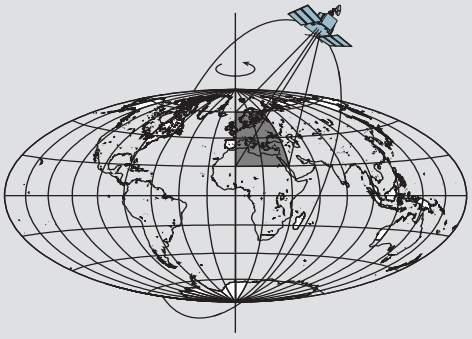
Heights, the Geopotential, and Vertical Datums

by

Christopher Jekeli



Report No. 459

Geodetic Science and Surveying
Department of Civil and Environmental Engineering and Geodetic Science
The Ohio State University
Columbus, Ohio 43210-1275

November 2000

Heights, the Geopotential, and Vertical Datums

Christopher Jekeli
Department of Civil and Environmental Engineering
and Geodetic Science
Ohio State University

Technical Report

Ohio Sea Grant Development Program, NOAA
Grant No. NA86RG0053 (R/CE-7-PD)
"Modern Vertical Datums and Lake Level"

November 2000

1. Introduction

At national, international, and regional levels, determining, monitoring, and understanding the changes in the water levels of the Great Lakes and providing vertical geodetic control to the region are important and necessary activities that relate to the economic utilization (commercial, agricultural, aquacultural), the ecological restoration and preservation, and the overall health of this many-faceted natural resource. Lake levels are measured within the construct and control of a height system. Several different types of height systems can be defined and have, in fact, been used in the Great Lakes region. The emergence in the last two decades of a variety of modern high-accuracy, three-dimensional and vertical positioning technologies based on satellite techniques provides new opportunities to monitor lake levels remotely over short as well as long periods; but they have also introduced new problems associated with establishing datum consistency between the modern techniques, as well as tying them to the more traditional land survey methods. This has particular importance in studying the long-term periodic and secular variations in lake levels, in extending and densifying vertical control around the region, and in defining the next International Great Lakes Datum (IGLD). To this seemingly straightforward problem of ensuring datum consistency comes the complication of post-glacial rebound (PGR) caused by the relaxation of the Earth's crust in the form of uplift following the retreat of the weighty ice sheets of the last ice age some 17,000 years ago. Several models have been developed (see Lindsay, 1996; also, Peltier, 1998; Simons and Hager, 1997) that predict about 5 cm per decade of north-south tilt of the Great Lakes region. In addition, there are other vertical displacements of the crust caused by periodic lunar and solar tidal deformations, local subsidence (oil and aquifer extractions), lake shore erosion, and seismic events. Lake level, itself, is affected by lake tides, the influx from the surrounding drainage basin, by prevailing winds and temperature gradients, and by atmospheric pressure changes.

In summary, lake water levels are determined and monitored in a geophysically dynamic environment, and it is clearly important to understand the origin of water level changes, whether due to changes in the amount and distribution of water or due to physical changes in the basins that contain the water. Toward this end the following report provides a systematic exposition of the definitions of heights, vertical datums, and the role played by the Earth's geopotential. Models are developed that relate modern height determination using satellite techniques to the traditional leveling procedures. This is accomplished within the more general temporally varying tidal potential and simplified models for post-glacial rebound are also included. The objective of the report is to provide the theoretical background for further studies related to lake level monitoring using modern satellite techniques. The discussion assumes that the reader is somewhat familiar with physical geodesy, in particular with the foundations of potential theory, but the development proceeds from first principles in review fashion. Moreover, concepts and geodetic quantities are introduced as they are needed, which should give the reader a sense that nothing is a priori given, unless so stated explicitly.

2. Heights

Points on or near the Earth's surface commonly are associated with three coordinates, a latitude, a longitude, and a height. The latitude and longitude refer to an oblate ellipsoid of revolution and are designated more precisely as *geodetic latitude and longitude*. This ellipsoid is a geometric, mathematical figure that is chosen in some way to fit the mean sea level either globally or, historically, over some region of the Earth's surface, neither of which concerns us at the moment. We assume that its center is at the Earth's center of mass and its minor axis is aligned with the Earth's reference pole. The height of a point, P, could refer to this ellipsoid, as does the latitude and longitude; and, as such, it designates the distance from the ellipsoid to the point, P, along the perpendicular to the ellipsoid (see Figure 1); we call this the *ellipsoidal height*, h_P.

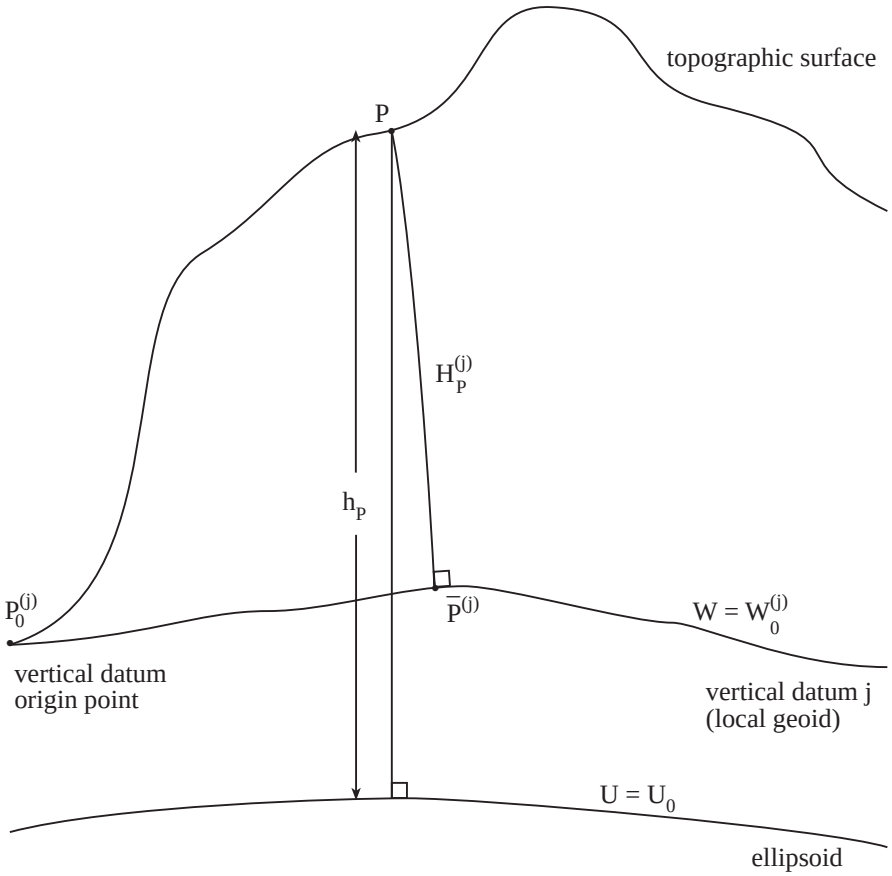


Figure 1: Ellipsoidal height versus orthometric height with respect to vertical datum, j.

However, in most surveying applications, the height of a point should refer to mean sea level in some colloquial sense, or more precisely to a *vertical datum* (i.e., a well-defined reference

surface for heights that is accessible at least at one point, called the *origin point*). We note that the ellipsoid surface is not the same as mean sea level, deviating from the latter on the order of 30 m, with maximum values up to 110 m, if geocentrically located. On the other hand, the ellipsoid may be identified as the "vertical datum" for ellipsoidal heights. In this case accessibility is achieved indirectly through the assumption that the ellipsoid is defined in a coordinate frame established by the satellite observations that yield the three-dimensional coordinates of a point.

A comprehensive discussion of the heights with respect to the traditional vertical datum, the geoid, cannot proceed without introducing the concept of *potential*; this is the topic of the next section.

2.1 Geopotential and Geoid

The *geopotential* is the gravitational potential generated by the masses of the Earth, including its atmosphere. In addition, the potential from other masses in the solar system may be considered separately, especially that of the sun and moon, as they introduce a time-varying field due to their apparent motions relative to the Earth; this will be done later. The geopotential, V_e, is expressed most conveniently in terms of spherical harmonic functions (an alternative formulation in terms of ellipsoidal harmonics is also sometimes used):

$$V_e(r,\theta,\lambda) = \frac{kM}{R} \sum_{n=0}^{\infty} \sum_{m=-n}^{n} \left(\frac{R}{r}\right)^{n+1} C_{nm}\, \bar{Y}_{nm}(\theta,\lambda)\,, \tag{1}$$

where (r,θ,λ) are the usual spherical polar coordinates (radial distance, co-latitude, longitude–these are not the geodetic coordinates mentioned above), kM is the product of Newton's gravitational constant and Earth's mass (including the atmosphere), R is a mean Earth radius, C_{nm} is a constant coefficient for degree n and order m, $\bar{Y}_{nm}$ is a fully-normalized, surface spherical harmonic function:

$$\bar{Y}_{nm}(\theta,\lambda) = \begin{cases} \bar{P}_{nm}(\cos\theta)\cos|m|\lambda\,, & m \geq 0\,; \\ \bar{P}_{n|m|}(\cos\theta)\sin|m|\lambda\,, & m < 0\,; \end{cases} \tag{2}$$

and $\bar{P}_{nm}$ is a fully-normalized Legendre function of the first kind. The series expression (1) is the solution to a boundary-value problem for the potential and holds only if the point with coordinates (r,θ,λ) is in free space (no mass). This is just a *necessary* condition; whereas, the convergence of the series to the true gravitational potential is guaranteed only for points outside a sphere enclosing

all masses (which represents a *sufficient* condition). In practice we have available a truncated version of the series ($n \le n_{max}$) that is used to approximate the potential at any point above the Earth's surface; this space exterior to the Earth, moreover, is approximated as being "free space". The effect of the atmosphere on the potential is, in fact, substantial, approximately $54\ m^2/s^2$ near the surface, but it is fully compensated to first order (Sjöberg, 1999) by mathematically "moving" the atmosphere radially inside the Earth.

