THE GEOID
WHAT'S IT GOT TO DO WITH ME?

R.E.Deakin
Department of Land Information
Royal Melbourne Institute of Technology
GPO Box 2476V, MELBOURNE, Vic
AUSTRALIA

(published in *The Australian Surveyor*, Vol. 41, No. 4, Dec. 1996, pp. 294-305)

ABSTRACT

The Global Positioning System (GPS), a US Department of Defense satellite positioning system, has become a widely used measuring tool in practical surveying applications, especially in the provision of coordinate control for construction projects, small scale mapping and cadastral survey coordination. To use GPS effectively, the surveyor must be familiar with the relationships between (i) local and geocentric datums, (ii) geodetic, Australian Map Grid (AMG) and three dimensional Cartesian coordinates and (iii) heights referred to the ellipsoid, the geoid and the Australian Height Datum (AHD).

These relationships, often complex, are usually expressed mathematically. The surveyor, therefore, must also be an applied mathematician - certainly by practice, if not by training. The aim of this paper is to provide an explanation of some of the mathematical theory associated with the reference surfaces and relationships mentioned above with particular attention to the geoid and GPS heighting.

INTRODUCTION

GPS has had (and will continue to have) the most far-reaching effects on the practice of surveying, arguably exceeding the combined effects of Electronic Distance Measurement (EDM) and the pocket calculator. In the space of a decade GPS receivers have reduced in size from two-man luggable boxes (not including heavy duty batteries and antennae) to small combined units mounted on plumbing rods. There have also been proportionate reductions in the cost of equipment and associated computer software.

With GPS, surveyors now have a tool which can serve as well in plane surveying applications (such as positioning of cadastral pegs) as in geodetic operations (such as national or state mapping control). In fact, GPS could be regarded as the universal surveying instrument, but its attraction comes at a cost: it forces upon the wise user a need to understand some of the basic concepts of *geodesy*. This word is associated with an historical division of surveying into two main branches: *geodetic surveying*, usually the domain of government mapping authorities and *plane surveying*, usually that of private surveying concerns.

In Australia, the emergence of GPS has coincided with the reduction of large government and semi-government survey organisations. In some cases they have disappeared altogether! This downsizing (a wonderful euphemism) has allowed private surveying firms the opportunity of involvement in projects formerly outside their province. These expanded business opportunities often require the surveyor to embrace the relatively new GPS technology and so there are economic, as well as professional imperatives for an understanding of the geodetic concepts implicit in the use of GPS.

This paper gives some historical information regarding the discovery of the Earth's "true" shape and the connection between this shape and gravity. A definition of the geoid is provided, together with explanations of Mean Sea Level (MSL) and the AHD, thus providing a connection between these height reference surfaces. The paper also shows how a geocentric ellipsoid can be used as a reference surface for both position and gravity; it describes gravity anomalies and their use in the determination of the geoid-ellipsoid separation – a vital requirement in GPS heighting. An attempt is made to present an overall perspective of some complex geodetic relationships and to be informative rather than definitive. References are provided to give the interested reader an opportunity for further study.

REFERENCE SURFACES

Surveying, in any form, requires the adoption of *reference surfaces* to which points surveyed on the irregular surface of the Earth may be related. These surfaces are used as a basis for the computation of bearings and distances between points, and areas of figures, as well as datums for heights.

In *plane surveying*, the extent of the survey is usually small in comparison with the Earth and the reference surface (as the name implies) is a plane, usually a local horizontal plane, and linear measurements on the Earth's surface are reduced to equivalent measurements on such a plane by the application of simple slope corrections. Angular measurements in the field (using a theodolite) are regarded as measurements between points on this plane. All cadastral surveying in Australia employs these simple principles of plane surveying.

In *geodetic surveying*, surveys are usually of such an extent that the Earth can no longer be considered a plane or even a sphere. Instead, an *ellipsoid* (an ellipse rotated about its minor axis) is adopted as an approximation of the size and shape of the Earth. In geodetic surveys, linear and angular measurements between points on the terrestrial surface are reduced to equivalent distances and angles between complementary points on the ellipsoid. *Geodesy* is the science of determining the size and shape of the Earth and the location of points thereon.

Surveying often involves the determination of heights of points; such heights are related to datums, which in this sense are in fact reference surfaces for heights. Height datums may be as simple as arbitrary datums, where heights of points in small survey areas are related to fixed points (benchmarks) with arbitrarily assigned heights, or as complex as national height datums, such as the AHD, where heights are related to an approximation of MSL around the coastline of Australia.

The determination of heights is closely allied to engineering works concerned with the control and flow of water. Water obeys the laws of physics and flows "downhill" from one *equipotential surface* to another - in fact, a "free" body of water will form its own equipotential surface. This natural occurrence is the reason why an equipotential surface is the most sensible datum for heights. It also provides the reason why most countries adopt some form of MSL as the datum for heights, since all waters of the Earth discharge to the oceans.

The *geoid* is a particular equipotential surface closely approximating the oceanic MSL. As such, the geoid can be regarded as a global reference surface for heights; much work in geodesy is therefore directed at determining the shape and location of the geoid. The geoid is not a tangible, physical surface - you can't dig it up or measure to it directly - and it is thus similar to the ellipsoid in being a purely mathematical conception. But, since it approximates MSL, the geoid does have an easily visualised companion. In Australia, the AHD, which is a practical attempt at estimating MSL around the coastline of Australia, is regarded as a working approximation of the geoid. More about this later - but first, some historical background on the shape of the Earth and the nature of gravity.

NEWTON, GRAVITY AND THE SHAPE OF THE EARTH

Equipotential surfaces, such as the geoid, are gently undulating surfaces related directly to the Earth's gravity field and so an understanding of them must begin with some words on Sir Isaac Newton (1642-1727), his discovery of the nature of gravity and his theory on the shape of the Earth.

