Chapter 6
Global Gravitational Models

Nikolaos K. Pavlis

6.1 Outline of the Chapter

This chapter discusses the development and use of *Global Gravitational Models* (GGMs), specifically those GGMs that are represented in the form of spherical (and/or ellipsoidal) harmonic coefficients. With the mathematical details having been presented in Chap. 3 of Part I of this book, the focus here is on the main concepts and considerations involved in the design and in the choice of alternative techniques and strategies that can be used to develop GGMs. Recent advances in geodetic techniques, in particular the availability of dedicated geopotential mapping missions on one hand and the availability of very high resolution GGMs on the other, provide the natural setting for the discussion that follows. Section 6.2 provides an introductory overview of the main concepts and distinguishes between *Global* and *Regional (or Local)* models, the latter being discussed in subsequent chapters within this part of the book. Section 6.3 discusses the aspects involved with the representation of GGMs and the characteristics of the data that are used to create the GGMs. Section 6.4 discusses the new satellite missions that are dedicated to the mapping of the gravitational field from space, and the advances and challenges that these missions introduce to GGM developments. Section 6.5 discusses the combination of the gravitational information obtained from satellites with the information obtained from surface data, which permit the development of very high resolution GGMs like EGM2008. Sections 6.2–6.5 provide the main concepts underlying the development of GGMs, omitting intentionally the mathematical and numerical details. In contrast, Sect. 6.6 discusses in some detail the specific mathematical and numerical procedures that may be used for the development of GGMs. For this purpose, two models are used as representative examples in Sect. 6.6 – EGM96, which represents the state-of-the-art *before* the availability of data from CHAMP and GRACE, and EGM2008, which represents currently the global model with the highest accuracy (developed *prior* to the availability of data from GOCE) and also the highest resolution. Section 6.7 discusses briefly the data requirements and the availability of the data necessary to develop GGMs. Section 6.8 deals with several

F. Sansò and M.G. Sideris (eds.), *Geoid Determination*, Lecture Notes in Earth System Sciences 110, DOI 10.1007/978-3-540-74700-0_6,

© Springer-Verlag Berlin Heidelberg 2013

aspects related to the use of a GGM and its by-products. The focus here is on the computation of the geoid, especially with regards to the treatment of permanent tide effects and the computation of height anomalies and geoid undulations referring to some specified ellipsoid of revolution and its normal gravity potential. Section 6.9 briefly discusses temporal (non-tidal) variations of the gravitational potential arising from the redistribution of mass within the Earth system, and the very recent advances in the monitoring and mapping of these variations from space, which resulted from the analysis of data from the GRACE satellite mission. Finally, Sect. 6.10 provides some outlook. This entire chapter is written in a way that focuses mostly on the concepts associated with global gravitational modeling, and the evolution of the art and science of the development of GGMs during the last 25 years or so. A rather extensive list of references is provided, so that the reader will be able to locate specific documents that provide the mathematical details associated with various aspects of GGM developments.

6.2 Introduction

A *Global Gravitational Model* (GGM) is a mathematical approximation to the external gravitational potential of an attracting body. We will focus here on the case where the attracting body is the Earth, although many of the concepts that we discuss apply equally well to other planets and celestial bodies. A GGM consists of a set of numerical values for certain parameters, the statistics of the errors associated with these values (as expressed, e.g., in their error covariance matrix), and a collection of mathematical expressions, numerical values, and algorithms that allow a user to perform:

1. **Synthesis**, i.e., computation of the numerical values of quantities related to the gravitational potential (*functionals* of the field), given the position of the evaluation point.
2. **Error Propagation**, i.e., computation of the expected errors of the computed functionals, as implied by the propagation of the errors of the parameters defining the GGM.

A GGM must be able to support such computations at arbitrary points, located on or above the Earth's surface, in a fashion that is both rigorous and efficient. In addition, a GGM should fulfill certain conditions stemming from the underlying physics. Namely, it should represent a scalar function of position that is harmonic outside the attracting masses and vanishes at infinity as the reciprocal of the distance between attracted point and attracting mass element. Moreover, the GGM should permit the computation of any functional of the field in a way that guarantees *self-consistency*. This means that the model should preserve the relationships (differential or integral) between the various functionals. A GGM has numerous uses, both operational and scientific (see also Tscherning 1983), including:

1. Orbit determination applications necessary for space surveillance (the detection, tracking, and orbit prediction of Earth-orbiting objects).

2. Inertial navigation applications for trajectory determination of airplanes and missiles.
3. Geoid undulation computations necessary to transform a geometric height, to an elevation referenced to an equipotential surface. This application has attracted great interest in recent years, because GPS positioning and gravimetrically determined geoid heights offer the possibility of determining orthometric heights and height differences without the need for the expensive and laborious spirit leveling (Schwarz et al. 1987).
4. Oceanographic applications that require the estimation of the Dynamic Ocean Topography (DOT) and its slopes, quantities that are directly related to ocean circulation. This application puts very stringent accuracy and resolution requirements on GGMs (Ganachaud et al. 1997).
5. A unique, accurate high resolution GGM may be used to provide the reference surface for the realization of a *Global Vertical Datum* (Rapp and Balasubramania 1992).
6. Geophysical prospecting applications where, in combination with other information (e.g., seismic data), a GGM may provide important constraints that aid the determination of underlying density distributions.

These and other applications represent integral parts of various civilian and military activities. Each of these applications has (in general) different accuracy and resolution requirements, as far as the supporting GGM is concerned. For example, due to the attenuation of the gravitational field with increasing altitude, a relatively low resolution GGM (e.g., a spherical harmonic expansion to maximum degree 70 or 90) is currently adequate for the precise orbit determination of most Earth-orbiting satellites. In contrast, accurate determination of the slopes of the equipotential surface (deflections of the vertical) demands a GGM of much higher resolution.

Geodesists have at various times developed "special purpose" models that optimize performance for a particular application (e.g., orbit determination of a particular satellite, or geoid undulation computation over a specific geographic region). Although such "tailored" models have found some uses in the past, the ultimate goal has always been the development of a *unique*, general purpose, GGM that addresses the different and diverse applications in an optimal manner, without over-performing in one application at the expense of its performance in others.

The development of a high-resolution GGM is a task that involves the optimal combination of a variety of data (satellite, land, marine, airborne). This is because a single data type with both global coverage and with uniformly high accuracy and high spectral sensitivity does not (yet) exist. The aforementioned data are of complementary character (in terms of spectral sensitivity and/or of geographic coverage), so that their optimal combination enables a GGM to satisfy the variety of applications described before. "Class" solutions of this type (e.g., EGM96) may include not only parameters that describe the gravitational potential, but also parameters that describe the Dynamic Ocean Topography, tides, Earth orientation and tracking station position parameters, as well as a plethora of "nuisance" parameters necessary to model completely the content of certain data types (e.g., biases and delays associated with certain satellite tracking data). The result of a successful GGM development effort is a model that can be used as a *standard* for numerous applications, over a substantial period of time.

6.2.1 Local and Regional Gravimetric Models

The accuracy and the resolving power of the data that were used in its development dictate the accuracy and resolution of a GGM. Geopolitical and/or proprietary issues many times prevent the individual or the team developing a GGM from having access to *all* the existing data. However, over some regions, data of higher accuracy and/or resolving power (geographically dense sets of gravity and elevation data) may be available to some individual(s) or may become available *after* the reference GGM has been developed. These data may be used in combination with the existing GGM to improve the accuracy and/or resolution of the determination of *one or more specific functionals of the field*, over the region where the detailed data became available. This local or regional "densification" can produce a specific local or regional gravimetric product or model.

Such densification has been among the favorite geodetic activities over many decades now, and represents the geodesist's way of creating a *multi-resolution* gravitational model resembling a "quilt": i.e., patches of fine detail (the *Local Gravimetric Models* – LGMs) are sewn on top of a more or less homogeneous piece of fabric (the reference GGM). *Geodesists do not necessarily have to re-evaluate the reference GGM every time a new set of data (a new patch) becomes available locally*. Such re-evaluation is mostly warranted if new and improved satellite data become available, spanning a sufficiently long time period, and/or if new terrestrial data (of higher accuracy and/or resolution) become available over areas with substantial geographic extent.

6.2.2 Global Versus Local Gravimetric Models: Similarities and Differences

It is useful to consider some important points related to the development and the nature of global and local gravimetric models.

- The most time-consuming, expensive, and laborious task in the development of both global and local gravimetric models is the data collection, validation, and pre-processing. In comparison, the time and effort required for the model estimation is almost negligible.
- Existing global gravitational models, developed using spherical harmonics as the representational basis, allow the computation of *any* functional of the field (geoid undulations, gravity anomalies, deflections of the vertical, second order gradients

of the potential) *anywhere* outside the attracting masses. These computed values are, of course, subject to commission and omission errors (see Sect. 6.6.2.4 for the definition of these terms). In contrast, currently available local or regional gravimetric models consist usually of geographic grids containing the estimated values of one or more *specific* functionals of the field (e.g., geoid undulations, deflections of the vertical), but *cannot* support the computation of arbitrary field functionals at arbitrary locations.
- Global gravitational models are accompanied by increasingly more complete and reliable error estimation. In contrast, existing local or regional gravimetric models are seldom accompanied by error statistics computed rigorously from the error estimates of the input data.
- Determination of a global gravitational model is *not* an interpolation problem. The gravimetric geoid surface is *not* directly observable. Multi-resolution representations that have been used in some studies to decompose and depict various levels of detail within a *given* geoid surface, address (at best) interpolation problems but fall short of addressing the much more challenging and important gravimetric estimation problems. The geodesist's main problem is how to determine the geoid (and other field functionals) from heterogeneous and noisy data – not how to interpolate it, once it has been determined.
- Determination of a local/regional model of a given gravimetric functional *may* reduce to an interpolation/extrapolation problem, if the functional of interest (e.g., gravity anomaly) is also observed *directly* within the area of interest.

The argument that the use of spherical (or ellipsoidal) harmonics as the representational basis for a GGM has the disadvantage that local data updates necessitate global updates of the model (re-computation of the GGM), would have been true if geodesists relied *only* on the GGM for *all* gravimetric applications. At present they do not. LGMs can be used to address efficiently local and regional data updates and applications. Furthermore, even if the re-computation of a GGM is required, due to some specific update of regional surface gravity data, such a re-computation can be done very efficiently and expeditiously at present, as long as the underlying satellite-only model does not have to be re-computed.

6.3 Signal Representation and Data Characteristics

Although geodesists have variously considered and studied the representation of the gravitational potential using point masses (Sünkel 1981b, 1983), finite element methods (Meissl 1981; Baker 1988) and splines (Sünkel 1984; Jekeli 2005), these approaches have seen only limited application in the representation of (especially) the "static" (i.e., time-averaged) gravitational field of the Earth. Spherical harmonics have prevailed as the standard form used for the representation of the gravitational potential globally, from the very early days of global determinations, to the present. Indeed, the set of coefficients of a spherical harmonic expansion of the gravitational potential has become pretty much synonymous to a GGM. Rapp (1998) provides a review of the geopotential modeling developments of the twentieth century, which includes an extensive list of references.