Because of Earth's rotation, static gravimeters on the Earth experience a centrifugal acceleration that (because of Einstein's Equivalence Principle) cannot be distinguished from the effect of gravitation. In geodesy, we term the combination of centrifugal and gravitational acceleration (both vectors) as *gravity* (the resultant vector). Correspondingly, it is convenient to define a *centrifugal potential* that generates the centrifugal acceleration:

$$\Phi(r,\theta) = \frac{1}{2}\,\omega_e^2\, r^2 \sin^2\theta\ , \tag{3}$$

where ω_e is Earth's rotation rate. The *gravity potential* is then simply

$$W(r,\theta,\lambda) = V(r,\theta,\lambda) + \Phi(r,\theta)\ . \tag{4}$$

A surface on which a potential function has a constant value is called an *equipotential surface*; and, the equipotential surface of W, $W(r,\theta,\lambda) = \text{constant} = W_0$, that closely agrees with mean sea level is known as the *geoid* (first introduced by C.F. Gauss in 1828 to help define the shape of the Earth, later in 1873 clarified by J.B. Listing as specifically associated with the oceans, and in modern views understood to vary in time due to mass deformations and redistribution; see also Grafarend, 1994). The distance between the ellipsoid and the geoid is known as the *geoid undulation*, or also the *geoid height*.

Immediately, we note that equation (1), in principle, cannot be used to compute the potential of the geoid, W_0, in land areas, because the geoid generally lies within the continental crust and the model (1) holds only in free space. In addition, we know that mean sea level and the land (and ocean bottom) surface are affected by the gravitational attractions of the sun and moon (and negligibly by the other planets), which must be recognized in our definition of the geoid. Both of these issues will be discussed later (Bursa et al. (1997) show that the potential *value* of the geoid, however, is not affected).

The relationship between the gravity vector and its potential is simply

$$\mathbf{g} = \nabla W\ , \tag{5}$$

where ∇ denotes gradient operator. Generally, we know from the calculus that the gradient is a vector pointing in the direction of steepest descent of a function, that is, perpendicular to its isometric lines–for example, in the case of the potential, it is the vector perpendicular to the

equipotential surfaces. The component of gravity perpendicular to the equipotential surface thus embodies its magnitude, and we may write

$$|\mathbf{g}| = g = -\frac{dW}{dn}\,, \tag{6}$$

where dn is a differential path along the perpendicular and the minus sign is a matter of convention (potential decreases with altitude, path length is positive upwards, and gravity magnitude is positive).

2.2 Dynamic height

The height of a point "above mean sea level" now is defined more precisely as a height with respect to the geoid, which is a well defined surface, in principle, although its accessibility has yet to be established. In fact, we will first define a *local* geoid (*local vertical datum*) with a single point, $P_0^{(j)}$, that is assumed to be on it and accessible (e.g., a tide gauge station), and where the gravity potential is $W_0^{(j)}$ (not necessarily a known value).

There are three types of geoid-referenced height; each fundamentally refers to the difference in gravity potential between the (local) geoid and the point in question. This potential difference is known as the *geopotential number*:

$$C_P^{(j)} = W_0^{(j)} - W_P\,, \tag{7}$$

where W_P is the gravity potential at the point, P. Any point has a unique geopotential number (with respect to the defined local geoid), and this, itself, appropriately scaled, can be used as a height coordinate of the point. Specifically, for an adopted constant, γ_0, we have

$$H_P^{dyn(j)} = \frac{C_P^{(j)}}{\gamma_0}\,, \tag{8}$$

which is the *dynamic height* of P (with respect to vertical datum, j). The scale factor is usually chosen as a nominal value of gravity at mid-latitude ($\gamma_0 = 9.806199203\ \mathrm{m/s^2}$; GRS80 value, Moritz, 1992). Noting that the magnitude of gravity is approximately the vertical gradient of the potential (equation (6)), we see that the dynamic height, defined by (8), looks like a height (it has units of distance). On the other hand, in fact, it has no geometric meaning; it is a purely physical quantity–it is the potential (in distance units) relative to the geoid. Clearly, the same constant scale factor must be used for all dynamic heights within a particular datum.

2.3 Orthometric Height

Seeking a geometric definition of height, that is, in terms of an actual vertical distance, we may proceed from the relationship between gravity and potential, (6), integrated as follows:

$$W_P = W_0^{(j)} - \int_{\bar{P}_0^{(j)}}^{P} g \, dn \,, \tag{9}$$

where the path of integration from the local geoid to P is arbitrary (the gravity field is *conservative*) and could, for example, be an actual path along the Earth's surface. Note, moreover, that the initial point is also arbitrary as long as it is on the datum surface since the local geoid is an equipotential surface (same $W_0^{(j)}$). With (7), we have the intermediate result:

$$C_P^{(j)} = \int_{\bar{P}_0^{(j)}}^{P} g \, dn \,, \tag{10}$$

showing that geopotential numbers can be determined from measurements of gravity and vertical increments between equipotential surfaces along the path; the latter are obtained with spirit leveling.

Consider now the special path that is always perpendicular to the equipotential surfaces of W. This is called the *plumb line*, and suppose that the path of integration in (10) is along this plumb line. Then

$$C_P^{(j)} = \int_{\bar{P}^{(j)}}^{P} g \, dH \,, \tag{11}$$

where dH is a differential element along the plumb line and $\bar{P}^{(j)}$ is at the base of the plumb line on the local geoid. Dividing and multiplying the right side by the total length of the plumb line, $H_P^{(j)}$, we obtain

$$H_P^{(j)} = \frac{C_P^{(j)}}{\bar{g}_P^{(j)}} \,, \tag{12}$$

where

$$\bar{g}_P^{(j)} = \frac{1}{H_P^{(j)}} \int_{\bar{P}^{(j)}}^{P} g \, dH \tag{13}$$

is the average value of gravity along the plumb line. $H_P^{(j)}$ is known as the *orthometric height* of P (with respect to the defined local vertical datum); it has a very definite geometric interpretation as a distance above the local geoid (along the plumb line, which is curved since the equipotential surfaces are not parallel); see Figure 1.

Unfortunately, while $C_P^{(j)}$ can be measured using (11), the value of $\bar{g}_P^{(j)}$ cannot be *computed* exactly (using Newton's Law of Gravitation) because this would require complete knowledge of the mass density of the crust (and it is not practical to measure g along the plumb line). Therefore, in theory, the orthometric height cannot be determined exactly, and its calculation depends on some density hypothesis, or model, for the crust. A commonly used model assumes constant crustal density and constant topographic height in the vicinity of the point, P. Then, the average gravity along the plumb line, between $\bar{P}^{(j)}$ and P, is obtained using the *Prey reduction*. This reduction models the value of gravity inside the crust by first removing a Bouguer plate of density, ρ, applying a free-air downward continuation using the normal gradient of gravity, and restoring the Bouguer plate. Each of these operations is linear in H; and, consequently, the average value of gravity along the plumb line is simply the average of the endpoint values of gravity, that is:

$$\bar{g}_P^{\text{Prey}(j)} = \frac{1}{2}\left[g_P + \left(g_P - 2\pi k\rho \, H_P^{(j)} + \frac{\partial \gamma}{\partial h} H_P^{(j)} - 2\pi k\rho \, H_P^{(j)} \right) \right]. \tag{14}$$

With nominal values for the density and gradient (Heiskanen and Moritz, 1967), we obtain

$$\begin{aligned} \bar{g}_P^{\text{Prey}(j)} &= g_P - 2\pi k\rho \, H_P^{(j)} + \frac{1}{2} \frac{\partial \gamma}{\partial h} H_P^{(j)} \\ &= g_P + (0.0424 \text{ mgal/m}) \, H_P^{(j)} . \end{aligned} \tag{15}$$

Utilizing (15) in (12) (i.e., $\bar{g}_P^{(j)} \approx \bar{g}_P^{\text{Prey}(j)}$) yields *Helmert heights*:

$$H_P^{\text{Helmert}(j)} = \frac{C_P^{(j)}}{\bar{g}_P^{\text{Prey}(j)}}, \tag{16}$$

which, in principle, requires an iteration on $H_P^{(j)}$; or, since (12) with (15) is just a quadratic in $H_P^{(j)}$, we may write

$$H_P^{\text{Helmert(j)}} = \frac{C_P^{(j)}}{g_P}\left(1 - (0.0424\ \text{mgal/m})\,\frac{C_P^{(j)}}{g_P^2} + (0.00180\ \text{mgal}^2/\text{m}^2)\left(\frac{C_P^{(j)}}{g_P^2}\right)^2 - \ldots\right), \qquad (17)$$

where higher-order terms are less than $O\left(10^{-10}\right)$.

2.4 Normal Height

It is possible to define a similar geometrically interpretable height that avoids a density hypothesis for the crust. This is accomplished by introducing an approximation to the gravity field that can be calculated exactly at any point. The *normal gravity field* suits this purpose. It is defined as the gravity field generated by an Earth-fitting ellipsoid that contains the total mass of the Earth (including the atmosphere), that rotates with the Earth around its minor axis, and that is, itself, an equipotential surface of the gravity field it generates. The gravitational part, V_{ellip}, of the normal field can be expressed as in (1), but because of the imposed symmetries of the ellipsoid and the boundary values, the series contains only even zonal harmonics (no dependence on longitude). The centrifugal part is also given by (3), and the total *normal gravity potential* is

$$U(r,\theta) = V_{ellip}(r,\theta) + \Phi(r,\theta)\ . \qquad (18)$$

On the ellipsoid, U is a constant, U_0, by definition. U can be calculated anywhere in the space above the ellipsoid using four constants that describe the size and shape of the ellipsoid, it mass, and its rotation. Nowadays, one typically uses:

$$\begin{aligned} &a = \text{semi–major axis of the ellipsoid ;}\\ &J_2 = \text{dynamical flattening (second–degree zonal harmonic coefficient) ;}\\ &kM = \text{Newton's gravitational constant times Earth's mass (including atmosphere) ;}\\ &\omega_e = \text{Earth's rate of rotation .} \end{aligned} \qquad (19)$$

The value of the potential, U_0, is completely determined by the adopted constants in (19); it is given by the Pizzetti formula (Heiskanen and Moritz, 1967):

$$U_0 = \frac{kM}{E}\tan^{-1}\frac{E}{b} + \frac{1}{3}\,\omega_e^2\,a^2\ , \qquad (20)$$

where b is the semi-minor axis of the normal ellipsoid, and E it its linear eccentricity. The *normal gravity vector* is, analogous to (5):

$$\boldsymbol{\gamma} = \boldsymbol{\nabla} U \,, \tag{21}$$

and can, likewise, be calculated exactly anywhere on or above the ellipsoid from the given constants (NIMA, 1997).

