The height datum/geodetic datum problem

F. Sansò and G. Venuti

DIIAR, Sez. Rilevamento, Politecnico di Milano, Via Valleggio, 12 I-22100 *Como, Italy. E-mail: iges@ipmtf4.topo.polimi.it*

Accepted 2002 January 8. Received 2002 January 8; in original form 2001 April 11

SUMMARY

The height datum problem is present in the geodetic literature since the times of Pizzetti, when it was realized that as only differences of the gravity potential can be derived from measurements, there was still one global parameter to be settled in order to determine a global model.

Several more realistic formulations of the problem have been introduced into the geodetic literature. After reviewing them, an ultimate formulation is attempted where only local data are given stemming from levelling, gravimetry, classical geodetic network observations, combined with global GPS-like observations. The problem of contemporaneous height datum/geodetic datum determination is shown to be solvable, in ellipsoidal approximation, only if the classical tide gauge and network orientation information is taken into account, specially if the local datum does not refer to very large areas, e.g. of continental size.

In fact we will show that when this classical information is missing for the datum problem referring to a fairly small area, it becomes, for instance, impossible to distinguish between a 'vertical' shift of the local ellipsoid and a change of the potential of the geoid, i.e. the choice of a slightly different gravity equipotential surface.

Key words: geoid, gravity, gravity anomalies, reference frames, satellite geodesy.

1 INTRODUCTION

Let us accept the definition of the geoid G as some specific equipotential surface of the actual gravity potential W of the Earth, $G \equiv \{P \mid W(P) = W_0\}$, such that it is lying, on average, close to the physical surface of the ocean (maximum absolute difference of the order of a few metres) when this is depurated from short periodic and tidal motions.

For the sake of simplicity we also assume a model of a rigid earth uniformly rotating around an axis fixed in space as well as with respect to its body and we stipulate that all the observations are reduced for the solid tidal effects, apart from the constant (the so-called Honkasalo) term.

G can then be univocally identified either by specifying an *a priori* value W_0, such that the stated conditions are satisfied, or by claiming that the equipotential surface at hand is the one that passes through a point $\bar{P}$ physically placed close to the sea surface (e.g. a tide gauge). In this first instance we don't know *a priori* any point of G and a good choice is $W_0 = U_0$, where U_0 is the potential of the normal field $U(P)$, used as our basic approximation of the actual gravity potential, when P lies on its defining ellipsoid.

This in turn can typically be done by fixing first the semimajor axis a, the angular velocity ω and the mass of the earth multiplied by Newton's constant, $\mu = kM$, as it can be very accurately estimated from satellite tracking. From the above quantities one can compute the constant (*cf.* Heiskanen & Moritz 1990)

$$m_0 = \frac{\omega^2 a^3}{\mu}. \tag{1}$$

At this point if one adds the spherical coefficient J_2, again fairly well known from its large effects on satellite dynamics, one has access to the eccentricity of the ellipsoid, $e = \sqrt{\frac{a^2-b^2}{a^2}}$, which can be computed by solving the exact equation

$$J_2 = \frac{e^2}{3}\left(1 - \frac{4}{15} m_0 \frac{e^3}{(3-2e^2)\arctan\frac{e}{\sqrt{1-e^2}} - 3e\sqrt{1-e^2}}\right) \tag{2}$$

stemming from a lengthy computation based on the explicit formulas of the normal field (e.g. *cf.* Heiskanen & Moritz 1990). From eq. (2) one finally has the famous explicit formula (adapted to the above quantities)

$$U_0 = \frac{\mu}{a}\left(\frac{1}{e}\arctan\frac{e}{\sqrt{1-e^2}} + \frac{1}{3}m_0\right), \tag{3}$$

and $W_0 = U_0$ is thus derived form our choice of a, ω, μ, J_2. In this approach the difficulty is that, since we don't know the coordinates of any physical point lying on G, we cannot claim that we are directly able to derive the values of W_P all over the surface, because the quantity physically accessible by measurements is the so-called geopotential number

$$C_P = W_{\bar{P}} - W_P, \tag{4}$$

computed along lines drawn on the Earth's surface, starting from some point $\bar{P}$ used as origin of the height system i.e. defining the specific height datum. To some extent equivalent to (4) would be giving the orthometric heights H_P related to C_P by the well-known formula

$$H_P = \frac{C_P}{\bar{g}} \tag{5}$$

where $\bar{g}$ is the mean value of the gravity along the plumbline from the surface point P down to the geoid (*cf.* Heiskanen & Moritz 1990, formula 4.48). So in this case we would be left with the need of estimating the difference

$$\delta W = W_0 - W_{\bar{P}}$$

from some additional information, what is nowadays creating a possible confusion in the geodetic literature. For that reason we prefer to go along the second approach and define the geoid as the equipotential surface through a given point $\bar{P}$ which is also the origin of the height system. This is indeed equivalent to the previous approach and presupposes that now $W_0 = W_{\bar{P}} \neq U_0$. Accordingly W_0 becomes one of the parameters that at the end of our process has to be estimated. The good point in this approach is that it forces the geodesists to think more in relative terms, realizing that when they are computing a high resolution geoid they are in fact determining a piece of one of a family of equipotential surfaces which are all lying close to the sea surface, in a distance of a few metres from one to the other. It is also important to state that all these surfaces are practically parallel to one another, since the distance between them is given by Brun's relation

$$d = \frac{\delta W}{\gamma} \tag{6}$$

so that

$$\text{Max}\left|\frac{\delta d}{d}\right| = \text{Max}\left|\frac{\delta\gamma}{\gamma}\right| \sim 5 \cdot 10^{-3},$$

i.e. for two surfaces in a distance of ∼2 m, the maximum distance variation (from pole to equator) is 1 cm.

Once G has been somehow defined we must have a suitable mathematical representation for it and a sound mathematical theory capable of retrieving this surface, which we use e.g. as reference for an orthometric height system, by some feasible numerical calculation relating it to the realistically achievable observations.

The description of G is done by the so-called geoid undulation N and we have to stress here that this is a function of both the point P on the geoid and the reference ellipsoid E with respect to which we compute it; N is in fact the height of P on E reckoned along the line orthogonal to E, passing through P.

It seems more or less one of the ordinary miracles of geodesy that the problem of determining the height datum and the ellipsoid E from surface data (of any kind) can be split into two parts when we reason in the so-called spherical approximation.