The Earth's external gravitational potential, V, at a point P defined by its geocentric distance (r_P), geocentric co-latitude (θ_P) and longitude (λ_P), can be expressed as:

$$V(r_P, \theta_P, \lambda_P) = \frac{GM}{r_P}\left[1 + \sum_{n=2}^{\infty}\left(\frac{a}{r_P}\right)^n \sum_{m=-n}^{n} C_{nm} Y_{nm}(\theta_P, \lambda_P)\right]. \qquad (6.1)$$

GM is the geocentric gravitational constant (the product of the universal gravitational constant, G, times the mass of the Earth including its atmosphere, M) and a is a scaling factor associated with the fully-normalized, unitless, spherical harmonic coefficients C_{nm} (a is usually numerically equal to the equatorial radius of an adopted mean-Earth ellipsoid). The surface spherical harmonic functions are defined as (Heiskanen and Moritz 1967, Sect. 1-14):

$$Y_{nm}(\theta_P, \lambda_P) = \overline{P}_{n|m|}(\cos\theta_P) \cdot \left\{ \begin{array}{c} \cos m\lambda_P \\ \sin|m|\,\lambda_P \end{array} \right\} \begin{array}{l} \text{if } m \geq 0 \\ \text{if } m < 0. \end{array} \qquad (6.2)$$

$\overline{P}_{n|m|}(\cos\theta_P)$ is the fully-normalized associated Legendre function of the first kind, of degree n and order $|m|$. In practice, the degree summation is truncated to some finite degree N, which depends on the resolving power of the available data. In turn, N defines (approximately) the resolution of the GGM. The goal of global high-resolution gravitational modeling is to estimate, as accurately as possible, the coefficients C_{nm}, through the optimal combination of gravitational information from a variety of data sources. Of equal importance is the estimation of reliable error estimates for the C_{nm} values. The estimated C_{nm} values can then be used to compute functionals of the field (e.g., geoid undulations, gravity anomalies, etc.) while their associated errors (and error correlation when available) can be propagated to yield the errors of the derived functional(s). *Before* the dawn of the new millennium and the availability of data from the satellite missions CHAMP and GRACE, four kinds of gravitational information were commonly available for the development of high-degree combination gravitational models like EGM96 (Lemoine et al. 1998):

1. Information obtained from the analysis of satellite orbit perturbations that are deduced from tracking data. This is of critical importance for the accurate determination of the low degree part of the model. *Satellite-only* models have progressed from solutions to degree 4 in the early 1960s, to models complete to degree 70 or 90 available at present. These advances were made through the availability of ever more accurate tracking data acquired over a continuously expanding constellation of Earth orbiters. Tracking data from approximately 40 satellites have been used in the development of the satellite-only solution supporting EGM96 (denoted EGM96S) (Lemoine et al. 1998). These data include

optical, radio Doppler and radio interferometric observations, Satellite Laser Ranging (SLR), Doppler Orbit determination and Radiopositioning Integrated on Satellite (DORIS), and Satellite-to-Satellite Tracking (SST) data from the Global Positioning System (GPS) and Tracking and Data Relay Satellite System (TDRSS) constellations to lower Earth orbiters. Despite these advances, these tracking data types are incapable of resolving the fine structure of the field, due to the attenuation of the gravitational signal with altitude. Moreover, the available satellites do not sample uniformly the range of inclinations and altitudes, which is a necessary condition for the de-correlation of the harmonic coefficients estimated from satellite tracking data only. This causes strong correlation especially among coefficients of higher degrees and necessitates the use of a priori constraints in the development of satellite-only models (Lerch et al. 1979).

2. Surface (land, marine, and airborne) gravimetric data that are in principle capable of resolving both long and short wavelength features of the gravity field. This however requires uniform global coverage with dense gravity data of uniformly high accuracy. The best available data sets circa 1996 (Kenyon and Pavlis 1996) represent information derived from over 4,000 sources of detail gravity data collected over several decades. The accuracy and density of point data vary substantially with geographic region, with extended regions (e.g., Antarctica) being practically void of gravity measurements. Gravity anomaly data are susceptible to various systematic errors (Heck 1990). These errors, in conjunction with the non-uniformity of coverage, degrade the long wavelength integrity of the gravitational information that can be extracted from surface gravimetry. Nevertheless, surface and airborne gravimetry presently provide the only data that can resolve short wavelength gravity features, especially over land areas. In addition, ship borne gravity measurements aid the separation of the geoid from the DOT signal when used in combination with satellite altimetry.
3. Satellite altimeter data have enabled an unsurpassed mapping of the field over the oceans, both in terms of accuracy and in terms of resolution. TOPEX/Poseidon (T/P) (Fu et al. 1994) (as well as its follow-on missions Jason-1 and Jason-2) routinely provides estimates of the Sea Surface Height (SSH) which, for the first time, are not significantly contaminated by radial orbit error (RMS radial orbit error at the $\pm 2\,\mathrm{cm}$ level). However, altimetric measurements are confined over the ocean areas bounded by the satellite's inclination, and furthermore provide a mapping of the sum of the geoid undulation plus the DOT. These aspects weaken somewhat the contribution of altimeter data in the determination of the long wavelength gravitational field and necessitate the appropriate modeling and estimation of the DOT when altimeter SSH data are used in combination solutions. There is however another way of incorporating altimeter data into a high-degree GGM, which is discussed next.
4. The combination of altimeter data from multiple missions, some of which have produced very closely spaced ground tracks, has provided a dense sampling of most of the ocean's surface. These data, in the form of SSH and/or SSH slopes, can be used to estimate ocean-wide sets of gravity anomalies, at very fine resolution (e.g., $2' \times 2'$ and $1' \times 1'$), as it is discussed in detail in Chap. 9.

Areal averages of these values can be merged with corresponding land and airborne gravity anomalies and gravity anomalies inferred from models of the topographic-isostatic potential (Pavlis and Rapp 1990), to produce a complete global equi-angular grid of gravity anomalies. The geometry of such grids allows the applicability of very efficient harmonic analysis (and synthesis) methods (Rizos 1979; Colombo 1981a), which have revolutionized the development and use of very high-degree spherical harmonic expansions. These approaches allow also efficient combination with *satellite-only* information, as was done, e.g., by Rapp and Pavlis (1990). However, incorporation of altimeter data into a GGM in this fashion requires some a priori knowledge of the DOT (or some other iterative approach – see Sect. 6.6.2.2), so that the altimetry-derived gravity anomalies are estimated from appropriately corrected SSH.

Satellite tracking, altimetric, and surface gravimetric data are of complimentary character both in a spectral as well as in a geographic sense. Their combination enables the determination of the gravitational field, over a wider band of its spectrum, with improved accuracy, than can be obtained by using any of the three data types alone. The particular means of combining these data, in order to develop a high-degree GGM, constitutes a solution strategy. A critical consideration in the design of a solution strategy is the treatment of altimeter data (Rapp 1993), i.e., if these data will be incorporated as in (3) or as in (4) above. OSU91A (Rapp et al. 1991) and EGM96 (Lemoine et al. 1998) represent the result of implementing a particular solution strategy, whereby altimeter data were used as in (3) for the determination of the low degree part of these models (maximum degree 50 and 70 respectively), and as in (4) for the higher degree part. The main disadvantages of this strategy are that: (a) the high degree GGM is obtained in a “piece-wise” fashion and, (b) a complete error covariance matrix exists only for the low degree portion of the model. N.K. Pavlis in (Lemoine et al. 1998, Chap. 8) discussed specific reasons for the selection of that particular estimation strategy. Certain characteristics of the above data types that are particularly important for their effective combination are discussed next.

- ***Information Content.*** The observables within the above four categories contain information not only about the gravitational field, but also about numerous other effects. Some of these effects are of interest in their own right (e.g., the DOT information contained within altimetric SSH), while others represent, at least as far as gravitational modeling is concerned, (more or less) systematic noise (e.g., the non-conservative forces acting on a satellite). In either case, effective incorporation of a particular data type into the combination solution requires precise modeling and optimal estimation of all the effects and signals contained within the observable. Otherwise, the estimated gravitational model can be severely corrupted by the mis-modeled (or un-modeled) systematic effects.
- ***Spectral Sensitivity Overlap.*** The development of a GGM through a least-squares adjustment combining different data types is meaningful, provided that the data used in the adjustment share some common degree of sensitivity to the gravitational signal over a certain portion of its spectrum (a range of

harmonic degrees). Otherwise, there is little "adjustment" being performed to data representing disjoint spectral bands. For example, existing satellite-only models have a narrow spectral sensitivity overlap with models recovered from surface gravity data alone. This complicates considerably the problem of optimal combination of these two data types. On the other hand, this also means that setting up and inverting extremely large linear systems corresponding to very high degree models may not be necessary, if a single data type (e.g., a complete global equi-angular grid of gravity anomalies) uniquely determines the higher degree portion of such a GGM.

- ***Relative Weighting.*** The optimal estimation of a GGM depends critically on the optimality of the relative weights assigned to the different data types. Considering the numerous sources of data that are involved, this is a very large *component of variance* estimation problem, complicated further by the fact that the extraction of gravitational information from satellites' orbit observations is a strongly non-linear problem. Although approximate solutions to this relative weight estimation problem have been used with considerable success (Lerch 1991), many times the experience and intuition of the model developer(s) guide the selection of appropriate data weights more than anything else.

6.4 The New Satellite Missions

The satellite data used for the development of all GGMs published by the end of the twentieth century represent tracking of "targets of opportunity", i.e., of spacecraft designed and equipped with instrumentation for applications other than the mapping of the gravitational field from space. As a result of three satellite missions, this situation has changed dramatically during the last few years. These three missions are CHAMP (Rapp et al. 1996), GRACE (Grace 1998), and GOCE (ESA SP-1233 1999). Table 6.1 summarizes the main characteristics of these missions.

A nice discussion regarding the concepts involved in these three mission scenarios can be found in (ESA SP-1233 1999, Sect. 2.3). Mapping of the gravitational field from space requires missions that adhere as much as possible to the following fundamental design constraints:

- Uninterrupted tracking in three spatial dimensions.
- Measurement or compensation of the effects of non-gravitational forces.
- Orbital altitude as low as possible, to enhance sensitivity to the gravitational signal, and inclination as high as possible, to permit (near) global coverage.
- Counteraction of the field's attenuation at altitude through the measurement of derivatives of the potential.

