distances (to obtain the plane value by multiplying it by the geodetic value over the spheroid) must be affected by the same value:

or

$$\frac{S'}{S} = K\left(1 + \frac{x_1^2 + x_1 x_2 + x_2^2}{6R^2}\right)$$

$$S' = K\left(1 + \frac{x_1^2 + x_1 x_2 + x_2^2}{6R^2}\right)S$$

Measured directions also require the application of a correction. This need arises from the fact that geodetic lines (on the spheroid) on being transferred to the plane, are represented by a slight concavity towards the central meridian.

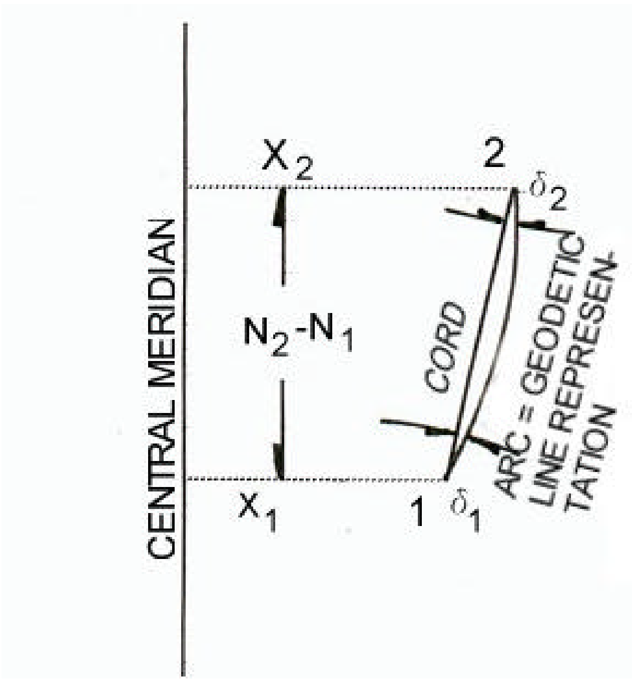


Fig. 6.14

Fig. 6.14 shows such curvature and the corrections which should be applied to pass from the arc, corresponding to the geodetic line, to the cord, corresponding to the side of a rectilinear figure on the plane. On accepting the correction sign to pass from the arc to the cord, we may see that:

$$(d_2 - d_1) = \frac{r''}{2R^2}(x_1 + x_2)(N_2 - N_1)$$

since absolute value addition for such corrections must be equivalent to the quadrilateral spherical excess the surface of which is 1/2 $(x_1 + x_2)$ $(N_2 - N_1)$ and r " is the typical constant to pass from radians to sexagesimal seconds ($\rho'' = 206265''$).

As arc curvature increases with **x** values, naturally the **x** of the known station point carries more weight than that of the observed point. Then:

$$d_1 = \frac{r''}{6R^2}(2x_1 + x_2)(N_1 - N_2)$$

$$d_2 = \frac{r''}{6R^2}(2x_2 + x_1)(N_2 - N_1)$$

and the difference between them leads to the first expression $(d_2 - d_1)$.

In general terms, if there is a need to reduce a series of directions to points Pi measured from a point Po, the corrections (along with their sign) are:

$$d_i = \frac{r''}{6R^2}(2x_0 + x_i)(N_0 - N_i)$$

It should be noted that westward of the central meridian **x** values are negative; thus the correction sign generating a change in concavity is automatically modified. On the assumption that the direction between the known station and the observed point are on different sides of the meridian, the change in the **x** sign will decrease the d value. This is logical since the geodetic line will have a curvature inversion in order to keep above concavity.

For calculations of corrections for both distances and directions, it is normal to make a preliminary calculation of the mark’s co-ordinates and ignore any deformations. Corrections are estimated using these provisional co-ordinates and then the final calculation is undertaken. In some cases provisional co-ordinates are used for adjustment; however this will not be dealt with further.

b) Intersection of Directions.

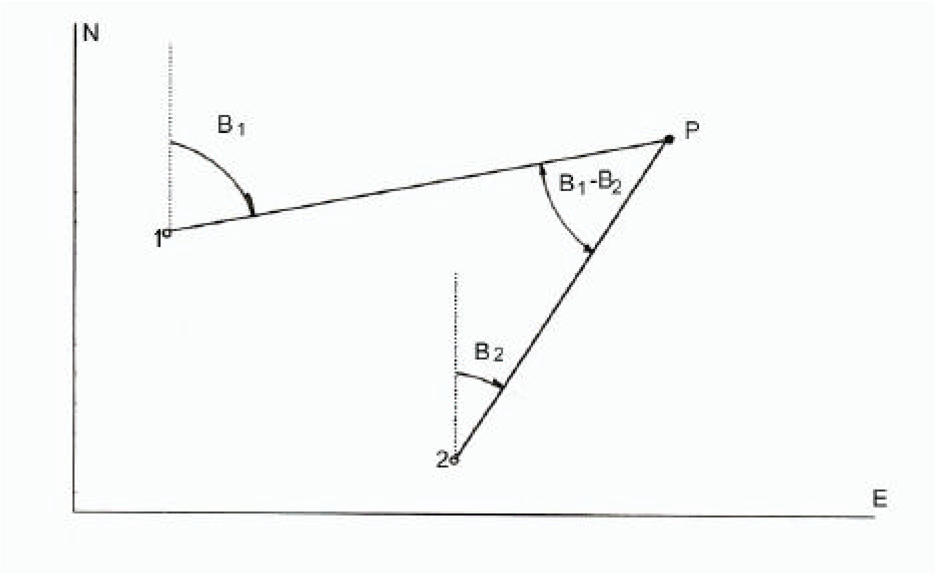


Fig. 6.15

Figure 6.15 shows an intersection of two directions, of which the grid bearings B_1 and B_2, are known. It may be the case that they have been obtained from line of sight observations 1 – 2 and 2 - 1.

There are several solutions and software to solve this problem. One of them is:

$$N = N_1 + \frac{[(N_1 - N_2)\operatorname{sen} B_2 - (E_1 - E_2)\cos B_2]}{\operatorname{sen}(B_1 - B_2)} \cos B_1$$

$$E = E_1 + \frac{[(N_1 - N_2)\operatorname{sen} B_2 - (E_1 - E_2)\cos B_2]}{\operatorname{sen}(B_1 - B_2)} \operatorname{sen} B_1$$

c) Intersection of Distances

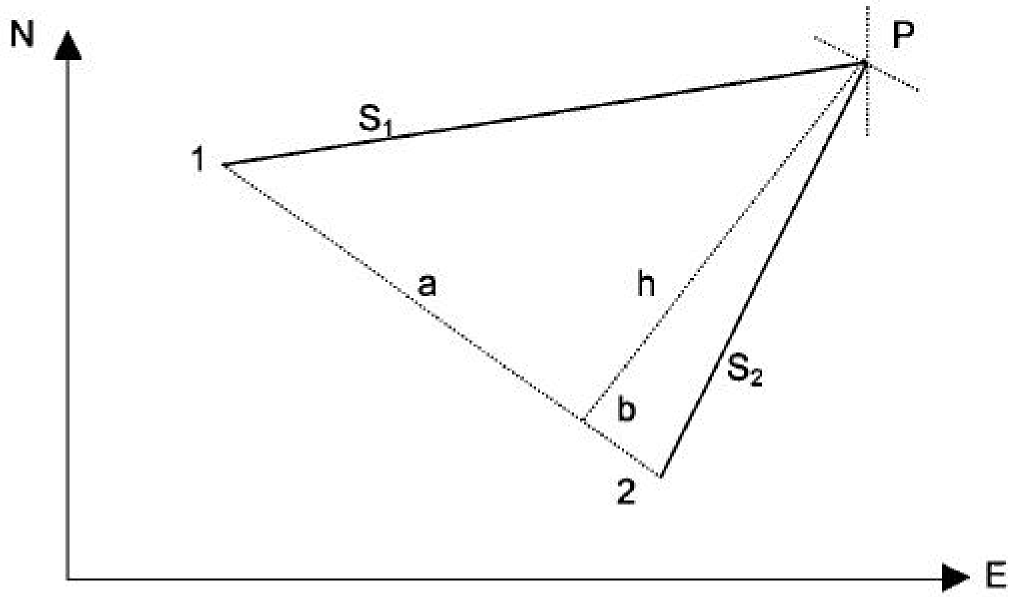


Fig. 6.16

This case is illustrated in Fig. 16, it has two mathematical solutions; therefore it is important to make it clear whether point P is on the left (this is the case in the figure), or on the right of line 1-2.

One solution is by applying the following calculations:

$$S_{12} = +\sqrt{(N_2 - N_1)^2 + (E_2 - E_1)^2}$$

$$\operatorname{sen} B_{12} = \frac{E_2 - E_1}{S_{12}}$$

$$\cos B_{12} = \frac{N_2 - N_1}{S_{12}}$$

$$a = \frac{1}{2}\left(S_{12} - \frac{S_2^2 - S_1^2}{S_{12}}\right)$$

$$b = \frac{1}{2}\left(S_{12} + \frac{S_2^2 - S_1^2}{S_{12}}\right)$$

$$h = +\sqrt{S_1^2 - a} = \sqrt{S_1^2 - b^2}$$

$$N = N_1 + a\cos B_{12} \mp h \operatorname{sen} B_{12}$$

$$E = E_1 + a \operatorname{sen} B_{12} \pm h \cos B_{12}$$

The lower sign is for the case when P is on the left of 1-2 and the upper sign is when it is on the right.

d) Resection

As stated in 2.2.4 above a resection occurs when directions or angles are measured:

from a point, the calculation of which is required, to three known control points. This situation, as well as nomenclature to be applied in the algorithm, is shown in Fig. 6.17.

Before proceeding, it should be noted that there are many graphical, numerical and mechanical solutions with which to obtain the station point position.

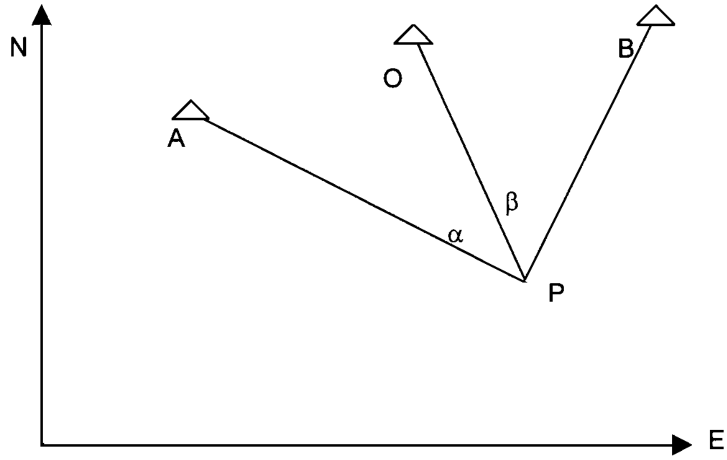


Fig. 6.17

With such numerical solutions it is essential that a method is available to detect cases close to indetermination as indicated in Fig.6.10.

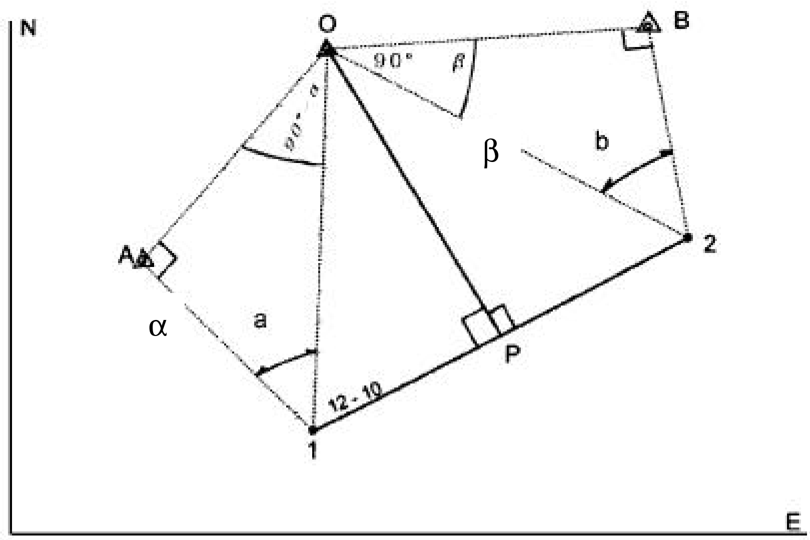


Fig. 6.18

The use of two auxiliary points 1 and 2 constitute the base for the algorithm, proposed below, is shown in Fig. 6.18.

The formulae to calculate the co-ordinates for these points can be obtained simply from:
When points 1 and 2 are too close to each other (for example less than a tenth of distances

$$N_1 = N_A - (E_0 - E_A)\cot g a$$
$$E_1 = E_A + (N_0 - N_A)\cot g a$$
$$N_2 = N_B - (E_B - E_0)\cot g b$$
$$E_2 = E_B + (N_B - N_0)\cot g b$$

AO or OB) it can be assumed that the network is close to indetermination.

Calculation for N and E co-ordinates of point P can be achieved by:

$$N = N_1 + \overline{10}.\cos .(\overrightarrow{12} - \overrightarrow{10}).\cos .\overrightarrow{12}$$
$$E = E_1 + \overline{10}.\cos .(\overrightarrow{12} - \overrightarrow{10}).\text{sen} .\overrightarrow{12}$$

Where

$\overline{10}$ = The distance from 1 to 0 = $\sqrt{(N_0 - N_1)^2 + (E_0 - E_1)^2}$

$\overrightarrow{12}$ = Grid bearing from 1 to 2 [tg $\overrightarrow{12}$ = $(E_2 - E_1) / (N_2 - N_1)$]

$\overrightarrow{10}$ = Grid bearing from 1 to 0 [tg 10 = $(E_0 - E_1) / (N_0 - N_1)$]

When calculating orientations ($\overrightarrow{12}$, $\overrightarrow{10}$) it is necessary for quadrants to be discriminated with signs ΔE and ΔN. For this purpose you may make use of the usual subroutines to pass from plane coordinates to polar coordinates.

Another way of solving the last part of the calculation is to obtain the coordinates of P such as the perpendicular base from 0 to segment 12 making use of the subroutines available in Computer-Assisted Design (CAD) programmes

Some checking calculations may be established though the most complete method is to calculate grid bearings from P to the known points (A, O, B) and then check

$$\alpha = \overrightarrow{PO} - \overrightarrow{PA}$$
$$\beta = \overrightarrow{PB} - \overrightarrow{PO}$$

2.2.6 Levelling and its Errors

Trigonometric levelling and the possible errors have been discussed in 2.2.3 (traverse). It should be noted that in the case of intersections a similar operation can be undertaken with the algorithms and the resultant calculations. It is also be possible to apply them for surveys using polar co-ordinates or EODM tachymetry, when it is particularly useful to have total stations which store (horizontal and vertical) ranges and directions to surveyed points. On processing such information and when ranges over 100 m are used, it is important to verify that the software application includes corrections for refraction and earth curvatures.

Direct levelling (with spirit or self aligning levels) is generally more precise. In the case of geodetic levelling, which requires the use of levels of higher sensitivity and stadia graduated on INVAR plates (an alloy of nickel and steel with a coefficient of expansion below 1 x 10^{-6} 1/°C) and other precautionary measures, error propagation may be below:

$$1\ mm\sqrt{K}$$

where **K** is the track distance expressed in Km.

If common topographic levels, wooden or plastic centimetre-graduated stadia with joints or couplings and instrument-stadia distances below 100 m, with equidistant stations (within 3 m), are used, you may attain accuracies of the order of:

$$7mm\sqrt{K}$$

for which it is assumed that every section between bench marks is measured in both directions with a tolerance of the order of:

$$\pm 3mm\sqrt{K} \quad \text{(geodetic) and } \pm 10\,mm\sqrt{k} \quad \text{(topographic)}$$

for both cases, without bias to any intermediate or even less accurate solution which may be adopted.

In hydrographic surveying, the highest accuracies are required to tie-in permanent tide stations followed by temporary tide stations, which are generally established during the survey, the calculation of levels for harbour facilities and standards for engineering works associated with water behaviour.

In an extensive hydrographic survey (more than 50 km) with no available local levelling datums, it is expedient to provide, as a minimum, a direct levelling line to which the tide stations can be related and leave a reference mark from which future trigonometric levelling can be conducted. When applying these provisions, the specifications in 2.1.6 should be considered and an analysis of the stability of the relationship of the tide station and mean sea levels is necessary.

When using satellite methods (GNSS) for altimetric purposes, the provisions in 2.1.6 and Chapter 2 need to be noted particularly the requirement for modelling corrections to pass from heights above the spheroid to values associated with sea level used in hydrographic surveys. Regardless of preset correction diagrams that may exist, it is necessary to adjust them to altimetric points as described in 2.2.1, including the provisions in Fig. 1, in connection with the relationship between altimetric marks. In other words, use of GNSS techniques for altimetric purposes should be limited to point interpolation rather than extrapolation. This concept is likely to evolve in the future but in 2004 there remains no likelihood of confidence in general correction models and still less in places where there is no guarantee that local observations have been carried out to create them.

2.3 Coastal and Harbour Ground Surveys

2.3.1 Application of Direct Topographic Methods

In general, coastal surveys which are a part of hydrographic surveys are mostly carried out by photogrammetry or other remote-sensing processes. In such cases the surveyor's main task, when processing information, consists of obtaining a proper interpretation of coastal features, that coastline delimitation poses no difficulty and that data on the ground control points are adequately provided. He must also ensure that aids to navigation signals and stations have their horizontal and vertical positions properly determined.

However, there are cases in which all this information must be obtained by direct topographic survey methods, i.e. by field observations and measurements. These cases are generally related to the need to represent certain areas on large scales (1:5000, 1:2000, 1:1000 ...). This often occurs in areas where there is a port infrastructure or where a harbour project, landing, water intake or other engineering works are being carried out or extended to occupy the inter-tidal zone and extending into the near shore strip.

The limited extent of such places as well as the required high degree of detail may require that such surveys are carried out by topographic measurements in the field.

2.3.2 Density of points to be surveyed

Firstly the required degree of detail must be established. The usual method is to set a scale according to the essential representational needs of the final product, in order to properly obtain the shape it may be necessary to survey a point every square centimetre. Nevertheless such a distribution shall not necessarily be strictly homogeneous. Priority should be given to sites where there is a significant change in slope or where there are outstanding characteristics: hillocks, holes, saddles, ridges, talwegs, etc.

Generally the survey of points on near-perpendicular lines to the shoreline provides much more useful information for good representation of shape than any other type of distribution.

For details which must be surveyed to allow representation of natural or artificial features, more or less independently of relief, the quantity of points should be adequate enough to be able to plot them at the intended scale, straight sections probably only require the surveying of turning points and if orthogonal, simplification may be greater still.

2.3.3 Applicable Methods

Satellite techniques (GNSS) are ideal for surveying horizontal positions. If they are intended to be extended to planimetric and altimetric positions, the provisions detailed in 2.2.1 should be observed. Generally the process is more advantageous when the density of the points to be surveyed is low (i.e. more than 50 or 100 m between them for scales of 1:5000, 1:10000, etc.). Ground-permitting, this process may be achieved by placing the rover station in a vehicle. The opportunity of processing information in a fully automated manner would rapidly improve the achievement of results.

EODM tachymetry is a particularly appropriate method for cases where, from a few stations, points with distances of 1000 m and above may be surveyed. Use of total stations with its capacity to store distances, directions (horizontal and vertical), attributes of surveyed points, etc., makes it possible to quickly process the information and generate the appropriate survey sheets which can be completed with additional data if required.

Stadia method tachymetry is ideally suitable for sites where the survey of a large number of points very close to each other (50, 20, 10 m) is required at relatively short distances (below 200 m) from every station. Reading of graticule lines is made on generally centimetre-graduated stadia.

Ground distance is obtained as **K.m**, where **K** is the stadia constant, usually 100, and "**m**" is the difference of stadia line readings. If an elevation angle a has been measured, the horizontal equipment ground distance to the stadia is:

$$K \bullet m \cos^2 a$$

and the relative elevation to the surveyed point equals:

$$\Delta H_{12} = i_1 - S_2 + K \cdot m \cdot \frac{1}{2} \operatorname{sen} 2a_1$$

where $\mathbf{i_1}$, $\mathbf{S_2}$, **y** and a_1 have the meanings given in 2.2.3 for trigonometric levelling.

In the case of lines with too great an incline of sight ($\alpha > 10°$), this method is not recommended for height transfer since the distance error (of the order 0.2%) and the sight's probable lack of verticality introduce considerable altimetric errors (this is less frequent in EODM tachymetry).

With special stadia having divisions of 5 cm or 10 cm, the survey ranges may extend to 500 m and above, although it is not advisable in the case of lines with too great an incline of sight for the above reasons.

All these procedures allow calculation, from the above formulae, of the 3 horizontal and vertical co-ordinates of the mark. In some cases these co-ordinates and orientations may be obtained by resection complemented with inverse trigonometric levelling, based on the adequacy of the formulae given in 2.2.3

In flat areas direct levelling is a simple and precise method of surveying. If necessary stadimetric distances (K.m) may also be used as well as horizontal directions which may be measured by some other instruments.

In relatively flat locations, for constructions with some orthogonal shapes measurement of perpendicular distances may be made using tapes and an optical square. Simple though it may be, it proves to be a useful method to be applied in some places such as docks, piers, moorings and other port buildings. This type of survey is usually be complemented with direct levelling in order to determine platform or floor elevations.

2.3.4 Relief Representation

Although the trend is to generate databases which provide a variety of applications for information through a Geographic Information System (GIS), implying the availability of a Digital Terrain Model (DTM), planimetric and altimetric measurements are frequently requested to be represented by contour lines. For this purpose, selection of a contour interval should be made at not less than four times the estimated error of elevations.

An alternative method of selecting the contour interval is by scales. In the case of very broken ground, the scale denominator thousandth part may be measured in metres (example: 5 m for 1:5000), but in the case of flat and featureless terrain, the values may decrease to a tenth part (0.5 m in the former example).

Both criteria should be harmonised and basically, the survey purpose as well as relief fluctuation in the area should be taken into account.

Several software packages are available for drawing contour lines from discreetly surveyed points. Some of them have proved to be very capable but it is expedient to adjust their drawing algorithms by incorporating some interpretation rules for the relief before the final version.

Fig. 6.19 shows how drainage lines tend to stress the contour line curvature while ridges tending to separate the water movement on the surface are gentler. These trends generally undergo changes and contours collectively representing the relief must keep some agreement.

The concepts mentioned above are valid for application in ground shape; however, not all of them are valid for seafloor application.

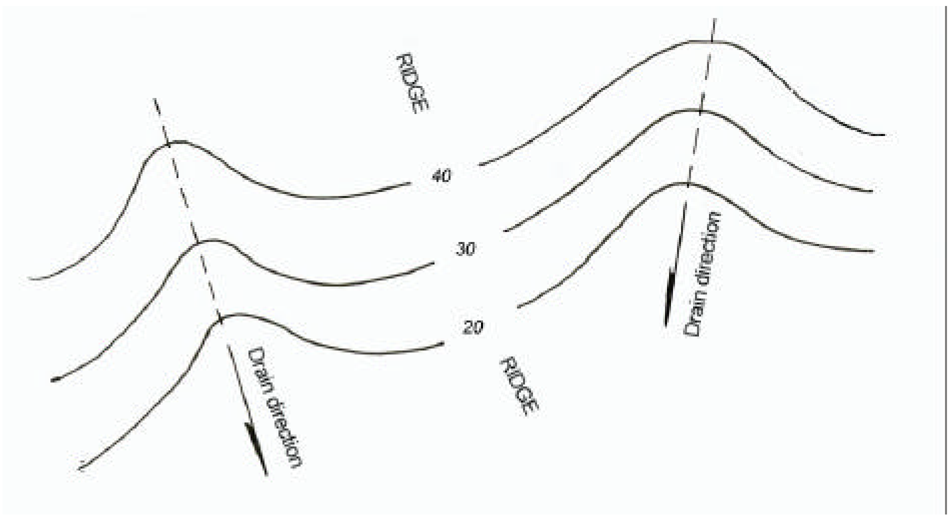


Fig. 6.19

With some geomorphological knowledge, the criteria may be generally improved for a better relief interpretation.

3. REMOTE SENSING

Some techniques for obtaining information through remote sensors, which capture the ground radiation, will be described in this section. This information is stored and then processed, thus generating products that provide topographic data.

If ground radiations originate from the reflected solar energy, sensors are called passive; if they are generated from reflected emission of devices associated with the sensor, sensors are called active.

The range of frequencies and wavelengths of electromagnetic waves for remote sensing is shown in Table 6.4:

Table 6.4

NAME	FREQUENCY (Hz)		WAVELENGHT (m)
Microwaves	3×10^9 to 3×10^{11}		10^{-1} to 10^{-3}
Thermal Infrared	3×10^{13} to 3×10^{12}		10^{-5} to 10^{-4}
Medium and near infrared	$4{,}3 \times 10^{14}$ to 3×10^{13}		0.7×10^{-6} to 10^{-5}
Visible Light	...		...
	4.6×10^{14}	Red	0.65×10^{-6}
	....		....
	5.4×10^{14}	Green	0.55×10^{-6}
	....		....
	6.6×10^{14}	Blue	0.45×10^{-6}
	...		...
Ultra Violet	3×10^{15} to 3×10^{16}		10^{-7} a 10^{-8}

Radio waves have the lowest frequency whilst x, gamma and cosmic rays have the highest. They also have other applications.

Among the passive sensors uses, which use visible light radiations and their close proximities, is the Photogrammetry. Since this technique began to be applied with light sensitive films, it has been used in hydrographic surveying since the beginning of the 20th century, it remains one of the most efficient ways of obtaining good information of the relief, especially for large scales (1:20000, 1:10000, 1:5000, ...)

From the 1970s and more intensively from the 1990s, remote sensing applications were extended from active and passive airborne and satellites sensors to other imagery processes. Satellite methods have not generally the same capacity as photogrammetry in the ground shape interpretation. However, they have additional capabilities for detecting the surface properties of soil and water covered areas. They also offer impressive updating capabilities, frequently at relatively low costs.

In photogrammetry, as well in other imagery processes, it is necessary to undertake ground control operations in order to achieve correct scale results and obtain good referenced positions. Ground control consists of locating, in the field, identifiable points based on the information provided by the sensors.

3.1 **Photogrammetry** (See 3.4 at Chapter 2)

Strictly speaking, photogrammetry is the technique that allows objects to be described in three dimensions from overlapped photographic images, taken from adjacent places. For hydrographic surveys, aerial photography with vertical axis through a metric camera is more useful.

3D description is achieved by stereoscopic viewing of virtual models and the measurements are taken with the use of specific instruments to achieve topographic representation. Of course, this technique requires ground control points obtained via field topographic methods or densification through an aerial photogrammetric process, also called aerotriangulation.

There are other products which are not 3D but may be considered to be part of photogrammetry. Among these are photo-plots, which can be obtained by simple assembling photographic images adjusted by rectification (scale and inclination).

3.1.1 Principles and applications of aerial photographs

The objective of aerial photographs is to provide information to obtain true ground representation including relief. This can be done through photogrammetric restitution or stereo compilation. Nevertheless, as previously stated in the concept of photogrammetry, there are other 2 D products which can be obtained from aerial photographs.

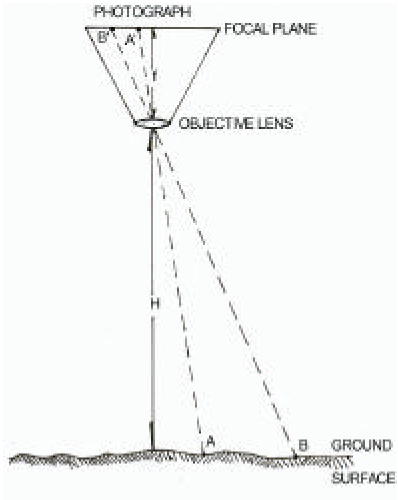


Fig. 6.20

To explain this it is necessary to take into account the basic expression for the aerial photograph scale:

$$S = \frac{A'B'}{AB} = \frac{f}{H}$$

Where the ratio among the focal length **f** and the flight altitude **H** is directly related to the image scale (see Fig. 6.20 for a camera with vertical axis).

Although objective lenses might be considered as a centred optical system with two nodes, the scheme is simplified with only one optical centre similar to a thin lens. That simplification is enough for the approximate calculation of the flight scale. See also that, as H>>f, it is assumed that the image is formed in the focal plane.

A change in the flight height causes a change in scale; a lack of verticality in the camera axis produces a change in the scale in different points of the camera. For example, a rectangle ABCD on the ground can be represented as a trapezium A´B´C´D´ on the photographic image where the scale of segments A´B´ is shorter than in C´D´.

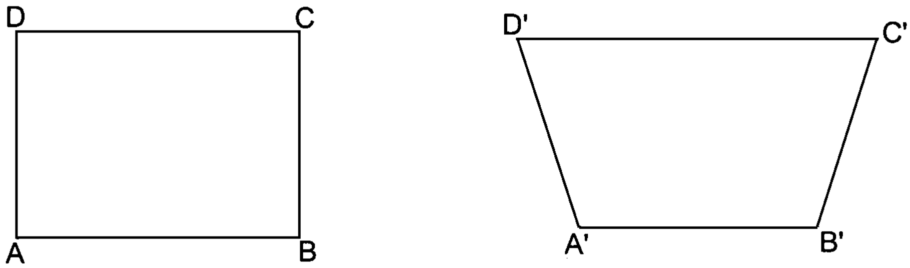


Fig. 6.21

Moreover, if there are features in the relief with significant vertical characteristics, the scale introduces other changes on each photograph. This can only be resolved by a 3D treatment like restitution. An adjustment of the change in the flight height and the axis orientation is possible by rectification through a 2D photographic process. Note that this adjustment is only possible on flat surfaces.

Special devices can fulfil the process of rectification mentioned above by using rectifiers provided with a camera, which project the image onto a board. The set allows a series of combined movements which enables the projected image to change and to include slopes according to focusing conditions. The current way to rectify is by making the projection of four well-distributed points coincide with their well-marked locations, as in Fig. 6.21. There are also equivalent 2D numerical procedures for solving this problem.

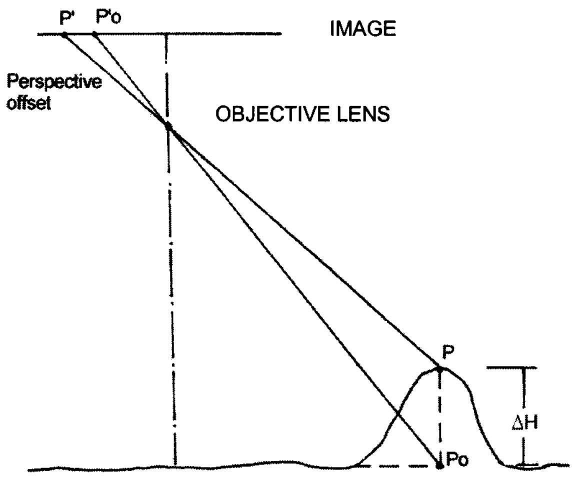


Fig. 6.22

The limits for these processes rest with the image of a point with a certain difference in elevation relative to the surrounding area, which experiences a perspective offset on the image (see fig 6.22). Note that, apart from the difference in elevation **? *H***, the distance of the elevated point from the camera vertical axis increases the offset, in other words, points near the vertical axis of the camera do not display significant offset effects.

An alternative way to generate photographic images free from this effect is to combine the photographic process with the 3D treatment, the product is called orthophotograph.

The best way to present ground photographic images is through an orthophotomosaic, which is an assembly of images forming a uniform-scale mosaic. The next in quality is the rectified mosaic adjusted as mentioned; the coarsest method is by assembling raw photographic images and accepting an approximate scale as a function of the average flight height.

The type of photo-plot should always be specified so that caution can be exercised with respect to the validity of the product metric.

3.1.2 Aerial photograph acquisition elements

Extraterrestrial solar radiation has a maximum range of wavelengths from 0.4 micrometers to 0.8 micrometers (1 micrometer= 10^{-6}m), which is between infrared and blue (see Table 6.4). The radiation changes when passing through the atmosphere, the soil reflection also impacts on the light spectrum received by the camera. Thus, the film and the emulsion must be carefully selected.

Among the black and white films (scale or tints of greys) orthochromatic emulsions are especially useful between 0.4 and 0.55 micrometers, panchromatic ones between 0.3 and 0.65 micrometers, with an additional increase in wavelengths of 0.6 and 0.9 micrometers. The most useful in aerial photogrammetry is the panchromatic emulsion. There are several types of three layer colour films, but they are more useful for photo interpretation, described later (3.18), than for photogrammetry.

There are a series of specifications regarding density, speed, resolving power, granularity and base stability which must be determined in order to achieve the best result in the prevailing conditions to meet the needs for the final product. The objective and filters to be used should be addressed in the analysis.

The objective lens is composed of an optical system where a good distortion correction is particularly required.

The image format commonly used is 23 cm x 23 cm with the focal distances (f) (see 3.1.1) detailed in Table 6.5:

Table 6.5

Camera Type	**f (mm)**
Super Wide Angle	85
Wide Angle	153
Intermediate angle	210
Normal Angle	305
Narrow Angle	610

Cameras with a shorter focal distance (f) require a better distortion rectification whilst the images are also more affected by atmospheric refraction. The Wide Angle is the most commonly used type of camera.

For photogrammetric purposes, an aerial camera must have a good determination of f, a rigorous correction of the distortion or other optical and mechanical conditions which can be checked by calibration. The camera is termed a metric camera if these conditions have been met. These cameras have an accurate system for ensuring verticality of the axis and assure the flatness of the film. They also have a proper dwell time control and allow an overlapping control between consecutive photographs (end lap), etc.

Although digital cameras generally enable high quality images for photographs, their development for use in photogrammetry is advancing rapidly but presently (2004) only non metric cameras are available.

An important component for aerial photography is the aerial survey platform. Criteria include suitable space for the camera and its attachments, have sufficient endurance, are able to operate at the required flight heights and required speeds, satisfy permissible vibration limits, etc.

Among other requirements it must have GNSS positioning, often with a differential capacity, a requirement for positioning synchronisation with the camera and multiple antennas for platform inclination checking.

3.1.3 Flight planning

Initially it is necessary to define the flight scale, meaning the scale of the camera, which, as it was stated in 3.1.2, has a format of 23cmx23cm. If the type of camera is defined, the scale also determines the flight height H=f/S (see 3.1.1 Fig. 6.20).

Although the scale may be enlarged five times to obtain good photogrammetric products to meet the hydrographic requirements, the analysis of the required altimetry accuracy should be conducted. It should be noted that the deviation of the elevation obtained by restitution reaches 200ppmxH (200 parts per million of the flight height=H/5000). Sometimes, this can make it unachievable and the altimetry requirements must be met by other means.

Having defined the flight scale, flight strips must be studied. In the simplest situations, the coastal band can be covered by a set of rectilinear strips (see Fig. 6.23)

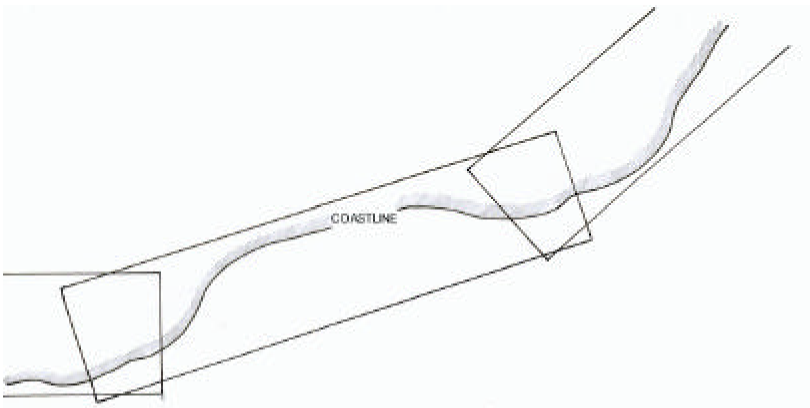


Fig. 6.23

When coastal features are extensive, wider inshore surveys are required. In this case, blocks of several strips must be planned (see Fig. 6.24).

Additionally end and side overlaps must be planned; generally, the end overlap is of 60% and the side overlap of 20%. When orthophotographs are required (see 3.1.1) or when ground features are so uneven that there is a possibility in gaps leaving some part of the information without stereoscopy, it may be necessary to enlarge the overlapping.

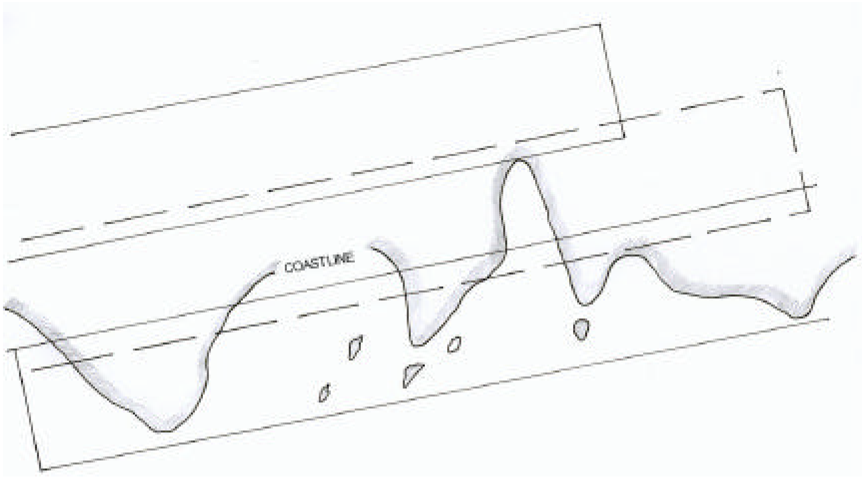


Fig. 6.24

The sun's altitude and angle during the flight should be considered, particularly in higher latitudes ($\varphi>50^{\circ}$) in wintertime.

To ensure that shadows do not interfere or impact on image quality, the sun altitude angle should be greater than 30º. The more uneven and cluttered the ground, the greater the elevation angle should be. The flight time may be limited by the time of the year and the latitude.

An additional limit for hydrographic surveys is that the flights should take place near low water to allow detection of features and dangers close to the foreshore in the inter-tidal zone.

The sky must be free of clouds below the flight height whilst many other meteorological conditions must be satisfied during the operation. All these limitations combine to make flight times longer and planning more complex.

The ground control and its densification by aerotriangulation must be considered when planning the flight. This provision is necessary to allow the opportunity of performing field tasks during the survey group's presents in the area.

The end overlap produces coverage as detailed in Fig. 6.25. If the overlap is of 60% or more there is a zone of 20% or more of triple overlap.

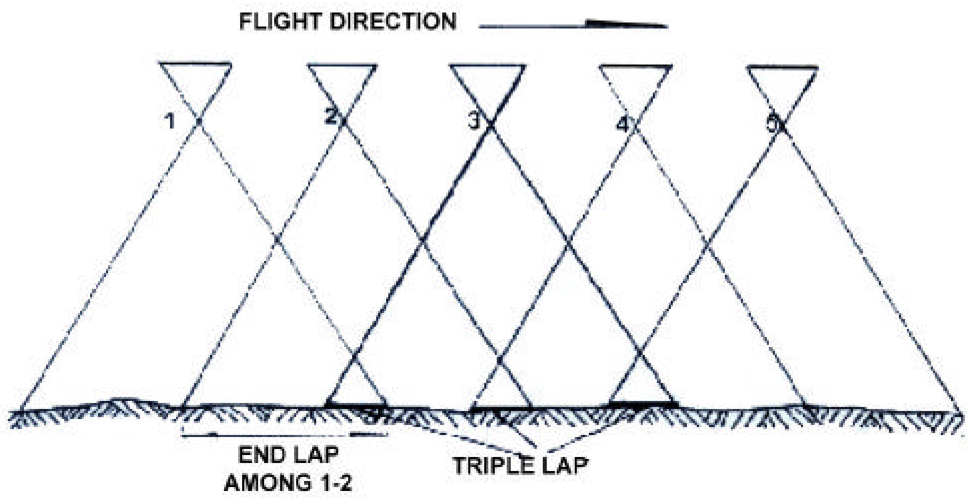


Fig. 6.25

In this zone, like in the side overlap (see fig 6.24), aerotriangulation can be carried out.

3.1.4 Restitution

The photogrammetric technique restitution is the basic process in the 3D treatment of the topographical information, generally aerial images are used. Restitution is carried on the optical, mechanical, analytical or digital adjoining photogram projection in the overlap zone, which allows stereoscopy observation.

In any version it is necessary to determine the relative and the absolute orientation of the model which replicates that part of the ground being surveyed.

A pair of photographs is oriented by intersecting five pairs of homologous rays corresponding to five ground points. This process is achieved by removing their parallaxes through motion projectors or by an equivalent digital process.

Prior knowledge of the co-ordinates of the selected points is not required; however it is expedient to choose them from the end overlap zone (see Fig. 6.26).

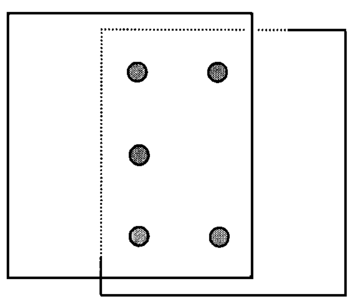


Fig. 6.26

Having done this, a 3D model is created, even though neither its position in the reference frame nor its scale has been defined. In other words, only a relative position of the photographs coincident with the cameras during the flight in an unknown scale and reference frame. It is possible to observe stereoscopically the entire model whilst holding the observed images location.

To assign a scale to this model and to express it in a reference frame compatible with the survey, at least 3D positions of two points (for example 1 and 2 from Fig.6.27) and the height of a third one must be known. However it is better to know the three co-ordinates of 1, 2, 3 and 4, which allows some verification.

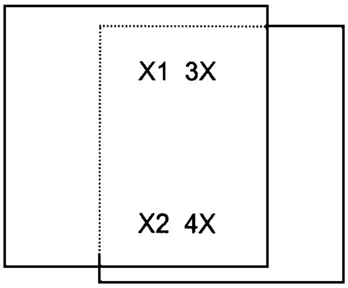


Fig. 6.27

Of course, this adjustment can be undertaken in an analogical way, by optical or mechanical means or numerically by analytical or digital stereo plotters.

With the models absolutely oriented, it is possible to obtain a topographic representation of the relief and description of features or infrastructures. The contour lines can be traced by analogical or digital means. In the latter case, it is possible to make a Digital Terrain Model (DTM) with a convenient density of stored points.

To obtain digital copies of the information, at present the simplest way is to scan aerial photographs through high-resolution scanners; however in the future, it will be available with information taken from digital cameras as mentioned at the end of 3.1.1.

3.1.5 Aerotriangulation

As has been described in 3.1.4, for the absolute orientation of a stereoscopic model in restitution, it is necessary to know the three sets co-ordinates of four points distributed as in Fig.6.27, although in principle the three sets of co-ordinates of 2 points and the vertical of a third point may be sufficient.

To achieve such control, whilst minimizing field work, an internal process has been developed by photogrammetry: aerotriangulation.

The first process in this technique consists in providing the ground control for the first model, determining its absolute orientation and then, passing to the second model adding a third image. Having completed all the movements in the third image projection, without modifying the previous during the process for the second model's relative orientation, it will be clear that absolute orientation has been transferred.

It is possible to repeat the process detailed above, however deformations could appear. Apart from the deviation, the effects due to terrestrial curvature and refraction of light rays must be considered. For this reason it is necessary to adjust the strip by adding ground control points.

Fig. 6.28

A strip with four start control points, four close control points and two intermediate pairs is shown in Fig. 6.28 (also see Fig. 6.23). The intermediate pairs should be present in six models in order to successfully solve for deformations and propagation of the deviation.

Both control points and tie points, to hold the restitution, must be present in the zone of triple superposition and when necessary with the side overlap.

Although the described distribution corresponds to analogical aerotriangulation processes, hydrographic experience shows that frequently a control is still valid in coastal surveying (see Fig. 6.23). This is also valid when strip adjustment is carried out through independent models using an analytic process. In this case, the normal method, after determining every relative orientation, is to note each model's co-ordinates and then to adjust them in numerical terms.

When there are several strips with side overlaps (see Fig. 6.24), block adjustment with independent models can be completed with certain advantages from the merged rigid set.

Points subjected to aerotriangulation are:

a. Ground control points;
b. Tie points;
c. Additional points for restitution control or items of detail requiring specific calculation.

Then, holding the co-ordinates of the ground control points fixed whilst taking into account their relation with tie points, a block adjustment can be undertaken. As a result, co-ordinates of the tie points and any additional points can be obtained and expressed in the ground control points' reference frame.

There are seven parameters for each transformation model in a normal process: one scale, three translation parameters and three orientation ones. Several software versions are available; the basic ones deal with the planimetric and altimetry processes separately. The more elaborated ones are based on 3D treatments with an important statistical analysis which tends to clean up the influence of out of tolerance deviations. With these kinds of blocks, the required amount of ground control points can be minimised. There is an integral utilisation for them and a strong linking between models to emphasise the set rigidity. With 5 + 0.2M ground control points, successful results can be achieved, M is the number of independent models that constitutes the block.

A block of independent models under adjustment is shown in Fig. 6.29. It must be remembered that apart from the number of ground control points, their distribution is important to assure an accurate and rigid network for the restitution.