Nevertheless this approach is not anymore sufficient when we come to the centimetre accuracy which is nowadays possible for both the point positioning in space by GPS and other space techniques and gravimetric geoid computations. In this paper we try to clarify how these two facts are intermingled at an ellipsoidal approximation level, also taking into account that we have to face (and solve) this problem every time we compute a high resolution geoid and not only for global models, as it was devised in earlier formulations. We shall conduct our derivations by ordinary approximation techniques trying to keep an accuracy of at least 1 cm in positions and at least 0.03 mGal in gravity. This defines what we mean by approximation throughout the paper. To simplify the reading of the paper all symbols used are listed in Table 1.

Table 1. List of symbols used in the paper.

$W(P)$	gravity potential at P
$g(P) = \lvert\nabla W(P)\rvert$	modulus of gravity vector
$U(P)$	normal field (Somigliana-Pizzetti)
$\gamma(P)$	modulus of normal gravity vector at point P in space
$\gamma_e(Q)$	the same computed on the reference ellipsoid, at Q
U_0	normal potential on the reference ellipsoid
$T(P) = W(P) - U(P)$	anomalous potential
$\bar{P}$	origin of the height datum
G	geoid i.e. equipotential surface through $\bar{P}$
H	orthometric height, with respect to G
h	ellipsoidal height
N	geoid undulation (vertical distance between G and the given reference ellipsoid)
ζ	height anomaly
Δg	free air gravity anomaly
$S(\psi)$	Stokes' function
$\delta W = U_0 - W(\bar{P})$	height datum parameter
$\underline{\nu}$	ellipsoidal normal = $[\cos\varphi\cos\lambda, \cos\varphi\sin\lambda, \sin\varphi]^+$
$\underline{e}_\varphi$	southward unit vector $= -\frac{\partial\underline{\nu}}{\partial\varphi}$
$\underline{e}_\lambda$	eastward unit vector $\frac{1}{\cos\varphi}\frac{\partial\underline{\nu}}{\partial\lambda}$

To conclude this section, let us claim that since following Molodensky's line of thought one can always write (*cf.* Heiskanen & Moritz 1990§8.2)

$$N - \zeta \cong -10^{-6}\Delta g_B \cdot \tilde{H} \tag{7}$$

(N = geoid undulation in m, ζ = height anomaly in m, Δg_B = Bouguer gravity anomaly in mGal, $\tilde{H}$ = approximate orthometric height in m); we consider the determination of N as something derived from the determination of the height anomaly ζ by solving a suitable boundary value problem.

2 THE HEIGHT DATUM PROBLEM: EVOLUTION OF THE CONCEPT AND ITS PRESENT-DAY UNDERSTANDING

It seems to the authors that the evolution of the definition of the height datum problem can be characterized by the following steps:

(a) global, single BVP, spherical approximation, single height datum problem,

(b) global, single BVP, spherical approximation, multiple height datum problem,

(c) global, single BVP, ellipsoidal approximation, multiple height datum-geodetic spherical approximation,

(d) global mixed BVP, spherical approximation, multiple height datum problem,

(e) local single BVP, spherical approximation, multiple height datum problem, and finally what we shall present here could be defined as,

(f) local, single BVP, ellipsoidal approximation, dual height datum, geodetic datum problem.

Let us quickly analyse the definition of these different problems: (a) goes back to Pizzetti, according to (Heiskanen & Moritz 1990,

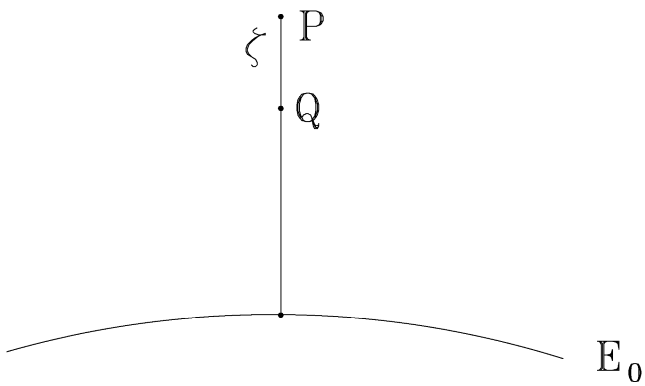

Figure 1. E_0 is the reference ellipsoid, P the point where we want the height anomaly ζ, Q the point on the same normal defined by eqs (8) and (9). While P is sweeping the earth's surface, Q describes the so-called telluroid.

§2.19). It stems from the fact that in the linearization of the basic BVP of physical geodesy we can define the telluroid point Q only through the equation

$$U(Q) = \tilde{W}(P) \tag{8}$$

where

$$\tilde{W}(P) = U_0 - C(P) = U_0 - [W(\bar{P}) - W(P)] = W(P) - \delta W \tag{9}$$

is the only quantity, as opposed to $W(P)$, which can be derived from real observations.

Note that (*cf.* Fig. 1) P and Q have to lie on the same normal to the reference ellipsoid E_0.

Linearization of eq. (9) leads to the more general relation of Brun's type

$$h_P - h_Q = \zeta = \frac{T - \delta W}{\gamma}. \tag{10}$$

Using this in the linearized equation

$$\Delta g = g(P) - \gamma(Q) \cong -\frac{\partial T}{\partial h} + \frac{\partial \gamma}{\partial h}\zeta$$

we get, in spherical approximation, the boundary relation

$$-\frac{\partial T}{\partial r} - \frac{2}{r}T = \Delta g - \frac{2\delta W}{r}, \tag{11}$$

which has indeed to be considered as a BVP for the Laplace equation, since $\Delta T = 0$ in the outer space.

Let us apply eq. (11) to a sphere of radius R. We recall that for any harmonic function $u(P)$ developed in a series of spherical harmonics converging on and outside a sphere of radius R, the first harmonic coefficient

$$u_{00} = \frac{1}{4\pi}\int u(R,\sigma)\,d\sigma$$

coincides also with $\mu = kM$, where M is the total mass (internal to the sphere) generating k. So if we stipulate that the normal potential U has the same mass (i.e. the same μ value) as W, we find that the anomalous potential T is expected to have a vanishing coefficient T_{00}. We see then from eq. (11) that one has

$$\delta W = \frac{R}{2}\Delta g_0 = \frac{R}{8\pi}\int \Delta g\,d\sigma, \tag{12}$$

which can be arranged into the classical Stokes formula to supply (*cf.* Heiskanen & Moritz 1990)

$$N = \frac{R}{\gamma \cdot 4\pi}\int \Delta g[S(\psi) - 1/2]\,d\sigma. \tag{13}$$

At that time the possibility of a true global computation of the geoid was still in the realm of theory, yet it is interesting to observe that in principle no additive information was needed to solve the problem, which might appear curious since we have suppressed one information, namely W_0, from the given data. Nevertheless we must realize that the additional information substituting W_0 is $\mu = kM$ and it appears into the problem when we claim that subsequently one has $T_{00} = 0$, which is in fact the relation determining δW.