Consider now the *normal plumb line* through P; it is the line that is always perpendicular to the equipotential surfaces of the normal gravity field. On that line is a point, Q, where the normal gravity potential equals the actual gravity potential at P:

$$U_Q = W_P \,. \tag{22}$$

Note that U and W refer to two distinct gravity fields, and the equality above is not a functional relationship, simply an assignment of values. We define the *normal geopotential number* of Q (also called *spheropotential number*) as follows:

$$C_Q^{\text{normal}} = U_0 - U_Q = U_0 - W_P = C_P^{(j)} + \left(U_0 - W_0^{(j)} \right), \tag{23}$$

where (7) was used and where, analogous to (11):

$$C_Q^{\text{normal}} = \int_{\bar{Q}}^{Q} \gamma \, dH^* \,, \tag{24}$$

and $\bar{Q}$ is the base point on the ellipsoid (not the geoid!) of the normal plumb line (see Figure 2). Also dH^* denotes a differential path element along the normal plumb line.

Dividing and multiplying the right side of (24) by the total length, H_Q^*, of the normal plumb line from the ellipsoid to Q, we obtain with (23):

$$H_Q^* = \frac{C_P^{(j)} + \left(U_0 - W_0^{(j)} \right)}{\bar{\gamma}_Q} \,, \tag{25}$$

where

$$\bar{\gamma}_Q = \frac{1}{H^*_Q} \int_{\bar{Q}}^{Q} \gamma \, dH^* \qquad (26)$$

is the average value of normal gravity along the normal plumb line. Now, formally, the point Q lies on the *telluroid* and the distance between the telluroid and the Earth's surface is known as the *height anomaly* at P, ζ_P. Conventionally, the distances H^*_Q and ζ_P are "reversed" along the normal plumb line, and the surface defined by the separation ζ_P from the ellipsoid is known as the *quasi-geoid*; see Figure 2. The shape of the quasi-geoid is similar to that of the geoid, but the quasi-geoid is not an equipotential surface in either the normal or the actual gravity field. If the point, P, is on the geoid (as it is approximately in ocean areas) and if the gravity potential value of the geoid is $W_0 = U_0$, then the telluroid point, Q, is on the ellipsoid. Thus the quasi-geoid equals the geoid at these points.

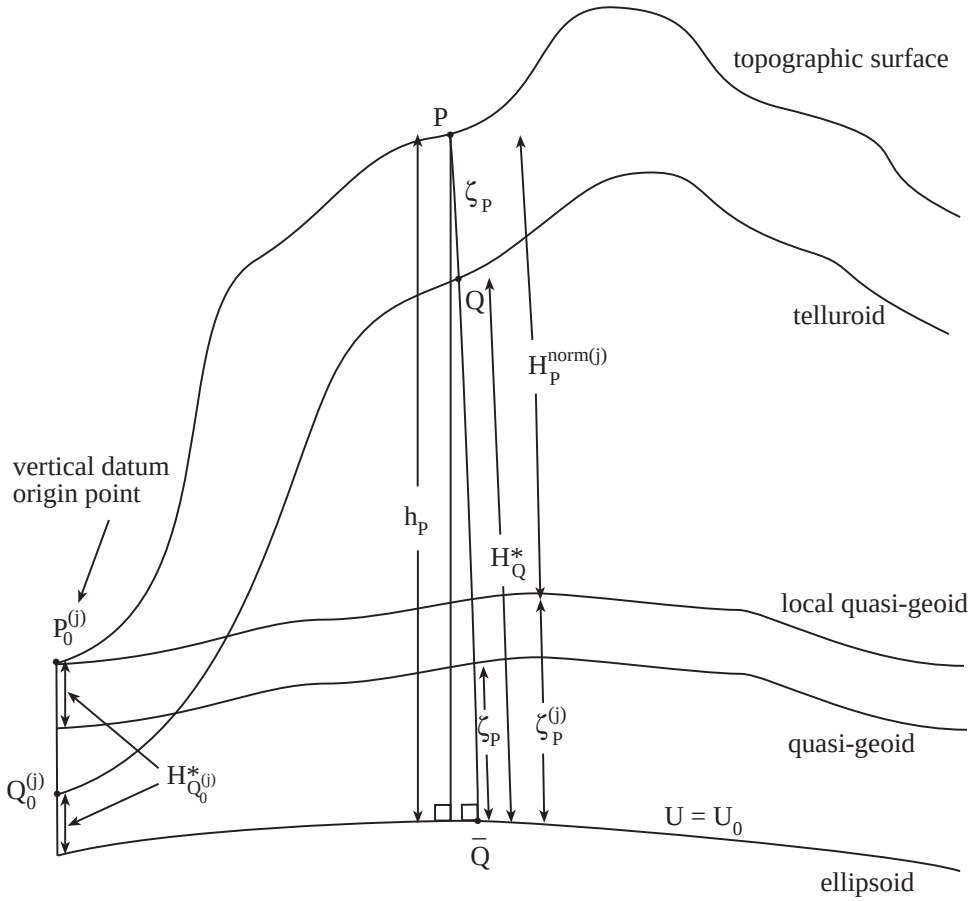


Figure 2: Normal height and height anomaly; also quasi-geoid versus local quasi-geoid.

We assume that the difference, $U_0 - W_0^{(j)}$, in (25) is generally not known (which is the case, at least historically). Consider the case when P is the origin point, $P_0^{(j)}$, of the local vertical datum. Then $C_{P_0^{(j)}}^{(j)} = 0$ and for the corresponding telluroid point, $Q_0^{(j)}$, we have

$$H_{Q_0^{(j)}}^{*} = \frac{\left(U_0 - W_0^{(j)}\right)}{\bar{\gamma}_{Q_0^{(j)}}} . \tag{27}$$

$H_{Q_0^{(j)}}^{*}$ is the distance from the ellipsoid to that point, $Q_0^{(j)}$, where $U_{Q_0^{(j)}} = W_0^{(j)}$, or, alternatively, the distance from the point, $P_0^{(j)}$, to the quasi-geoid.

We now define a *local quasi-geoid* that contains the datum point $P_0^{(j)}$ and is parallel to the quasi-geoid by the amount $H_{Q_0^{(j)}}^{*}$ (Figure 2). The distance from the local quasi-geoid to the point P is then given the name, *normal height*, and is expressed by (25) relative to $H_{Q_0^{(j)}}^{*}$, and now with new notation:

$$H_P^{\mathrm{norm(j)}} = \frac{C_P^{(j)}}{\bar{\gamma}_Q} . \tag{28}$$

Using an approximate form of normal gravity as a function of (ϕ,h), Heiskanen and Moritz (1967) derive the average normal gravity, $\bar{\gamma}_Q$, using (26):

$$\bar{\gamma}_Q \approx \gamma \left[1 - \left(1 + f + \frac{\omega_e^2 \, a^2 \, b}{kM} - 2f \sin^2\phi \right) \frac{H_P^{\mathrm{norm(j)}}}{a} + \left(\frac{H_P^{\mathrm{norm(j)}}}{a} \right)^2 \right], \tag{29}$$

and, substituting this into (28) and inverting, they find an expression for the normal height:

$$H_P^{\mathrm{norm(j)}} \approx \frac{C_P^{(j)}}{\gamma} \left[1 + \left(1 + f + \frac{\omega_e^2 \, a^2 \, b}{kM} - 2f \sin^2\phi \right) \frac{C_P^{(j)}}{a\,\gamma} + \left(\frac{C_P^{(j)}}{a\,\gamma} \right)^2 \right]. \tag{30}$$

We note for future reference that

$$H_P^{\mathrm{norm(j)}} = H_Q^{*} - H_{Q_0^{(j)}}^{*} . \tag{31}$$

Also, we can define the *local height anomaly*, identifying the separation of the local quasi-geoid from the ellipsoid (Figure 2):

$$\zeta_P^{(j)} = \zeta_P + H^*_{Q_0^{(j)}} = \zeta_P + \frac{\left(U_0 - W_0^{(j)}\right)}{\gamma_{Q_0^{(j)}}} \,. \tag{32}$$

2.5 Review of Heights

It is important to realize that the normal height depends in the first place, like the dynamic and orthometric heights, on the geopotential number at P, $C_P^{(j)}$ (which is the actual geopotential value with respect to the local geoid potential). Unlike orthometric heights, normal heights can be determined exactly, albeit some iterative procedure is required in the computation of $\bar{\gamma}_Q$, since it depends also on the normal height. And, unlike dynamic heights, normal heights have a definite geometric interpretation, as the vertical distance of P above the local quasi-geoid. Finally, we will see that geopotential models, such as (1), yield the height anomaly more readily than the geoid undulation (in fact, the latter, in theory, is not determinable because we do not know the crustal mass density); and, therefore, the quasi-geoid is theoretically more realizable from geopotential models than the geoid. One disadvantage of the normal height is its arcane definition, being the height above the quasi-geoid and not the geoid (which is usually, though with some error, identified as mean sea level).

A disadvantage of both orthometric and normal heights is that neither indicates the direction of flow of water. Only dynamic heights possess this property. That is, two points with identical dynamic heights are on the same equipotential surface (of the actual gravity field) and water will not flow from one to the other point. Two points with identical orthometric heights lie on different equipotential surfaces (since, generally, $\bar{g}_{P_1} \neq \bar{g}_{P_2}$); and, water *will* flow from one point to the other, even though they have the same (orthometric) height. The same statement holds for normal heights, although, because of the smoothness of the normal field, the effect is not as severe. For these reasons the International Great Lakes Datum of 1985 is based on dynamic heights (Zilkoski et al., 1992).

