Refinement of the Geoid from GEOS-C Data

Bernard H. Chovitz
NOAA/National Ocean Survey
Rockville, Md. 20852

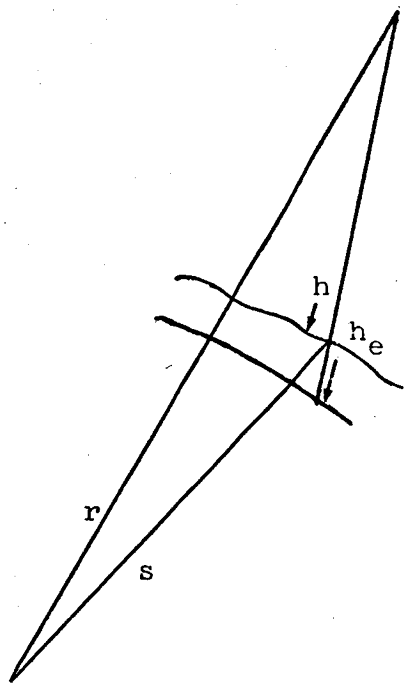
I begin with two assumptions: first, the accuracy of the GEOS-C altimeter is known; second, the altimeter measures the distance between the satellite and the geoid, (that is, the geoid is coincident with sea level). In the context of GEOS-C, the first assumption is definitely false. In fact, the primary objective of the GEOS-C altimeter experiment is to verify the accuracy of the altimeter itself. This is as it should be; the altimeter opens up such a fruitful source of data, that it is most important to determine just how good this data is. However, it is hoped that this question can be resolved, so that the data then can be used for geodetic and geophysical application. With respect to the difference between sea level and the geoid, any time-invariant effects (like currents) or long-period effects (like tides) will be an order of magnitude smaller than the fine structure in the geoid separation (of the order of 5 to 10 meters) which cannot be discerned by dynamical satellite analysis but which may be realizable from altimetry.

The basic principle of geoid determination from satellite altimetry over the oceans is as follows (fig. 1). By tracking, the height of the satellite above the ellipsoid, h_e, is obtained. The satellite's height above the geoid (using assumption 2 above), h, is obtained by altimetry. Then the geoid height, $N = h_e - h$.

The question arises: since the height of the geoid above the ellipsoid depends on the determination of a dynamic orbit, and this in turn depends on the knowledge of the gravitational field, which is equivalent to knowing the geoidal height, isn't this a circular approach? The answer is, no, because the variations in N are of much shorter wavelength than their effect on the orbit, and hence the orbit is not appreciably affected by neglect of these short wave variations.

A further step in addition to the determination of the localized ocean geoid is the use of the altimetry data to refine the global gravity field. This will yield a better reference orbit and determination of h_e, and thereby improve the value of N. The altimetry provides data for observation equations which can be added to observation equations obtained from tracking for the improvement of parameters relating to the orbit and the gravitational field.

Figure 1



From fig. 1,

$$\vec{h} = \vec{r} - \vec{s}$$

where $\vec{h}$, $\vec{r}$, $\vec{s}$ are vectors, $\vec{r}$ and $\vec{s}$ being the geocentric position of the satellite and sub-satellite ocean surface point, respectively. For the purpose of writing a linearized observation equation, the small angles between these vectors are neglected, and their magnitudes are taken in the relation

$$h = r - s.$$

This approximation can be recovered by iteration.

Then the observation equation for the measured altitude h, is

$$h_{obs} + \delta h = h_{calc} + \frac{\partial h}{\partial p} \Delta p$$

where p is a vector of parameters and δh is due to the imperfection in the observation. Then

$$h_{obs} + \delta h = h_{calc} + \frac{\partial r}{\partial p} \Delta p - \frac{\partial s}{\partial p} \Delta p.$$

Also, $$r = r(E, X)$$

where E is a set of orbital parameters, and X a set of gravitational parameters (non-gravitational effects being neglected or considered as perfectly known).

$$s = A_e \,(1 + B^T X)$$

represents the radius of a point on the geoid expressed in terms of a scaling factor (which in this case can be taken to be the earth's equatorial radius, A_e) and the set of gravitational parameters X, oriented by the vector B. (For example, if X were the usual spherical harmonic coefficients, B would be a set of spherical harmonics).

$$\text{Then} \quad \frac{\partial r}{\partial p}\,\Delta p = \frac{\partial r}{\partial E}\,\Delta E + \frac{\partial r}{\partial X}\,\Delta X$$

$$\text{and} \quad \frac{\partial s}{\partial p}\,\Delta p = \Delta A_e(1 + B^T X) + A_e\, B^T\,\Delta X$$

$$\approx \Delta A_e + A_e B^T \Delta X$$

finally yielding

$$h_{obs} + \delta h = h_{calc} + \frac{\partial r}{\partial E}\;E + \left(\frac{\partial r}{\partial X} - A_e B^T\right)\Delta X - \Delta A_e.$$

The form of this observation equation is due to Kaula (unpublished). A similar formulation can be found in Lundquist et. al., [1969].

