OC 3902 Gravity Models Notes

James R. Clynch

2001, 2002

I. Spherical Harmonics as Expansion Function for Gravity

The gravity field of the earth is represented as a potential function. This function has a Newtonian part and a rotational part.

$$W \quad = \quad V + \Phi$$

The rotational part, Φ, is a nice analytic function. It is standard procedure to model the Newtonian part, V, in a series expansion.

In choosing some function to use for the expansion two things are considered:

1. Mathematical properties of completeness and orthogonality.

   Completeness means that the function set can completely represent the gravity field. For example, a set that only had variations in the equatorial plane would not be complete.

   Orthogonality means that the different functions are "perpendicular" to each other. This is taken in an abstract sense. The "inner" product between any two functions in the set is zero unless the two are the same. For this case the inner product is defined as the integration over the sphere.

   Orthogonality has a practical advantage in extending a series expansion to include more terms for better accuracy. If you fit a function f(x) with a linear equation,

   $$f = a + b\ x,$$

   then decide that a quadratic is needed

   $$f = a' + b'\ x + c' x^2,$$

   the coefficients that you already have change. That is a' is different from a, and b' is different from b. The coefficients for powers of x are not orthogonal. But if you use orthogonal functions the coefficients do not change. (If you want a "power series like" fit that is orthogonal, you use the Hermite polynomials.)

2. The coefficients should decrease rapidly after the first few. In practical terms this means that the functions share the approximate symmetry of the function. For properties of the earth this means spherical symmetry.

The functions that meet these criteria are the spherical harmonics, $Y_{nm}(\theta, \lambda)$. Because approximate spherical symmetry is common in nature, the spherical harmonics show up many places in physics, astronomy, and even engineering. These functions are algebraic equations in the sins and cosines of the two angles. These are usually written as functions of the angles in the spherical polar coordinate system. The first angle is the co-latitude (90 degrees - latitude). The second argument is the longitude.

The main variations in the earth's gravity field for low values of n, are in the m=0 terms. These are called the zonals. They have no dependence on longitude. They are only a function of latitude. They are largest for small n (which corresponds to large features) because the equatorial bulge is the largest non-spherical feature. The terms with m not zero include variations with longitude. They are called the tesserials.

The spherical harmonics are a combination of the associated Legendre functions, P_{nm} , and sine and cosine functions of m*ë. In geodesy the explicit form is commonly used. These can be connected to the complex form used in physics and mathematics via

$$Y_{nm}(\theta,\lambda) \quad = \quad K_{nm} P_{nm}(\cos(\theta))\,[\cos(m\lambda) + i\sin(m\lambda)] \ ,$$

where K_{nm} is some constant that depends on which conventions you have chosen. There are various normalizations of these functions. Geodesy usually uses what is called Schmidt Normalization. These functions usually are written with a bar over them.

In geodesy the latitude is used for the first argument. This just changes the cos(co-latitude) into sin(latitude) and vice versa. The geocentric latitude is used in the expansion of the Newtonian potential of the earth.

II. Interpretation of Newtonian Gravity Expansion Equation

The expansion for the potential V is:

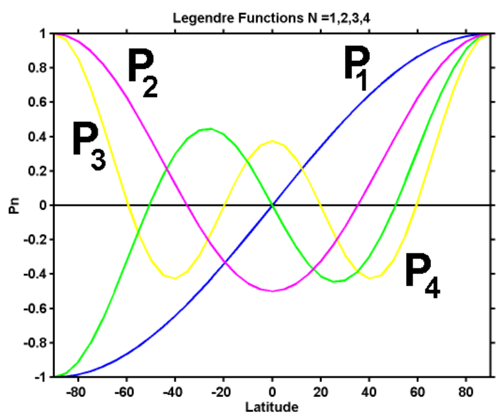
$$V = \frac{GM}{r}\left[1 + \sum_{n=2}\sum_{m=0}^{n}\left(\frac{a}{r}\right)^{n}\overline{P}_{nm}(\sin\phi')\left(\overline{C}_{nm}\cos m\lambda + \overline{S}_{nm}\sin m\lambda\right)\right]$$