Table 1 gives a summary of the three types of heights, as well as variations of these based on different assumptions and approximations. The reader is cautioned that the nomenclature of the approximate forms and the approximations, themselves, are not universal and can lead to confusion. The given formulas correspond to definitions given in NGS (1986).

Table 1: Height definitions according to NGS (1986).

ellipsoidal height (geodetic height)	h_P	$h_P = h_P(x_P, y_P, z_P, a, f)$
dynamic height	$H_P^{dyn(j)} = \dfrac{C_P^{(j)}}{\gamma_0}$	$\gamma_0 = \gamma(\phi_0)$, $\phi_0 = 45°$
normal dynamic height		$C_P^{(j)} \approx \int_{P_0^{(j)}}^{P} \gamma \, dH^*$, $\gamma_0 = \gamma(\phi_0)$
orthometric height	$H_P^{(j)} = \dfrac{C_P^{(j)}}{\bar{g}_P^{(j)}}$	$\bar{g}_P^{(j)} = \dfrac{1}{H_P^{(j)}} \int_{\bar{P}^{(j)}}^{P} g \, dH$
Helmert height		$\bar{g}_P^{(j)} \approx g_P - 2\pi k\rho \, H_P^{(j)} + \dfrac{1}{2} \dfrac{\partial \gamma}{\partial h} H_P^{(j)}$
Niethammer height		$\bar{g}_P^{(j)} \approx \bar{g}_P^{Prey(j)} + \delta g_t$ †
normal orthometric height		$C_P^{(j)} \approx \int_{P_0^{(j)}}^{P} \gamma \, dH$, $\bar{g}_P^{(j)} \approx \gamma\left(\phi_P, \frac{1}{2} H_P^{(j)}\right)$
normal height (Molodensky height)	$H_P^{norm(j)} = \dfrac{C_P^{(j)}}{\bar{\gamma}_Q}$	$\bar{\gamma}_Q = \dfrac{1}{H_Q^*} \int_{\bar{Q}}^{Q} \gamma \, dH^*$
quasi-dynamic height		$\bar{\gamma}_Q \approx \gamma(\phi_0) + \dfrac{H_P^{(j)}}{2} \dfrac{\partial \gamma}{\partial h}\Big\|_{h = H_P^{(j)}}$
height anomaly	$\zeta_P^{(j)}$	$\zeta_P^{(j)} = \dfrac{T_P}{\gamma_Q} + \dfrac{\left(U_0 - W_0^{(j)}\right)}{\gamma_{Q_0^{(j)}}}$
geoid height (geoid undulation)	$N_P^{(j)}$	$N_P^{(j)} = \dfrac{T_{\bar{P}^{(j)}}}{\gamma_{Q^{(j)}}} + \dfrac{1}{\gamma_{\bar{Q}}} \left(U_{\bar{Q}} - W_0^{(j)}\right)$

† δg_t accounts for the terrain relative to the Bouguer plate in a modified Prey reduction.

Clearly, given sufficient gravity information, the dynamic, orthometric, and normal heights can be transformed from one to the other, because they all depend firstly on the geopotential number. From (8), (12), and (28), we have

$$C_P^{(j)} = \gamma_0 \, H_P^{dyn(j)} = \bar{g}_P^{(j)} \, H_P^{(j)} = \bar{\gamma}_Q \, H_P^{norm(j)} \, . \tag{33}$$

The difference between a length along the plumb line and a corresponding length along the perpendicular (or normal) to the ellipsoid is due to the curvature of the former. The curvature results in a deflection of the plumb line from the ellipsoid normal (deflection of the vertical, Θ) that is usually of the order of 10 arcsec and in rare cases can reach 1 arcmin. With reference to Figure 3, we see that the corresponding height difference is

$$\delta h \approx h \sin\Theta \tan\Theta . \qquad (34)$$

This is a negligible effect for all topographic heights of the Earth (even for the extreme case of $\Theta = 1$ arcmin and $h = 10000$ m, we obtain $\delta h < 1$ mm). Thus we can treat all geometrically interpretable heights as lengths along the ellipsoidal normal, which considerably simplifies comparisons and conversions among the different heights.

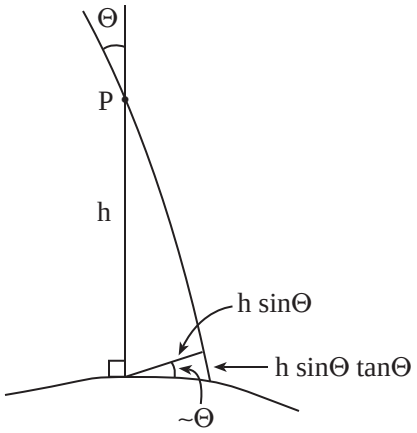


Figure 3: The difference between lengths along the curved plumb line and along the straight ellipsoid perpendicular.

3. Models for $W_0^{(j)}$

We wish to find a means to determine the potential value of the local vertical datum, j. Then we are able to relate different datums around the world and also define a world vertical datum. Clearly, if we have an estimate of the geopotential (e.g., EGM96, Lemoine et al., 1998) and we know the (geocentric) coordinates of a point on the surface of the local vertical datum (the local geoid), then it is simply a matter of evaluating the gravity potential at this point to find $W_0^{(j)}$. The problem generalizes if the point is not on the local geoid, but we know its height with respect to the datum and thus, again, have the requisite coordinates to evaluate the gravity potential. Thus, overall, we require the following data to make an estimation of $W_0^{(j)}$: an estimate of the gravity potential

function, W; the geocentric coordinates of a point (e.g., r,θ,λ), and the height of this point with respect to the datum.

We start with an expansion of the normal gravity potential in a Taylor series along the ellipsoid normal:

$$U_P = U_Q + \left(h_P - h_Q\right) \left.\frac{\partial U}{\partial h}\right|_{h=h_Q} + \frac{1}{2!}\left(h_P - h_Q\right)^2 \left.\frac{\partial^2 U}{\partial h^2}\right|_{h=h_Q} + \cdots . \tag{35}$$

Now, with

$$\gamma_Q = -\left.\frac{\partial U}{\partial h}\right|_{h=h_Q} \tag{36}$$

and with (22), we obtain:

$$h_P - h_Q = \frac{1}{\gamma_Q}\left(W_P - U_P\right) - \frac{1}{2\,\gamma_Q}\left(h_P - h_Q\right)^2 \left.\frac{\partial \gamma}{\partial h}\right|_{h=h_Q} + \cdots . \tag{37}$$

The left side is the height anomaly, ζ_P. The second term (and higher-order terms) on the right side can be ignored; since $\left|h_P - h_Q\right| < 110\ \text{m}$ and the vertical gradient of gravity is 0.3086 mgal/m, it amounts to less 2 mm. Defining the *disturbing potential*, T_P, at point P:

$$T_P = W_P - U_P , \tag{38}$$

we then obtain from (37) the height anomaly in *physical* rather than geometric terms:

$$\zeta_P = \frac{T_P}{\gamma_Q} . \tag{39}$$

From (32), we obtain the local height anomaly, also in terms of the disturbing potential:

$$\zeta_P^{(j)} = \frac{T_P}{\gamma_Q} + \frac{\left(U_0 - W_0^{(j)}\right)}{\gamma_{Q_0^{(j)}}} . \tag{40}$$

If the point, P, is on the local geoid ($P = \bar{P}^{(j)}$), then we obtain similarly as in (37):

$$h_{\bar{P}^{(j)}} - h_{Q^{(j)}} = \frac{T_{\bar{P}^{(j)}}}{\gamma_{Q^{(j)}}}, \tag{41}$$

where $Q^{(j)}$ is the point at which $U_{Q^{(j)}} = W_{\bar{P}^{(j)}}$. The *local geoid undulation* is obtained as follows. First note that by again applying (35) ($P \to Q^{(j)}$ and $Q \to \bar{Q}$), we have

$$\begin{aligned} h_{Q^{(j)}} - h_{\bar{Q}} &= \frac{1}{\gamma_{\bar{Q}}}\left(U_{\bar{Q}} - U_{Q^{(j)}}\right) \\ &= \frac{1}{\gamma_{\bar{Q}}}\left(U_{\bar{Q}} - W_{\bar{P}^{(j)}}\right). \end{aligned} \tag{42}$$

Thus, with the geoid undulation, $N_P^{(j)} = h_{\bar{P}^{(j)}} - h_{\bar{Q}}$, and $W_{\bar{P}^{(j)}} = W_0^{(j)}$ we obtain

$$N_P^{(j)} = \frac{T_{\bar{P}^{(j)}}}{\gamma_{Q^{(j)}}} + \frac{1}{\gamma_{\bar{Q}}}\left(U_{\bar{Q}} - W_0^{(j)}\right), \tag{43}$$

Equations (39), (40), and (43) are manifestations of the (*generalized*) *Bruns' formula*.

From leveling (and gravity data) we obtain the normal height, $H_P^{norm(j)}$, with respect to the local quasi-geoid according to (11), (26), and (28). From GPS we obtain the ellipsoidal height, h_P; which allows us to compute at P the actual potential from a model (e.g., EGM96), assumed errorless for the moment. Also, the normal potential can be computed at P, hence T_P can be computed according to (38). From Figure 2, we have

$$h_P = H_P^{norm(j)} + \zeta_P^{(j)}. \tag{44}$$

Using the expression (40) for the local height anomaly, we can solve for the local geoid potential:

$$W_0^{(j)} = U_0 - \gamma_{Q_0^{(j)}}\left(h_P - H_P^{norm(j)} - \frac{T_P}{\gamma_Q}\right). \tag{45}$$

All quantities on the right side are either given or measured, and the left side is the actual potential of the local geoid. To verify this equation, select the point, P, to be the origin point, $P_0^{(j)}$, of the local vertical datum; then (45) reduces to

$$W_0^{(j)} = U_0 - \gamma_{Q_0^{(j)}} \left(h_{P_0^{(j)}} - \frac{W_{P_0^{(j)}} - U_{P_0^{(j)}}}{\gamma_{Q_0^{(j)}}} \right)$$

$$= W_{P_0^{(j)}} + U_0 - U_{P_0^{(j)}} - \gamma_{Q_0^{(j)}} h_{P_0^{(j)}} \qquad (46)$$

$$= W_{P_0^{(j)}} ,$$

as it should be (to first-order approximation).

