ON THE GLOBAL VERTICAL DATUM AND ITS ROLE IN MARITIME BOUNDARY DEMARCATION

Petr Vaníč ek
Department of Geodesy and Geomatics Engineering
University of New Brunswick
P.O. Box 4400
Fredericton, N.B. CANADA
E3B 5A3

ABSTRACT

Offshore boundary demarcation, according to the United Nations Convention on the Law Of the Sea (UNCLOS), often requires a knowledge of sea level behaviour. This sea level is specifically defined as the "local sea level" which is only remotely related to the global vertical datum as understood in geodesy. Why do we then need a global vertical datum in maritime boundary demarcation? The only reason is to put the maritime boundaries on a solid geometrical foundation, as we do elsewhere in geodesy. In this contribution, we discuss the definition of a global vertical datum and its relation to the local sea level behaviour, sea surface topography, and some other geodetic considerations as encountered in the UNCLOS applications.

GLOBAL VERTICAL DATUM

The global vertical datum is one of the basic geodetic notions, and every honest-to-goodness geodesist is supposed to know what it is. Yet, we submit that there is a considerable divergence of opinion as to what constitutes a global vertical datum, never mind a local vertical datum, which is still a somewhat different and even more controversial notion. We will probably all agree that there are at least two basically different definitions of the global vertical datum:

i) It is a geocentric reference ellipsoid used as a reference surface for geodetic (ellipsoidal, geometric) heights;
ii) It is a specified equipotential surface (the geoid) of the actual gravity potential W of the earth (possibly including the constant term of the actual tidal potential)

$$W = W_0 = \text{const.} \tag{1}$$

as a reference surface for orthometric or dynamic heights.

We note that this specified equipotential surface is also the reference surface for normal heights at sea, as the geoid and quasigeoid at sea are one and the same surface.

Let us have a closer look at the first definition. To begin with, a geocentric reference ellipsoid is often defined as an equipotential surface of the (artificial) normal gravity potential and, in this respect, it is superficially similar to the second definition. The position, orientation, shape, and size of the geocentric reference ellipsoid are exactly and uniquely defined in a meaningful geocentric coordinate system and do not change with time. This fact gives us the possibility to exactly obtain the geodetic height h (reckoned above the geocentric reference ellipsoid) of a point from known geocentric coordinates, say x, y, z, of that point. This is a distinct advantage for someone who uses geocentric coordinates determined by one of the space techniques such as the Global Positioning System (GPS), Very Long Baseline Interferometry (VLBI), or Satellite Laser Ranging (SLR). We shall not discuss here the role of non-geocentric reference ellipsoids (or regional horizontal datums, the more frequently employed term) as vertical datums.

While the geocentric reference ellipsoid is a very good choice of a global vertical datum for points determined by space techniques, it is not a very good choice for a global vertical datum for practical applications because geodetic heights h are not physically very meaningful. Height "consumers" expect to see heights of points at the sea shore to be close to 0, having been trained to think in terms of "heights above the sea level". As we all know, geodetic heights of sea shore points range within almost 200 metres.

The second definition is, of course, much more satisfactory from the practical point of view. The sea level surface follows quite closely an equipotential surface, so much so, that up until relatively recently, the mean sea level surface was considered to be equivalent to the geoid. We know now that this equivalence is valid only in the first approximation (see below), and that in applications requiring sub-metre accuracy, the difference must be taken into account.

The closeness of the two surfaces is responsible for the choice of the particular equipotential surface to be used as the global vertical datum, i. e., the choice of the geoid. To put it more formally: the geoid is that particular equipotential surface which "corresponds to within about 1-2 m to the time-averaged ocean surface" [Lambeck, 1988], or "which best fits, in the least squares sense, mean sea level" [National Geodetic Survey, 1986]. This means that it is the mean sea level (MSL) surface that implies the value W_0 of the potential associated with the geoid. Following this definition, we can rewrite eqn. (1) as

$$W = W_0 \text{ (MSL)} \quad , \qquad (2)$$

understanding that the value W_0 of the potential on the geoid changes with time if MSL changes with time (see below), i. e.,

$$W(t) = W_0\,[MSL(t)]\ . \tag{3}$$

While the above definition of the geoid is preferred by those who use the geoid as a vertical datum on land, there is yet another definition of the geoid used in geodesy. This alternative definition specifies a particular value W_0 for the geoid and assumes this value to be constant in time, see, e. g., [Bursa, 1992, p. 195]. This latter definition is preferred by more theoretically oriented geodesists. At this point we must make clear that even this latter definition does not describe a time invariant geoid; the gravity field changes in response to temporal variations in the density distribution. Thus, following the latter definition, eqn.(1) should read:

$$W(t) = W_0\ , \tag{4}$$

showing that the shape of the geoid varies in time even when the value of its potential does not. These temporal changes are, however, relatively small and even for large mass transfers such as the one represented by the post-glacial rebound of the Laurentide (eastern Canada), the change in the geoid is at most 1 millimetre per year [Sjøberg et al., 1990].