The approach in Item (b) goes back to the late seventies/early eighties and authors such as Colombo (1980); Rummel & Teunissen (1981); Heck (1989); Xu (1990); Xu & Rummel (1991); Rapp & Balasubramania (1992) and Balasubramania (1994) among others have worked on the problem. Common to all these authors is the idea that the modification (as in eq. 11) of the standard geodetic BVP should be applied to as many areas $\{A_i\}\, i = 1, 2, \ldots, n$, as there are origins $\{\bar{P}_i\}$ of different height datums. When $\cup_i A_i$ covers the whole earth surface and by using the characteristic functions

$$\chi_{A_i}(P) = \begin{cases} 1 & P \in A_i \\ 0 & P \notin A_i \end{cases}$$

one can then write, in spherical approximation,

$$\begin{cases} \Delta T = 0 \\ -\dfrac{\partial T}{\partial r} - \dfrac{2}{r}T\Big|_S = \Delta g - \frac{2}{r}\sum \delta W_i \chi_{A_i}(P) \end{cases}, \tag{14}$$

with (*cf.* eq. 9)

$$\delta W_i = W(\bar{P}_i) - U_0.$$

Working again at the level of Stokes approximation and on a sphere of radius R, we can write the solution of eq. (14) as

$$T(P) = \tilde{T}(P) - \frac{2}{R}\sum_i \delta W_i S_{A_i}(P) \tag{15}$$

where

$$\tilde{T}(P) = \frac{1}{4\pi}\int S(\psi_{PQ})\Delta g(Q)\,d\sigma_Q$$

$$S_{A_i}(P) = \frac{1}{4\pi}\int_{A_i} S(\psi_{PQ})\,d\sigma_Q,$$

let us note here explicitly that indeed $S_{A_i}(P) \neq 0$ even when $P \notin A_i$.

Now assume that in each area A_i we have a number of (possibly permanent) stations of space geodesy, e.g. permanent GPS stations,

$$P_{is} \in A_i \quad s = 1, \ldots, n_i, \quad i = 1, \ldots, n. \tag{16}$$

All the space geodetic observations can be analysed together, as the International GPS Service does for the international GPS network, providing geometric positions of the stations in a unified geodetic datum. Therefore, in particular, at these points we know the ellipsoidal heights in a consistent geocentric geodetic datum. At the same points we know as well the approximate heights $h(Q_{is})$, derived from $\tilde{W}(P_{is}) = U(Q_{is})$.

In addition, as in eq. (10), the relation

$$h(P_{is}) - h(Q_{is}) = \zeta(Q_{is}) = \frac{1}{\gamma}T(Q_{is}) - \frac{1}{\gamma}\delta W_i \tag{17}$$

has to hold.

Let us note that what is done here in terms of $h_P = h_Q + \zeta_P$ could be repeated with the other decomposition $h_P = N_P + H_P$.

Substituting eq. (15) into (17) yields a system of observation equations

$$h(P_{is}) - h(Q_{is}) = \frac{1}{\gamma}\tilde{T}(P_{is}) - \frac{2}{R\gamma}\sum_k \delta W_k S_{A_k}(P_{is}) - \frac{1}{\gamma}\delta W_i \tag{18}$$

which are typically more than the unknowns $\{\delta W_i\}$, if $n_i > 1$.

One is therefore led to apply a least-squares procedure to estimate δW_i, which has been done in a number of simulated studies. On purely theoretical grounds it seems interesting to observe that if both types of data $h(P_{is})$ and the gravity anomaly field Δg are affected by some noise, then the correct estimation tool should include a re-estimate of T too, falling in the framework of overdetermined BVP's. This has been considered in Migliaccio *et al.* (1989). It is worth nothing that in addition to the observation eq. (18) one should also add the condition ($S(A)$ = surface measure of A)

$$T_{00} = 0 \Rightarrow \Delta g_{00} = \frac{2}{R}\sum_i \delta W_i S(A_i), \tag{19}$$

which has to hold if $\mu = kM$ is assumed to be known.

Additionally, when considering a unified height datum over large areas (e.g. North America or Europe) one should be aware that probably significant distortions have been introduced to connect partial levelling networks.

In the approach in Item (c), Sansò and Usai (1995) have reformulated the problem globally as in (b), but adding the effects due to the change of geodetic datum (reference ellipsoid) when we consider different areas. Since we shall return to this item at the end of the paragraph, we won't dwell on it here.

The approach in (d), presented in Lehman (2000), is the only one, for the moment, taking into account that typically geodetic data on land and ocean differ significantly leading, for the globe, to the formulation of mixed BVP's, the so-called Altimetry Gravimetry problems. The analysis is more directed to study the uniqueness of the solution of the modified BVP's so we shall not go into details here apart from underlining that in the future, considering the global problem in the form of mixed BVP is mandatory if we want to be realistic.

The local approach to the height datum determination, (e), has been proposed by Milbert (1995) and Forsberg (2000). This approach is undoubtedly appealing because it is reducing the problem to the use of a realistic data set and also because it exploits the full power of modern approaches to the geoid estimation, like the so-called Least Squares Collocation Theory (*cf.* Heiskanen & Moritz 1990), i.e. the application of the principle of minimizing a mean square estimation error in a class of estimators linear in the observations, represented as functionals of the unknown anomalous potential T, which in turn is considered as a random field endowed with an isotropic covariance function.

Let us accept that a statistical approach like collocation can provide a reasonable local solution of the BVP depending on the available data in many areas, A_i, generally not covering the whole Earth surface, one can use the observation equations of gravity anomalies, localized at points P_{ij},

$$\Delta g(P_{ij}) = \left(-\frac{\partial T}{\partial r} - \frac{2}{r}T\right)_{P_{ij}} + \frac{2}{r}\delta W_i, \quad P_{ij} \in A_i \tag{20}$$

as well as the 'vertical' observations from GPS

$$\zeta(P_{is}) = \frac{1}{\gamma}T(P_{is}) - \frac{1}{\gamma}\delta W_i, \quad P_{is} \in A_i. \tag{21}$$

These equations can then be treated in the typical collocation approach with T, corrected by a known global model for long wavelengths and by the residual terrain correction to account for short wavelengths, as an unknown random field and δW_i as unknown parameters. This approach has to be pursued even if data are given in a single area, if, afterwards, we want to be able to compute the transformation between the local geodetic datum and a geocentric datum. Otherwise our height anomalies will all be known up to an arbitrary (almost) constant bias.