All three missions above have in common the high-low Satellite-to-Satellite Tracking component (SST-*hl*) from the GPS (and GLONASS in the case of GOCE) constellation, and the measurement of non-gravitational forces by the on-board accelerometers. These data permit highly accurate orbit determination for all three

Table 6.1 Main characteristics of three satellite missions

Mission	Status	Orbit	Mission objective	Instrumentation, tracking, and comments
CHAMP	Launched on 7/15/2000 Active	Alt. = 450 km $e \approx 0.004$ $i = 87°$	Gravity and Magnetic fields Atmospheric Limb Sounding Ionosphere Sounding	3-axis STAR accelerometer GPS and SLR Altitude will decay from 450 km (BOL) to 300 km (EOL)
GRACE	Launched on 3/17/2002 Active	Alt. = 485 km $e \approx 0.001$ $i = 89°$	Gravity field and its temporal variation	3-axis accelerometers (1 per s/c) GPS and SLR K-band inter-satellite ranging between the 2 s/c
GOCE	Launched on 3/17/2009 Active	Alt $\approx$ 250 km $i = 96.7°$ Sun-Synchronous	Gravity field (Especially static)	Six 3-axis accelerometers forming the gradiometer GPS/GLONASS and SLR

missions, and in addition may enhance the gravitational field determination at very long wavelengths (very low degrees). In addition to that, GRACE involves the continuous measurement of the range between two identical satellites that "chase each other", which constitutes a low-low SST formation (SST-*ll*). GOCE's accelerometer array on the other hand provides the measurements necessary to determine the gravitational tensor (i.e., the 3×3 matrix of second order spatial derivatives of the gravitational potential) at altitude. GOCE is unique in the sense that it will provide boundary data at altitude covering the entire Earth, *except* for two polar caps of $\sim 6.7°$ radius (due to the satellite's inclination). The data from each of these three missions result in different sensitivities to the gravitational spectrum. Simulation studies examining the geopotential recovery attainable from these (and other) mission scenarios were reported e.g., by Sneeuw and Ilk (1997). Figure 6.1 depicts the degree amplitude spectra (square root of the degree variance) of the geoid undulation signal and its error as predicted from EGM96S and from CHAMP, GRACE, and GOCE mission simulations. In the same figure an estimate of the degree amplitude spectrum of the DOT and of geoid undulation effects predicted from a postulated model of vertical datum inconsistencies are shown. The latter is just one of several systematic error sources possibly affecting terrestrial gravity anomaly data (Heck 1990), but not necessarily the dominant source of error, as an analysis by Pavlis (1988) indicated.

Two main questions arise when considering the data from these satellite missions:

1. What is the optimal way of analyzing the data from these missions?
2. What is the optimal way of combining their data with existing data, e.g., from surface gravimetry and from satellite altimetry, in order to develop high-degree combination gravitational models?

1. ***Data Analysis.*** In the case of CHAMP the gravitational information is extracted from the analysis of the perturbations of a low Earth orbiter, in a fashion similar to other existing satellite missions. However, CHAMP's low orbit, in conjunction

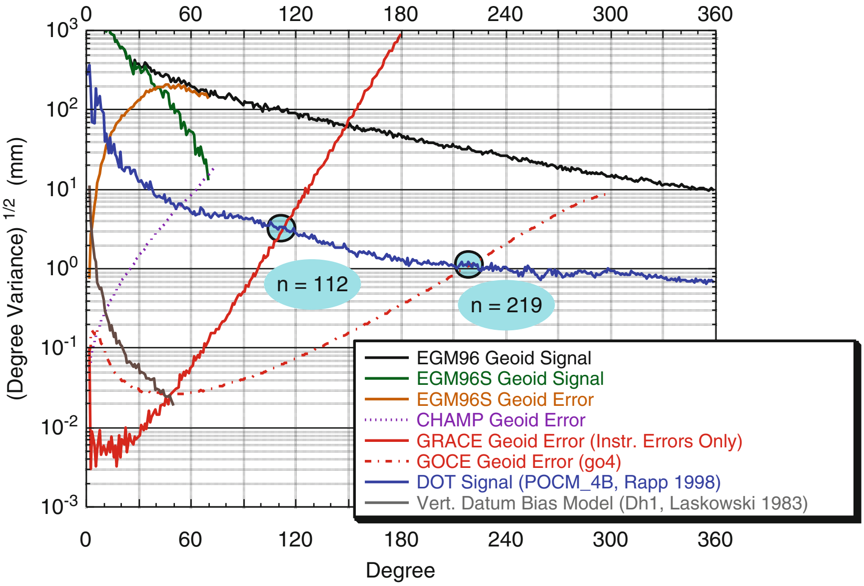

Fig. 6.1 Degree amplitude spectra

with the accelerometer data and with the availability of nearly global tracking data coverage, enabled *for the first time* the determination of an accurate long wavelength global gravitational model from a *single* satellite's data. Indicative of this "new state of affairs" is that a very preliminary solution (complete to degree and order 91) was already developed based on a *single* month's worth of CHAMP data *only* and was presented during the 2001 meeting of the International Association of Geodesy (IAG) by Reigber et al. (2001). Although significantly better models that include GRACE data have by now surpassed considerably this preliminary solution, it served as a good example of the improvements that were to follow.

Compared to CHAMP, GRACE added the SST-*ll* component, which permitted higher resolution gravitational information to be extracted from the analysis of the orbital perturbation *differences* along the line-of-sight of the two low orbiting satellites. One can use traditional orbit perturbation analysis methods to process the GRACE data and derive a GGM, e.g., in spherical harmonics. GGM01S (Tapley et al. 2004) and GGM02S (Tapley et al. 2005) were estimated following such a procedure. This analysis scenario, albeit costly, is within current computational capabilities for models extending to degree and order 180 or so. Geodesists have also considered alternative analysis methods for GRACE-type missions (e.g., Wolff 1969; Colombo 1981b; Jekeli 1999a; Rowland et al. 2002). Such methods provide higher computational efficiency at the cost of committing certain approximations. Luthcke et al. (2006) reported monthly gravitational solutions determined from GRACE inter-satellite range-rate data *alone*, which

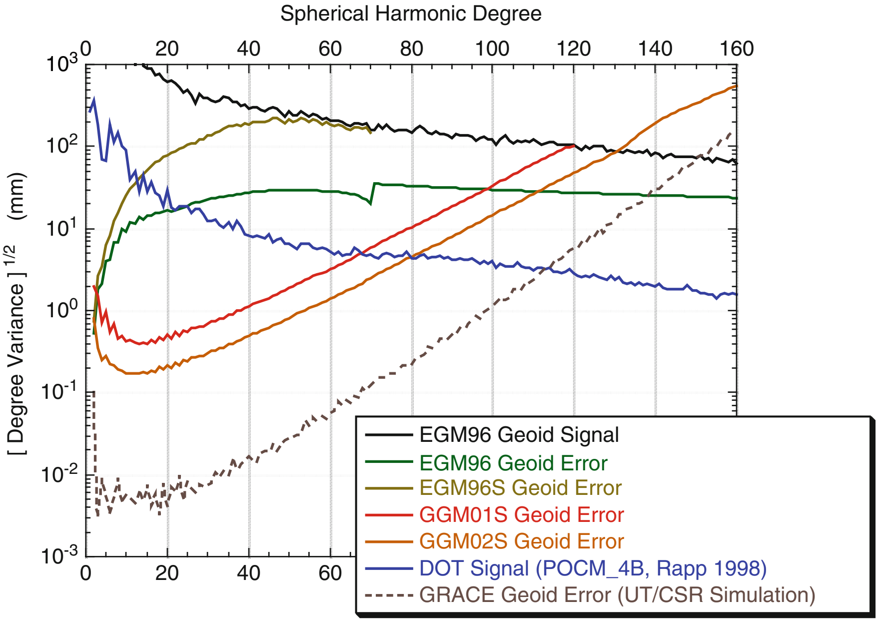

Fig. 6.2 Degree amplitude spectra comparing actual (GGM01S, GGM02S) versus simulated GRACE performance

are significantly less affected by certain systematic errors ("stripping") than those solutions that incorporate simultaneously the GPS data (SST-*hl*) in their development. Setting aside for a moment the details of optimal GRACE data processing, it is important to recognize here the quantum leap that has been accomplished with the GRACE mission. Approximately 14 months of GRACE data *alone* have been used to develop GGM02S, whose cumulative geoid undulation error to degree 70 is less than ±1 cm (Tapley et al. 2005, Fig. 6.2). By comparison, the cumulative geoid undulation error to degree 70 for EGM96, which required the combination of data from tens of satellites, along with surface gravity and satellite altimetry, was ±19 cm (Lemoine et al. 1998, Table 10.3.2-1).

As far as GOCE is concerned, numerous investigations of the various aspects of its data analysis and of the development of a GGM from them can be found in ESA (2000). One particular issue that receives increased attention relates to the polar gaps and their impact on analysis schemes that exploit regularity and completeness in the data coverage (e.g., block-diagonal normal equation formation schemes).

2. ***Data Combination.*** While the existing satellite gravity-mapping missions (CHAMP and especially GRACE) have already delivered (or promise to deliver as in the case of GOCE) quantum leaps in the accuracy and resolution of the *satellite-only* gravitational models (see Fig. 6.1), there is still a need to combine that information with terrestrial gravity and satellite altimetry data in an

optimal fashion. This is required so that a seamless extension of the gravitational spectrum can be achieved, taking advantage of the rich high frequency content of the surface and altimetric data. The higher resolution of GRACE- and GOCE-based *satellite-only* models will significantly increase the spectral overlap with surface gravity and altimetry. This will enable for the first time the estimation of high-resolution models of the DOT and of the significant reduction of systematic errors present in surface gravity data. Whether these estimates will be obtained within *comprehensive* combination solutions (i.e., solutions where gravitational potential coefficients are estimated *simultaneously* with several other parameter sets), or using some other approach is (to some extent) still under investigation. Comprehensive solutions have the advantage that they provide a complete error covariance matrix associated with *all* estimated parameters. They are however computationally demanding, as we elaborate in the next paragraph.

Let us focus for a moment on the parameter sets corresponding to potential coefficients, DOT, and systematic surface gravity errors. Let us also assume for the sake of this argument that all three sets will be represented using spherical harmonic coefficient sets complete to degrees M (potential), L (DOT), and K (surf. gravity errors), respectively. One can deduce approximate values for M and L (e.g., from Fig. 6.1), but not for K. The spectral (and geographic) behavior of the surface gravity systematic errors and their maximum resolvable degree will have to be estimated *directly*, e.g., from a preliminary comparison of surface gravimetry with the information implied by GRACE and GOCE. In contrast to the DOT, for which spectral and geographic estimates exist e.g., from the analysis of oceanographic models (see Fig. 6.1), and can be used to guide the selection of L, only very limited information exists regarding the spectral and geographic behavior of surface gravity errors (Pavlis 2000). There are good reasons to try to estimate these surface gravity systematic errors optimally, and also to try to explain their origin. Such estimates may reveal not only problems associated with the gravity anomaly (and/or elevation) data themselves, but also possibly problems related to the (pre-) processing and modeling of gravity anomaly data (e.g., analytical continuation). Furthermore, direct estimation of these effects may help resolve some outstanding questions related to the weighting of surface gravity relative to the satellite information, and should shed new light on the spectral distribution of surface gravity errors (at least to the degree resolvable by the satellite information). To illustrate the challenges implied by the new satellite missions, we consider here the development of a (*hypothetical*) GGM, following a procedure similar to that used for the development of the degree 70 part of EGM96 (see Lemoine et al. 1998, Chap. 7). With spherical harmonics as the representational basis for the three parameter sets discussed above, each of the new satellite missions implies certain values for M, L, and K. Using Fig. 6.1 as our guide (at least for M and L), we conclude that:

- This combination scenario can be readily implemented in the case of CHAMP.
- In the case of GRACE, such a scenario implies $M \approx 140$ (potential coefficients), $L \approx 110$ (DOT coefficients), and a K value that will probably not exceed ~ 90

(surf. gravity error coefficients). The total number of parameters to be estimated in this case will be $O(40\,\mathrm{k})$. Despite its size, this problem is well within our current computational capabilities, as Pavlis and Kenyon (2003) have already demonstrated. One should also notice that using the error curves of Fig. 6.1 to estimate M, L, and K may result in values that are larger than those implied by the *actual* (as opposed to the *simulated*) performance of these missions. For the case of GRACE, Fig. 6.2 allows a comparison between *actual* (GGM01S, GGM02S) and *simulated* performance.

