CHAPTER 2
POSITIONING

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.

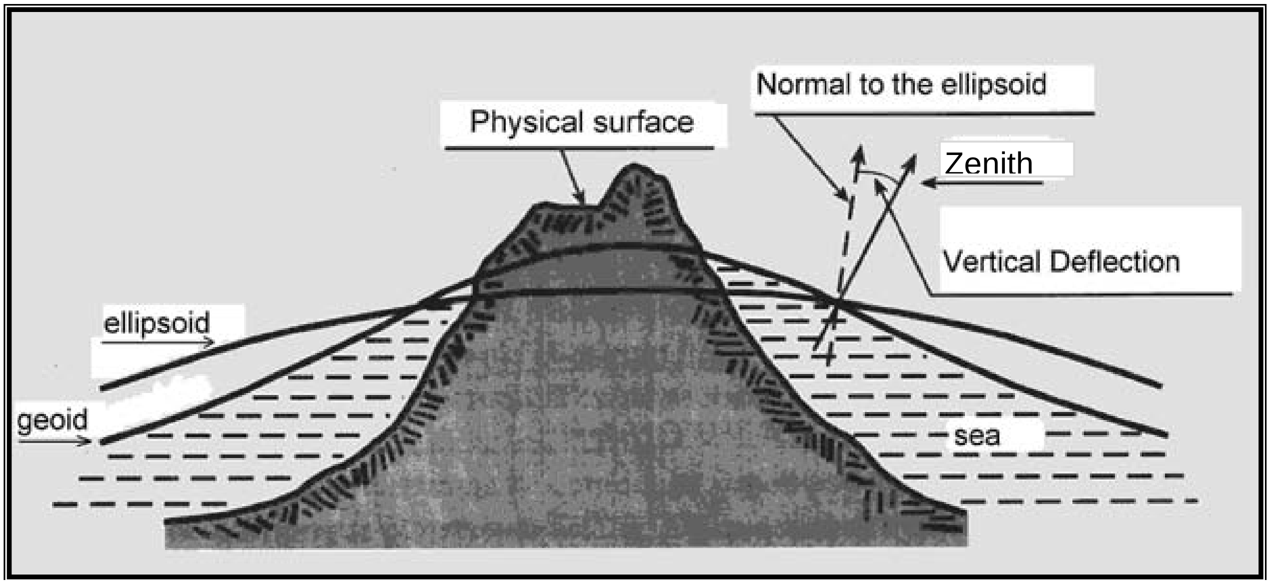


Fig. 2.1 "The Earth"

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:* $e = [(a^2 - b^2)^{1/2} / a]$.

Figure 2.1 shows the general relationship between geoid, ellipsoid and the physical shape of the earth. Figure 2.2 shows the structure and parameters of the ellipsoid.

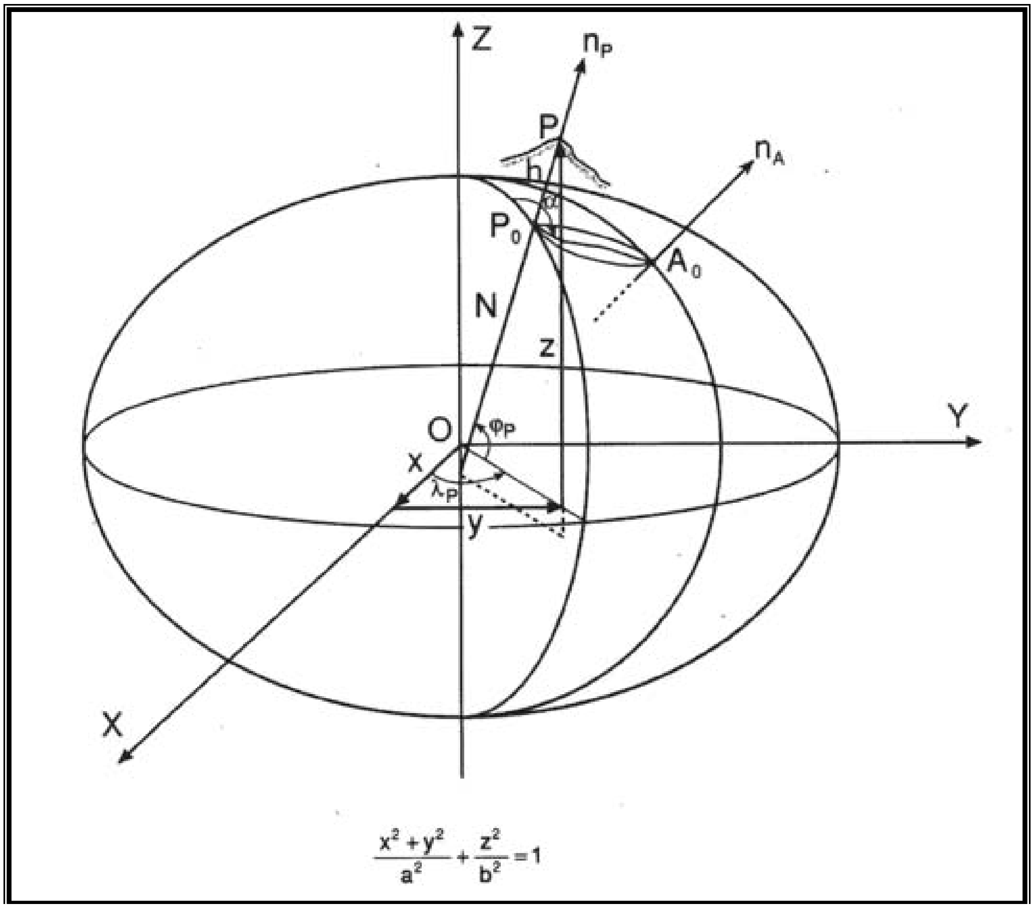


Fig. 2.2 "The Ellipsoid"

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 8 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 (local ellipsoid) and the vertical (local geoid / mean sea level). Figure 2.3 attempts to show this relationship.

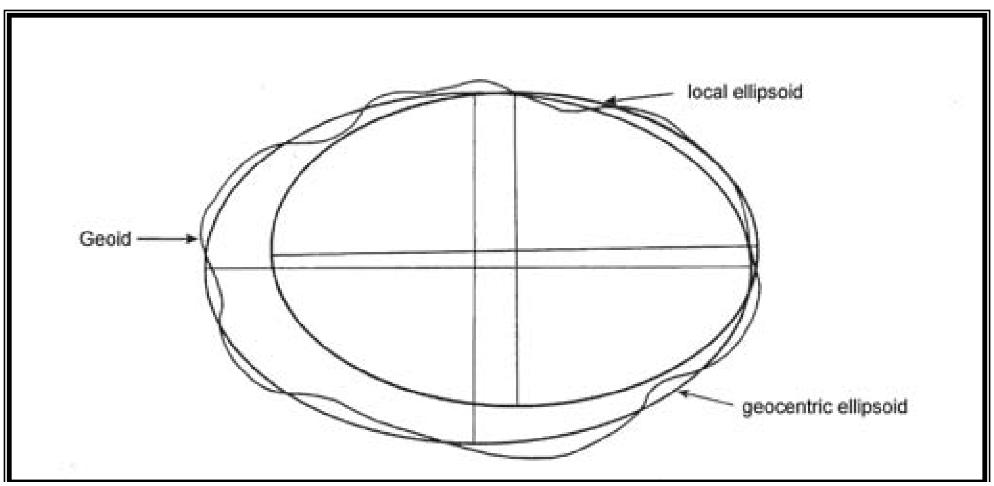


Fig. 2.3 "Datum orientation"

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. See Figure 2.4 for a graphical depiction of the relationship between 2 ellipsoids.