We come now to a precise definition of the problem we want to treat in this paper (Item f). For the sake of simplicity we shall treat it with only two geodetic datums, one geocentric, based on the ellipsoid E_0, one local, based on an ellipsoid E_L rototranslated with respect to E_0. Hereafter we shall use the indexes 0 and L for quantities referring to E_0 and E_L respectively.

The data, given in a local area A, are:

(i) Gravimetric/leveling data

$$\begin{cases} \tilde{W}(P) = U_0 - C(P) \quad \forall P \in A \\ g(P), \end{cases} \tag{22}$$

where gravimetry directly supplies $g(P)$ and, combined with levelling, $C(P)$; to this we add the knowledge of a normal field $U(P)$ with the exact value of μ, of a global (truncated) model $T_M(P)$ and a digital terrain model, enabling us to compute a residual terrain potential T_{RTC} and its functionals.

(ii) Classical geodetic networks data

$$\{\varphi_L(P), \lambda_L(P)\} \tag{23}$$

propagated to any point P in A, as it is usually done in mapping practice, and given in the local geodetic datum.

(iii) GPS data, which can be collected for a number of stations $P_s \in A$ in the vectors

$$\underline{r}_0(P_s) \Leftrightarrow \varphi_0(P_s), \lambda_0(P_s), h_0(P_s), \tag{24}$$

we assume that the origin of the local height system, $\bar{P}$, is also among the P_s, so that in $\bar{P}$ we know

$$\underline{r}_0(\bar{P}) \tag{25}$$

as well as

$$\underline{r}_L(\bar{P}) \tag{26}$$

as it comes from the fact that by definition

$$\zeta_L(\bar{P}) = 0. \tag{27}$$

In this equation we are implicitly assuming that $\bar{P}$ is a tide gauge point and that the local datum E_L is chosen so as to pass through $\bar{P}$. Our purpose is to find the transformation between the two geodetic datums as well as $T(P)$ and δW, what allows us to compute $\forall P \in A$

$$\zeta_L(P) = \frac{T_L(P) - \delta W}{\gamma}, \tag{28}$$

and then $\underline{r}_L(P)$ and $\underline{r}_0(P)$ too.

In addition transforming ζ_L to ζ_0 and then this to N_0, through eq. (7), we can for any new GPS point derive the corresponding orthometric height (GPS levelling) from

$$H_P = h_{0P} - N_{0P}, \tag{29}$$

the height H_P will then be referred to the height datum with origin $\bar{P}$.

Table 2. Approximate formulae for the normal field used through the paper; relative accuracy in the range 10^{-5}, 10^{-6}.

$U(\varphi, h)$	$=$	$U_0 - \gamma_e(\varphi)h - 1/2\gamma_e'(\varphi)h^2$
$\gamma(\varphi, h)$	$=$	$\gamma_e(\varphi) + \gamma_e'(\varphi)h + 1/2\gamma_e''(\varphi)h^2$
$\gamma_e(\varphi)$	$=$	$\Gamma_0(1 + 5,30244 \cdot 10^{-3}\sin^2\varphi)$
$\gamma_e'(\varphi)$	$=$	$-\Gamma_0'(1 - 1,457 \cdot 10^{-3}\sin^2\varphi)$
$\gamma_e''(\varphi)$	$=$	$\Gamma_0''\sin 2\varphi$
$\underline{\gamma}(\varphi, h)$	$=$	$-\gamma(\varphi, h)\underline{\nu} - H(\varphi)\underline{e}_\varphi$
U_0	$=$	$62,636878 \cdot 10^6\ \mathrm{m^2\,s^{-2}}$
Γ_0	$=$	$0,9780327 \cdot 10^6$ mGal
Γ_0'	$=$	$0,30877 \cdot \mathrm{mGal\,m^{-1}}$
Γ_0''	$=$	$72 \cdot 10^{-9}\ \mathrm{mGal\,m^{-2}}$
$H(\varphi)$	$=$	$0,814 \cdot 10^{-3}\sin 2\varphi\ \mathrm{mGal\,m^{-1}}$

3 THE SOLUTION

We shall propose the solution of the height datum/geodetic datum problem in 10 steps. To be more specific in the rest of the paper we shall adopt the approximate formulae in Table 2, which are well suited for our purposes, especially to compute orders of magnitude.

3.1 Step 1: Transformation between E_0 and E_L

By $\underline{r}_0$ we mean the algebraic position vector, $\underline{r}_0 = [x_0, y_0, z_0]^+$, in terms of the Cartesian triad attached to the ellipsoid E_0, with Z_0 along the symmetry axis of E_0. An analogous notation holds for $\underline{r}_L$. The small rototranslation between E_L and E_0 can then be expressed by

$$\underline{r}_0 = \underline{r}_L + \underline{t} + \underline{\varepsilon} \wedge \underline{r}_L \tag{30}$$

where $\underline{t}$ is the position of the centre of E_L in terms of E_0 coordinates, while $\underline{\varepsilon}$ is the infinitesimal rotation vector. The inverse of eq. (30) is

$$\underline{r}_L = \underline{r}_0 - \underline{t} - \underline{\varepsilon} \wedge \underline{r}_0. \tag{31}$$

To be more specific we shall assume that $|\underline{t}|$ is at most of the order of 10^2 m and $\underline{\varepsilon}$ is also such that $R|\underline{\varepsilon}| = 0(10^2\ \mathrm{m})$ with $R \equiv 6.300$ km (i.e. $|\underline{\varepsilon}|$ is of the order of a few arcseconds).

3.2 Step 2: Anomalous potential

Let us first recall that by $U(\underline{r})$ we mean the normal potential computed according to

$$\underline{r} \to (\varphi, \lambda, h) \to U(\varphi, h), \tag{32}$$

which is a function analytically well specified. Accordingly, when we have to express a normal potential attached to one of the two ellipsoids we have only to change the argument $U(\underline{r}_0)$, $U(\underline{r}_L)$ because U is always the same function.