Re-defining the unknown parameter as

$$\Delta H_0^{(j)} = \frac{U_0 - W_0^{(j)}}{\gamma_{Q_0^{(j)}}} , \qquad (47)$$

we have the following model

$$\Delta H_0^{(j)} = h_P - H_P^{norm(j)} - \zeta_P . \qquad (48)$$

A similar model can be derived that involves the orthometric height. In this case, we must assume that the gravity potential model can be evaluated at the local geoid point, $\bar{P}^{(j)}$ (the coordinates of $\bar{P}^{(j)}$ come from the ellipsoidal height obtained by GPS and the orthometric height, $H_P^{(j)}$, obtained by leveling and gravity (and a density assumption); and we still need a density assumption to compute the potential inside the mass). Then from Figure 2, we have

$$h_P = H_P^{(j)} + N_P^{(j)} ; \qquad (49)$$

and, if we substitute (43), we again obtain an equation for the local geoid potential:

$$W_0^{(j)} = U_{\bar{Q}} - \gamma_{\bar{Q}} \left(h_P - H_P^{(j)} - \frac{T_{\bar{P}^{(j)}}}{\gamma_{Q^{(j)}}} \right) . \qquad (50)$$

Because of the density hypotheses needed to compute both the orthometric height and the disturbing potential, this model is less useful for precise determinations of the local geoid potential. If orthometric heights are available at certain points of the datum, then it would be advantageous to first convert these to normal heights, using (33), and then apply the model (48).

This section concludes with formulas for the orthometric and normal heights of a particular local vertical datum, determined on the basis of the gravity potential and ellipsoidal heights. Often, nowadays, the determination of the geoid undulation (or, more correctly, the gravity potential

function) is justified on the basis that it provides orthometric heights (or, normal heights) if ellipsoidal heights are given, the latter being readily determined by satellite methods such as GPS. Clearly, equations (44) and (49) are the basis for these determinations. Note, however, that it is the geoid undulation for the local geoid, or the local vertical datum, that is required, which implies that in addition to the gravity potential function, W (or T), the potential value of the local geoid must be known. That is, substituting (43) into (49) yields

$$H_P^{(j)} = h_P - \frac{T_{\bar{P}^{(j)}}}{\gamma_{Q^{(j)}}} - \frac{1}{\gamma_{\bar{Q}}}\left(U_{\bar{Q}} - W_0^{(j)}\right). \tag{51}$$

Similarly, for the normal height with respect to the local vertical datum, we substitute (40) into (44) to obtain

$$H_P^{norm(j)} = h_P - \frac{T_P}{\gamma_Q} - \frac{1}{\gamma_{Q_0^{(j)}}}\left(U_0 - W_0^{(j)}\right). \tag{52}$$

Thus, we see the importance of knowing the local geoid potential, $W_0^{(j)}$, in order to determine orthometric (or normal) heights in a local vertical datum with GPS and the disturbing potential function.

4. Local Determinations

Determinations of $W_0^{(j)}$ can be made wherever we know the geocentric coordinates (primarily the geometric height), as well as the potential-related height of points, provided we have knowledge of the geopotential. The model relating these quantities is given by (48), repeated here in slightly different form:

$$F_k\left(\Delta H_0^{(j)}, \mathbf{h}, \mathbf{H}^{norm(j)}, \boldsymbol{\zeta}\right) = \Delta H_0^{(j)} - h_k + H_k^{norm(j)} + \zeta_k = 0\,, \tag{53}$$

where the points on the Earth's surface are now indexed, $k = 1, ..., K$, and the vectors in the argument of F_k are

$$\mathbf{h} = \begin{pmatrix} h_1 \\ \vdots \\ h_K \end{pmatrix}, \quad \mathbf{H}^{norm(j)} = \begin{pmatrix} H_1^{norm(j)} \\ \vdots \\ H_K^{norm(j)} \end{pmatrix}, \quad \boldsymbol{\zeta} = \begin{pmatrix} \zeta_1 \\ \vdots \\ \zeta_K \end{pmatrix}. \tag{54}$$

The complete model comprises K equations of the type (53). It is linear in the unknown datum potential, $W_0^{(j)}$, as well as in the other quantities, here all interpreted as observations. The height anomaly, related to the disturbing potential (39), is derived from many other types of observations, including gravity anomalies, satellite perturbation observations, satellite altimetry, and any other observations that are combined to form the model for the disturbing potential. However, we will treat the disturbing potential estimate as an observation with error variance (or, one could also interpret it as a random parameter with an a priori variance).

Let

$$\ell_0 = \left(\mathbf{h}_0^T \quad \left(\mathbf{H}_0^{\text{norm(j)}}\right)^T \quad \boldsymbol{\zeta}_0^T \right)^T \tag{55}$$

be the vector of all observations and let $X_0 = \left(\Delta H_0^{(j)}\right)_0$ be an initial value of the unknown parameter (say, $X_0 = 0$). Because of its linearity the model (53) can be written as a Taylor expansion with only linear terms:

$$A\,\delta X + B\,\delta\ell + Y_0 = 0\,, \tag{56}$$

where $\delta X = X - X_0$, $\delta\ell = \ell - \ell_0$, and

$$\mathbf{Y}_0 = \begin{pmatrix} F_1\left(\left(\Delta H_0^{(j)}\right)_0, \mathbf{h}_0, \mathbf{H}_0^{\text{norm(j)}}, \boldsymbol{\zeta}_0\right) \\ \vdots \\ F_K\left(\left(\Delta H_0^{(j)}\right)_0, \mathbf{h}_0, \mathbf{H}_0^{\text{norm(j)}}, \boldsymbol{\zeta}_0\right) \end{pmatrix}; \tag{57}$$

and the coefficient matrices are given by

$$A = \left(\frac{\partial F_k}{\partial\left(\Delta H_0^{(j)}\right)}\right) = \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix}, \quad B = \left(\frac{\partial F_k}{\partial\left(h_m, H_m^{\text{norm(j)}}, \zeta_m\right)}\right) = \left(-I \quad I \quad I\right), \tag{58}$$

where each identity matrix, I, in B is a $K \times K$ matrix.

If C is the variance-covariance matrix of the observation errors, then the least-squares estimate of the unknown parameter is given by

$$\delta\hat{X} = -\left(A^T \left(B\,C\,B^T\right)^{-1} A\right)^{-1} A^T \left(B\,C\,B^T\right)^{-1} Y_0\,. \tag{59}$$

We may assume that C is diagonal if no better information on the error statistics is available. Moreover, if the variances of each observation type are invariant, given respectively by σ_h^2, σ_H^2,

and σ_ζ^2, then we have simply

$$\delta\hat{X} = -\frac{1}{K}\sum_{k=1}^{K} F_k\left(\left(\Delta H_0^{(j)}\right)_0, \mathbf{h}_0, \mathbf{H}_0^{norm(j)}, \boldsymbol{\zeta}_0\right), \tag{60}$$

or, as expected,

$$\Delta\hat{H}_0^{(j)} = \frac{1}{K}\sum_{k=1}^{K}\left(\left(h_k\right)_0 - \left(H_k^{norm(j)}\right)_0 - \left(\zeta_k\right)_0\right). \tag{61}$$

Such determinations of the vertical datum potential were made, for example, by Rapp (1994) for various datums around the world, by Grafarend and Ardalan (1997) for the Baltic Sea area, and by Bursa et al. (1999) for North America. In each case a spherical harmonic model for the geopotential was used to compute the height anomaly (or geoid undulation). In (Rapp and Balasubramania, 1992) more accurate geoid undulations were computed on the basis of local gravity data. Determinations of the time-variation of the vertical datum potential have also been attempted (Ardalan and Grafarend, 1999). When monitoring a vertical datum in time, of course, the time-variation of the gravity field potential must be understood. This is the object of Section 6.

5. Global Determinations

The discussion above pertains to the case where we wish to find the value of the vertical datum (local geoid) potential, $W_0^{(j)}$. So far the determination of the geoid potential value, W_0, has not been specifically addressed, mainly because the accessibility of the geoid has not been established–we can define it (although there are different options), but the question of how to access it is more difficult to answer. Even the classic Gauss/Listing definition noted in Section 2.1 is not satisfactory in view of the temporal aspects associated with the tidal potential (see Section 6.

There are various ways of defining the geoid as an accessible equipotential surface. Heck and Rummel (1991) give an excellent review that includes the method derived on the basis of the discussion above. That is, we either have one vertical datum which then defines the accessible geoid, or (what is more realistic in a global sense) we have a set of vertical datums and the geoid is defined in some precise (accessible) way in relation to these datums.

For two datums, j and k, we can find the difference $W_0^{(j)} - W_0^{(k)}$ and thus are able to relate all datums to each other or to a global datum, the geoid. Suppose there are J datums, j = 1, ..., J ; then a model for the local geoid potential in terms of normal heights is given by (45) or (50). To relate the local geoid value to a global value, W_0, we introduce new unknowns, $\Delta W_0^{(j)} = W_0 - W_0^{(j)}$. Then from (45):

$$H_P^{norm(j)} = h_P - \frac{T_P}{\gamma_Q} - \frac{1}{\gamma_{Q_0^{(j)}}}\left(U_0 - W_0 + \Delta W_0^{(j)}\right);\ j = 1, ..., J\ . \tag{62}$$

Assume for the moment that the disturbing potential and the ellipsoidal height are known. Then the unknowns are W_0 and $\Delta W_0^{(j)}$, $j = 1, ..., J$. The observations are normal heights in each datum.