- Treating the GOCE case however in the same fashion could result in parameter sets that will be approximately four times as large as the set for GRACE. Combination of GOCE data with surface gravity and altimetry, in a comprehensive least-squares adjustment fashion, may require the development of new innovative approaches that economize on number of parameters and/or result in patterned normal equation systems that adhere to efficient formation and inversion algorithms. Whatever these techniques might be, they should allow also the efficient computation of the error variance and covariance of the various recovered fields (and of their functionals), as a function of geographic location, as well as in a spectral form.

6.5 Beyond the Sensitivity of Satellite Data

There is obviously a limit to the gravitational information that can be extracted from space-borne sensors, which implies a limit to the resolution (maximum degree) of satellite-only models. Observations made on the Earth's surface (or on airplanes flying at low altitudes) can extend the resolution of gravitational models considerably. Surface and airborne data like gravity anomalies, gravity disturbances, etc. are therefore capable of supporting the development of much higher resolution gravitational models, than those developed based on satellite data only. One way of developing such high degree models (the way used to develop the degree 71–360 part of EGM96) involves first the formation of a regular grid of areally averaged values of some functional of the field that covers completely the globe. These values represent data derived on the basis of other primary observables, using techniques like Least Squares Collocation (LSC). Surface gravity anomalies are a suitable choice for such a functional, both from a spectral sensitivity and from an availability viewpoint. Very high degree and order spherical harmonic models are then developed from the analysis of such global anomaly grids, using efficient harmonic analysis techniques like those put forward by Colombo (1981a). Wenzel (1998, 1999) reported such expansions complete to degree and order 1,800. Pavlis et al. (2005) reported the development and evaluation of the PGM2004A preliminary gravitational model, extending to maximum degree and order 2,160, and identified some aspects related to its development that required further improvement. Computational efficiency was *not* one of them. An expansion to degree 2,160, given a complete $5' \times 5'$ grid of area-mean gravity anomaly values required approximately 30 min to execute on a Sun Fire v480 workstation with four 6.2 GHz Ultra SPARC III processors, and this time includes the computation of the values of the integrals of Associated Legendre functions. Expansions of this kind can augment models developed via comprehensive combination adjustments, thereby defining "composite" high degree GGMs (like EGM96). Certain aspects of such "composite" models may be criticized. These aspects are discussed next.

1. ***Piece-wise nature.*** The fact that "composite" models (as EGM96) are not developed via a single combination solution least-squares adjustment is considered a drawback of this approach. But is it really necessary to strive for such a single step adjustment? One should recognize that beyond the maximum degree resolvable from satellite sensor data, the gravitational information is *uniquely* determined from surface data (including altimetry-derived gravity anomalies). This implies (as mentioned before) that no "adjustment" is really being performed over this high degree spectral band; therefore a single step approach may be more of a nicety rather than a necessity. What is necessary though is that the transition from the low to the high degree spectral band is seamless in terms of both signal and error. In essence, this piece-wise approach, with spherical (or ellipsoidal) harmonics as the representational basis, may be viewed as a (limiting) case of a "remove-compute-restore" gravimetric approximation that is performed globally and in the spectral domain. In this case, the "remove" step corresponds to the low-pass filtering of surface gravity and satellite altimetry data using a preliminary high-degree expansion, which is done to minimize aliasing effects (see Pavlis 1988, Sects. 5.2.4, 5.2.5); the "compute" step refers to the (relatively) low degree part of the field that is developed through the comprehensive combination solution; and finally the "restore" step corresponds to the augmentation of the low degree combination solution with the high-degree expansion coefficients.
2. ***Lack of complete error covariance matrix.*** This is a critical shortcoming of the currently available high resolution GGMs. Using spherical harmonics as the representational basis implies that in order to obtain propagated error estimates, *with geographic specificity*, for derived functionals of the field, one has to form and propagate complete error covariance matrices corresponding to the maximum degree of the model. Clearly this is a very computationally demanding proposition for existing models (to degree and order 360), let alone ultra high-degree expansions like EGM2008 (Pavlis et al. 2008) that will be discussed in more detail in Sect. 6.6.2. A model complete to degree and order 2,160 involves $\sim$4.7 million coefficients. This would be the dimension of the (symmetric) error covariance matrix that needs to be formed, to allow conventional error propagation. A much more efficient solution to this problem has been developed and presented initially by Pavlis (2005). This technique recognizes again that beyond a certain degree M (corresponding to the resolution of the *satellite-only* model), the GGM is uniquely determined from surface gravity (terrestrial, airborne, and altimetry-derived) data. This implies that complete error covariance matrix propagation may *only* be necessary for the portion of the model up to degree M. Using band-limited kernels (beyond degree M and up to the

maximum degree of the model) within integral formulas, one is able to compute the error contribution of the harmonics beyond M in a much more efficient manner, through global convolutions. The propagated error components for the different spectral bands are subsequently added (in a quadratic sense), assuming no data error correlation across the different spectral bands.

3. ***Data pre-processing requirements.*** The development of high degree GGMs using the procedures discussed previously does require a considerable effort for the pre-processing of available surface (land, marine, and airborne) data, as well as for the prediction of altimetry-derived gravity anomalies. The harmonic analysis approaches developed by Colombo (1981a) are best suited for the analysis of complete grids of a *single* data type, referring to a surface of revolution (e.g., a rotational ellipsoid). Furthermore, the efficiency of such estimators is based in part on rather strong assumptions concerning the signal and error covariance functions of the data (homogeneity and isotropy). Since the actual measurements do not comply with such configurations in general, several pre-processing steps are required to transform the primary observables to quantities that adhere to the requirements of the estimator (at least to a certain degree of approximation). LSC prediction of area-mean values of gravity anomalies from (the combination of) measurements acquired on land, sea, and air, aims to produce a *single* data type out of the several types of (possibly overlapping) measurements that may be available over a given area. The same technique is used to derive gravity anomalies from dense sets of altimetric SSH data. Analytical continuation aims to artificially reduce these gravity anomalies to quantities that refer to a surface of revolution (e.g., the reference ellipsoid). These artificial quantities, when analytically continued in the opposite direction, are supposed to reproduce the input gravity anomalies. The prediction of area-mean gravity anomalies on a regular grid, and their analytical continuation, produce derived data that adhere to the *geometric* requirements of the estimator. LSC, which could be used to derive a GGM without much need for pre-processing of the *original* data, requires the formation and inversion of a matrix whose size equals the number of observations. This is an impossible task for the foreseeable future. *Efficient* techniques that can make use of the *original* data with minimal pre-processing requirements are (still) desirable.

The treatment of the *stochastic* properties of the data is even more complex than the treatment of *geometric* requirements. Availability of error variance estimates is pretty much the best that an analyst can hope for, and most times these estimates reflect data precision rather than data accuracy. The majority of gravimetric approximation studies (both global and local) either neglects completely any error correlation between the data, or attempt to account for it in empirical (many times not well justified) ways. It seems reasonable (at least to this author) to consider the error of currently available area-mean surface gravity anomaly data as composed of two main components: (a) a long wavelength component originating from systematic errors e.g., in the base network, the "ties" to it, long wavelength errors in elevations, etc., and, (b) a short wavelength component reflecting things like the accuracy and density of local data, the local roughness of the field and of the topography, etc. The former component is expected to have relatively low standard deviation (order of ±2 mGal) but very long correlation lengths (continental or even global scale); the latter may have standard deviations that in certain regions exceed ±30 mGal, but its correlation lengths are expected to be short (a few tens of km). The best way to account for the long wavelength component of these errors is probably by direct estimation of systematic errors in combination solutions with *satellite-only* models from missions like GRACE and GOCE. How to treat the short wavelength component with rigor and efficiency, both in the estimation of a GGM and in the subsequent error propagation, remains still an open question.

6.6 State-of-the-Art Global Gravitational Modeling

In this section we discuss the main aspects of the development of two global gravitational models, representative of the state-of-the-art at the respective time of their development: EGM96 (Lemoine et al. 1998), which represents the state-of-the-art before the availability of data from CHAMP and GRACE, and EGM2008 (Pavlis et al. 2008; 2012), which represents currently (2010) the model with the highest resolution and accuracy, prior to the anticipated availability of data from GOCE. The choice of these two models also permits a comparison of the approaches followed in their development. Such a comparison reveals the critical changes in model development, which were brought about by the availability of GRACE data on one hand, and of high quality $5' \times 5'$ area-mean gravity anomalies (from the combination of terrestrial and altimetry-derived data sources) on the other.

In theory, the estimation of a combination solution complete to some (arbitrary) high degree and order could be carried out as follows:

(a) Form separate normal equations from each individual data type, to a maximum degree and order that corresponds to the resolution of the available data and their sensitivity to the gravitational signal.
(b) Treat satellite altimeter data as "direct tracking" observations, i.e., ranges from the spacecraft to the ocean surface whose upper endpoint senses (through the orbit dynamics) attenuated gravitational signals (static and time-varying), while their lower endpoint senses the combined effects of geoid undulation, DOT as well as tides and other time varying effects, without any attenuation. In this manner, altimeter data contribute to the estimation of the satellite's orbit, as well as the estimation of the DOT and of the potential coefficients.
(c) Combine the various normal equations (with appropriate relative weights) and invert the resulting system, to estimate the combination solution to its high degree, along with its full error covariance matrix.