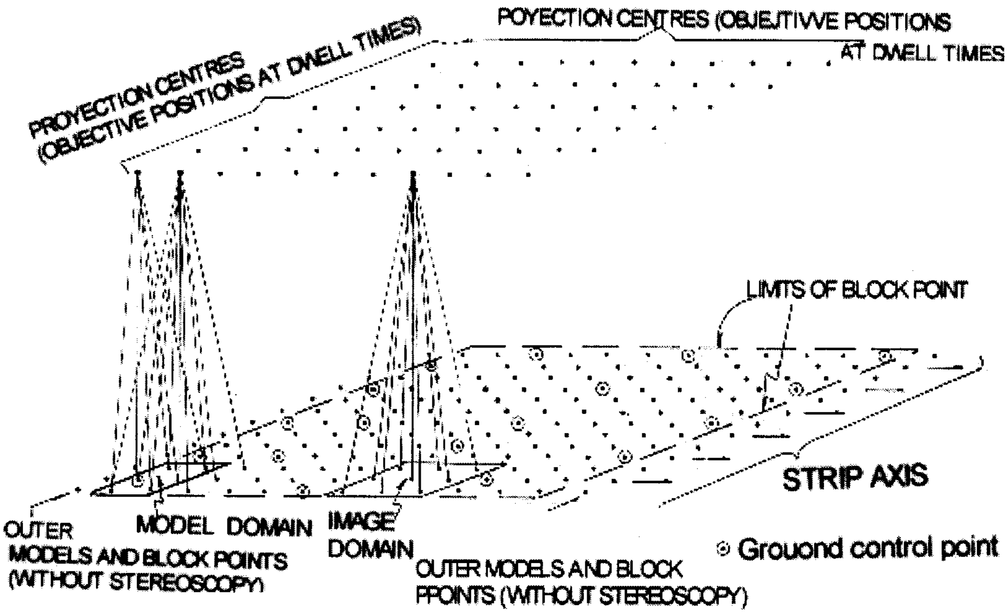


Fig. 6.29

For simplicity, only a few rays from the perspective centres to the points under aerotriangulation are shown. These perspective centres are associated with the objective position at the moments of exposure. The ground control points, of which some are coincident with tie points, but not in every case, are also indicated.

Fig. 6.29 is also useful to demonstrate the linking that can be achieved through intercepting homologous rays.

Even though they have been chosen by stereoscopy observation, the measuring of plane co-ordinates inside each image, without stereoscopic processing, sets up these rays. Through this method, at least nine points from the photogram are usually measured with a distribution as shown in Fig. 6.30. The stability of a block adjusted through this bundle block adjustment technique is higher than that achieved through strips or independent models. Occasionally, a first adjustment is conducted through independent models and then, with these provisional co-ordinates, the last adjustment is undertaken through bundle of homologous rays.

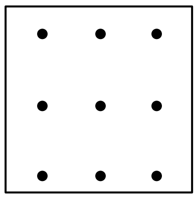


Fig. 6.30

In the block adjustments, by independent models or by bundle, apart from the three co-ordinates for each point being processed, co-ordinates for perspective centres are also created.

There are cameras which can be synchronised with GNSS systems. They have the capacity to receive differential corrections; the position of the centres can be introduced in the block adjustment. Thus, the number of ground control points can be reduced. Systems with three or more antennas are being developed in order to extend the capacity to the calculation for orientation.

There are other means to increase aerotriangulation capacity and whilst minimising ground works; the obtaining of images at smaller scales is among them. It has certain validity for horizontal co-ordinates but it is not capable, as yet, for solving altimetry requirements accurately. Some transversal strips at the same or smaller scale are also used. At the present time all these procedures for small photogrammetric scales (1:20000, 1:50000, 1:100000, ...) are avoided and replaced with the described technique of GNSS positioning through the projection centres with minimum ground control.

3.1.6 Ground control

One of the main tasks of topographic field surveying is the creation and marking of ground control points.

Although in 3.1.5 some guidance for selecting the points to allow aerotriangulation are given, it is necessary to consider the specified requirements of those responsible for restitution or the aerotriangulation processes. The objective of aerotriangulation is the control of the restitution described in 3.1.4

Selecting of control before the flight is theoretically possible, deciding on the positioning points which are to be photographed. However, the control is frequently selected after the flight by positioning identified points on the photographed images. This is a way of avoiding problems caused by the short life of artificial marks.

Apart from obtaining the values of the co-ordinates of the control points, their plots must be completed. An initial impression for that information can be acquired from photogram copies or on a photo-plot. Sometimes the feature is pricked through the image with details written down on the back. However, this is not always sufficient and it is necessary to add a description with graphs to clarify the chosen detail and to fix its position and the reference level for the vertical co-ordinate. This is important because sometimes the appropriate detail to fix the horizontal position has no well-defined level. For example, the corner of a building is a good point with which to refer a horizontal position but the ground level or both must be identifiable to give good vertical control.

In every case the description obtained in situ must be compatible with that which can be obtained by stereoscopic information. To do this it is useful to have a stereoscopy and accurate image copies to analyse this information or to observe it through a stereo plotter to provide the description to be used in aerotriangulation.

The accuracy in the position of ground control points must be carefully studied, taking into consideration the desired aerotriangulation results to control the restitution. A maximum deviation of 100ppm (100parts per million) of the flight height (that is H/10000) in the three co-ordinates can be accepted. In cases where difficulties arise, acceptable alternatives must be available and analysed.

Apart from the issues on the instructions for distribution of ground control points, accounting for the aerotriangulation adjustment, it is expedient to clarify that the provision of **x**, **y** and **z** co-ordinate points around the outside edge of the block are more useful; some internal points may be limited to the vertical **z** co-ordinate only.

3.1.7 Stereo plotter generalities. Digital processing.

A simplified scheme of a stereo plotter is shown in Fig. 31. It has two photogram supports (on film or in digital format) on which co-ordinates **x´y´** and **x´´y´´** can be read. It also has an observation device (represented by two eyepieces) which has two (optical, optical/mechanical, electronic) paths inside it to allow each eye to partially see each image, making the stereoscopic model available for measurements. The paths have floating marks with which to form a point which can be seen in 3D near the model. These marks can be moved on the model in the direction of the flight through the X control, transversal to the flight with the Y control and vertically via the Z control.

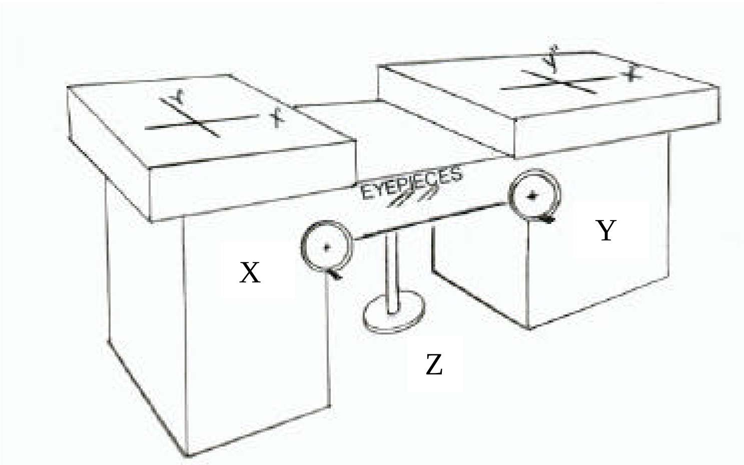


Fig. 6.31

As is indicated in Fig. 6.31, controls in **x** and **y** are operated through cranks and **z** through a pedal.

Through codifiers related to the movements in **x**, **y** and **z**, these co-ordinates can be registered. An independent model aerotriangulation process can then be applied to them. To conduct the process through ray bundles, it is necessary to have an instrument with codifiers to register photogram co-ordinates (**x´**, **y´**, **x´´** and **y´´**)

Apart from the required accuracy to distinguish 100ppm of the flight height (H/10000), an instrument suitable for aerotriangulation must have all the essentials to register and to codify.

Naturally, all the stated registration, codification elements and others related to orientation and inner equipment performance, should be connected to a computing system, particularly in analytical and digital versions (see 3.1.4).

In the new digital versions (soft or video plotters) a computer monitor is used to display the data required to carry out the observations detailed above (see Fig. 6.51). Both photograms are projected alternatively onto the monitor, the operator views one in each eye through special observation equipment (anaglyph, polarised lens or other electro-optical means), which create a stereoscopic image and therefore the ability to make the required measurements. Other peripherals are connected as indicated in Fig. 6.32. In a digital stereo plotter, the image is provided by a stereo camera (CCD = charge coupled device).

Fig. 6.32 shows a data flow diagram in a digital stereo photogrammetric system.

In electromechanical restitution devices the plotter gives the final output version of the work, additionally the plot was produced analogically without the aid of a computer process. In digital versions data output consists of files containing a precise format for future graphic manipulation (soft copy), such as a Geographic Information System (GIS). In these digital cases, the use of a plotter is a supplement providing a global overview of the aerophotogrammetry process.

The use of soft copy files is extremely convenient for hydrographic survey processing. The information from the photogrammetric process can be superposed, compared and made compatible with other data which is generated from the topographic field work, previous work or near shore bathymetric data.

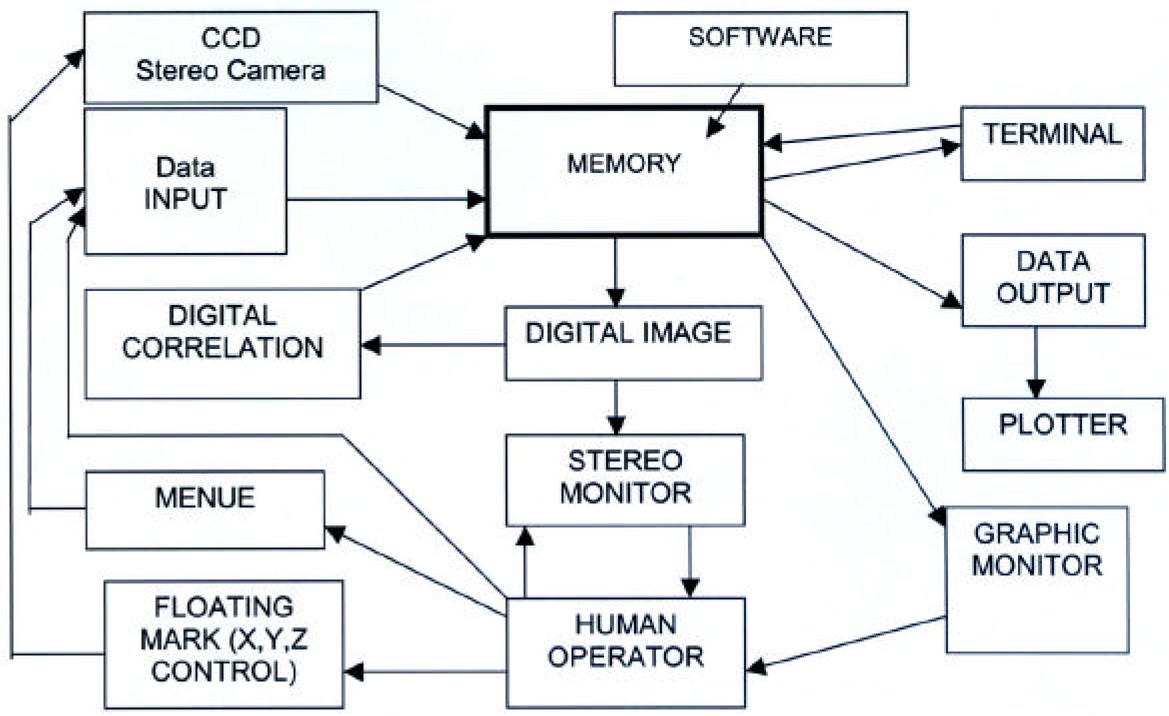


Fig. 6.32

3.1.8 Photo interpretation

Photo interpretation involves the examination of photographic images, sometimes supported by stereoscopic observation, which allows the identification of objects and features, as well as certain soil properties, vegetation, etc., in order to obtain a qualitative description of its character, use or behaviour.

In many cases, the relation between flight height and smooth topographic features is not sufficient for the relief to be viewed with sufficient detail to define the drain lines, as indicated in Fig.6.19. Nevertheless, it can be achieved through a minutely precise analysis of aerial photographs, thus, it is possible to detect the existence of temporary courses, separated by watersheds, whose features are clearer than in the relief interpretation. This is a typical example where photo interpretation can achieve more accurate descriptions than photogrammetry with small scale imagery, although this procedure should not be extrapolated.

Generally, image interpretation, photo interpretation being a particular case, can be undertaken more successfully by experts in their particular field. For example, a coastal engineer can make better conclusions regarding the behaviour of a beach than a surveyor, because he can analyse wave refractions and certain erosion processes.

In certain cases, very detailed contour lines can be traced assisted by images obtained from different periods, not only above the high waterline and the inter-tidal zone, but for the existence of permanent vegetation or the lines left by the flood tide currents before the backrush or the water image changes tonality with the depth.

An accurate combination between the calibration and the observation can be attained by comparison of some topographic or bathymetric measurements. This can produce excellent results. Nevertheless it is necessary to prove a strict correlation between the detected evidences, for example tone changes, and the measurements. If this is not verified, the basis for the interpretation must be revised. Sometimes, the behaviour of thematic phenomena is incorrectly interpreted as the presence of shoaler zones.

A photo interpreter's experience and the checking of doubtful details in the field allow photo interpretation to be a very useful procedure as a complement to topographic surveys.

3.1 Not Photogrammetric Remote Sensing Imagery

In this section only non-photogrammetric systems and methods will be considered. As stated previously, the term "Remote Sensing" is applied to the detection of objects and the determination of their position and some properties without making actual physical contact with them. Although the term remote sensing covers all the techniques for making observations at distance, such as those based on acoustics, gravity and aero-magnetics, at the present time, the normal sense of the term has been restricted to that of the electromagnetic energy.

A generic remote sensing system is composed of four basic elements (Chuvieco, 1995) (Fig. 6.33):

Sensor system: sensor and platform (including the rocket vehicle which transports it until in the definitive operation orbit);

Scene: is the ground area covered within a certain time by the sensor;

Source of Energy: is the Sun (for the passive sensors) or generated by the sensor (for the active sensors).

System of Process, Sale, Interpreter and Final User: involves the reception-capture station, antenna, tracking system, sales, distribution agency, interaction with the client and finally the final user (i.e. government agency, defence, university, domestic service companies, etc.).

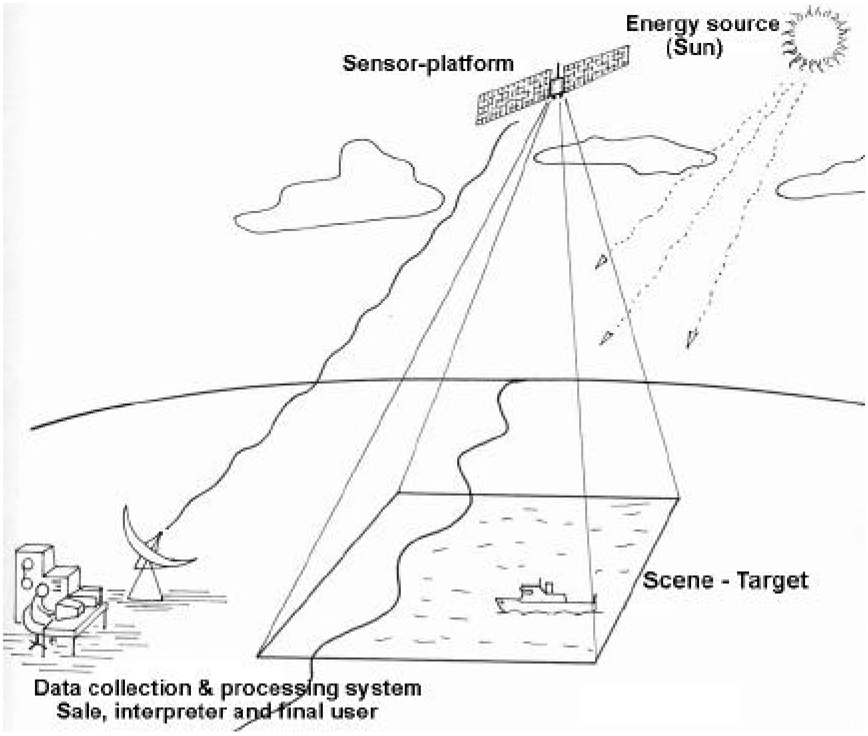


Fig. 6.33 "Remote sensing system (satellite passive sensor case)"

3.2.1 Satellite and sensors for Earth Resources Remote Sensing

The satellites employed in earth resources remote sensing use 2 types of **orbits** (Fig. 6.34):

a. **Equatorial geostationary orbit:** the satellite is at circa 36000 km distance and is over a fixed point on the Equator. These satellites only look at the Earth's surface in a particular way for a dedicated function; i.e. European meteorologic al satellite METEOSAT, the American GOES, etc.

b. **Cuasi-polar sun-synchronous orbit:** the satellite uses much lower orbits (700 to 1200 km) and it always passes over the Equator at the same time (sun-synchronous), moving a certain distance along the Equator and passing near the poles.
i.e. SPOT, LANDSAT, NOAA, METEOR, JERS, ERS, RADARSAT, etc.

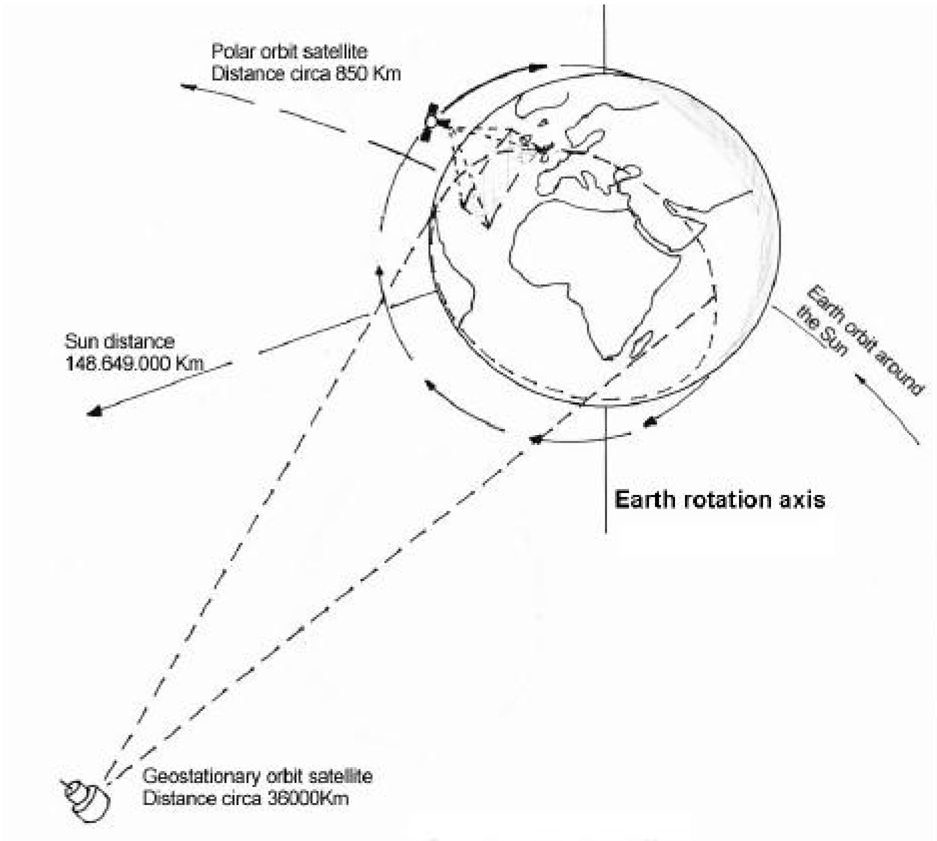


Fig. 6.34 "Main satellite orbits"

The **instrument-sensors** can be classified according to system methodology:

a) In the origin of the **energy source** used, they are divided into (Fig. 6.35):

Passive: the instruments capture the radiated energy emitted from the area of interest and generate a corresponding and measurable electric signal. The energy source is the Sun.
Examples: MSS and TM LANDSAT, AVHRR NOAA, HRV SPOT, MMRS SAC-C.

Active: the sensors emit an energy beam and register the backscattered proportion from the Earth's surface. They are able to obtain images in any meteorological or light conditions, since the energy source is self generated and independent of the Sun.
Examples: SAR ERS, JERS and RADARSAT.

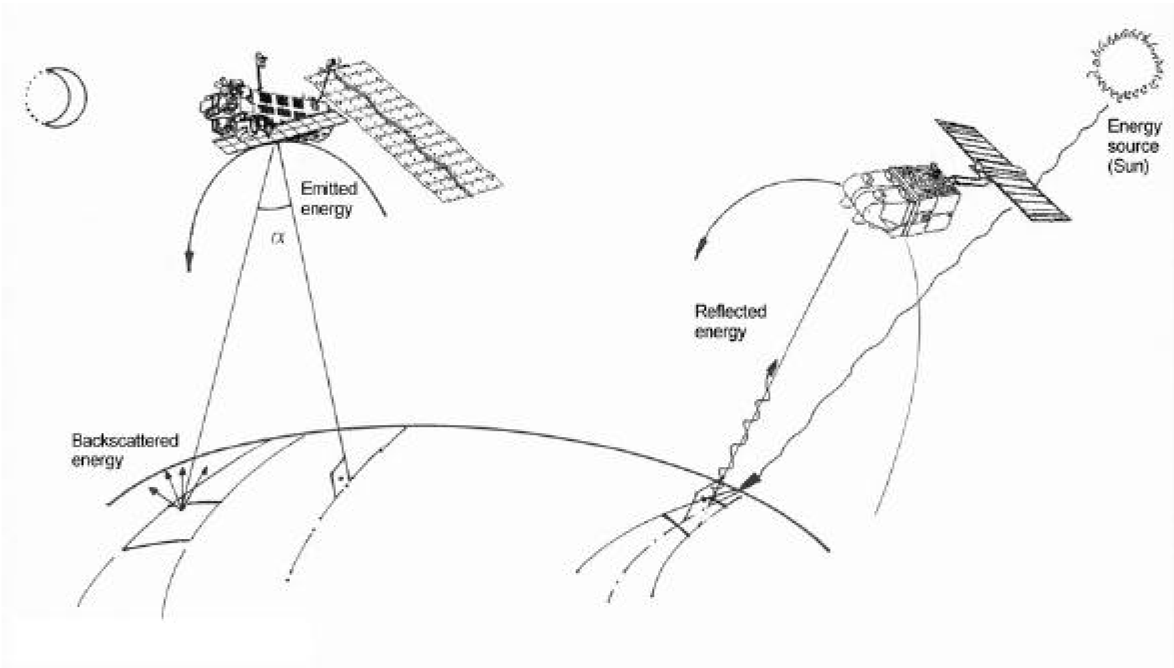


Fig. 6.35 "Passive and active sensors"

b) The useable sections of the **electromagnetic spectrum**:

Optic: this includes the visible spectrum of the human eye (0.4 µm – 0.7 µm) and the reflected or near infrared (0.7 µm – 3 µm).
Examples: MSS LANDSAT, HRV SPOT, MMRS SAC-C.

Thermal: corresponds to the thermal or emissive infrared (7 µm – 15 µm).
Examples: AVHRR NOAA, TM LANDSAT.

Microwave: the longer length waves (mm to cm) used mainly by the radars.
Examples: SAR ERS, JERS and RADARSAT.

Sources of electromagnetic radiation which may be used for remote sensing can be natural such as the sun, earth and atmosphere or artificial sources such as flash lamps, laser and microwave emitters.

The main natural energy source is the sun, whose radiated energy reaches a maximum (peak) at a wave length of 0.47 µm (green visible). On its way to the Earth, sun energy passes through the atmospheric layer and undergoes complex interactions, summarised in the effects of absorption, reflection, scattering and emission (Fig. 6.36):

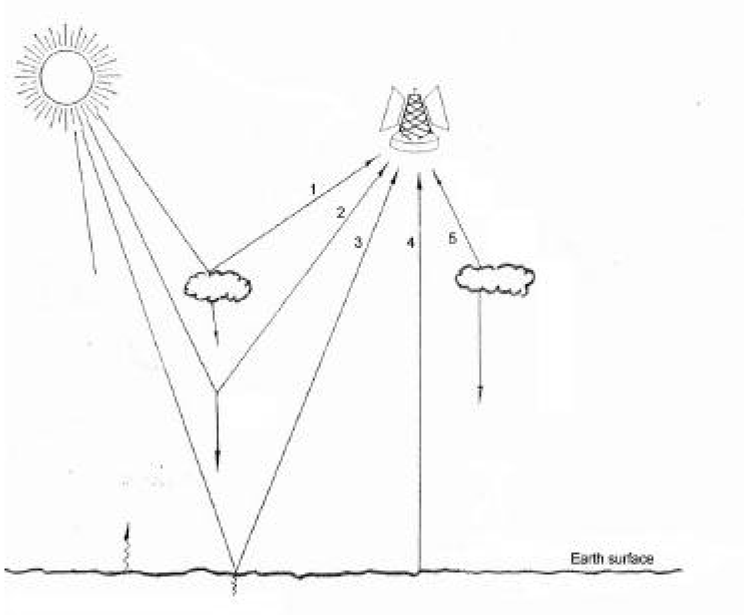


Fig. 6.36 "Radiation received by the sensors"

The different components are:

1. Radiation reflected by the atmosphere;
2. Radiation scattered by the atmosphere;
3. Radiation reflected by Earth surface;
4. Radiation emitted by Earth surface;
5. Radiation emitted by atmosphere.

Only a small part of the energy captured by the sensor is used to extract information regarding the Earth's resources: that reflected and/or emitted by the Earth's surface. The rest should be filtered to enable additional information to be extracted.

There are zones in the spectrum, which have a better electromagnetic radiation path; they are called "atmospheric windows" (Fig. 6.37). In these zones the absorption is lower, so the transmitted energy is higher. The main windows are:

0.4 – 0.7 µm in the visible;
3.5 – 5.5 µm and 8 - 14 µm in the thermal IR

The sensor captures and measures the electromagnetic energy coming from the area of interest in discrete spectrum bands. The measurement of the intensity of the transmitted energy from a target in each band is called spectral response or the "spectral signature" of this target.

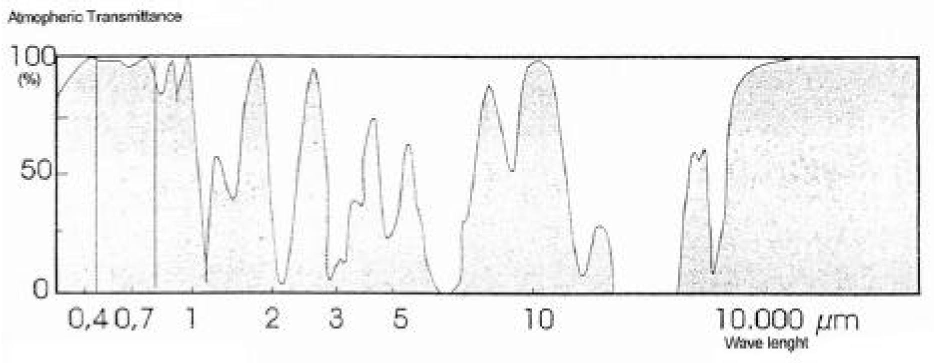


Fig. 6.37 "Atmospheric windows"

3.2.2 Main Remote Sensing Systems

The main remote sensing systems can be classified as follow:

PASSIVE SENSORS:

- Photographic systems
- Return Beam Vidicon Systems
- Opto-mechanical scanners
- Opto-electronical scanners

ACTIVE SENSORS:

- Radar systems

- **Photographic systems**

The photographic cameras were the first sensors able to receive multispectral pictures from space. They continue to be a method often used for remote sensing, particularly from air platforms. Their basis of operation is the impression of a scene onto photosensitive films via an optic system which allows control of the exposure conditions.

Most important characteristics are:

a. Film type : The most commonly utilised is the panchromatic film, on which the whole visible spectrum can be captured on a single emulsion. Radiation corresponding to the near or reflected infrared (IR) is captured in grey tones with the infrared film.

b. Number of objectives: Multiple observations can be carried out with two different constructions, incorporating several lens, each one of them with an appropriate filter, either in a single camera, which enable impressions of the same image in different bands of the spectrum, or by assembling several cameras on the same platform, each one with different filters and appropriate films. (See Fig. 38)

c. View angle: In vertical photography (the most often employed), the images are captured approximately orthogonal to the land surface (5° deviation being permitted) and in oblique photography, a viewing angle of less than 90° (used in studies of the relief, urban infrastructure, etc.).

d. Observation height: The height (**H**) is highly variable, depending whether it is air or space photography and the relationship with the focal distance (**f**) determines the scale (**S**) of the obtained photogramm (See 3.1.1).

$$\boxed{S = f / H}$$

An example of space photography can be cited in the panchromatic and IR pictures taken from the Space Shuttle during the European Spacelab Program (1983). Indeed, with the metric camera RMK 20/30 some stereoscopic pictures were obtained of several regions of the world, at a scale of 1: 820.000 at 250 km high with an approximate resolution of 20 to 30 m, these were used mainly for cartographic purposes (Konecny, 1986).

More recently, cameras like the MKF-6 (Fig. 6.38), on board the Soyuz space laboratory, have allowed the capture of pictures of high resolution and in 6 bands of the visible and near IR spectrum (Chuvieco, 1995). Also on board the Soyuz, cameras, as the KFA 1000, with focal distance of 1 m approximately and at 351 km of distance, achieved geometric resolutions from 5 to 10 m.

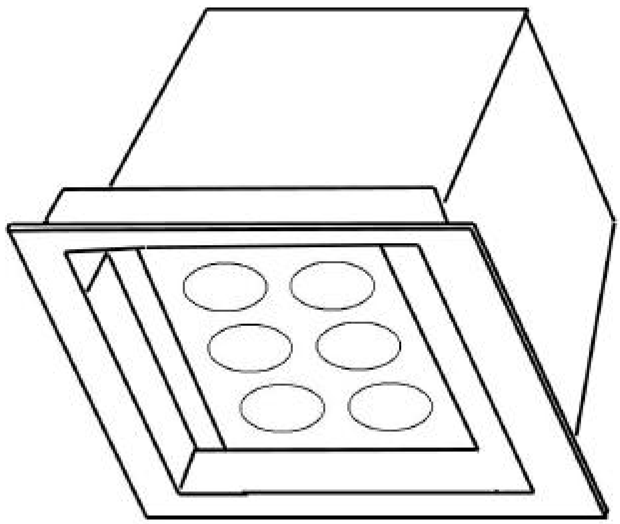


Fig. 6.38 "MKF-6 multispectral camera"

- **Return Beam Vidicon Systems**

Return Beam Vidicon (RBV) was a passive sensor similar to a television camera tube. This sensor failed early on the first ERTS (called LANDSAT after) and never came into routine use.

Two RBV cameras observed the whole surface in instantaneous form, using colour filters to provide multispectral bands centred in the blue-green, yellow-red and red-IR spectra in the first two LANDSAT satellites.

A forth RBV camera on Landsat-3 was a panchromatic (0.505 - 0.750 µm) version that provided four adjoining images at 30 m resolution.

This type of system has been used in TIROS and LANDSAT satellites, among others.

- **Opto – mechanical scanners systems**

These types of scanner are opto-mechanical instruments, where a mechanically driven optical element, generally a rotating or oscillating mirror, is used to deflect an optical beam to the detectors at right angles to the line of flight. The axis of rotation or oscillation of the mirror is parallel to the line of flight or orbit.

As examples, aircraft Daedalus scanner uses a rotating system and the LANDSAT satellite series utilises an oscillating system in their Multi Spectral Scanner (MSS) (Fig. 6.39).

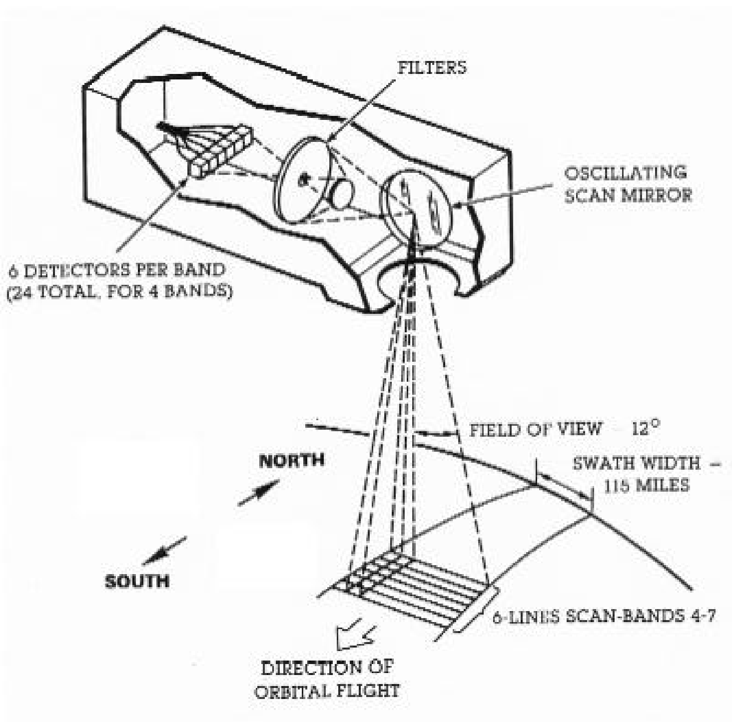


Fig. 6.39 "LANDSAT MSS (after NASA, 1997)"

MSS LANDSAT consists of a mobile mirror that oscillates perpendicularly to the flight direction. The received radiation passes to a set of detectors which amplify and convert it to an electric signal. This signal is converted to a numeric value which can be stored on board or transmitted to the network of reception ground antennas.

In summary, the sensor transforms an analogue signal, the received radiation, into a digital value, generating digital images. These radiation values can be translated, again, in radiation levels, knowing the calibration coefficients of the sensor and the conditions of acquisition.

The number and attributes of the detectors, which contain the scanning equipment, is fundamental in understanding the characteristics of the resulting image.

The signal sent by the optic system to these detectors is re-sampled at regular intervals, so only a numeric value is recorded at each certain distance. That interval marks the size of the minimum information unit acquired by the sensor; it is termed a “pixel” (picture element). The signal detected by each pixel has a direct relationship to the type of observed surface. If the signal originates from a homogeneous surface, the value of the pixel will correctly define it; in the case of a heterogeneous surface, the result will be an average of the characteristics of the area observed.

In many scanners, the received signal breaks down on board into several wavelengths, each one directed to a special type of detector sensitive to this energy range. These are known as multispectral scanners, because they are able to detect the same land surface utilising different spectrum bands.

The advantages of the multispectral scanners, in relation to the simple photographic sensors, are (Chuvieco, 1995):

a. They enable enlargement of the detected spectrum band to wavelengths longer than the visible one. The emulsions are limited to the range 0.4 to 0.9 μm, while the scanners can embrace from 0.4 to 12.6 μm, including the medium and thermal infrared;

b. Easier calibration and radiometric correction of data;

c. Ability to undertaken systematic and extensive coverage due their capacity to transmit data in real time;

d. Digital recording of the information, which improves their reliability and allows computer processing.

Disadvantages are the limited area resolution and the need for specific image processing systems.

Examples of these systems are the Advanced Very High Resolution Radiometer (AVHRR) in the satellites TIROS-NOAA and the Multi-Spectral Scanner (MSS) LANDSAT.

A more sophisticated multispectral imaging sensor, named the Thematic Mapper (TM), has been added to LANDSATs 4 to 7. Although similar in operational modes to the MSS, the TM consists of seven bands which have differential characteristics, adding bathymetric, geological and thermal capabilities with improved geometric resolution.

- **Opto – electronic scanners systems**

In the optic-electronic scanners, also called “pushbroom”, the oscillating mirror is eliminated, due to a chain of detectors which cover the whole field of vision of the sensor. These detectors are energised by the orbital movement of the satellite, enabling them at each instant to survey a complete line, which moves simultaneously with the platform. The solid detectors which make-up an optical-electronic scanner are termed “Charge Couple Devices” (CCD) (Fig. 6.40).

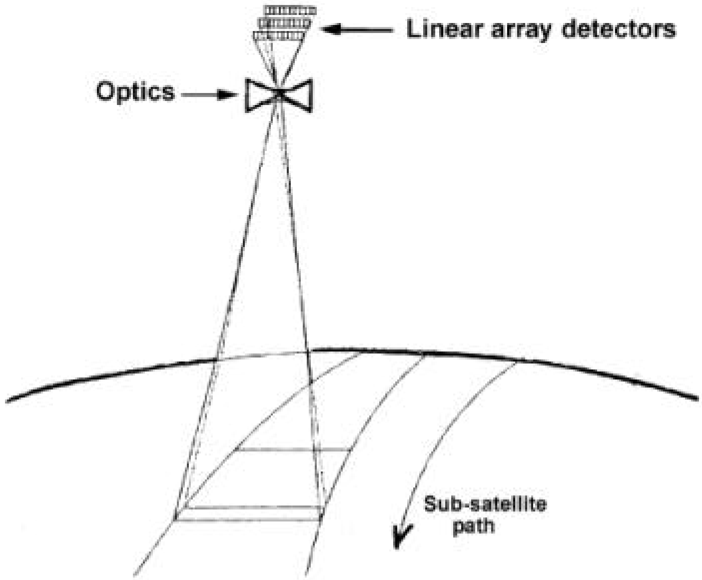


Fig. 6.40 "CCD system (after Chuvieco, 1995)"

With this type of sensor the space resolution of the system increases compared with conventional scanners, whilst eliminating the moving parts. It is also not necessary for the sensors to be interrogated by pixel, but rather by line, which in turn makes it more responsive for detection and emission of data.

Examples of this system are the High Resolution Visible (HRV) sensor of the French SPOT satellite, the German MOMS and the sensors of Indian IRS-1 and Japan MESSR MOS-1.

- **RADAR systems**

The RADAR (**RA**dio **D**etection **A**nd **R**anging) enables information about the topography, roughness, land cover and moisture of the scene to be acquired using an active radiometer of microwaves, which works in a spectral band between 0.1 cm and 1m. Due to their ability to operate in any atmospheric and light condition, they are increasingly used. There are important differences between how a radar image is formed and what is represented in that image compared to optical remote sensing imagery. To interpret radar imagery, it is necessary to understand the radar configuration, the energy associated with radar remote sensing, the way in which that energy interacts with surface targets and the way in which this interaction is represented in the image (Davidson, 1997).

The radar principle of operation is based on the emission of a pulse (beam) of microwaves (radio) toward the scene or target. The incident energy is backscattered by the scene towards the radar, which measures the intensity (detection) and the time lapsed between emission and reception (range) (Fig. 6.41).

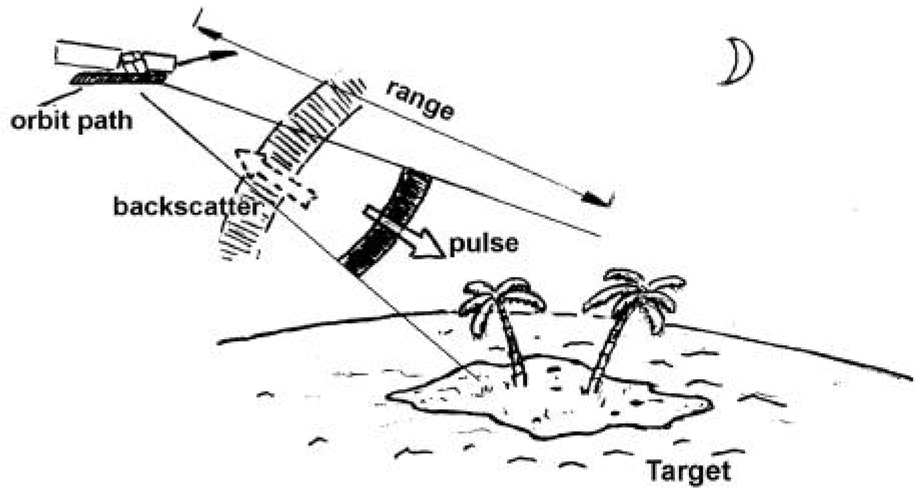


Fig. 6.41 "RADAR basic operation"

The **Synthetic Aperture Radar (SAR)** is the radar type most used in satellites. The principle is based on the Doppler effect, which affects the recorded observations when a relative motion between an object (target) and sensor is observed in the pulses from the terrestrial surface target caused by successive moments of the satellites orbit trajectory. The resulting resolution is equivalent to that which would be obtained with an antenna of similar length at the distance between the extreme points from which backscatter is received from the same target (Fig. 6.42).

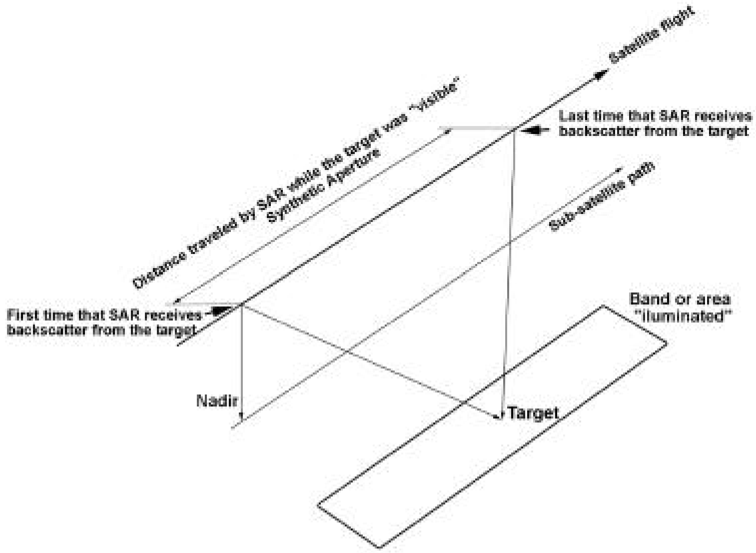


Fig. 6.42 "Synthetic Aperture Radar concept"

Examples of SAR natural resources sensors are the European ERS and ENVISAT, the Canadian RADARSAT and the Japan JERS.

3.2.3 Image Structure and Support

An image is generated from the energy captured by the sensor, which converts it to an analogical signal. Then it is processed and stored as a numeric value. The regular storing interval of the signal determines the images' unit of information. This minimum segment of data, represented by a single digital value, is termed a "pixel" (Picture Element), and, as was previously stated, it depends on the sensor geometric resolution. The pixel is characterised by a Digital Number (DN), resulting from the digital codification of the radiation detected for that range of the spectrum or band.

The numeric image is a geometric array (matrix), of two dimensions. In each pixel **Pij** (elementary point of the matrix) there are three associate values:

a. their line co-ordinate **Li**;
b. their column co-ordinate **Cj**;
c. the physical measure made by the receiver in that pixel in a range of wavelength: DNij.

A multispectral image is constituted by **k** arrays, called channels or bands. In this case, the image becomes a three-dimensional array, incorporating the band like third dimension. For example, an MSS LANDSAT multispectral image possesses four channels MSS_k, where k = 1, 2, 3, 4. (Fig. 6.43). The radiometric intensities of a channel are numeric recounts with values whose limits are between 0 and 255, allowing up to 256 possible values in general. These values are coded in bytes or 8 bits.

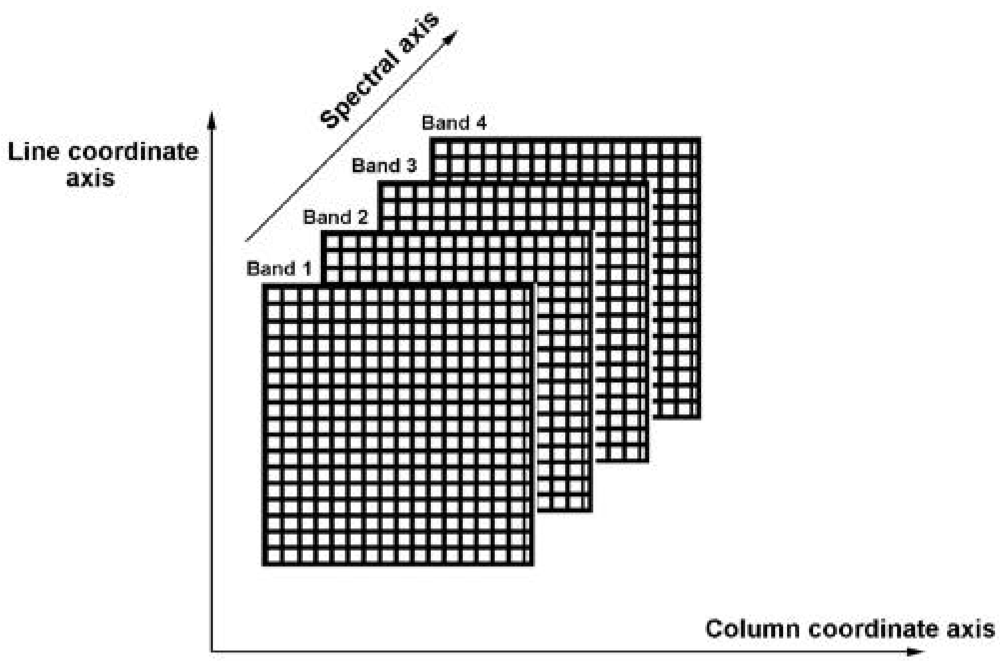


Fig. 6.43 "Multispectral image array structure"

The fundamental principle of visualization of a digital image is to associate a colour or a tone of grey to each radiometric value, conserving at the same time the matrix representation of the image. There are two visualization possibilities: the visualization of a single channel and the visualization of several channels in additive synthesis of colours.

In the case of unique channel visualization, a correspondence between intensity (DN) and tone of grey is defined, so that the minimum intensity (0) is assigned to black and the maximum intensity (255) to white, assigning the intermediate values to different tones of grey. The histogram of a numeric image is a graphic representation of the frequency of appearance of the different levels of radiometric intensity (DN); thus expressing the pixel distribution as a function of their radiometric intensity. The histogram allows one to know the distribution of the pixels in the image for the interval values from 0 to 255.

To improve a digital image it is possible to modify the correlation between numeric values and the range of grey or colour. The objective is to increase the global contrast of the image. It is done by replacing the original values, between the minimum and maximum levels, with new values distributed within the 255 levels, as a way of using all the available grey levels in the visualisation. This can be achieved by applying a lineal function (straight line), adapting the image according to the curve of the accumulated histogram or by other viable methods of distribution, the most common using exponentials, the lineal segments, etc.

For colour visualisation, the principle is the same as that for visualisation in black and white (B/W). The only differences being there is a colour associated with each channel numeric value and no grey intensity. Thus there should be a proper palette of colours defined.