The Earth can be regarded as a viscous fluid body, of varying density, rotating about its axis in space and slightly "squashed" at the poles due to the combined effects of gravitational and centrifugal forces. This revolutionary proposition was first put forward by Newton in his *Philosophiae naturalis principia mathematica (The Mathematical Principles of Natural Philosophy)* published in 1687. In the *Principia,* Newton - the Lucasian Professor of mathematics at Trinity College, Cambridge - set down his famous statement:

> "*That the forces by which the primary planets are continually drawn off their rectilinear motions, and retained in their proper orbits, tend to the sun; and are inversely as the squares of the distances of the places of those planets from the sun's centre.*" (Fauvel & Gray 1987, p.397)

We now refer to this as Newton's *Universal Law of Gravitation*, often stated as: masses attract each other with a force inversely proportional to the square of the distance between them.

In the *Principia,* Newton also stated his three laws of uniform motion; gave mathematical proof of Kepler's laws of planetary motion (which Kepler deduced from observation) and described his use of *fluctions* (or infinitesimal changes) in determining tangents to curves - thus inventing calculus! Newton's remarkable work in the fields of gravity and the calculus was done in a two-year period from 1665-66, when he was staying at his birthplace in Lincolnshire to escape from the plague then infesting Cambridge. It remained unpublished for twenty years until his friend, Edmund Halley (1656-1742), enquired of him about the possible shapes of orbits of comets. When he informed Halley that he had solved this problem some years before, (using his theory of gravitation) and that the orbits were elliptical, periodic and thus predictable, Halley insisted that Newton publish his work. With the publication of the *Principia,* Newton's place in history was assured. He remained at Cambridge until 1696 when he accepted the positions of firstly Warden, and later, Master of the Royal Mint. In 1705 he was knighted by Queen Anne and he died in 1727. His passing, as Voltaire suggested, was the occasion for national mourning, and the poet Alexander Pope penned his *Epitaph:*

> "*NATURE, and Nature's Laws lay hid in Night. God said,* Let Newton be! *and all was* Light."
> (Fauvel & Gray 1987, p.415)

Newton's reasoning on the shape of the Earth can be explained as follows. The Earth can be regarded as consisting of an infinite number of small masses. For a body at rest on the surface of the Earth, the *gravitational force* on the body is the resultant of all the forces of attraction between the body and the infinite number of masses making up the Earth. This resultant force acts along a line directed towards the centre of the Earth, but due to the varying mass density of the Earth and local attractions (perhaps caused by mountainous regions) gravitational forces acting on bodies at a number of points on the Earth's surface are not always directed towards the same "earth centre". As the Earth is rotating at a constant angular velocity, a body on its surface is also subjected to a *centrifugal force* – the same force acting on your clothes in the spin dryer. This centrifugal force, proportional to the perpendicular distance of the body from the Earth's spin axis and directed outwards along this line, is a maximum at the equator, decreasing to zero at the poles. Both gravitational and centrifugal forces act on a body at rest on the Earth's surface and the resultant force is known as the *gravity force,* or simply, *gravity.* [Gravity is a vector having both direction (the direction of gravity) and magnitude (the value of gravity)]. At the equator, the gravitational and centrifugal forces act in opposite directions, whilst at the poles, only the gravitational force has an effect. This means that gravity is greater at the poles than at the equator, and since the Earth is a deformable fluid body, it must be, therefore, slightly flattened at the poles.

This practical effect of Newton's law of gravitation was confirmed by the measurements of two survey expeditions organised by the French Academy of Sciences between the years 1735-43. Newton's propositions were opposed by the French, led by the Royal Astronomer, Jean Dominique Cassini and his son Jacques, who had analysed the length of a meridian arc measured near Paris and deduced that the Earth was prolate (flattened at the equator) rather than oblate (flattened at the poles) as predicted by Newton's theory. To settle the dispute between the English and French "camps", a survey expedition was sent to Peru[1] in 1735 to measure an arc of a meridian near the equator by the method of baselines and triangulation. In 1736-37 another expedition, under the leadership of P.L. Maupertuis (1698-1759), went to Lapland (northern Sweden) to measure a meridian arc near the pole. [The expeditions also included the scientists A.C. Clairaut (1713-1765) and P. Bouguer (1698-1758), both still honoured in geodesy for their theorems on gravity]. The triangulation measurements, combined with careful astronomic observations of latitude, confirmed that the distance on the surface of the Earth subtended by a degree of latitude was indeed longer near the pole than at the equator. This was a triumph for Newton's theory, as well as for Maupertuis himself, who henceforth was known as the *grand aplatisseuer* ("great flattener"), becoming president of the Berlin Academy of Science and basking for many years in the sun of his fame at the court of Frederick the Great. Alas, Maupertuis' glory was short lived; in 1750 he set down a general principle unifying the laws of the universe and combined it with a proof of the existence of God! His theory caused great controversy which reached a climax when Voltaire lampooned the unhappy president in the *Diatribe du docteur Akakia, médecin du pape* (1752). Neither the King's support nor Euler's defence could bring succour to Maupertuis' sunken spirits, and the deflated mathematician died not long afterwards in Basel in the home of the Bernoullis (Struik 1987, p.127).

THE GRAVITY FIELD AND EQUIPOTENTIAL SURFACES

The Earth's gravity field is a *vector field,* meaning that there is a triplet of numbers assigned to every point. These numbers represent the x,y,z Cartesian components of the gravity vector at the point. In geodesy, however, it is much more convenient to work with a *scalar field* where a single-valued function is assigned to every point. This scalar field is known as the *gravity potential* W and the derivatives of the scalar function $W(x,y,z)$ are the components of the gravity vector at that point. In the same way as gravity is the vector sum of the Earth's gravitational force of attraction and the centrifugal force, the Earth's gravity potential W is the scalar (or algebraic) sum of its *gravitational potential* W_g and its *centrifugal potential* W_c:

$$W = W_g + W_c \qquad (1)$$

[1] This was the Spanish Viceroyalty of Peru, much larger than the present Peru. Headquarters of the expedition were at Quito, now in Ecuador.

(The gravitational potential and the centrifugal potential are often called respectively the potential due to mass attraction and the rotational potential.)