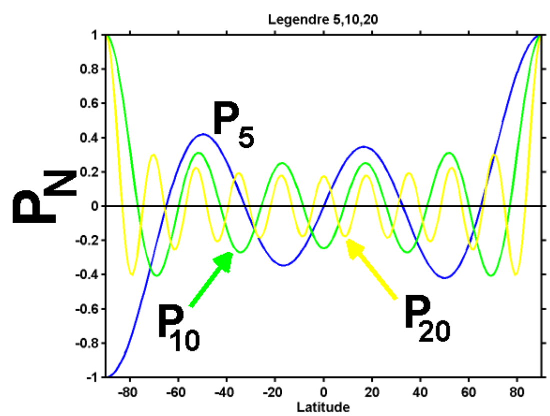
All the physical information is in the coefficients C_{nm} and S_{nm} . They represent the potential field. The angular variations are in the $\overline{P}_{nm}(\sin\phi')$, $\cos(m\lambda)$, and $\sin(m\lambda)$. The radial variation is in the $(a/r)^n$ term.

A few things are important about this equation. First notice that if the sum is zero (all the C_{nm} and S_{nm} are zero) the potential for a spherical earth is recovered. This is the meaning of the “1” inside the brackets.

The degree, the value of n, is the most important index. It is closely related to the spatial information - the scale of differences that are represented by a coefficient. This defines the spatial resolution of a particular set of coefficients. The second component, the order m, is related to the shape and orientation of the bumps of the function. If you rotate the axes and the coefficients change, but the total effects of all the values with the same degree (n) remains the same.

The maximum number of bumps in a spherical harmonic is n+1. This sets the spatial scale of the information represented. For example, the zonal functions for the first few degrees are shown in the following figures. A second figure shows a few higher degree zonals. The figures are labeled with "P_n" because the zonals (m=0 functions) are identical with the Legendre polynomials with $\cos(\phi')$ as the argument.





This means that the shortest piece of spatial information, on angular scale is about

$$\text{Angular scale} = 360 \text{ deg} / (n+1).$$

(Some people give a value half this size, it depends on the definition of resolution used.)

For the terms with n = 360, that is $C_{360,m}$ and $S_{360,m}$, the information is on a scale of 1 degree or about 110 km. For the terms of 18^{th} degree, the scale is 20 degrees or 2000 km. These resolutions are quite coarse for a lot of applications. These models are very useful for satellite work though.

The factor $(a/r)^n$ is controls how important the effects are as the altitude increases. It plays an important role in satellite problems. Here a is the semi-major axis of the earth. This term is close to one on the surface of the earth. Going up a 1000 km, where the low orbit satellites reside, reduces it to 1/1.15 . Even this relatively small fraction is quite small when raised to a moderate power. The effects of the higher degree terms fall off with altitude. The higher the degree (value of n), the faster they decrease in importance. To make a very good model of low earth orbits, you need about 36^{th} degree. For GPS, which has a value a/r = ¼, 12^{th} degree is good enough.

Because of the orthogonality property of the Y_{nm} , the method of finding the Cnm and Snm is straightforward.

$$\begin{aligned} A_{nm} &= N_{nm} \oiint V\, Y_{nm}(\phi', \lambda)\, d\Omega \\ C_{nm} &= \mathrm{real}(A_{nm}) \\ S_{nm} &= \mathrm{imag}(A_{nm}) \end{aligned}$$