Such an "ideal" approach would permit the most rigorous modeling of the observables and would allow the greatest flexibility in terms of data weighting. A combination solution to degree 360, if performed as outlined above, would require the formation of full (symmetric) normal matrices (from satellite altimetry and surface gravimetry) for approximately 130,000 parameters (considering *only* the gravitational potential coefficients). For maximum degree 2,160, there would be approximately 4.7 million such parameters involved. Such computational tasks are beyond our present computational capabilities. Therefore, at present, one may choose between, or combine, two main solution strategies to attack the problem:

- ***Solution Strategy (A)*** Apply the "ideal" estimation strategy outlined above, to obtain a combination solution for the lower degree part of the field, up to a maximum degree that is computationally manageable. Apart from reasons of computational feasibility, this maximum degree should enable the appropriate modeling of the gravitational signal contained in the currently available satellite tracking data. Furthermore, since the DOT signal is of long wavelength nature, the benefits of "direct" altimetry are almost entirely retained here. To avoid aliasing effects however, the contribution to the altimetry and surface gravity data from the coefficients beyond the solved-for degree has to be filtered out of the data prior to the normal equation formation. This may be done using a pre-existing high-degree solution (Pavlis 1988; Denker and Rapp 1990). Hereon, we will refer to this type of solution as the *comprehensive* low degree combination model. The obvious shortcomings of this approach are the relatively low maximum attainable degree (approximately 200 at present) and its computational demands. Some models developed using this approach (or similar ones) include JGM-1 and JGM-2 (Nerem et al. 1994) and JGM-3 (Tapley et al. 1996), the part of EGM96 up to degree and order 70 (see Lemoine et al. 1998, Chap. 7), and EIGEN-GL04C (Förste et al. 2008).
- ***Solution Strategy (B)*** Consider that one is willing to make the following two approximations:

  (i) The orbits of the altimeter satellites, whose data are included in the combination solution, are perfectly known (at least radially). This approximation is justifiable if one is working with altimeter satellites supported by T/P-class precise orbit determination. Moreover, after the availability of GRACE-based gravitational models for precise orbit determination of altimeter satellites, errors arising from gravitational model inaccuracies do not dominate the orbit error budget of these satellites. Errors due to, e.g., mis-modeling of non-conservative forces acting on the spacecraft are likely to be more significant nowadays. In this regard, to allow the orbits of altimeter satellites to contribute (through their dynamics) to the determination of gravitational parameters within a combination solution may not be a desirable approach nowadays, because the effects of orbit errors of non-gravitational origin could corrupt the solved-for gravitational parameters.
  (ii) The DOT is known a priori, e.g., from an Ocean Circulation Model (OCM) or from a previous low degree comprehensive combination solution.

The implication of (i) is that satellite altimetry does not have to be treated as "direct" tracking anymore, which simplifies the problem considerably, since now orbit dynamics are not involved in the altimeter data processing. One is left with a "surface" problem, where the geoid (N) and the DOT (ζ) signals have to be separated, given the "observed" SSH (h), which is their sum. If in addition, the approximation (ii) is introduced, then altimetry contributes to the combination solution "observed" geoid heights (N) over (parts of) the ocean.

In addition to the above two approximations, a key issue here is that altimetric information may also be provided in a *gridded* form. This is possible through the use of a Mean Sea Surface (MSS), obtained from multiple altimetric missions. The success of T/P has significantly improved the accuracy of MSS data sets (especially at long wavelengths). This is accomplished by adjusting the SSH data from other altimetric missions (e.g., ERS-1, ERS-2, GEOSAT, SEASAT), to the surface defined by T/P, using cross-over minimization techniques. Such MSS data sets have been developed by, e.g., Yi (1995), Kim et al. (1995), Anzenhofer et al. (1996) and Wang (2001), and more recently by Andersen et al. (see Chap. 9 for details). One may also have available gridded, altimetry-derived gravity anomaly values.

Such values have been estimated using various techniques, on an ocean-wide basis by, e.g., Rapp and Basic (1992), Andersen and Knudsen (1998), Trimmer and Manning (1996) and Sandwell and Smith (1997, 2009) among others.

The two simplifying approximations discussed above and particularly the availability of altimetric information in gridded form (especially in the form of gravity anomalies), make applicable an alternative class of high-degree combination solution techniques. These, combine the satellite-only information, with potential coefficient information obtained from the analysis of *complete*, regular grids of functional(s) of the disturbing potential (e.g., N, Δg), and are based on the highly efficient harmonic analysis algorithms originally studied and put forward by Colombo (1981a). These algorithms exploit the regularity of the data grids and the symmetry properties of Legendre and trigonometric (sine/cosine) functions. Using Fast Fourier Transform (FFT) techniques, one may process data arrays residing over latitude bands that are symmetric with respect to the equator, in a highly efficient manner. Estimators of this type are the (simple) Numerical Quadrature (NQ), the Block-Diagonal (BD) least-squares adjustment, and the Optimal Estimation (OE) technique. Models developed using the NQ approach include OSU86E/F (Rapp and Cruz 1986a) and OSU89A/B (Rapp and Pavlis 1990). BD techniques of varying sophistication have been used to develop GPM2 (Wenzel 1985), DGFI92A (Gruber and Bosch 1992), GFZ95A (Gruber et al. 1996), and EGM2008 (Pavlis et al. 2008). OE was used to develop the OSU86C/D models (Rapp and Cruz 1986b).

In the following sections we discuss in some detail the development approaches used for EGM96 and EGM2008 respectively.

6.6.1 EGM96

The two solution strategies (A) and (B) discussed above have their respective advantages and disadvantages. EGM96 (Lemoine et al. 1998) employed a comprehensive solution to degree 70, augmented by a BD solution from degree $n = 71$–359, while the $n = 360$ coefficients were obtained from a NQ model. In the following sections we describe in some detail the "building blocks" that were used to form the EGM96 high-degree model. Although EGM96 has by now been surpassed in terms of performance by more recent models like EGM2008, its development strategy still serves as a didactic example of the particular techniques that were used to model and combine optimally the data that were available at the time of its development.

6.6.1.1 The EGM96S Satellite-Only Model

The estimation of potential coefficients from satellite tracking data is a non-linear problem that involves the simultaneous estimation of the orbit, tracking station coordinates, tide parameters, polar motion and Earth rotation parameters as well as numerous nuisance parameters which may be measurement type or satellite specific (e.g., measurement biases and drifts, atmospheric drag and solar radiation pressure scale factors, etc.). The problem is further complicated by the fact that each satellite samples the gravitational field effects in a particular manner, dictated by its orbital characteristics (altitude, inclination, eccentricity) and the type of tracking data (e.g., Doppler versus ranges). Empirical acceleration parameters that may be necessary to estimate accurate orbits, many times absorb useful gravitational signal as well, so the analyst has to make appropriate trade-offs with extreme care, to ensure an optimum solution. A satellite-only solution involves the processing of tracking data segmented initially by "arcs" of various time spans depending on the satellite and the tracking data type. Once the estimation of the initial state parameters for an orbital arc has converged, normal equations for all the parameters (arc-specific and common) are formed. EGM96S involved the formation of approximately 2,000 such normal equation sets. These were subsequently combined by satellite and/or measurement type, while arc-specific parameters were successively eliminated through back-substitutions. Thus, one was left with "combined" normal equations, which now involved only the parameters common to all satellites and all data types. In EGM96S, this process resulted in approximately 40 sets of "combined" normal equations, which involved $\sim$12,300 parameters. Addition of these normal equations (appropriately weighted) resulted in a single, final set of normal equations. Its inversion defined the satellite-only model and its associated error covariance matrix.

The most critical aspect in this combination of normal equations is the weight assigned to each set of them. In a relative sense, weights should be such that the solution does not "over-fit" any particular satellite/data type at the expense of the others. In an absolute sense, they should yield an a posteriori error covariance matrix, which would accurately reflect the quality of the model. To "calibrate" the weights one may use the subset solution technique of Lerch (1991). One data type (or satellite) at a time is withheld from the solution, and the changes of the potential coefficients are compared to the changes predicted by the corresponding formal error estimates (complete versus subset solution). Weight calibration is a time consuming, iterative task and requires one to start with a preliminary set of weights, which should be close enough to the optimal set, to ensure convergence and minimize the number of iterations needed to achieve it. The experience of the analyst is indispensable here. This subset solution technique ensures primarily the internal consistency of the solution. Comparisons with external data, independent from the satellite-only model, provide the best means to test the reliability of the propagated error estimates of the model, in an absolute sense.

Particularly valuable to the development of EGM96S were SST data (high-low mode) from the GPS satellites to T/P, EUVE and GPS/Met, as well as TDRSS tracking of EUVE. These data provide continuous, precise tracking of the low orbiter and are more sensitive to high frequency geopotential effects, than traditional (pre-CHAMP) tracking data types. Calibration of the weights of these data proved to be a particularly challenging task.

The development of an accurate and well-calibrated satellite-only model is the most critical (and arguably the most complicated) part of the combination model development. Satellite-only models that were developed before the dedicated gravity-mapping missions (CHAMP, GRACE, and GOCE) include EGM96S (Lemoine et al. 1998) and GRIM4-S4 (Schwintzer et al. 1997).

It is important to recognize here that the normal equation matrices associated with these pre-GRACE satellite-only models were fully occupied, and rather ill-conditioned due to high correlations present among the coefficients of higher degrees. The inversion of these matrices usually required the use of some a priori constraint in the form of a power law (e.g., Kaula's rule). In order to preserve the integrity of the least-squares adjustment used to derive the combination solution, one had to consider the satellite-only normal equations in their complete (fully-occupied) form. Any block-diagonal approximation of these normal equations would result in estimation errors that could not be tolerated. This situation changed dramatically with GRACE, due to its global coverage and uniform data accuracy, which simplified the development of combination solutions dramatically, as we will see when we discuss the development of EGM2008.

6.6.1.2 The EGM96 Comprehensive Low-Degree Combination Solution

This solution involves the combination of the final satellite-only normal equations with normal equations developed from terrestrial gravity data and from satellite-altimeter data treated as "direct" tracking. Since altimetry enters here as direct tracking, and since the surface gravimetric data are introduced as a totally independent data type (i.e., no error correlation between the surface gravity, altimetry and satellite tracking data is considered), the surface gravity normal equations have to be developed based on gravimetric information independent of both the tracking and of the altimeter data. This requires the exclusion of any altimeter-derived anomalies from the file used to develop the surface gravity normal equations. The requirement for independence from the tracking data is slightly violated because of the way that "fill-in" anomalies are computed. In the following sections we describe the development of the surface gravity and altimetry normal equations.

The Surface Gravity (Low Degree) Normal Equations

The gravity potential of the Earth, W, is defined to be the sum of the gravitational potential, V, given in (6.1), plus the centrifugal potential Φ arising due to the rotation of the Earth. Consider the gravity potential U of a rotating equipotential ellipsoid of revolution (*Somigliana-Pizzetti* normal field). The disturbing potential $T(r_P, \theta_P, \lambda_P)$ is defined as (Heiskanen and Moritz 1967, Eq. 2-137):

$$T(r_P, \theta_P, \lambda_P) = W(r_P, \theta_P, \lambda_P) - U(r_P, \theta_P). \tag{6.3}$$

Due to (6.1) we have:

$$T(r_P, \theta_P, \lambda_P) = \frac{GM}{r_P} \sum_{n=2}^{\infty} \left(\frac{a}{r_P}\right)^n \sum_{m=-n}^{n} C_{nm} Y_{nm}(\theta_P, \lambda_P), \tag{6.4}$$

where we have assumed that the ellipsoid has the same mass and rotational speed as the actual Earth, and is centered at the Earth's center of mass. In (6.4), C_{nm} are now the remainders of the coefficients appearing in (6.1), after subtraction of the even degree zonal harmonic coefficients of the normal gravitational potential. Consider a quantity Δg^c that fulfills:

$$\Delta g^c = -\left(\frac{\partial T}{\partial r}\right)_Q - \frac{2}{r_Q} T, \tag{6.5}$$

where Q is a point on the *telluroid* (Heiskanen and Moritz 1967, p. 292). Substitution of (6.4) into (6.5) yields:

$$\Delta g^c(r_Q, \theta_Q, \lambda_Q) = \frac{GM}{r_Q^2} \sum_{n=2}^{\infty} (n-1) \left(\frac{a}{r_Q}\right)^n \sum_{m=-n}^{n} C_{nm} Y_{nm}(\theta_Q, \lambda_Q). \tag{6.6}$$