An *equipotential surface* of the Earth is a surface upon which the gravity potential is constant, or

$$W(x,y,z) = constant. \quad (2)$$

Movement along such a surface involves no change in potential and thus no "work" (in the conventional static sense) is done. Therefore, this movement cannot go with, or against, the direction of the force field (the gravity field) and the consequence of this is that *the direction of gravity must be everywhere perpendicular to an equipotential surface.*

Equipotential surfaces are often called *level surfaces* and the perpendicular is known as the *vertical* or the *plumb line.* The local *tangent* plane to an equipotential surface at a point is a *horizontal* plane - as defined by a carefully levelled theodolite at that point – and the vertical plane of the same carefully levelled theodolite is tangential to the vertical (or plumb line) at that point. This means that angular measurements made with a theodolite are directly related to the gravity vector at the point of observation, as are spirit level observations of height differences. Heights so related to particular equipotential surfaces (such as the geoid) are known as *orthometric heights.*

Equipotential surfaces cover the Earth like the layers of an onion; they do not cross each other nor are they parallel to each other, except as a first approximation. They are continuous (ie, they do not have any breaks), they have no sharp edges and are convex everywhere with smoothly varying radii of curvature. A sectional view of equipotential surfaces surrounding the Earth would show them as oblate curves spaced closer together at the poles than at the equator with verticals as curved lines intersecting each surface at right angles. This is a consequence of gravity being stronger at the poles and corresponds to the definite relationships between equipotential surfaces and gravity, viz. (i) the direction of gravity and the equipotential surface are mutually perpendicular, and (ii) the spacing of the surfaces is directly related to the magnitude of gravity. It should be noted that equipotential surfaces are curved in every direction and with respect to each other, hence the verticals (or plumb lines) to these surfaces are space curves having both curvature (bend) and torsion (twist).

THE GEOID AND MEAN SEA LEVEL

Carl Friedrich Gauss[2] (1777-1855) was the first to propose that an equipotential surface corresponding to mean sea level be considered as the mathematical surface of the Earth. The term *geoid* was first used in 1873 by J.B. Listing (1808-82) who defined it as being *"the equipotential surface of the Earth's gravity field which would coincide with the ocean surface if the latter were undisturbed and affected only by the Earth's gravity field"* (NGS 1986). This definition is now regarded as deficient since it assumes that the oceanic surface specified is an equipotential surface - more about this below. A commonly accepted "modern" definition of the geoid is *"the equipotential surface of the Earth's gravity field which best fits, in the least squares sense, mean sea level"* (NGS 1986).

MSL is an empirical determination based on long term measurements of tidal heights recorded by tide gauges. A knowledge of the reasons why the sea level rises and falls, and of other factors affecting MSL, is important in understanding the differences between it and the geoid.

Tidal forces acting on the Earth are due to the gravitational attractions of the Moon, the Sun and the planets (in order of decreasing effect) combined with centrifugal forces. For example, the mass of the Moon exerts an attractive force on the Earth, and centrifugal forces are created by the Earth-Moon couple rotating about a common centre of gravity where the rotation rate is approximately $2\pi/28$ radians/day and the centre of gravity of the couple is 4720 km from the Earth's centre. Tidal forces (which are in addition to gravity as defined earlier) cause the equipotential surfaces of the Earth to be deformed into prolate shapes bulging in the directions of the Earth-Moon and Earth-Sun axes. Water will constantly adjust its level to coincide with an equipotential surface responding to a tidal force; this is easily seen in the case of the rise and fall of sea level (*ocean tides*), but tidal forces also cause very small movements of land masses (*solid Earth tides)* as well as periodic changes in the "depth" of the Earth's atmosphere. In this paper, unless indicated otherwise, "tides" refer to ocean tides, and those other (and much smaller) tides will be ignored.

---

[2] Gauss is regarded as the prince of mathematicians and according to Melluish (1931, p.40) ".. seems to have left no branch of mathematics unadorned by his researches". In his *Theoria motus corporum coelestium* (1809), Gauss used his theory of *least squares* to calculate the orbit of the minor planet Ceres from a small number of observations and predicted its future position in the heavens. As the director of the astonomical observatory at Göttingen from 1807 until his death in 1855, Gauss was actively interested in geodesy and published seminal works on potential theory and conformal mapping (Struik 1987, p.143).

A number of tidal cycles must be taken into account in any determination of MSL. Doodson and Warburg (1941, p.39) list the following "short" cycles:

(a) *lunar/solar semidiurnal* and *diurnal* cycles of 12 and 24 hours due to the Earth's rotation about its axis;
(b) *lunar monthly* cycles varying between 27.2 and 29.5 days due to the Moon's rotation about the Earth;
(c) *solar annual* cycles due to the Earth's rotation about the Sun;

and three "long" cycles of:

(i) 18.03 years, known as the Saros, in which solar and lunar eclipses repeat themselves;
(ii) 18.61 years, the period of revolution of the Moon's nodes[3]; and
(iii) 19.00 years, the Metonic[4] period of lunar phases.

These *luni-solar* tidal cycles (the planets being regarded as having a negligible effect) give rise to different high and low water means, depending on the period of observation. Any "accurate" MSL should be derived from at least twelve months of tide gauge readings and, if possible, a nineteen year period of observation.

Tides are not the only factors affecting the determination of MSL. Bomford (1980, p.249) lists others as:

- the location of tide gauges (estuarine and bay versus ocean);
- the prevailing wind (offshore or onshore);
- uniform changes in sea level due to an increase or decrease in polar ice (eustatic changes);
- changes in the barometric pressure and density of the Earth's atmosphere; and
- changes in temperature, salinity and currents.

The question as to whether MSL is an equipotential surface is often posed, and Bomford (1980, p.250) notes that if the free surface of a uniform liquid at rest closely approximates an equipotential surface then, MSL fails to meet these conditions in several respects:

(i) its surface is overlain by air, whose pressure varies – it is, therefore, not quite a "free" surface;
(ii) the wind applies a horizontal force to the surface;
(iii) the density of the water varies, principally with its temperature and salinity;
(iv) the sources of water – rain, rivers, and melting ice – do not coincide with the areas where water is lost by evaporation;
(v) these inequalities cause ocean currents which act towards the restoration of equilibrium, but with a time lag.