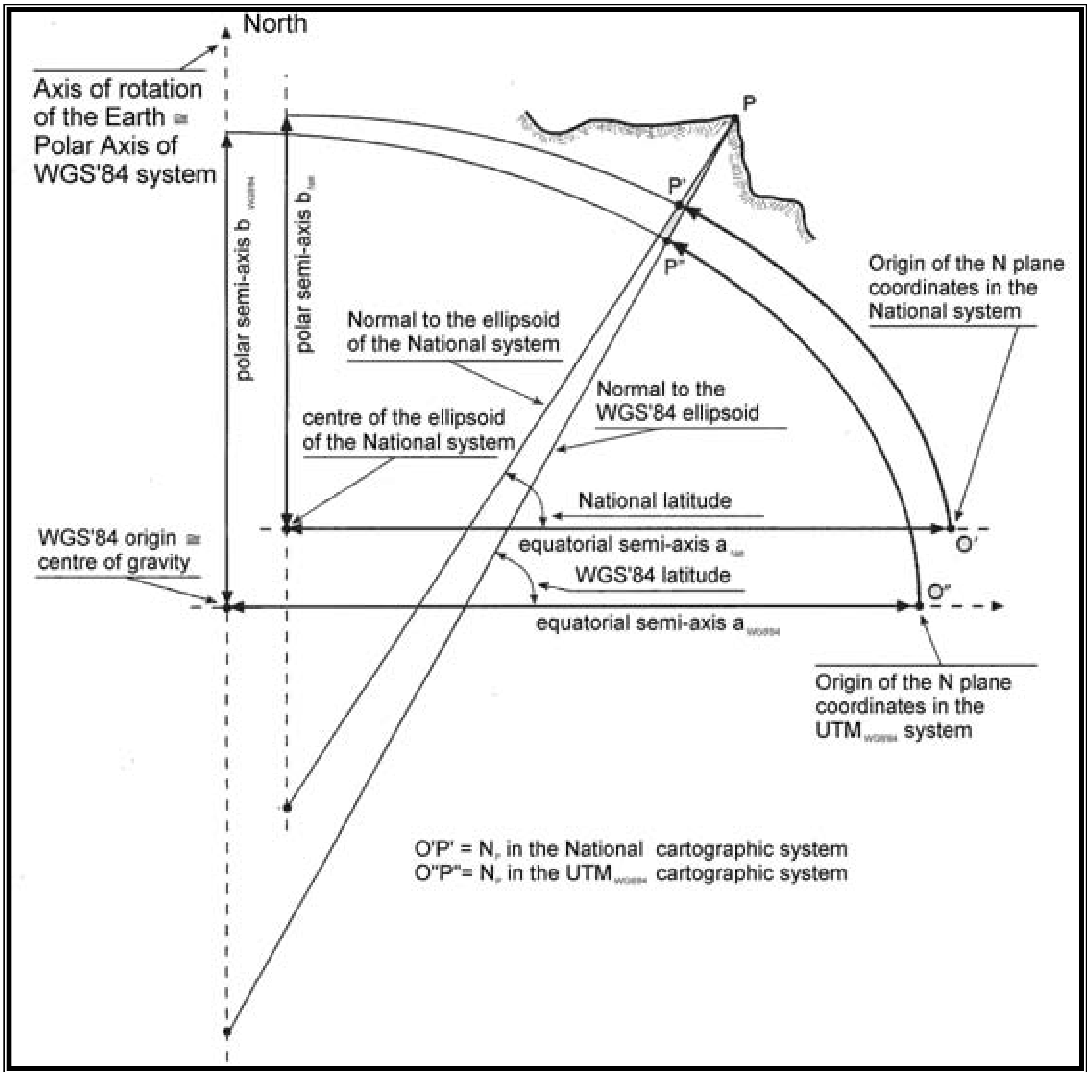


Fig. 2.4 "Horizontal Datum orientation"

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.

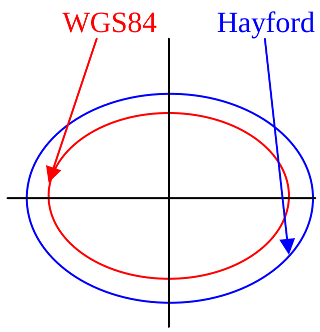
2.2.3 Datum transformation

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:

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

They connect the geodetic co-ordinates **φ**, **λ**, & **h** related to the ellipsoid with semi-axis '**a**' and ellipticity (or compression or flattening) **α**, 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 ($\geq$ 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**.

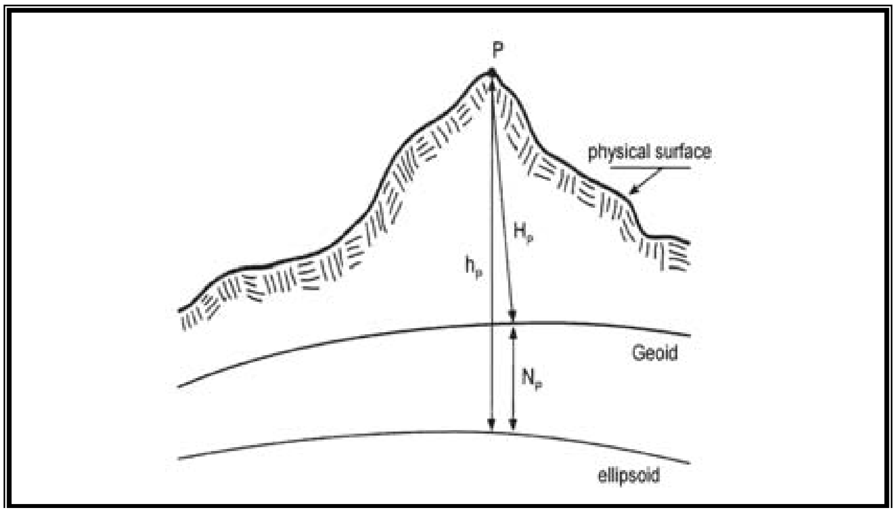


Fig. 2.5 "Vertical Datum"

In a first approximation, to within a few millimetres:

$$h_p = H_p + N_p \quad \textbf{(2.3)}$$

$\mathbf{h_p}$ 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 distorted 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:

$$\begin{aligned} m_l &= dl'/dl \\ m_s &= dS'/dS \\ m_\alpha &= \alpha' - \alpha \end{aligned} \qquad \textbf{(2.4)}$$

where with **dl'**, **dS'**& **α'** being the geometric elements belonging to the paper and with **dl**, **dS** & **α** the corresponding elements for the ellipsoid (**α'- α** 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 and 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.

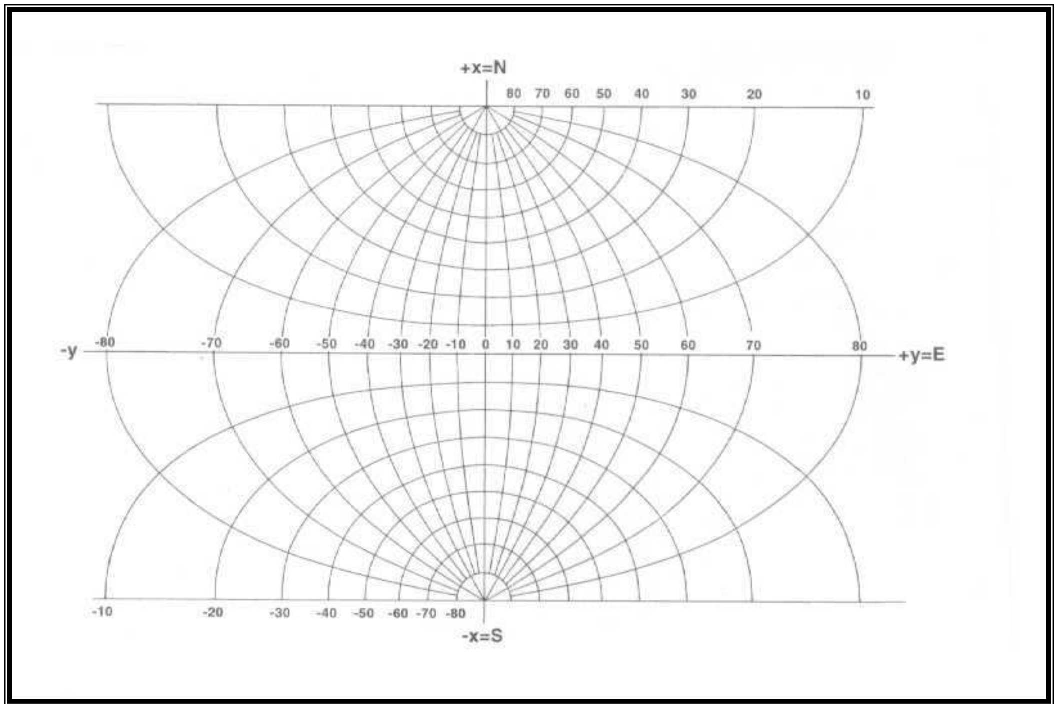


Fig. 2.6 "Geographic grid"

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).