The quantity Δg^c is related to the *Molodensky* surface free-air gravity anomaly $\Delta g \quad (= |\vec{g}_P| - |\vec{\gamma}_Q|)$, obtained from scalar gravimetry, by:

$$\Delta g^c = \Delta g - \left(\varepsilon_h + \varepsilon_\gamma + \varepsilon_P\right)_Q . \tag{6.7}$$

$(\varepsilon_h + \varepsilon_\gamma + \varepsilon_P)_Q$ are ellipsoidal corrections (Pavlis 1988). These, along with atmospheric and other corrections are applied to the observed gravity anomalies beforehand. Equation 6.6 refers to point values. Gravitational model estimation currently employs area-mean values over equi-angular cells, although (as Jekeli 1996 has pointed out), the use of area-mean values defined over spherical caps may be a preferable approach. In addition, the gravity anomaly may be analytically continued from the telluroid to the reference ellipsoid. Analytical continuation may be done using a Taylor expansion approach, whereby the anomaly on the ellipsoid Δg^e is related to Δg^c by:

$$\Delta g^e = \Delta g^c - \sum_{k=1}^{\infty} \frac{1}{k!} \frac{\partial^k \Delta g^e}{\partial h^k} \cdot h^k. \tag{6.8}$$

If the Taylor series in (6.8) is truncated to the linear term only, one obtains the (linear) gradient solution to the downward continuation problem:

$$\Delta g^e \approx \Delta g^c - (\partial \Delta g^e / \partial h) \cdot H^*. \tag{6.9}$$

The normal height, H^*, is seldom available in practice. It is usually approximated by the orthometric height H. This approximation however, introduces non-negligible systematic errors in the analysis of surface gravimetric data (Pavlis 1988). The free-air gravity anomaly gradient $(\partial \Delta g^e / \partial h)$ may be evaluated using detailed elevation information (assuming linear correlation between the free-air anomaly and the elevation), or from a preliminary high-degree model (from which one may also compute higher-order terms in the Taylor series of (6.8), in an iterative fashion). Downward continuation can also be performed by the iterative numerical solution of Poisson's integral (see Wang 1987, 1988 for more details). These reductions, and the discretization of the area-mean values over equi-angular cells, produce a grid of values on the ellipsoid, which may be modeled using solid spherical harmonics as:

$$\overline{\Delta g}^e_{ij} = \frac{1}{\Delta \sigma_i} \frac{GM}{(r_i^e)^2} \sum_{n=2}^{M} (n-1) \left(\frac{a}{r_i^e} \right)^n \sum_{m=-n}^{n} C_{nm} \cdot IY_{nm}^{ij}, \tag{6.10}$$

where the subscripts (i, j) identify the location of the cell in a two-dimensional array, defined by parallels and meridians, covering the ellipsoid. r_i^e is the geocentric distance to the center of the (i, j)th cell, and:

$$\Delta \sigma_i = \Delta \lambda \int_{\theta_i}^{\theta_{i+1}} \sin \theta d\theta = \Delta \lambda \cdot (\cos \theta_i - \cos \theta_{i+1}), \tag{6.11}$$

$$IY_{nm}^{ij} = \int_{\theta_i}^{\theta_{i+1}} \overline{P}_{n|m|}(\cos \theta) \sin \theta d\theta \cdot \int_{\lambda_j}^{\lambda_{j+1}} \left\{ \begin{array}{l} \cos m\lambda \\ \sin |m| \lambda \end{array} \right\}^{d\lambda} \begin{array}{l} \text{if}\, m \geq 0 \\ \text{if}\, m < 0. \end{array} \tag{6.12}$$

Equation 6.10 represents a mean value whose frequency content is restricted to maximum degree and order M. However, real data (e.g., the 1° × 1° mean values that were used in EGM96) are not band limited. To reduce aliasing effects, one may remove from the equi-angular mean values, the contribution beyond the solved-for degree M, using a preliminary high-degree model complete to some degree $N_{\max}$ (obviously one needs $N_{\max} \gg M$). Schematically:

$$\overline{\Delta g}^e (n = 2 \to M) \approx \overline{\Delta g}^e (n = 2 \to \infty) - \overline{\Delta g}^e (n = M + 1 \to N_{\max}). \tag{6.13}$$

Equation 6.10 is the mathematical model that underlies the observation (and subsequently the normal) equations formed from terrestrial (1° × 1°) gravity data. Details on the normal equation formation can be found in Pavlis (1988). The weight assigned to each individual gravity anomaly should be considered carefully, so that the terrestrial-only solution does not over-fit the areas covered with the most accurate and dense gravity data. Note also that (6.10) assumes that the entire signal present in terrestrial gravity data is of gravitational origin, i.e., (6.10) does not account for any systematic errors that may be present in (near) global anomaly databases.

The "Direct" Altimetry Normal Equations

The altimeter range measurement, ρ_{alt}, can be modeled as Marsh et al. (1990):

$$\rho_{alt} = h_{sat} - (h + \Delta h + \varepsilon) + b, \tag{6.14}$$

where ρ_{alt} is the observed range (corrected for instrument offsets) from the instantaneous sea surface to the satellite's center of mass, and:

h_{sat} is the distance from the surface of the reference ellipsoid to the satellite's center of mass,

h is the instantaneous sea surface height above (or below) the reference ellipsoid,

Δh is the sum of various instrumental, environmental and geophysical corrections,

ε is the instrument noise, and,

b is a bias term arising from the constant and time-varying instrument bias.

The instantaneous sea surface height can be modeled as:

$$h = N(n = 2 \rightarrow M) + N(n = M + 1 \rightarrow \infty) + \zeta_t + t_b + t_o + h_a, \tag{6.15}$$

where:

$N(n = 2 \rightarrow M)$ is the geoid undulation contribution up to degree and order M,

$N(n = M + 1 \rightarrow \infty)$ is the geoid undulation contribution beyond degree and order M,

ζ_t is the instantaneous Dynamic Ocean Topography,

t_b, t_o are the solid-Earth and ocean tides, respectively, and,

h_a is the ocean's response to the atmospheric loading.

Notice that the absence of zero degree from the geoid undulation implies that the bias term b also contains the difference between the semi-major axis of the adopted reference ellipsoid, and that of an *ideal* mean-Earth ellipsoid, with respect to which the geoid undulation averages globally to zero. Combination of (6.14) and (6.15), yields an observation equation which relates ρ_{alt} to the quantities of interest. h_{sat} depends on the potential coefficients, as well as on the initial state vector, and possibly other parameters that influence the orbit dynamics. Therefore, starting from some approximate values, altimeter data can be used to differentially improve the satellite's orbit (primarily in the radial direction), the low degree part of the potential coefficients (the $N(n = 2 \rightarrow M)$ part) and to estimate the DOT (ζ_t). The contribution $N(n = M + 1 \rightarrow \infty)$ can be filtered out (approximately) from the altimeter data using again a preliminary high-degree model, in a fashion similar to what was described above for the gravity anomalies. Δh is usually provided along with the data (Geophysical Data Records) or is computed from suitable models. h_a may be approximated as an "inverted barometer" response, while b, t_b and t_o may be included in the differential correction (estimation) process. The differential improvement of orbital parameters can be performed using numerical integration, in the exact same manner as is done for other tracking data types (Marsh et al. 1990; Nerem et al. 1994). Alternatively, linear perturbation theory may be used, to improve the radial orbit accuracy (Denker and Rapp 1990; Rapp et al. 1991).

The DOT, ζ_t, is composed of time-invariant and time-dependent parts, i.e.:

$$\zeta_t = \bar{\zeta} + \zeta(t), \tag{6.16}$$

where $\bar{\zeta}$ represents the mean value over some time interval and $\zeta(t)$ may contain annual, semi-annual and seasonal periodic constituents, as well as quasi-periodic effects (e.g., El Niño or La Niña effects). When data from non-contemporaneous missions are analyzed, more than one set of $\bar{\zeta}$-related parameters may be necessary. $\bar{\zeta}$ and $\zeta(t)$ may be represented in terms of surface spherical harmonics. In this case, the data gap generated by the presence of land areas (where $\bar{\zeta}$ and $\zeta(t)$ are not defined) requires some special consideration, in order to prevent large oscillations from occurring over land areas in the recovered $\bar{\zeta}$ and $\zeta(t)$ fields. Such oscillations may be avoided using, e.g., some a priori constraint, or some appropriately selected fictitious values over land, or by employing some alternative representation for these fields. Alternative representations for these ocean-specific signals have been proposed and studied by Hwang (1991) and Sanchez et al. (1997).

The normal equations from the satellite tracking data, the surface gravity data and the "direct" altimetry, can now be combined to estimate the low-degree comprehensive combination solution and its associated error covariance matrix. In the case of EGM96 this solution extended to degree and order 70. The maximum degree of the $\bar{\zeta}$ and $\zeta(t)$ representations which can be resolved, depends primarily on the accuracy of the satellite-only model and of the available marine gravity data. $\bar{\zeta}$ models to degree 20 and $\zeta(t)$ representations to degree 10 for annual and semi-annual constituents were estimated in EGM96. The relative weighting of the three sources of gravitational information (satellite, surface gravity, and altimetry) are again critical, especially when one employs long time spans of altimeter data acquired over repeat ground tracks. Examples of comprehensive low-degree models include EGM96 (to degree 70), TEG-3 (Tapley et al. 1997) and GRIM4-C4 (Schwintzer et al. 1997). The last model however, did not incorporate altimetry as "direct" tracking; it combined the GRIM4-S4 normal equations with normal equations obtained from the analysis of a global 1° × 1° grid of mixed Δg (mostly over land) and N values obtained from altimetry, where the Levitus (1982) model was used to define a priori the DOT.

6.6.1.3 The High-Degree Combination Solution

As discussed in Sect. 6.6, a combination solution to high-degree (e.g., 360 or higher) may be performed by combining the satellite-only normal equations, with gravitational information obtained from the analysis of *complete* global grids of functionals of the field (e.g., N and/or Δg observations). These approaches rely on the exploitation of symmetry properties of the data grids, and take advantage of the applicability of FFT algorithms for the efficient formation of normal equations from the gridded data. Two of the techniques originally studied by Colombo (1981a), the (simple) Numerical Quadrature and the Block-Diagonal least-squares adjustment, were applied during the development of EGM96. These techniques combined the satellite-only information, with a global $30' \times 30'$ "merged" file of mean Δg (appropriately corrected and reduced to the ellipsoid). Other groups (e.g., GFZ) have variously incorporated global sets of *both* N and Δg data simultaneously into the high-degree combination solution. A disadvantage of their approach is that extensive areas have to be "filled-in" with synthetic pseudo-observations particularly of N, but also of Δg, to achieve global complete coverage in the respective files (Gruber et al. 1996). We review next the techniques implemented during the EGM96 model development.