For instance if $\underline{r}_{0P}, \underline{r}_{LP}$ refer to the same point P in space, by using eq. (30) we have to the first order,

$$U(\underline{r}_0) - U(\underline{r}_L) = \underline{\gamma}(\underline{r}_L) \cdot (\underline{r}_0 - \underline{r}_L) = \underline{\gamma}(\underline{r}_L) \cdot (\underline{t} + \underline{\varepsilon} \wedge \underline{r}_L). \tag{33}$$

Now, recalling that

$$T_0(P) = W(P) - U(\underline{r}_{0P}) \quad \text{and} \quad T_L(P) = W(P) - U(\underline{r}_{LP}),$$

we get immediately

$$T_L(P) - T_0(P) = \gamma \cdot (\underline{t} + \underline{\varepsilon} \wedge \underline{r}). \tag{34}$$

Let us explicitly remark here that, as customary in geodesy, we can neglect the index $0, L$ for quantities which are already multiplied by infinitesimals of the first-order, like $\underline{r}$ in (35).

3.3 Step 3: Telluroids

A telluroid is an approximate surface used for the linearization of the relevant geodetic BVP; it has to be computed with known data. In our case, corresponding to a foot-point on the ellipsoid with coordinates (φ, λ), we have a height $\tilde{h}$ determined by solving the equation

$$\tilde{W}(P) = U(\varphi_P, \tilde{h}). \tag{35}$$

Here $\tilde{W}(P)$ refers to the approximate true potential which is a known function according to eq. (22).

In a pure spherical approximation, eq. (35) becomes

$$\tilde{h}_P = \frac{\mu}{\tilde{W}(P)} - R \quad (R = \text{radius of the reference sphere}) \tag{36}$$

showing explicitly that $\tilde{h}_P$ is datum independent. In the case eq. (35) a simple differentiation shows that a change of datum would give

$$\delta h = \frac{1}{\gamma}\frac{\partial U}{\partial \varphi}\delta\varphi$$

and, for $\delta\varphi$ of the order of few arcseconds and an altitude $h \sim 2000$ m, this, (*cf.* Table 2), is much less than 1 mm and therefore negligible.

Accordingly we can write an explicit approximate formula for $\tilde{h}(P)$ (*cf.* Table 2)

$$\tilde{h}(P) = \frac{U_0 - \tilde{W}(P)}{\gamma_e(\varphi_P)} - \frac{1}{2}\frac{\gamma_e'(\varphi_P)}{\gamma_e(\varphi_P)}\left(\frac{U_0 - \tilde{W}(P)}{\gamma_e(\varphi_P)}\right)^2 \tag{37}$$

which is, for any practical purpose, datum independent. So the same function $\tilde{h}(\varphi, \lambda)$, given by eq. (37), can be used to construct a telluroid $\tilde{S}_L$ attached to E_L or the telluroid $\tilde{S}_0$ attached to E_0, i.e. the two telluroids do not coincide.

3.4 Step 4: Height anomalies

Recalling eqs (10) and (34), we can write

$$\begin{aligned} h_0(P) - h_L(P) &= \tilde{h}(P) + \zeta_0(P) - \tilde{h}(P) \pm \zeta_L(P) = \zeta_0 - \zeta_L \\ &= \frac{T_0 - \delta W}{\gamma} - \frac{T_L - \delta W}{\gamma} = \frac{1}{\gamma}(T_0 - T_L) \\ &= -\frac{\underline{\gamma}}{\gamma} \cdot (\underline{t} + \underline{\varepsilon} \wedge \underline{r}) = \underline{\nu} \cdot (\underline{t} + \underline{\varepsilon} \wedge \underline{r}). \end{aligned} \tag{38}$$

3.5 Step 5: Linearized gravity observation equations

We can use eqs (10) and (11) to write

$$\begin{cases} \Delta g_L = -\dfrac{\partial T_L}{\partial h} + \dfrac{\gamma'}{\gamma}T_L - \dfrac{\gamma'}{\gamma}\delta W = BT_L - \dfrac{\gamma'}{\gamma}\delta W \\ \Delta g_0 = -\dfrac{\partial T_0}{\partial h} + \dfrac{\gamma'}{\gamma}T_0 - \dfrac{\gamma'}{\gamma}\delta W = BT_0 - \dfrac{\gamma'}{\gamma}\delta W \end{cases} \tag{39}$$

which yields

$$\Delta g_L - \Delta g_0 = B(T_L - T_0) = B[\underline{\gamma} \cdot (\underline{t} + \underline{\varepsilon} \wedge \underline{r})]. \tag{40}$$

An easy but lengthy computation and a judicious simplification of eq. (40) gives

$$\Delta g_L - \Delta g_0 = H(\underline{t} \cdot \underline{e}_\varphi + \Re\underline{\varepsilon} \cdot \underline{e}_\lambda) \tag{41}$$

where $H(\varphi)$ is as in Table 1, $\underline{e}_\varphi, \underline{e}_\lambda$ are defined in Table 1 and

$$\Re = a\{1 - e^2\sin^2\varphi\}^{-1/2}.$$

Though small, the term (41) cannot be neglected since it can amount to ~10 cm for translations $\underline{t}$ and rotations $\Re\underline{\varepsilon}$ up to 100 m.

3.6 Step 6: The BVP

Let us start here by observing that T_0, T_L are both harmonic functions. For the geocentric T_0 this is standard; on the contrary, for T_L, one has to observe that the two centrifugal potentials of $W(P)$ and $U(\underline{r}_L)$ do not cancel but give rise to a term (considering only the translation)

$$\frac{1}{2}\omega^2\left(X_0^2+Y_0^2\right)-\frac{1}{2}\omega^2\left[(X_0-t_X)^2+(Y_0-t_Y)^2\right]$$
$$=\omega^2\left(X_0t_X+Y_0t_Y\right)-\frac{1}{2}\omega^2\left(t_X^2+t_Y^2\right)$$

which is linear in X_0, Y_0 and therefore harmonic. Nevertheless, for the same reason, we expect T_L not to be regular at infinity.

It follows that if we want to write and use the BVP theory for an anomalous potential, it is safer to do it for T_0 rather than for T_L, though the final approximations implied are known to be very small.

Exploiting eqs (39) and (40) we have

$$\begin{cases}\Delta T_0=0\\ BT_0=\Delta g_L-H(\underline{t}\cdot\underline{e}_\varphi+\Re\underline{\varepsilon}\cdot\underline{e}_\lambda)+\dfrac{\gamma'}{\gamma}\delta W.\end{cases}\tag{42}$$

For clarity, we should emphasize that while the Laplace equation in (42) has to hold in the exterior of the telluroid $\tilde{S}_0$ (see Step 3), the boundary condition can hold only on that part of $\tilde{S}_0$ which corresponds to the data area A.