There are conventions for the channel colour combination. For instance, the MSS LANDSAT normalised False Colour Composite (FCC) image assigns the blue colour to the green band (centred at 0.55 μm), green to the red visible band (0.65 μm) and red to the near-photo IR band (0.75 μm).

There are alternatives for the **image recording format**. In general, the image contains a “header” file, which indicates the recording format, the sensor type, geographical location of the area, date, solar position, data of corrections and calibration of the image. The more frequent recording formats are:

BSQ (Band Sequential): DN follows a sequential order, leaving the origin (line 1, column 1) until the final pixel of the first band; the succeeding bands then follow on.

BIL (Band Interleaved by Line): The DN are ordered by line. The first line of band one begins, then the first line of the second band and continuing on with the other bands. Once all the bands are completed, it passes to second line of the first band, the second line of the second band, etc.

BIP (Band Interleaved by Pixel): The format is similar to the previous method, except in this case the DN are ordered by pixel. The first pixel is recorded at the origin of each band, then the second pixel, and so forth until completing the image.

The **available image support**, requesting the negative or positive film or photographic page, varies according to the space programme. Negative film is the most versatile product, since it enables the generation of all the amplifications considered necessary at the desired scale. Positive film is very useful for photographic reproduction and impressions of the image. Paper is the most commonly issued analogue media as it guarantees the ability for a direct interpretation of high quality images; however the scale in which it is presented is rigid. LANDSAT TM image negative films (23x23cm) are offered at a scale of 1:500.000, while other programmes offer images at a scale of 1:1.000.000, additionally amplifications can be obtained in paper at scales of 1:250.000 and 1:100.000.

Presently magnetic Computer Compatible Tapes (CCT), Exabyte tapes and Compact-Disk Read Only Memory (CD-ROM) are the most useful digital media types.

3.2.4 Interpretation and Processing Fundamentals

Image interpretation refers to the techniques required to define, to recognise and to identify objects or phenomena in an image and to interpret their meaning. To conduct these tasks, it should be considered an essential part of preparation work to define the parameters and methods to be used.

The scale is one of first parameters to be defined and is linked to the previously mentioned objectives; the scale defines the minimum unit of information which should be included in the map, termed the Minimum Cartographic Unit (MCU). It is recommended that the MCU not be less than 4 mm^2 at scale of the map. Thus, the work scale should be related directly with the most suitable type of sensor to undertaken the project. In accordance with the International Cartographic Association guidance, the most suitable scale limits for the different sensors are:

LANDSAT – MSS	**1 : 200.000**
LANDSAT – TM	**1 : 100.000**
SPOT – HRV	**1 : 50.000**

To summarise image interpretation factors, the following should be considered:

- **Sensor-platform system**: The most appropriate sensor type depends on the objectives and the level of precision required for the project; i.e. global mapping (planispheres) will be carried out starting with sensors of low space resolution (NOAA AVHRR or SAC-C MMRS) while those requiring a larger scale will use sensors that offer greater spatial detail (LANDSAT TM or SPOT HRV). However, in other cases, the spatial resolution is secondary to time or spectral resolution; if the studied phenomenon is very dynamic in time, such as the detection of spills of oil in the sea, it would be suitable to use sensors of higher temporary resolution, whilst sacrificing space precision. In other projects, spectral dimension will be more important such as for studies of ocean colour starting from optic sensors.

- **Image capture date**: The most suitable moment to acquire images will be when the phenomenon to be studied has its highest discrimination ahead of others of similar spectral behaviour; i.e. riverside area mapping with large tidal amplitude and extended beaches, it will be most suitable to choose low tides due to having the highest quantity in detectable coastal details, thus image capture should be planned with an analysis of local tide predictions.

- **Image support**: Selection of media, on which the interpretation is carried out, depends on the techniques to be applied. If a visual analysis, then three main aspects should be considered: the material support of the image, the scale and the band number or selected combination of bands.

Photographic film or paper is ideal for analogical (visual) interpretation, while Exabyte, floppy disks, CCT or CD-ROM are best for digital processing. Additionally, the ideal number of bands for a project depends on the phenomenon being mapped or monitored.

- **Selection of the analysis method**: Image analysis methods can be visual or digital. Each has its own advantages and disadvantages. Visual treatment requires lower inversion than digital processing; however, computer processes presents lower unitary costs with larger areas, while visual interpretation follows linear costs.

In summary, when undertaking complex works the results of both methods are suitable, although digital methods are gaining in importance due to advances in image processing, via computer equipment (hardware) and programs (software).

Remote sensing **image visual interpretation** is based on the same skills used in classic air-photo interpretation. VIR and SAR images interpretation are similar in that the same interpretation key is used. When SAR images are employed, the unique properties of radar imagery must be remembered and incorporated into the interpretation process.

Main elements of visual interpretation used are:

- **Scale**: It is the relationship between lineal dimensions in the image and on the terrain (ground).

$$\boxed{\mathbf{S = Image \ / \ Ground}}$$

In general, scale (**S**) is expressed as a division with numerator equal to "**1**" and denominator "**D**":

$$\boxed{\mathbf{S = 1 \ / \ D}}$$

- **Shape and size**: Shape and size are linked directly with scale; shape refers to the spatial form of an object or area, shape can help to distinguish between natural and cultural features.

The size of a feature can be helpful in distinguishing features from each other in relative terms. The scale is one factor, which influences the size of an object or feature present on the image. Shape, size and scale are fundamental for the definition of the patterns.

- **Tone**: The tone refers to the energy intensity received by the sensor for a certain band of the spectrum. In a photographic product, pixels with dark tones indicate those areas where the sensor detected a low sign, while the clear areas are those of high radiation values. The tone is related closely with the spectral behaviour of the different land covers for the specific spectrum band in which it works.

In radar images, tone results from target backscatter, tonal variations are usually functions of the strength of the radar backscatter from the ground; i.e. smooth water surfaces appear dark because they act as a specular reflector with the energy being reflected away from the sensor.

- **Colour**: In VIR images the colour is a consequence of the selective reflectivity from the objects to different wavelengths. Those surfaces with high reflectivity in visible short wavelengths and lower in the rest appear with blue colour, while those which absorb short wavelengths and reflect long ones appear with a red tint. If the sensor captures information in the bands of the blue, green and red spectrum, a composition in natural colour can be obtained.

SAR images are mono-band and so they are viewed in grey tones.

- **Texture**: Texture is the frequency of tonal or colour changes. It refers to the apparent roughness or softness of the image region, representing the space contrast between the elements from which it is composed.

The texture of the image comes from the relationship between the size of the objects and the resolution of the sensor. In general, texture is classified in thick, medium and fine texture. In SAR images, it may be classified as smooth, fine, grainy, linear, speckled and flecked.

Contrast is the relationship between clear and dark areas or the tone relationship between an object and the surrounding objects.

- **Shadow**: Usually shadow links the dimensions of the object (mainly its height) with the angle of incidence of the energy (Sun or waves beam).

In SAR images, shadows indicate relief type. Shadow length can be used to estimate height, while their projection indicates spatial form.

Finally, visual interpretation is carried out by evaluating all the above mentioned parameters and comparing the characteristics of the displayed objects with well known patterns (i.e. land cover, drainage net and urban infrastructure, etc.).

The process of identifying or helping to identify features through local and regional context is called association. For instance, terrain landscape and Antarctic Sea features tend to form associations through well-understood natural relationships and processes (i.e. floes, fractures and glacial landforms).

- **Pattern**: Pattern represents the orderly spatial arrangement or repetition of features. Spacing, density and orientation are indicative of pattern; for example river (or watershed) net is linked to relief, dendritic pattern is representative of an undulating area (hills, mountains), while meander patterns represent flat or plain areas.

3.2.5 Image pre-processing and complementary data

The satellite images, obtained by a third party, will have been processed by the acquiring institution in order to standardise the available products. A base treatment is conducted and, at the client's request, additional optional treatments improving the geometry and radiometry of the product can be completed, better adapting it to the objectives of its future application.

The process type and their denomination are characteristic of each system. In general, it is organised in a progressive hierarchy of corrections, such that each level includes all the previous ones whilst adding others.

For example, for HRV SPOT products, there are the following treatment levels (Fig. 6.44):

Level 1A: General basic level for all the images, where a calibration of sensors in each spectral band has been conducted. There is no geometric correction. These images are used in fine radiometric studies.

Level 1B: Systematic deformations caused by terrestrial rotation, panoramic effect, drift effect and incidence angle are corrected. There are corrections for geometric origin, but that may have influenced the radiometry, because a re-sampling is carried out.

Level 2: Geometric and localisation corrections are carried out using internal and external data. The internal data employees are: data of restored orbit, geometry of the instantaneous field of view (IFOV) and satellite altitude restitution auxiliary data. The external data are: parameters of the selected plane representation system, rectification medium altitude selected and ground control points (GCP) co-ordinates.

There are 2 sub-levels, in functions for the use and not of the GCP:

Level 2A: Two-dimensional corrections are carried out to transfer the scene to a certain cartographic projection (Mercator, Transverse Mercator, Lambert, etc. - see Chapter 2). Data from satellite altitude and geometry of the IFOV (instantaneous field of view) are used, without using GCP.

Level 2B: The geometric correction uses GCP, obtaining a higher precision than in Level 2A.

Level 3: The geometric deformations produced by the relief are considered. It requires data from the Digital Elevation Models (DEM). The absolute position precision is of the order of 0.5 pixels. As result, an orthomorphism is obtained.

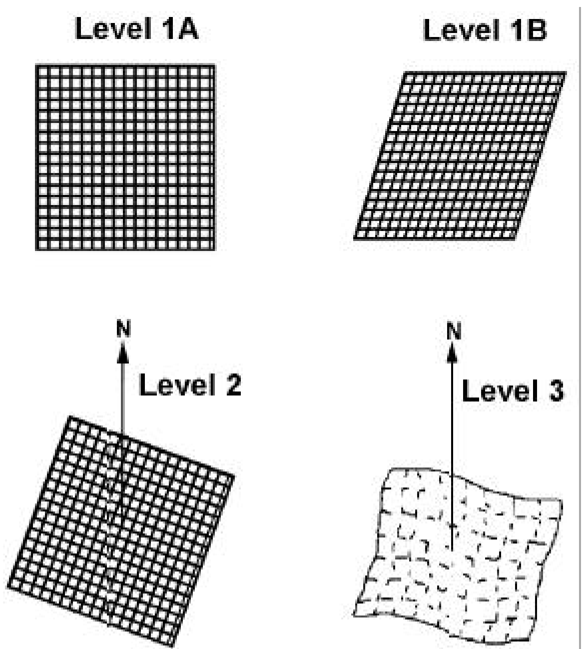


Fig. 6.44 "SPOT image processing levels (adapted from GDTA, 1993)"

Complementary data
Image processing requires complementary data to be used for its geometric corrections and preparation of the final cartographic product. The geo-codification can be conducted without the GCP but the resultant position is only relative, if a link to a reference system is required, the use of GCP will be required.

Ground control points (GCP) are points with previously known co-ordinates and clearly identifiable in the image. They are used to define the co-ordinate transformation equations from the image to the required geodetic reference system and cartographic projection. They can be obtained from cartographic documents (charts) or starting from topographical field tasks. In certain cases the land reflective signals (transponders) can be installed, which are easily detectable in the image, thus enabling the geo-codification in areas with few natural or artificial details.

Additional cartographic data (type of coasts, navigation obstacles, coastal details and urban, port and road infrastructure) can be used for the interpretation of the image or to enhance the final product. They can be obtained from digitalisation of existent charts, from geo-referenced databases (GIS) or from complementary surveys.

3.2.6 Image processing

Satellite images present geometric and radiometric distortions, which are dependent upon the sensor type, platform and the capture conditions. In hydrographic applications, information from multiple sources is very frequently used. Therefore, to standardise, and thus be able to compare and integrate the acquired data with other information, it should be normal procedure to rectify and to restore (rectification and restoration) the satellite images. The levelling and correction process depends on the evaluated images and on the application to which the final product is to be put. In some cases it can be enough just to correct the systematic errors and then co-register the images with other previously geo-referenced data; in other cases, the images will be corrected and re-sampled in a cartographic projection with a given scale. The complete correction of an image involves the initialprocessing of the raw image data to eliminate the geometric distortions, the radiometric calibration and the reduction of the actual data noise.

When images of diverse sources (i.e. LANDSAT TM, SPOT PAN, etc) are used, the processes of geometric correction, rectification, radiometric calibration and enhancement are prior requirements for the image fusion and they assure the compatibility on a pixel-by-pixel basis. The radiometric enhancement is as important as the geometric integrity in all aspects of mapping with images because the quality of the resulting final fused image depends on the precision of the geometric corrections in each participant image (Pohl, 1996). This should be given particular consideration given the frequent employment of mosaiced images to complete sectors of charts.

<u>Geometric treatments</u>

The geometric distortions can be classified as systematic (predictable and correctable) and accidental (random). The systematic errors are easily recoverable by applying formulae derived from modelling the sources of distortion. The accidental errors are corrected by applying polynomials with points of ground control (GCP) conveniently distributed in the image.

Geometric corrections can be grouped in following processes:

Co-registration (or registration simply): it is the adjustment of an image taking as a reference another image, using a polynomial transformation between common points in both. It is used when comparing two data sets, without utilising the cartographic projection (absolute position).

Geo-reference: it consists of the assignment of co-ordinates to the pixels of the image by means of the definition of the transformation equations.

Geo-codification: it involves the passage from image to map by means of the application of the transformation equations. The image becomes a chart, where each pixel has its corresponding geographical co-ordinate pair. The geo-codification is central to integrate images from diverse sources, achieving the integral compatibility of its data on a pixel-by-pixel basis.

Polynomial adjustment

Polynomial rectification is a relatively simple method of geometrically correcting the images. It consists of the transformation of the original image based on a group of appropriately distributed points with well-known co-ordinates. It is necessary that the points have co-ordinates in both systems: origin (**x** & **y**) and final (**X** & **Y**).

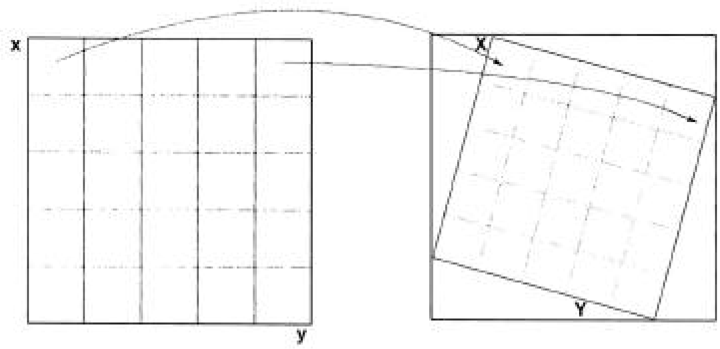


Fig. 6.45

The number of well-known points reflects the order of the polynomial in use; as the order increases, a higher number of points with well-known co-ordinates are required. A system of equations, whose coefficients are obtained by a least squares adjustment method, is produced.

A polynomial of first order (linear) requires 6 well-known co-ordinated points, it corrects for translation, rotation, scale, inclination, perspective and oblique distortions in the image (Fig. 6.46).

$$X = a_0 + a_1 x + a_2 y$$

$$Y = b_0 + b_1 x + b_2 y$$

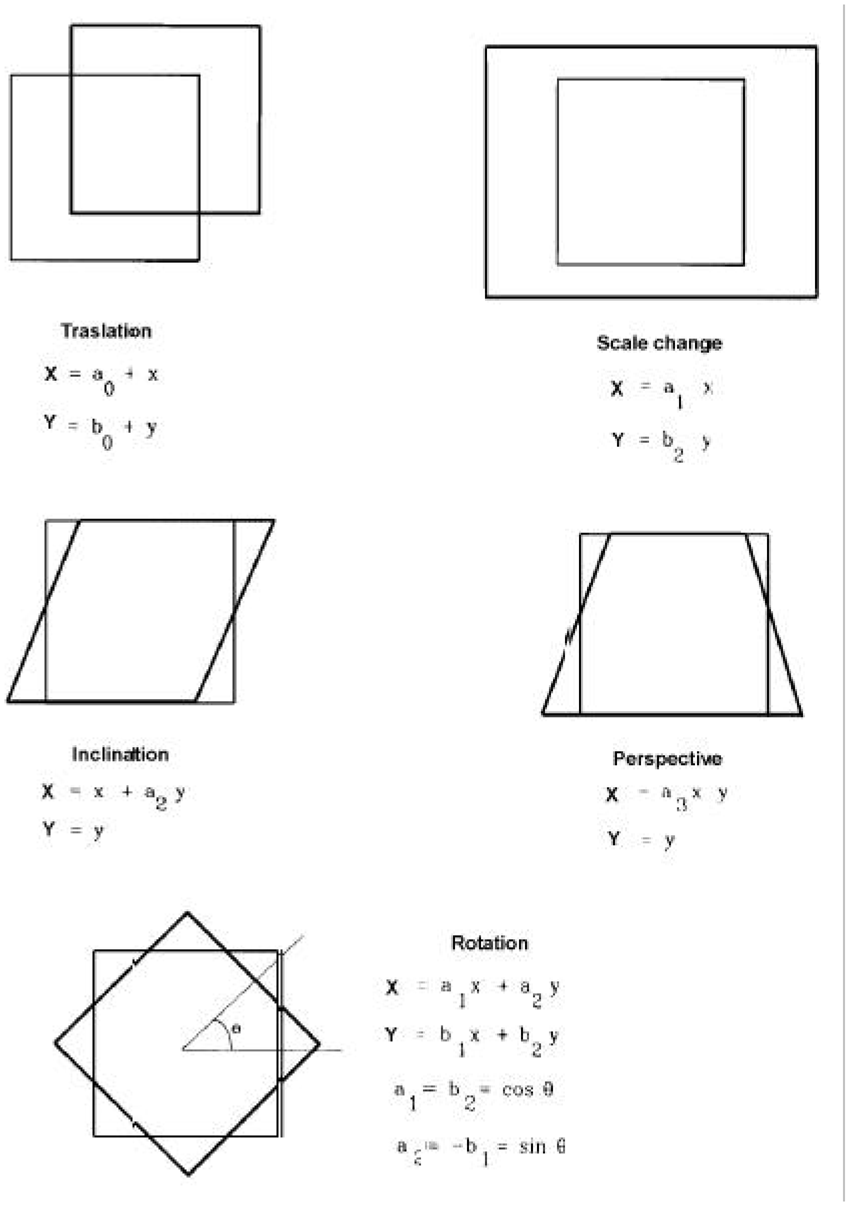


Fig. 6.46

A polynomial of second order requires 12 well-known co-ordinated points; it adds torsion and convexity parameters.

$$X = a_0 + a_1 x + a_2 y + a_3 x^2 + a_4 xy + a_5 y^2$$

$$Y = b_0 + b_1 x + b_2 y + b_3 x^2 + b_4 xy + b_5 y^2$$

Applying these equations, the original image can be transformed, resulting in an image shifted, rotated, scaled and warped.

The polynomial approach only locally corrects the image, since it depends on the distribution of points and their precision. In general it assures a correct image in flat featureless lands, but it is not very useful for very undulating areas.

The effectiveness of the adjustment comes from the evaluation of the residuals (deviations); the more often used indicator is the Root Medium Square error (RMS).

The assignment of the proper DN to the new position (**X** & **Y**) pixel can be made following different algorithms: nearest neighbour, bilinear interpolation and cubic convolution.

Nearest neighbour assigns to each pixel of the transformed image the DN of the nearest pixel in the original image. It is the quickest solution, but some linear features (roads, riversides, etc) can appear as fractured lines in the transformed image.

Bilinear interpolation calculates the measured average of the 4 nearest pixels. Here the distortion of the lineal features is smaller but the spatial contrast is diminished.

Cubic convolution considers the DN of the 16 nearest pixels. It produces a better transformed image but it requires considerably larger calculating capacity.

In summary, the choice of the method depends on the final use and objective of the project, on the available computer resources (hardware, software) and on the GCP availability. Also, image processing differs depending on whether the image comes from an optic (VIR) or from a microwave (SAR) system.

Geometric effects in VIR images

In general, optic images are more distorted by the process for obtaining the image itself than by the physical characteristics of the scene. For some sensors, such as the TM LANDSAT for example, the aspect (shape) and terrestrial rotation movement are important factors to consider during the correction of their images. In general such processes are undertaken by the supplier (see 3.3.5).

For the geo-codification, as described previously (see 3.3.4), object data of well-known co-ordinates (GCP) is a necessity and, in general, an adjustment by means of polynomials is beneficial.

Geometric effects in SAR images

SAR is a system very sensitive to the physical-chemical and geometric aspects of the target. SAR emits an energy beam that strikes the target surface sideways, creating a particular geometry with their images (Fig. 47) that can be summarised in the following terms:

- Altitude: distance between the satellite and the sub-satellite point on the surface of the Earth;
- Nadir: intersection of the vertical from the satellite with the terrestrial surface;
- Azimuth: direction, relative to North, of the trajectory of the satellite Nadir point on the terrestrial surface;
- Range vectors: vectors that connect the SAR to the ground, corresponding to each measured range sample at a single pulse transmission time;

- Slant range: the distance from the sensor to a target located in the range direction;
- Range direction: range vectors direction (perpendicular to azimuth);
- Ground range: the slant range projected onto the Earth's surface;
- Incidence angle: the angle between the radar range vector and the local vertical direction (Earth normal);
- Local incidence angle: angle between the radar range vector and the normal to the surface of each land element.

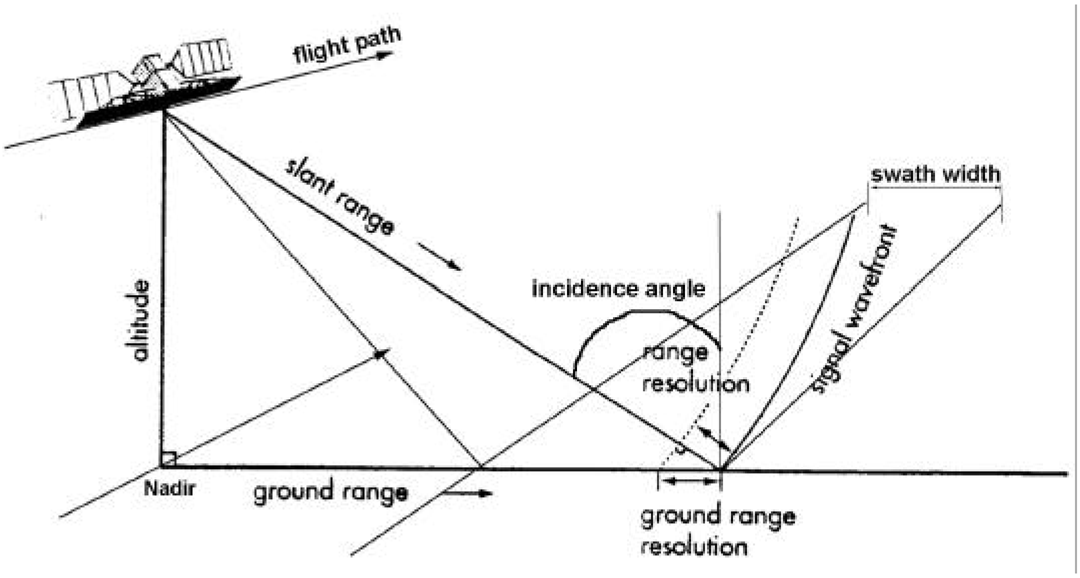


Fig. 6.47 "SAR imagery geometry (adapted after Raney 1992)"

The main parameter is the local incidence angle (Fig. 6.48). It can be seen that the geometry of signal-target interaction is a function of the land slope, which causes various distortions which differentiate it from a true orthogonal projection.

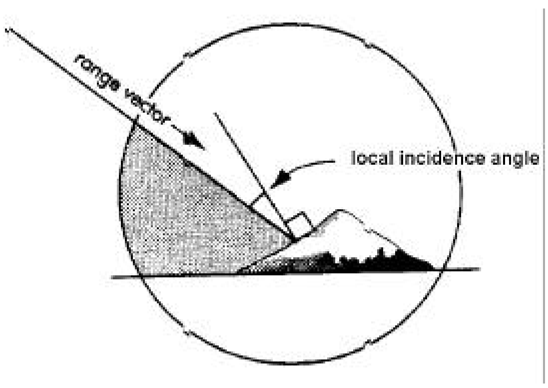


Fig. 6.48 "Local incidence angle (adapted after Raney 1992)"

Main distortions are called "foreshortening", "layover" and "shadow" effects (Fig. 6.49).

The effect “foreshortening” occurs when the local incidence angle is smaller than the angle of incidence but greater than zero. This distortion produces an effect which makes the viewed slope appear to be shortened and leaning towards the sensor.

In cases of small incidence angles or very abrupt relief, the radar signal returns from the peak of the mountain before that from the base, producing the effect of “layover”. In these cases the local incidence angle is larger than the angle of incidence.

The “Shadow” occurs on the slopes which are not illuminated by the radar signal. These areas appear very dark (without information) in the images.

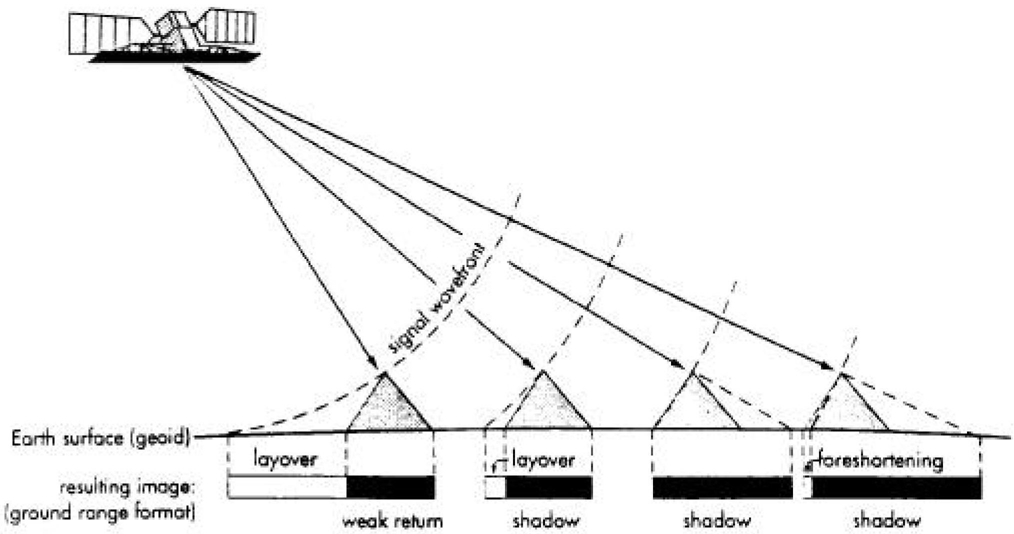


Fig. 6.49 "Distortions due to the relief (adapted after Lillesand and Kiefer 1987)"

<u>Radiometric treatments</u>

These techniques are useful to improve the radiometry, so that the features of interest appear clearer and more understandable to the interpreter. They are additional methods to those previously mentioned and they help in the interpretation of topographical features.

A common method involves the manipulation of the statistic of the image, represented by its histogram, which details the spectral frequency for each band of the image.

To improve the interpretation of the image, the association between the numeric values and range of grey or colour is modified with the intention of increasing the global contrast in the image (histogram stretching). This is equivalent to altering the current digital values for minimum (MIN) and maximum (MAX) with new values distributed within the 255 levels to make use of all the levels of grey possible.

The distribution can be conducted in several ways, the most frequent being a lineal distribution, which involves the values between MIN and MAX being redistributed in a straight line between 0 to 255 (Fig. 50).

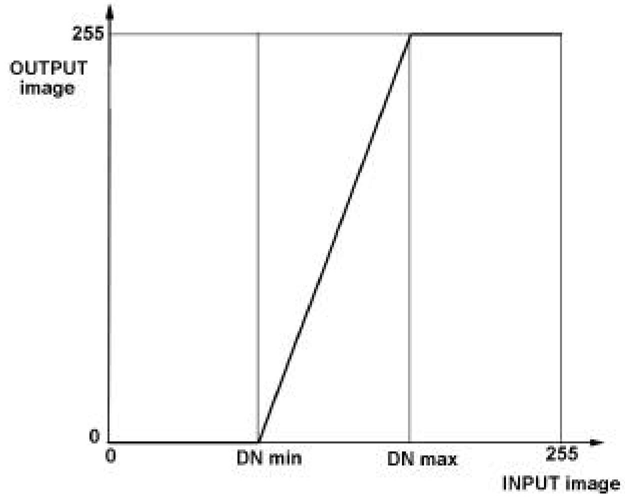


Fig. 6.50

Another means for assisting in the interpretation of the images is the application of spatial filters. The spatial filters are used to select or to mask a range of values, inside the total DN scale. These filters suppress certain frequencies, depending on the filter type. The low-pass filter reduces the range of DN in an area, reducing details and smoothing the general aspect of the image. The high-pass filter enhances the gradients of DN, i.e. the edges, which are used to better detect roads, railroads, riversides, etc.

The ideal filter is one which smoothes homogeneous areas whilst at the same time preserves limits and texture; it should maintain the arithmetic average and diminish the standard deviation.

In SAR images, adaptive filters are examples of high-pass filters and are employed to minimise effects of speckle. There have been numerous algorithms developed in recent times: Lee, Kuan, Frost and Gamma MAP.
In summary, filters are important for enhancing features and they can contribute to topographical interpretation, if they are used carefully and with discretion.

Radiometric distortions in VIR images

The sources of radiometric distortions are the atmosphere (due to dispersion and absorption effects), the sensor (effect of "striping") and the solar illumination. These effects are described in detail by Lillesand and Kiefer (1994) and Richards (1986).

Among the different components of the atmosphere, the effect of water vapour in form of haze can be reduced by applying a modification of the histogram.

A frequent problem is the presence of cloud which reduces the detectable data in optic images. Areas covered by clouds present a centre with high DN (white, near 255) with diffused and grey borders. Their corresponding shadows are also detected with very low DN. Resolution is via the application of diverse methods (thresholding and density slicing) and masks but they require a careful control, since some artefacts can be produced.

In general, radiometric distortions created by the sensor are small compared with the atmospheric influences. The most important of these distortions are those generated by the detectors, which manifest in the form of strips (striping). It is a recurrent effect, which can be removed after the interpretation or processing of the image (Crippen, 1989).

The correction for the different solar illuminations requires a ratio between bands (band rationing). This correction is not usually applied, since the solar illumination produces an effect which facilitates the visual interpretation for cartographic updating.

Radiometric distortions in SAR images

SAR images present their own characteristics, which require particular treatments and calibration. These characteristics are related to speckle, the process based on multiple views, the range of DN and radiometric specific corrections.

The "**speckle**" is a characteristic phenomenon of radar images; it is produced by interference of the coherent beam by various individual reflectors. The backscattered energy represented in a pixel is formed from the contributions of a large number of individual reflectors, such tree and vegetation foliage. Interference of the returning waves to the SAR produces variations in the grey level of the neighbouring pixels, creating a grainy appearance to the image. Speckle occurs in active systems which use coherent waves and it limits the radiometric resolution of SAR images (Hoeckman, 1990, Schumann, 1994).

Since it is a random effect, it cannot be totally eliminated. The impact can be diminished using various procedures, which reduce the spatial resolution.

The process based on multiple views (**multi-look processing**) is a radar signal process, employed to reduce the speckle. It can be achieved either by processing the signal to create independent, single look images at reduced resolution, identifying them and then averaging the image power to form a multi-look image or it can be accomplished by processing the data to full resolution and spatially averaging the developed image. The image signal-to-noise ratio is preserved in multi-look processing. Multi-look processing requires special hardware and software, so it is usually performed at the data reception facilities.

The range of the pixel digital number "DN" depends on the dynamic range of radar signatures in the scene and on the digital coding used to create the image. Often SAR data are delivered in 32 or 16 bits per pixel, however many display and software packages only handle 8 bit range data. Ranges of 16 or 32 bits require high storage and processing capacities. Other conditions (human eye resolution, display and/or printing capacity, etc) make it more convenient to transform the final data to 8 bit range, thus these data are expressed in a range from 0 to 255 grey levels. The process of conversion to 8 bits is named "**scaling**".

Frequently additional radiometric enhancement is necessary in order to use the full range (0 to 255). This process, "**stretching**", increases the image contrast allowing the better detection of diverse features.

For **SAR images calibration** in particular, 2 types of radiometric processing can be applied:

- Absolute calibration: establishes a relationship between the DN in the SAR image and the target backscattering, independent of the time. It is used when DN should be compared between 2 or more images, for example for thickness (age) of sea ice, environmental effects, etc.

- Relative calibration: it establishes the same relationship between DN and backscattering, but only within the image. This results in a target having the same brightness regardless from where in the SAR image it is taken.

Generally radiometric calibration is carried out at the data acquisition facility.

3.2.7 Altimetry

Land and coastal altimetric information is of great assistance to the hydrographer. Description of the relief facilitates a clearer understanding of coastal topography, islands, port and signal infrastructure, etc. High resolution satellite systems allow the representation of relief by numerous diverse means. Presently the cartographic representation of relief has been generally via numeric modelling (Terrain Numeric Model TNM) and its digital versions (Digital Elevation Model DEM or Digital Terrain Model DTM).

Procedures have been developed to process various types of data (space photography, VIR sensors, SAR, altimeters), with different formats (analogical, digital) and for diverse methods (shadowing, stereoscopy, interferometry, polarimetry) taking advantage of the different characteristics of the sensors and the images (geometry, radiometry, phase), applying several types of technologies (analogical, analytic, digital) and of processing (interactive, automatic).

Among the methods, stereoscopic ones have been those which initially have spread more thoroughly for cartography due to the precursor of well-developed stereo air-photogrammetry (See 3.1).

Coming from the latest advances in the computer stereo vision, considerable advances have been achieved in the satellite stereoscopy; additionally radar image stereoscopy has had an important stimulus in the last 20 years.

From the launching of the ERS-1, interferometric techniques were extended using previously developed parametric models. With the inception of the RADARSAT-1 in 1995, radargrammetry was consolidated between the different methodologies for altimetric application, using it alone or complemented with VIR images (Toutin, 2000).

- **<u>Stereoscopic methods</u>**

Stereo methods are similar procedures to those used in air-photogrammetry (see 3.1.7), in which two images are used for the construction of the three-dimensional stereo model.

A digital stereo-plotter allows the measurement of features using two floating marks (one for each stereo pair image), which enable the views to be fused to give 3D cartographic co-ordinates (Toutin, 1995).

The processing of the stereo pair requires the use of digital restitution equipment and specific software. At the present time compact systems exist, based on PC computers which allow the stereo restitution of different digital images types (air, space, VIR, SAR). (Fig. 6.51)

Fig. 6.51 "Digital stereo-plotter scheme"

There are a variety of combinations to capture both images, which can be obtained in the same or contrary directional passes, diverse angles of incidence, etc.

The HRV-SPOT established system has a mobile device installed in the optic equipment, which facilitates the observation of the same area in successive passes (Fig. 6.52).

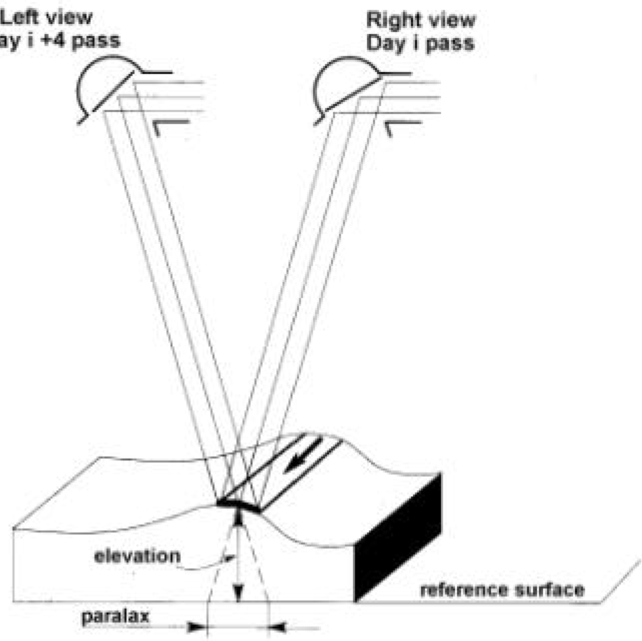


Fig. 6.52 "HRV SPOT stereo aptitude"

The MOMS system allows the capture of images in the same pass, by means of cameras in forward, after and nadir directions. The serial images are taken at intervals of 20 seconds, from three different viewing points (Fig. 53).

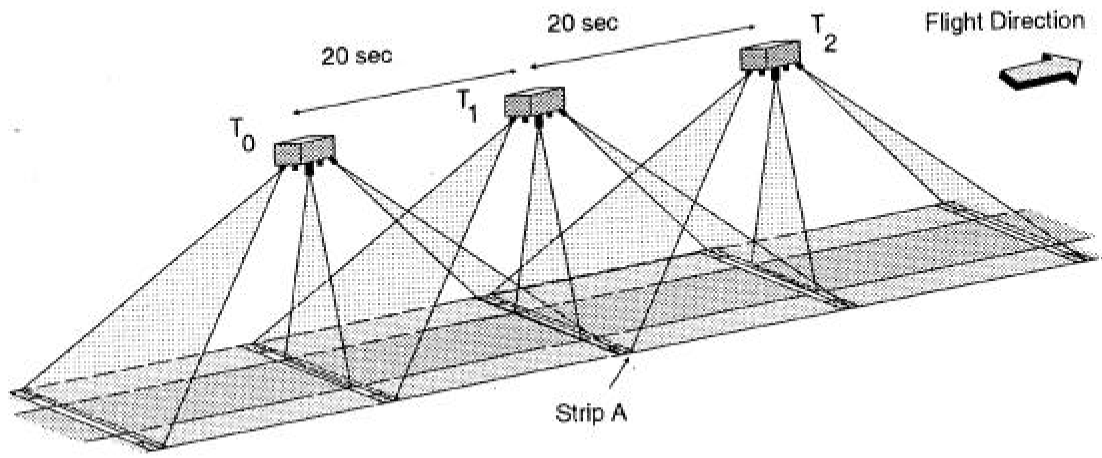


Fig. 6.53 "MOMS-02 stereo geometry (after DARA, 1994)"

Another system is the new HRS (High Resolution Stereoscopic) instrument in the SPOT-5 which has two telescopes looking forward and after in the direction of the orbital trajectory.

The forward-looking telescope captures images at a viewing angle of 20° ahead of the vertical. Ninety minutes later, the after-looking telescope acquires the same ground area at an angle of 20° behind the vertical (Fig. 6.54).

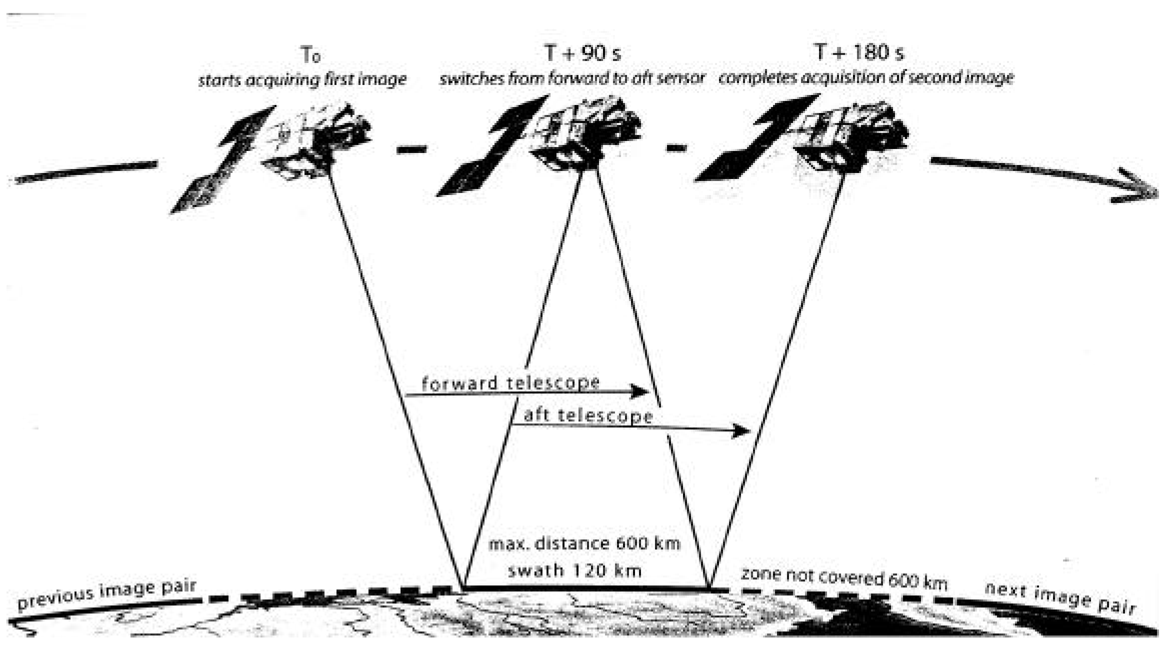


Fig. 6.54 "HRS SPOT-5 stereo geometry (after SPOT IMAGE, 2002)"

- **Radargrammetry**

Radargrammetry is a technique similar to photogrammetry which uses images obtained from the radar signal. A pair of images is acquired and using their correlation a DEM is generated. In this case, it should be noted that the angle of incidence is complementary to that corresponding to an optic image. The absolute precision is of the order of the pixel size. As with stereoscopy, diverse configurations exist (Fig. 6.55).

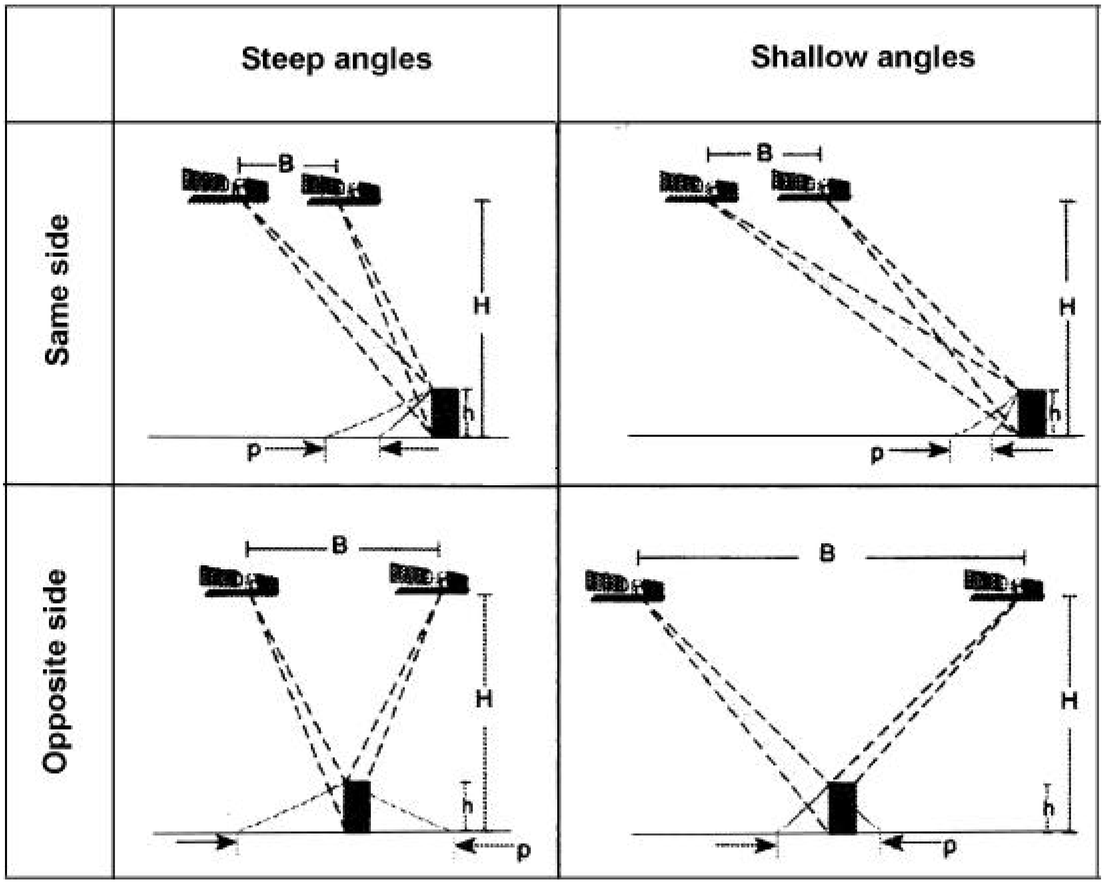


Fig. 6.55 "Diverse SAR stereo configurations (after Toutin, 2001)"

- **Interferometry**

With knowledge of phase of the radar signal, a channel of phase difference and a channel of phase coherence (constant phase angle) can be generated. They are useful in interpreting the interferometric information.

Two scenes, taken during two close passes of the satellite and at a distance called "base", are required. The base should be less than the maximum value, depending on the frequency of the electromagnetic wave and between to 0.5-1 km. The environmental conditions (wind, rain, etc.) should be as similar possible as for both capture occasions.

One of the scenes, the primary, is used as the reference for the calculations. The other, the secondary, together with the primary is used to calculate a channel of phase difference, called an interferogram, and one of coherence, which is an indicator of the degree of dependability of the phase measurements. The phases should then be developed, adopting a resolution and transforming the fringes of the interferogram in terrain level curves.

The problems of this method:

a. The measurement is ambiguous; since the phase difference is know accurately but not the quantity (number) of complete wavelengths to the distance radar-target.

b. The phase depends on the radio-electric characteristics of the target. If they are modified between the scenes, coherence will not be achieved. However, if the fringes of the interferogram can be correctly built, it indicates that the target has remained unalterable.

3.2.8 Cartographic application

Over the last decades the considerable potential of satellite images, especially in the optic domain for cartographic updating, has been realised. According to the ISPRS, the mapping requirements utilising space images, can be divided in three categories:

- Planimetric precision;
- Altimetric precision;
- Detectability (Konecny, 1990).