The Numerical Quadrature (NQ) Technique

The orthogonality relations of the surface spherical harmonics (Heiskanen and Moritz 1967, Sect. 1–13) constitute the underlying principle of the NQ approach. In theory, one has:

$$C_{nm} = \frac{1}{4\pi\gamma(n-1)} \iint\limits_{\sigma} \Delta g(\theta,\lambda) Y_{nm}(\theta,\lambda) d\sigma. \tag{6.17}$$

Equation 6.17 requires the existence of gravity anomalies continuously covering the sphere. In practice, one has discrete area-mean values of gravity anomalies, on an equi-angular grid on the ellipsoid ($\overline{\Delta g}^{e}_{ij}$). The discretization of the surface integral in (6.17), and the consideration of the ellipticity of the surface on which the $\overline{\Delta g}^{e}_{ij}$ values reside (see Jekeli 1988), lead to:

$$\begin{aligned} C^{T}_{nm} = \frac{1}{4\pi a\gamma(n-1)} \sum_{i=0}^{N-1} r^{e}_{i} \sum_{k=0}^{s'} \frac{L_{nmk}}{\bar{S}_{n-2k,|m|}(b/E)} \frac{\overline{IP}^{i}_{n-2k,|m|}}{q^{i}_{n-2k}} \\ \cdot \sum_{j=0}^{2N-1} \overline{\Delta g}^{e}_{ij} \begin{Bmatrix} IC \\ IS \end{Bmatrix}^{j}_{m} \quad \begin{matrix} m \geq 0 \\ m < 0 \end{matrix} \quad . \end{aligned} \tag{6.18}$$

The complete derivation of (6.18) can be found in Rapp and Pavlis (1990). The estimation of the complete high-degree set of geopotential coefficients is performed here as a two-step procedure. First, the global set of $\overline{\Delta g}_{ij}^{e}$ provides, through (6.18), a "terrestrial" estimate, C_{nm}^{T}, of those harmonic coefficients present in the satellite-only model. In addition, (6.18) is used to propagate the error variances of $\overline{\Delta g}_{ij}^{e}$ to C_{nm}^{T}, and thus form their complete error covariance matrix, $\mathrm{Cov}(C_{nm}^{T})$. Based on the harmonic coefficients of the satellite-only model, C_{nm}^{S}, and their associated error covariance matrix, $\mathrm{Cov}(C_{nm}^{S})$, and their "terrestrial" counterparts, a least-squares adjustment is performed to estimate a unique set of coefficients (and its associated error covariance matrix), essentially as a weighted average of the two independent estimates. The formulation of this adjustment process is described in full detail in Rapp and Pavlis (1990, Sect. 2.3). This adjustment provides also a global set of adjusted gravity anomalies. In a second step, the adjusted gravity anomalies are input to (6.18) to yield the complete set of harmonic coefficients up to the high degree (360 or higher). The error variances of these higher-degree coefficients may be obtained by quadratic addition of the propagated anomaly error and an estimate of the sampling error (ibid., (50)–(53)). This general procedure was originally proposed by Kaula (1966) and has been used in several high-degree models developed at The Ohio State University (Rapp 1981; Rapp and Cruz 1986a; Rapp and Pavlis 1990).

Composite quadrature weights $1/q_n^i$ were introduced by Colombo (1981a, p. 76) as an efficient way of approximating the harmonic analysis results obtainable using Optimal Estimation (their latitude dependence was suggested by Katsambalos (1979). Pavlis (1996) introduced the following set of composite de-smoothing factors q_n^i, which avoid the discontinuities of Colombo's (ibid.) original set:

$$q_n^i = \begin{cases} (\beta_n^i)^2 & 0 \le n \le L/2 \\ (\beta_n^i)^{L/n} & L/2 < n \le L \\ \beta_L^i & L < n \end{cases} . \tag{6.19}$$

$L(= \pi/\Delta\lambda)$ is the Nyquist degree implied by the sampling interval $\Delta\lambda$, and β_n^i is the Pellinen operator computed by:

$$\beta_n^i = \frac{1}{(1-\cos\psi_0^i)}\frac{1}{(2n+1)}\left[P_{n-1}(\cos\psi_0^i) - P_{n+1}(\cos\psi_0^i)\right], \tag{6.20}$$

where ψ_0^i is the semi-aperture of a spherical cap having the same area as the equi-angular block on the ith latitude band. It is computed by Colombo (1981a, p. 85):

$$\psi_0^i = arc\cos\left[\frac{\Delta\lambda}{2\pi}(\cos\Delta_{i+1} - \cos\Delta_i) + 1\right], \tag{6.21}$$

where δ is the reduced co-latitude (Heiskanen and Moritz 1967, Sect. 1-19). Introduction of the de-smoothing factors of (6.19) enabled Pavlis (1996) to extend

NQ models (developed using 30′ × 30′ $\overline{\Delta g}^{e}_{ij}$) to degrees higher than 360, without experiencing large jump discontinuities at the Nyquist degree (360) implied by the 30′ × 30′ data sampling.

The Block-Diagonal (BD) Least-Squares Adjustment Technique

Equation 6.10 could be used to form normal equations from a global set of 30′ × 30′ $\overline{\Delta g}^{e}_{ij}$ data. These normal equations could then be combined with the satellite-only normal equations, to yield the combination solution. Rapp (1967) proposed originally this approach. However, to implement this technique to a high degree, one has to deal with the extremely high computational demands of such a task. This may be accomplished by forming, instead of the full (symmetric) "terrestrial" normal matrix, a suitable approximation of it. This approximation should be simple enough, to allow numerical implementation, and, on the same time, rigorous enough to maintain the most important characteristics of the full matrix. Colombo (1981a) has shown that if:

(a) The data reside on a surface of revolution (e.g., a rotational ellipsoid),
(b) The grid is complete and the longitude increment constant,
(c) The data weights are longitude-independent,
(d) The data weights are symmetric with respect to the equator, then, zero elements in the normal equations formed in the least-squares estimation will occur as prescribed by (see also (Pavlis 1988) for details):

$$[\mathbf{N}]_{C_{nm}C_{rs}} = 0 \quad if \quad \{m \neq s\} \ or \ \{m = s \ and \ n - r = 2k + 1\}. \tag{6.22}$$

Note that in this notation the order subscript is a signed integer, whose sign identifies the type of coefficient (positive: cosine, negative: sine). If condition (d) does not hold true, then:

$$[\mathbf{N}]_{C_{nm}C_{rs}} = 0 \quad if \quad \{m \neq s\}. \tag{6.23}$$

The sparsity patterns implied by (6.22) and (6.23) will be referred to as BD1 and BD2 respectively. In addition, a BD3 pattern will be considered defined by:

$$[\mathbf{N}]_{C_{nm}C_{rs}} = 0 \quad if \quad \{|m| \neq |s|\}, \tag{6.24}$$

which admits non-zero off-diagonal elements across coefficients of different type within a given order. It is instructive to consider the computational efficiency implied by these patterns. Table 6.2 provides relevant statistics for a solution complete from degree and order 0 to degree and order 360, excluding degree $n = 1$ terms. Such a solution involves 130,318 unknown coefficients, and the upper (or lower) triangular part of the (symmetric) full "terrestrial" normal matrix has 8,491,455,721 elements.

Table 6.2 Statistics of normal matrices related to an expansion complete to $N_{max} = 360$ (excluding degree $n = 1$ coefficients) using different sparsity patterns

Statistic	Sparsity pattern		
	BD1	BD2	BD3
Total number of non-zero elements	7,905,721	15,746,100	31,362,241
Percentage of full matrix elements	0.09	0.19	0.37
Number of blocks	1,440	721	361
Num. of unknowns in largest block	181	360	718
Num. of elements in largest block	16,471	64,980	258,121

The enormous computational savings that can be inferred from Table 6.2 make the BD approximations very attractive estimation strategies. These savings however come at the expense of the rigor exercised in the development of the model. The real-world anomaly data to be analyzed comply *only* with the conditions (a) and (b) above (in fact, to comply even with the (a) and (b) conditions, "filling-in" techniques and analytical continuation have to be employed, since the original $30' \times 30'$ data file is neither complete, nor residing on any surface of revolution). Furthermore, the normal equations of the EGM96S satellite-only model do not conform to any such sparsity pattern. BD3 is the most rigorous of the three patterns, while being well within our present computational capabilities. In EGM96 we therefore chose to form the "terrestrial" normal equations according to BD3. We did not however alter the data weights to enforce compliance with (c) or (d). Furthermore, the satellite-only normal equations were *not* truncated, since this was found to degrade the combination solution unacceptably, at the lower degrees (Lerch et al. 1993).

The BD technique may be viewed as an intermediate type of approach between the rigorous *comprehensive* least-squares adjustment procedure and the NQ procedure. The BD approach combines some of the advantages of the other two approaches, while avoiding their critical shortcomings. Pavlis et al. (1996a) discuss some of the analytical differences between the NQ and the BD techniques, both from the harmonic analysis and from the combination solution points of view. Three important aspects of the BD approach require some discussion here.

1. ***Ordering and Grouping of the Unknowns.*** The particular ordering of the unknown potential coefficients within the vector of parameters has a tremendous impact on the efficiency with which the combined (satellite-only plus terrestrial) normal equations can be inverted. To illustrate this, let us consider for a moment that a "high-degree" solution complete to $N_{\max} = 6$ is to be developed, in a least-squares combination with a satellite-only model complete to degree $N_{sat} = 4$. In this case, the terrestrial normal equations involve 46 unknowns (complete set to $N_{\max} = 6$, excluding $n = 1$ terms), while the satellite-only normal equations 22 unknowns. The unknown coefficients are ordered first by increasing order (m), then by type (C then S), and lastly by increasing degree (n). This is denoted ordering pattern "V" in Pavlis (1988, Table 6.3). Adhering to the sparsity pattern BD3, the *terrestrial* normal equations take the form shown in Fig. 6.3a, where

gray areas indicate non-zero elements. This type of normal matrix can be set up and inverted very efficiently, thus providing the terrestrial-only estimates of the coefficients and their associated BD error covariance matrix. For an analysis to $N_{\max} = 359$, (which is the maximum degree resolvable from $30' \times 30'$ mean values using least-squares) this matrix contains 360 diagonal blocks, the largest one having dimension 716×716 (corresponding to order $m = 1$), while the smallest one having dimension 2×2 (corresponding to $m = 359$).