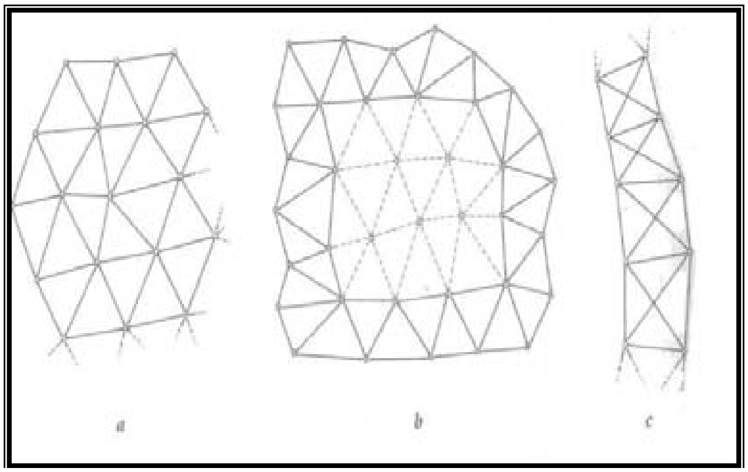


Fig. 2.7

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: α_1, β_1, γ_1 of the ABC triangle; α_2, β_2, γ_2 of the ABD triangle, and so on.

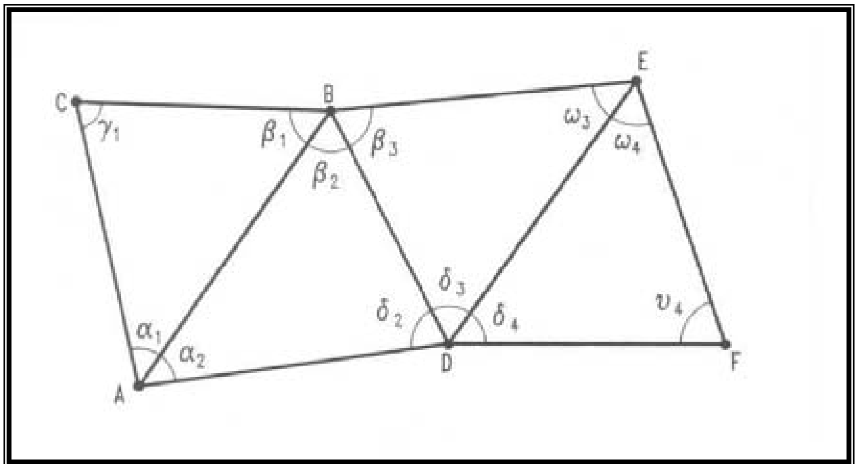


Fig. 2.8

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:

$$\varepsilon_\alpha = \left| \sum \alpha_i - 180° \right| \le t_\alpha \qquad \textbf{(2.5)}$$

where the summation $\boldsymbol{\Sigma\alpha_i}$ 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:

$$\overline{AB} = \overline{AC} \cdot \frac{\sin \gamma_1}{\sin \beta_1} \tag{2.6}$$

$$\overline{BC} = \overline{AC} \cdot \frac{\sin \alpha_1}{\sin \beta_1} \tag{2.7}$$

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. $\boldsymbol{\alpha_2}$), has to be equal to the difference of the two angles of direction measured on the AD base and on the AB base:

$$\alpha_2 = (AD) - (AB) \tag{2.8}$$

from which:

$$\alpha_2 - (AD) + (AB) = 0 \tag{2.9}$$

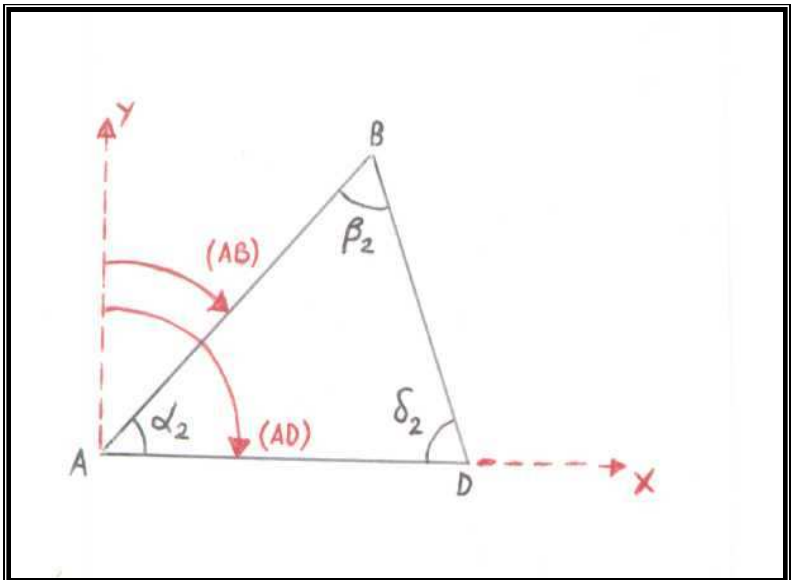


Fig. 2.9

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:

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

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

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:

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

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:

$$\alpha_2 - (AD) + (AB) = \nu_i \tag{2.13}$$

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 ($\mathbf{X_D}^\circ$, $\mathbf{X_A}^\circ$, $\mathbf{Y_D}^\circ$, $\mathbf{Y_A}^\circ$), 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 ($\mathbf{x_D - x_A}$) and ($\mathbf{y_D - y_A}$) 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 principles of least squares enable the writing of the algebraically determined system of normal equations for the unknowns $\mathbf{x_i}$ and $\mathbf{y_i}$.

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 ($\boldsymbol{\mu}$) 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 $\mathbf{P_\alpha}$ and $\mathbf{P_d}$) emphasises that:

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

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

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:

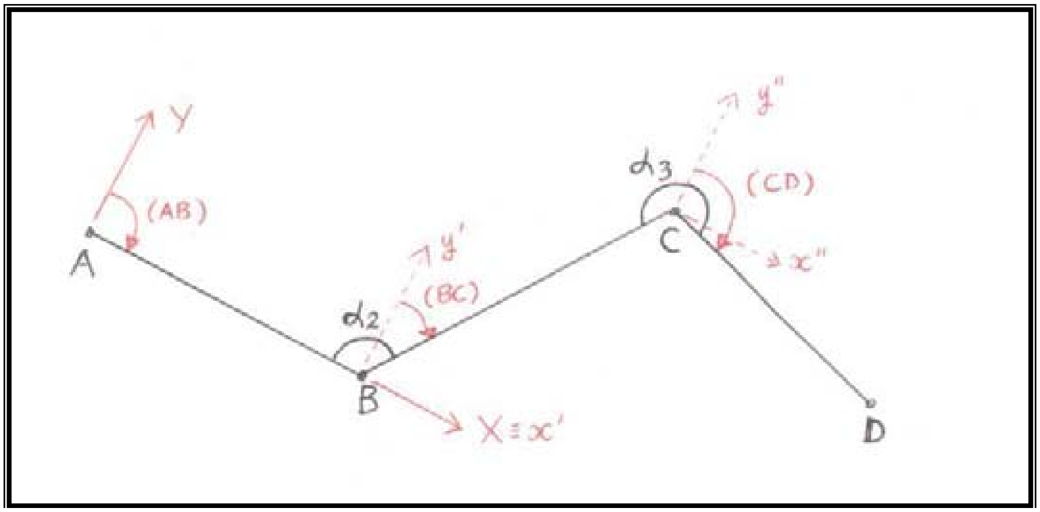


Fig. 2.10

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:

$$(BC) = (AB) + \alpha_2 \pm 180° \tag{2.18}$$

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:

$$x_{C(B)} = \overline{BC} \cdot \sin(BC) \tag{2.19}$$

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

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:

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

and so on for the following points.

It is important to notice that having the number of the measurements (angles α_A α_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 α_A α_B α_C α_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.