The more demanding **planimetric precision** is linked with the cartographic resolution of +/- 0.2 mm, which generates requirements for the scales (1/D) the more frequent of which are (Table 6.6):

D	Planimetric Precision
25.000	+/- 5 m
50.000	+/- 10 m
100.000	+/- 20 m
200.000	+/- 40 m

Table 6.6

The **altimetric precision** requirement (**h**), knowing that the equidistance (contour interval "**e**") is e = +/- 5 · h, is shown in the Table 6.7.

E	H
20 m	+/- 4 m
50 m	+/- 10 m
100 m	+/- 20 m

Table 6.7 (Konecny, 1990)

The **detectability** refers to the possibility of detecting objects starting with the digital interpretation of the images. It is a requirement that the object covers at least 1.5 pixels, which creates to the following minimum dimensions for detectable objects (Table 6.8).

Object – Target	Dimension
Urban infrastructure	2 m
Land ways	2 m
Drainage network	5 m
Vial infrastructure	10 m

Table 6.8 (Konecny, 1990)

Taking into consideration the main commercial satellite systems and the most common cartographic scales, the following usage chart can be produced (Table 6.9):

Satellite sensor	**Ground resolution**	**Chart scale**
QUICK BIRD	0.7 m	1/3500
IKONOS	1 – 4 m	1/5000 – 1/20 000
SPOT PAN	10 m	1/50 000
Landsat 7 ETM	15 m	1/75 000
SPOT XS	20 m	1/50 000 – 1/100 000
RADARSAT 1 SAR	8 – 30 m	1/30 000 – 1/100 000
ERS SAR	30 m	1/100 000
Landsat TM	30 m	1/100 000
Landsat MSS	80 m	1/250 000
SAC-C MMRS	175 m	1/875 000

Table 6.9

MAIN PRESENT SATELLITE SYSTEMS AVAILABLE

The following systems list is not exhaustive. It details some of the most regularly used systems for mid-high scale cartography.

Satellite System/ Series **Country**	**Orbit type, Altitude, Recurrent period, Inclination**	**Scene area, Ground resolution, Modes**	**Sensors Spectral bands**
LANDSAT USA	Sun-synchronous 705 km 16 days 98.2°	185 x 185 km MSS: 80 m TM: 30 m ETM+:	Multispectral Scanner (MSS) Band 1: 0.5 – 0.6 µm (green) Band 2: 0.6 – 0.7 µm (red) Band 3: 0.7 – 0.8 µm (near IR) Band 4: 0.8 – 1.1 µm (near IR) Thematic Mapper (TM) Band 1: 0.45 – 0.52 µm (blue) Band 2: 0.52 – 0.60 µm (green) Band 3: 0.63 – 0.69 µm (red) Band 4: 0.76 – 0.90 µm (near IR) Band 5: 1.55 – 1.75 µm (near IR) Band 6: 10.4 – 12.5µ m (thermal IR) Band 7: 2.08 – 2.35 µm (mid IR) Enhanced Thematic Mapper (ETM+)
SPOT France	Sun-synchronous 832 km 23 days 98.7°	60 x 60 km XS: 20 m P: 10 m	High Resolution Visible (HRV) Multi Band (XS) Band 1: 0.49 – 0.59 µm (green) Band 2: 0.61 – 0.68 µm (red) Band 3: 0.79 – 0.89 µm (near IR) Panchromatic (P) 0.51 – 0.73 µm
IRS India	Sun-synchronous 900 - 904 km 22 days 99.5°	148 x 148 km LISS-I: 73 m LISS-II: 36.5 m	Linear Imaging Self Scanning (LISS-I) Band 1: 0.45 – 0.52 µm (blue) Band 2: 0.52 – 0.569 µm (green) Band 3: 0.62 – 0.68 µm (red) Band 4: 0.77 – 0.86 µm (near IR) Linear Imaging Self Scanning (LISS-II) Consists of 2 cameras, same as above, with swath width 74 km per camera (145 km together)

Satellite System/ Series **Country**	**Orbit type, Altitude, Recurrent period, Inclination**	**Scene area, Ground resolution, Modes**	**Sensors Spectral bands**
MOS Japan	Sun-synchronous 909 km 17 days 99°	100 x 90 km MESSR: 50 m	MESSR Band 1: 0.51 – 0.59 µm (green) Band 2: 0.61 – 0.69 µm (red) Band 3: 0.72 – 0.80 µm (near IR) Band 4: 0.80 – 1.10 mµ (near IR)
JERS Japan	Sun-synchronous 568 km 44 days 97.7°	75 x 75 km OPS: 18 x 24 m SAR: 18 x 18 m	Optical Sensor (OPS) Visible and Near Infrared (VNIR) Band 1: 0.52 – 0.60 µm (green) Band 2: 0.63 – 0.69 µm (red) Band 3: 0.76 – 0.86 µm (near IR) Band 4: 0.76 – 0.86 µm (near IR) Short Wave Infrared (SWIR) Band 5: 1.60 – 1.71 µm Band 6: 2.01 – 2.12 µm Band 7: 2.13 – 2.15 µm Synthetic Aperture Radar (SAR) 1.275 GHz (L-Band) HH
ERS Europe	Sun-synchronous 777 km 3 - 35 days 98.5°	AMI works in three modes. In Image mode: 100 x 100 km 30 x 30 m (3 looks)	Active Microwave Instrument (AMI) Synthetic Aperture Radar (SAR) 5.3 GHz (C-Band) VV Incidence angle fix: 23°
RADARSAT Canada	Sun-synchronous 798 km 3 - 24 days 98.6°	Several modes In standard mode: 100 x 100 km 30 x 30 m (3 looks) In fine mode: 50 x 50 km 11 x 8 m (1 look)	Synthetic Aperture Radar (SAR) 5.3 GHz (C-Band) HH Incidence angle selectable: In standard mode: 20° – 50° In fine mode: 37° – 48°

Satellite System/ Series Country	Orbit type, Altitude, Recurrent period, Inclination	Scene area, Ground resolution, Modes	Sensors Spectral bands
IKONOS USA	Sun-synchronous 681 km 1-3 days 98.1°	Several incidence angles Multi-band: 4 m (with nominal angle 26°) Panchromatic:1 m (with nominal angle 26°)	Band 1: 0.45 – 0.52 μm (blue) Band 2: 0.52 – 0.60 μm (green) Band 3: 0.63 – 0.69 μm (red) Band 4: 0.76 – 0.90 μm (near IR) Panchromatic band: 0.45 – 0.90 μm

ACRONYMS

AIRSAR	AIRborne SAR sensor, (J P L)
ASPRS	American Society for Photogrammetry and Remote Sensing
AVHRR	Advanced Very High Resolution Radiometer
CCD	Charge Coupled Device
CCRS	Canadian Centre for Remote Sensing
DGPS	Differential GPS
DLR	German Aerospace Research Establishment
DN	Digital Number
DTM	Digital Terrain Model
EDM	Electronic Distance Measurement
EODM	Electro-Optic Distance Measurement
ERS	European Remote Sensing Satellite
ESA	European Space Agency
ETM	Enhanced Thematic Mapper
GALILEO	European (ESA) Global Positioning Satellite System
GBAS	Ground Based Augmentation System (Reference System for differential satellite positioning)
GCP	Ground Control Point
GIS	Geographic Information System
GLONASS	Global Navigation Satellite System (Russia)
GNSS	Global Navigation Satellite System (GPS + GALILEO + GLONASS)
GPS	Global Positioning System (USA)
HRV	High Resolution Visible
IFOV	Instantaneous Field Of View
IHO	International Hydrographic Organization
IHS	Intensity Hue Saturation
IR	InfraRed
IRS	Indian Remote Sensing satellite
ISPRS	International Society for Photogrammetry and Remote sensing
JERS	Japanese Earth Resources Satellite
JPL	Jet Propulsion Laboratory
KFA 1000	Kosmologisher Fotoapparat with 1000 mm focal length
LASER	Light Amplification by Stimulated Emission of Radiation
Lat	Latitude
Long	Longitude
LUT	Look-Up Table
MSS	MultiSpectral Scanner
NHO	National Hydrographic Office
NOAA	National Oceanic and Atmospheric Administration
pp.	Pages
PAN	Panchromatic
ppm	Part per million (1×10^{-6})
RBV	Return Beam Vidicon
RGB	Red Green Blue
RS	Remote Sensing
RSI	RadarSat International

RTK	Real Time Kinematic (Precise GNSS rapid method)
S-44	Special Publication 44 (IHO Standards for Hydrographic Surveying)
SAR	Synthetic Aperture Radar
SBAS	Satellite Based Augmentation System (Reference system for differential satellite positioning)
SPOT	Satellite Pour l-Observation de la Terre (France)
SSMI	Special Sensor Microwave Imager
TM	Thematic Mapper
USFAA	United States Federal Aviation Association
UTM	Universal Transverse Mercator
VIR	Visible and near InfraRed
WAAS	Wide Area Augmentation System
WGS	World Geodetic System
WGS 84	World Geodetic System 1984
XS	Multispectral

REFERENCES

(List intended to help the Chapter 6 reader in finding more information through printed matter or web pages)

ALBERZ J. KREILING W., (1989).	*"Photogrammetric Guide"*	Wichmann, Karlsruhe (Germany)
ASPRS ,(1980).	*"Manual of Photogrammetry"*	American Society for Photogrammetry and Remote Sensing. Bethesda, Maryland, (USA)
ASPRS ,(1983).	*"Manual of Remote Sensing. 2 Volumes"*	American Society for Photogrammetry and Remote Sensing. Bethesda, Maryland, (USA) The Sheridan Press.
ASPRS ,(1996).	*"Digital Photogrammetry"*	American Society for Photogrammetry and Remote Sensing. Bethesda, Maryland, (USA)
BOMFORD G. (1980).	*"Geodesy 4h Ed"*	Claredon Press, Oxford (UK).
CHUECA PAZOS Et. Al. (1996).	*"Tratado de Topografía (3 Volumes)"*	Paraninfo, Madrid (Spain)
Chuvieco E. **(**1995).	*"Fundamentos de Teledetección Espacial"*	Editorial RIALP, Madrid, Spain, 453 pp.
CNES**, (**2002).	*"HRS puts terrain into perspective"*	SPOT Magazine N°34, 1st semester 2002, pp 10-11.
FEDERAL GEODETIC CONTROL COMMITTEE (1984).	*"Standards and Specification for Geodetic Control Networks"*	NOAA Rockville Maryland (USA)
GDTA**, (**1995).	*"Aspects stéréoscopiques de SPOT, Cahier A2 MNT"*	Les Cahiers Pedagogiques du GDTA, France, 93 pp.
GERMAN SPACE AGENCY, (1994).	*"MOMS-02-D2 Data Catalogue"*	DARA, München, Germany.
HOFMANN WELLENHOF Et.Al. (2001).	*"GPS, Theory and Practice. 5th. Ed "*	Springer, Wien (Austria), New York (USA)

IHO (1998).	*"IHO Standards for Hydrographic Surveys. S-44 4th. Ed "*	IHB, Monaco. There are also French and Spanish versions available
IHO (1994).	*"Hydrographic Dictionary 5th Ed. S–32"*	IHB, Monaco. Also available are Spanish version (1997) and French version (1998).
JOECKEL R., STROBER M. (1995).	*"Elektronische Entfernungs und Richtungsmessung, 3th. Ed",*	Wittwer, Stuttgart (Germany).
KONECNY, G., (1990).	*"Review of the latest technology in satellite mapping. Interim report, Inter-commission Working Group I/IV on International Mapping and Remote Sensing Satellite Systems of ISPRS, Vol.14"*	Hanover, Germany, pp. 11-21.
LANGERAAR W. (1984).	*"Surveying and Charting of the Seas"*	Elsevier. Amsterdam (The Netherlands), Oxford (UK) New York (USA) Tokyo (Japan)
LAURILA S. (1976).	*"Electronic Surveying and Navigation"*	J. Wiley & Sons, New York (USA)
LEICK A. (1995).	*"GPS Satellite Surveying. 2nd. Ed".*	Wiley Chichester, Brisbane. New York (USA) Toronto (Canada) Singapore.
LILLESAND, T.M. and KIEFER, R.W., (1987).	*"Remote sensing and image interpretation, 2nd edition"*	John Wiley and Sons, Inc., New York, 721 p.
MEISENHEIMER D. (1995).	*"Vermessungsinstrumente Aktuell"*	Wittwer, Stuttgart (Germany)
NASA, (1997).	*"The Remote Sensing Tutorial".*	Goddard Space Flight Centre, NASA Web Production. Written by: Nicholas M. Short, Sr.
OLLIVER F. (1995).	*"Instruments Topographiques"*	Eyrolles, Paris (France)
OLIVER C. and S. QUEGAN (1998).	*"Understanding Synthetic Aperture Radar Images"*	Artech House, Norwood, Massachuset (USA)
POHL, C., (1996).	*"Geometric aspects of multi-sensor image fusion for topographic map updating in the humid Tropics"*	ITC Publication Number 39, The Netherlands, 214 pp.
RANEY, R.K., (1992).	*"Course notes; unpublished notes"*	Canada Centre for Remote Sensing, Ottawa, Canada.

RICHARDUS P. (1977).	*"Project Surveying".*	Balkema. The Netherlands.
SEEBER G. (1993).	*"Satellite Geodesy"*	W. de Gruyter Berlin (Germany) New York (USA)
SEEBER G. (2003).	*"Satellite Geodesy 2nd. Ed "*	Walter de Gruyter (Berlin - NY)
TORGE W. (2001).	*"Geodesy"*	W. de Gruyter Berlin (Germany) New York (USA)
TURNBULL D. (2001).	*"The Evolution of an Object - Oriented Geospatial Information System Supporting Digital Nautical Chart Maintenance at the NIMA"*	Bulletins Hydr. Int. Jul. Aug. - Sep; IHO, Monaco
TOUTIN, Th., (1998).	*"Evaluation de la précision géométrique des images de RADARSAT"*	Journal Canadien de télédétection, 23(1):80-88.
TOUTIN, Th., (1997).	*"Single versus stereo ERS-1 SAR imagery for planimetric feature extraction"*	International Journal of Remote Sensing, 18(18):3909-3914.
TOUTIN, Th. and B. RIVARD, (1997).	*"Value-added RADARSAT Products for Geoscientific Applications"*	Canadian Journal of Remote Sensing, 23(1):63-70.
TOUTIN, Th., (1995).	*"Generating DEM from stereo images with a photogrammetric approach: Examples with VIR and SAR data"*	EARSeL Journal Advances in Remote Sensing, 4(2):110-117.
WOLF R., BRINKER R.C. (1994).	*"Elementary Surveying 9th. Ed. "*	Harper Collins College Publishers New York (USA) There is available also a Spanish version "Topografía", Alfaomega, México (1998)

URL ADDRESSES

COUNTRY	INSTITUTION	WEB SITE
	European Space Agency	http://www.esa.int
	International Society on Photogrammetry and Remote Sensing	http://www.isprs.org
	Fédération Internationale de Géometres	http://www.Fig.net
	International Association of Geodesy	http://www.gfy.ku.dk/iag/
Argentina	Comisión Nacional de Actividades Espaciales	http://www.conae,gov.ar
Austr. – N.Z.	Australian - New Zealand Land Information Council	http://www.anzlic.org.au
Australia	Commonwealth Scientific & Industrial Research Organization	http://www.csiro.au
Australia	Surveying and Land Information Group	http://www.auslig.gov.au
Bolivia	Centro de Levantamientos Aeroespaciales y SIG	http://www.clas.unmss.edu.bo
Brazil	Instituto Nacional de Pesquisas Espaciais	http://www.inpe.br
Canada	Centre For Remote Sensing	http://www.ccrs.nrcan.ca
Canada	Radarsat International	http://www.rsi.ca
Canada Chile	Geodetic Survey Agencia Chilena del Espacio	http://www.geod.emr.ca http://www.agenciaespacial.cl
China	China Academy of Space Technology	http://fas.org/nuke/guide/china/contractor/cast.htm
France	Centre National d'Etudes Spatiales	http://www.cnes
France	Group pour le Développement de la Téledetection Aerospatiale	http://www.gdta.fr
Germany	Institute für Erdmessung, Hanover University	http://www.ife.unihannover.de
Germany	Institute für Angewandte Geodäsie	http://www.gibs.leipzig.ifag
Germany	Karlsruhe University	http://www.ipfr.bau.verm.uni.karlsruhe.de
Germany	GPS Information Bulletin Board System	http://www.gibs.leipzig.ifag.de

COUNTRY	INSTITUTION	WEB SITE
Germany	Deutsches Zentrum für Luft und Raumfahrt	http://www.dlr.de
India	Indian Space Research Organization	http://www.isro.org
Italy	Agenzia Spaziale Italiana	http://www.asi.it
Japan	National Space Development Agency	http://www.nasda.go.jp
Russia	Russian Space Science Internet	http://www.rssi.ru
Spain	Instituto Nacional de Técnica Aeroespacial	http://www.inta.es
Spain	Org. for Cartography and Geodesy	http://www.cartesia.org
Spain	Valencia University	http://www.miranda.tel.uva.es
Switzerland	Astronomical Institute Berne University	http://www.aiub.unike.ch
UK	Nottingham University	http://www.ccc.nottingham.ac.uk
UK	British National Space Centre	http://www.bnsc.uk
USA	Ohio State University (Centre for Mapping)	http://www.cfm.ohio.state
USA	Maine University	http://www.spatial.maine.edu
USA	Geological Survey (EROS)	http://edc.usgs.gov
USA	Earth Observation Handbook	http://www.eohandbook.com
USA	Goddard Space Flight Centre (NASA)	http://www.gsfc.nasa.gov
USA	Nat. Ocean. And Atm. Adm. Central Library	http://www.lib.noaa
USA	Nat. Aeronautic and Space Adm.	http://www.nasa.gov
USA	National Oceanic and Atmospheric Adm.	http://www.noaa.gov
USA USA	Geological Survey Professional Survey (review)	http://www.usgs.gov http://www.profsurvey.com
USA	Department of Defence	http://www.defenselink.mil
USA	National Geodetic Survey	http://www.ngs.noaa.gov
USA	Institute of Navigation	http://www.ion.org

COUNTRY	INSTITUTION	WEB SITE
USA	Jet Propulsion Laboratory	http://www.jpl.nasa.gov
USA	Naval Observatory	http://www.usno.navy.mil
USA	GPS Interface Control Document	http://www.navcen.usc.mil/gps
USA	Interagency GPS Executive Board	http://www.igeb.gov
USA	Texas University	http://www.host.cc.utean.edu
USA	GPS Nav. Inf.	http://www.navan.uscg.mil/gps
USA	California - Los Angeles University	http://www.cla.esc.edu
USA	American Society for Photogr. and R.S.	http://www.asprs.org
USA	National Imagery and Mapping Agency	http://www.164.214.2.59
USA	GPS issues	http://www.206.65.196

BIBLIOGRAPHY

(Edited or digital source information used in the preparation of Chapter 6).

ASPRS, (1983).	*"Manual of Remote Sensing"*	American Society of Photogrammetry and Remote Sensing. 2 volumes. The Sheridan Press, USA, 2420 pp.
CHUVIECO E. (1995).	*"Fundamentos de Teledetección Espacial"*	Editorial RIALP, Madrid, Spain, 453 pp.
CURAN P.J. (1985).	*"Principles of remote sensing"*	Longman, London, England.
CURLANDER J.C. and R.N. MCDONOUGH, (1991).	*"Synthetic Aperture Radar Systems and Signal Processing"*	John Wiley and Sons, Inc., Toronto
DRURY S.A., (1990).	*"A Guide to Remote Sensing"*	Oxford Science Publications, Oxford, USA, 199 pp.
ELACHI C. and F.T. ULABY, (1990).	*"Radar Polarimetry for Geoscience Applications"*	Artech House, Boston
ELACHI C. (1988).	*"Spaceborne Radar Remote Sensing: Applications and Techniques"*	IEEE Press, New York
FAO, (1990).	*"Remote sensing applications to land resources"*	FAO RSC Series 54, Rome, Italy.
FITCH J.P. (1988).	*"Synthetic Aperture Radar"*	Springer-Verlag, New York
HENDERSON F.M. and A.J. LEWIS, EDS. (1998).	*"Principles and Applications of Imaging Radar, Manual of Remote Sensing, Third Edition, Volume 2"*	John Wiley & Sons, Inc., Toronto
KNEISSL M. (1956).	*"Handbuch der Vermessungskunde Band III (Hohenmessung, Tachymetrie)"*	Metzer, Stuttgart (Germany)
KNEISSL M. (1958).	*"Handbuch der Vermessungskunde Band IV (Mathematische Geodäsie)"*	Metzer, Stuttgart (Germany)
KNEISSL M. (1963).	*"Handbuch der Vermessungskunde Band II (Feld und Land Messung, Abstekungsarbeiten)"*	Metzer, Stuttgart (Germany)
MAGUIRE D. et al (1991).	*Geographic Information System Principles and Applications"*	John Wiley & Sons N.Y.
NASA, (1997).	*"The Remote Sensing Tutorial"*	Goddard Space Flight Centre, NASA Web Production. Written by: Nicholas M. Short, Sr.

OLIVER C. and S. QUEGAN (1998).	*“Understanding Synthetic Aperture Radar Images”*	Artech House, Norwood, Mass.
RINNER K., BENZ F. (1966)	*“Handbuch der Vermessungskunde Band VI (Die Entfernunsmessung nit Elektromagnetische Wellen und ihre geodätische Anwendung)”*	Metzer, Stuttgart (Germany)
RINNER K., BENZ F. (1971).	*“Handbuch der Vermessungskunde Band III a,3 Volumes (Photogrammetrie)”*	Metzer, Stuttgart (Germany)
RUSSELL - WOLF (1984).	*“Elementary Surveying”*	Harper and Row Publishers, New York (USA)

The following texts of the REFERENCE LIST were also used:

ALBERZ J. KREILING W (1989)
ASPRS (1996)
BOMFORD G. (1980)
CHUECA PAZOS Et Al (1996)
HOFMANN WELLENHOF Et Al (2001)
IHO (1998)
IHO (1994)
LANGERAAR W. (1984)
MEISENHEIMER D. (1995)
SEEBER G. (1993)
TORGE W. (2001)
WOLF R, BRINKER R.C. (1994)

CHAPTER 6 – ANNEX A
ALGORITHMS FOR THE TRANSVERSE MERCATOR REPRESENTATION

1. PRESENTATION

The Mercator Transverse Representation (see chapter 2, 2.5.4 and 2.5.5) is a useful medium in which to transfer geodetic co-ordinates (latitude, longitude) to the plane. The use of plane co-ordinates (**x** & **y** or **N** & **E**) with small corrections related to the measured distances and angles is suitable for topographic survey purposes and also for some detailed hydrographic surveys.

2. GEOMETRIC GEODESY AND MATHEMATICAL CARTOGRAPHIC CONCEPTS

Before the study of these concepts, the reader should be familiar with Chapter 2: 2.4 and 2.5, with particular attention to subparagraphs 2.5.4 and 2.5.5.

Taking the earth rotation ellipsoid as the reference surface, with '**a**' semi-major (equatorial) axis and '**b**' semi-minor (polar) axis, it is possible to define:

$$f = \frac{a - b}{a}$$ (flattening, also described as "a" in chapter 2: 2.2.3)

$$\varepsilon = \frac{\sqrt{a^2 - b^2}}{a}$$ (first excentricity, also described as "**e**" in chapter 2: 2.1.1))

$$\varepsilon' = \frac{\sqrt{a^2 - b^2}}{b}$$ (2nd. excentricity)

with elemental algebraic procedures it is feasible to verify the following relationships:

$$f(2 - f) = \varepsilon^2$$

$$(1 - f)^2 = 1 - \varepsilon^2$$

with the described constants, the computation of the curvature radius and arcs of lines on the ellipsoidal surface are possible:

$$M = a(1 - f)^2\left[1 - f(2 - f)\,\mathrm{sen}^2\varphi\right]^{-3/2} = a(1 - \varepsilon^2)\left[1 - \varepsilon^2\,\mathrm{sen}^2\varphi\right]^{-3/2}$$

$$N = a\left[1 - f(2 - f)\,\mathrm{sen}^2\varphi\right]^{-1/2} = a\left[1 - \varepsilon^2\,\mathrm{sen}^2\varphi\right]^{-1/2}$$

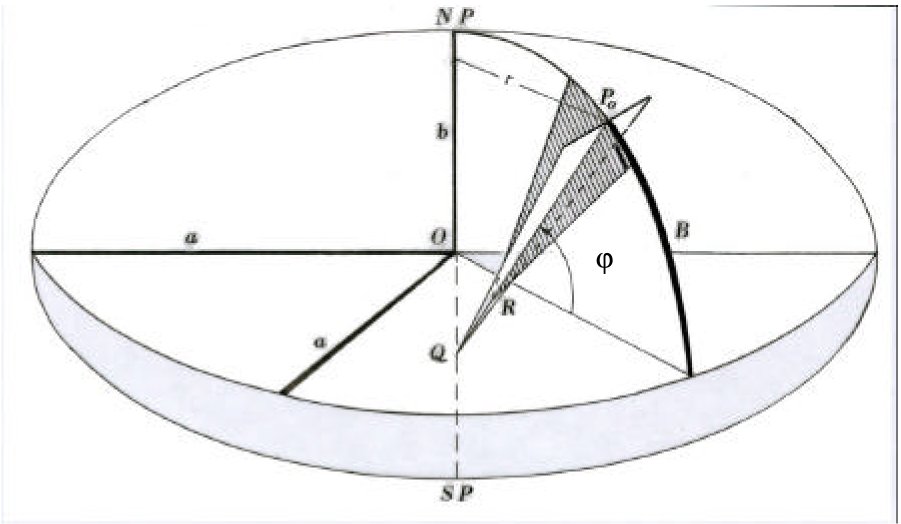
$$r = N\cos\varphi$$

$$\Delta p_{12} = r(\lambda_2 - \lambda_1)$$

$$B = \int_0^{\varphi} M d\varphi = \alpha\varphi + \beta \operatorname{sen} 2\varphi + \gamma \operatorname{sen} 4\varphi + \delta \operatorname{sen} 6\varphi + \dots$$

where (see Fig 6A.1):

M - Meridian curvature radius
N - Normal (to the meridian) section curvature radius
R - Parallel curvature radius
φ - Geodetic (ellipsoidal) latitude
Δp_{12} - Parallel are between longitudes λ_1 and λ_2 at φ latitude, (λ_2 - λ_1) expressed in radians
B - Meridian arc from equator to the φ latitude (for the first term = $\alpha\varphi$, expressed in radians)
$\alpha = a\,(1 - 1/4\,\varepsilon^2 - 3/64\,\varepsilon^4 - 5/256\,\varepsilon^6) = a\,(1-f)^2\,(1 + 3/2\,f + 33/16\,f^2 + 85/32\,f^3)$
$\beta = -a\,(3/8\,\varepsilon^2 + 3/32\,\varepsilon^4 + 45/1024\,\varepsilon^6) = -a/2(1-f)^2\,(3/2\,f + 3\,f^2 + 285/64 f^3)$
$\gamma = a\,(5/256\,\varepsilon^4 + 45/1024\,\varepsilon^6) = a/4(1-f)^2\,(15/16\,f^2 + 75/32\,f^3)$
$\delta = -a\,(35/3072\,\varepsilon^6) = -a/6(1-f)^2\,(35/64\,f^3)$



Po O = N
Po R = M
N ≥ M

Fig. 6A.1

The following table contains the described constants for two commonly used ellipsoids with the Q values (meridian arc, B, from equator to the pole) added

$$Q = \int_0^{\pi/2} M d\varphi$$

ELLIPSOID	MADRID 1924	WGS 84
A	6378388 m	6378137 m
F	1/297	1/298.2572236
$\varepsilon^2 = f(2 - f)$ α	0.0067226722 6367654.500 m	0.0066943800 6367449.146 m
β	-16107.035 m	-16038.509 m
γ	+ 16.976 m	+ 16.833 m
δ	- 0.022 m	- 0.022 m
Q	10002288.30 m	10001965.73 m

The mathematical form to generate a representation of the ellipsoid on the plane is:

$$x = x(\varphi, \lambda)$$
$$y = y(\varphi, \lambda)$$

and these formulae provide the properties for this transformation. For a conformal or othomorphic representation it is necessary to replace the latitude with a new variable called "isometric latitude" or

$$q = \int_0^\varphi \frac{M}{N \cos\varphi} d\varphi$$

"meridional part"

The origin of this function is the MERCATOR representation of the earth ellipsoid on the plane, starting from a circular cylinder whose axis of orientation coincides with the semi-minor axis **'b'** of the rotation ellipse and the surface tangent to its respective equator (See Fig. 6A.2)

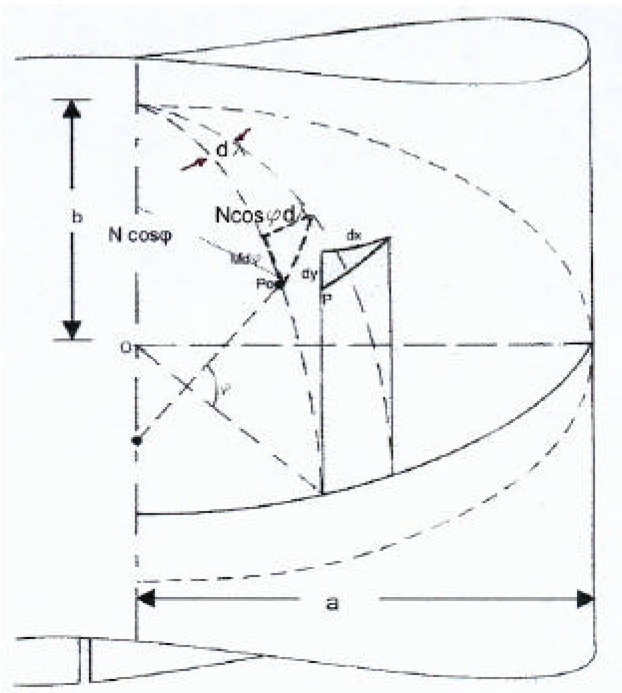


Fig. 6A.2

Taking the '**y**' axis with the projection of the longitudinal origin meridian ($\lambda = 0$) on the cylinder, with y = 0 for $\varphi = 0$ and the '**x**' axis representing the equator, with x = 0 for $\lambda = 0$, it is possible to show:

$$\mathbf{x = a}\ l$$

(isometry on the tangent line = equator) but, in this case, the '**y**' should satisfy the following differential rate (see Fig. 6A.2)

$$\frac{dy}{Md\varphi} = \frac{dx}{N\cos\varphi d\lambda} = m$$

where **m** is coincident with $\mathbf{m_l}$ given in 2.4, chapter 2. Also:

$$\frac{dy}{Md\varphi} = \frac{ad\lambda}{N\cos\varphi d\lambda} = \frac{a}{N\cos\varphi} = m$$

and

$$y = a\int_0^{\varphi} \frac{M}{N\cos\varphi} d\varphi = aq$$

solving the integral it is possible to express:

$$q = \ln\left[\left(\frac{1-\varepsilon\,\mathrm{sen}\,\varphi}{1+\varepsilon\,\mathrm{sen}\,\varphi}\right)^{\varepsilon/2} \mathrm{tg}\left(\frac{\pi}{4}+\frac{\varphi}{2}\right)\right]$$

Fig. 6A.3 shows a partial representation of the meridians and parallels grid and also of a geodetic line (minimum distance track over the ellipsoidal surface) among points A and B for this transformation (x = $a\lambda$, y = aq)

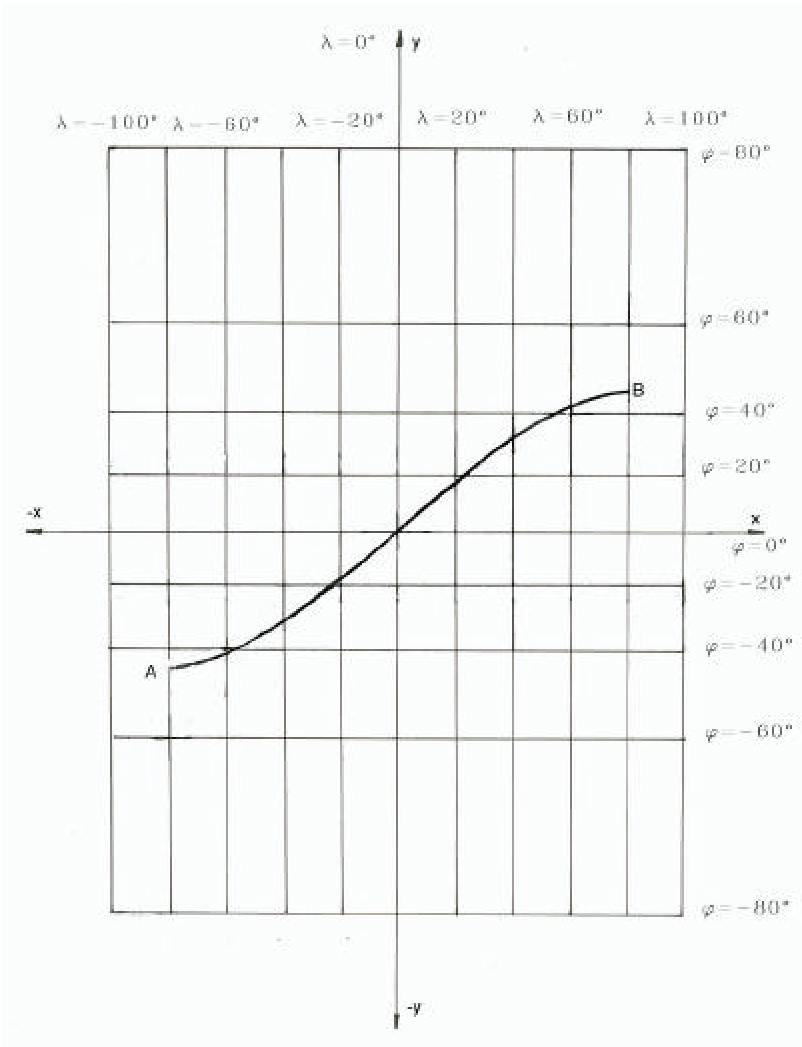


Fig. 6A.3

Algorithms based in these principals, but with other assumptions, are useful for nautical charting, but for further deliberations in this annex it is sufficient to remember that:

$$x = a\lambda$$

$$y = aq$$

$$q = \int_0^{\varphi} \frac{M}{N\cos\varphi} d\varphi$$

is a conform transform of the ellipsoid to the plane, in conclusion the principals of the analytical functions is applicable:

$$\mathbf{y + ix = f(q + i\,l)} \qquad \mathbf{(2.1)}$$

(taking **y** to the north and **x** to de east) where i = $(-1)^{1/2}$ and the Cauchy - Riemann conditions should be satisfied:

$$\frac{\partial y}{\partial q} = \frac{\partial x}{\partial \lambda}$$

$$\frac{\partial x}{\partial q} = -\frac{\partial y}{\partial \lambda}$$

this is possible because **q**, l and **x**, **y** are two plane co-ordinate pairs.

For a better comprehension of this subject, consultation of a mathematical text in complex variables and their application to conform transfer between two plane domains is recommended.

The general relation 2.1, the Cauchy - Riemann conditions and the following considerations are valid for all conformal transformations (not only for the described Mercator expressions).

Other mathematical formulae for the generic conformal representation come from differential expressions of x = x (φ, λ) and y = y (φ, λ):

$$dx = \left(\frac{\partial x}{\partial \varphi}\right) d\varphi + \left(\frac{\partial x}{\lambda}\right) d\lambda$$

$$dy = \left(\frac{\partial y}{\partial \varphi}\right) d\varphi + \left(\frac{\partial y}{\partial \lambda}\right) d\lambda$$

when φ = constant (parallel arc), the square of the plane differential distance $dx^2 + dy^2$ gives, with the correspondent ellipsoidal element (N cos φ dλ) the square of linear deformation rate:

$$m^2 = \frac{\left(\frac{\partial x}{\partial \lambda}\right)^2 + \left(\frac{\partial y}{\partial \lambda}\right)^2}{N^2 \cos^2 \varphi}$$

and also for λ = constant and elementary meridian arc M df , issues:

$$m^2 = \frac{\left(\frac{\partial x}{\partial \varphi}\right)^2 + \left(\frac{\partial y}{\partial \varphi}\right)^2}{M^2}$$

is also valid:

$$m^2 = \frac{\left(\frac{\partial x}{\partial \varphi}\right)^2 + \left(\frac{\partial y}{\partial \varphi}\right)^2}{M^2} = \frac{\left(\frac{\partial x}{\partial \lambda}\right)^2 + \left(\frac{\partial y}{\partial \lambda}\right)^2}{N^2 \cos^2\varphi} \qquad \textbf{(2.2)}$$

Arising from the same differential expressions and taking the rates:

$$\frac{dx}{dy} \qquad \text{(for } \varphi = \text{constant)}$$

$$\frac{dy}{dx} \qquad \text{(for } \lambda = \text{constant)}$$

it is possible to obtain the grid declination 'g' formulas (g is the angle between the cartesian axis and the lines of the respective meridians and parallels representation -see Fig. 6A.4)

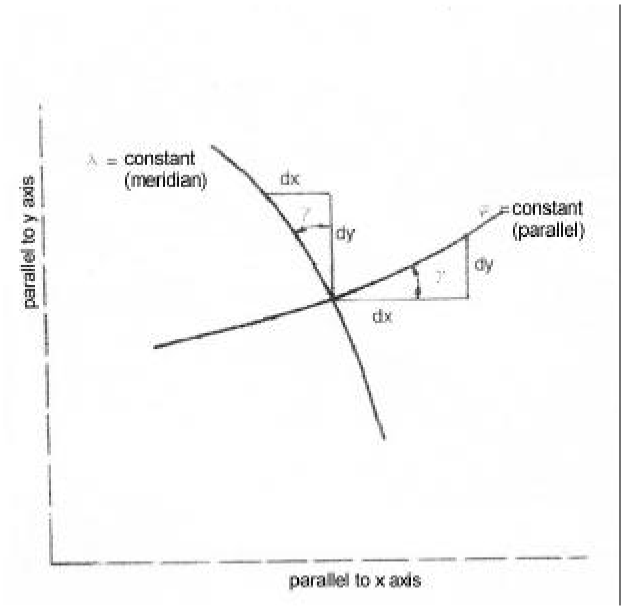


Fig. 6A.4

$$\operatorname{tg}\gamma = \frac{\left(\dfrac{\partial x}{\partial \varphi}\right)}{\left(\dfrac{\partial y}{\partial \varphi}\right)} = \frac{\left(\dfrac{\partial y}{\partial \lambda}\right)}{\left(\dfrac{\partial x}{\partial \lambda}\right)} \qquad \textbf{(2.3)}$$

in these formulae the sign of g (or tgγ) is not considered.

3. GAUSS-KRÜGER FORMULAS

To start in the development of conformal representation with minimum deformation upon a NORTH - SOUTH strip, an elliptical cylinder tangent to a central meridian will be assumed (see Fig. 6A.5).

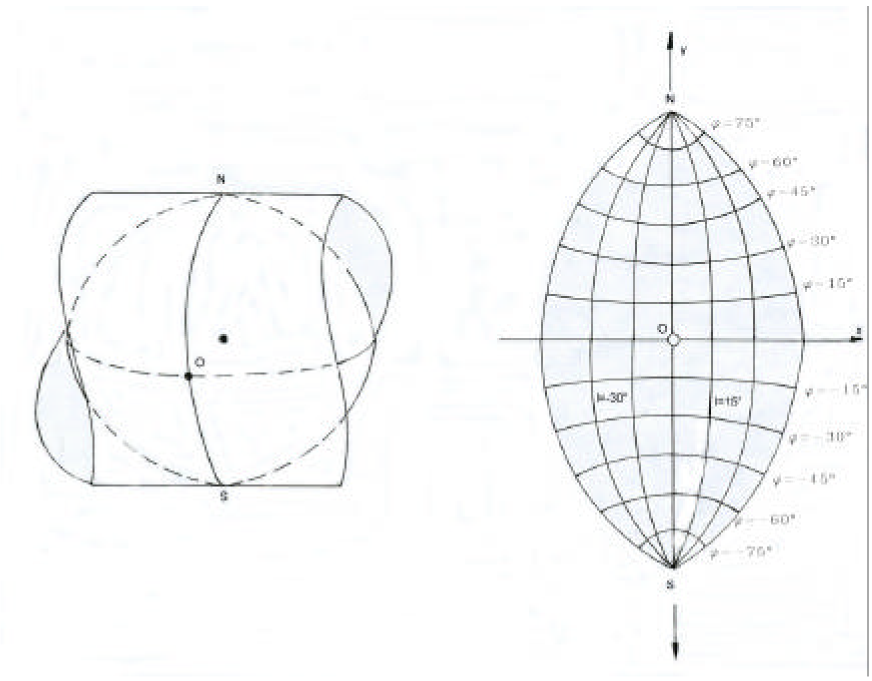


Fig. 6A.5

A more extensive representation of the meridians and parallels grid is given in Fig. 2.6 (2.5.4 Chapter 2), but for the following consideration we start from the former Fig. 6A.5.

For this case, the 2.1 formula will be transformed as follow:

$$\mathbf{f(q + il) = y + ix}$$

where **l** is the longitude referred to the central meridian:

$$\mathbf{l} = l - l_0 \qquad \textbf{(2.4)}$$

Adopting a Taylor series development of the function issues:

$$f(q+il) = f(q) + \frac{df}{dq}(il) + \frac{d^2f}{dq^2}\frac{(il)^2}{2!} + \frac{d^3f}{dq^3}\frac{(il)^3}{3!} + \frac{d^4f}{dq^4}\frac{(il)^4}{4!} + \ldots$$

and separating the real and imaginary parts produces the generic expressions for the conformal

$$x = f(q) - \left(\frac{d^2f}{dq^2}\right)\frac{l^2}{2} + \left(\frac{d^4f}{dq^4}\right)\frac{l^4}{24} + \ldots$$

$$\mathbf{y} = \left(\frac{df}{dq}\right)\cdot l + \left(\frac{d^3f}{dq^3}\right)\frac{l^3}{6} + \ldots$$

representation of the described strip

Taking the equidistance along the central meridian (l=0) it is correct to take:

$$y(l=0) = f(q) = B = \int_0^{\varphi} M d\varphi$$

and then:

$$\frac{df}{d\varphi} = M$$

also, remembering that:

$$q = \int_0^{\varphi} \frac{M}{N\cos\varphi} d\varphi$$

it is possible to obtain:

$$\frac{d\varphi}{dq} = \frac{N\cos\varphi}{M}$$

and also:

$$\left(\frac{df}{dq}\right) = \left(\frac{df}{d\varphi}\right)\left(\frac{d\varphi}{dq}\right) = N\cos\varphi$$

Coming from these principals it is possible to obtain the following derivates

$$y = B + \frac{N \operatorname{sen}\varphi \cos\varphi}{2} l^2 + \frac{N \operatorname{sen}\varphi \cos^3\varphi}{24}\left(5 - tg^2\varphi + 9\eta^2 + 4\eta^4\right)\cdot l^2 + \ldots$$

$$x = N\cos\varphi l + \frac{N\cos^3\varphi}{6}\left(1 - tg^2\varphi + \eta^2\right)\cdot l^3 + \ldots$$

(2.5)

Where **B** and **N** are given in the formulae at beginning of 2 (in this Annex) and h² is:

$$\eta^2 = \varepsilon'^2 \cos^2\varphi = \frac{f(2-f)}{1-f(2-f)}\cos^2\varphi$$

l is given by (2.4) and for its application in (2.5) should be expressed in radians.

With the consideration of (2.2), (2.3) and (2.4) comes also:

$$\gamma = \operatorname{sen}\varphi l + \ldots$$

(2.6)

$$m = 1 + \frac{\cos^2\varphi(1+\eta^2)}{2} l^2 + \ldots = 1 + \frac{x^2}{2R^2} + \ldots$$

Where:

$$R = \sqrt{MN} = \frac{a(1-f)}{[1 - f(2-f)\operatorname{sen}^2\varphi]}$$

(**R** is the best spheric al radius suitable to the ellipsoid at j latitude)

For the inverse computation (to obtain j , l arising from **x**, **y**) the following formulas are useful:

$$\varphi = \varphi_1 - \frac{tg^2\varphi_1}{2}\left(\frac{x^2}{M_1N_1}\right) + \frac{tg\,\varphi_1}{24}\left(5 + 3\,tg^2\varphi_1 + \eta_1^2 - 9\eta_1^2\,tg^2\varphi_1\right)\left(\frac{x^4}{M_1N_1^3}\right) + \ldots$$

(2.7)

$$l = \frac{x}{N_1\cos^2\varphi_1} - \left(\frac{1 + 2\,tg^2\varphi_1 + \eta_1^2}{6\cos\varphi_1}\right)\left(\frac{x}{N_1}\right)^3 + \ldots$$

$$\lambda = \lambda_0 + l$$

where j_1 is the latitude that made possible $B(\varphi) = y$

The representation with these algorithms was applied by GAUSS at beginning of the XIX century for the HANNOVER Kingdom. 100 years after this, Dr. L. KRÜGER made an explanatory analysis and extension of the expressions applying several strips to Germany. Also similar criteria were extended to other countries.

With the tangent cylinder, where m = 1 by the central meridian of the strip, the zone width should be smaller than 200 km each side because in this case the linear deformation rate ($m = 1 + x^2/2R^2 +...$) overcomes the 1.0005 value, this is 0.5 m in 1 km.

With this limitation, the use of these plane co-ordinates is very convenient for topographic charting purposes and also for many control networks computations. For better results, a correction to the measured elements should be made (see 2.2.5 a in Chapter 6).