The result is that MSL departs from an equipotential surface by amounts which are more or less constant with time. It is therefore, at best, only an approximation of the geoid (Bomford, loc. cit.).

As a final comment, it is worthwhile noting that MSL for a region is determined from tide gauges located around relatively shallow coastlines and does not include tidal measurements of the open ocean. This lack of open ocean tidal information (Done 1984, sec.3.1, p.29; Doodson & Warburg 1941, sec.12.2, p.100) adds a further reason why no regional MSL can be a truely coincidental part of the "global" geoid.

THE AUSTRALIAN HEIGHT DATUM

Prior to 1972, there was a multitude of levelling datums in operation across Australia, each adopted by government and regional instrumentalities to suit their own particular purposes. For example, in Melbourne, there was the Board of Works datum, the Tramways datum, the Harbour Trust datum and the Dandenong Valley Authority datum (to name just a few), a similar plurality being evident in other states of the Commonwealth. To unify the various datums, the Division of National Mapping (now AUSLIG - Australian Surveying and Land Information Group) undertook a massive campaign of levelling across Australia. This program, which commenced in 1945, culminated in 1971 with the simultaneous least squares adjustment of 97,320 km of primary levelling connected to thirty tide gauges around the Australian coastline. MSL for 1966-68 was assigned a value of zero at the tide gauges and the *Australian Height Datum (AHD)* is that surface which passes through MSL at the tide gauges and through points with zero AHD height vertically below the other basic junction points in the levelling network (Roelse et al. 1971, pp.1 & 48; NMC 1986, p.60).

---

[3] The nodes are the intersections on the celestial sphere of the Moon's orbital plane and the *ecliptic* - the plane of the Earth's orbit around the Sun.

[4] The cycle discovered by the Athenian astronomer Meton in which the Moon returns (nearly) to the same apparent position with respect to the Sun, so that new and full moons occur at the same dates in the corresponding year of each cycle. (SOED, 1993)

ELLIPSOIDS AND SPHEROIDS

As Newton predicted, and as the French proved by measurement, the Earth has a slightly flattened spherical shape, with a polar diameter less than an equatorial diameter in the ratio of 299:300 (approximately). Such shapes are known as *spheroids;* the SOED (1993) defines a spheroid as: "*A body resembling or approximating to a sphere in shape;* esp. *one formed by the revolution of an ellipse about one of its axes.*" This special case of a spheroid is known as an *ellipsoid of revolution* and it may be *prolate* (rotated about its major axis) or *oblate* (rotated about its minor axis). In geodesy, an oblate ellipsoid of revolution – simply known as the *ellipsoid* – is used as the mathematical approximation of the Earth's shape and its minor axis is parallel to (or coincident with) the Earth's axis of revolution. In Australia, and indeed in most geodesy texts, the terms ellipsoid and spheroid are taken to be synonymous.

The size and shape of an ellipsoid can be defined by one of three geometric relationships:

(i) specifying a and b, the lengths of the major and minor semi-axes respectively; or

(ii) specifying a and f, where f is the *flattening* where $f = \frac{a-b}{a}$; or

(iii) specifying a and e^2, where e is the *eccentricity* where $e^2 = \frac{a^2-b^2}{a^2}$.

Figure 1 shows the relationships between Cartesian coordinates x,y,z and *geodetic* coordinates ϕ, λ, h of a point P related to an ellipsoid whose semi-major axis is $OE = a$ and semi-minor axis is $ON = b$. The geodetic coordinates are defined as follows:

(i) ellipsoidal height h is the distance from the ellipsoid to the point P, measured along the *normal* to the ellipsoid passing through P and intersecting the minor axis at H;

(ii) geodetic longitude λ is the angle, measured in the plane of the equator, between the meridian plane of Greenwich, and the meridian plane containing P; and

(iii) Geodetic latitude ϕ is the angle, measured in the meridian plane, between the ellipsoidal equator and the normal passing through P.

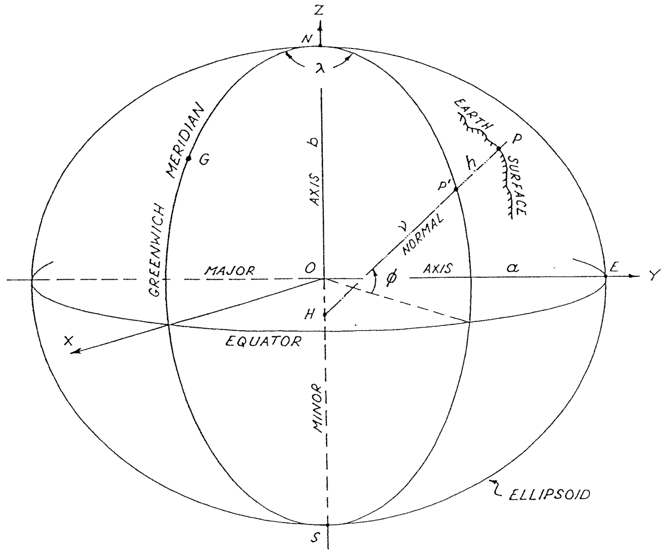

Figure 1

The Cartesian coordinates (x,y,z) of $P(\phi,\lambda,h)$ on an ellipsoid of semi-major axis a and flattening f may be calculated by the following formulae:

$$x = (\nu + h)\cos\phi \cos\lambda \qquad (3.1)$$

$$y = (\nu + h)\cos\phi \sin\lambda \qquad (3.2)$$

$$z = (\nu(1-e^2)+h)\sin\phi \qquad (3.3)$$

where

$$\nu = HP' = \frac{a}{\sqrt{1-e^2 \sin^2\phi}}$$

ν (nu) is the radius of curvature of the ellipsoid in the prime vertical plane (see below)

$$e^2 = f(2-f)$$

$$b = a(1-f)$$

$$OH = \nu e^2 \sin\phi$$

The inverse computation of (ϕ,λ,h) given (x,y,z) can be made using the following:

$$\cos\lambda = \frac{x}{r} \qquad (4.1)$$

$$\tan\phi = \frac{z + \nu e^2 \sin\phi}{r} \qquad (4.2)$$

$$h = \frac{r}{\cos\phi} - \nu \qquad (4.3)$$

where

$$r = \sqrt{x^2 + y^2}$$

Note: In equation (4.2), functions of the latitude appear on both sides of the equation, thus requiring an iterative solution for ϕ. A first approximation for ϕ may be obtained from $r \tan\phi = z$.