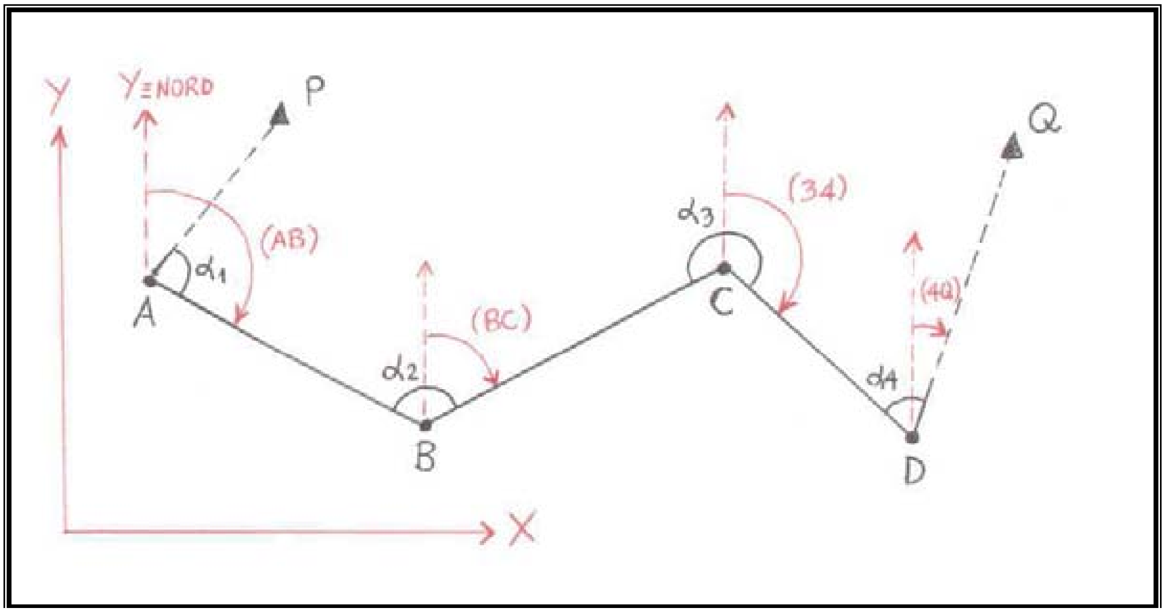


Fig. 2.11

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:

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

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

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

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

$$(DQ) = arctg\,[(X_Q - X_D)/(Y_Q - Y_D)] \qquad \textbf{(2.24)}$$

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:

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

With the unavoidable presence of accidental errors in the measurements of angles α_A ,α_B ,α_C ,α_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 $\Delta\alpha$. The (2.25) then becomes:

$$(DQ)' - (DQ) = \Delta\alpha \qquad \textbf{(2.26)}$$

remembering that $\Delta\alpha$ has to be smaller than an angular tolerance established for the project.

c. Calculation of the adjusted angles of direction:

$$\begin{aligned} (AB) &= (AB)' - u\alpha \\ (BC) &= (BC)' - 2u\alpha \\ (CD) &= (CD)' - 3u\alpha \\ (DQ) &= (DQ)' - 4u\alpha \end{aligned} \qquad \textbf{(2.27)}$$

gives **uα**, 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:

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

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:

$$\begin{aligned} \sum x' - (X_D - X_A) &= 0 \\ \sum y' - (Y_D - Y_A) &= 0 \end{aligned} \qquad \textbf{(2.29)}$$

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:

$$\begin{aligned} \Delta x &= \sum x' - (X_D - X_A) \\ \Delta y &= \sum y' - (Y_D - Y_A) \end{aligned} \qquad \textbf{(2.30)}$$

Defining **ΔL** as:

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

Δx and **Δy** have to be such that **ΔL** is not greater than an established linear tolerance.

f. Calculation of the partially adjusted co-ordinates:

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

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.

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

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.

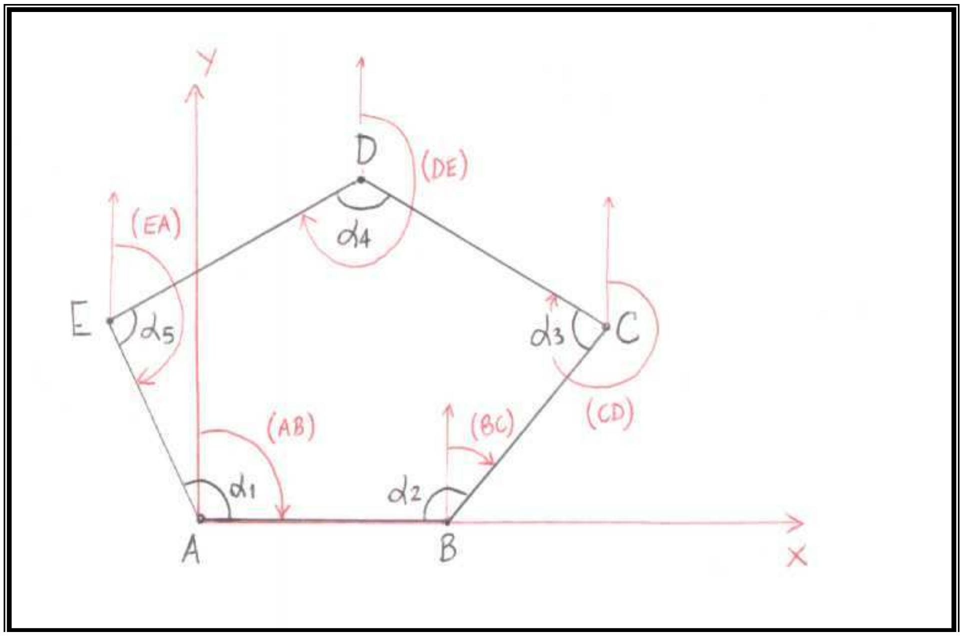


Fig. 2.12

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 **Δα**) are generated:

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

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

$$\alpha_1 = \alpha_1' - \mathrm{u}\alpha$$
$$\alpha_2 = \alpha_2' - \mathrm{u}\alpha \qquad \textbf{(2.35)}$$

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 $\mathbf{\Delta x}$ and $\mathbf{\Delta y}$ (side closing errors). Defined the quantity $\mathbf{\Delta L}$ as:

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

$\mathbf{\Delta x}$ and $\mathbf{\Delta y}$ have to be such that $\mathbf{\Delta L}$ 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).

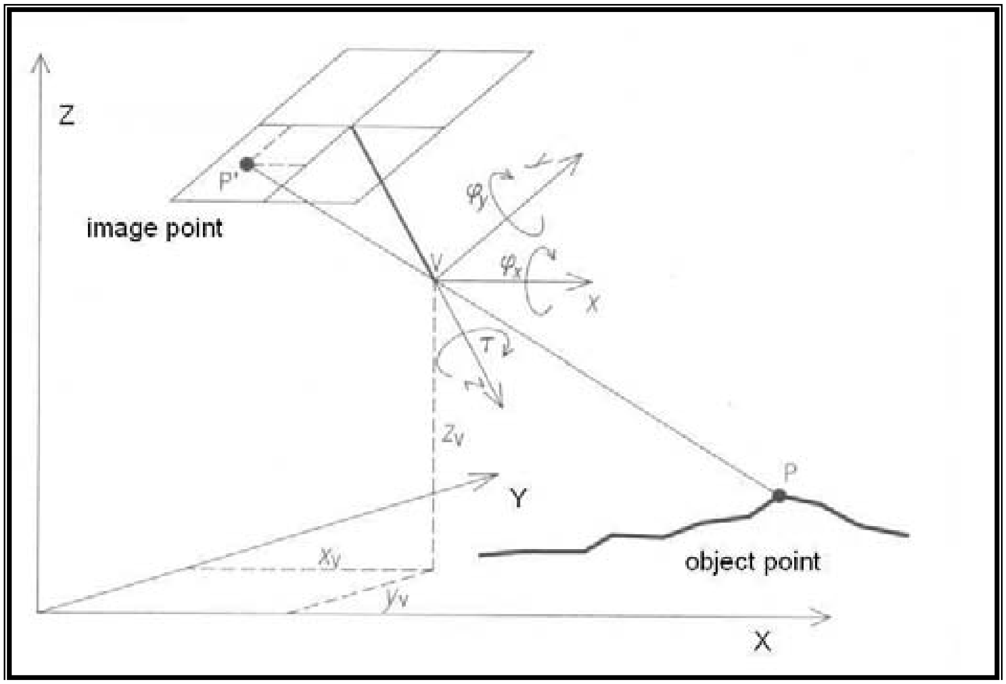


Fig. 2.13 "Twelve parameters determination for an Analogue Restitution"

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.