4. GENERAL TRANSVERSES MERCATOR REPRESENTATIONS (See 2.5.4 and 2.5. at Chapter 2)

The Gauss-Krüger representation, after World War 2, was also called TRANSVERSE MERCATOR and it was utilised increasingly in many countries. For this reason, several constants and coefficients were adopted; for **N** (North co-ordinate) and **E** (East co-ordinate) are valid:

$$N = Y_0 + Ky$$

$$E = X_0 + Kx$$

and, consequently **(2.8)**

$$m = K\left(1 + \frac{x^2}{2R^2} + ...\right)$$

K is a coefficient (below 1) to reduce the linear deformation rate and allows the zone width extension (i.e. 300 km each side of central meridian), particularly for topographic charting in scales smaller than 1:100000 (1:200000 ...)

Y_0: is called FALSE NORTHING,
X_0: FALSE EASTING
K: SCALE FACTOR AT CENTRAL MERIDIAN

the application of **K** coefficient makes on the central meridian, it appear as a negative linear deformation, i.e. for K = 0.9998, the ellipsoidal distances contracts 20 cm/km and the isometric lines are transferred in two parallel lines to the described meridian image. This tangent surface is replaced by a secant elliptical cylinder.

For the U.S. world wide Universal Transverse Mercator Grid System (UTM) the following values are adopted:

K = 0.9996
Y_0 = 0 or 10000000 (for North or South Hemisphere respectively)
X_0= 500000 for each central meridia n

and the zones are distributed at 6° longitude intervals, according to the following table

ZONE (Z)	CENTRAL MERIDIAN (LONGITUDE)	APPLICATION RANGE (LONGITUDE)
31	3°	0° to 6
32	9°	6° to 12°
.	.	.
.	.	.
.	.	.
50	117°	114° to 120°
.	.	.
.	.	.
.	.	.
60	177°	174° to 180°
1	183° (-177°)	180° to 186° (-174°)
.	.	.
.	.	.
.	.	.
20	297°(-63°)	294° (-66°)to 300° (-60°)
.	.	.
.	.	.
.	.	.
29	351° (-90°)	348° (-12°) to 354° (-6°)
30	357° (-3°)	354° (-6°) to 0°

The Zone number (Z) may be computed starting from the central meridian longitude with the following formulae:

$$Z = 30 + \frac{CM + 3}{6}$$ (East Hemisphere)

$$Z = \frac{183 + CM}{6}$$ (West Hemisphere)

in the second formula (West), the negative value of the longitude (of the central meridian) should be taken.

There are many software programs to solve the calculation of Transverse Mercator transformation, with the algorithms described in 3 and 4. of this Annex, or other alternative mode. However it is desirable to have good subject knowledge of the nature of the linear deformation rate.

In many cases it is suitable to select the general Transverse Mercator scheme adopted for the country representation but in particular cases there are chances to select the better plane representation procedure. For this purpose, it is important to remember that the Transverse Mercator representation is particularly suitable for application in a North - South strip, where the West - East width is less than 400 km (200 km to each side of the central meridian).

After the selection of the central meridian, taking into account the reduction of any distance for the selected area to this line, there is still the possibility of choosing a **K** coefficient (see formulas 2.8) for a better linear deformation rates distribution in the whole representation domain.

For topographic purposes, included coastal delineation, aids to navigation positioning, inshore feature descriptions and special harbour surveys, a linear deformation rate under 0.2m/km is better; this is '**m**' values between 0.9998 and 1.0002.

CHAPTER 6 - ANNEX B
(COMMERCIAL EQUIPMENT EXAMPLES)

1. INTRODUCTION

In this Annex, is list of addresses, telephone numbers and web sites of some of the providers of equipment used in topographic and remote imagery surveying.

However the analysis of commercial booklets or web pages is the recommended way of keeping abreast of the available technology, price information and future product launches.

2. ADDRESSES AND WEB PAGES

Below is a list set of companies or institutions that provide equipment, products and services related to the subjects covered in the CHAPTER VI, it is by no means exhaustive and it is hoped that in future versions of this manual a more comprehensive directory may be created.

INSTITUTION NAME	EQUIPMENT, PRODUCTS OR SERVICES	ADDRESS/COUNTRY	WEB SITE or E-MAIL
AGFA GEVAERT	Photogrammetric Films	B-2640 Mortsel BELGIUM	www.agfa.com
ANEBA, Geoinformatica	Topographic software (CARTOMAP)	Nicaragua 48. 2°, 6° 08029 - Barcelona SPAIN	www.aneba.com
ASAHI PRECISION	Theodolites, Levels, Total Stations (Pentax)	2-5-2 Higashi Oizumi Nerima-ku, Tokyo JAPAN	www.pentax.co.jp
CLARK LABS	Cartographic Software, GIS (IDRISI)	Clark University 950 Main Street Worcester, MA 01610-1477 USA	http://www.clarklabs.org
EARTH RESOURCE MAPPING (ERMAPPER) ERDAS	imagery products and software for GIS database Images Processing software	4370 La Jolla Village Drive suite 900 San Diego CA USA USA	www.ermapper.com www.earthetc.com www.esdas.com
ESRI	GIS (ArcInfo, ArcView)	USA	www.esri.com info@esri.com
EURIMAGE	Imagery Products	Viale e. D'Onofrio 212, 00155 Rome, Italy	
GARMIN Int.	GPS Navigators	1200E 151 st., Street Olathe, KS 66062 – USA	www.garmin.com

INSTITUTION NAME	EQUIPMENT, PRODUCTS OR SERVICES	ADDRESS/COUNTRY	WEB SITE or E-MAIL
GEOMATECH	Geomatics, GIS and cartography services and assistance	2, rue Philippe Lebon, BP 102, 44612, Saint Nazaire, FRANCE	geomatech@wanadoo.fr
GODDARD SPACE FLIGHT CENTER	Remote Sensing Assistance	USA	http://www.gsfc.nasa-gov
Hewlett-Packard	Hardware	USA	www.hp.com
Institute Cartografic de Catalunya	Cartographic, geomatics, photogrammetric and remote sensing services	Parc de Montjuic s/n, 08038 Barcelona, España	www.icc.es
INTERGRAPH CORPORATION	Soft/Hardware and images for Cartographic processing	P.O. Box 6695 Mailstop MD IW17A2 Huntsville Al 35894-6695 USA	http://imgs.intergraph.com www.intergraph.com
ISM Europe S.A.	Photogrammetric software and hardware, and services.	Passeig de Fabra i Piug 46, 08030, Barcelona, ESPAÑA	sales@ismeurope.com www.ismeurope.com
ITC	Photogrammetric and Cartographic assistance	Hengelostraat 99 P.O BOX 6 THE NETHERLANDS	www.itc.nl ilwis@itc.nl
KODAK, GR-OUPE ALTA	Films for Photogrammetry GIS, Cartography, Remote processing	Hant Monts Inc 3645, Boulevard Sainte-Anne Beauport (Quebec) CANADA G1E3L1	www.kodak.com www.mb-gepair.com www.groupealta.com
LEICA GEOSISTEMS AC	Total Stations Levels, Theodolites, GNSS, Photogramm. Cameras, Stereo plotters, Scanners	CH.9425 Heerbrugg SWITZERLAND	www.leica-geosistems.com
MAPINFO	Software for Cartography, Photogrammetry and GIS	USA	www.mapinfo.com
MicroImage, Inc	Software, Image processing TNT MIPS software	11th. Floor, The Sharp Tower 206 south 13th street Lincoln. NE 68508-2010 USA	www.microimages.com
OMNISTAR, INC.	Worldwide (satellite based) DGPS Service	8200 Westglen Dr. 77063-Houston, TX USA	www.omnistar.com
PCI GEOMATICS	Software for cartography and GIS	50 west Wilmon Street, Richmond Hill, Ontario CANADA L4B1M5	www.pci.on.ca sales@pci.on.ca
P.GEERDERS Consultancy	Marine and coastal remote sensing applications services	Kobaltpad 18, 3402 JL, Ijsselstein, THE NETHERLANDS	pgcons@wxs.nl plaza.wxs.nl/pgconsult/

INSTITUTION NAME	EQUIPMENT, PRODUCTS OR SERVICES	ADDRESS/COUNTRY	WEB SITE or E-MAIL
RADARSAT International	Images and Image Products.	CANADA	www.rsi.ca
RESEARCH SYSTEMS	ENVI Software	USA	www.rsinc.com
SITEM S.L.	Photography and satellite image processing, Digital Elevation Models cartography	Aragó 141-143, 08015 Barcelona, ESPAÑA	www.sitem-consulting.com
SOKKIA CO.LTD.	Total Stations Level, Theodolites	20-28, ASAHICHO 3-C HOME, MACHIDA,TOKIO,194-0023 JAPAN	www.sokkia.co.jp
SPOT Image	Remote Sensing Images, Products, etc.	FRANCE	http://www.spotimage.com
THALES NAVIGATION TRIMBLE NAVIGATION	GNSS (ASHTECHMAGUELLAN) GNSS, Total Stations, Theodolites, Levels, geodetic and topo-cartographic software	471 El Camino Real Santa Clara, CA 950050 – USA 645 North Mary Ave. Sunnyvale, CA 94088-3642 USA	www.ashtech.com www.trimble.com
XYZ Sistemas Industriales S.A.	Cartographic and data base handling for their use with Internet Mapper application	Av. Infantes 105, 39005 Santander, Cantabria. ESPAÑA	www.imapper.com
Z/I Imaging Corporation	Cameras, Scanners, Stereo-plotters, GIS	301 Chochran Road, Suite 9 Huntsville AL USA 35824	www.ziimaging.com

CHAPTER 7
HYDROGRAPHIC PRACTICE

by LCdr. Bob WILSON and LCdr. David WYATT (UK)

1. INTRODUCTION

The execution of a surveying operation, from its inception to the final submission of the results, is a continuous process, all parts of which must be methodically and meticulously carried out if full value is to be gained from it. The most accurate and carefully gathered data will be wasted if not processed carefully and collated and rendered in a clear and understandable manner.

All information must be gathered, validated, checked and rendered in a logical and uniform manner using clear unambiguous terms so that the data may be recovered and understood both immediately after the survey and in the future.

The use of sophisticated computer systems and instrumentation in all areas has not lessened the surveyor's responsibility. The need for rigorous quality management is as essential now as ever, but made the more difficult by the rapidly increasing volume and complexity of data gathered. The surveyor must understand the principles of the equipment he controls, be meticulous in ensuring that any data input are entirely correct and examine with care the data output before they are passed to the next stage. Only by these means will the data rendered be of the highest quality and be able to fulfil its needs until the task can be repeated perhaps decades, if not centuries, later.

There is no substitute for practical experience where theoretical knowledge can be turned into practical skills; time spent in the field gathering data under the guidance of an experienced hydrographic surveyor will highlight the many difficulties and complex problems that need to be overcome to ensure the final product meets the initial requirement. The use and applications, to which highly complex sophisticated equipment can be best applied, can only be learnt whilst involved in the practical tasks of a hydrographic survey.

The processes are discussed in more detail in the following sections. The points covered are not exhaustive, and the surveyor must use his judgement and experience to expand or contract the list as considered appropriate. This chapter will consider general principles applicable to any survey.

2. HYDROGRAPHIC SURVEY PLANNING

As will be seen, survey planning is a complex process requiring considerable attention to detail, a flexible approach, good management and effective decision-making. If the planning is thorough, the chances are that the survey will be good too.

2.1 The Hydrographic Project

Survey planning is the term used to cover the whole process of the development of a hydrographic project from its inception, its subsequent issue to a designated unit, the detailed planning within that unit of how to conduct the survey, and the final submission of data to the Hydrographic Office.

Survey planning must, therefore, involve a coherent procedure which will consist of the following stages:

a. survey requirement.
b. preparation of a hydrographic survey specification. (To include a review of existing data.)

c. issue to a designated unit.
d. programme planning of that unit.
e. assessment of the task within that unit.
f. reconnaissance requirements.
g. resource allocation.
h. detailed survey planning.
i. estimation of time required.
j. final programme planning and approval.
k. liaison with outside authorities.
l. administration planning.
m. daily planning.
n. plans for compilation and checking of data.
o. plans for rendering of data.

The development of the survey requirement varies greatly from country to country. The final survey specification is assigned to a designated survey unit which has responsibility for the remaining planning requirements. A 'designated unit' might be a government-surveying vessel, an independent survey party working either in a chartered vessel or using boats, or a commercial company working under contract. Each nation will have its own planning process.

2.2 Assessment of the Survey Task

Before detailed planning can take place, the surveyor must be very clear about the aim of the survey, and who will be the primary user of the information. In general every survey should cover the immediate needs of the user as well as benefiting others.

Having studied the survey requirement, accompanying data and graphics of the area to be surveyed, the surveyor must first decide whether any additional data is required and propose any changes he considers necessary to the survey task.

Once all the basic data is held and the limits and scale of the survey have been fixed, the main surveying tasks can be established. The specifications for the survey will be stated directly in the survey job specification, key amongst these should be the specified Order of the survey as defined IHO Publication S-44.

The main task assessment points are as follows:

a. establishment of geodetic control;
b. method of positional control and calibration of navaids;
c. sounding criteria including interlining policy;
d. sonar search category;
e. tidal datum and observations;
f. wrecks and obstructions;
g. seabed sampling;
h. oceanographic observations;
i. tidal stream observations;
j. geophysical observations;
k. coastline and topography;
l. lights and buoys;

m. sailing directions and views;
n. radio Signals;
o. ancillary observations (vertical photography, overfalls, measured distances, leading lines, shore magnetics, natural phenomena, etc);
p. passage observations.

Existing Data. The surveyor should be provided with copies of the largest scale published chart and all previous surveys to survey scale, together with their Reports of Survey (RoS). These should be studied carefully, together with the relevant sections of tidal data, sailing directions, aerial photographs and topographic maps.

Resource Allocation. From the list of points in 2.2.5, and a study of previous data, the size of the task can be assessed. Detailed planning will take place after this process, but the initial study will reveal what resources are required to meet the task. The following list shows some of the considerations that should be made when planning resources:

- Weather predictions and sea state. These will affect the size of the vessel to be used for the task and the ability to use boats for inshore work and examinations.

- Size of the shallow water area. This will dictate how much boatwork is required. If boats are necessary, the time taken to complete the task will be highly dependent on sea state; 2.2.7.1 and 2 should, therefore, be considered together.

- Use of helicopters. It may be that the unit has a helicopter available, but if not is one required for access to remote sites?

- Logistics. The endurance of the surveyor's own resources will dictate fuel, water and stores requirements. Maintenance of equipment is another consideration.

- Manpower. The number and specialisation of personnel required to meet each task must be assessed. The following factors are also relevant: changes of personnel; mail and communications; leave and recreation; medical facilities ashore and afloat; shore support and transport; shore accommodation and monetary arrangements.

- Topography. This will dictate the resources required to access sites ashore.

- Detached Boat Camps - A detached boat camp may be ordered by survey specification in which case the planning list above should be followed. However, consideration should be given to detaching a boat to conduct inshore work and shoal investigations if a suitable harbour or sheltered mooring exists. Time spent lowering and hoisting boats is unproductive.

Likely constraints on the conduct of the survey should also be investigated. Investigate what fishing activity is likely to affect survey progress together with the constraints imposed by danger, firing and practice areas, shipping lanes and marine traffic choke points.

A field reconnaissance may be required to expedite the survey. See Section 3

2.3 Detailed Survey Planning

Once the size and scope of the task has been assessed and the necessary resources required to conduct it decided upon, detailed survey planning can begin. A number of activities can be planned to run in parallel and a good surveyor will try to reduce the overall time required to achieve the aim. A comprehensive list of required actions is given in the following paragraphs, but it must be remembered that every survey will be different and additional items may have to be inserted, or listed ones deleted.

2.4 Horizontal Control

The survey specification will detail the horizontal reference for the survey and list details of existing co-ordinated geodetic stations together with their descriptions if held:

Decide how best to achieve the accuracy standards of horizontal control set out in survey specifications. The survey specification will detail the choice of navaid, more than one may be ordered. In rare cases, it may not be possible to achieve the stated standards with the navaids available and relaxations may have to be sought from the Hydrographic Office.

Once the choice of navaids has been determined, their sites need to be chosen. Use whatever network analysis systems are available. Decide how best to co-ordinate new stations. Consider access to sites and any reconnaissance requirement. Decide how to power navaids and work out how often site replenishment will be required. Note the authorities that must be approached for permission to use the chosen sites, authorise frequency clearance and land boats and helicopters. This will include sites for shore marks if visual fixing methods are to be employed.

Decide on how and where the chosen navaids will be calibrated and whether re-calibration will be necessary during the course of the survey.

Most modern surveys use some form of GPS for control ashore and afloat. Where DGPS is used for control afloat it must be validated. In remote locations ashore, point positioning within 20cms should be achievable within 24 hours of data capture if data can be transmitted to the Hydrographic Office for comparison with the nearest ITRF monitoring site. Otherwise establishing a new station will require connection to an existing network.

2.5 Vertical control

The survey specification will detail the datum to which soundings are to be reduced and its relation to existing land datums, a list of any existing benchmarks should also be provided. The following planning points should be considered:

Decide where to observe tidal heights if not ordered in the survey specification. Decide on the siting of additional poles and gauges if required and plan the laying and recovery of offshore gauges if appropriate. Ensure that gauge and pole sites do not dry out at low water, plan additional gauges and poles if this is unavoidable.

Decide how best to establish chart datum on the pole/gauge from existing benchmarks or from transfer of datum or observation and analysis. Plan to connect a newly established datum to a land based levelling system if applicable.

Establish the nature of the tide and expected ranges, and the affect this will have on fieldwork. Where there is a possibility of tide poles or gauges drying out an additional pole or gauges should be planned to allow tidal data to be recorded throughout the survey.

Decide whether co-tidal adjustments will be necessary. If they are, determine the tidal factors from tide tables and the appropriate co-tidal chart, or locally produce them from the best data available. Seek advice from the Hydrographic Office if necessary.

2.6 Tidal Streams

Establish the expected maximum rate and direction of tidal stream within the survey area.

Determine the requirement for full tidal stream observations and how online observations can be conducted.

Identify charted overfalls, eddies and freshwater springs and plan to observe them.

2.7 Sounding

The following general considerations should be taken into account when conducting initial planning:

By examining the largest scale charts of survey area and previous surveys locate all critical or controlling depths and prepare sounding comparison overlay.

Plan main survey line spacing, direction and sounding speed. For SBES, where possible, lines should be perpendicular to general direction of contours.

Plan cross line direction, normally at right angles to main sonar and sound line direction, and plan to run these at the start of the survey as a crucial quality control measure.

Estimate likely spatial or temporal changes in sound velocity regime and plan initial sound velocity probe coverage.

Estimate sounding error budget and compare to the survey specification.

Vessel speed is to be assessed for the expected range of depths in the survey area and the type of echo sounder in use. Compare vessel speed with speed required for towing sonar to determine optimum and maximum survey speed.

During the planning of sonar and sounding lines, a list of all planned lines should be kept.

For SBES surveys additional lines should be considered inside the 10m contour. Additional lines should be run parallel to a jetty or wharf.

Lines of sounding should be planned and run along recommended tracks, leading lines, in possible anchorages and off headlands passed close to hand by vessels on normal passage.

If using SBES for offshore surveys particular attention should be paid to the sounding of depths <40m, where the least depth should be obtained over all seabed features. Interlines should be run in depths <40m unless the seabed is flat and featureless and no dangers are shown to exist by complete coverage by high definition towed side scan sonar. A full explanation should be given in the RoS when areas <40m are not interlined.

2.8 Side Scan Sonar

The following general considerations should be taken into account when conducting initial planning:

Tidal stream will have a significant influence over sonar line direction when using towed systems, often a compromise will need to be reached between the best sounding and sonar line directions. In some cases bathymetric and sonar data will have to be gathered separately.

Inspect the Wreck List supplied with the survey specification data and identify those with positions listed as approximate and thus requiring disproving searches or special attention. The limit of the area of this search may extend outside the given limits of the stated survey area. See IHO Publication S-44 Chapter 6.

Plot Wreck List information, other dangers and depth contours on planned track plots.

When surveying in or near oilfields or exploration areas careful note is to be taken of 500 metre safety zones, seabed installations and possible pipelaying operations to ensure the safety of the sonar towfish.

Sonar lines should be planned to run within 20 degrees of the prevailing tidal stream or current. In areas of strong tidal flow, a direction much less than 20 degrees may have to be adopted to ensure that the sonar towfish follows the ship's track closely.

The sonar line spacing should be planned in accordance with the survey requirements.

Ensure that any disproving search areas lying on the outer edge of the survey area are covered. Additional lines are to be planned to run outside the area in order to ensure complete ensonification of the area, with appropriate overlap.

Whenever a survey includes a channel, recommended track or leading line in restricted waters it should be swept by sonar. When planning such sweeps allowance should be made to accommodate the largest vessels likely to use these tracks paying particular attention to turning areas and where a track changes course.

2.9 Seabed Sampling

Seabed samples should be obtained as required throughout the entire survey area. See IHO Publication S-44 4.1.

The survey specification may require the retention of a percentage of all samples obtained. Due allowance should be made for this requirement.

2.10 Coastline Delineation, Conspicuous Objects and Topography

The requirement for mapping of the coastline and other topography will be defined in the survey specification.

The High Water line shown on maps cannot always be relied upon as the coastline for hydrographic surveys.

Using charts and any photo plots supplied with the survey specification, identify those areas adequately covered and those requiring additional work. Where no modern charts, maps or air photo plots exist, all coastline and topographic detail which will be of use to the mariner should be fixed accurately.

The surveyor should attempt to obtain copies of any relevant local modern charts, maps and geodetic data additional to those supplied with the survey specification. Any such data should be rendered to the Hydrographic Office at the end of the survey.

Determine the means of delineating inadequate areas and identify equipment to be used to define such areas appropriate to the scale of the survey.

2.11 Ancillary Observations

Geophysical Observations The survey specification should detail which geophysical observations are required, but generally magnetic and gravity observations can be taken concurrently with bathymetry. The survey specification will cover line spacing in detail. If magnetic anomalies are charted, plan to observe and report them. Plan magnetic observations ashore if ordered in the survey specification.

Lights and Buoys

Establish which lights should be visible from the survey ground and plan to check their characteristics.

Establish the number of buoys that require to be fixed.

Air Photography If air photography is ordered, then plan to fly it during advantageous tidal and weather conditions. The subsequent photographs may be used for coastlining and topography.

Sailing Directions and Views The amendments to Sailing Directions can normally be compiled during the course of the survey and additional time to observe information for inclusion should not be required. Plan to check all existing photographic views and to photograph new ones as ordered in the survey specification. Plan to check harbour facilities and the facilities for small craft.

Radio Stations Plan to check the accuracy of published data.

Jetties, Wharves and Landing Stages Plan to check the details of jetties, wharves and landing stages. This can normally be done during boat sounding operations.

2.12 Surveying Team Organization

The senior surveyor will normally prepare a Survey Order Book which will give an overall plan, stating how the survey is to be tackled, and detailing responsibilities for planning and execution of the work. These orders should be updated regularly to inform the whole team of the short term priorities and to give the framework within which more detailed day-to-day planning can be made. When both the ship and her survey boats are working together on a daily basis the activity becomes particularly intensive and complicated. It will then be vital to have a flexible and well thought out plan to co-ordinate times of boat lowering/hoisting, change of boat's crews, provision of victuals, and instructions for those in charge of boat work.

Shortage of manpower will always be a problem at the start of the survey, with observing and tidal parties away simultaneously and boats crews and shore parties on standby to land nava ids etc. Good shore transport and/or helicopters will greatly assist the successful start of a survey.

The Bridge and Chartroom organisations require careful planning and structuring ensuring that the data is acquired and handled in the most efficient manner.

The pressure on all surveying units for increased productivity is considerable. Good management and planning as well as positive leadership is fundamental to the success of the survey.

2.13 Compilation and Checking of Data

Quality Control should be built into the plan at every stage, with nominated checkers of all data required to keep abreast of the incoming work.

Plan the allocation of drawing and compilation tasks. Ensure that accompanying fair records are compiled and checked as the survey progresses.

In a large survey, it is generally better to complete one area in all respects before moving to the next. This will ensure that complete data can be rendered should the unit be withdrawn from the survey for some over-riding reason.

It may be convenient to allocate the writing of separate sections and annexes of the Report of Survey to individuals.

All transcripts and records should be compiled as the survey progresses and not left until afterwards if at all possible.

The comparison of surveyed data with previous surveys is a most important consideration; it should proceed with the fieldwork and the planning should take account of additional investigation and disproving searches that may arise from differences between charted and surveyed data.

2.14 Data Rendering Requirements

The data required to be rendered to the Hydrographic office will vary greatly depending on national policy and requirements. In general it will include:

a. bathymetric data set in digital or graphic (Fair Sheet) format;
b. navigational track data in digital or graphic format;
c. sonar data set in digital or graphic format;
d. seabed texture data set in digital or graphic format;
e. report of survey.

Hydrographic Offices should appraise all the survey data rendered and dispatch a critique within two months of rendering the data. Points raised from the Hydrographic Office should be answered as quickly as possible while the survey is still fresh in the mind.

2.15 Operation programme Development

The total number of planned days taken to complete a survey must now be fitted into the requirements for port calls, vessel maintenance, passage time, exercises etc. Each vessel will have its own operating cycle and from this skeleton a work cycle can be fleshed out and passed for approval if necessary. If the estimation of time required shows that the survey cannot be conducted within the broad timescale allotted in the survey specification the matter must be represented for programme modification or reduction of the size of the survey task.

2.16 Operation Duration and Cost Estimates

There are no hard and fast rules for arriving at a precise time required for completing a survey. An experienced surveyor can sum up the requirement after studying the survey specification and arrive at a good estimate without recourse to a mathematical sum. However, the format provided at Appendix 1 to this chapter will provide a reasonable answer and can be adjusted to suit the needs of any survey. During the detailed planning stage the surveyor should maintain a tote of line mileage, numbers of wrecks, numbers of bottom samples required etc. This data can then be used to compile the time required format.

2.17 Liaison with Outside Authorities

As soon as the survey specification is received, letters should be sent to a number of outside authorities giving the broad details of the survey specification and the timescale of the survey, together with a request to use facilities if appropriate. This can be followed by further letters with more detail once the detailed planning has been conducted, if considered necessary. A list of examples is shown below. Survey specifications are often sent to a number of agencies direct from the Hydrographic Office, and the survey specification covering letter should indicate those organisations that have already been informed:

a. fishing authorities;
b. local landowners;
c. coastguards;
d. lighthouse authorities;
e. local defence forces;
f. firing or exercise range operating authorities;
g. oilfield operating authorities;
h. local government representatives;
i. naval attachés;
j. local chart/land survey departments;
k. helicopter operating authorities;
l. religious authorities.

In addition, if a detached boat party is to be landed or shore parties are to be landed from sea or operate from a local port or settlement, the following should also be considered:

a. local police;
b. Mayor, Town Chief or Headman;
c. harbour authorities;
d. local Service establishments.

Follow up visits may need to be made during either the advanced survey reconnaissance or on arrival. Security implications should always be considered.

3. SURVEY RECONNAISSANCE

3.1 General Reconnaissance

Reconnaissance is needed prior to any survey to acquire the necessary data to permit the best and most economical survey to be carried out. The information collected should cater for the design, planning, organisation and observations of the proposed task. The reconnaissance may be carried out immediately before the survey, or many months in advance.

The reconnaissance is important; a bad one can result in wasted time and effort later, when much more expensive assets are likely to be involved. It should also be complete because a poor reconnaissance will inevitably result in a poor plan.

The surveyor called upon to do the reconnaissance should possess experience, commonsense, a sound knowledge of all equipment available, and have no preconceived ideas about the method by which the task will be carried out. The actual observations can safely be left to less experienced surveyors once the major decisions have been taken.

3.2 The Geodetic Reconnaissance

The purposes of the reconnaissance can be summarised as follows:

a. establish local contacts in person;
b. visit all proposed stations - select actual sites. Recover existing control stations;
c. confirm inter-visibilities;
d. decide upon final network design (re-analyse if necessary);
e. permanently mark geodetic stations;
f. describe geodetic stations;
g. prove the proposed observing plan (instruments/targets required). Prepare detailed observing programme;
h. prove the administrative plan for the main survey, amend as necessary.

For each new geodetic station, the following information will be required:

a. accessibility by road, rail, boat, foot or helicopter. Time for access (e.g. on foot from road) and recommended route;
b. visibility from station and requirements for any subsequent clearing;
c. description of the station, magnetic bearings to other visible stations;
d. photographs of the station, surroundings, and panoramic photos from the station;
e. local factors, customs etc;
f. likely visibility and meteorological conditions.

3.3 The Tidal Reconnaissance

Whenever possible it is advisable to use established or previously used tide stations for commonality of data. When selecting a site for a tide gauge and tide pole the following must be considered:

- Ease of Erection Consider which is the easiest place to erect a tide pole and gauge, some places are easier than others, and some places are impossible;

- The Station Must Not Dry Out. The zero of the Tide Pole and Tide Gauge pressure sensor should not dry out. If this is unavoidable a secondary pole and gauge should be established below the level of the first gauge or pole;

- Ease of Reading. The pole or gauge must be sited such that it can be read at all times;

- Security. Avoid situations where the tide pole and particularly the tide gauge will be likely to be interfered with by the public, e.g. Fishing boats berthing;

- Shelter. The pole or gauge sensor should be sited away from the more severe effects of weather sea and swell;

- Protection. Ideally the tide gauge recorder should be placed in a lockable building;

- Impounded Water. Water that is restricted in movement by a sandbar or basin will not be at the same level as the open sea. Therefore a station should be selected that reflects the true level of the sea at the survey area;

- Proximity of Bench Marks. Select a station near to two benchmarks if possible, to avoid time spent on long levelling runs;

- Accessibility. If a tide watcher is employed, accommodation should be close at hand. If a detached Boat Party is in operation the tide station should be close to where the boat is moored, or close to the tide party base.

4. DATA ACQUISITION

The depths shown on a nautical chart are its most important feature and the mariner must be able to rely implicitly on accurate bathymetry to avoid danger. The greatest care must be taken to ensure that soundings are precisely positioned. An error in position is often more misleading than an error in depth, for a mariner is more likely to navigate clear of a charted danger than to rely on the accuracy of its charted depth and deliberately navigate over it.

The disciplines of sounding error monitoring, data checking and Quality Control (QC) are continuous procedures that need to be sustained throughout the entire hydrographic surveying process. Similarly the generation of the final report should commence on completion of the planning stage and be uninterrupted during the remaining phases of the survey; it should not be left until the end when all data acquisition has been completed.

4.1 Horizontal Control and Calibration

4.1.1 Introduction

The Hydrographic Specification will detail the horizontal datum to be used during the survey; if, after the planning and reconnaissance (paragraph 2.4), there are insufficient co-ordinated geodetic stations, secondary stations, landmark and navaids then additional horizontal control should be generated within the area and sub-areas to meet the required accuracies for positioning at sea.

The methods selected for providing control offshore will dictate to a great extent the preparatory work required onshore. Numerous shore stations may be needed for visual fixing in surveys of small areas close inshore, whereas only two stations may be required for surveys controlled by local area DGPS. In either case, the stations should be as close to the high water line as possible to avoid inaccuracies in the electromagnetic patterns caused by variable propagation conditions over land paths.

Satellite position fixing is capable of achieving great accuracies with GPS Relative Positioning Techniques – code phase Differential GPS (DGPS) and Real Time Kinematic (RTK) carrier phase DGPS – with only one GPS reference station, which gives greater flexibility on site selection and deployment than is the case with terrestrial methods. DGPS corrections can be obtained from Radio Beacon Navigation Service (Beacon-IALA) and a variety of WAAS (Wide Area Augmentation Systems) via commercial services (Landstar, Seastar, Omnistar, Skyfix etc.) and free service (EGNOS); these systems provide good accuracies in positioning without the need for a reference station onshore, however GPS receiver calibrations and real time checks on geometry (GDOP) should be conducted during the survey.

4.1.2 Horizontal Control Ashore

Control for offshore surveys can usually be generated by extending the established geodetic network in the vicinity. Failing this it will be necessary to determine a datum position, azimuth and scale to allow the new stations to be fixed relative to each other.

Conventional land surveying techniques should be employed, these are summarized below, detailed explanations can be found in Chapter 2 with reference texts listed in the bibliography:

a. the determination of absolute position of a datum point (A);
b. the orientation of the network by azimuth observations (at A to B);
c. the determination of scale by baseline measurement (from A to B);
d. the extension of the network by traverse, triangulation or trilateration to the required stations, with intermediate stations fixed by resection or intersection.

Operations a., b. and c. will be required only where no established geodetic network exists. This is rarely necessary; the techniques for astronomical or satellite GPS observations ashore, performed for geodetic surveys, are beyond the scope of this manual.

Angular observations are made by theodolite or sextant with distance measured by mechanical, optical or electromagnetic (EDM) means or both by Total Stations. Subsequent computations may be carried out on the reference spheroid in terms of latitude and longitude, or on the grid and projection in rectangular co-ordinates using plane trigonometry.

GPS observations, carried out using geodetic dual frequency receivers or by RTK DGPS technique, can produce better accuracies in the determination of a baseline (see paragraph 6.1 Chapter 2), however it should be remembered that the position co-ordinates obtained are referred to the WGS 84 ellipsoid and a compatible grid and projection. Datum transformation from WGS 84 must be performed (see paragraph 2.2.3 Chapter 2), if the hydrographic survey is to be conducted in a local horizontal and vertical datum.

The positional accuracies of marks for primary shore control points and secondary stations are specified in IHO S-44.

4.1.3 Horizontal Control at Sea

General Description of Positioning Systems

Terrestrial positioning methods include traditional land-based techniques such as:

a. sextant resection positioning;
b. triangulation/intersection positioning;
c. visual positioning methods;
d. tag line positioning methods;
e. range-azimuth positioning methods;
f. land based electronic positioning systems.

Since the early 1990's most of these terrestrial positioning methods have been largely replaced by satellite-based positioning systems, namely GPS and more accurate code phase Differential GPS (DGPS) and Real Time Kinematic (RTK) carrier phase DGPS. Within isolated project areas, where satellite GPS methods may be inaccessible or impractical, one of the traditional terrestrial survey techniques may be needed to provide survey control. Examples of such cases may include:

a. small dredging or marine construction projects where only a limited amount of depth coverage is required;
b. areas under bridges, in deep-draft harbour berths or near dams where GPS satellite view is masked;
c. intermittent, low-budget projects where traditional terrestrial positioning techniques may prove more economical than equipping a fully automated DGPS-based hydrographic survey system;
d. quick reconnaissance surveys, where meeting a specific positional accuracy standard is not required.

Procedural methods and Quality Control (QC) criteria for some of these terrestrial survey techniques are detailed in this manual primarily for reference purposes.

Horizontal Positional Accuracy

All the positioning methods, summarized at Appendix 2 in Table 7.1 “Horizontal Positioning Systems and Selection Criteria”, are capable of meeting the required minimum standards of horizontal accuracy for a selected survey Order, detailed in IHO S-44, provided that distances from the shore-based reference point and the vessel are within normal system operating limits. The operating limits vary with the type of positioning system, procedures employed and the environment in which it is being used. In general, the positional accuracy of all systems will degrade as a function of distance from the baseline reference points, some faster than others. Users must fully assess and evaluate the resultant accuracy of any positioning method, including DGPS, to ensure its suitability for the survey to be undertaken.

Selection of Positioning Systems

The accuracies predicted for positioning systems employed in hydrography are generally quoted with reference to normal use of the equipment within their operational limits and the different classes of surveys. Table 7.1 shows the criteria for selection and employment against the orders of hydrographic surveys, as defined in the IHO S-44, for positioning systems with their anticipated positional accuracy. Suitability of a particular technique for a survey should be guided by the tasking authority taking these limitations into account. The Table assumes a standard project area located within 25 miles of the coastline or shoreline reference point (horizontal control) or up to 200 meters water depth. Criteria for performing surveys within these ranges should conform to the standards contained in IHO S-44 and in this Manual.

Generalised accuracy ranges achievable with each type of system are also shown in this and other manuals, including operators' equipment manuals; the extreme variations are a result of factors discussed elsewhere in this manual and relevant chapters of the above mentioned equipment manuals. The indicated maximum accuracy range is generally that which could be expected with the equipment being employed within its normal operating limits and conditions. In some cases the accuracy range covers those prescribed for Special, 1st and 2nd Order surveys; this indicates that project-dependent factors (geometry, distance offshore, etc.) must be considered in order to select the most appropriate equipment for a particular order of survey or project site.

Track Control

The methods highlighted in paragraph 4.1.3.2 will prove the surveyor with a position at sea, in addition, he must ensure tha t his vessel follows the desired track over the seabed making correct allowance for effects of tidal steams, currents and wind drift and therefore thought must be given to this requirement when planning positional control. The chosen fixing method will often also provide track information, such as a left/right indicator displayed on a device of positioning system or on the monitor of special HW/SW automated data acquisition and control, however, particularly in close range work, supplementary aids must sometimes be provided for the steering of the vessel.

In traditional visual methods or old EPS techniques, a real time plot of the vessel's track is kept manually or by a track plotter with the survey data superimposed after reduction in the post processing stage. In this case plotting sheets should be prepared with collector overlays to be used to generate a running record of survey progress.

Whatever method is employed, it will have an impact on the planning and execution of survey and must be considerer within the overall plan from the outset.

4.1.4 Field Preparation

General Description

A field reconnaissance of the survey area will save considerable time during the data gathering stage of the task. The positions selected for survey marks should be visited, their suitability confirmed and descriptions written. Once the survey team arrives on site, equipment will need to be installed ashore and in the vessel, all of which may require field calibration and checking.

Working within the "strategic" framework drawn up at the Hydrographic Office, the surveyor-in-charge must refine the plan and if necessary review the deployment of personnel and equipment for optimum utilisation within the overall project. Any adjustments to the initially agreed plan should be discussed with the Hydrographic Office and suitable methods for monitoring progress and achievement of key mile stones should be put in place.

Observation Planning

The greatest care should always be taken when observing the framework of the geodetic system, every opportunity taken to obtain checks on the observations and to detect weakness in observing techniques, observers and equipment. All calculations must be completed and very fully checked before proceeding with the fieldwork dependent on the accuracy of co-ordinates derived from such primary observations.

The surveyor should identify the optimum observing periods, using a mission planning program, to achieve the order of standard for the survey. Instrument selection should be such that observations of the appropriate type and standard are obtained, calibration data should be checked and the details recorded for inclusion in the Report of Survey.

Site Selection

Considerable care should be taken on network creation, site selection and density, installation of the reference stations and the techniques for measurement of angles and distances, to ensure the necessary accuracy in the positioning to meet the survey Order. The type of survey being undertaken (harbour and approach, littoral, coastal or off-shore), the selected positioning system (visual /EDM/EPS/Satellite), the number of the LOPs and their geometry within the survey area will all have an influence on the final decision.

The site selection should be based on:

a. accessibility of the site by land or from sea;
b. the ability to occupy the station or the necessity to create an eccentric station;
c. proximity to the shore or coastline with clear views to seaward;
d. inter-visibility to adjacent sites, clear of structures likely to cause interference of EDM/EPS signals and unobstructed receipt of the satellite signals;
e. availability of mains power or space to co-locate portable power supplies, such as batteries/solar panels and generators;
f. site security and ability to leave equipment untended;
g. site elevation and suitability for chosen positioning system.

Beacon Deployment and Inspection

Check lists, created by the surveyor-in-charge from the equipment manuals, should be followed during the installation of the ground reference stations (EPS, DGPS or RTK GPS) or during the use of visual/EDM tools for measuring angles/distances (sextant, theodolite, EDM, total station) to ensure the correct operation of the system and similar techniques are used throughout the survey.

The type of the deployed ground reference stations (EPS, DGPS or RTK GPS) will determine the frequency of inspection necessary to verify correct operation; this is also the case for unmonitored total stations operating in automatic mode.

4.1.5 Alignment and Calibration of Positioning Systems

General Description

The type of system or tool selected will dictate the procedure adopted to verify performance against the anticipated limits to ensure the achieved positional accuracy matches the select survey Order requirements, as expressed in Table 7.1.

The alignment/calibration procedures and techniques, detailed in the user manual (or operator manual), should always be followed at the beginning and end of a survey and when deemed necessary to verify system performance in the field, particularly if performance or accuracy is suspect. These checks should be conducted, as far as possible, within the survey area at the expected ranges and against a previously calibrated higher order system or navigational aid or between co-ordinated control stations. All total stations, EDM systems and prisms used for primary control work should be serviced regularly, checked frequently over lines of known length and in date for periodic factory calibration.

Angular Measurement

Care should be taken to ensure that the correct observing techniques for angular measurement systems (sextants, theodolites, total stations) are used and that instruments are set up to minimise errors. Instruments should be in-date for calibrations and service; standard zeros for the appropriate order of observations should always be used and careful recording techniques to avoid blunders.

Each station selected for use should be visited and carefully checked against the station description, the distances to fixed reference points should be confirmed to determine if the station mark has been displaced. New stations should be checked for inter-visibility to the survey area and other stations and be linked to 3 established stations. The use of eccentric stations should be avoided whenever possible. Any amendments to the planned observing scheme due to unsuitability of the sites should be re-analysed to ensure that the standard for the order of survey is being met. All stations used should be marked and full descriptions recorded before observing at or to them.

When determining heights by angular measurement reciprocal heighting is to be used whenever possible. Before moving the observing instrument, verify the data recorded to ensure the observations, both angular and distance, are to the standard required, if the standards are not met, re-observe the entire set of observations.

The verified final angular and distance observations should be adjusted to the grid as appropriate for each type of observation by using an approved computer program and then compute the most probable position and error ellipse data. The error ellipse of each new station position should be carefully examined to determine the quality of the final position. Network analysis should be conducted.

Distance Measurement

When using distance measuring systems (EDM, EODM, total stations, etc.), all the procedures described in the operator/equipment manuals should be followed and a comparison check conducted against a geodetic baseline or a higher order system with greater or equal accuracy to that required by the order of survey for establishing position.

2D Measurement

As with distance measuring systems, the guidance in the user/equipment manuals should be followed for 2D positioning systems with appropriate calibration and comparison checks conducted against higher order systems or geodetic baselines/networks.

When planning the use of microwave EPF systems to validate GPS positional data prior to the commencement of survey operations, care should be taken to ensure that established stations are all on a common datum. Navigation systems should be calibrated and verified by comparison with an alternative precise positioning system at the start of each survey and a validation undertaken at the end.

Satellite Measurement (3D)

When using GPS satellite systems, observing procedures articulated by the Hydrographic Office and detailed in the user guides should be followed with great care to ensure the equipment is operated to its maximum capability for the various modes of positioning SPS, PPS, Differential and RTK available. All systems should be verified prior to fieldwork and a closing validation conducted on completion of observing secessions against a geodetic baseline, high order geodetic control net or a system with greater or equal accuracy to that required by the order of survey.

4.1.6 Horizontal Control Methods and Equipment

4.1.6.1 Sextant Resection Positioning

General Description

Sextant positioning involves the simultaneous observation of two horizontal angles between three known objects from which the position of an offshore point is resected (see Figure 7.1). Sextant positioning is totally performed aboard the survey vessel and does not depend on electronics, communications, or shore-based support. Under certain conditions (i.e., close to targets or for near static position fixes) it can be relatively accurate when properly conducted by an experienced team. In general, however, sextant positioning under dynamic vessel conditions is no longer considered accurate for most applications.

Hydrographic marks for sextant-controlled surveys may be located by sextant fixes or by sextant cuts. Lower than third-order traverse methods may be used, if the distance from a basic or supplemental control station does not exceed 4 km for hydrographic surveys at scales of less than 1: 10,000 or 2 km for larger scale surveys.

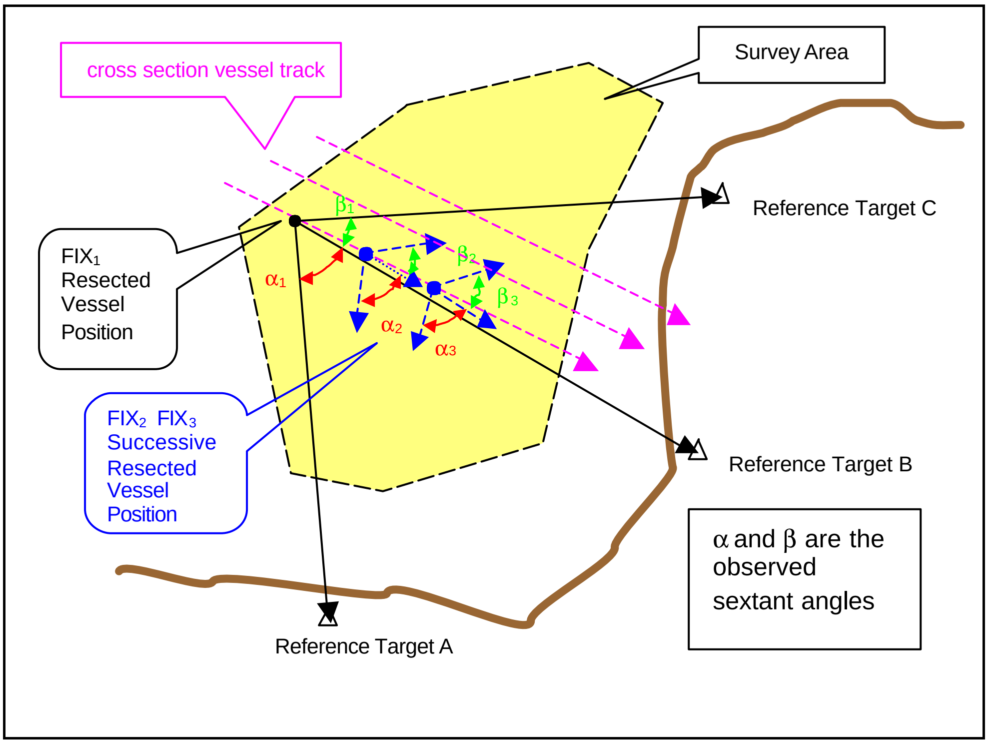


Fig. 7.1 "Sextant Resection positioning"

A single sextant angle may be used in conjunction with a fixed range LOP, as shown in Figure 7.2 (Hopper dredge positioning). In the past this was a common technique for locating hopper dredges.