To state the problem in its most comprehensive form involves two further considerations. First the gravitational parameters, X, have purposely been written in ambiguous form, because many of the detailed solutions to this problem proposed up to now have advocated functions for X which are deliberate alternatives to the conventional spherical harmonic approach. The essential difficulty with spherical harmonic coefficients is that they are integrated averages over the entire surface, and thus the higher degree harmonics can have no meaningful physical correlation with specific portions of the earth's surface. A second consideration is the insertion of all possible data sources for an overall solution. This means taking advantage of gravity data on land, and the tracking data itself.

Let us consider an approach due to Koch [1970]. Since altimetry yields geoid heights, N, as data, the inverse of Stokes' formula can be employed [Molodensky et. al., 1962, p. 50]

$$\Delta g_s = -\gamma\left(\frac{N_s}{R} + \frac{1}{2\pi}\iint_\sigma \frac{N - N_s}{r^3}\, d\sigma\right)$$

where the subscript s denotes the point of measurement, r is the distance between s and the surface elements $d\sigma$ of the sphere σ of radius R, γ is normal gravity, and N is the geoid height at $d\sigma$. To apply this formula the geoid heights N must be known over the entire globe; however, altimetry will not be available over land. But Stokes' formula itself is available:

$$N_s = \frac{1}{4\pi R\gamma} \iint_\sigma \Delta g . S(\psi)\, d\sigma$$

where ψ is the spherical arc between s and $d\sigma$, $S(\psi)$ is Stokes' function, and Δg is the gravity anomaly on $d\sigma$. This formula depends on knowledge everywhere of Δg which has been obtained mainly on land (and is even sparse in many areas there). But gravity anomalies closest to the fixed point have the greatest influence on the geoid undulations, and approximate values for Δg on the oceans should suffice to give a good initial set of N on the continents. Then successive approximation between these two formulas should yield representative values of Δg_s over the oceans.

This preliminary approach has both mathematical and physical deficiencies. The former lies in the fact that the conditions for convergence of the scheme are not specifically known and proven. However, physical intuition leads us to believe that failure of convergence would be due mainly to a lack of sufficiently well-distributed data. This could be overcome by using statistically obtained, instead of observational, data, although this alternative is not desirable. However, there are also deficiencies due to imperfect physical assumptions. The use of Stokes' formula and its inverse presupposes that the Earth has been "regularized", that is, there are no masses outside the geoid. Thus all topography is neglected. Over broad regions and in the middle of the oceans, this will not mean much, but over special areas of interest--like sea trenches, and the continental shelf regions--this approximation must be accounted for.

This can be accomplished by introducing two sets of integral equations, one of which uses N, the other Δg, as observational data:

$$\frac{R^2}{\gamma} \iint_\sigma \frac{\chi}{r}\, d\sigma = N_s$$

$$2\pi\, \chi \cos^2 \alpha_s - \frac{3}{2}\; \frac{1}{R} \iint \frac{\chi}{r}\, d\sigma - \iint \frac{H-H_s}{r^3}\, \chi\, d\sigma = \Delta g_s$$

The derivation of these equations may be found in Koch [1970] and Molodensky et. al. [1962, Ch. 5]. H is the topographic height and α is the deflection of the vertical. The unknown in these equations is the parameter χ which expresses the anomalous gravitational field as a simple density layer on the reference surface. The practical method for solving these equations is to replace the integration by a summation over a set of surface elements with a single density, χ_i, corresponding to each surface element σ_i. This yields a set of linear equations in χ_i $(i = 1 \ldots, n)$ where n is the number of surface elements, which can be treated as observation equations in the usual fashion, taking advantage of redundant data (s>n), and employing pertinent weights.

Young [1970] tackles the same problem as Koch in considering worldwide data consisting of a mix of gravity anomalies on land, and geoid heights (from altimetry) at sea. Young sets up a function

$$\zeta = \frac{r}{2} \frac{\partial T}{\partial r}$$

where T is the anomalous potential. By the so-called fundamental theorem of geodesy [Heiskanen and Moritz, 1963, p. 88], there is obtained

$$\zeta = - \frac{r}{2}\, \Delta g - \gamma N.$$

Young has two purposes; first, to exhibit uniqueness and existence proofs for the determination of T, and second to provide an algorithm for the computation of T. The choice of ζ satisfies these purposes in the following way:

$$T = \frac{1}{4\pi} \iint K\zeta\, d\sigma$$

is the formulation of the Neumann (or second boundary-value) problem, which can be solved on the sphere by representing the kernel K in terms of spherical harmonic functions. Furthermore, to begin the algorithm, one can set the initial ζ equal to

$-\frac{r}{2}\,\Delta g$ on land, and to $-\gamma N$ at sea. The algorithm then proceeds by solving for T in terms of spherical harmonic corrections δC directly from the integral expression. Practically, this is done by a summation over a set of surface subdivisions, similar to Koch's formulation. However, since spherical harmonics are directly involved in the kernel, each summation term itself is an integral of the form

$$\int_{\phi_1}^{\phi_2} P_n^m(\sin\phi)\cos\phi\, d\phi$$

where $P_n^m(\sin\phi)$ is a spherical harmonic function of the latitude ϕ. Recursion formulas for this are available to expedite the computation. The algorithm proceeds by computing corrections to ζ in terms of the current δC until convergence is reached.