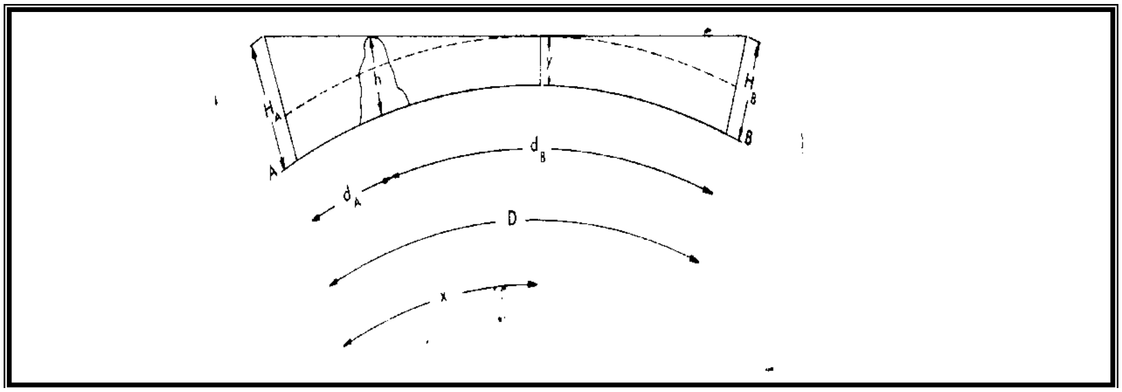


Fig. 2.14 "Intervisibility of Geodetic Station"

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

Therefore,

$$
\begin{aligned}
& H_A - y = Kx^2 \\
& H_B - y = K(D - x)^2 \\
& \textit{Whence} \\
& x = \frac{D}{2} - \left(\frac{H_B - H_A}{2KD} \right) \quad \textit{and} \quad y = H_A - Kx^2
\end{aligned}
\tag{2.37}
$$

$$
\begin{aligned}
& h = y + K(d_A - x)^2 \\
& \textit{Therefore,} \quad h = \frac{d_A H_B}{D} + \frac{d_B H_A}{D} - K d_A d_B
\end{aligned}
\tag{2.38}
$$

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:

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

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

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

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

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

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

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:

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

from which:

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

where:

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

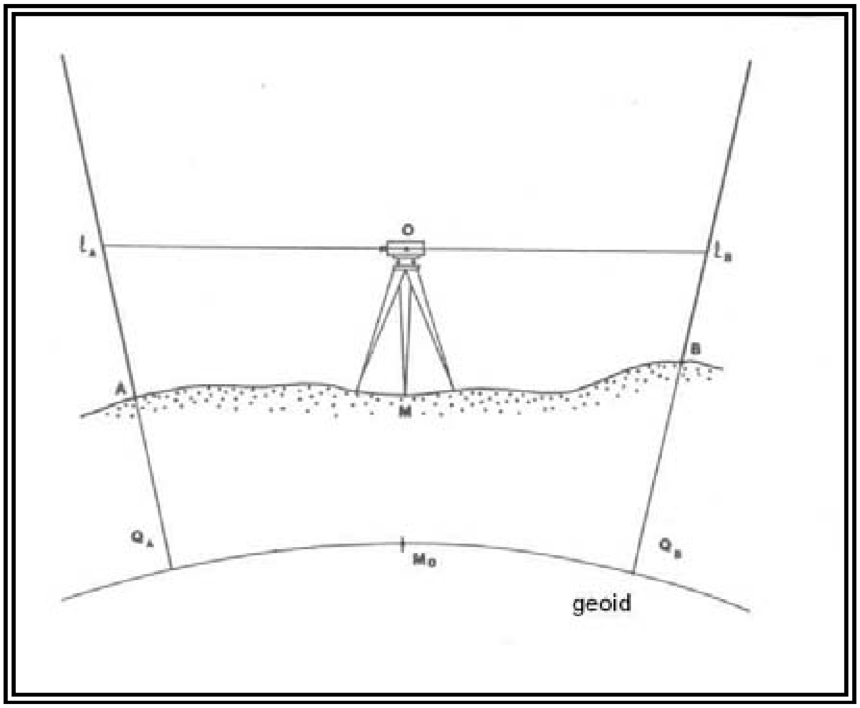


Fig. 2.15 “Geometric levelling”

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.

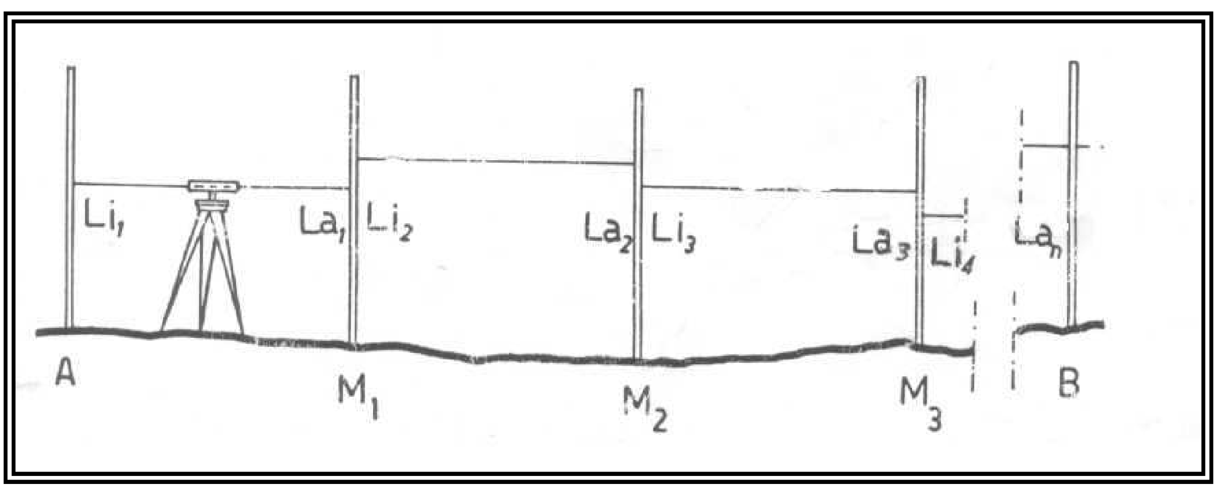


Fig. 2.16 “Level difference among several points”

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:

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

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 $\boldsymbol{\sigma}$. Since the total difference in level is equal to the sum of the partial difference of elevations, subsequently determined, the mean error $\boldsymbol{\sigma_t}$ of the whole levelling is:

$$\sigma_t^{\,2} = \sigma_1^{\,2} + \sigma_2^{\,2} + \sigma_3^{\,2} + \ldots\ldots + \sigma_n^{\,2} = n\sigma^2 \qquad \textbf{(2.44)}$$

from which:

$$\sigma_t = \sigma\sqrt{n} \qquad \textbf{(2.45)}$$

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:

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

where:

- p_i: the measure weights inversely proportional to the distances;
- ν_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:

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

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.