The ellipsoid is a surface of varying curvature and planes intersecting the ellipsoid create *elliptical* curves of intersection having changing *radii of curvature,* rather than the constant radii characteristic of circular curves. Intersecting planes containing the normal to the ellipsoid create special curves of intersection known as *normal section curves,* and at any point on an ellipsoid there are an infinite number of possible normal section curves of intersection. Two normal section curves are of interest in geodesy: (i) the meridian curve, which has the least radius of curvature ρ (rho), and (ii) the prime vertical curve, which has the greatest radius of curvature ν (nu). Every curved surface has these *principal curvatures*[5] in mutually perpendicular directions known as the *principal directions* (Lauf 1983, p.19). (Two special cases are a *sphere* which has principal curvatures which are constant and equal, and a *cylinder* which has one principal curvature finite and constant and the other infinite.)

For a latitude ϕ on an ellipsoid a, e^2 the equations for the principal radii of curvature are

$$\rho = \frac{a(1-e^2)}{\sqrt{(1-e^2 \sin^2\phi)^3}} \tag{5.1}$$

$$\nu = \frac{a}{\sqrt{(1-e^2 \sin^2\phi)}} \tag{5.2}$$

The radius of curvature of a normal section curve having an azimuth α (alpha) is

$$R_\alpha = \frac{\rho\nu}{(\rho \sin^2\alpha + \nu \cos^2\alpha)} \tag{5.3}$$

and a mean radius of curvature is

$$R_m = \sqrt{\rho\nu} \tag{5.4}$$

These equations for the radii of curvature ρ, ν, R_α and R_m coupled with the geometric relationships linking a, b, f and e^2 are the fundamental formulae required for any calculations of distances and directions between points on an ellipsoid (NMC 1986, p.16).

[5] Gauss, in 1803, published his theorems on curvatures of surfaces in *Disquisitiones generales circa superficies curvas,* roughly translated as "A general discourse on curved surfaces", in which he provided the foundation for the branch of mathematics now known as the differential geometry of curved surfaces (Struik 1987, p.143).

THE GEOCENTRIC EQUIPOTENTIAL ELLIPSOID

As noted above, among the mathematicians who went with Maupertuis to Lapland to measure the meridian arc was Alexis Claude Clairaut (1713-65). Clairaut had already published a mathematical paper on the theory of space curves; on his return from Lapland he published his *Théorie de la figure de la terre*[6] in 1743. This became the standard work on the equilibrium of fluids and the attraction of ellipsoids of revolution (Struik 1987, p.128). It contained one of the most striking formulas of *physical geodesy:–*

$$f + f^* = \frac{5}{2}m \tag{6}$$

where

$$f = \frac{a-b}{a}, \quad f^* = \frac{\gamma_P - \gamma_E}{\gamma_E} \quad \text{and}$$

$$m \approx \frac{\omega^2 a}{\gamma_E} = \frac{\textit{centrifugal force at equator}}{\textit{gravity at equator}}$$

f is the geometric flattening of the ellipsoid, f^* is an analogous quantity known as the *gravity flattening,* γ_E and γ_P are the values of gravity at the equator and pole respectively and ω is the Earth's angular velocity.

Clairaut's equation, although containing an approximation for the constant m, showed that the geometric flattening f can be derived from f^* and m, which are purely dynamical quantities obtained by gravity measurements; in other words, *the flattening of the Earth can be obtained from gravity measurements* (Heiskanen & Moritz 1967, p.75).

Clairaut's work, with subsequent refinements by mathematicians and geodesists of the 18th, 19th and 20th centuries, forms the basis of the *theory of the geocentric equipotential ellipsoid.* This theory can be used to show that an homogeneous ellipsoid, concentric with and having the same mass as the Earth, and rotating with the same angular velocity about the Earth's axis, will generate a theoretical gravity field known as the *normal gravity field.*

[6] Theory of the figure of the Earth.

The surface of this ellipsoid (also called the *normal ellipsoid*) is an equipotential surface of the normal gravity field and the magnitude and direction of *normal gravity* can be computed on or above the normal ellipsoid. Hence, a geocentric equipotential ellipsoid can be used as a reference surface for both position and gravity.

Since the early 1900's, the parameters of many "best fit" ellipsoids have been determined from observations of the various effects of the Earth's gravity field. These observations have included:

(i) direct measurements of gravity and the determination of *gravity anomalies;*

(ii) astro-geodetic differences between observed astronomic and computed geodetic values of latitude, longitude and azimuth at selected stations in trigonometric networks (astronomic values are related to equipotential surfaces of the Earth's gravity field); and

(iii) measurements to satellites whose orbital characteristics are directly related to the Earth's gravitational attraction.

In 1924, the International Union of Geodesy and Geophysics (IUGG) adopted an ellipsoid as the best representation of the size and shape of the Earth. This ellipsoid, known as the International Ellipsoid of 1924 was considered at the time, to be the best fit of the geoid, on a global basis. The geometric parameters of the ellipsoid were determined by the American geodesist Hayford in 1909 from astro-geodetic data in the United States.

Since that time there have been several revisions of parameters and the currently accepted ellipsoid is defined by the Geodetic Reference System 1980 (GRS80), adopted at the XVII General Assembly of the IUGG in Canberra, December 1979 (BG, 1988). The GRS80 is based on the theory of a geocentric equipotential ellipsoid defined by the following four parameters:

- equatorial radius of the Earth
  $a = 6378137\,m$
- geocentric gravitational constant of the Earth (including the atmosphere)
  $GM = 3986005 \times 10^{8}\,m^{3}s^{-2}$
- dynamical form factor of the Earth (excluding the permanent tidal deformation)
  $J_2 = 108263 \times 10^{-8}$
- angular velocity of the Earth
  $\omega = 7292115 \times 10^{-11}\,rad\,s^{-1}$

The gravitational constant GM is the product of the Newtonian gravitational constant G and the total mass of the Earth M. The dynamical form factor J_2 is a function of the Earth's moments of inertia (polar and equatorial), the radius a and the mass M. It is also linked to the dynamical constant m in Clairaut's equation (6).