It should be clear that there is an indeterminacy among the unknowns, even if we have many observed normal heights since all normal heights belonging to one vertical datum can only determine one potential value, $\Delta W_0^{(j)}$. Therefore, we must introduce additional information in the form of a constraint on the unknowns. This constraint can take many forms. For example, we could let $W_0 = W_0^{(1)}$. In this case we define the geoid to be the first vertical datum and all other datums are determined with respect to this one. Or, we could define the geoid according to some predefined potential value, such as $W_0 = U_0$. Then all local datums will be determined relative to this global geoid. This predefined value could also come from a fit of mean sea level to an equipotential surface (Bursa et al., 1998); then it may have a variance associated with it. Finally, we could impose an inner constraint:

$$\sum_{j=1}^{J} \Delta W_0^{(j)} = 0 \qquad \Leftrightarrow \qquad \frac{1}{J}\sum_{j=1}^{J} W_0^{(j)} = W_0\ . \tag{63}$$

In this way the geoid is defined to be the "mean" of all local vertical datums. We could also include different weights for each datum.

6. Temporal Effects

The Earth's solid masses, in fact, are elastic to some extent, and they consequently are subject to displacement under various internal and external influences. Also, the oceans and the atmosphere are fluid; and a change in the distribution or location of any these masses with time implies a temporal change in the gravitational potential, and hence in the gravity potential (in an Earth-fixed coordinate frame). Aside from this, any equipotential surface belongs to the field not only generated by the geopotential (and centrifugal potential), but also by the gravitational potentials of other bodies in the solar system, especially the sun and moon. Inasmuch as these bodies move with respect to the Earth (due to Earth's rotation and the orbital motions of the Earth and Moon), the equipotential surface accordingly changes in time. The mass displacements are also largely due to the tidal forces of the sun and moon. Therefore, we consider in detail the *tidal potential*, which is the source of a direct effect and several indirect effects.

In an Earth-fixed coordinate system, the tidal potential due to an external point mass body, designated generically as B, is given approximately by (Torge, 1991) as

$$V_B(r,\psi,\lambda) = D_B(r)\left[\cos^2\psi \cos^2\delta_B \cos^2 2t_B + \sin 2\psi \sin 2\delta_B \cos t_B + 3\left(\sin^2\psi - \frac{1}{3}\right)\left(\sin^2\delta_B - \frac{1}{3}\right)\right], \tag{64}$$

where t_B is the hour angle of the body:

$$t_B = \lambda + t_G - \alpha_B \tag{65}$$

and (α_B,δ_B) are the right ascension and declination of the body, while t_G is the hour angle of the vernal equinox at the Greenwich meridian (i.e., Greenwich sidereal time). The coordinates of the point of evaluation are geocentric radial distance, r, and geocentric latitude, ψ, and longitude, λ. The radial dependence of V_B is given by

$$D_B(r) = \frac{3}{4} kM_B \frac{r^2}{r_B^3} \tag{66}$$

where r_B is the mean distance between the Earth and the body. D_B is known as *Doodson's coefficient* and reflects that (64) is a second-degree *interior* harmonic potential (point of evaluation is inside sphere containing the body: $r \le r_B$).

The tidal potential (64) varies in time, as viewed at a point on the Earth, with different periods, such as fortnightly (due to the moon) or semi-annually (due to the sun) as the coordinates (α_B,δ_B) vary, and diurnally because of Earth's rotation described by t_G. There is also a constant part, the average over time, that is not zero; this is the *permanent tide* due to the approximate coplanarity of the Earth-sun-moon system. Averaging (64) over time (i.e. over the hour angle, $0 \le t_B \le 2\pi$), we see that the first two terms have zero mean. Alternatively, we can average over the ecliptic longitude, ν_B, since all bodies of the solar system orbit roughly on the ecliptic, as does the sun viewed by Earth with respect to inertial space. Neglecting the inclination of the body's orbit with respect to the ecliptic, it is easily seen that

$$\sin\delta_B = \sin\varepsilon \sin\nu_B, \tag{67}$$

where $\varepsilon \approx 23° 44'$ is the obliquity of the ecliptic. The average of the third term in (64) over all ν_B ($0 \le \nu_B \le 2\pi$), therefore, yields the *permanent tidal potential*:

$$\bar{V}_B(r,\psi) = D_B(r)\left[\left(3\sin^2\psi - 1\right)\left(\frac{1}{2}\sin^2\varepsilon - \frac{1}{3}\right)\right]. \tag{68}$$

The sun and moon are the only extraterrestrial bodies of consequence and one may denote the

tidal potential as

$$V_t(r,\psi,\lambda) = V_{sun}(r,\psi,\lambda) + V_{moon}(r,\psi,\lambda) \, . \tag{69}$$

Using nominal parameter values, Doodson's coefficients evaluated at Earth's mean radius, $R = 6731$ km , for the sun and moon are

$$D_{sun}(R) = 1.21 \ m^2/s^2 \, , \quad D_{moon}(R) = 2.63 \ m^2/s^2 \, . \tag{70}$$

According to the model (68), the permanent tidal potential (combined effect of sun and moon),

$$\bar{V}_t(r,\psi) = -\, 0.252 \left(D_{sun}(r) + D_{moon}(r) \right) \left(3\sin^2\psi - 1 \right) , \tag{71}$$

varies in latitude on the Earth's surface from about $-\, 1.95 \ m^2/s^2$ at the poles to $0.98 \ m^2/s^2$ at the equator.

Indirect effects on Earth's gravitational potential arise from the displacement of the terrestrial masses under the influence of the tidal potential. These are changes in the potential due to the ocean tide (that displaces the water level, typically less than 1 m in the vertical, up to several meters near coasts), the Earth tide (several decimeters in the vertical), atmospheric tide, and secondary indirect effects due to loading of the ocean and atmosphere on the Earth's surface. The total gravitational potential thus constitutes the following:

$$\begin{aligned} V = V_e + V_t \\ + V_{\text{solid Earth tide}} + V_{\text{ocean tide}} + V_{\text{atmospheric tide}} \\ + V_{\text{ocean loading}} + V_{\text{atmospheric loading}} + V_{\text{other mass redistributions}} \, , \end{aligned} \tag{72}$$

where it is recognized that some parts are better known than others.

The indirect effect on the gravitational potential, due to the solid Earth tides which are displacements of masses of the quasi-elastic Earth, is modeled by a fraction of the direct effect. It can be shown (Lambeck, 1988, p.254) that the tidal potential including the indirect effect is given by

$$V_t'(r,\psi,\lambda) = \left(1 + k_2 \left(\frac{R}{r} \right)^5 \right) V_t(r,\psi,\lambda) \, , \tag{73}$$

where $k_2 = 0.29$ is Love's number (an empirical number based on observation). Equation (73) assumes that the elastic response to the tidal potential is instantaneous; whereas, in. reality, there is

a lag, which to a first approximation is constant. The permanent tide component including the indirect effect of the Earth tide now is given by

$$\bar{V}_t'(r,\psi) = \left(1 + k_2 \left(\frac{R}{r}\right)^5\right) \bar{V}_t(r,\psi) , \tag{74}$$

where $\bar{V}_t$ is given by (71). It is also called the mean tide potential. However, the Love number, k_2, does not adequately represent the indirect effect in the mean and for geodetic applications it is recommended that no correction for this permanent effect be applied to observed quantities (McCarthy, 1992).

The Earth tides cause both vertical and horizontal displacements of the Earth's crust. The vertical displacement, δh, is modeled as a fraction of the direct tidal potential using Love's number, h_2:

$$\delta h(\psi,\lambda) = h_2 \frac{V_t(R,\psi,\lambda)}{g} , \tag{75}$$

where g is the magnitude of gravity. With $h_2 \approx 0.49$ and considering Doodson's coefficients, (70), the vertical displacement is of the order of 0.2 m. Again, the time-averaged component of this indirect tidal effect is responsible for a permanent crustal deformation, putatively given by

$$\overline{\delta h}(\psi) = h_2 \frac{\bar{V}_t(R,\psi)}{g} . \tag{76}$$

But, again, the Love number, h_2, suitably describes the deformation only for the periodic components. Therefore, geodetic observations should not be corrected for this term. Specifically, we note that the mean crustal deformation and mean Earth-tide potential cannot be observed independently and it is natural to retain these effects within all geodetic observations. On the other hand, periodic tidal effects should be removed from observations that must be combined, or should hold, for different time epochs.

Several interesting and important distinctions should be made here in regard to the tidal effects on the geoid. We first note the fundamental linearized relation between potential and gravity, obtained from (6):

$$W_2 - W_1 = - g \left(n_2 - n_1\right) , \tag{77}$$

where g is the magnitude of gravity, n is the distance along the plumbline, and the indicated differences are assumed to be relatively small (points 1 and 2 are close). If the geoid, as an equipotential surface, is defined solely by its potential, W_0, then a change in the potential, as given by (73), implies that the equipotential surface with the potential value, W_0, has been displaced. This displacement is equivalent to a change in geoid undulation (with respect to some predefined

ellipsoid) that according to (77) is given by

$$\delta N(\psi,\lambda) = n_2 - n_1 = -\frac{1}{g}\left(W_0 - \left(W_0 + V_t'(R,\psi,\lambda)\right)\right) = \left(1 + k_2\right)\frac{V_t(R,\psi,\lambda)}{g}\,, \tag{78}$$

so that the potential of the geoid remains W_0. Using (70) and (71), the permanent tidal effect on the geoid that accounts for the sun and moon is given by

$$\overline{\delta N}(\psi) = -\,0.099\left(1 + k_2\right)\left(3\sin^2\psi - 1\right). \tag{79}$$

If N represents the instantaneous geoid, then the geoid without any tidal effects, that is, the *non-tidal geoid*, is given by

$$N_{nt} = N - \delta N\,. \tag{80}$$

The *mean geoid* is defined as the geoid with all but the mean tidal effects removed:

$$\bar{N} = N - (\delta N - \delta\bar{N})\,. \tag{81}$$

This is the geoid that could be directly observed, for example, using satellite altimetry averaged over time, if mean sea level is corrected for the mean sea surface topography. For theoretical reasons (the geoid determined from a solution to a boundary-value problem in potential theory requires that there are no external masses, such as the sun and moon), it is desirable to define a geoid with all but the mean Earth-tide effect removed. This is called the *zero geoid*, identifying it as the geoid that retains the permanent Earth-tide effect, but no other tidal effects:

$$N_z = N - \left(\delta N + 0.099\,k_2\left(3\sin^2\psi - 1\right)\right). \tag{82}$$

The difference between the mean and zero geoids is, therefore, the permanent component of the direct tidal potential. We note that, in principle, each of the geoids defined above, has the same potential value, W_0, in its own field. That is, with each correction we define a new gravity field and the corresponding geoid undulation defines the equipotential surface in that field with potential value given by W_0. This is fundamentally different than what happens in the case when the geoid is supposed to define a vertical datum.