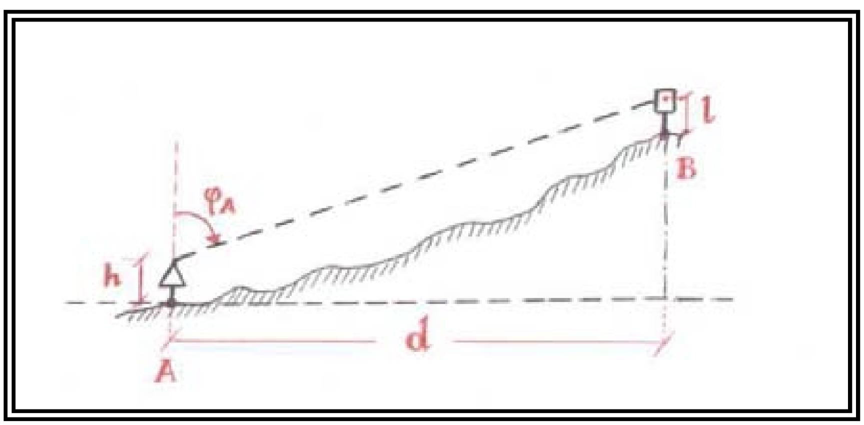


Fig. 2.17 “Trigonometric levelling”

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.

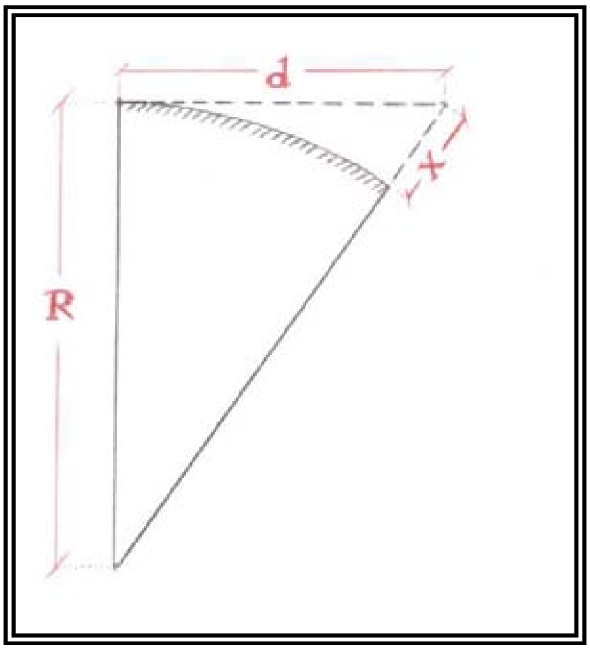


Fig. 2.18 "Correction for sphericity"

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:

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

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 always tends downwards.

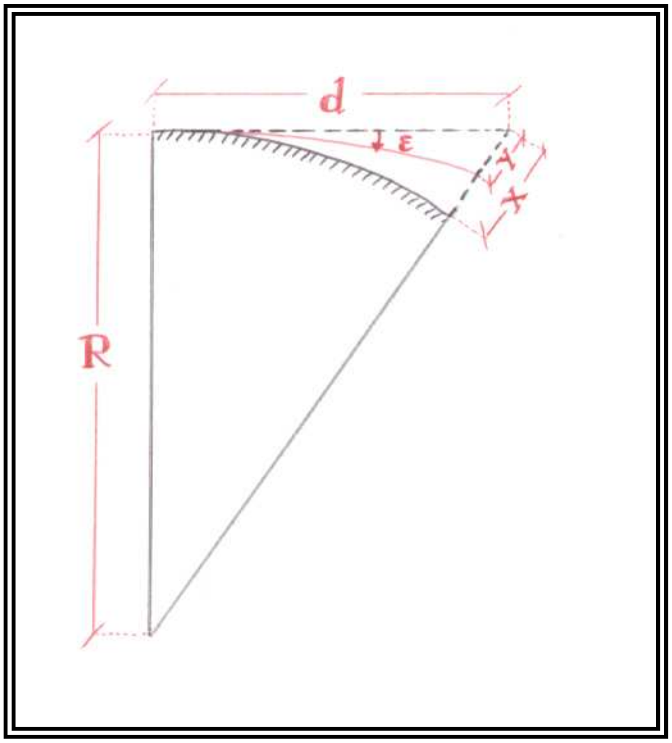


Fig. 2.19 "Correction of refraction"

where:

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

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

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

and therefore, replacing the expression of ε 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 quantity **(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:

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

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: $\mathbf{\Delta_{AB}}$ 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:

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

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

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

in which $\mathbf{m_d}$, $\mathbf{m_{\varphi A}}$ 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**, $\mathbf{\varphi_A}$ and **K**, it is obtained that:

$$\frac{\partial \Delta_{AB}}{\partial d} = \cot \varphi_A + (1-K)\cdot \frac{d}{R} (\text{with the second negligible term})$$
$$\frac{\partial \Delta_{AB}}{\partial \varphi_A} = -\frac{d}{\sin^2 \varphi_A} \qquad \textbf{(2.57)}$$
$$\frac{\partial \Delta_{AB}}{\partial K} = -\frac{d^2}{2\cdot R}$$

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 **α**, the angle of inclination, ($\alpha = 90° - \varphi_A$). With α=0°, the error removes itself. It is however always small (i.e. for $\alpha = \pm 10°$ and d = 5km, the error will be 1.6 cms).
b. in the second term, assigning to **α** a mean value of 10°, the error will depend on the mean error of **φ_A,** the zenithal angle, and from **d**, the distance, (i.e. m$\alpha = \pm 10°$ 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 **δN** is:

$$\delta N = N_{WGS84} - N_{localgeoid} \qquad \textbf{(2.58)}$$

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 **δN** values for at least of three points of which the height is known in both reference systems, with the following expression:

$$a_1 X_i + a_2 Y_i + a_3 = \delta N_i \qquad \textbf{(2.59)}$$

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)[24], 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 ($\boldsymbol{\nu_{air}}$) of the air which depends on temperature, pressure and humidity: $\mathbf{\{c_{air}=c_0/\nu_{air}(t,p,h)\}}$

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.

---

24 (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 **σ = 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.

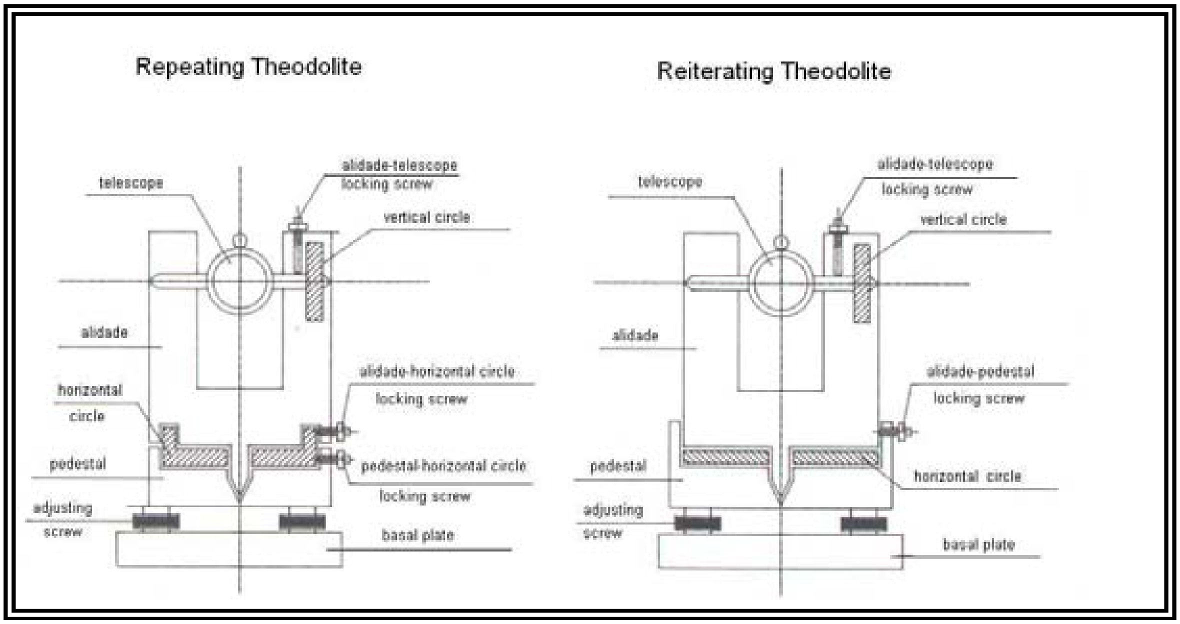


Fig. 2.20 "Theodolites"

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 circular 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 a minimum of 24 satellites, although there are often more, 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 = 10.23 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 (λL1≅19 cm) and 120 x f0 = 1227.60 MHz (λL2 ≅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 1 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. *Induced 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:

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

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 phase measurements for the satellites in sight and transmits its observations and position to the mobile receivers. At the same time the rover stations also perform phase measurements on the same satellites whilst processing the raw data received from the reference station; each rover then computes its position (double and triple differences). 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 signal related 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[25] and Accuracy[26] 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:

---

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

[26] "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[27]: 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[28];
b. Radial error or root mean square error in the distance (**1 σRMS** or **1 DRMS**): with the assumption of equality of the standard deviations around two dimensions (σ_x σ_y)[29], of orthogonality between the axes x and y, of normal and not correlated distributions of error, the following relationship is valid:

$$\mathrm{DRMS} = \sqrt{\sigma_x^2 + \sigma_y^2} = \sqrt{2 \cdot \sigma^2} = 1{,}414 \cdot \sigma \qquad \textbf{(2.65)}$$

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

---

[27] "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)

[28] (see also probable error: IHO S32 – fifth edition 1994 – # 1689)

[29] (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 **Δt** or a difference in phase **Δϕ**. Such measurements are translated into differences of distances (hyperbolic LOPs) or direct distances (circular LOPs) respectively with the relationships:

$$\Delta d = c \cdot \Delta t \qquad \textbf{(2.66)}$$

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

where:

- **Δd**: is the difference of distance;
- **c**: is the speed of propagation of the electromagnetic waves;
- **Δt**: is the difference of measured time;
- **Δϕ**: is the difference of measured phase;
- **f**: is the frequency of the wave on which the measurement is effected Δϕ;
- **n**: is the number of integer cycles of the received wave.

An error in the measurement of **Δt** or **Δϕ** 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 :

$$d_{RMS} = \frac{\sqrt{2\sigma^2}}{\sin\alpha} = \frac{1{,}414 \cdot \sigma}{\sin\alpha} \qquad \textbf{(2.68)}$$

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

Considering therefore constant **σ** 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 **α**. 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.

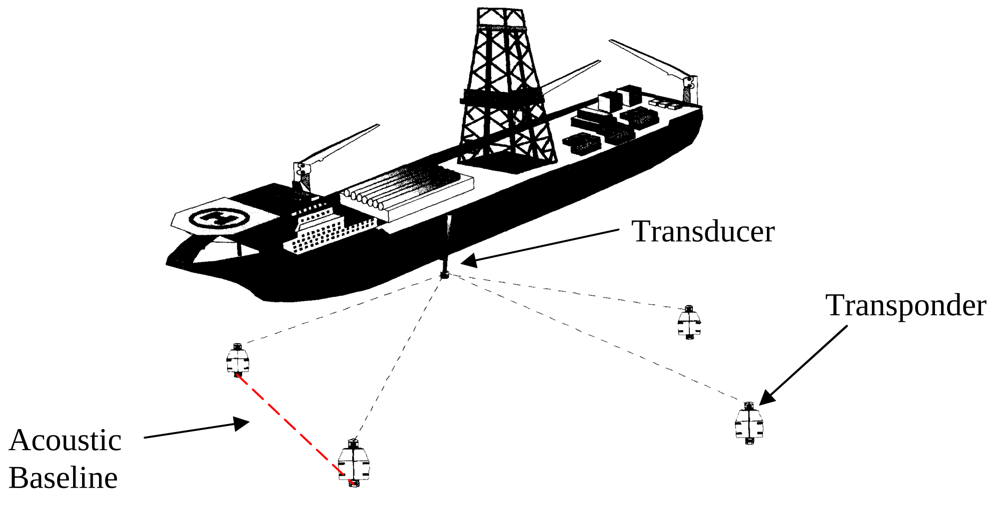


Fig. 2.21 “Long Baseline Meted”

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.

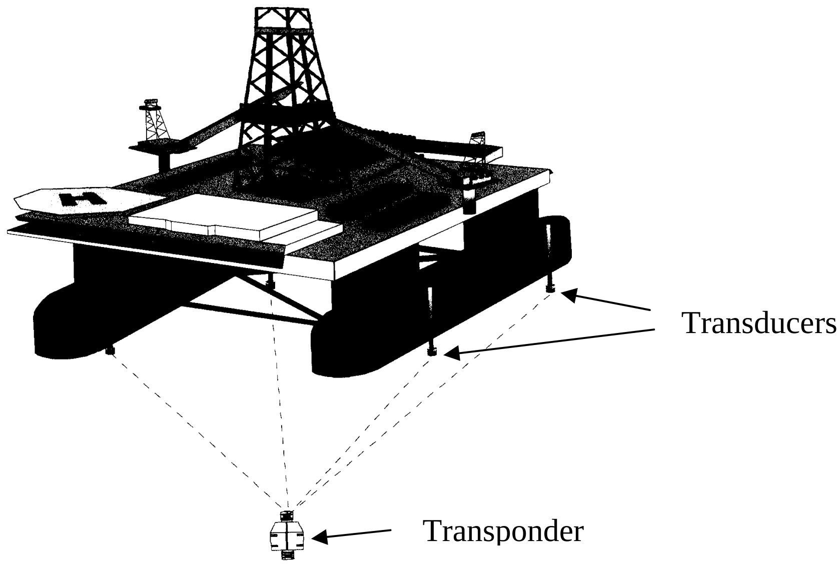


Fig. 2.22 "Short Baseline Meted"

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.

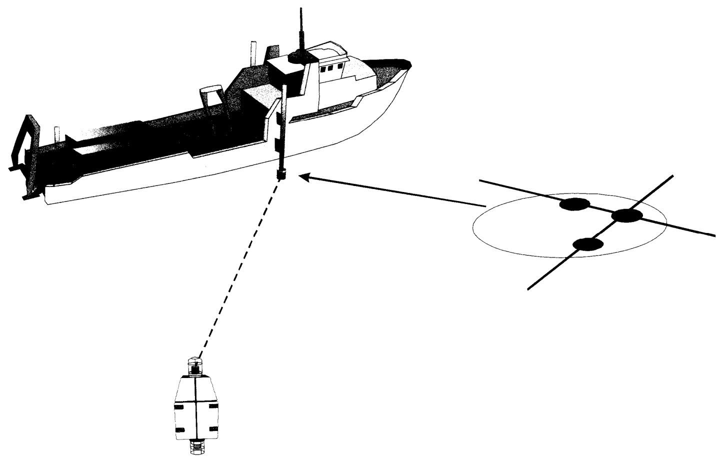


Fig. 2.23 "Ultra Short Baseline Meted"

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)

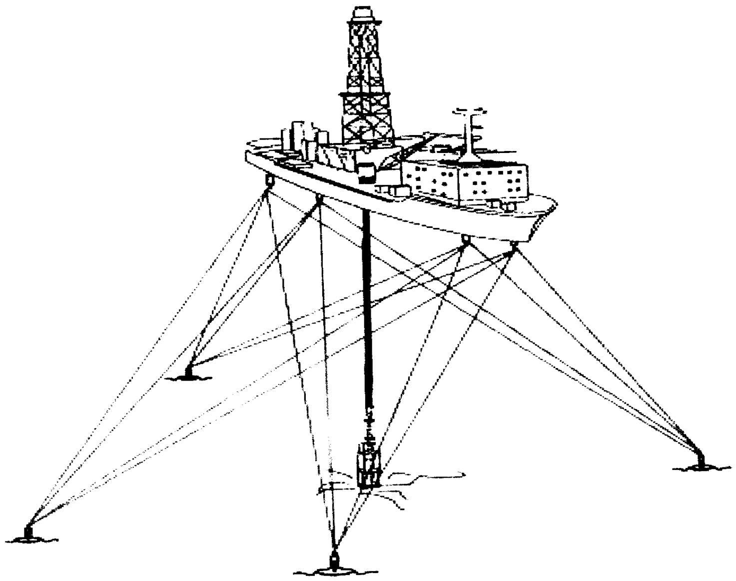


Fig. 2.24 "Combined System (LSUBL)"

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:

$R = \frac{ct}{2}$ and the horizontal distance (Y) can be determined by: $Y = R \sin\theta$

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: $Y = D \tan \theta$

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.