All geometric constants of the GRS80 ellipsoid (the *normal ellipsoid*) can be computed from the four defining parameters. In addition, physical constants, such as the gravitational potential of the ellipsoid (*normal potential*) and the magnitude of gravity on the ellipsoid (*normal gravity*) can also be computed. Hence the GRS80 ellipsoid is a convenient reference surface for gravity as well as position. Formulae, definitions, parameters, constants and other information related to the GRS80 are given in *The Geodesist's Handbook 1988* (BG, 1988).

The development of equipotential geocentric ellipsoids is a continuing process; densification of gravity data (especially over the oceans) and refinements of the values of GM and J_2 from satellite orbital data will allow improvement of the ellipsoidal model so that GRS80 be superseded in its turn.

GRAVITY ANOMALIES AND THE GEOID-ELLIPSOID SEPARATION *N*

Absolute values of gravity can be determined by either *pendulum* or *free-fall devices* which require precise timing of the observed motions of bodies which are then converted into gravity values by using appropriate equations of motion. The instrumentation required is large and cumbersome and is only used to establish reference stations at a limited number of locations around the world. The establishment of these *global gravity networks* has been coordinated by the IUGG and the current network is known as the *International Gravity Standardisation Network 1971* [IGSN 1971] (Vanicek & Krakiwsky 1986, p.535). Densification of the gravity network is achieved by using portable devices, called gravimeters, which measure *gravity differences* between points. New points in the network are connected to base stations via measurement loops and the gravity differences are corrected for instrumental errors, any "misclosures" in the gravity loops being removed by adjustment. This technique should be familiar to surveyors, since it is based on the same practical principles used to densify levelling networks such as the AHD. Australia is covered by a network of gravity stations and gravity measurements are continually being made for various geophysical research purposes - particularly mining, where *gravity anomalies* may indicate the locations of ore bodies beneath the Earth's surface. The word *anomaly* is generally taken to mean any variation from a "normal state"; in geodesy (and geophysics), a gravity anomaly Δg is defined as the difference between the "observed" value of gravity g and the value of normal gravity γ (gamma):

$$\Delta g = g - \gamma . \qquad (7)$$

Since gravity is usually measured at points on the Earth's surface, its value will be "affected" by elevation (since gravity decreases as elevation increases), and also by the surrounding topography. The topographic effect on gravity measurements (and theodolite measurements, which are related to equipotential surfaces of the Earth's gravity field) was discovered by Bouguer during the expedition to Peru in 1735, when he determined that the mass of the Andes was affecting theodolite observations. Corrections for elevation and/or topography will yield several different gravity anomalies, each one suitable for specific purposes. In geodesy, when gravity observations are used to determine the size and shape of the *geoid*, it is usual for the observations to be corrected only for height above sea level. The corrected observations, less the values of normal gravity (see equation 7) give *free-air gravity anomalies*, which are regarded as the differences between gravity "observed" on the geoid and normal gravity. The term *free-air* signifies that the correction is determined by assuming that the point of observation is suspended in "free-air" above the geoid.

[It is interesting to note that the application of free-air corrections has the effect of "moving" the mass of the Earth (between the observation points and the geoid) inside the geoid, thus affecting the total mass distribution of the Earth, which in turn causes a slight change in the shape of the geoid to a (practically) parallel surface known as a *co-geoid*. This is a "Catch-22" situation, since gravity anomalies are used to determine the shape of the geoid. In this paper, the theoretical difference between geoid and co-geoid will be ignored.]

Free-air gravity anomalies are used to determine the *geoid-ellipsoid separation N* (see the following section for an explanation), often computed at regular grid intervals over regions of interest and known as a *geoid model*. N is the link between *ellipsoidal height h* (measured along the normal) and *orthometric height H* (measured along the vertical from the geoid to P). This relationship is shown in Figure 2 and given by the equation

$$h = H + N \qquad (8.1)$$

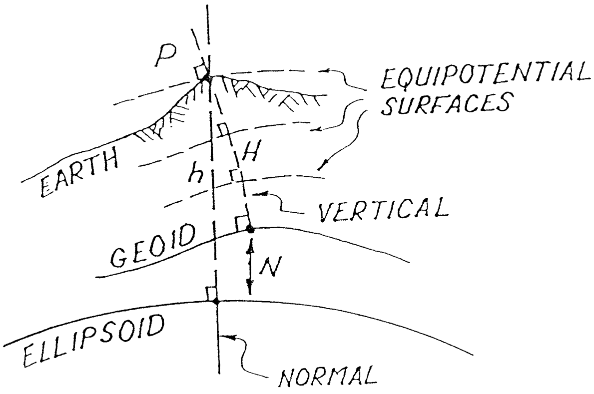

Figure 2
(The vertical is a curved line intersecting all equipotential surfaces at right angles)

The diagram is hugely exaggerated and, in reality, N varies quite slowly, for example, approximately 60m from Melbourne to Cairns (Kearsley & Govind 1991, Fig.2, p.33), which is roughly equivalent to 3 cm per km.

In practice, surveyors obtain ellipsoidal height differences between points Δh from an adjustment of their GPS measurements and ΔN values from geoid models such as AUSGEOID93[7] (Steed & Holtznagel 1994, p.22). These can be converted to orthometric height differences ΔH by

$$\Delta H = \Delta h - \Delta N . \qquad (8.2)$$

Therefore, if one point in the survey has a known AHD level value, other AHD values can be obtained by adding and subtracting appropriate values of ΔH. This presupposes that the surface of the AHD is approximately parallel to the geoid in the area of interest. Steed and Holtznagel (1994, pp.22-27) have compared AHD levels obtained by GPS and traditional spirit levelling techniques and concluded that GPS heighting is capable of producing results within third-order spirit levelling limits, provided that suitable N values are used.