On stable offshore vessels and other platforms, multiple sextant angles can be observed to several targets (*Redundant sextant resectioning*). The resultant fix can be adjusted by onboard software using least squares adjustment techniques with results being quite accurate (less than ±1 m in some isolated cases).

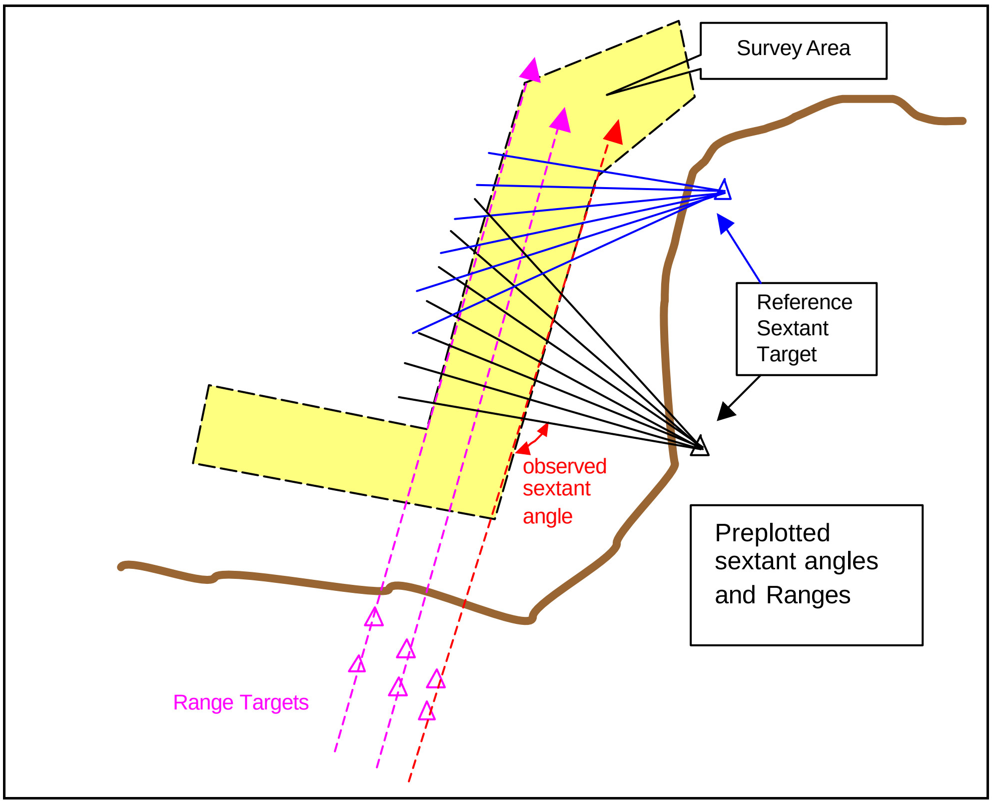


Fig. 7.2 "Hopper dredge positioning"

Accuracy and Quality Control

The two observed sextant angles form the loci of circles, the intersection of which is the vessel's position. Each angle forms a circle defined by three points: the two shore control points/targets and the vessel. The geometry of these two intersecting circles is a primary factor in determining the strength of a sextant resection, as the two intersecting circles converge on each other, the resultant position weakens drastically. In the best conditions, dynamic positional accuracies were rarely better than ±5 m (95% RMS), average accuracies were generally in the 10 to 20 meter range.

The simplest method for estimating *resection accuracy* at any point is to move each angle by its estimated accuracy and assess the resultant change in position. This is readily done when automated resection computing software is available, or by noting the position shift in a station-pointer. Positional accuracy should be accessed at various points in the work area. In performing sextant resection positioning the following QC factors must be considered:

a. sextant angles precision;
b. observer synchronization;
c. vessel velocity and motion;
d. observer experience and fatigue;
e. type of targets.

Due to design and handling, internal sextant instrument calibration is not particularly stable; therefore observers should continuously check the calibration of their sextants. This is usually done periodically during the survey, typically at the end of each survey line.
Few opportunities exist to perform quality assurance (QA) checks on sextant positioning. When more than three targets were visible, different resection positions could be compared from an anchored position.

Sextant fixes at distances approaching the limit of visibility of the marks are likely to be weak because the angles or rates of change are small. The sextant must be in perfect adjustment, and the angles measured and read with extreme accuracy, to the nearest 30 seconds of arc if necessary. If the sum of the two angles frequently approaches 180º with one angle often being very large and the other very small, the rate of change of angle will be rapid when the vessel is moving; thus, particular care must be taken to ensure simultaneous observations; the effects of errors introduced by failure to observe angles simultaneously is minimized when the distance of the marks from the observe are small.

4.1.6.2 Triangulation/Intersection Positioning

General Description

An offshore vessel or platform can be positioned (triangulated) by transit or theodolite angles observed from base line points on shore. This technique may have an application in areas where electronic positioning systems cannot be deployed or where increased positional accuracy is required. As indicated in Figure 7.3, two (or more) shore-based transit or theodolite observers are required. Due to the higher precision and stability of the instruments, the resultant positional accuracy can be quite good. Theodolite stations should meet the accuracy requirements for special order or order 1 surveys. The angle of intersection at the vessel should be such that a directional error of 1 minute from a theodolite station will not cause the position of the vessel to be in error by more than 1 mm at the scale of the survey; angles greater than 30º and less than 150º will usually ensure meeting this condition. Triangulation techniques are often used to supplement electronic distance measurement (EDM) or DGPS positioning of fixed offshore structures (piers, bridges, rigs, etc.) both during construction and subsequent deformation monitoring.

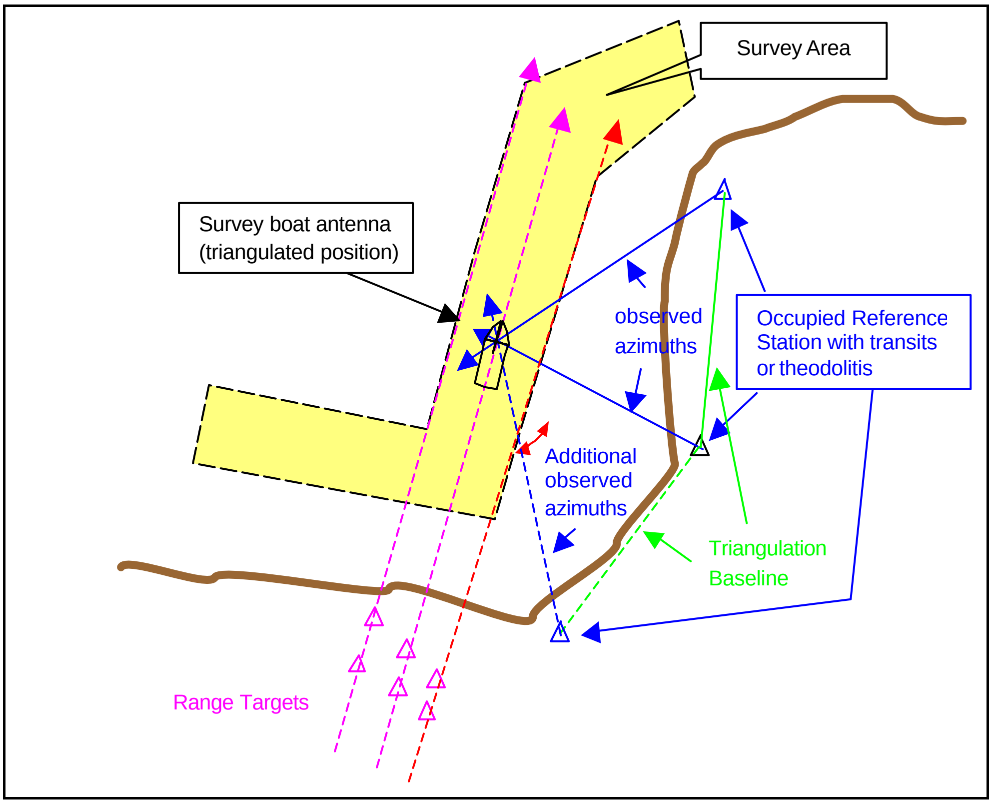


Fig. 7.3 "Triangulation/Intersection positioning"

Accuracy

Triangulation/Intersection positional accuracy depends on the tracking accuracy of the system in use; it is related to the geometric strength of the intersection from two angles or azimuth directions and varies throughout the survey area because the angular standard errors for each instrument vary as a function of distance between the instrument and the survey vessel. The average of the standard errors of each angular measurement at the offshore vessel position, together with the computed range from each observing reference point, gives an estimate of the triangulated/intersected positional accuracy.

Multiple azimuth intersection techniques, allowing three or more additional angular observations, enables increased accuracy with redundancy provided by the additional measurements from other shore stations.

Often azimuth alignments are combined with simultaneously EDM or GPS range measurements and a least squares adjustment technique is performed, if automated acquisition is running during the survey operations

Quality Control and Quality Assurance

QC is performed with periodic backsight checks during the course of the survey. Independent QA should be performed with a third instrument, which is not very easy to perform in practice and normally an EDM or GPS system is used to make checks.

4.1.6.3 Visual Positioning

General Description

This traditional method was often used to locate a hopper dredge relative to known shore features or flags and is still used for a few applications, such as horizontal and vertical alignment of construction equipment, rigs, barges, etc.

Relative visual positioning techniques is now rarely performed, given the availability of microwave EPS, range-azimuth and GPS positioning methods; it is generally suitable only for non-navigation reconnaissance work where identifiable features (navigational aids, beacons, day markers, bridges and other structures or map features) on the supplied drawings, navigation charts or maps are assumed to be accurate for this standard of survey.

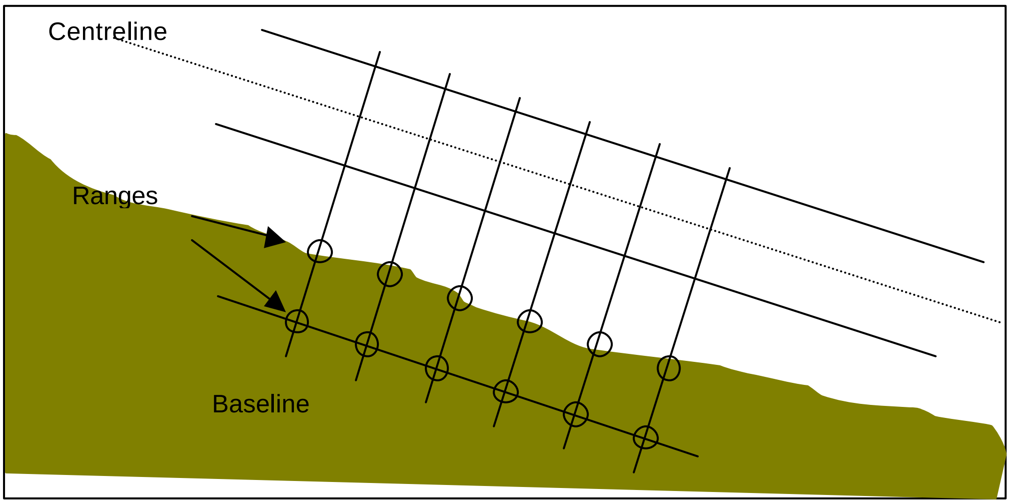


Fig. 7.4 "Range poles, flags, and/or lasers set ashore for relative positioning"

The main points of this method are:

a. the boat maintains constant survey speed between all the identifiable objects or range intersections;
b. fixes are taken when the survey boat is passing abeam or lateral to an identifiable object;
c. positions are interpolated between fixes;
d. vessel speed is assumed to be constant between fixes, which are assumed to be free of errors;
e. position determination may be obtained by intersection of shore points and ranges established by sighting across such features;
f. results should be used with caution due to the approximate nature of data and the marginal accuracy of such a survey.

Accuracy and Quality Control

Accuracy is difficult to estimate and QC is rarely performed, when using visual positioning techniques.

4.1.6.4 Tag Line Positioning

General Description

This traditional method was often used before the 1970s to monitor dredging progress of navigation projects and traditional channel cross-section surveys and in subsurface investigation for channel obstructions and channel clearance sweep surveys. Tag Line techniques were replaced by microwave EPS and range-azimuth techniques and now have been superseded by GPS positioning methods.

Within limited distances off the baseline and with proper execution, a tag line controlled survey is an accurate and stable method of performing hydrographic surveys and other investigative work for marine design and construction:

a. a calibrated wire rope is employed, stretched perpendicular from berths or hubs on a baseline to the survey boat;
b. is maintained around berthing areas for critical site investigation work; where GPS signals are blocked for such surveys (however an electronic total station is preferred);
c. usually requires no electronics or communication devices.

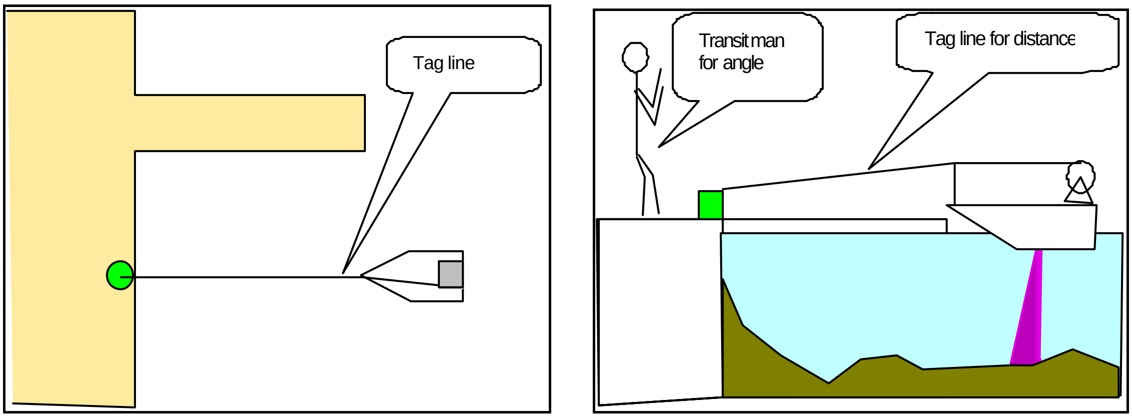


Fig. 7.5 "Tag line surveys"

Techniques

A tag line survey is simply a method of running cross sections from a fixed baseline.

Different survey techniques should be performed, depending on type of operations and instruments adopted as:

a. static observations – tag line length observations are made when the boat is properly aligned on the section and the wire is pulled taut to minimise sag;
b. dynamic or continuous tag line surveys – some tag line surveys are conducted in a dynamic mode using analogue echo sounders;
c. baseline boat tag line extension methods – tag lines may be anchored to a floating vessel (baseline boat) that has previously been positioned by tag line or other means;
d. constant range methods – a tag line may be used to maintain a constant range from the baseline hub;

e. baseline layout for tag line survey – intermediate points or baseline for controlling tag line work are set using standard construction survey techniques and standards;
f. tag line alignment methods – visual range flags, right angle prisms, transits, theodolites, sextants, and total stations are used for maintaining lateral alignment control of the survey boat, which can be the weakest part in the performance of tag line surveys, especially if strong currents are present;
g. data recording procedures – tag line survey and related depth measurements may be recorded on worksheets or in a standard field survey book, survey data are plotted in either site plan or section formats;
h. survey boats – any type of survey boat, equipped with tag line man-hold or powered winches, may be used to conduct tag line surveys. Generally boats' length range from 5 to 8 metres and drafts are less than 0.40 metres are essential to work in shallow water areas and to provide ease of beaching.

Accuracy and Calibration Requirements

Accuracy: the positional accuracy of a point positioned by tag line may be computed using the estimated accuracy of the alignment and distance measurements; similar to that done with range-azimuth survey methods.

Calibration: flagged tag line intervals must be periodically calibrated every 3 to 6 months or after breaks against a chained measure or EDM system.

4.1.6.5 Range-Azimuth Positioning

General Description

This, once widely used positioning technique, is based on the intersection of azimuth-range measurements, generally performed from the same shore reference station, similar to a forward traverse computation, see Figure 7.6. Now days it is employed only where GPS positioning cannot be obtained due to satellite masking. The main features are:

a. angle observations (azimuth) can be measured by transits, theodolites, or total stations;
b. distance observations (range) can be measured by EPS devices (laser or infrared EDM, microwave EPS or total stations);
c. data can be manually observed, noted on field book and voice-relayed to the boat by radio or digitally recorded and transmitted by radio-modem to the boat;
d. typically used within 5 km of coastline and/or reference station;
e. high relative accuracy is achievable depending on the equipment used (best accuracies are achieved by automated theodolites/EDM or total stations);
f. periodic calibration or a third LOP measurement (angle or range) is essential for redundancy;
g. a small team is required to perform the survey (relatively efficient);
h. boats 5 to 8 metres long often is used;

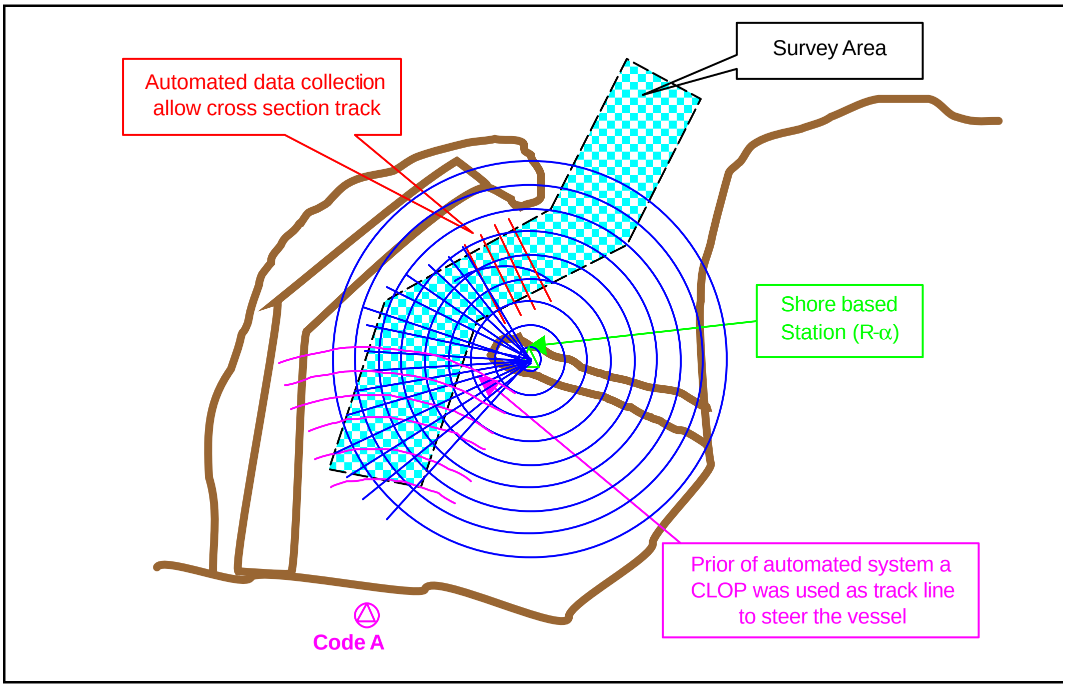


Fig. 7.6 "Range-Azimuth positioning"

i. theodolite with laser or infrared EDM and total station are highly accurate as range-azimuth systems for Special Order survey areas within 2 km of the reference point;
j. microwave based EPS will rarely meet positional accuracy standards for Order 1 (2 m or 5 m);
k. dynamic alidade or transit stadia distances meet positional accuracy standards for Order 1 (accurate to 5 meters) within ranges of only 30-50 metres, depending on conditions.

Quality Control Procedures and Requirements

Angular orientation:

a. the lower plate of the tracking device must initially be referenced, relative to the survey project, to the grid azimuth of the reference back sight (000° line of sight);
b. additional reference lines of sight (landmarks) should be taken as a redundant reference orientation;
c. all reference orientation points should be selected as the farthest and most reliable visible controls, sighted on and relative errors resolved onsite;
d. all orientation measures and grid azimuth computation must be recorded on a field book.

Periodic orientation checks:

a. periodic orientation checks on the initial reference back sight (000° line of sight) should be performed during the survey (normally every 20/30 fixes or 5/10 minutes or at the end of each survey line) to ensure that no horizontal or vertical misalignment occurred to the instrument;

b. additional reference lines of sight (landmarks) should be taken, this should normally completed at the beginning and at the end of the survey session;
c. periodic readjustment and re-levelling of the instrument should be performed as required after these checks;
d. all periodic checks and re-levelling operations should be noted into the field book;
e. if the orientation check indicates a significant misalignment, all fixes taken since the previous orientation check must be rejected and the measurements rerun.

Quality assurance checks:

a. independent positional checks are rarely available as with most visual survey positioning methods;
b. carrier phase RTK-DGPS techniques allow independent positional checks, but these will be performed with geodetic receivers in static mode and in the topographic range field;
c. for critical navigation surveys, position checks should always be done with the vessel as near as possible to a reference control point.

4.1.6.6 Electronic Positioning

General Description

A variety of systems have been developed, most of which have become obsolete since GPS became fully operational. However, the basic operating concepts behind land-based Electronic Positioning Systems (EPS) and related trilateration positioning (including GPS) have not significantly changed.
Land-based (or terrestrial) positioning systems use time difference and trilateration techniques to determine a position.

Electronic Positioning Systems (EPS)

In general EPS are classified according to their operating frequencies or the bandwidth, see Table 7.2 of Appendix 3 to Chapter 7, which determine the operating range and accuracy, and thus a system's applicability for a particular type of work. In general, the higher the frequency of the system and the shorter the wavelength, the greater the achievable accuracy possible in the resolved position, see Table 7.2 of Appendix 3 to Chapter 7.

Medium-frequency Electronic Positioning Systems (RAYDIST/DECCA):

a. systems were first developed in 1950s but they are no longer used;
b. systems operated by time/phase differencing methods, resulting in either circular or hyperbolic lattices (time differences);
c. systems required repeated calibration to resolve whole-wavelength (lane) ambiguities and continual monitoring during the course of the survey to resolve lane or cycle slips, similar to the integer ambiguity determination requirements for modern day DGPS;
d. onsite calibration was essential to maintain accuracy, but in very far offshore surveys calibration was often impossible;
e. visual positioning techniques were used to calibrate these systems.

Low-frequency Electronic Positioning Systems (LORAN-C):

a. the primary marine and airborne navigation system for over 40 years;
b. a low-frequency time-differencing hyperbolic system;

c. suitable only for general navigation or reconnaissance surveys (Order 3 when calibrated);
d. daily near-site or onsite calibration is critical if any semblance of absolute accuracy is to be maintained;
e. absolute positional accuracy is about + 450 metres (+ 0.25 mile) at best, without onsite calibration.

Range-Range EPS

These microwave EPS (hyperbolic or circular) were introduced in the 1970s and remained the primary positioning system up until the mid 1990s, thereafter their use declined when differential GPS techniques became available for large areas. Nowadays microwave EPS (Range/Range) are still in use in those areas where poor GPS signal coverage occurs.

a. Trilateration is the process performed by the Range/Range microwave EPS when determining the co-ordinates by intersection of measured distances from two (or more) shore control points:

   i. a Circular Line of Position (CLOP) is associated with the distance from each shore station;
   ii. each pair of CLOP generates two intersection points, which are generated each side of the connecting baseline between the two shore stations points;
   iii. each EPS uses its own method to solve this ambiguity, by orientation to initial reference point co-ordinates or by referencing the calculated position relative to the azimuth of the baseline;
   iv. initially EPS distances were visually observed and steered, performing manual data logging in a workbook or field book with manual fixing on a plotting sheet;
   v. modern EPS use built-in automated data acquisition systems that log the ranges and compute the relative positions, subsequently sending this data to a helmsman display unit and a track plotter;
   vi. present-day EPS, and also GPS systems, transmit the raw digital data to a PC running a suitable Acquisition and Control Process package, capable of synchronizing position and sounding in logged data files, whilst performing real-time positioning QC and tracking the position in several windows selected on the main operator and helmsman monitors.

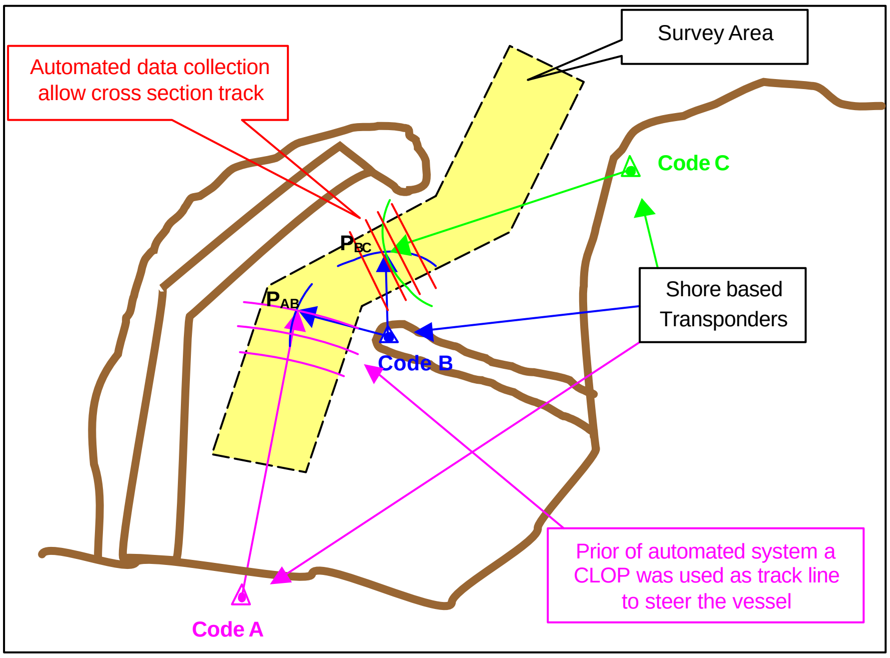


Fig. 7.7 "Two Range Intersection"

b. Constant range tracking provides a good backup capability should failure occurred in the automated positioning and guidance system. When no automated techniques are available, the vessel follows a CLOP:

i. by keeping a constant range from one reference station;
ii. by fixing at range intercepts from the other reference station;
iii. by proceeding at low speed, to produce a better accuracy in positioning and ease the helmsman's task of following the curve range pattern;
iv. survey lines are circular, not aligned to the project co-ordinate system and often not orthogonal to the main bathymetry.

c. Automated range-range tracking:

i. the range intersection co-ordinates are automatically computed from the precise co-ordinates of the shore stations;
ii. the point co-ordinates are transformed relative to the project alignment co-ordinate system (station-offset);
iii. analogue/digital course indicators or left-right track indicators receive positional data, allowing the tracking of any cross section or offset range;
iv. position fixes are taken manually by the observer from a receiver or track plotter, recording co-ordinates on in suitable workbook;
v. at each fix, the depth is marked on the paper trace of the analogue echo-sounder and the value noted in the workbook;

vi. the correlation between position and sounding will be performed during the post-processing stage;
vii. digitised depth data are correlated in real time with positional data in a data acquisition software system at regular preset intervals.

Range-Range Accuracy

The intersection accuracy is a function of two factors:

a. The range accuracy of distances (or standard deviation σ);
b. The angle of intersection that varies relative to the baseline, positional accuracy varies as the vessel changes its position in the survey area.

Quality Control

Main quality control criteria to be considered on Microwave EPS accuracy:

a. α angle of intersection has major effect on position determination and should be between 45° and 135°;
b. σ isn't constant with distance from a shore station and in general is in the order of ±3 m rather than the ±1 m or 2 m stated by manufacturers for ideal or well calibrated conditions;
c. the average positional accuracy (σ ±3 m) can vary from 5 to 10 metres.

Multiple Range Positioning

Multiple range positioning techniques. (ie. Racal Micro Fix, Sercel Syledis, Motorola Falcon VI)

The position is determined from the computed co-ordinates of the intersections of three or more simultaneously observed range circles.

The CLOPs don't intersect at the same point because each range contains observational errors:

a. three different co-ordinates result from three observed ranges and six separate co-ordinates result from four observed ranges;
b. an adjustment of these co-ordinates gives the final position and is normally performed on-line at each update cycle generally by a least-squares minimisation technique or simply by the strongest angle of intersection or not weighted average of all the intersecting coordinates.

The positional data are then transformed to a project-specific co-ordinate system as described for a two-range system:

a. the use of multiple ranging minimises uncertainties with the vessel position being obtained by adjustment of these ranges to a best fit; an on-line accuracy assessment is accomplished by evaluating the positional intersection misclosure of CLOPs which contain errors (see Figure 7.8);

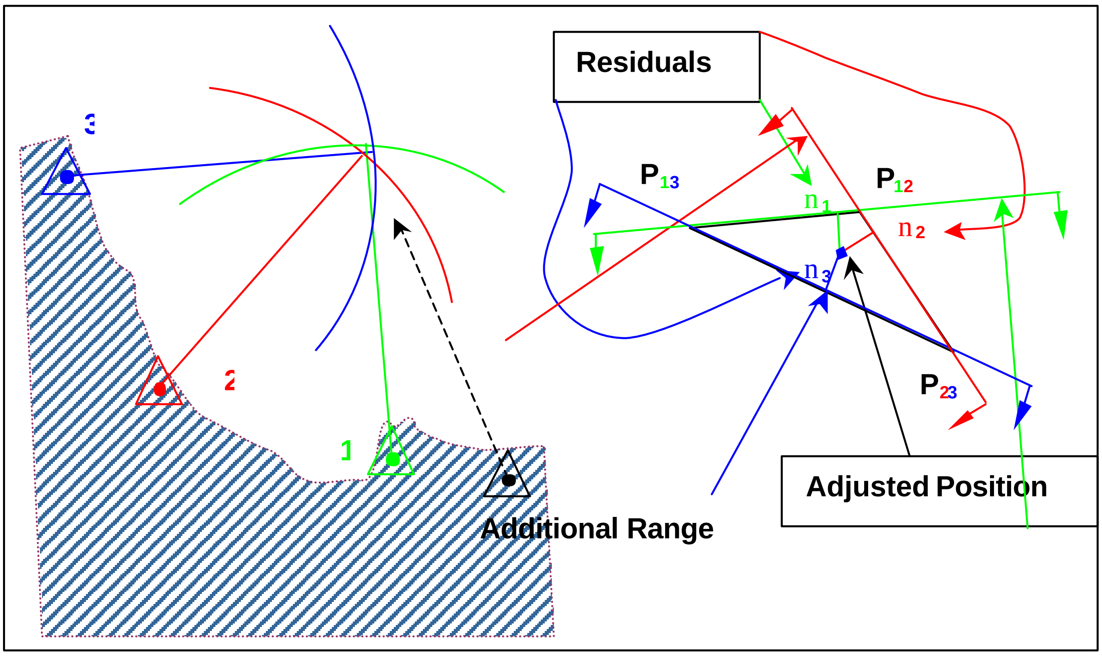


Fig. 7.8 " Multiple CLOPs Intersection"

b. an assessment of the range measurement accuracy may be obtained by computing the residual range errors (*v*) for each position (automated software, using a least-squares type of adjustment, can provide an accuracy estimate of the positional RMS error at each position update);

c. automated EPS give an alarm when RMS exceed prescribed limits, assuming a constant initial standard error within the survey area.

Calibrations and Quality Control

The process of calibrating a microwave EPS is performed by the following basic steps:

a. independent determination of the vessel's antenna location;
b. comparison of differences between the observed microwave distances and the distances computed from an independent measuring system (if observing the direct distances);

or

b. comparison of differences between the observed microwave co-ordinates and those computed from the independent system;
c. performing a series of independent calibrations (*Repeated Calibrations*), the correction to the EPS, to be applied a the console or stored into position computation software, is represented by the mean range difference.

The systems and methods used to perform an independent calibration would include:

a. EDM calibration – series of EDM distance readings is directly compared with the simultaneously observed microwave ranges and corrections are then applied;

b. baseline calibrations – the simplest microwave EPS calibration method, the survey vessel is positioned at a point on the baseline between two shore stations and the computed distance is compared with the combined observed ranges from the microwave system, range corrections are computed and applied. This method should be repeated at various points on the baseline and should be performed between all pairs of shore stations;

c. total station instrument calibration – the observed co-ordinates of an automated positioning system can be directly compared with the more accurate co-ordinates obtained from total station EDM measurements;

d. triangulation intersection – the most accurate method of microwave calibration is performed in a dynamic environment. Three theodolites are used for this high-accuracy triangulation calibration, a series of 5 to 10 measurements, intersection fixes, to the moving survey vessel. For each series of measurements, the triangulated positions are computed, inversed, and converted to grid distances which are compared against the simultaneously observed microwave ranges. An estimate of the statistic validity of the mean range difference must be calculated as previously explained;

e. sextant resection – this method is valid only when resection geometry is ideal, near the shore line and a very slow vessel motion. A numbers of simultaneous resection angles (5 to 10) and corresponding microwave EPS distances are observed with three sextants centred near the EPS antenna to minimise eccentric errors. Resection computation should performed using suitable software providing a quality indication of the resection based on the geometry and estimated standard error of the observed angles, to judge whether applying a mean correction to the range is statistically appropriate.

f. General QC criteria for EPS:

   i. the static calibration does not simulate the dynamic survey condition;
   ii. the calibration must be performed within or close as possible to the survey area, to simulate the real conditions within the project area;
   iii. the accuracy of the independent calibration procedure followed must be better or at least equal to the calibrating microwave EPS;
   iv. the multi-path residual effects can be reduced but not eliminated by calibration procedures due to survey vessel antenna location and orientation;
   v. calibrations of pulsing microwave EPS are valid only for the particular range measurement system used;
   vi. the more accurate measurement systems used to calibrate EPS must also be independently checked, or verified, to prevent blunders (GPS, Total station, theodolites, etc);
   vii. calibration procedures must be consistent during the course of a project.

Some of these basic criteria, described for performing EPS calibrations, are also applicable to GPS positioning techniques.

4.1.6.7 Global Positioning System (GPS)

General Description

During the 1990s the Global Positioning System (GPS) has become the worldwide standard positioning and navigation system and has replaced almost all other techniques. Poor GPS satellite coverage only occurs in isolated instances over relatively small areas, in these cases the traditional terrestrial methods will need to be employed. Differential GPS systems allow for world wide coverage, they do not always require the site selection/deployment effort required for terrestrial systems however careful pre-survey calibration and post survey validations are still required; accuracies now exceed those of any other hydrographic survey positioning system.

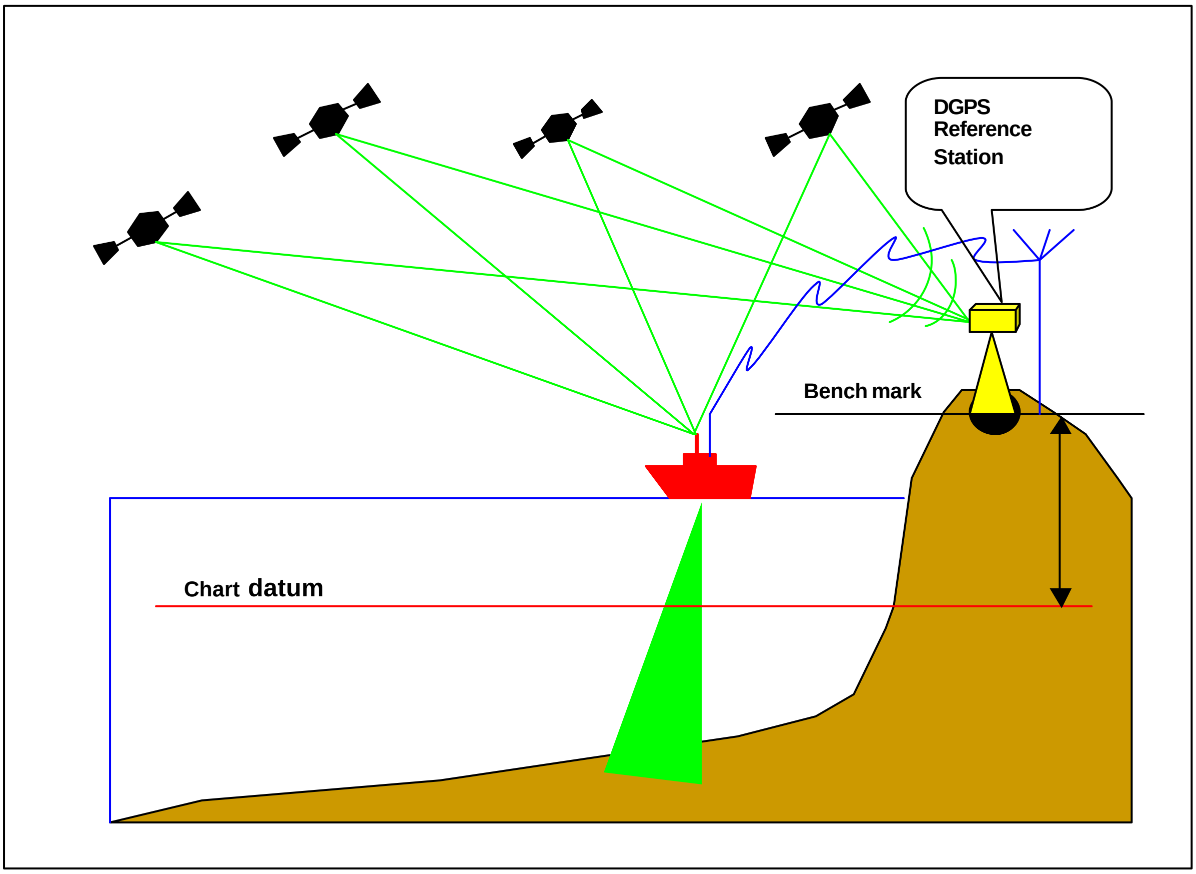


Fig. 7.9 "Differential GPS positioning of a hydrographic survey vessel"

The system consists of two absolute positioning services, the Standard Positioning Service (SPS) and the Precise Positioning Service (PPS):

a. SPS is available to civilian users, using the C/A-code on the L1 carrier, it provides absolute accuracies of 10-30 meters in absolute positioning mode;

b. PPS was developed for the U.S. military and other authorized users; it uses the P(Y)-code on the L1 and L2 carriers providing an absolute accuracy of 5-15 meters in absolute positioning mode.

For many applications, this absolute positioning does not provide sufficient accuracy. Differential GPS (DGPS) is a technique which can provide relative positioning with an accuracy of a few meters using code phase measurements to a few millimetres with carrier phase measurements. DGPS requires two or more GPS receivers to be recording measurements simultaneously and processing software to reduce or eliminate "common errors". With a reference system positioned at a known survey control point, DGPS data can be used to determine baselines between stations and establish the positions of other receivers within the same reference system. Code and carrier phase DGPS, when operating in a differential mode, can be tracked in real-time for the positioning of moving platforms, dredges, survey boat and vessels (see Figure 7.9) to provide real-time information at accuracies required for hydrographic surveying and/or dredge positioning.

Tracking Techniques (see paragraph 6.1.4.1 of Chapter 2)

Phase tracking techniques are:

a. carrier phase tracking;
b. code phase tracking.

Accuracies

The absolute range measurement accuracies or standard deviation achievable with GPS depends on:

a. the type of code used (C/A or P);
b. a three-dimensional (3-D) confidence ellipsoid describing the uncertainties in all three geocentric co-ordinates, when coupled with the GDOP of the satellites during the position determination;
c. the time and the location of the changing satellite geometry.

The nominal accuracy statistics for a GPS user are defined by error propagation techniques. The user range measurement accuracies refer to geocentric co-ordinates, which can be transformed to a local datum, and to a 3-D covariance matrix, which defines and assess the dimensions (direction or co-ordinate) of the error ellipsoid in the reference system.

The more usual methods to describe error measures are listed below:

a. 2-D (horizontal) GPS positional accuracies are normally estimated using a Root Mean Square (RMS) radial error statistic;

b. 3-D GPS accuracy measurements are most commonly expressed by Spherical Error Probable, or ***SEP***. This measure represents the radius of a sphere with a 50% confidence or probability level and the spheroid radial measure only approximates the actual 3-D ellipsoid, representing the uncertainties in the geocentric co-ordinate system.

c. for 2-D horizontal positioning, ***CEP*** (Circular Error Probable) is commonly used as the probable or statistic error measure, represented the radius of a circle containing a 50% probability of position confidence;

Accuracy Comparisons

It is important that GPS accuracy measures clearly identify the statistics from which they are derived. A "20 meter" or "5 meter" accuracy statistic is meaningless unless it is identified as being 1-D, 2-D, or 3-D, together with an applicable probability level. In addition, absolute GPS point positioning accuracies are defined relative to an earth-centred co-ordinate system/datum. This co-ordinate system will differ significantly from local project or construction datum. Nominal GPS accuracies may also be published as design or tolerance limits, and actual accuracies achieved can differ significantly from these values.

Relative Accuracy Measures

Hydrographic surveys performed according to IHO/S-44 Standards are concerned with absolute world-wide (j, l, h) positional accuracy at 95% confidence level, but normally engineering, construction, and dredging surveys are concerned with local project coordinates (X, Y, h), and with ensuring high accuracy within a local construction project. The relative accuracy measure is expressed typically in parts per million (ppm) as a function of the distance between two points or receivers and usually given at the one-sigma standard error (or 68% standard deviation) level.

Dilution of Precision (DOP)

GPS errors resulting from satellite configuration geometry can be expressed in terms of Dilution of Precision (DOP), which is the geometric strength of the configuration of satellites observed during the survey session. In mathematical terms, DOP is a scalar quantity used in an expression of a ratio of the positioning accuracy (or of the standard deviation of one co-ordinate to the measurement accuracy). DOP represents the geometrical contribution of a certain scalar factor to the uncertainty (i.e., standard deviation) of a GPS measurement. In 2-D (horizontal) positioning it refers to the HDOP factor.

Reference Datum

In general differential survey methods are concerned with relative co-ordinate differences, but in absolute positioning and for navigation purposes we must consider variations within a global reference system used by the NAVSTAR GPS. Thus, GPS co-ordinate differences or transformations from the World Geodetic System 84 (WGS 84) reference system must be applied to any type of local reference datum. On the North American continent the co-ordinates of WGS 84 are highly consistent with North American Datum 83 (NAD 83). The European Terrestrial Reference Frame (ETRF89) is a realisation of WGS 84 for the European Continent. Each nation in Europe has determined their own transformation to tie the ETRF to the local datum.

Error Sources (see paragraph 6.1.3 of Chapter 2) **and Calibration Requirements**

The accuracy of GPS is a function of errors and interferences on the GPS signal and the processing technique used to reduce and remove these errors. Similar to range-range microwave systems, GPS signals are highly affected by humidity and multi-path; additional errors are caused by the 20,000 km path through the ionosphere and troposphere layers. It should be remembered that satellite signals can be altered for US national security reasons by the use of Anti Spoofing (AS). Differential techniques close to a reference station can eliminate most of these errors, however the further the remote operates from the reference station, the less similar will be the errors received by both receivers.

DGPS operation has no prescribed calibration requirements (check list), unlike microwave or R/A systems; the major blunders to check for are:

a. incorrect project datum or geodetic reference datum;

b. incorrect master station co-ordinate values;
c. incorrect measurement of antenna height values;
d. DGPS mode not selected in the unit;
e. RTCM-104 input/output format not selected.

Positioning Methods

Two general operating methods, used to obtain GPS positions for dynamic horizontal control, have a variety of applications for hydrographic surveys at sea:

a. absolute point positioning;
b. relative positioning (DGPS).

In general, absolute point positioning involves only a single passive receiver and is not sufficiently accurate for precise surveying or hydrographic positioning requirements. It is, however, the most widely used military (PPS) and commercial (SPS) GPS positioning method. Relative (Differential) positioning requires at least two receivers and can provide the accuracies required for basic land surveying and offshore positioning.

Absolute Point Positioning (Pseudo-Ranging)

The GPS receiver generates a navigation solution by pseudo-ranging, measuring an approximate distance (pseudo-range) between the antenna and the satellite by correlation of a satellite-transmitted code and a reference code created by the receiver, no corrections are made for errors in synchronization between the clocks in the transmitter and the receiver. The distance which the signal has travelled is equal to the velocity of the transmission from the satellite multiplied by the elapsed time of transmission. Additional delays (errors), which can affect positional accuracy, are caused by tropospheric and ionospheric conditions. To create a GPS 3-D position, at least, four pseudo-range observations are required to resolve the constant clock biases (t) contained in both the satellite and the ground-based receiver.

The solution of four pseudo-range observation equations, containing four unknowns (X, Y, Z and t.), gives the solution for the 3-D position of a point (for a 2-D location only three pseudo-range observations are needed), it is highly dependent on the following accuracies:

a. the accuracy of the known co-ordinates of each satellite (i.e., X_s, Y_s, and Z_s);
b. the accuracy of the modelled atmospheric delays (d);
c. the accuracy of the resolution of the actual time measurement process performed in a GPS receiver (clock synchronization, signal processing, signal noise, etc.);
d. the accuracy of an absolute point position is a function of the range measurement accuracy and the geometry of the satellites (DOP).

Dilution of Precision (DOP) is a description of the geometrical contribution to uncertainty in a GPS determined point position and is roughly related to the physical orientation of the satellites relative to the ground receiver along with the range measurement accuracy.

Static Solution - as with any measurement process, repeated and redundant range observations to the satellites at varying orientations will enhance the overall positional accuracy and reliability. In static mode with a stationary GPS antenna, range measurements to each satellite may be continuously measured over the varying orbits of the satellite(s). The altering satellite orbits create changing positional intersection geometry over the same ground position. In addition, simultaneous range observations to numerous satellites can be adjusted using weighting techniques based on the strength of intersection and pseudo-range measurement reliability.

Dynamic Solution - In dynamic mode where the GPS antenna is moving, range measurements to each satellite are unique due to the altering orbital locations of the satellite(s). The varying satellite orbits and the vessel speed cause changing positional intersection geometry over the moving GPS antenna position.