So strictly speaking eq. (42) is not a BVP and in fact it is a problem with a non-unique solution. Nevertheless the knowledge of a global model T_M and of a residual height model of the terrain, with the corresponding residual potential correction T_{RTC}, can help in finding a very acceptable solution by a so-called remove-restore concept (*cf.* Moritz 1980). The idea is that by subtracting from Δg the terms corresponding to very long wavelengths (calculated from T_M) and those with very short wavelengths (calculated from T_{RTC}) one is left with a fairly smooth field in the area A, displaying very little correlation with the field in distant areas. At this point one can apply any solution method, from the use of Stokes' function, to the use of a stochastic approximation technique like collocation (*cf.* Moritz 1980) to compute the relevant residual potential and then add back to it the contribution of T_M and T_{RTC}.

What is relevant here is that to a kind of BVP

$$\begin{cases}\Delta u=0\\ Bu=f\quad\text{on}\quad A,\end{cases}$$

we attach a linear operator S such that

$$u=S\{\chi_A f\}$$

where $\chi_A(P)=1$ on A and $\chi_A(P)=0$ elsewhere.

If we now use this operator S for the problem (42), exploiting its linearity one gets

$$\begin{aligned}T_0&=S\{\chi_A\Delta g_L\}-S\{\chi_A H\underline{t}\cdot\underline{e}_\varphi\}\pm S\{\chi_A H\Re\underline{\varepsilon}\cdot\underline{e}_\lambda\}\\&\quad+S\left\{\chi_A\frac{\gamma'}{\gamma}\delta w\right\}=\tilde{T}_0-\underline{t}\cdot\underline{\eta}_1-\underline{\varepsilon}\cdot\underline{\eta}_2+\delta Wk\end{aligned}\tag{43}$$

where we have put

$$\begin{aligned}&\tilde{T}_0=S\{\Delta g_L\},\quad k=S\left\{\frac{\gamma'}{\gamma}\chi_A\right\}\\&\underline{\eta}_1=S\{H\chi_A\underline{e}_\varphi\},\quad \underline{\eta}_2=S\{H\Re\chi_A\underline{e}_\lambda\}.\end{aligned}\tag{44}$$

One has to stress that $\tilde{T}_0$ is exactly the local solution computed by ignoring any problem of height or geodetic datum and using only the gravity anomaly Δg_L obtained from the normal potential formula and local coordinates.

Moreover $\underline{\eta}_1$, $\underline{\eta}_2$, k are very smooth functions of the computation point P, so that the connection terms in (42) appear mostly as a linear trend on A.

3.7 Step 7: Space (GPS) observations; the planimetric part

As we said in Section 2, geodetic space observations, in particular in GPS observations which are the most frequent, are able to provide us with the geometric positions of points S with respect to a geometric ellipsoid E_0. The position of P is collected in the vector $\underline{r}_0$ which we consider as the outcome of the space observation.

Let us use the decomposition (Fig. 2)

$$\underline{r}_P=\underline{r}_Q+h\underline{\nu}_P\tag{45}$$

for both datums E_0 and E_L.

We have then

$$\begin{aligned}\underline{r}_{0P}-\underline{r}_{LP}&=\underline{r}_{0Q}-\underline{r}_{LQ}+h_{0P}(\underline{\nu}_{0P}-\underline{\nu}_{LP})+(h_{0P}-h_{LP})\underline{\nu}_{LP}\\&=\underline{t}+\underline{\varepsilon}\wedge\underline{r}_P\end{aligned}\tag{46}$$

Recalling eq. (38) and defining the vertical projector P_ν

$$P_\nu\underline{a}=(\underline{a}\cdot\underline{\nu})\underline{\nu},\tag{47}$$

we can rearrange eq. (46) as

$$(I-P_{\nu i})\left(\underline{t}+\underline{\varepsilon}\wedge\underline{r}_{P_i}\right)=\left(\underline{r}_{0Q_i}-\underline{r}_{LQ_i}\right)+h_{0P_i}\left(\underline{\nu}_{0P_i}-\underline{\nu}_{LP_i}\right).\tag{48}$$

We can observe that this is an observation equation for the tangential component of $\underline{t}+\underline{\varepsilon}\wedge\underline{r}$ at the GPS *type* station P_i, because the second member is fully known from observations, according to the scheme

GPS $\rightarrow\varphi_{0i},\lambda_{0i},h_{0i}\rightarrow\underline{r}_{0Q_i},h_{0P_i},\underline{\nu}_{0P_i}$

Classical $\rightarrow\varphi_{Li},\lambda_{Li}\rightarrow\underline{r}_{LQ_i},\nu_{LP_i}$.

3.8 Step 8: Space (GPS) observations of levelling type

From GPS observations we indeed know h_{0P_i} and at the same time from $\tilde{W}(P_i)$ we know $\tilde{h}_{P_i}$. Then we know $\zeta_{0P_i}=h_{0P_i}-\tilde{h}_{P_i}$ too, and we can write (*cf.* eq. 10)

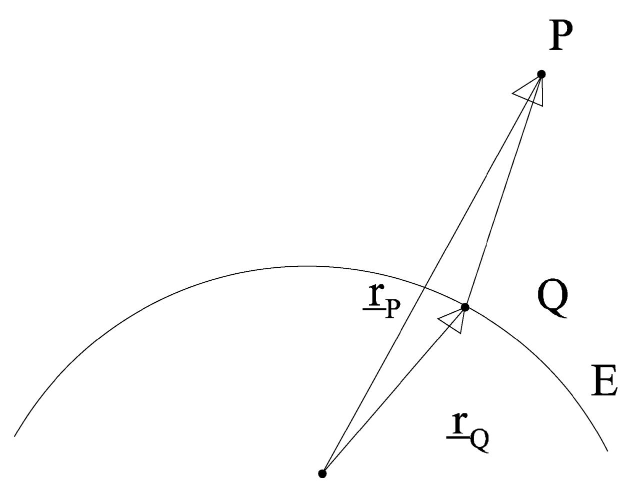

Figure 2. Decomposition of the position vector of P according to the geocentric datum.

$$\zeta_{0P_i} = \frac{1}{\gamma} T_0(P_i) - \frac{1}{\gamma}\delta W. \tag{49}$$

By using eq. (43) in eq. (49) we receive

$$\zeta_{0P_i} = \frac{1}{\gamma}\tilde{T}_0(P_i) - \frac{1}{\gamma}\underline{t}\cdot\underline{\eta}_{1i} - \frac{1}{\gamma}\underline{\varepsilon}\cdot\underline{\eta}_{2i} + \frac{\delta W}{\gamma}[k_i - 1] \tag{50}$$

Note: before continuing we have to observe that for a rough understanding of the results obtained up to now, we can think in terms of spherical approximation and see that eq. (48) has basically a geometric content, to be used for the determination of $\underline{t}, \underline{\varepsilon}$, while eq. (50) has more a physical content to be used for the determination of the height datum shift $(1/\gamma)\delta W$. As a matter of fact in spherical approximation $\underline{\eta}_1 = \underline{\eta}_2 = 0$ so that (50) depends on δW only.