GEOPOTENTIAL MODELS AND GEOID MODELS

Analogous to the Earth's gravity potential W, which is the sum of its gravitational and rotational potentials, (see equation 1) the equipotential

[7] A grid of computed N values over Australia, produced by AUSLIG, which can be interpolated to give values at desired locations.

potentials, (see equation 1) the equipotential ellipsoid (GRS80) has a *normal* gravity potential U, the sum of its own gravitational and rotational potentials:

$$U = U_g + U_c. \qquad (9)$$

Subtracting U from W (noting that the two rotational potentials are the same) gives the *disturbing potential* T:

$$T = W_g - U_g. \qquad (10)$$

The gravitational potential W_g can be computed from a *spherical harmonic* expansion (Torge 1980, pp.28-31):

$$W_g = \frac{GM}{r}\left[1+\sum_{n=2}^{\infty}\left(\frac{a}{r}\right)^n \sum_{m=0}^{n}\left\{P_n^m(t)\left(C_n^m \cos m\lambda + S_n^m \sin m\lambda\right)\right\}\right] \qquad (11.1)$$

where

- r is the radial distance from the geocentre;
- $t = \sin\psi$ and ψ is the geocentric latitude;
- λ is longitude;
- n and m are integers known as the degree and order respectively;
- C_n^m and S_n^m are *geopotential coefficients* (of nth degree and mth order); and
- $P_n^m(t)$ are *associated Legendre functions.*

Equation (11.1) is known as a *geopotential model* and coefficients C_n^m and S_n^m up to degree and order 360 (approximately 130,000 terms) have been determined from an empirical analysis of gravity and satellite altimetry data. The current data set, compiled by Dr Richard H. Rapp of the Ohio State University, is designated OSU91A. An improved set of coefficients, using the latest data, is expected to be available in 1997; four preliminary models are currently being tested by an International Association of Geodesy (IAG) working group led by Dr Michael Sideris of the University of Calgary.

Normal gravitational potential U_g can be computed from another spherical harmonic expansion (Heiskanen & Moritz 1976, p.73):

$$U_g = \frac{GM}{r}\left[1-\sum_{n=2}^{\infty}\left\{\left(\frac{a}{r}\right)^n J_n P_n(t)\right\}\right] \qquad (11.2)$$

where n is even, and
J_n are *normal potential coefficients* derived from the dynamical form factor J_2.

Whilst equations (11.1) and (11.2) look formidable, they are simply gigantic summations grist to the computer's mill – and T (equation 10) can be computed with relative ease. Combining T with γ (normal gravity) in *Bruns' formula*[8] gives the geoid-spheroid separation N:

$$N = \frac{T}{\gamma}. \qquad (12)$$

Gravity anomalies Δg are linked to the disturbing potential by the differential equation (Heiskanen & Moritz 1967, p.89):

$$\Delta g = -\frac{\partial T}{\partial r} - \frac{2}{r}T. \qquad (13)$$

This equation was solved by *Stokes*[9] who gave:

$$T = \frac{R}{4\pi}\iint_\sigma \Delta g S(\psi)\, d\sigma \qquad (14.1)$$

and from *Bruns' formula* ($N = T/G$, where G is a mean value of gravity over the Earth):

$$N = \frac{R}{4\pi G}\iint_\sigma \Delta g S(\psi)\, d\sigma \qquad (14.2)$$

where

- R is the mean radius of the spherical Earth;
- G is the mean value of gravity over the Earth;
- $S(\psi)$ is known as *Stokes' function* and ψ is the angular distance between points;
- $d\sigma$ is an element of solid angle; and
- $\iint_\sigma$ is an integration over the full solid angle.

Equations (14.1) and (14.2) both known as *Stokes' integral,* (Heiskanen & Moritz 1967, pp.84-94) show how the disturbing potential T and the geoid-ellipsoid separation N are related to gravity anomalies Δg. The equations may be interpreted in the following manner:

---

[8] This formula, together with other famous equations of physical geodesy, was given by Ernst Heinrich Bruns (1848-1919) in *Die Figur der Erde,* Berlin, 1878.

[9] Sir George Gabriel Stokes (1819-1903), Lucasian professor of mathematics, Cambridge and the foremost British authority of his time on the principles of geodesy. In his study of 1849 he generalized the theory relating the Earth's shape to the strength of gravity (obtaining Clairaut's equation as a particular result) and published his famous equation linking the geoid-ellipsoid separation to gravity anomalies.

number of points Q_i covering the Earth's surface. Gravity anomalies are known at all points Q_i, as are the spherical distances ψ_i from P. The sum of the product $\Delta g\, S(\psi)$ for all Q_i, multiplied by $R/4\pi$ gives the disturbing potential T; division by G gives the geoid-ellipsoid separation N.

In theory, solutions for N using *Stokes' integral* require gravity data covering the entire Earth; this of course is not available (nor is it ever likely to be) and in practice only gravity anomalies over a relatively small region of the Earth are used. *Stokes' integral* solutions may be improved by using "reduced" gravity anomalies; obtained by subtracting a theoretical anomaly, derived from a geopotential model[10], from the "observed" anomaly. These reduced anomalies are then used to compute a small value N' that can be added to N computed from the geopotential model. This technique, which should give more "accurate" values for N, was used to produce the *geoid model* AUSGEOID93 (Kearsley & Govind 1991; Steed & Holtznagel 1994), a grid of computed N values covering Australia.

SUMMARY

Surveyors are constantly making measurements related to the Earth's gravity field: they have been for centuries, theodolite and spirit levelling observations being the most obvious. Until very recently however, the inter-relation between surveying measurements and gravity (or equipotential surfaces) has largely been ignored by the profession, mainly because the bulk of surveying work has been "plane" surveying over relatively small areas using traditional methods, such that the connection can be conveniently ignored. The small minority concerned with geodetic surveying, in which the connection cannot be ignored, was just that – a very small minority – but GPS is changing all that. Nowadays, the surveyor must be familiar with the idea of the ellipsoid and geoid, the differences between them and the relationships between these surfaces and the AHD. Even if surveyors aren't using GPS in their daily operations, the community, who rightly regard surveyors as experts in measurement, does expect them to have a sound theoretical as well as technical knowledge of GPS and how it can be used to their advantage (as clients). Understanding relevant new technology is a necessary part of the professional activity of a modern surveyor.