7. Vertical Datums and the Tidal Potential

If the local geoid is a vertical datum that must be *realized* in terms of a monument on the Earth's surface, then because the crust is displaced in the vertical by (75), the potential of the datum changes not only by the effect (73), but also by the effect of the displacement as given according to (77). Therefore, the potential of the datum at the point of realization changes by

$$\delta W_0^{(j)} = \left(1 + k_2 - h_2\right) V_t(R,\psi_0^{(j)},\lambda_0^{(j)}) \,. \tag{83}$$

Clearly, from geometrical considerations the geoid undulation at the datum origin point changes only by (75). Since datums are supposed to provide a stable reference for heights, we must redefine the vertical datum within the presence of the tidal potential, first of all, as being realized at an origin point on the *mean topographic surface* (or, sometimes called the *mean crust*). The mean topographic surface is defined with respect to the ellipsoid as follows:

$$\bar{h} = h - \left(\delta h - \overline{\delta h}\right), \tag{84}$$

where the terms within the parentheses are given by (75) and (76), and h is the instantaneous ellipsoidal height. We note that the "zero topography" is the same as the mean topography, since there is no "direct effect" on the topography.

Next, we must identify the potential of the datum. To be consistent with the realization of the origin point, it should be the *mean potential* realized on *the mean topography*. This is designated $\overline{W}_0^{(j)}$, and excludes all but the permanent direct and indirect tidal effects, since the latter cannot be determined:

$$\overline{W}_0^{(j)} = W_0^{(j)} - \left(1 + k_2 - h_2\right)\left[V_t\left(R,\psi_0^{(j)},\lambda_0^{(j)}\right) - \overline{V}_t\left(R,\psi_0^{(j)}\right)\right], \tag{85}$$

where $W_0^{(j)}$ is the instantaneous potential at the instantaneous datum origin. The datum thus realized and with the potential, $\overline{W}_0^{(j)}$, is called the *mean datum*. The other types of datums, the zero, non-tidal, and even instantaneous datums can also be defined, but may be arcane or impractical. Indeed, each requires that we first define corresponding origin points that, in principle, should be realizable (otherwise it would not adhere to a proper definition of a datum). A *zero datum* would have the same origin point as the mean datum; but it would exclude the permanent direct tidal potential. Although this is computable and for theoretical purposes should be *excluded* in determinations of the potential from solutions to boundary-value problems, for vertical datum definitions, it seems more appropriate to adhere to physical reality as closely as possible (the mean tidal effects of the sun and moon *are* present). The *non-tidal datum* should refer to the non-tidal topography, but this cannot be realized since we do not know the permanent

crustal deformation. Also, it makes little sense to define an *instantaneous datum* since it would refer to a reference surface that contains significant temporal variations. Therefore, we consider only the mean datum.

The geoid undulation for the mean (and zero) datum at the origin point is defined by $\bar{N}_0^{(j)}$. At other points it is the distance from the ellipsoid to the equipotential surface with.potential given by $\bar{W}_0^{(j)}$ in the "mean field", that is, the potential field with all but the permanent direct and indirect tidal effects removed. However, with regard to height datums, the geoid undulation need not be known. The need to know the datum potential arises only if orthometric (or normal) heights should be determined with satellite methods (see (51) and (52)), or if different datums should be connected. In these cases, we need to develop models analogous to (48), (50), (51), and (52) that include the tidal effects. It is noted that the traditional leveling procedures (which are not instantaneous) may reasonably be expected to yield *mean* orthometric or normal heights, that is heights with all but the mean tidal effects removed. The extent to which tidal effects are not averaged out by the leveling procedure causes systematic errors and vitiates the theoretical ability to determine a height by this method.

We consider again the disturbing potential, T_P, at a fixed point of the Earth's surface. It is affected by the tidal potential including its indirect effect, (73), but it is not sensitive to the vertical displacement. Indeed, the gradient of the disturbing potential is the *gravity disturbance*, $\delta g = -\partial T/\partial h$; and, therefore, the effect of the vertical displacement, (75), is $-h_2 V_t(R,\psi,\lambda)\, \delta g/g$; and, the total tidal effect on the disturbing potential is

$$\begin{aligned} \delta T_P &= \left(1 + k_2 - \frac{\delta g}{g} h_2\right) V_t(R,\psi,\lambda) \\ &\approx \left(1 + k_2\right) V_t(R,\psi,\lambda)\,, \end{aligned} \tag{86}$$

since $\delta g/g = O\left(10^{-5}\right)$. Disturbing potential models may be given in various tide systems; for example, the EGM96 model is given in the non-tidal system; others like TEG are given in the zero-tide system. In fact, it seems more logical to define the geopotential model in the zero-tide system since the permanent indirect effects are not observable and not correctly modeled using the Love number, k_2. Thus, we consider only the *zero disturbing potential* which is the disturbing potential with all but the permanent Earth-tide removed:

$$\overset{0}{T}_P = T_P - \left[V_t(R,\psi,\lambda) + k_2\left(V_t(R,\psi,\lambda) - \bar{V}_t(R,\psi)\right)\right], \tag{87}$$

where T_P is the instantaneous disturbing potential.

Introducing the tidal potential and its indirect effects on the determination of the normal heights, we see from (7), (30), and (83) that

$$\begin{aligned}\delta H_P^{\text{norm(j)}} &\approx \frac{\delta C_P^{(j)}}{\gamma} \\ &= \frac{\left(1 + k_2 - h_2\right)}{\gamma}\left(V_t\left(R,\psi_0^{(j)},\lambda_0^{(j)}\right) - V_t(R,\psi,\lambda)\right).\end{aligned} \tag{88}$$

This assumes, of course, that the normal height is determined instantaneously. The *mean normal height* is given by removing (88) and restoring the permanent components

$$\overline{H}_P^{\text{norm(j)}} = H_P^{\text{norm(j)}} - \frac{\left(1 + k_2 - h_2\right)}{\gamma}\left[V_t\left(R,\psi_0^{(j)},\lambda_0^{(j)}\right) - V_t\left(R,\psi,\lambda\right) - \overline{V}_t\left(R,\psi_0^{(j)}\right) + \overline{V}_t\left(R,\psi\right)\right], \tag{89}$$

where $H_P^{\text{norm(j)}}$ is the instantaneous normal height.

Starting with the instantaneous model (45), repeated here for convenience:

$$W_0^{(j)} = U_0 - \gamma_{Q_0^{(j)}}\left(h_P - H_P^{\text{norm(j)}} - \frac{T_P}{\gamma_Q}\right), \tag{90}$$

we find with (75), (86) and (88) that

$$\begin{aligned}\delta W_0^{(j)} &= -\gamma_Q\left(\delta h_P - \delta H_P^{\text{norm(j)}} - \frac{\delta T_P}{\gamma_Q}\right) \\ &= \left(1 + k_2 - h_2\right) V_t\left(R,\psi_0^{(j)},\lambda_0^{(j)}\right),\end{aligned} \tag{91}$$

which agrees with (83) and implies that the same model holds, as expected since it is linear, whether the quantities are all non-tidal or all instantaneous. However, we wish to determine the potential, $\overline{W}_0^{(j)}$, of the mean datum, supposedly using mean normal heights, instantaneous ellipsoidal heights, and the zero disturbing potential as observations. The instantaneous ellipsoidal heights are determined by GPS (if they were obtained over a longer observation period, then the model must be altered accordingly), and the zero disturbing potential is computed on the basis of a solution to a boundary value problem in potential theory. With (85), (87), and (89), the model (90) becomes

$$\overline{W}_0^{(j)} = U_0 - \gamma_Q\left(h_P - \overline{H}_P^{\text{norm(j)}} - \frac{{}^{0}T_P}{\gamma_Q}\right) + h_2\left[V_t\left(R,\psi,\lambda\right) - \overline{V}_t\left(R,\psi\right)\right] + \overline{V}_t\left(R,\psi\right). \tag{92}$$

The last term is the difference between the mean and zero disturbing potential and the penultimate

term accounts for the difference between the mean and instantaneous ellipsoidal height.

To determine the datum potential without the theoretical difficulty of obtaining a mean normal height (through leveling), one could use the mean normal height at the datum origin, where, by definition it is known (e.g., equal to zero). Once determined, the datum potential can be used to incorporate new mean normal heights into the datum using ellipsoidal heights and the geopotential:

$$\bar{H}_P^{norm(j)} = \left(h_P - \frac{{}^{0}T_P}{\gamma_Q} \right) - \frac{U_0 - \bar{W}_0^{(j)}}{\gamma_Q} - \frac{h_2}{\gamma_Q}\left[V_t(R,\psi,\lambda) - \bar{V}_t(R,\psi) \right] - \frac{1}{\gamma_Q}\,\bar{V}_t(R,\psi)\,. \qquad (93)$$

For comparison, see (52). A similar model can be developed for orthometric heights.

8. Further Investigations and Applications to IGLD

The definition and realization of vertical datums has so far been restricted to the static (or instantaneous) case and to the case associated with the direct tidal potential and its indirect effect due to Earth tides. However, as noted with (72), there are other indirect tidal effects, as well as other temporal variations induced by geodynamic phenomena. For example, Lambeck (1988) gives the potential due to the ocean tides (including the loading effect on the solid Earth) and states that the amplitudes are less than 15% of the solid Earth tidal effect. Similarly, atmospheric variations and their loading on the Earth's surface (including the ocean) need to be considered, as do the polar motions of the figure and spin axes of the Earth with respect to the reference pole (International Reference Pole, IRP). Finally, geodynamic phenomena that create significant vertical crustal motion over longer time intervals include post-glacial rebound, local subsidence, and tectonic uplift. Appropriate models and numerical investigations are required to assess the importance of these effects on the precise realization of vertical datums, primarily for local applications, but also for global unifications.