Now (48) allows a reasonable determination of both $\underline{t}, \underline{\varepsilon}$ only when the space stations (connected in a unique datum) are fairly apart from one another so that the horizontal projections $(I - P_{\nu_i})$ are referred to planes with a wide spread of attitudes. If this is not the case, i.e. when the area A is fairly small, we immediately realize that the vertical component $\underline{t}\cdot\underline{\nu}$ can only be very poorly determined. In addition if $\underline{r}_{P_i}$ are more or less parallel to some average direction $\bar{\underline{\nu}}$, the component of $\underline{\varepsilon}$ along $\bar{\underline{\nu}}$ will not be adequately determined too, because the contribution of $(\underline{\varepsilon}\cdot\bar{\underline{\nu}})\bar{\underline{\nu}} \wedge \underline{r}_{P_i}$ will in general be too small.

These two small area effects however do not have a quite symmetrical impact on our problem; in fact $\underline{t}\cdot\underline{\nu}$ is really needed only if we want to determine the full geometric transformation, specially the height part as we can see from eq. (38). On the other hand, if A is small even if we do not determine $(\underline{\varepsilon}\cdot\bar{\underline{\nu}})$ this has very little impact in the geometric transformation

$$\underline{r}_0 = \underline{r}_L + \underline{\varepsilon}\wedge\underline{r}_L$$

as far as $\underline{r}_L$ is very close in direction to $\bar{\underline{\nu}}$.

Therefore, to cope with this situation, we shall add other two steps, integrating the observation eqs (38) and (50) in such a way that $\underline{t}\cdot\bar{\underline{\nu}}$ can be conveniently estimated (step 9) and, if we want to know better the whole $\underline{\varepsilon}$, also $\underline{\varepsilon}\cdot\bar{\underline{\nu}}$ can be well estimated (step 10).

3.9 Step 9: The tide-gauge condition

Assume that $\bar{P}$, the origin of the height datum, is a tide gauge and it is included in the GPS stations too. Then, as claimed in eqs (25), (26) and (27), we know both $\underline{r}_{0\bar{P}}$ and $\underline{r}_{L\bar{P}}$, so that we can add to (48) the observation equation

$$\underline{r}_{0\bar{P}} - \underline{r}_{L\bar{P}} = \underline{t} + \underline{\varepsilon}\wedge\underline{r}_{L\bar{P}}. \tag{51}$$

As it is obvious (51) is certainly suited to determine $\underline{t}\cdot\underline{\nu}_{\bar{P}}$.

3.10 Step 10: The azimuth condition

Traditionally the orientation of a local geodetic datum is done by equating an ellipsoidal to the corresponding astronomical azimuth (Heiskanen & Moritz 1990, Chapter 5). The effect of this is to place the axes $\underline{e}_{0Z}, \underline{e}_{LZ}$ parallel to one another. If we assume this as a condition, we see that

$$\underline{\varepsilon} = \varepsilon\underline{e}_Z = \varepsilon\sin\varphi\underline{\nu} + \varepsilon\cos\varphi\underline{e}_\varphi. \tag{52}$$

It is then easy, after substitution into (48), to realize that ε can be well determined, even for a small zone A, apart from the pathological case of the poles where $\cos\varphi \cong 0$. This basically answers to the problem we have formulated.

4 SUMMARY OF THE RESULTS AND DISCUSSION

In Section 3 we have established linearized and suitably approximated (according to our definition at the end of Section 1) observation equations, involving the unknowns of our problem, namely:

$T_0(P)$: anomalous potential with respect to a normal field attached to a geocentric reference ellipsoid,
$\underline{t}, \underline{\varepsilon}$: rototranstlation parameters for the transformation, $E_0 \leftrightarrow E_1$,
δW: height datum parameter.

Since following the 10 steps the reader can easily lose the point we deem it useful to summarize them making more explicit the way in which the estimation problem is solved. Basically our observation equations, and the constraints involving all the unknowns, are:

$$\begin{cases} \Delta T_0 = 0 \\ -\dfrac{\partial T_0}{\partial h} + \dfrac{\gamma'}{\gamma}T_0 = \Delta g_L - H(\underline{t}\cdot\underline{e}_Q + \Re\underline{\varepsilon}\cdot\underline{\ell}_\lambda) + \dfrac{\gamma'}{\gamma}\delta W \end{cases} \tag{53}$$

here (*cf.* 42) the Laplace equations has to hold in the exterior of the telluroid $\tilde{S}_0$, (see eq. 37 and subsequent comment), while the boundary relation (53) has to hold on the part of $\tilde{S}_0$ that corresponds to the area A. Then we have the 'horizontal' GPS equations (*cf.* eq. 48)

$$\left(I - P_{\nu_i}\right)\left(\underline{t} + \underline{\varepsilon}\wedge\underline{r}_{P_1}\right) = \left(\underline{r}_{0Q_i} - \underline{r}_{LQ_i}\right) + h_{0P_i}\left(\underline{\nu}_{0P_i} - \underline{\nu}_{LP_i}\right) \tag{54}$$

where $\underline{r}_{0Q_i} - \underline{r}_{LQ_i}$ and $\underline{\nu}_{0P_i} - \underline{\nu}_{LP_i}$ are computed from $(\varphi_{LQ_i}, \lambda_{LP_i}, \varphi_{0P_i}, \lambda_{0P_i})$ and h_{0P_i} are directly known from GPS. To (54) we can add the 'vertical' GPS equations (*cf.* 49)

$$\frac{1}{\gamma}T_0(P_i) - \frac{1}{\gamma}\delta Wm = \zeta_{0P_i} = h_{0P_i} - \tilde{h}_{0P_i} \tag{55}$$

where $\tilde{h}_{0P_i}$ is nothing but the telluroid height derived from eq. 37.