The NAVSTAR GPS satellite system gives two levels of absolute positioning accuracy:

a. *Standard Positioning Service (SPS)*. The SPS is capable of achieving real-time 3-D absolute positional information in the order of 10-30 m (95% confidence level on horizontal accuracy). US DoD has implemented Anti-Spoofing (AS), which interchanges the P code with a classified Y code, thus denying the SPS user the higher P code accuracy;

b. *Precise Positioning Service (PPS)*. Non-military PPS users must be authorized by the US DoD to have a decryption device capable of deciphering the encrypted GPS signals. This authorization must be obtained from the National Security Agency (NSA). USACE is an authorized user; however, actual use of the equipment has security implications. PPS users can obtain real-time absolute 3-D positional accuracy in the order of 16 m SEP (or 5-15 m at 95% confidence level on horizontal accuracy).

The US DoD security action does not significantly impact a hydrographic user operating in a differential positioning mode.

Absolute positioning (SPS/PPS) only provides real-time absolute positional accuracies and will not satisfy the IHO/S-44 hydrographic surveying requirements for Special Order and Order 1. It does have general navigation applications and will eventually replace LORAN-C and other navigation systems for ships and aircraft.

Differential Point Positioning (DGPS)

Differential positioning is the technique used to position one point relative to another, both receiving stations simultaneously observing the same satellites. Since errors in the satellite position (X_s, Y_s, and Z_s) and atmospheric delay estimates (d) are effectively the same, they can be ignored to a large extent. This method can be performed by using code or carrier phase measurements and can provide results in real-time or post processed.

a. *DGPS Code Phase tracking*. The technique consists of two GPS receivers; one set up over a known point and one moving from point to point or placed on a moving survey vessel, measuring pseudo-ranges to at least four common satellites. Since the positions of the satellites are known and one of the receivers is over a fixed known point, a “known distance” can be computed for each observed satellite. This “known distance” can then be compared against the “measured distance” (or Pseudo Range) to obtain a Pseudo Range Correction (PRC), which is computed for each satellite being tracked at the fixed point. Each PRC can then be applied to the moving or remote receiver to correct the measured distances. Code phase tracking has primary applications to real-time positioning systems with meter-level accuracies. It is sufficient for hydrographic survey positioning which meets IHO S-44 requirements for Order 1 surveys, since meter-level positioning suffices for the vast majority of these purposes.

b. *DGPS Carrier Phase tracking*. This is the most accurate GPS survey technique and the relative positional accuracies are of the order of two to five parts per million (ppm) between two GPS receivers (one at a known reference point and the other at an unknown location or aboard a moving platform). The tracking method uses a similar formulation of pseudo-ranges used in code phase tracking systems described above, but with a more complex process when the carrier signals are tracked. The short wavelength (19 cm) necessitates the adding of an ambiguity factor to the solution equations to account for the unknown number of whole carrier cycles over the pseudo-range. Carrier phase tracking provides for a more accurate range resolution due to the short (19 cm) wavelength and the ability of a receiver to resolve the carrier phase down to about 2 mm. This method is referred to as real-time kinematic or RTK and provides 3D positions accurate to a few centimetres over ranges up to approximately 20 kilometres. It is applicable to hydrographic survey positioning and meets IHO/S-44 requirements for Special Order surveys and may be employed with either static or kinematic receivers.

c. *Advantage* of the code phase (DGPS) over the carrier phase(RTK):

  i. wavelengths are much longer than the carrier wavelengths, eliminating the ambiguity problem.

d. *Disadvantage* of the code phase (DGPS) over the carrier phase(RTK):

  i. longer wavelengths decrease the system accuracy;
  ii. longer wavelengths are more affected by signal multipath.

Real-Time Dynamic DGPS Positioning System (Code Phase)

The system in general includes:

a. reference station equipment (master);
b. communications links;
c. rover station equipment (remote user).

There are several DGPS services that provide real-time pseudo-range corrections:

a. radio beacon navigation services (Beacon IALA System);
b. commercial satellite subscription services;
c. commercial land base DGPS network services (telephone or mobile phone links);
d. local DGPS systems.

Local DGPS systems are normally installed or used by the agency responsible for a survey, where other services do not provide coverage or sufficiently good accuracy to meet survey requirements.

Reference Station

The reference receiver consists of a GPS receiver, antenna, and processor that:

a. is placed at a known co-ordinated station with an unobstructed view of the sky from at least 10° above the horizon;
b. GPS antenna should be erected clear of objects likely to cause multi-path or interference (avoid areas with antennas, microwave towers, power lines, and reflective surfaces);

c. measures timing and ranging information broadcast by the satellites;
d. computes and formats, every 1 to 3 seconds, the pseudo range corrections (PRC) then transmits, broadcasting via the communications link, to the operator's equipment in the offshore vessel; the recommended data format is that proposed by the Radio Technical Commission for Maritime Services (RTCM) Special Committee 104 v 2.0;
e. performs QC functions and determines the validity and quality of the computed PRCs.

Communications Link

a. the communications link is used as a transfer medium for the differential corrections, the type is dependent on the user's requirements and the minimum transmission rate should be 200 bits per second (bps);

b. communication links operating at Very High Frequency (VHF), Ultra High Frequency (UHF) and High Frequency (HF) are viable systems for the broadcast of DGPS corrections, with ranges extending out from 20 to 50 km (VHF/UHF) and up to 200 km (HF), depending on local propagation conditions and site elevation. The disadvantages of UHF and VHF links are their range being restricted to line of sight and the effects of signal masking from islands, structures and buildings, multi-path and licensing issues;

c. communication links require a reserved frequency of operation to avoid interference with other activities in the area, all frequencies need authorisation for use within each nations geographical area of responsibility;

d. several companies offer subscriptions for satellite communications, telephone or mobile phone communications systems, capable of being used for the transmission of the PRCs;

e. satellite and phone communication systems are less limited in range than UHF/VHF systems but are usually higher in price.

User Equipment

Using the technology of differential pseudo-ranging, the position of a survey vessel can be calculated relative to the reference station with a receiver (the user equipment) that consists of real-time code phase tracking DGPS, antenna and processor:

a. should be a multi-channel single frequency (L1) C/A code GPS receiver;
b. is capable of receiving the differential corrections from the communications link in the RTCM SC-104 V.2.0 format and then applying those corrections to the measured pseudo-range;
c. receiver update rate should be 1 to 3 seconds;
d. output from the rover receiver should be in the NMEA-183 format as the most widely used format for input into hydrographic survey software packages;
e. equipment should be capable of maintaining positional tolerances for surveys at speeds up to 10 knots;
f. receivers should not bias the position during vessel turns due to excess filtering.

Separation Distances

a. the differential tropospheric and ionospheric corrections are not presently applied to internal solutions of most GPS receivers and these errors contribute to horizontal position errors on average 0.7 m per every 100 km;
b. the type of data link in use will be a limiting factor on the separation distance between the reference station and mobile receiver, the reference station may need to be moved from one point to another so that the minimum separation requirements are maintained.

Satellite Geometry

The Horizontal Dilution of Position (HDOP) is the critical geometrical component that:

a. in Order 1 and 2 surveys HDOP <5;
b. the 24 Block II GPS satellites constellation maintains a HDOP of approximately 2 to 3 most of the time.

Other DGPS Services (Radiobeacon Navigation Service and Commercial WAAS)

Radiobeacon Navigation Service

The main function of the Radiobeacon Navigation Service is to provide aids to navigation in navigable water covered by the service; the object is to substitute Loran-C and Omega systems, which were used as the primary navigation systems for offshore marine navigation, with full coverage from GPS for more accurate positioning. Many nations have commissioned real-time positioning systems for their coastal areas, rivers and lakes regions, utilising DGPS and marine radiobeacon technology; there is a desire for other maritime governments to expand the coverage to all offshore waterways and eventually have entire world coverage.

a. System set-up and configuration:

   i. GPS Radiobeacon:
      - N. 2 GPS L1/L2 geodetic reference station receivers with independent geodetic antennas to provide redundancy and a marine radiobeacon transmitter with transmitting antenna;
      - N. 2 combined L1 GPS/Modulation Shift Key (MSK) receivers used as integrity monitors, each one utilises an independent GPS antenna and a MSK near-field passive loop antenna.

b. Site location:

   i. the location of the reference stations' GPS antennas are known co-ordinated geodetic control points based on the ITRF [i.e. ETRF (European Terrestrial Reference Frame) datum for Europe and NAD 83 (North American Datum of 1983) datum for USA/Canada];
   ii. GPS C/A-code pseudo-range corrections are computed and transmitted via a marine radiobeacon;

iii. the system on board user vessels consists of a marine radiobeacon receiver and a GPS receiver (or an integrated GPS/Radiobeacon receiver) with the ability to accept and apply pseudo-range corrections; with accuracies of less than 5 m achievable dependent on the type and quality of the user's GPS receiver, distance from the reference station and the satellite geometry.

c. Data transmission (data types):

i. the corrections and other information are transmitted using the Type message of the Radio Technical Commission for Maritime Services Special Committee 104 (RTCM SC-104) version 2.1 data format;
ii. more detailed descriptions of these message types are explained in the Broadcast Standard available for the Radiobeacon Navigation Service of each nation;
iii. corrections are generated for a maximum of nine satellites tracked by the reference station GPS receiver at a minimum elevation angle of 7.5° above the horizon; if more than nine satellites are observed above 7.5°, the corrections are based on the nine satellites with the highest elevation angles;
iv. satellites below 7.5° elevation are masked due their susceptibility to multipath and spatial decorrelation;
v. corrections are normally transmitted at a 100 or 200 baud rate;
vi. corrections can be considered valid for a period of 15 seconds from generation;
vii. using corrections more than 30 seconds old, particularly for positioning of a moving platform, may cause spikes.

d. Availability and reliability of the system:

i. the system maintains a broadcast availability that exceeds 99.7%, in designed coverage areas, assuming a healthy and complete GPS constellation;
ii. signal availability, in most areas, will be higher due to the overlap of broadcast stations;
iii. each site is equipped with two integrity monitors (i.e. a GPS receiver with a MSK radiobeacon) that are mounted over known positions, they receive the pseudo-range corrections from that site and compute a position that is compared to the known location to determine if the corrections are within the expected tolerance;
iv. the corrected positions calculated by the integrity monitors are sent via phone lines to the control monitoring stations, which notify users via the type 16 message of any problems with a radiobeacon site within 10 seconds of an out-of-tolerance condition.

e. Coverage:

i. an updated coverage map can be found at the Radiobeacon Navigation Service web site of each participating nation, under the DGPS section.

f. User requirements and equipment:

i. to receive and apply the pseudo-range corrections generated by the reference station, the user needs to have a MSK Radiobeacon receiver with antenna and, at a minimum, a L1 C/A code GPS receiver with antenna or a combined MSK radiobeacon and GPS receiver with a combined MSK and GPS antenna, a more expensive option.

The MSK receiver demodulates the signal from the reference station and generally will automatically select the reference station with the strongest signal strength to track or it will allow the user to select a specific reference station. Since the reference station generates corrections only for satellites above a 7.5° elevation, satellites observed by the user's GPS receiver below a 7.5° elevation will not be corrected.

g. Position QC tolerance checks and calibrations:

i. most precise DGPS augmentation systems are capable of providing sub-meter accuracies at reasonable distances from the nearest reference station, however, at increasing distances spatial decorrelation errors (due to differing ionospheric/tropospheric conditions) can induce systematic positional biases;
ii. in general, under nominal atmospheric conditions, less than 5 m RMS (95%) positional accuracy may be achieved at distances upwards of 200 miles;
iii. to confirm positional accuracy is better than the 5 m tolerance, a static check position should be obtained by occupying a known survey point near the project area;
iv. when operating with the Radiobeacon Navigation Service, static positions should be observed from different radiobeacon reference stations to ascertain if positional systematic biases are present, in practice the closest beacon will normally be with minimal biases;
v. when large or ambiguous positional biases occur in a project area, it may be necessary to establish a local DGPS network (code or RTK carrier) for observing different beacon positions in a static comparison;
vi. a similar process should be performed when using Commercial DGPS Wide Area Augmentation Systems (WAAS, GLONASS, EGNOS, GALILEO, MSAS etc).

Real-Time Dynamic RTK DGPS (Carrier Phase)

General

The DGPS carrier phase is capable of yielding decimetre accuracy on a moving vessel within the geographic area (20 km.), both horizontally and vertically. This technology, known as ‘On The Fly’, can provide real-time elevations of survey vessels.

Current kinematic techniques allow for the ambiguities to be resolved while the receiver is in motion and provides accuracies in the range from 2 to 5 cm. This method of carrier phase positioning is commonly referred to as real-time kinematic or RTK surveying.

A real-time kinematic (RTK) DGPS positioning system is based on DGPS carrier phase technology similar to the kinematic techniques currently used for static GPS geodetic surveys where millimetre level accuracies are achieved. RTK procedures allow for the movement of a GPS receiver after the initial integer ambiguity (i.e., whole number of wavelengths) between satellites and receiver has been resolved, as outlined in Chapter 2.

Accurate real-time elevations (and depths related to the GPS antenna height) can be directly obtained without observing tidal measurement data, if adequate motion compensation sensors are used and project tidal datum modelling has been undertaken (see Figure 7.10).

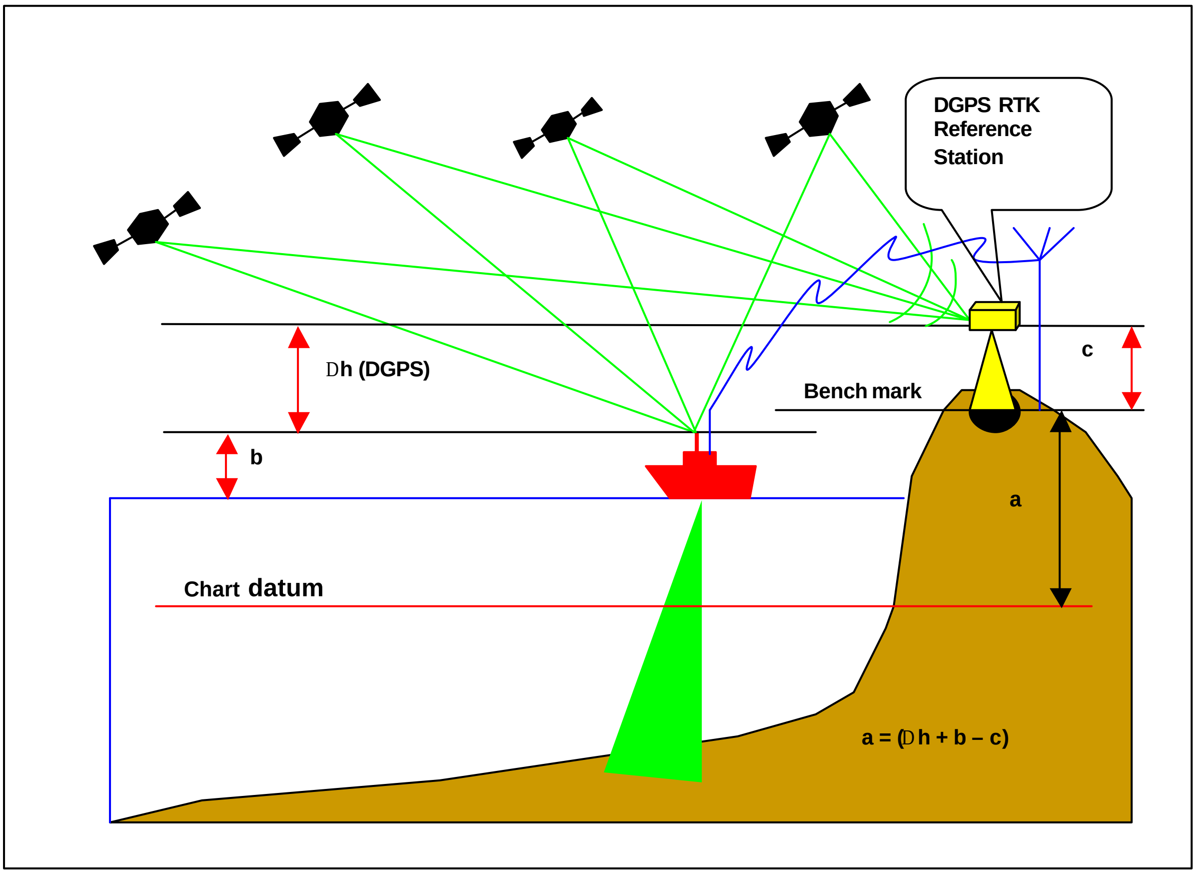


Fig. 7.10 "Principle of Real Time Kinematic DGPS elevation determination"

If elevations are obtained through the use of RTK DGPS techniques, Geoidal-ellipsoidal modelling and tidal modelling procedures are mandatory and should be performed before RTK surveys can be conducted.

Reference Station

The carrier phase positioning system is very similar to the current code phase tracking technology. A shore GPS reference station is located over a known survey mark tied to the local horizontal/vertical geodetic reference framework; however, the reference station must be capable of collecting both pseudo-range and carrier phase data from the GPS satellites and will consist of:

a. a carrier phase dual frequency receiver capable of full wavelength L1/L2 with cross-correlation technique during times of P-code encryption tracking,
b. a GPS rover receiver with its associated antenna and cables, high-speed processor and communications link capable of, at least, a 1 second update rate;
c. the location will be the same as for a code phase tracking DGPS system;
d. the processor will compute the pseudo-range and carrier phase corrections and format the data for the communications link;
e. the corrections will be formatted in the RTCM SC-104 v.21 format (CMR) for transmission to the rover receiver.

Communications Link

a. the carrier phase positioning system requires a minimum data rate of 4800 baud and differs in the volume of data which has to be transmitted from the code phase tracking DGPS system, which requires a baud rate of 300;
b. this high data rate limits the coverage area for high-frequency broadcast systems;
c. VHF and UHF frequency communications systems are well suited for this data rate.

User Equipment

The user equipment on the survey vessel is:

a. a carrier phase dual-frequency full-wavelength L1/L2 GPS receiver with a built in processor, capable of resolving the integer ambiguities while the vessel is underway;
b. an associated geodetic GPS antenna, that reduces the effects of multipath on the GPS signal;
c. a communications link to receive data from the reference station;
d. the minimum update rate from the reference station to the vessel(s) should be 1 second;
e. the position output (NMEA 183) carrier phase DGPS data should enable the performance of real-time navigation and recording of the vessel's true position needed for survey processing.

RTK systems are not designed to be used for surveys in excess of 20 km from the reference station.

Ambiguity Resolution

a. if the system remains in the RTK mode, real-time sub-decimetre 3D positioning should be available from the rover receiver;
b. both reference station and the remote station receivers must maintain lock (continuous GPS data) on at least four satellites;
c. if the number drops to below four satellite, the ambiguities will again be resolved after the system reacquires lock on a sufficient number of satellites, the system will work in DGPS mode or Autonomous mode during this period.

4.2 Vertical Control and Calibration

4.2.1. General Description

The datum to which depths are to be reduced is fundamental to any bathymetric survey and the Hydrographic Specification will contain full details of how this is to be established together with details of established benchmarks. If the datum is not defined, the existing Chart Datum should be used if at all possible.

The necessity to either establish a new datum or to transfer datum should be carefully considered, any new or transferred datum must be related to the local survey datum through existing or newly established benchmarks, for which full details should be recorded and rendered to the Hydrographic Office. Particular care is required for surveys in rivers and river estuaries; guidance is available in the Admiralty Tidal Handbook Volume 2.

Using the data supplied with the Hydrographic Specification, the location for tidal stations should be determined. If conducting a resurvey, the tide station should be established in the position of the old station if at all possible. If multiple stations are required the distance between stations should not be too great and in any case no more than 10 miles. The Hydrographic Specification will provide guidance on the placement of offshore gauges.

Data to assist in the creation of a co-tidal chart will be included in the Hydrographic Specification. A co-tidal and co-range chart should be produced as described in chapter 5.

Pre-calibrated tide poles and gauges should be established at the desired locations. The tide poles should be connected to the sounding datum via the land levelling system and witness marks installed as a future means of quick visual check on the integrity of the pole. If no local benchmarks are available, for whatever reason, at least 2 new marks should be established and their details fully recorded.

A comparison of tide pole and tide gauge readings over a 25 hour period should be taken to both establish sounding datum on the gauge and to ensure its correct operation. Thereafter checks should be conducted at regular intervals during the survey.

Calculation of Mean Sea Level (MSL) using 39 hours of observation should be conducted at the beginning and end of the survey. Due to daily atmospheric and weather influences, results should be within 0.3 metres of MSL quoted in the Tidal Tables, which will provide additional confidence in the observed tidal data.

When an established gauge is used, the setting must always be checked independently to ensure that the zero corresponds to the stated figure.

Observed tidal data should be inspected each day to ensure that the observations meet the Hydrographic Specification standards. Whenever possible continuous tide readings should be obtained for the entire duration of the survey. Where continuous readings are not obtained, care should be taken at the start and end of each survey period with co-tidal time differences to ensure that the tidal data covers sounding operations.

4.2.2 Tidal Modelling for RTK Surveys

The survey area must have details of an appropriate tidal datum to meet the requirements of the project to be undertaken. The reason for establishing a tidal datum in the survey area is to update knowledge of MLLW (or Chart) Datum and to enable the benefits of RTK DGPS technology to be realised by performing the survey without using tide gauges.

The main requirements are:

a. perform wide area GPS static surveys in the selected area;
b. install sufficient tide gauges in the area to obtain details of tidal datum at these gauge sites computed form long observation periods of data;
c. perform GPS tidal measurements in the survey area at the same time to obtain a comparable data set of GPS water measurements against conventional tide gauge measurements;
d. anchor a survey vessel fitted with a RTK Rover Receiver for 25 hour periods in sufficient locations to generate intermediate datum points within the area, to allow correlation between the conventional tide gauge methods and the GPS tidal datum method, and to check any changes in ellipsoid heights between the RTK stations and the gauge sites over a full tide cycle of 28 days;
e. use a suitable software configuration in the hydrographic survey package which allows for the ellipsoid separation values to MLLW to be used to compute tidal height measurements from the waterline of the survey vessel.

The whole project area must be related to the tidal measurements from the nearest primary tide gauge used to measure the Mean Lower Low Water (MLLW) for the area, also a reference is needed to incorporate tidal datum measurements performed in the survey area. The GPS ellipsoid reference frame and the local vertical datum must be used over the entire survey area.

Tidal Datum Diagrams

Two different tidal data can be realised:

a. a traditional tidal relationship for the area is represented by the Mean Lower Low Water Tide Surface relative to the local vertical datum, which must provide the MLLW reference with an acceptable tolerance (standard S-44) and should theoretically be parallel to the local geodetic reference surface in absence of currents;

b. an ellipsoidal tidal datum diagram for the area is represented by the Kinematic GPS MLLW Tide Surface obtained from the ellipsoid height values.

The Ellipsoid Height Surface values and the GPS reference station used to measure the ellipsoid-MLLW separation enables Kinematic GPS hydrographic surveys to be conducted without the use of tide gauges.

Location of the GPS Reference Station

A permanent GPS reference station (Figure 7.11) must to be established close to the shoreline for hydrographic surveys in harbours and related approach channel areas. An antenna height $\boldsymbol{h_1}$ in meters (negative) should be entered into the GPS receiver during GPS hydrographic surveys. If the GPS reference station antenna is moved, the value is invalid. If the antenna must be moved, the vertical difference D$\boldsymbol{H}$ between the bottom of the antenna and the reference benchmark must be re-measured and confirmed that the benchmark is (ellipsoid height $\boldsymbol{h_2}$ in meters) below the ellipsoid. Levelling runs starting from the benchmark should be conducted through the old antenna location and the new antenna location.

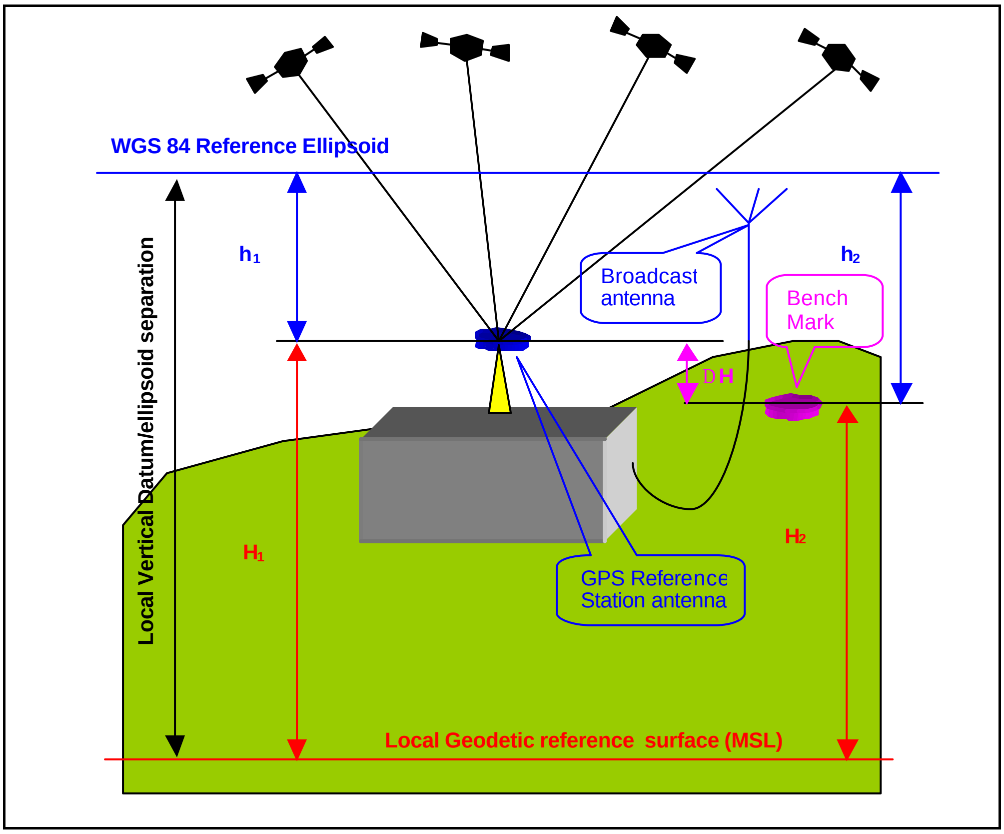


Fig. 7.11 "RTK DGPS reference station parameters"

h_1 = ellipsoid height of GPS antenna below the Ellipsoid WGS 84 surface
h_2 = ellipsoid height of benchmark below the Ellipsoid WGS 84 surface
H_1 = orthometric height of GPS antenna above Local Geodetic reference surface (VD)
H_2 = orthometric height of benchmark of the Local Geodetic reference surface (VD)
ΔH = vertical difference between the bottom of GPS antenna and the reference benchmark measured by geometric levelling

Resultant RTK DGPS Elevation Accuracy

The resultant absolute project accuracy is estimated to be less than 10 cm. The absolute accuracy refers to the MLLW relative to the Local Geodetic reference vertical datum. A local project modelling of the ellipsoid-geoid separation should be attempted for the project. A suitable computer program should used by entering the surveyed horizontal positions to compute the Local Geodetic reference/ellipsoid WGS 84 separations.

Real Time Kinematic Measurement at Sea

The GPS antenna phase zero measurement down to the water line of the vessel is the most important vertical measurement on the survey vessel. In a static condition, the measurement is as shown in Figure 7.12. Underway the vessel motion through the water will change these figures, however vessel squat is not entered as a correction in the survey system as the transducer depth is reduced by the same amount the antenna height is reduced.

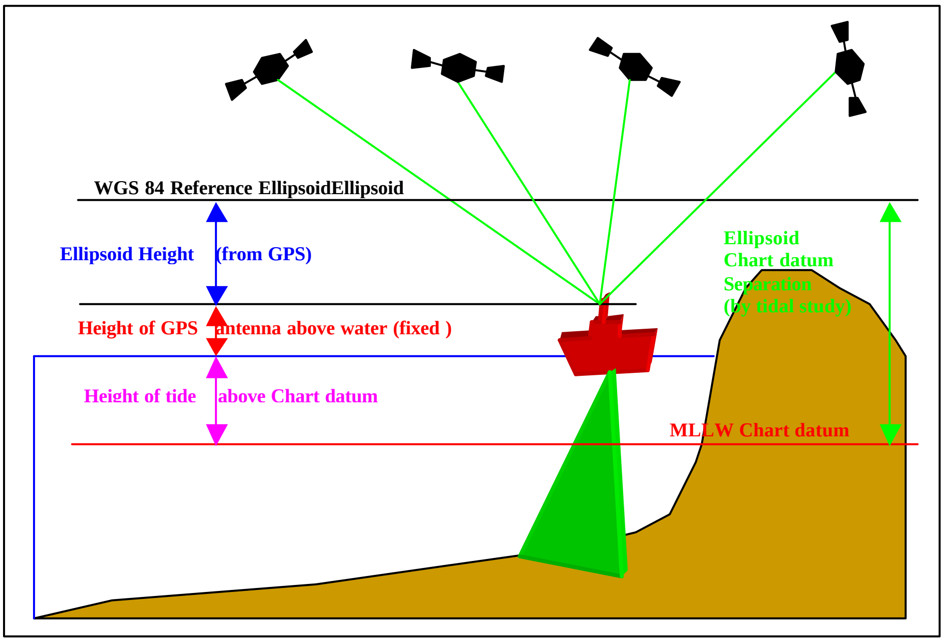


Fig. 7.12 "Real Time Kinematic measurement on survey vessel"

Survey Procedures, Test and Processing

a. Conventional method:

i. a number of established acoustic tide gauges should be running and downloaded to produce a time series from the data sets, which should be referenced to the station numbers;
ii. the gauge data should be used to eliminate actual time differences in the area from average time differences between the gauges;
iii. resultant average error when using one tide gauge should be appreciated, the maximum distance at which the data is valid and where it does not exceed standards listed in S-44 4th Edition 1998;
iv. the separation distance between gauges should not normally exceed double maximum distance highlighted in iii above.

b. RTK DGPS carrier phase method:

i. the use of tide gauges during the hydrographic survey are not required;
ii. a MLLW surface should be generated by suitable software from the tidal datum diagram;
iii. the survey vessel must be equipped with a GPS Rover Receiver capable of applying OTF GPS carrier phase corrections received from Reference Station;
iv. vessel data (layout, draft, squat etc) must be obtained from conduct of reference measurements on the survey vessel.

A test on the RTK GPS Tides separation value must be performed, creating a difference ellipsoid/MLLW matrix by suitable software. A sufficient number of cross-section lines must be run in an area between the two nearest acoustic tide gauge sites, which are recording tidal data.

c. Two survey processing methods are possible:

   i. the conventional method uses the horizontal GPS co-ordinates only (and not the vertical one) and the reduced depths relative to the MLLW will be obtained applying the tide gage data to the raw soundings;
   ii. RTK GPS method generates an accurate measurement of depths related to the GPS antenna height without observing tidal measurement data. The GPS depths are directly referred to the Kinematic GPS MLLW Tide surface. A random number of depths from each line should be selected for comparison with the GPS depths reduced by the tide gauge data.

4.3 Environmental Observations

The direction and rate of the tidal stream should be observed wherever it is of navigational significance and where there is no evidence that observations have been made previously. Positions, and full requirement, for observations will be articulated in the Hydrographic Specification but additional stations should be included if considered necessary.

Observations should be made using a current meter, current profiler or a floating log-ship. Observations should be made at a depth appropriate to the average draught of shipping using the area or as directed. Observations should not be taken during abnormal weather conditions.

In predominantly semi-diurnal areas, observations should be conducted over a single period of 25 hours at Springs. In areas where the diurnal equality is large, 30 days observations, using a current meters to enable harmonic analysis to be conducted, are required. Should it not be possible for such protracted observations to be taken then sufficient measurements should be obtained to enable a description to be inserted in the Sailing Directions and tidal stream arrows to be shown on the chart.

In addition to standard observations, information of a less formal nature may be available from local sources, especially if it may affect low-powered vessels or yachts. Data obtained should include the estimated maximum rates at Springs and the directions of tidal streams assessed by the best possible means. In areas of strong tidal streams, especially in the vicinity of banks, rocks shelves and in narrow passages, eddies and overfalls may occur which can be of considerable significance especially to small, or under-powered craft. The limits of these phenomena should be fixed, at Springs, on both directions of the tidal stream.

The initial observations of sound velocity should be conducted to allow determination of the spatial and temporal variations across the entire survey area. A grid of observation points should ensure representative sampling is conducted over the whole survey area in a methodical and timely fashion. This data, together with other environmental factors such as climate, fresh water inflow, any seasonal variations and seabed topography, will determine the frequency at which SV profile observations are conducted. The use of moving vessel profilers, undulating profilers and hull mounted probes will reduce the need for static observations to be performed; however water depth and vessel size may limit the ability to make use of such equipment.

There are set intervals at which SV data should be obtained and applied, guidance should be provided in the Hydrographic Specification and by the Hydrographic Office. The importance of correct SV when using MBES can not be understated.

4.4 Line Guidance

General Description

In positional terms, the process of data acquisition can be summarized in Figure 7.13. Once line data base and direction have been decided the surveyor needs to know his position along the selected line at all times.

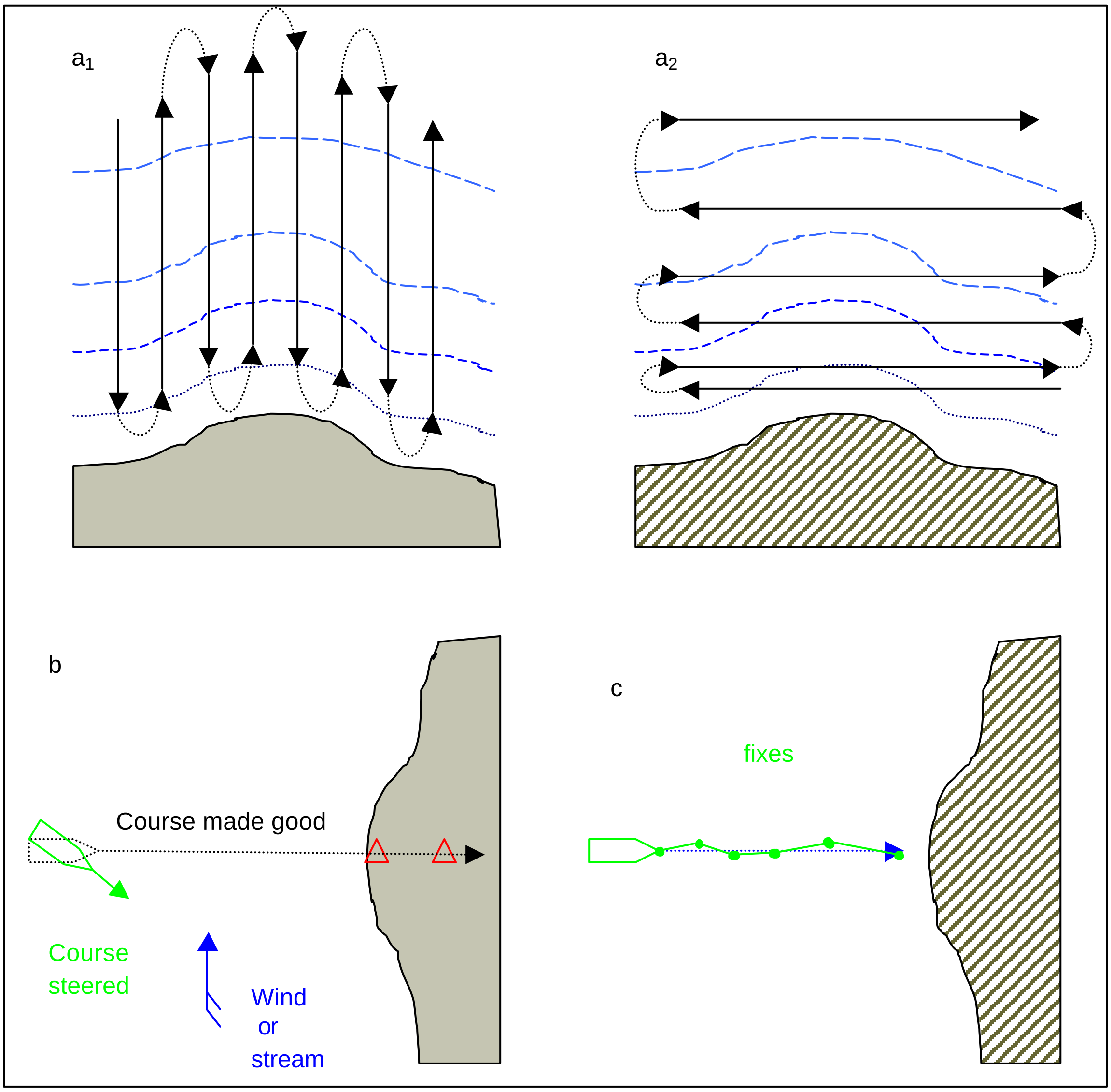


Fig. 7.13 "Sounding lines and related track control"

a. The decision on line orientation and spacing for methodical sounding of an area will be influenced by the equipment to be used. Figure a_1 depicts Singlebeam Echo Sounder (SBES) with closely spaced lines crossing the depth contours at right angles, Figure a_2 illustrates Multibeam Echo Sounder (MBES) or Side Scan (S/S) Sonar with lines spaced to have a minimum overlap parallel to the depth contours;

b. Make good the selected track;

c. Determine the actual track made good, the vessel's position is fixed at set intervals with the track followed assumed to lie along the line joining the fixes;

A traditional EPS position is updated continually and, by observing a track plotter or left/right indicator, the slightest deviation from the chosen track can be detected and corrected. Further, the track plotter record, if marked to show time intervals along the track, enables the precise plotting of other acquired data. Visual fixes, on the other hand, are periodic events and the assumption that positions at times other than the fix instants will lie on the lines joining fixes will be increasingly erroneous as the interval between fixes increases. Sextant, theodolite, and total station fixes conducted during a harbour survey, may be taken at intervals of a few seconds and the departure from the intended track can be controlled by transit (as is in the above example) or by a variety of other methods.

Modern EPS and satellite fixes are taken every second or less, providing continuous positional information connected to a left/right indicator or to an automated data acquisition system, which offers a means of determining, in real time, the ship's course and speed over ground, with an accuracy directly related to the chosen positioning system.

In surveys of approximately 25 km^2 or less, seabed acoustic transponder arrays, positioned and laid using conventional positioning methods, can be used in conjunction with echo-sounder or sonar ranging to provide continuous positional information and, consequently, control of track.

Visual Line Guidance

When visual fixing methods are used, the ship's track is almost invariably plotted manually, with lines joining fix positions representing the track followed by the vessel. Therefore, the fixing interval and vessel speed are selected that the fixes are sufficiently close (about 3-4 cm. apart on paper) for inaccuracies to be assumed negligible at the scale of the survey (i.e. departures of the vessel from the line joining fixes will be unplottable). Line control during survey operations is usually achieved independently.

In the most difficult case of the off-shore survey with no visible shore control, track guidance will be by compass or better gyrocompass. This is never entirely satisfactory except on very small scale surveys, course adjustment being necessarily delayed until the fix has been plotted. An alternative method of line guidance in these circumstances is to steer the vessel around the arc of a fixed angle subtended between two marks or following a circular/hyperbolic LOP of a traditional EPS chain. These methods are superior to compass/gyrocompass courses but they can be difficult on large scale surveys with shore marks relatively close to the survey area and where the arcs are of small radius requiring constant, large variations of course. Now days, in off-shore surveys, GPS or EPS techniques are exclusively used.

Other methods for visual line guidance and track control are:

a. Natural transits – by keeping an object near the shore line with another further inshore in the direction of the track to be steered, the helmsman should be able to maintain the line more easily and accurately than by compass course. Any suitable feature may be used (bushes, fence or telegraph posts, huts, parts of buildings etc.), the transit marks should be spaced far enough apart, about one third the length of the survey line, to offer sufficient sensitivity.

b. Artificial transits – the same principles apply as for natural transits; artificial marks, placed to meet the required line spacing, enable more precise steering and may be essential for large scale work off a barren shore; this method is particularly useful when undertaking large scale harbour and wharf surveys, where transits at right angles can be erected to provide line guidance and fixes at set intervals to meet the survey requirements;

c. 180° collimator prism – this robust and simple instrument enables the helmsman to sight forward and back marks simultaneously so that, in harbour or river surveys, the vessel can be steered along the line joining points on opposite banks.

d. Direction from shore – the direction of the intended line is taken off the plot by station pointer as an angle from a reference object or directly from the field sheet by intersection of LOPs of a lattice. The required direction is then observed by theodolite or sextant with the vessel directed along the line by the shore observer using hand flags, lights or radio communications link. When surveying across a river, basin or berthing area, the shore observer can sight an object on the opposite shore on the line to be run, enabling him to follow the vessel's progress by eye.

e. Starring – by planning the survey lines as radial lines centred on a shore mark, that mark may form the front mark of a transit, the helmsman picking up a new (natural or artificial) back mark for each successive line. Alternatively the *"direction from shore"* method may be used, the shore observer having to occupy only one station. This method is particularly suitable for surveying around headlands and promontories.

EPS Line Guidance

Control may be achieved simply by planning the survey lines along either the range circles or range difference hyperbolae depending on the type system in use. If the lines are steered along pattern lane readings, any departure from the line is immediately obvious, and the vessel may be fixed at the intersection of the line by a second pattern. In a hyperbolic pattern the lines and the fixing intervals will diverge or converge but lane expansion is usually negligible. Allowance may be made for lane expansion, either by changing the pattern intervals used or by running interlines to maintain minimum spacing of lines. A left/right indicator will show the vessel's position relative to the line and provides clear guidance to the helmsman.

Almost all manufacturers of short and medium range EPS systems offer the facility of track plotter as peripheral equipment, which is particularly useful when lines cannot be run along lane boundaries as is often the case in dredging or pipe laying work. The lattice can be plotted out and the selected line followed by a plotter pen, fixes may be marked on the plotter track as a check but it is more usual to keep a separate manual plot since the lattice is often distorted on the track plotter, some types of lattice are shown as a rectangular grid and the required scale of the survey will rarely be that of the plotter.

Automated Line Guidance

Appendix 4 of Chapter 7 (page 8 and 9) outlines a typical automated hydrographic systems configuration for an Operation Room of a survey ship and the general hardware configuration on a survey boat.

In general the hardware configuration of an automated hydrographic data acquisition system is similar on both survey ship and survey boat, with a suitable hydrographic data acquisition software to control, manage, acquire and store in a specific survey data format digital data from the positioning system and echo-sounder system (SBES/MBES/SSS). Modern hydrographic data acquisition software should provide a helmsman display allowing the vessel to be steered, either manually or automatically, along pre-planned survey lines.

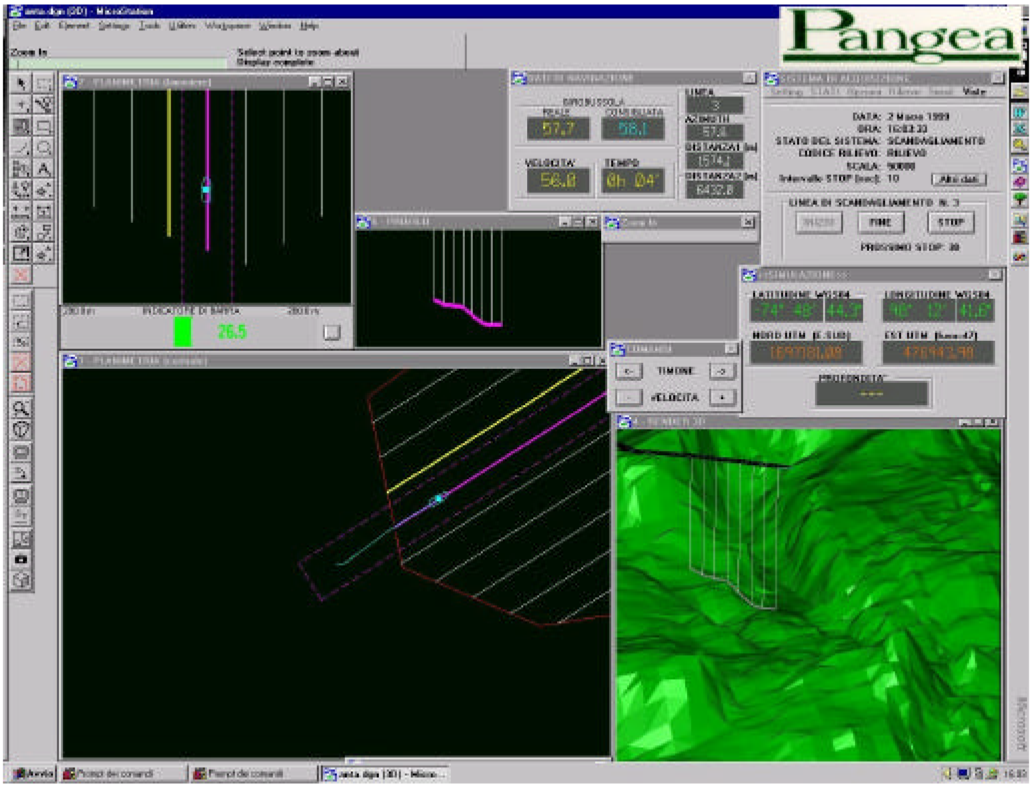


Fig. 7.14 "Video image displayed from acquisition software"

Figures 7.14 and 7.15 show some typical video image suitable for a helmsman display providing the following information:

a. the survey line followed by the ship geo-referenced with the real time position updated at 1 second intervals;
b. a left/right indicator;
c. digital information received and managed by the acquisition software (position co-ordinates, depths, COG, heading, SOG, line number and fixes, distances from start and end line, etc.).

All this information allows the helmsman and the surveyor to control and monitor the acquisition process along the selected survey line to cover the area. The acquisition software operator manual should contain all the instructions and procedures to manage the automated track control, which in general is conceptually similar for every software manufacturer.