The connections between gravity and the geoid, and gravity and the equipotential ellipsoid (GRS80) are important. Both reference surfaces, the geoid for height, the ellipsoid for position and gravity, are equipotential surfaces: the former, an equipotential surface of the Earth, the latter, an equipotential surface of the normal (model) earth. Both are determined from analyses of the Earth's gravity field. This paper has provided an explanation of gravity, and how the famous Newtonian Universal Law of Gravitation led to a better understanding of the Earth's shape, together with Clairaut's discovery of the link between the Earth's geometric flattening f and its gravity flattening f^*. These explanations have been given some historical "treatment" to show the clear connection between surveying (geodesy) and the great mathematicians and scientists of the 17th and 18th centuries, although only a few have been mentioned in the text. A knowledge of the rich history of surveying, however small, is useful in understanding these complex topics.

The inter-relationship between the geoid, Mean Sea Level (MSL) and the Australian Height Datum (AHD) is often confusing. There is no *physical* connection between the three surfaces – although it could be argued that the AHD is connected to the particular MSL of 1966-68 defined by thirty tide gauges around Australia – but the following practical links exist:

(i) the geoid is the particular equipotential surface of the Earth's gravity field which best fits, in the least squares sense, MSL;

(ii) MSL is an empirical determination based on long term measurements of tidal heights; the "accuracy" of its estimation is dependent on the period of observation and the location of tide gauges. It is not an equipotential surface – at best, MSL is only an approximation of the geoid; and

(iii) the AHD is a "practical" surface approximating MSL around the coastline of Australia.

Hence, for many applications, the AHD is regarded as a reasonable determination of the geoid over Australia, which explains the often-seen words "... AHD heights are orthometric heights ..." This of course is not correct – the statement should include the word "approximate".

Estimation of orthometric heights from GPS measurements requires a knowledge of the value of the geoid-ellipsoid separation N. This value cannot be determined exactly, but instead, must be

[10] It is possible to "create" a set of spherical harmonic coefficients of T by subtracting normal potential coefficients from corresponding geopotential coefficients (see equations 11.1 & 11.2) This spherical harmonic model of T can then be numerically differentiated with respect to r and equation (13) used to compute a gravity anomaly related to the geoid defined by the geopotential model.

interpolated from pre-prepared geoid models, or computed from geopotential models. In either case, surveyors will be using models based largely on gravity anomalies. This paper has provided information on the nature of gravity anomalies and has attempted to shed some light on geopotential models and integration techniques used to determine *N*. These complex numerical procedures are really applied mathematics and any explanation of the processes (apart from the superficial) will necessarily involve the quotation of formulae. In this paper, they are provided as an adjunct to the explanation and are not intended to cloud the issues in mathematical mystique!

Finally, two questions and answers:

1. Is an orthometric height *H* (GPS ellipsoidal height *h* less the geoid-ellipsoid separation *N*) a reasonable estimation of AHD height?

   This can only be answered by comparative field checks. GPS heighting surveys should, where possible, include "old" points of known AHD height so that some degree of confidence can be assigned to the estimation of AHD height at "new" points. This is good survey practice. Recent studies (Steed & Holtznagel 1994) indicate that GPS heighting is capable of producing results within third-order spirit levelling limits provided that suitable *N* values are used.

2. What has the geoid got to do with me?

   The geoid is the sensible reference surface for heights derived from GPS measurements and geopotential/geoid models. In the future, when GPS heighting has become as commonplace as spirit levelling, surveyors will have quite a lot to do with the geoid.

ACKNOWLEDGMENTS

I would like to thank Peter Done (*Lt-Cdr* RN Retd) for his thoughtful comments and proof-reading.

REFERENCES

BG, 1988, The Geodesist's Handbook 1988, *Bulletin Géodésique,* Vol.62, No.3.

Bomford, G. 1980, *Geodesy,* 4th edn, Clarendon Press, Oxford.

Done, P. 1984, The Geoid and the Vertical Datum Problem, *The Hydrographic Journal,* No.32, April 1984, pp.28-29.

Doodson, A.T. & Warburg, H.D. 1941, *Admiralty Manual of Tides,* HMSO, London.

Fauvel, J. & Gray, J. 1987, *The History of Mathematics: A Reader,* Macmillan, London.

Heiskanen, W.A. & Moritz, H. 1967, *Physical Geodesy,* W.H.Freeman and Co., London.

Kearsley, A.H.W. & Govind, R. 1991, Geoid Evaluation in Australia, A Status Report, *The Australian Surveyor,* Vol.36, No.1, pp.30-40.

Lauf, G.B. 1983, *Geodesy and Map Projections,* TAFE Publications, Collingwood, Australia.

Melluish, R.K. 1931, *An Introduction to the Mathematics of Map Projections,* University Press, Cambridge.

NGS 1986, *Geodetic Glossary,* National Geodetic Survey, Rockville, MD, USA, Sept. 1986.

NMC 1986, *The Australian Geodetic Datum Technical Manual,* Special Publication 10, National Mapping Council, Canberra.

Roelse, A., Granger, H.W. & Graham, J.W. 1971, *The Adjustment of the Australian Levelling Survey 1970-1971,* Technical Report 12, Division of National Mapping, Canberra, June 1971.

SOED 1993, *The New Shorter Oxford English Dictionary,* Clarendon Press, Oxford.

Steed, J. & Holtznagel, S. 1994, AHD Heights from GPS using AUSGEOID93, *The Australian Surveyor,* Vol.39, No.1, pp.21-27.

Struik, D.J. 1987, *A Concise History of Mathematics,* 4th edn, Dover, New York.

Torge, W. 1980, *Geodesy,* Walter de Gruyter, New York.

Vanicek, P. & Krakiwsky, E.J. 1986, *Geodesy: The Concepts,* 2nd edn, North-Holland, Amsterdam.