Already at this point if we have a sufficient spread of the vectors $\underline{\nu}_i$ in eq. (54) (i.e. if GPS stations P_i lying in A have separations of an order of magnitude comparable to the Earth's radius) (53), (54) and (55) are capable of determining our unknowns.

The solution should be achieved on a rigourous basis, as observed at the end of point (b) in Section 2, by including for all the equation a noise model and treating them as in a form of an overdetermined boundary value problem. An equivalent solution can be found by eliminating part of the unknowns and part of the equations. This can be done by using (53), to eliminate T_0, i.e. by determining T_0 as function of Δg_L (observed), $\underline{t}, \underline{\varepsilon}$ (unknowns) and substituting it into (55).

Since the problem (53) is linear we can always find a linear operator $S\{\cdot\}$ such that (*cf.* the discussion in step 6, Section 3)

$$T_0 = S\{\chi_A\Delta g_L\} - \underline{t}\cdot S\{\chi_A H\underline{e}_q\} + \underline{\varepsilon}\cdot S\{\chi_A H\Re\underline{e}_\lambda\} + \delta W S\left\{\chi_A\frac{\gamma'}{\gamma}\right\}. \tag{56}$$

In this equation, derived when we know a global model T_M and of a residual height model for the area A, if we give names to the known quantities (*cf.* eq. 45)

$$\tilde{T}_0 = S\{\chi_A\Delta g_L\}, \underline{\eta}_1 = S\{\chi_A H\underline{e}_Q\}, \underline{\eta}_2 = S\{\chi_A\Re\underline{e}_\lambda\}$$

$$k = S\left\{\chi_A\frac{\gamma'}{\gamma}\right\},$$

we recognize that the unknown T_0 is expressed as the actual local solution of the geodetic BVP, $\tilde{T}_0$, computed ignoring any datum problem, suitably corrected by terms depending on the other unknown parameters $\underline{t}$, $\underline{\varepsilon}$, δW. Substituting in (55) yields

$$-\underline{t}\cdot\underline{\eta}_{1i}-\underline{\varepsilon}\cdot\underline{\eta}_{2i}+\frac{1}{\gamma'}[k_i-1]\delta W=\zeta_{0P_i}-\frac{\tilde{T}_0(P_i)}{\gamma_i}. \tag{57}$$

These equations together with (54) constitute the reduced observation system for the unknowns $\underline{t}$, $\underline{\varepsilon}$, δW, which could now be estimated via least squares. A rigourous application of least squares would require, in particular, the propagation of observational noise from Δg_L into $\tilde{T}_0(P_i)$ in eq. (57). Also, if the points $\{P_i\}$ in A cannot span long distances one is forced to introduce more information in order of well estimating the parameters. In particular, to determine $\underline{t}\cdot\underline{\nu}$ one can use for $\bar{P}$ the 'tide gauge' eq. (38)

$$h_{0\bar{P}}-h_{L\bar{P}}=\underline{\bar{\nu}}\cdot(\underline{t}+\underline{\varepsilon}\wedge\underline{r}_{\bar{P}})$$

considering that $h_{0\bar{P}}$ is known by GPS while $h_{L\bar{P}}=0$.

Finally if the full vector $\underline{\varepsilon}$ has to be estimated for a small area A one can use the re-parametrization of $\underline{\varepsilon}$ in terms of one variable only, (52), which derives from imposing the classical condition of equality of an astronomical azimuth with the corresponding azimuth in the local geodetic network in A. Before closing let us remind the reader that the solution of this problem provides us with both an enhanced estimate of the geoid over A and the full geometric transformation between the global reference system based on E_0 and with coordinates $(\varphi_o, \lambda_0, h_0)$ and the local one based on E_L and described in terms of the classical coordinates $(\varphi_L, \lambda_L, H)$.

Finally, What we proved here is not that the spherical approximation, with its splitting of the determination of δW from $\underline{t}$, $\underline{\varepsilon}$, is wrong; rather we have proved that it is not sufficient, at the 1 cm level, if the global rototranslation produces shifts of coordinates of the order of 100 m. Nevertheless one could always think of using a purely spherical approximation to determine gross estimates of δW, $\underline{t}$, $\underline{\varepsilon}$ and then one could use a new local datum E'_L much closer to E_0, to reproduce the whole computation and come to a final estimate. Naturally the main road stays for us in using the approach presented here.

REFERENCES

Balasubramania, N., 1994. Definition and realization of a global vertical datum, Ohio State University Report no. 427.

Colombo, O.L., 1980. A world vertical network, Ohio State University Dep.—Geod. Sci.

Forsberg, R., 2000. Draping of geoids to GPS observations, IGeS Geoid school Johor, Malaysia, IGeS Publications, Milan.

Heck, B., 1989. A contribution to the scalar free boundary value problem of physical geodesy, *Manu. Geod.*, **14,** 87–99.

Heiskanen, W. & Moritz, H., 1990. *Physical Geodesy,* Inst. of Phys. Geod., TU Graz.

Lehman, R., 2000. Altimetry gravimetry problems with free vertical datum, *J. Geod.*, **74,** 327–334.

Migliaccio, F., Sacerdote, F. & Sansò, F., 1989. Overdetermined BVP with parameters, *2nd Hotine Marussi Symposium on Mathematical Geodesy,* June 5–8, 1989, eds Sacerdote, F., Sansõ, F., 635–645.

Milbert, D.G., 1995. Improvement of a high resolution geoid height model in the United States by GPS height on NAVD88 benchmarks, *IGeS Bulletin,* **4,** 13–36.

Moritz, H., 1980. *Advanced Physical Geodesy,* H. Wichmann Verlag, Karlsruhe.

Rapp, R.H. & Balasubramania, N., 1992. A conceptual formulation of a world height system, Ohio State University Report no. 421.

Rummel, R. & Teunissen, P., 1981. Height datum definition, height datum connection and the role of the geodetic boundary value problem, *Bull. geod.*, **62,** 477–498.

Sansò, F. & Usai, S., 1995. Height datum and local geodetic datum in the theory of geodetic boundary value problems, *Allgemaine Vermussungs-nachreich ten, Wichmann, Heft* **8–9,** 343–385.

Xu, P., 1990. Monitoring the sea level rise, *Delft Reports (new series),* 90.1.

Xu, P. & Rummel, R., 1991. A quality investigation of global vertical datum connection, *Netherland Geodetic Commission,* **34,** *New Series*.