The question that one is immediately tempted to ask is: How different would the geoid defined through the MSL surface be from the geoid defined by a constant in time value of the potential? In the past ten thousand years, the two geoids have diverged by as much as some 80 metres [Peltier et al., 1978]. The present rate of divergence is dictated mostly by the global eustatic water rise which is somewhere between 1 and 2 millimetres per year [Emery and Aubrey, 1991, Table 8]. Clearly, it would be very advisable for an international body such as the International Association of Geodesy (IAG) to coin only one definition of the geoid and provide geodesists who desire to work with the other definition with a transformation algorithm from one to the other concept.

Let us finally turn to the relation between the first definition of global vertical datum (the reference ellipsoid) and the second definition (the geoid). The relation is given simply by the geoid-ellipsoid separation, usually called the geoidal height (above the reference ellipsoid). Geoidal height (of which geoid?) is the link between the two basic definitions of the global vertical datum. Since the first global vertical datum is time invariant and the geoid varies with time, the geoidal heights must, of course, vary with time. These variations are largely unknown, mostly even unacknowledged. This is not surprising, when we realise that geoidal heights are globally known only to an accuracy of 1 metre or worse. We reiterate here that we are still talking about global rather than regional cases; regional cases are generally quite different.

LOCAL SEA LEVEL

In the UNCLOS, the maritime boundary is, one way or the other, referred to some land points (base points) that have to be identified in nature. We shall not discuss here the prescribed procedures for this identification. We shall only state that these base point are referred to local sea level, in fact to local "low water line" [United Nations, 1983, Article 5]. The local sea level variations that one has to take into account in the context of determining the low water as well as other species of local sea levels, are those caused by local sea tides; we shall assume that the sea tide phenomenon is sufficiently well known so we shall not describe it here. We note that the custom of using the local sea level as a reference for boundary demarcation originates in the custom of using this reference in bathymetric charting as practised by national hydrographic services the world over.

We must also note, that the sea level variations, discussed in the previous section, are significantly slower than the tidal variations and consequently are ignored altogether. One may presume that the justification for doing so would be that, in the context of maritime boundary determination, we are interested only in a "short term behaviour" (a few tens of years, perhaps, but the UNCLOS is mute on this point) and even the cumulative effect of those long-term variations would be relatively small (a few decimetres).

To determine the various species of local sea level, we have to know the behaviour of the local sea tide. The best way of studying the behaviour of the local sea level is to conduct the measurements of its temporal variations, by means of tide-gauges, and a subsequent analysis of these measurements. If the record of the local tidal variations is sufficiently long, then the analysis yields sufficiently accurate results so that not only can we describe the local sea level behaviour during the recording period, but can also predict its behaviour in the near future.

Sometimes, it is possible to predict the local sea level behaviour to a sufficient accuracy from records collected at some nearby points. Such interpolation of tidal behaviour must be done with a lot of caution though; tidal behaviour may vary significantly and irregularly even along one coastline. The tidal behaviour of the local sea level is dictated not only by the tidal regime of the water body adjacent to the coast but also by the shape of the coastline and that of the sea bottom in the vicinity of the point of interest [Stanley and Swift, 1976]. We note that the determination of local tidal behaviour is usually considered to fall under the auspices of national hydrographic services.

In this treatise, we shall concentrate only on the local mean sea level (MSL) from which the other species of local sea level at the point of interest can be derived mathematically, once the behaviour of the local sea level can be modelled and predicted from the results of the above discussed tidal analysis. Because we wish to discuss the relation between the local sea level and a global vertical datum, we have to take into consideration that the local MSL is subject to some secular variations in time.