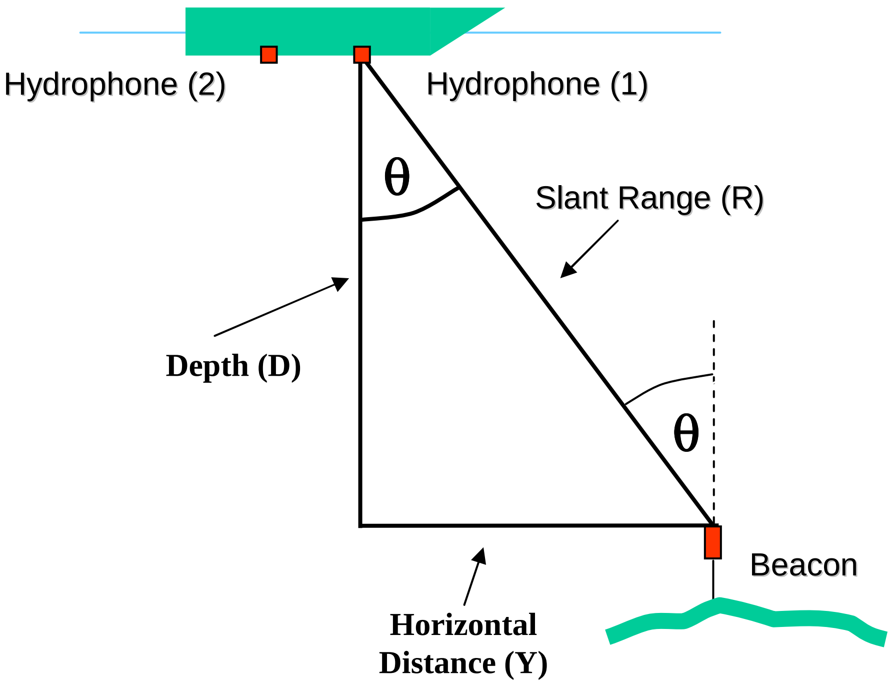


Fig. 2.25 "Range determination"

Angular Measurement

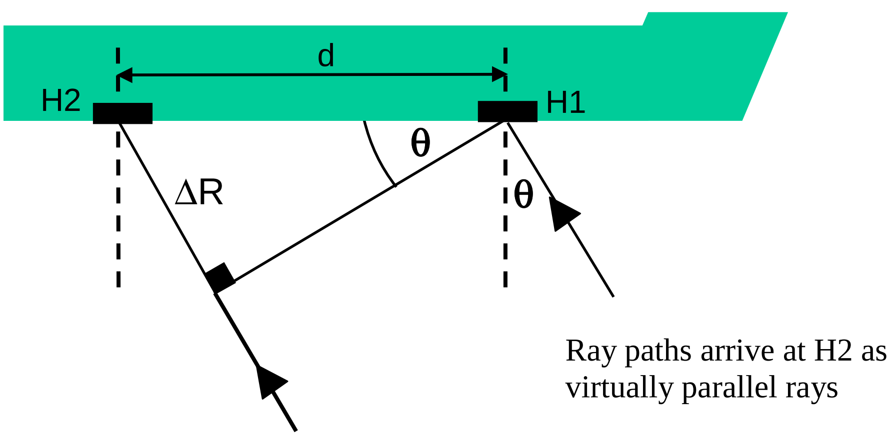


Fig. 2.26 "Angular Measurement"

- Provided Sound Velocity is known: $\Delta R = c\Delta t$ and $\sin\theta = \dfrac{c\Delta t}{d}$

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.

$$\sin\theta = \frac{c\Delta t}{d}$$

$$Y = R\sin\theta \quad = R\frac{c\Delta t}{d}$$

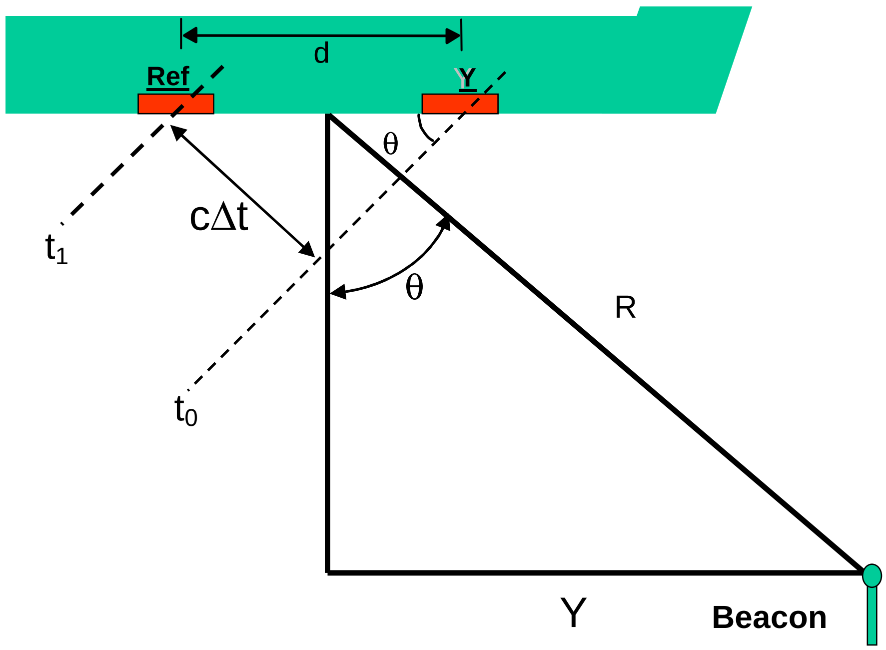


Fig. 2.27

Calculating position in 2 planes

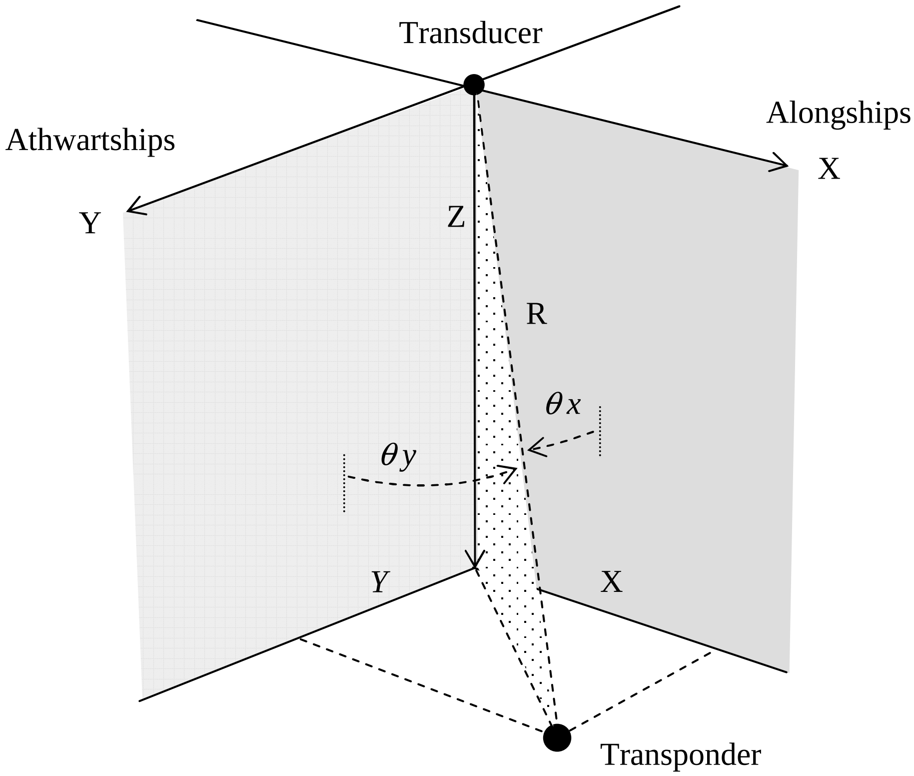


Fig. 2.28

$X = R \sin \theta x$ $Y = R \sin \theta y$ and therefore $z = R(1 - \sin^2 \theta y - \sin^2 \theta x)^{1/2}$

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, 5th Edition – 2008
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, 5th Edition – 2008
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*	U.S. Army Corps of Engineers,

	"Topographic Surveying"	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