The International Great Lakes Datum of 1985 (IGLD85) is a vertical datum based on a single origin point (Pointe-au-Père, Rimouski, Québec) where all heights are given as dynamic heights (and converted to orthometric heights to be consistent with the North American Vertical Datum of 1988). An important issue for the Great Lakes community concerns the historical and future water levels of the lakes which are being monitored by several tide gauges along the perimeters of the lakes. These tide gauges measure water level with respect to the IGLD85. Therefore, to assess the long-term fluctuations of the water levels, an accurate assessment of the variation and/or stability of the vertical datum needs to be made. The models developed in Sections 3 and 7 provide the means to determine the datum potential and thus to monitor its time history, provided that accurate dynamic heights, ellipsoidal heights, and disturbing potential values are available.

Ellipsoidal heights can be determined using vertical GPS positioning and normal heights come

from traditional leveling. An alternative is to measure ellipsoidal heights of the lake level with satellite altimeters, which currently are slightly more accurate in the long term than the GPS-derived ellipsoidal heights of ground points. Corresponding normal heights of the lake level can be determined from the tide-gauge stations, whose normal heights presumably are known, by assuming that the lake level, averaged over time, is an equipotential surface (and normal heights would need to be converted to dynamic heights; see (33)). To realize this condition may require a model for the dynamic lake surface topography. The last ingredient, the disturbing potential, is already quite accurate at wavelengths of several hundred kilometers and longer (about the dimensions of the lakes), but will become more accurate by orders of magnitude over the next few years as the GRACE and GOCE satellite missions are realized.

Secular vertical crustal motion complicates the monitoring of lake level changes over long time intervals (decades) and causes heights and their datums to vary in time. For the open ocean,the following model describes the change in sea level, δn_t, with respect to a land-based tide gauge, in terms of a *eustatic* change in sea level, δw (due to a change in water volume, for example, from ice melt), and crustal uplift, δp (for example, due to PGR):

$$\delta n_t = \delta w - \delta p \, . \tag{94}$$

Here the small change in the geoid slope due to the uplift (which then also affects the water level) has been neglected, being two orders of magnitude smaller in value than the PGR in the Great Lakes region (Chao, 1994). The change in water level, δn_e, with respect to a pre-defined ellipsoid (again for the open ocean) is simply equal to the eustatic change in water level:

$$\delta n_e = \delta w \, . \tag{95}$$

Thus, satellite altimetry determines directly the eustatic change, while tide-gauge measurements include a component due to PGR (or other vertical crustal changes).

The situation is reversed for a hypothetical closed lake with neither inflow nor outflow of water. In this case the lake level will rise with the PGR of the land surface and the tide-gauge measurement sees directly the eustatic change in water level, while the altimeter-derived change in water level with respect to the ellipsoid includes both the eustatic change and PGR. To what extent the Great Lakes behave like the open ocean, in which case the models (94) and (95) hold, is a question that remains to be answered. Potentially, both models must be amended to exclude or include, respectively, a fraction of δp to account for more complex boundary conditions. Certainly PGR plays a role if both satellite altimetry and tide-gauge measurements are used to determine the eustatic change in lake levels over the long term.

For corresponding secular changes in heights of terrestrial points and in the datum to which they refer, we note that the total vertical change in the land surface is the sum of a change in the orthometric height and a change in the datum origin:

$$\delta p = \delta H^{(j)} + \delta N_0^{(j)}, \tag{96}$$

where, again, the change in the slope of the datum is neglected. The left-hand side is also the change in ellipsoidal height and can be determined by repeated vertical satellite positioning (which is all that is needed in the models for the eustatic change in lake level). To monitor the stability of the vertical datum, i.e., the change, $\delta N_0^{(j)}$, on the right-hand side of (96), requires repeated vertical satellite positioning of the origin point whose orthometric height by definition does not change. Then the orthometric (or normal) heights at other points in the datum can be monitored by repeated satellite vertical positioning and by using the disturbing potential according to the model (51) or (52). Of course, periodic re-leveling of points in the datum would accomplish the same objective, but is less cost-effective.

9 . Summary

This report reviews the fundamental definitions of heights and vertical datums, specifically motivated by the modern technique of determining heights using accurate satellite vertical positioning in combination with an accurate model for the geopotential. It is shown that the determination of heights in such a manner requires knowledge of the potential value of the vertical datum (as opposed to leveling procedures that do not require this). Furthermore, to determine the potential of a vertical datum ideally requires normal heights (defined at the origin point of the datum, or determined elsewhere by leveling) rather than orthometric heights, as this avoids the complication of assuming a density model for the crust. The models associated with these procedures are also developed within the context of the temporally varying field of the tidal potential, which leads to a more fundamental distinction between a vertical datum (local geoid) and the global geoid. That is, the global geoid, by definition, has always the same potential value but its surface varies (varying geoid undulation); while the vertical datum has an origin changing only because of crustal deformation and the potential varies due to the direct and indirect tidal effects.

The models thus developed also form the basis for monitoring the stability of vertical datums under the influence of geodynamic vertical crustal deformations, such as caused by post-glacial rebound. This has obvious implications in the monitoring of lake levels that are tied to a particular vertical datum. Preliminary models and procedures are indicated.

10. References

Ardalan, A.A. and E.W. Grafarend (1999): The first test for W0-dot, the time variation of W0, based on three GPS campaigns of the Baltic Sea Level Project. Reports of the Finnish Geodetic Institute, 99:4.

Bursa, M. (1995): Geoidal potential free of zero-frequency tidal distortion. Earth, Moon, and Planets, 71, 59-64.

Bursa, M., K. Radej, Z. Sima, S.A. True, and V. Vatrt (1997): Determination of the geopotential scale factor from TOPEX/Poseidon satellite altimetry. *Studia geoph. et geod.*, **41**, 203-216.

Bursa, M., J. Kouba, K. Radej, S.A. True, V. Vatrt, and M. Vojtiskova (1998): Mean Earth's equipotential surface from TOPEX/Poseidon altimetry. *Studia geoph. et geod.*, **42**, 459-466.

Bursa, M., J. Kouba, K. Radej, V. Vatrt, M. Vojtiskova, and S.A. True (1999): Determination of the geopotential at the tide gauge defining the North American Vertical Datum 1988 (NAVD88). *Geomatica*, **53**(3), 291-296.

Chao, B.F. (1994): The geoid and Earth rotation. In *Geoid and Its Geophysical Interpretation*, P. Vanicek and N.T. Christou (eds.), CRC Press, Boca Raton.

Grafarend, E.W. (1994): What is a geoid? In: *Geoid and its Geophysical Interpretation*, P. Vanicek and N.T. Christou (eds.), pp.1-32, CRC Press, Boca Raton.

Grafarend, E.W. and A.A. Ardalan (1997): W0 - an estimate in the Finnish Height Datum N60, epoch 1993.4, from twenty-five GPS points of the Baltic Sea Level Project. *J. Geodesy*, **71**(11), 673-679.

Heck, B. and R. Rummel (1991): Strategies for solving the vertical datum problem using terrestrial and satellite geodetic data. In: Sea Surface Topography and the Geoid, Proceedings of the IAG Symposia, 104 (ed. H. Sunkel, T. Baker), pp.116-127, Springer Verlag.

Heiskanen, W.A. and H. Moritz (1967): *Physical Geodesy*, Freeman and Co., San Francisco.

Lambeck, K. (1988): *Geophysical Geodesy*. Clarendon Press, Oxford.

Lemoine, F.G. et al. (1998): The development of the joint NASA GSFC and the National Imagery Mapping Agency (NIMA) geopotential model EGM96. NASA Technical Report NASA/TP-1998-206861, Goddard Space Flight Center, Greenbelt, Maryland.

Lindsay, D (1996): web site page: http://www.lre.usace.army.mil/IGLD.1985/fig1map.html.

McCarthy, D.D. (1992): IERS Standards (1992). IERS Technical Note 13, Observatoire de Paris, Paris, France.

Moritz, H. (1992): Geodetic Reference System 1980. *Bulletin Géodésique*, **66**(2), 187-192.

NGS (1986): *Geodetic Glossary*. Publication of the National Geodetic Survey, NOAA/NOS, National Geodetic Information Center, Rockville, MD.

NIMA (1997): Department of Defense World Geodetic System 1984. Tech. Report TR-8350.2, third edition, National Imagery and Mapping Agency, Bethesda, MD.

Peltier, W. (1998): Postglacial variations in the level of sea: implications for climate dynamics and solid Earth geophysics. *Rev. Geophys.*, **36**(4), 603-689.

Rapp, R.H., 1987: An estimate of equatorial gravity from terrestrial and satellite data. *Geophys. Res. Lett.*, **14**, 730-732.

Rapp, R.H., 1994: Separation between reference surfaces of selected vertical datums, *Bulletin Geodesique*, **69**, 26-31.

Rapp, R.H., 1995: Equatorial radius estimates from Topex altimeter data. In: Festschrift to E. Groten; M. Becker, G. Hein, and R. Rummel (eds.), Institute of Geodesy and Navigation, Univ. FAF, Munich.

Rapp, R.H., 1997: Use of potential coefficient models for geoid undulation determinations using a spherical harmonic representation of the height anomaly/geoid undulation difference. *J. Geodesy*, **71**(5), 282-289.

Rapp, R.H. and N. Balasubramania (1992): A conceptual formulation of a world height system. Report No.421, Dept. Geodetic Science and Surveying, Ohio State University, Columbus, Ohio.

Sjöberg, L.E. (1999): The IAG approach to the atmospheric geoid correction in Stokes' formula and a new strategy. *Journal of Geodesy*, **73**(7), 362-366.

Simons, M and B.H. Hager (1997): Localization of the gravity field and the signature of glacial rebound. *Nature*, **390**, 500-504, 4 Dec. 1997.

Torge, W. (1991): *Geodesy*. Walter de Gruyter, Berlin.

Zilkoski, D.B., J.H. Richards, and G.M. Young (1992): Results of the general adjustment of the North American Vertical Datum of 1988. *Surveying and Land Information Systems*, **52**(3), 133-149.