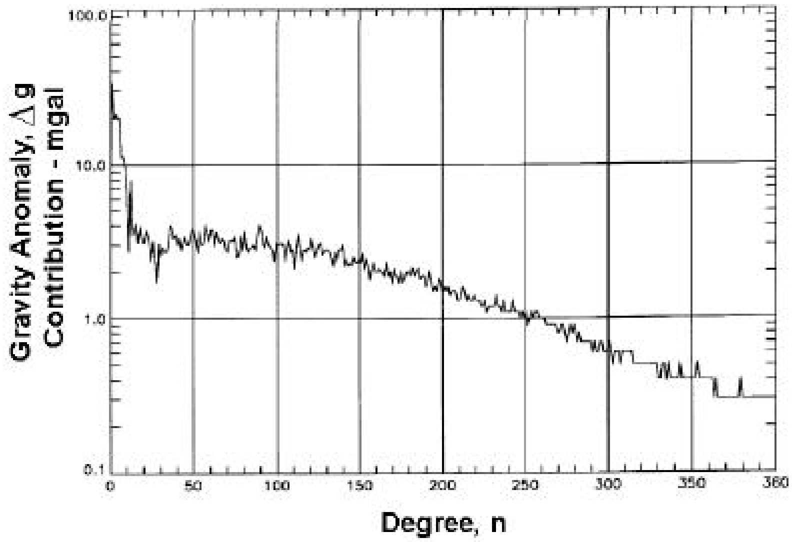
Where the N_{nm} is some normalization constant and the integral is over the surface of the earth. Gauss's theorem can be used to convert the integral over the surface of V to an integral over the volume of the density, ρ,

$$A_{nm} = N'_{nm} \iiint_{\text{Volume}} r^n Y_{nm}(\phi', \lambda) dV \quad .$$

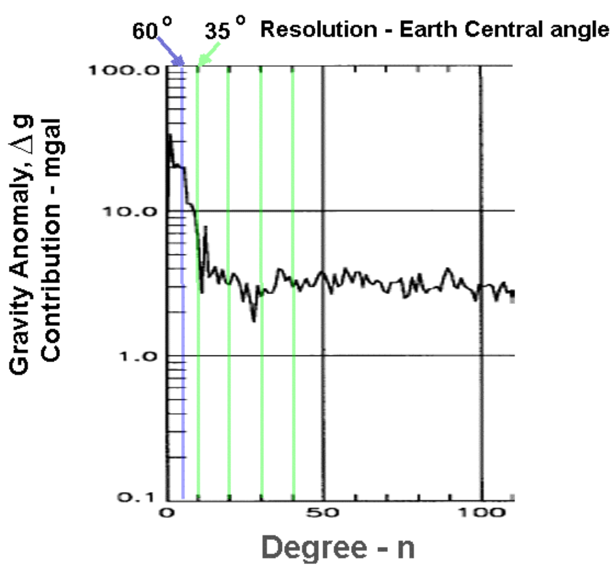
The n,0 terms all are symmetrical for rotations about the polar axis. They do not depend on the longitude. These zonal terms usually have the largest coefficients (C_{nm} , S_{nm}) of the family "n" for small n. (There are a couple of exceptions at degree 2 and 3). The earth is fairly symmetrical on long wavelengths (long distance scales). At larger n, mountain ranges start to be effective and the zonals are not necessarily dominant. The terms with m not zero are called the tesserials.

Because there are three or four common normalizations, you cannot determine the relative effects of the C's and S's by comparing them. In order to find a meaningful measure of the effects at different scales or wavelengths, the effect of the coefficients on the gravity anomalies is computed. Because the Y_{nm} are orthogonal, the variances of the gravity anomalies from each degree can be summed to get the overall variance for that degree. This gives a measure of the variations as a function of spatial wavelength (distance scale).

This degree average gravity anomaly for the Earth Gravity Model, 1996, EGM96, is shown below. Notice that the scale is logarithmic. There is a dramatic drop in the value about degrees 6 to 10 which is shown in the expanded scale on the right figure. The continents and the "density blobs" below them generate these long wavelength values. Beyond this there is a gentle falloff as the degree increases.