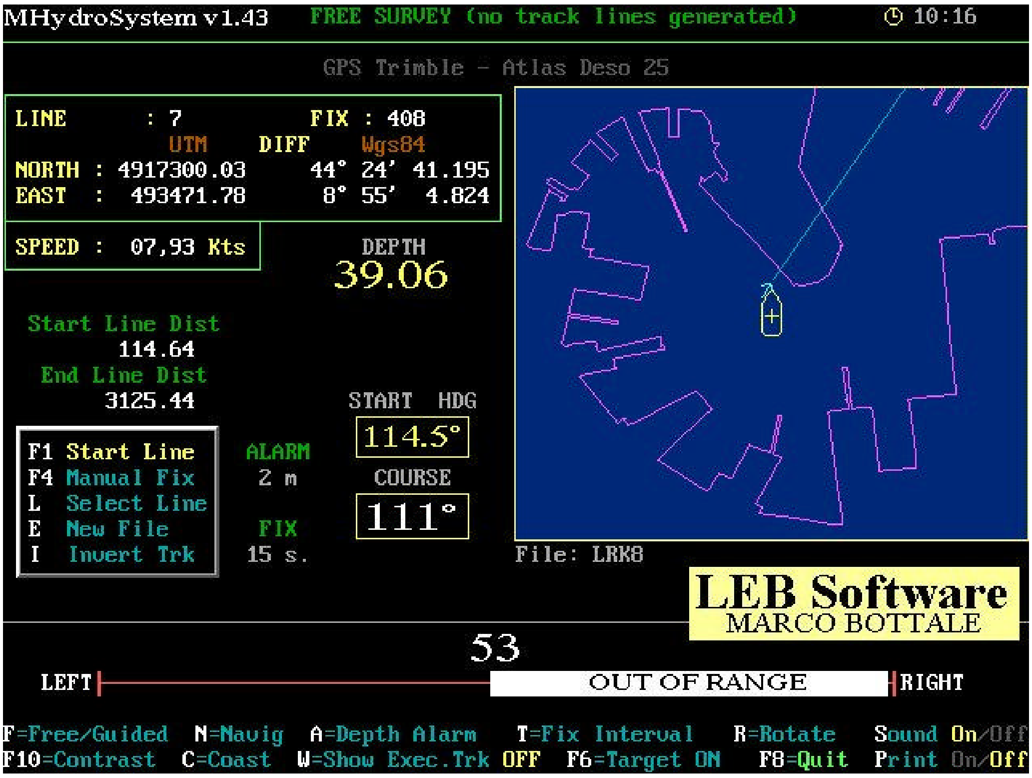


Fig. 7.15 "Video image displayed from acquisition software"

4.5 Check Lines

Crosslines should be run, at the beginning of the survey, perpendicular to the main sounding lines and, whenever possible, at a different stage of the tidal cycle and in good sea conditions. Sufficient sonar information from crosslines should be gathered to enable a statement to be made in the RoS of the extent of sand ripples in the survey area and their direction.

Sounding crossovers should be compared as the survey progresses as a check against gross errors, co-tidal modelling error or equipment malfunction. On-line MBES displays should be used to verify the repeatable performance of sounding by monitoring adjacent swathes; this should also be used to ensure that the coverage and swathe overlaps are being maintained, additional lines should be run to fill any gaps in coverage.

4.6 Main Lines

The sonar sweep type should be articulated in the Hydrographic Specification, which should also detail the % coverage, % overlap and object detection criteria to be achieved. Careful inspection of the Hydrographic Specification limits and those of adjacent modern surveys is required to ensure that there are no gaps between them.

When conducting sonar and sounding lines concurrently, a careful balance must be maintained between the conflicting requirements. Every effort must be made to ensure that there are no gaps in the sonar sweep and therefore it will generally be necessary to subordinate the needs of bathymetry to those of sonar. Additional sounding only lines may be required to assist with the delineation of contours and critical features, where line keeping error is >25% of line spacing, the gaps should be filled with additional lines.

An *a priori* sounding error budget/estimation should be created and compared to the Hydrographic Specification requirements. If significant environmental variance is encountered or changes to equipment are made to those initially used, the error budget/estimation process should be repeated and these new results used as the basis for comparison.

Main survey line spacing, direction and sounding speed should run to meet the requirements of the Hydrographic Specification and the laid down criteria. The optimum and maximum survey speeds should be assessed taking into account the depth range and systems in use for the survey; the maximum ping rate and range scale should be utilised commensurate with the depth of water. Normally sounding lines should be run perpendicular to the general direction of contours; however sonar lines should be run within 20º of the prevailing tidal stream or current. In areas of strong tidal flow, a direction much less than 20º may have to be adopted to ensure that the sonar towfish follows the ship's track.

When surveying with MBES only, line orientation and spacing will be dictated by the criteria detailed in the Hydrographic Specification and the customer requirements. Object detection and thus data density will determine % overlap between adjacent swathes, which will be a function of water depth; thus line spacing will be influenced by water depth, data density and objection detection criteria and % overlap, which will give the % coverage achieved. It is most likely that line spacing will vary across the survey area, particularly if there are marked differences in depths; careful monitoring during survey operations will be required to ensure the desired objectives are being met.

Fix interval and logging intervals should be set as required by the scale of the survey and the survey processing system in use.

When using SBES, additional sounding lines should be run at standard spacing perpendicular to the contours from the 10 metre contour into the shore in order to determine the 5 and 2 metres contours and the drying line. Additional lines should be run parallel to and at distances of 2, 5 and 10 metres off jetties or wharves.

When sounding over sandwaves, operations should take place following periods of calm weather and neap tides when sandwave amplitudes are greatest. Where possible re-surveys of sandwave areas should follow the same tracks to detect changes in sandwave profiles.

Care should be taken delineating the drying line, particularly with regard to off lying banks, rocks and shoals, as the position of such features may have international legal implications. When surveying within harbours and boat havens, drying heights and the location of foul ground, in areas where small craft anchor or take the ground, should be accurately delineated.

Whenever a survey includes a channel, recommended track or leading line in restricted waters, it should be swept by sonar. Allowance during such sweeps should be made to accommodate the largest vessels likely to use these tracks paying particular attention to turning areas and where a track changes course.

Regular checks should be conducted to ascertain sonar performance, suggested occasions are:

a. On first streaming;
b. Once per day when operating in areas of featureless seabed;
c. After maintenance or repair;
d. After changing towfish or fins;
e. If the performance is in any doubt.

If it appears that the sonar system is not working to its maximum range, due either to water conditions or to material inadequacies which cannot be rectified, the sonar sweep should be modified to ensure complete coverage.

A magnetometer should be deployed throughout the basic sonar sweep to provide additional evidence of the existence of ferrous metal on or below the seabed.

It is of utmost importance that the sweep is thorough, that no gaps appear in it, and that every significant seabed feature or artificial obstruction is located.

4.7 Inter-lines and Investigations

Items in the supplied list of wrecks requiring a disproving search should be searched for out to a radius of 2.5 miles from the listed position. The limit of the area of this search may extend outside the given limits of the Hydrographic Specification; however such an extension is essential to avoid the anomaly of having a PA wreck lying inside an area considered to be fully surveyed. Ensure that any disproving search areas lying on the outer edge of the survey area are covered; additional lines should be planned to run outside the area in order to ensure complete insonification of the area with an appropriate overlap.

Careful note should be taken of the 500 metre safety zone around seabed installations and pipe laying operations when surveying in or near oilfields or exploration areas to ensure the safety of any towed equipment.

For an object charted PA, the sonar search should be conducted in 2 directions at right angles and extend to at least 2.5 miles from the datum position. If there is a high degree of confidence that the initial search in one direction was entirely thorough and that the sonar equipment was operating satisfactorily, consideration may be given to dispensing with the search in the second direction.

Objects whose positions have been previously established, but which cannot be found during the survey, need a very detailed investigation to disprove them. Where such objects fall within the survey area and a sonar search is completed to a radius of ½ mile around the listed position, this will be considered sufficient. A magnetometer should also be deployed. When there is no doubt about the geographical position of a wreck after many repeat surveys, the above radius may be reduced. Consideration should be given to the use of a wire sweep.

Each contact should be closely examined using sidescan sonar; should the contact be confirmed, its position and least depth by close sounding should be established. A minimum of 4 good runs, comprising 2 perpendicular pairs, should be achieved. In the case of wrecks, one pair of tracks should be parallel to the axis of the wreck and one pair perpendicular to it.

Data regarding the contact can be obtained from the use of sonar, echo sounder, magnetometer, wire sweeping, diver or a combination of these. Each contact should have the following detail:

a. Position;
b. Least depth;
c. Nature of the object;
d. Length, breadth and orientation;
e. Depth, length and orientation of scour;
f. Debris field length and orientation;
g. Strength of magnetic field.

Examination of the supplied wreck list may aid identification of the object; however caution should be exercised in too freely linking newly discovered wrecks with those contained in the wreck list. Disproving searches may still be necessary in charted positions. The use of divers may be helpful in identifying wrecks and reporting their state and attitude; in particular, they may be able to locate high points, which may not have been distinguished on the echo sounder or sonar.

The least depth over wrecks and obstructions must be established, in certain circumstances this will require the use of wire sweeping, which should be conducted after the position, size, orientation and probable least depth have been determined by sonar and echo sounder. Wire sweeping should be considered in the following circumstances:

h. As directed in the Hydrographic Specification for specific wrecks;
h. If the least depth is likely to be less than 40 metres;
h. Where depths around the wreck are significantly different from those charted;
h. When salvage/dispersal work has taken place since the last survey;
h. Sonar indication of protruding masts and structures;
h. Areas charted as foul within an anchorage;
h. Wrecks in areas of strong tidal streams and seabed mobility;
h. Where the position of the wreck is significantly different from that charted.

Care should be taken to ensure that the whole area of the wreck is covered by the wire sweep, albeit in several laps, and that there are no gaps between the sweeps. It is not sufficient to cover only the areas which appear to be high points on the sonar.

Particular attention should be paid to the sounding of depths <40 metres, where the least depth should be obtained over all seabed features. Interlines should be run in depth <40 metres unless the seabed is flat and featureless and no dangers are shown to exist by complete coverage by high definition towed sidescan sonar or MBES.

4.8 Ancillary/miscellaneous Observations

Seabed samples should be obtained at regular intervals throughout the entire survey area. Additional samples should be taken in all likely anchorages, on all banks, shoals and seamounts, particularly where these are likely to be unstable, and in the channels between them. Approximately 10% of the samples obtained should be retained for rendering with the final survey.

Seabed samples should normally be obtained throughout the survey area prior to commencing the area sonar sweep to provide a method of ground truthing the sonar picture and to enable more accurate interpretation for seabed textures.

Before arrival in the survey all fixed and floating navigation marks and aids should be identified from the largest scale charts, List of Lights and List of Radio Signals. On arrival all uncharted or listed navigation marks and aids should be identified.

The position of each new fixed or floating navigation mark and any suspected of being out of their charted/listed position should have its position determined; floating marks should be fixed on the full ebb and flood tides, however if an unambiguous position of the central mooring sinker/anchor can be obtained from swathe bathymetry, it is permissible to quote this as the charted position.

For lit navigational aids the following should be recorded:

a. The height of the focal plane;
b. Light and sound characteristics;
c. Light sectors and obscured arcs;
d. The structure shape and colour;
e. The top mark shape and colour.

Colour photographs of all marks should be taken for inclusion in Sailing Directions.

Harbour authorities should be consulted over changes found to navaids to determine if such changes are permanent or temporary. Details of any planned changes should be obtained.

The details of port radio operations, including Ship Reporting Systems (VTS, VTM, VTIS, etc), radio pilot services, radio navigation aids (including (Aero) radio beacons, radar beacons, etc), coast radio station services (ie. public correspondence details, navigational warnings and weather information broadcasts, schedules, etc), GMDSS facilities, together with Search and Rescue procedures, are to be obtained.

Every opportunity should be taken to obtain details of natural phenomena during the course of the survey. Such phenomena includes:

a. Deep scattering layer is the biological layer, consisting of plankton, and other small marine organisms, and the larger fish that feed on them. Certain of these have swim bladders which respond to echo-sounder and sonar transmissions, causing scattering of the sound waves, which may have considerable affect on sonar operations. Reports of the phenomenon are therefore important and should be rendered;

b. Marine bioluminescence is caused chiefly by marine animals varying in size from microscopic organisms to quite large fish, squid and jellyfish. It is more commonly encountered in warm waters than in cold, and is of considerable interest to marine biologists and military scientists. Reports on the phenomenon are therefore important and should be rendered;

c. Discoloured water is generally recognised to be almost always biological in origin. Water samples and secchi observations from such areas are of considerable interest and should be obtained and reported;

d. Marine life reports should be made to cover whale movements and those of other marine mammal species, which are of considerable interest in anti-submarine warfare and to marine biologists. The presence of commercial fishing activity in the survey area is of importance as the sound generated by engines and deployed fishing gear may significantly affect ambient noise levels. Also the occurrence of a fishing fleet may indicate the presence of large fish populations; sound scattering caused by shoals of fish may inhibit sonar performance. Sightings should be included in the RoS.

5. COASTLINE DELINEATION

5.1 Coastlining General

The accurate delineation of the coastline (shoreline) and coastal features is an essential feature of a Hydrographic Survey, since the mariner is often required to fix his position by bearings and angles or ranges to promontories and similar features on the coastline. Generally, in hydrographic surveying, the coastline is defined with respect to a HW datum.

The coastline, except in the most rapid and cursory surveys, must always be walked over in the field when its nature permits. Many small river mouths and streams have been missed by the practice of taking a boat along the coast and only landing at various discreet locations.

In some instances, adequate land survey maps based on modern air photography or satellite imagery will provide data which can be used to assist in plotting the coastline. Occasionally air photography will be flown specially for a survey and an air "photo plot" produced at the appropriate scale. This however does not obviate the need to walk the coastline in the field.

All land survey maps and air photo plots should be checked in the field before use is made of them on the Bathymetric Sheet and Tracings to Accompany. Where no maps are suitable, the coastline must be properly surveyed with key features being fixed by a regular method dependant upon the scale of the survey.

A surveyor delineating the coastline, even when checking a map or photo plot, should plot the Drying Line as best as possible and should always note the nature of the foreshore. The best way to find the Drying Line is by reduced soundings but every part of the foreshore should, if possible be seen at least once at low water in order to see whether there are dangers which might have escaped attention. This is particularly necessary where the range of the tide is great.

5.2 Coastal Details Required

The surveyor should carry out the following tasks:

a. Delineate and fix the coastline by the best methods available;
b. Fix and describe all objects CONSPICUOUS and Prominent to the mariner, which are not already fixed, and check existing marks/features on charts and publications are positioned and described correctly, even though they may be a little inland;
c. Fix and describe or indicate on the chart all objects and features of the coastline which would assist the mariner to fix himself and identify the coastline. In large scale surveys this will include very minor detail which can only be seen close inshore;
d. Measure and estimate the heights of all such features, some features can be described in general terms such as "Low red cliffs, 5 to 6 metres high";
e. Fix all islands, visible offshore dangers and obtain their heights, also fix adjacent floating marks (buoys not on chart);
f. Describe the composition of the beach between the low water line (drying line) and the high water line as well as above the high water line. The appropriate symbols should be inserted on the Bathymetric Sheet;

g. Indicate established landing places along the coast. Fix and describe groynes, sewer outfalls and anything that might constitute a danger to landing. Piers and jetties should also be fixed and a full description obtained, which should include type of structure, depths alongside, height of deck above the HW datum and facilities available;
h. Details of harbours should be obtained giving berthing facilities and supplies available, this information is to be included with the Sailing Directions;
i. Where appropriate, correct spelling of place names should be obtained from reliable local sources and checked with names shown on existing maps and charts & publications.

In addition the surveyor may concern himself with topographical detail near the coast. The amount of detail will depend on the time available, the scale of the survey and whether topography is going to be obtained by another field party or by other means, such as air photography.

5.3 Detail Of Concern To The Mariner

CONSPICUOUS objects/marks. - Mariners use bearings to peaks, churches, chimneys, windmills, masts, permanent buildings etc. They will be visible from quite a distance offshore.

Prominent marks. – Again, Mariners will use bearings to peaks, churches, chimneys, windmills, masts, permanent buildings etc. These marks will be visible from quite close to the foreshore/coastline.

Headlands, Islands, Offshore features. - Mariners use bearings to left and right tangents and vertical heights to cliff tops etc.

Harbour and Port facilities. - Determine dimensions and heights of jetties, orientation and depths alongside, type of construction, mooring and berthing facilities, small craft facilities, visiting craft details, marinas, yacht clubs, fuelling berths etc.

Principle Land Features:

a. Natural Objects - Hill summits, rivers, lakes, marshes, woods, contours etc.
b. Artificial Objects - Buildings, towns, flag staffs, roads, railways, factories etc.
c. Contours - Sufficient to indicate height and shape of coastal region.
d. Nature of the Foreshore and Near shore Topography - To assist in recognition and in selecting landing places.

Lights. - Details must be checked in the field and against entry in Light Lists.

Sailing Directions. - Full written descriptions of the coastline and details of port and harbour facilities.

5.4 Topography

Where recent land surveys maps, aerial photographs or plots from photographs are available, they should be thoroughly checked in the field and any discrepancies recorded.

The topography shown on large scale charts should be also be checked in the field to update detail which is not normally shown on maps and which is not visible on aerial photos, paying particular attention to coastal details e.g., beacons, flag staffs, groynes, etc. Charted objects which no longer exist should be recorded, preferably as deletions on a copy of the chart, which should be forwarded with the results of the survey.

No topographic detail is to be shown on the Bathymetric Sheet unless it has been surveyed in the field, or its existence and position on a map or air photo plot have been confirmed by the surveyor. Newly surveyed detail must be included, detail from other sources which is found to be correct may be transferred to the Bathymetric Sheet, at the surveyor's discretion to give a balanced, more complete presentation, but it must be inserted with meticulous care so that subsequent queries concerning its authenticity do not arise.

Where no modern maps or air photography exist, all topographical detail which will be of use to the mariner must be accurately fixed and shown on the Bathymetric Sheet. In particular, all features that may be used to fix a ship's position, whether visually or by radar, should be plotted and, where practical, coordinated. Unless time and resources are available, effort should not be expended in recording minor topographical detail of no direct interest to the mariner, or which is not visible from seaward.

Any changes that are found should be reported.

5.5 Delineation Of The Drying Line

The drying line is in general the basis for determining the limits of the territorial sea and associated baselines and its careful delineation is most important because of the affect it may have on fishery limits, observance of pollution regulations and the licensing of offshore mineral extraction, as well as on the delineation of international boundaries. Surveyors should therefore take special care that the drying line of the mainland, islands and of all drying features, is adequately delineated even though it may be of little interest to the mariner.

The survey of an area involving drying features must be considered incomplete if the drying line is not adequately delineated, unless specific instructions are issued to the contrary.

5.6 Heights Of Land Features

The height of all land bordering the coast must be obtained even if the elevation is small, in particular every small islet and rock which shows above water must have its height noted against it. If there are no summits available, the heights of the tips of tree or islands must be found and recorded. Similarly the height of all tall artificial features such as masts, chimneys etc.

Cliffs must have their heights noted, and their colour if in any way remarkable.

Heights of objects should be referred to MSL (or its equivalent), if this is impractical, the actual height of the object may be shown as a legend. For special purposes or in river surveys, other criteria for the reference plain may be used.

5.7 Charting The Foreshore

The main concern when charting the foreshore is to fix the position of all dangers which may have been missed by routine sounding, such as rock outcrops or man made obstructions. Objects that lie beyond the drying line will have to be intersected. Pipelines or cables prominent on the beach may extend out to sea beyond the drying line and be a hazard when anchoring. Finally the nature of the foreshore should be recorded.

5.8 Coastline Overlay

A properly prepared coastline overlay should contain:

a. Title of the Survey. Survey specification number, scale, ship's name and any other subsidiary identification;
b. Position of all co-ordinated marks to be used;
c. Grid and Geographical positions pricked and circled in ink with value of at least 2 of each positions;
d. Diagonal scale;
e. True and magnetic meridians;
f. A note of the value of least plottable distance, which is ¼ mm on paper, regardless of scale (1 : 10 000 = 2½m).

5.9 Use Of Air Photo Plots

An air photo plot of the coastal area within the limits of a Hydrographic survey will sometimes be prepared in the Hydrographic Office when:

a. Existing mapping of topography is outdated, poor or non-existent;
b. Features of hydrographic interest, eg underwater rocks, are either shown inadequately, or not included, on topographical maps, and are evident on photography held;
c. Drying heights can be accurately and efficiently established by photogrammetric means.

When an air photo plot has been supplied in support of a survey specification, the following procedure can be adopted:

a. The photo plot is to be thoroughly checked in the field for errors in photo interpretation, shape and position. Difficulties experienced in photo plot compilation and discrepancies with existing data, will be highlighted in the photo plot report that accompanies the photo plot. Special attention must be given to resolving these differences;
b. Data on the photo plot which has not been verified in the field must not be included on the Bathymetric Sheet. It should be noted that the water line derived from air photographs rarely coincides precisely with Chart Datum, especially in areas with gently sloping beaches and there may be a need to adjust it before insertion on the Fair Sheet;
c. Amendments to the photo plot must be shown by marking corrections on a copy of the photo plot;
d. When required, air photographs supplied with the plot are to be marked by pricking and identification on the reverse side, to show additional control established for the survey.

Existing air photography can be of considerable use in coastline delineation. Previously used ground control points should be checked against the survey reference system, remembering that the flight scale may need to be expanded to 4 or 5 times to match the survey scale. Note that ground elevation errors in photogrammetry are about 1/5000 of the flight height whilst horizontal errors are considerably smaller.

When coastline delimitation is the objective, rectilinear strips orientated parallel to the coastline should be planned. Overlapping strips should include common control points and the flights should co-inside with low water times.

For large-scale surveys (less than 1: 50 000) where photogrammetry will be used to obtain ground relief and other topographic details, additional 3d control will be required. Errors in co-ordinated ground control points should less than half the aerotriangulation errors for points used during photogrammetic interpretation.

For strip aerotriangulation adjustment, 4 planialtimetric (3D) control points are required. Further inner control points should be added to each strip.

In more complex flights with a large number of parallel strips, the aerotriangulation enables a block adjustment to be made. The planialtimetric ground control points may be in the order of 5+0.2M (where M = the model number of the aerotriangulation process). Additional altimetric control points may be required in areas where accurate heighting of features is required.

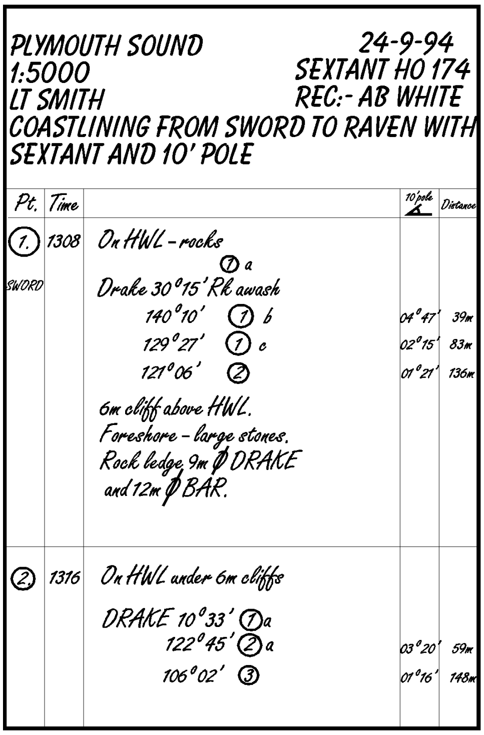
For smaller scale surveys (greater than 1 : 50 000) a lesser number of ground control points will be required, some with only horizontal co-ordinates known. This is also true for satellite imagery.

Photogrammetry or satellite imagery does not negate the need for the surveyor to fully explore the coastline in the field.

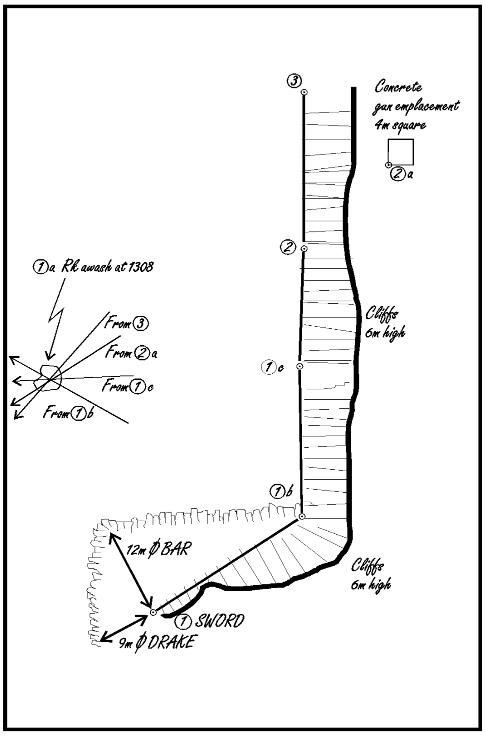
5.10 Coastlining Methods

Whichever method is used, points must be fixed on the HWL so close on paper that detail can accurately be drawn in between them. Spacing on paper depends, therefore, on the complexity of the coastline.

Typical recording for a traverse using sextant and 10' pole is detailed below. Similar recording would be employed when resecting, or using a combination of both methods. Note the diagram shown embraces all angles and distances on the recording page.



(Opposite page)



5.11 Plotting The Coastline

It is usual to draw the coastline by plotting the resection fixes or traverse points directly on to a coastlining overlay, although a millimetric plotting sheet may also be used. It may also be appropriate to compute the turning points of a traverse and to adjust them prior to plotting. Prick through known co-ordinated points used to start and close the traverse, the remaining coastline details are then plotted by hand.

The graphic method of plotting a coastline is given below:

a. Plot the points by protraction of angle and distance. It will be found most convenient to use a large circular protractor of transparent plastic, springbow dividers, pricker and well sharpened 4H pencils. It is important to remember to avoid the use of short zeros on the paper for aligning the protractor. When plotting rays, ensure that they are drawn longer than required and that prick holes made at the edge of the protractor are marked so that they can be used again for aligning the protractor if required. It is important to do this when the reference object is the back mark of the traverse and therefore the zero for plotting would be very short if the back mark prickholes were used for aligning the protractor or station pointer. If there is an acceptable misclosure the traverse should be adjusted graphically. If there is a large misclosure it will often be found to be due to a single gross error rather than an accumulation of minor ones. The plotting should be checked carefully before abandoning hope and re-surveying the section in the field;

b. The intermediate points, tangents and plotting shots are then drawn in from the corrected positions of the turning points; care must be taken to zero on the corrected positions of the back marks, drawing lines through them upon which to align the protractor;

c. The detail of the coastline between the intermediate points is then inserted using the sketch map; air photographs and existing maps are of great assistance as is the information noted at the ends of lines by boat-sounders;

d. The coastline is then inked-in using the appropriate colours and symbols (see Int 5011). When depicting cliffs or steep coasts, the base of the cliffs must be correctly charted but the tops can be drawn far enough inland to permit the pen work to be inserted; similar exaggeration is necessary with certain other symbols but great care must be taken with those representing low water features to ensure that the seaward limit is correctly charted.

5.12 Coastline Delineation Report

Coastline survey work should be reported in accordance with Appendix 2 to this chapter, however some of the more important aspects are detailed below:

a. State the aim of the work, noting whether it was to update an existing chart, for a new chart or another special purpose. Indicate what other sources of material were used in addition to field observations by the surveyor on the ground;

b. A brief summary of the methods used and the measurements taken to establish control for the coastline survey, including any air photo control points;

c. Highlight any particular difficulties encountered and any additional work required; provide an indication for the time scale before resurvey may be necessary;

d. Provide an assessment of the co-ordination errors of surveyed objects and marks;

e. Describe all new control points established with an indication of life expectancy, detail previous control points reoccupied and assessed material state; full station descriptions should be rendered for all points;

f. Comprehensive details of the results obtained should be rendered; all plots, photographic views, video records, harbour and port information reports, samples obtained and survey records to be rendered should be listed;
g. The sources from which names have been obtained should be reported, if locally produced modern maps have been used to confirm names, copies should be forwarded with the Report of Survey;
h. The survey should comment on the completeness of the survey and indicate where further work is required.

As with all survey records, it is vital that all coastline survey records are cross-checked to ensure no anomalies exist and that consistency between quoted details is achieved. All records to be rendered must be carefully checked and listed to ensure a clear understanding of the survey can be generated back at the Hydrographic Office.

6. DATA PROCESSING

The care and attention devoted to work in the field must be extended to all aspects of data processing and to the careful and legible annotation of all original material used to generate the final records. The underlying principle to be observed in compiling records of any survey is that they must be entirely intelligible to any person having a sound knowledge of the type of survey concerned. The preparation of all data in the established manner, neatly, concisely and accurately, is absolutely vital; terminology should be in accordance with the definitions in IHO Publication S-32.

6.1 Bathymetric

The output of the bottom detection process, when using MBES, is an accurate time and angle of arrival for each measured depth. These 2 parameters are used as inputs to the ray tracing algorithms which convert them into an accurate depth and along/across track distance using information about the SV profile to make the calculation, which is normally carried out in near real time.

The sheer volume of gathered data with MBES invariably means that area based processing is the only realistic method. SBES can be processed in a similar manner, however line-by-line processing with direct comparison against the echo trace, if generated, is often more appropriate.

The use of satellite navigation systems has reduced the amount of post-processing and track editing necessary, however careful online QC of the navigation system output and statistics is vital to ensure that the positional standards are achieved.

SV column profile data should be obtained at regular intervals during survey operations, particularly if continuous SV observations are not being acquired and applied. The interval between SV observations will be determined by the environmental dynamics assessed from the initial temporal and spatial data obtained in accordance with paragraph 4.4.5.

All algorithms used for data editing should be recorded and included in the RoS to enable a clear understanding of the processing procedure to be gained by the Hydrographic Office. The smoothing and filter parameters may be detailed in the Hydrographic Specification or standardised by the Hydrographic Office, any variations should be justified in the RoS. Excessive filtering and smoothing should be avoided.

The editing and processing procedure must follow a logical path with a clear audit trail to enable all actions and parameters to be checked and approved, a robust QC routine must be in operation throughout, which should allow for comparison against previous surveys, published charts or validated survey data, and against recently obtained adjacent survey data. Careful inspection of the crossline/checkline data should be undertaken, a statistical plot should be produced and any differences > v2 x sounding error budget should be investigated.

Observed tide gauge and pole readings should be reduced to sounding datum at the tide station using the values obtained for each when they were installed. Any co-tidal time and range factors should be applied to reduce the observed tides to the values at the survey area.

A daily comparison of the tide gauge and tide pole should be carried out by comparing simultaneous readings. After reduction to sounding datum, simultaneous tide pole and tide gauge readings should be compared to ensure that the tide gauge was recording correctly.

Reduced tide gauge readings should be plotted and the resulting tide curve compared to the predicted curve at the tide station to ensure consistency of data and to ensure that the tide gauge was recording correctly. Unless 'steps' in tides are expected, curves may be smoother, large or repetitive steps should be noted in the RoS and the tide gauge checked for any malfunctions.

When using telemetry links with established gauges, a comparison of actual gauge and telemetry values should be conducted during the 25 hour pole/gauge comparison and then at intervals throughout the survey.

Data density should be aligned with the object detection requirements, which will determine the grid size to be used. In surveys for navigational charting, shoal bias is the normal criteria; however there are occasions when mean depth will be appropriate. The Hydrographic Specification and the Hydrographic Office policy should provide guidance on the requirements.

The Hydrographic Specification should detail the final survey product presentation including contour intervals; however full use should be made of the numerous visualisation tools available to aid in the checking and QC of the survey. Some of the visualisation formats are:

a. Sounding plots;
b. Contour plots;
c. Digital colour sounding and contour displays;
d. 3D colour depth surfaces;
e. 3D grey scale sun illuminated surfaces.

If 3D grey scale sun illuminated surfaces are created, the surface should be viewed from 2 perpendicular directions to highlight any anomalies or artefacts which may need further investigation.

Note should be made of all features and least depths for comparison against sonar records, if obtained. In any event, an assessment should be made as to whether further investigation is required.

6.2 Seafloor Characterisation

The texture data deduced from the sonar trace or backscatter data will be merged into a mosaic either using an automated system or by creating a hand drawn seabed texture collector. In both cases interpretation will be guided by the grid of seabed samples taken earlier in the survey.

Initial detail should include the crests of sandwaves with heights, the positions of obstructions with heights and the start and end of rock outcrops and pinnacles. Texture detail should be plotted to define the limits of texture boundaries. Care should be taken to ensure conformity of texture details to previously survey areas.

Instances will occur when additional seabed samples are required to clarify complex seabed texture areas. Sufficient additional seabed samples should be obtained until confidence has been achieved that the seabed has been accurately classified.

The quality and totality of sonar coverage should be determined by inspection of:

a. A plot of the ship's track, inspection will reveal any gaps in coverage due to poor line keeping or excessive towing speed, these gaps should be re-run;

b. Line quality control data which will indicate whether the maximum tow speed has been exceeded, areas of excessive should be re-run;

c. Sonar traces should be inspected to ensure that the towfish was deployed at the correct height above the seabed and that good data was being recorded, any areas with suspect data should be re-run.

6.3 Feature Detection

Wreckage or artificial obstructions which stand proud of the surrounding seabed may constitute a hazard to shipping or to submarines navigating over continental shelf areas. All such objects must be located, examined and recorded.

During the initial sonar sweep of the area, sonar traces must be carefully examined and all contacts likely to represent obstruction carefully noted. Contacts should be recorded methodically recorded:

a. Sonar trace number;
b. Julian day and time;
c. Contact number (should be consecutively through the survey);
d. Fix details;
e. Port/Starboard channel;
f. Slope range;
g. Layback;
h. Height of towfish above seabed;
i. Assessment of contact;
j. Further action required (ie. investigate, interline, quick look, no further action).

The magnetometer trace and depth data should be carefully examined to provide supporting evidence.

On completion of an examination the records should be carefully checked to ensure that the process has been conducted thoroughly. The following points should be considered when assessing the thoroughness of a examination:

a. As long as a wreck, foul or obstruction continues to be a hazard to navigation or other marine activity, it must appear on the chart;
b. Any objects described as giving 'non-sub' echoes or which constitute a 'foul' on the seabed must be found, classified, fixed and recorded; whether dangerous to shipping or not, they must be disproved;
c. The onus is to classify or disprove every charted wreck, foul, obstruction or contact previously described as 'non-sub'; unless disproved beyond doubt, they must remain on the chart.

The satisfactory examination of every significant object located during a survey is a major factor in deciding whether an area has been fully surveyed.

6.4 Ancillary/miscellaneous Observations

Tidal stream, gathered during the survey via whatever methods, should be assessed for validity and consistency. Where previous tidal stream data exists, the new observations should be compared to ensure continuity and uniformity. Where no previous data exists, observations should be inspected to ensure that they are in agreement with tidal streams experienced during the survey, which can be assessed from comparisons between the courses and speeds set against the courses and speeds made good.

It is usual for the analysis of observed tidal stream data, for inclusion on charts, to be conducted by the Hydrographic Office.

The mean position of floating navaids should be calculated from the observed ebb and flood positions, unless a position of the sinker/anchor can be deducted from the swathe bathymetry.

The position, characteristic, sectors and physical description of each fixed or floating navaid should be compared against the published chart, the relevant List of Lights and Sailing Directions as a gross error check. It is important to verify that the derived positions for navaids meet the required standards.

The final list of fixed and floating navaids observed for and checked within the survey area should be compared with the original listing created from the published charts, List of Lights and Sailing Directions, in accordance with paragraph 4.8.3, to ensure complete coverage of all navaids.

Any variations from the published data and confirmed with responsible authorities in accordance with paragraph 4.8.7 should be reported immediately by signal/e-mail to the Hydrographic Office and followed up with a Hydrographic Note.

The details of port radio operations obtained during the survey, in accordance with paragraph 4.8.8, should be checked against the relevant List of Radio Signals and Sailing Directions.

Details of any marine life, bioluminescence, secchi, deep scattering layer observations should be rendered. Details of other features, such as pock marks and brine lakes, and any sediment samples should be given with a description on how the observations were made.

If ocean fronts, eddies or internal waves have been investigated, details of their locations, the type of feature, the methods used and the sensors employed should be provided. Comment on how the data has been rendered and any conclusions made.

6.5 Compliance with the Plan

An assessment should be made of the completeness of the survey and its compliance with the Hydrographic Specification and the original plan. Any areas which require further investigation, including areas incompletely surveyed or the Hydrographic Specification requirements have not been achieved, should be identified and what actions are required to rectify the shortfalls, which may be due to equipment limitations or physical conditions. Any further work required should be highlighted and recommendations should be made on how it can be successfully approached in the future.

7. DATA RENDERING

7.1 The Report of Survey

When survey material of any sort is rendered to the Hydrographic Office, it must be accompanied by a report, in some form or other, of how it was obtained. In a few cases, such as Hydrographic Notes, this may be relatively brief, but in the vast majority of cases, the Report of Survey forms the core of the survey data, and should remark on every aspect of the survey and on all other data being rendered with it. For conventional bathymetric surveys, it is often divided into two parts; an example of their contents is contained at Appendix 2 to this chapter.

The Report of Survey is the principal means by which the Surveyor in charge approves the contents of ALL survey records and is thus a very important document, and the Surveyor must take considerable care in its presentation. It must give a clear and comprehensive account of how the survey was carried out, the results achieved, the difficulties encountered and the shortcomings. Emphasis within Part 1 should be placed on the analysis of achieved accuracies and whether the specifications called for in the survey specification and IHO Publication S-44 standards have been met. Part 2 contains the necessary technical discussion to support opinions expressed in Part 1. It should be borne in mind that it is often just as important to say what was not done and why, as to say what was done and how.

A thorough Report of Survey can reduce the need for subsequent correspondence between Hydrographic Office and survey unit, which otherwise may be necessary to elucidate points which have not been covered in a less exhaustive report. The example at Appendix 2 provides an outline of the material that it is felt advisable to include along with a format suitable for bathymetric surveys. It is useful to note against any paragraphs which are not applicable to a particular survey a brief statement of the form 'No ……. observations were conducted'.

The Report of Survey is as much a fair record of the survey as any other, and must be compiled and present with as much care, neatness and accuracy. The manner and format in which it is rendered to the Hydrographic Office will vary according to national requirements.

The Report of Survey and associated survey data will be rendered to the Hydrographic Office where it will undergo a rigorous validation and appraisal process. It is recommended that a complete copy of the rendered data should be kept in the survey unit until all queries have been answered. It should be remembered that the verified Report and data set will be archived and remain as the definitive data source for the creation of future products.

7.2 Data Requirement

The Hydrographic Specification will articulate the reason for the survey and the primary product requirement of the customer, ie safety of navigation or cable/pipeline route surveys will be bathymetry driven, minewarfare and archaeological surveys will be object detection led whilst environmental surveys may be seabed texture and water column based. The Hydrographic Office will detail what data is to be rendered on completion of the survey and to what time scale.

7.3 Data Format and Density

Most Hydrographic organisations have detailed standards for data format and data density to meet their requirements. The Hydrographic Specification will detail any variations to these standards in the data format, which will depend upon the survey system to be used and the systems available for verification and validation of the rendered data. The modifications to the data density will be stipulated in the Hydrographic Specification.

Modern visualisation tools have allowed significantly greater flexibility to the surveyor in data presentation, both digitally and graphically. Care must be taken to ensure that the inherent cautiousness of hydrographic surveying is not blinded by the multi-coloured imagery which can be created with relative ease and minimal human interaction.

7.4 Media Requirement

The Hydrographic Specification will detail in which media the survey data is to be rendered and whether it is a fully digital survey or if fair graphics are required by the customer. The common media used are:

a. DVD;
b. CD-Rom;
c. DAT Tape;
d. Fair graphics on Ozatex/Cronaflex;
e. Paper records.

Whatever the media, great care should be taken with the transmission and handling to ensure that the data arrives at its final destination uncorrupted and undamaged. Much of the data will be unique, will form the basis for amending and maintaining charts and publications until the area is resurveyed (probably very many years in the future), and will become part of the nation's archives as public records. It follows that this material must be afforded the highest degree of security at all times, for to lose a part or all of it would clearly be very expensive in time, effort and material.

REFERENCES

Edition 1/04 2004	*Hydrographic Quality Assurance Instructions for Admiralty Surveys*	UK Hydrographic Office
17th Edition 1996	*General Instructions for Hydrographic Surveys*	UK Hydrographic Office
11th Section 1968	*Admiralty Manual of Hydrographic Surveying Volume 1*	UK Hydrographic Office
1969 to 1973	*Admiralty Manual of Hydrographic Surveying Volume 2*	UK Hydrographic Office
PP24 (Parts 1 & 2) 1987	*The Use of Side-scan Sonar for Hydrographic Surveying and the Gathering of Bottom Texture Information*	UK Hydrographic Office
5th edition – 1994	*"Hydrographic Dictionary" S–32*	International Hydrographic Organization, Monaco
4th edition – 1998	*"IHO Standards for Hydrographic Survey"* *S–44*	International Hydrographic Organization, Monaco
29 April 1998	*"IHO Standards for Hydrographic Survey" Supplement to S–44* *Draft enclosed in letter IHB File N. S3/7198*	International Hydrographic Organization, Monaco
1 January 2002	*EM 1110-2-1003"Hydrographic Surveying"*	U.S. Army Corps of Engineers, Department of the Army, Washington.
1 July 2003	*EM 1110-1-1003"NAVSTAR Global Positioning System Surveying"*	U.S. Army Corps of Engineers, Department of the Army, Washington.
Fourth Edition 4 July 1976	*"Hydrographic Manual"*	Melvin J. Umbach Rockville, Md. U.S. Department of Commerce National Oceanic and Atmospheric Administration (NOAA) National Ocean Service (NOS)
March 2003	*NOS Hydrographic Surveys "Specifications and Deliverables"*	U.S. Department of Commerce National Oceanic and Atmospheric Administration (NOAA) National Ocean Service (NOS)

A.A. 2002/2003	“*Lezioni di Idrografia teorica ed operativa*”	Luigi Sinapi Napoli
1976	“*Electronic Surveying and Navigation*”	Simo H. Laurila John Wiley & Sons, Inc New York (USA)
January 1983	“*Electronic Surveying in practice*”	Simo H. Laurila John Wiley & Sons, Inc New York (USA)
1991	“*Radionavigation system*”	Börje Forssell Prentice Hall International (UK) Ltd
July 1974	“*Sea Surveying*”	Alam E. Ingham John Wiley & Sons, Inc New York (USA)

BIBLIOGRAPHY

Edition 1/04 2004	*Hydrographic Quality Assurance Instructions for Admiralty Surveys*	UK Hydrographic Office
17th Edition 1996	*General Instructions for Hydrographic Surveys*	UK Hydrographic Office
11th Section 1968	*Admiralty Manual of Hydrographic Surveying Volume 1*	UK Hydrographic Office
1969 to 1973	*Admiralty Manual of Hydrographic Surveying Volume 2*	UK Hydrographic Office
PP24 (Parts 1 & 2) 1987	*The Use of Side-scan Sonar for Hydrographic Surveying and the Gathering of Bottom Texture Information*	UK Hydrographic Office
PP25 1990	*The assessment of the Precision of Soundings*	UK Hydrographic Office
1998	*“La georeferenziazione delle informazioni territoriali”*	Luciano Surace Estratto dal "Bollettino di geodesia e scienze affini"
Prima Edizione – 2000	*“GPS Principi Modalità e Tecniche di Posizionamento”*	A. Cina Celid
Prima Ristampa – 2001	*“Topografia”*	L. Costa Cooperativa Libraria Universitaria – Genova
Terza Edizione – 1949	*“Manuale di Idrografia per la costruzione delle carte marine*	Manoia G. Romagna Accademia Navale di Livorno
II 3100. Quinta Edizione – 1992/Prima Ristampa – 1998	*“Manuale dell’Ufficiale di Rotta”*	Istituto Idrografico della Marina, Genova
NorMas FC 1028 Seconda Edizione – 1978	*“Norme di Massima per i Rilievi Idrografici”*	Istituto Idrografico della Marina, Genova
5th Edition – 1994	*“Hydrographic Dictionary” S–32*	International Hydrographic Organization, Monaco
4th Edition – 1998	*“IHO Standards for Hydrographic Survey”* *S–44*	International Hydrographic Organization, Monaco
29 April 1998	*“IHO Standards for Hydrographic Survey” Supplement to S–44* *Draft enclosed in letter IHB File N. S3/7198*	International Hydrographic Organization, Monaco

1 January 2002	*EM 1110-2-1003 "Hydrographic Surveying"*	U.S. Army Corps of Engineers, Department of the Army, Washington.
1 June 2002	*EM 1110-1-1004 "Geodetic and Control Surveying"*	U.S. Army Corps of Engineers, Department of the Army, Washington.
1 July 2003	*EM 1110-1-1003 "NAVSTAR Global Positioning System Surveying"*	U.S. Army Corps of Engineers, Department of the Army, Washington.
31 August 1994	*EM 1110-1-1005 "Topographic Surveying"*	U.S. Army Corps of Engineers, Department of the Army, Washington.
Fourth Edition 4 July 1976	*"Hydrographic Manual"*	Melvin J. Umbach Rockville, Md. U.S. Department of Commerce National Oceanic and Atmospheric Administration (NOAA) National Ocean Service (NOS)
March 2003	*NOS Hydrographic Surveys "Specifications and Deliverables"*	U.S. Department of Commerce National Oceanic and Atmospheric Administration (NOAA) National Ocean Service (NOS)
A.A. 2002/2003	*"Lezioni di Idrografia teorica ed operativa"*	Luigi Sinapi Napoli
1976	*"Electronic Surveying and Navigation"*	Simo H. Laurila John Wiley & Sons, Inc New York (USA)
January 1983	*"Electronic Surveying in practice"*	Simo H. Laurila John Wiley & Sons, Inc New York (USA)
1991	*"Radionavigation system"*	Börje Forssell Prentice Hall International (UK) Ltd
July 1974	*"Sea Surveying"*	Alam E. Ingham John Wiley & Sons, Inc New York (USA)