From the brief discussion of the two seemingly very different concepts of a "vertical datum" used in geodesy and in maritime boundary delimitations and from the fact that these two concepts are used in very different applications, one may reach an intuitive conclusion that there is really no need and no way to even try to tie the two together. Why, one may ask, should we even try to formulate a transformation between the two kinds of "vertical datums"? The answer is twofold:

i) To be able to use space techniques in maritime boundary delimitations. As we have seen earlier, the geocentric reference ellipsoid is the natural global vertical datum for space-technique determined heights, which would then have to be then transformed into heights above local sea level.

ii) To unify the height systems used for heights on dry land with the height/depth systems used in maritime boundary delimitation (and in bathymetric charting). It is not difficult to see that following the currently used practices, there is a discontinuity of a varying magnitude between heights used in maps and depths used in bathymetric charting. As a matter of interest, we note that the Canadian Hydrographic Service is now pursuing a study aimed at such possible unification of vertical datums in Canada.

TRANSFORMATIONS BETWEEN GLOBAL VERTICAL DATUMS AND LOCAL SEA LEVEL

Let us begin with the transformation between a geocentric reference ellipsoid and a geoid. This transformation was already discussed above in the context of a global geoid, and we have seen that there are really three problems/questions associated with this transformation:

i) We have to first decide just which definition of the geoid we wish to work with;

ii) We have to decide how we wish to deal with the temporal variations of the geoid, i. e., forget about them? attempt to model them?

iii) We have to come to terms with the relatively large errors of the global geoid models.

Concerning the last problem, it can be somewhat alleviated by taking an appropriate regional geoid model, if one is available for the region of interest, instead of the global model. Such a regional geoid would tend to be, naturally, more accurate (see, e. g., [Vaníč ek and Martinec, 1994]) because a regional model does not suffer from the cutoff of the high frequencies above the highest degree of the global model. It would also tend to have the latest gravity, satellite altimetry, and/or deflection of the vertical data included in the solution.

The first problem, we believe, should be dealt with by adopting an appropriate convention, preferably on an international level. In the sequel, we will assume that the geoid defined through the MSL has been adopted and its time variations, including the eustatic water rise, accounted for. After these steps have been taken care of, we can transform the geodetic height h of the point of interest obtained by one of the space techniques into a height, say orthometric, above the geoid. We note that in maritime boundary delimitation, the heights we deal with tend to be very small and it does not matter very much just which kind of height we consider (orthometric, dynamic, or normal); they will be practically the same.

Once we have the height above the geoid (the one defined through the MSL surface), we have to transform it into a height above the local MSL. The local MSL (point) lies, by the definition of the MSL surface, on the MSL surface. Thus the transformation we are interested in here, is the transformation between the equipotential surface that approximates the MSL surface the closest and the MSL surface itself. The separation of these two surfaces is called the Sea Surface Topography (SST), a term which reflects the terminology used on land, where the "land surface topography" is also referred to the geoid.

As we know, the surface of a body of homogeneous fluid in equilibrium is an equipotential surface. Oceans, however, are neither homogeneous nor are they in equilibrium, and their surface is only approximately an equipotential surface. Water in the oceans varies in density due to variations in temperature, salinity content and suspended particle content. The water is also subject to meteorological forces such as wind shear and barometric pressure variations, to river discharge, and to other phenomena. Then, there are the steady and the turbulent flows, which contribute a lot to the SST. For a complete list of causes of SST, the reader may consult Montgomery [1937-38], and Warren and Wunch [1981]. We note that the other phenomenon mentioned above, the tide, is not contributing to SST which deals with the sea level (**mean** sea level) from which all tidal variations have been removed.

Globally, the SST ranges between -1.5 and + 1.5 metres [Levitus, 1982]. It is customary to divide the SST into a permanent part and a time varying part:

$$SST(t) = SST_0 + \delta SST(t) \,. \qquad (5)$$

To illustrate what these two terms stand for, the permanent temperature difference between polar and equatorial regions contributes to the permanent part while the sea level variation in response to passing atmospheric highs and lows contributes to the time varying part. Strictly speaking, of course, nothing about the earth is ever permanent and on a geological time scale, even the "permanent" part will change. We have to realise that the eustatic water rise which we spoke about earlier, does not affect the SST; it affects the MSL surface as a whole and not the separation between the MSL and the geoid. It thus has to be taken care of during the transformation from the geocentric reference ellipsoid to the geoid.