Young provides necessary conditions for the uniqueness and existence of a solution for his method. As long as the zeroth harmonic is given, a solution exists regardless of the relative distribution of the gravimetry and altimetry. The computational procedure, however, does not provide for the use of redundant data, and involves more complicated computations than Koch's method.

The most comprehensive attack on the problem combines altimetry, gravimetry, and tracking data into one simultaneous solution. This has been outlined by Koch [1970] in connection with the density layer method of expressing the geopotential. The integral equation expressing the geoid height, N, as a function of χ is introduced into the observation equation for the altimetry measurement h_{obs}. This is combined with integral equations in Δg and with the conventional tracking data observation equation. Computational complexity is proportional to the size of the surface elements chosen. This particular approach is very flexible since the size can be varied according to the specific use being made. The satellite orbit is not sensitive to high frequency undulations (except in special cases of resonance); hence the residual field can be approximated by a coarse subdivision. On the other hand, to obtain the detailed structure, a finer subdivision will be required. A common solution of all data (altimetry, gravity, and tracking) can employ both the fine and coarse mesh. Final values of N and Δg are computed directly from the corresponding integrals using the final set of χ_i. If desired, spherical harmonic coefficients can also be obtained from the χ_i.

Lundquist et. al. [1969] have concentrated on the problem of best expressing the geopotential. This method employs "sampling" functions which are linear combinations of spherical harmonics, such that each function peaks strongly in the neighborhood of a particular point. If the formulation is to be equivalent to a spherical harmonic expansion up to degree n, then $(n+1)^2$ such points are chosen. The rationale behind this method lies in the simplification in the computational procedure over the conventional spherical harmonic representation of the gravity field. The coefficients of these functions are those designated by X in the altimetry observation equation exhibited earlier, and their improvement ΔX is obtained by using just this equation. Paraphrasing from Lundquist et. al. [1969], the sampling function coefficients over ground points will maintain their initial values, obtained from the best information available otherwise. However, there appears to be no reason why a further set of observation equations for Δg in terms of sampling function coefficient parameters could not be added, so that the method would be conceptually as complete as the other two. In addition, the approaches of both Young and Lundquist et. al. should be amenable to the addition of tracking data in a simultaneous solution.

It is plausible to assume that all these methods are equally reliable in having the theoretical capability of yielding valid results. The superiority of one over the other will probably be in computing efficiency.

The amount of altimetry data points recoverable from GEOS-C is potentially very large. Assuming one measurement per second for a 20 minute altimeter run each revolution over a two-year lifetime, the number of data points is of the order of 10^7. A more conservative estimate, mentioned by Hudson [1971], is $5.5(10^5)$ data points based on 1500 hours of data. Since there are approximately 36,000 1° squares (subdivisions whose area is the same as a 1° x 1° square at the equator) over water, there will be on the average 15 data points per 1° square. In general, the oceans will be covered by altimetry better than the land by gravity, provided that the coverage is uniform.

Statistical problems will emerge. Since the satellite travels about 7km per second, the points falling within a degree square (100 x 100km) are likely to occur over one or two individual revolutions, and thus present correlation problems. Should aggregation be practiced as in the case of Dopper data of which there is an excess? In fact this is the method employed on land where the Δg are aggregates obtained from individual gravity measurements.

The way to first proceed probably will be to obtain a uniform solution for the global geoid employing large size subdivisions, say 10° x 10°. The altimetry could be aggregated more consistently over a block of this size. Such a solution should be sufficiently accurate to obtain an orbit for the purpose of securing the geocentric position of the satellite which can serve as a geoidal reference against each altimeter measurement.

References

Heiskanen, W. and H. Moritz, Physical Geodesy, W. H. Freeman & Co., San Francisco, 1967.

Hudson, E. H., A Geodetic and Oceanographic Satellite Altimeter System, AIAA Space Systems Meeting, Denver, Colorado, 1971.

Koch, K. R., Gravity Anomalies for Ocean Areas from Satellite Altimetry, Proceedings of the Second Marine Geodesy Symposium, Marine Technology Society, Washington, D. C., 1970.

Lundquist, C. A., G. E. O. Ciacaglia, K. Hebb, and S. G. Mair, Possible Geopotential Improvement from Satellite Altimetry, Smithsonian Astrophysical Observatory Special Report 294, Cambridge, Mass., 1969.

Molodensky, M. S., V. F. Eremeev, and M. I. Yurkina, Methods for Study of the External Gravitational Field and Figure of the Earth (translated from the Russian, 1960), Israel Program for Scientific Translations, Jerusalem, 1962.

Young, Ronald Ging-wei Eng, Combining Satellite Altimetry and Surface Gravimetry in Geodetic Determinations, Massachusetts Institute of Technology TE-37, Cambridge, Mass., 1970.