EGM 96 Contribution To Gravity Anomaly As A Function of Degree, n



Irregularities Fall off Below Continent Size Features

III. Gravity Models

A. WGS84 and GPS

There has been a gravity model associated with all the World Geodetic Systems. The latest DoD model is WGS84, which has been updated but not renamed. This model used hundreds of thousands of gravity measurements and a very large volume of satellite data. It was expensive data to collect and the computation stretched the computers of the time.

Originally the gravity model in WGS84 was only partially released to the public. The gravity coefficients were released to 18th degree. The undulations were released every 10 degrees of latitude and longitude. In addition the N's were rounded to the meter level. This limited the spatial frequencies that could be derived and made meaningful deflections unobtainable. This model has been called WGS84U. The entire WGS84 gravity model, which was 180th degree, was declassified in the early 1990's.

WGS84U is important because if it's effects on the Global Positioning System, GPS, receivers. At the time GPS receivers were being first deployed in the military and built by the civilians, only WGS84U could be used in an unclassified receiver. This was the only official DoD data available to the civilians. This model of the N's was put into all the military and civilian receivers of the day. It was used to convert from the satellite measured ellipsoidal heights to orthometric heights. This introduces a lot of error.

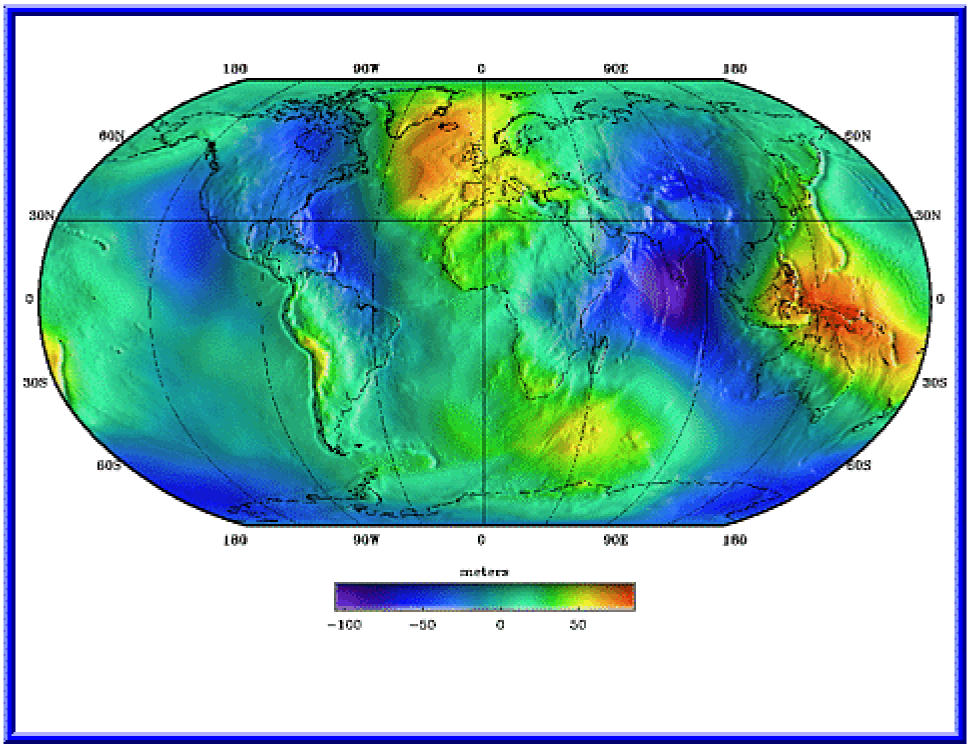
This condition still exists in 2000. Only a few high-end civilian receivers have modern, more complete undulations. This is true of all current military receivers. Therefore you must be wary of msl heights from GPS receivers.

B. Newer Models

The US National Geodetic Survey, NGS, produced a undulation and gravity model of North America, mainly from gravity measurements, called GEOID93. This was followed by a combined NGS, NASA, DoD model called Earth Gravity Model 1996 (EGM96). EGM96 is publicly available. A Geoid of North America for 1999 has been released and further models are to be expected.