How well can we quantify the SST at present? Not very well! Only the small (spatial) scale, i. e., large in spatial extent, features are known to any accuracy, and it is probably fair to say that the value of SST at any given place and time cannot be given any more accurately than to some 2 to 3 decimetres. With some special instrumentation and a well conceived experiment focusing on one locality, the SST error for this locality could be driven down to about 1 decimetre. The SST determination is usually considered to be a problem that falls under the auspices of physical oceanography. Geodesists, however, should be equally interested in the results. Since they do have some useful tools to contribute to oceanographic studies of SST [Vaníč ek, 1991], they should consider getting more actively involved with these studies

Given the value of SST as a function of location and time, it can be subtracted from the height above the geoid to give us the height above the local MSL. The tidal variations at this location are then used to transform the local MSL to any other desired species of local sea level, which completes the chain of transformations needed.

CONCLUSIONS AND RECOMMENDATIONS

In this paper, we have discussed the two basically different definitions of global vertical datums, as used in geodesy. In addition, we have pointed out that the geoid, as understood in geodesy, is also defined in two different ways, and we recommended that this dichotomy be solved by the IAG specifically adopting only one of these definitions and suggesting an appropriate transformation between the two. This calls for a study of the eustatic water rise by geodesists, physical oceanographers, and geophysicists. We have shown that both geoids vary with time, albeit in very different ways.

After discussing briefly the behaviour of local sea level, which serves, one way or another, as the vertical "datum" for maritime boundary delimitations, we have looked at possible reasons why it would be desirable to formulate sound transformation equations between the local sea level and the geocentric reference ellipsoid, as a global vertical datum. We have identified two such reasons: It would make possible the use of heights determined by one of the space techniques in maritime boundary delimitations, and it would unify the height systems used on dry land with those used in bathymetric charting.

We have discussed at length the "missing link" between the geoid and the local mean sea level, the sea surface topography. Even though the determination of sea surface topography is a task for physical oceanographers, we would recommend that geodesists should get more involved in using levelling networks as an alternative tool to those used by oceanographers.

Transformations between (geodetic) global vertical datums and local sea level datums clearly require cooperation among geodesists, physical oceanographers, geophysicists, and hydrographers. None of these sciences can do it alone. Has the time arrived to tackle this problem seriously?

I would like to express my thanks to Drs. Sue Nichols and Alfred Kleusberg, for their thoughtful comments on an earlier version of the manuscript and to Ms. Wendlyn Wells, for her editing and manuscript preparation.

REFERENCES

Bursa, M., 1992. Parameters of common relevance of astronomy, geodesy, and geodynamics. *Bulletin Géodésique*, Vol. 66, No. 2, pp. 193-197.

Emery, K.O., and D.G. Aubrey, 1991. *Sea Levels, Land Levels, and Tide Gauges*. Springer-Verlag.

Lambeck, K., 1988. *Geophysical Geodesy*. Clarendon Press, Oxford.

Levitus, S., 1982. *Climatological atlas of the world oceans*. NOAA Professional Paper No. 13, Washington, D.C., U.S. Department of Commerce.

Montgomery, R.B., 1937-38. Fluctuations in monthly sea level on eastern U.S. coast as related to dynamics of western North Atlantic ocean. *J. Mar. Res.*, 1(2), pp. 165-185.

National Geodetic Survey, 1986. *Geodetic Glossary*. NOAA Technical Publication, Washington, D.C., U.S. Department of Commerce.

Peltier, W.R., W.E. Farrell, and J.A. Clark, 1978. Glacial isostasy and relative sea level: A global finite element model. *Tectonophysics,* 50, pp. 81-110.

Sjøberg, L.E., P. Vaníč ek, and M. Kwimbere, 1990. Estimates of present rates of land and geoid uplifts in eastern North America. *Manuscripta Geodaetica,* Vol. 15, No. 5, pp. 261-272.

Stanley, D.J., and D.J.P. Swift (Eds.), 1976. *Marine Sediment Transport and Environmental Management*. Wiley & Sons, New York.

United Nations, 1983. *UNCLOS*. New York.

Vaníč ek, P., 1991. Vertical datum and NAVD 83. *Surveying and Land Information Systems*, Vol. 51, No. 2, pp. 83-86.

Vaníč ek, P., and Z. Martinec, 1994. The Stokes-Helmert scheme for the evaluation of a precise geoid. *Manuscripta Geodaetica*, Vol. 19, No. 2, pp. 119-128.

Warren, B.A., and Wunsch, C. (Eds.), 1981. *Evolution of Physical Oceanography*. MIT Press, Cambridge, Mass., U.S.A.