These EGM's built on the work of the ocean height satellites such as Topex-Poseidon (T/P). An initial T/P model was developed using the Ohio State model OSU89B as a start. The T/P data as well as data from ERS1 and ERS2 was critical to the development of the later EGMs.

The geoid height is determined by the gravity field. The world wide height, from EGM96 is shown here.

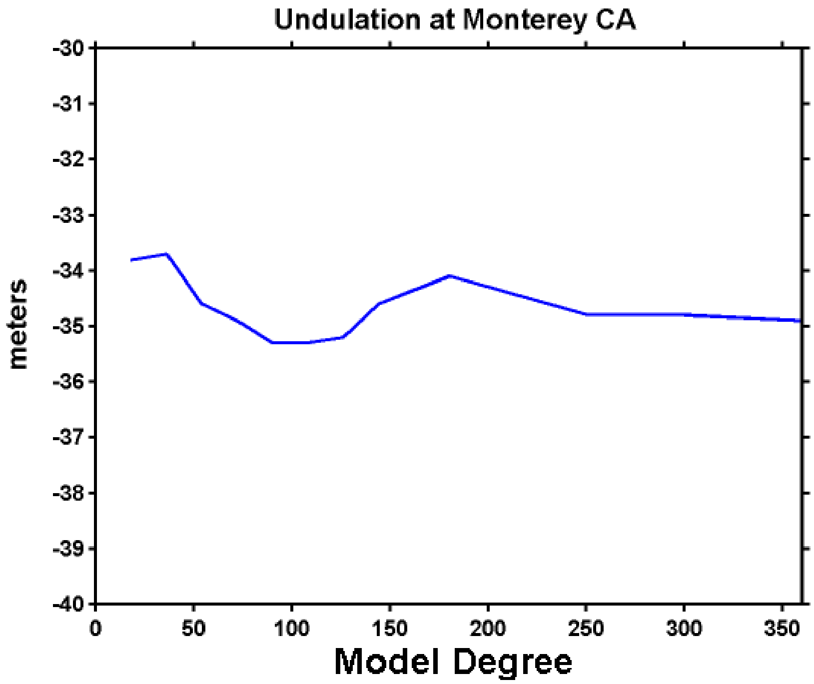


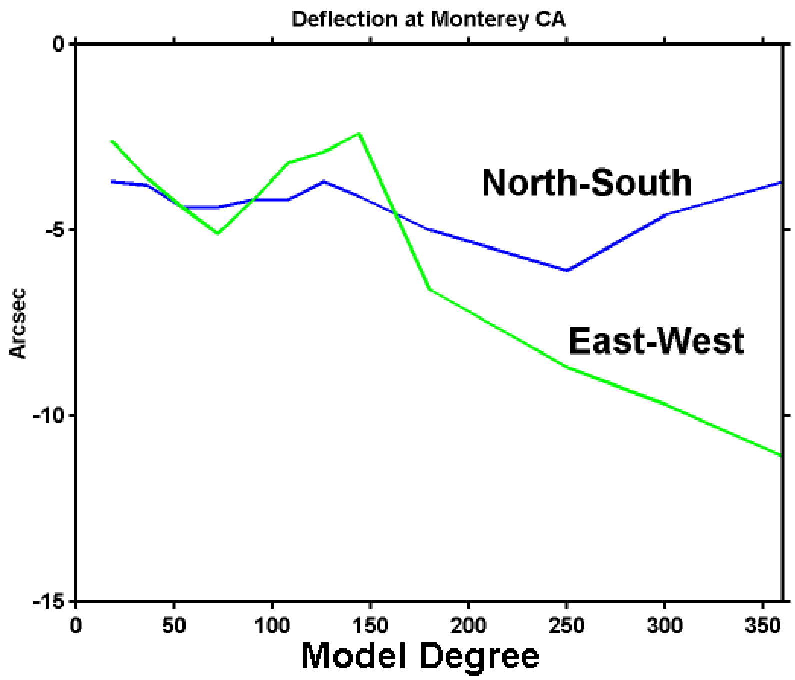
C. Utility of the Models

All of these models have their limits. They are excellent for some applications, but poor for others. In general they are very useful for satellite orbit work. This has been the driving force behind model improvement. They also are good for geoid heights. But the limits on the spatial resolution is important. Even an degree 360 model, the best available today (2002), is averaging values over distances of 100 km.

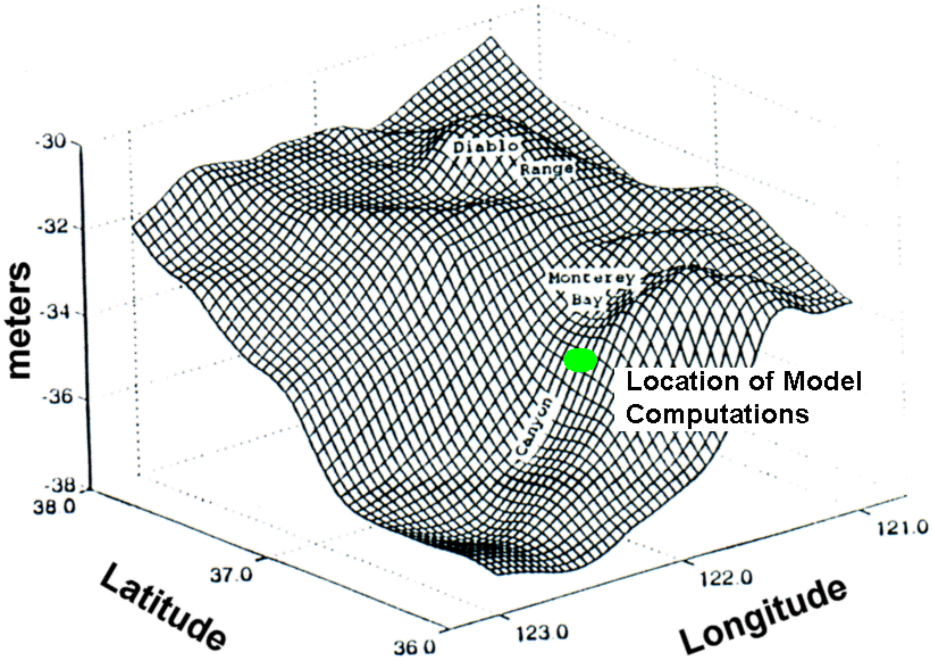
As an example of this, the geoid height and the deflection of the vertical were computed for a point near the Naval Postgraduate School (36.5 N, 122 E) with the EGM96 model. The computation was cut off at different degrees. Below are the results plotted versus the degree.

Notice that the geoid height, or the undulation of the vertical, stabilizes at about degree 180. This is an averaging scale of about 200 km. However the deflection does not stabilize. This is the normal to the geoid. This can vary quite a lot, without the geoid varying much. The east-west slope changes radically once the averaging scale gets small enough to "feel" the sharp break at the continental shelf.





In order to illustrate what is really occurring, a high precision local geoid is shown in the next figure. This was produced by incorporating over 10,000 local gravity measurements into a high precision EGM. (A Stokes integral was done.) This method accurately produces the small scale variations, but not the long scale. The EGM provides the long wavelength components and the local gravity measurements provide the small scale values. The long scale, essentially a bias and tilt, were taken from the full WGS 84 model.



Precision Geoid
Monterey Bay Area

The location of the test point is shown. On this figure, one degree of latitude is 100 km (equivalent to degree 360). Clearly there are bumps that significantly affect the normal – which is the deflection of the vertical – on scales much smaller than can be modeled with today's best EGM's.

Version of July 17, 2002