However, if one conforms to this ordering of unknowns, the *combined* (terrestrial plus satellite-only) normal equations take the form shown in Fig. 6.3b. In this figure, black areas indicate the non-zero elements in the combined normal equations, which arise from the satellite-only normal equation contribution (overlaid on the structure of Fig. 6.3a). It is obvious that the "V" type of ordering of the unknowns creates a large (although sparse) block in the combined normal equations, which would have to be treated as a full matrix. In the real-world (EGM96) case, where $N_{sat} = 70$, this block would have dimension $45,787 \times 45,787$. Clearly, a different ordering of the unknowns is required, whereby the coefficients present in the satellite-only model would be grouped together. Two ways to accomplish this are:

Forward grouping	**Reverse grouping**
Group 1: $n \leq N_{sat},\ m \leq N_{sat}$	Group 1: $n > N_{sat},\ m > N_{sat}$
Group 2: $n > N_{sat},\ m \leq N_{sat}$	Group 2: $n > N_{sat},\ m \leq N_{sat}$
Group 3: $n > N_{sat},\ m > N_{sat}$	Group 3: $n \leq N_{sat},\ m \leq N_{sat}$

Inside each group, the coefficients are ordered following the same pattern "V" as before. Figure 6.3c, d show the structure of terrestrial and combined normal equations respectively, for the forward grouping, while Fig. 6.3e, f for the reverse grouping. Bosch (1993) studied the structure of the combined normal equations resulting from the forward grouping and proposed an algorithm for the solution of such a system. However, the reverse grouping possesses a very significant advantage over the forward one. Namely, the Cholesky factor of the matrix in Fig. 6.3f preserves the structure of the upper (or lower) part of the original matrix. This enables a very efficient solution of the combined system, and provides the possibility to compute selected elements of the inverse of the combined normal equations (error covariance matrix). This was recognized by Chan and Pavlis (1995), and independently by Schuh (1996).

2. ***Reference Values and Aliasing Effects.*** Consider the rigorous (complete) set of normal equations obtained from the merged $30' \times 30'$ $\overline{\Delta g}^{e}_{ij}$ data, denoted by:

$$\mathbf{N} \cdot \hat{\mathbf{X}} = \mathbf{U}, \tag{6.25}$$

where $\hat{\mathbf{X}}$ represents the adjusted coefficients of the disturbing potential, i.e., remainders after subtraction of the even zonal harmonics of the normal ellipsoidal field. The BD3 truncated version of the normal equation system may be written as:

$$\tilde{\mathbf{N}} \cdot \hat{\hat{\mathbf{X}}} = \mathbf{U}. \tag{6.26}$$

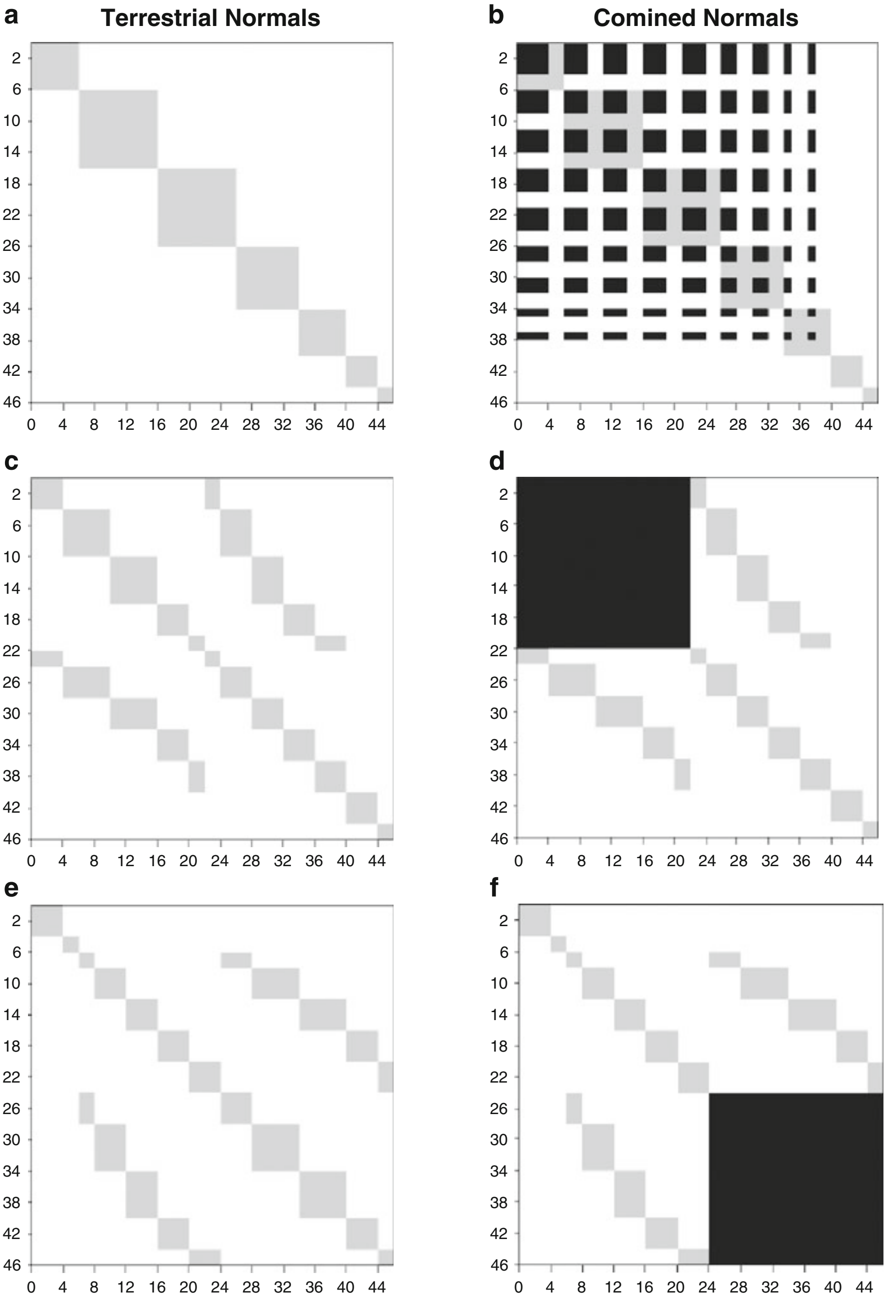

Fig. 6.3 Structure of terrestrial and combined normal equations for different orderings and groupings of the unknown coefficients

Notice that in both cases the right hand side vector **U** is the same (and is computed rigorously). The difference in the estimates of the unknowns (rigorous minus BD3 approximation) is therefore:

$$\mathbf{dX} = \hat{\mathbf{X}} - \hat{\hat{\mathbf{X}}} = (\mathbf{N}^{-1} - \tilde{\mathbf{N}}^{-1}) \cdot \mathbf{U}. \tag{6.27}$$

Equation 6.27 indicates that the magnitude of **dX** may be reduced either by reducing the magnitude of the term $(\mathbf{N}^{-1} - \tilde{\mathbf{N}}^{-1})$ (i.e., by providing a better approximation to the normal matrix), or, for a given approximation of the normal matrix, by reducing the magnitude of **U**. One may reduce the magnitude of **U** by modeling residual anomalies, after subtraction of a high-degree reference model, instead of the original $30' \times 30'$ $\overline{\Delta g}_{ij}^{e}$ values, which refer to the normal (ellipsoidal) field. Moreover, as it is discussed by Pavlis et al. (1996a), the BD solution (unlike the NQ one) may be affected significantly at the high degrees by aliasing. This arises from signal present in the $30' \times 30'$ data, which corresponds to harmonics beyond those solved-for. In order to reduce the magnitude of **dX** in (6.27), and also reduce aliasing effects, Pavlis et al. (1996b) suggested removing a "reference" $\overline{\Delta g}_{ij}^{e}$ value from the original $30' \times 30'$ anomalies. This reference value can be computed from a preliminary NQ model to $N_{\max} = 460$. The BD combination solution is then performed using residual anomalies, and yields a set of coefficient corrections to $N_{\max} = 359$. Addition of the reference model coefficients (to $N_{\max} = 359$) yields the final BD high-degree expansion. This is the procedure that was followed in the development of EGM96, for the portion of that model from degree $n = 71$ to $n = 359$.

3. ***Use of an A Priori Constraint.*** Un-modeled long wavelength systematic errors that are (still) present in global gravity databases, coupled with our inability to numerically account for error correlations between the $30' \times 30' \overline{\Delta g}_{ij}^{e}$ data, necessitate some down-weighting of these data in present combination solutions. This aims to preserve the highly accurate long wavelength information contributed by the satellite-only normal equations. This down-weighting, in conjunction with the use of diagonal anomaly weight matrices, has the undesirable side effect that the propagated errors of the combination solution at the high degrees ($n > 250$) are too pessimistic. This was the case for both the OSU89A/B models (Rapp and Pavlis 1990), and for the OSU91A model (Rapp et al. 1991). The BD approach offers a possible (at least partial) remedy to this problem. This may be accomplished by introducing some a priori information for the coefficient *corrections* (relative to the NQ reference model), for $n > 70$. This a priori information may be provided in the form of some anomaly degree variance model. Although this approach helps produce a more realistic error spectrum at the higher degrees ($n > 250$), it is a rather simplistic and empirical way of attacking the underlying problem. Furthermore, it has the undesirable side effect of "dampening" also the power of the signal itself, a characteristic of the EGM96 model that attracted some criticism (see also Jekeli 1999b). More study is needed to develop better ways of treating systematic errors in global gravity anomaly data sets either by direct estimation and/or by using more sophisticated error modeling.

6.6.2 EGM2008

The main reason for the choice of the solution strategy implemented in the development of EGM96 was the fact that its satellite-only component (EGM96S) was accompanied by a variance-covariance matrix that was fully-occupied. This was due to the fact that the heterogeneous tracking data from the (approximately) 40 satellites that were used to derive EGM96S to degree and order 70, were incapable of de-correlating adequately the spherical harmonic coefficients up to this degree and order. Therefore, in order to preserve the integrity of the least-squares adjustment used to combine EGM96S with the surface gravity and altimetry data, one had to consider the satellite-only normal equations in their complete (fully-occupied) form. Any block-diagonal approximation of these normal equations would result in estimation errors that could not be tolerated. This situation changed dramatically with GRACE. Due to the global coverage and uniform accuracy of the GRACE data, the corresponding normal equation matrix could be safely approximated by a block-diagonal matrix, without significant loss of accuracy. In addition, after the availability of satellite-only models from GRACE, there is really no need to incorporate altimeter data into the combination solution in the form of “direct” tracking. Instead, a preliminary model based on GRACE data and a MSS, can be used to estimate a preliminary model of the DOT. This DOT model could then be used to correct the altimeter data, and estimate from them an ocean-wide set of altimetry-derived gravity anomalies. These anomalies can be “merged” with terrestrial and airborne data to form a complete global gravity anomaly grid. The gravitational information implied by these gridded data could then be combined (in a least-squares adjustment) with the satellite-only model from GRACE, to derive the combination solution, up to the high degree (2,159), in a single step. The entire process may be iterated, using the high-degree combination solution to derive the next estimate of the DOT, and so on. This is essentially the approach that was used to develop EGM2008 (Pavlis et al. 2008). Two iterations of the estimation of the altimetry-derived gravity anomalies were performed, which was sufficient for the process to converge. Despite its iterative nature, this approach permits the development of very high-degree combination solutions in an efficient manner, and moreover in a single adjustment step, thereby avoiding the “piece-wise” nature of models like EGM96.

In terms of data complement and solved-for parameters, EGM2008 resembles the OSU89A/B solutions (Rapp and Pavlis 1990). The gravitational information of a satellite-only model (accompanied by its complete error variance-covariance matrix) is combined with the corresponding information from the analysis of a complete set of area-mean gravity anomalies given over a global equi-angular grid, to estimate a set of potential coefficients complete to a harmonic degree commensurate with the resolution of the gravity anomaly data. In EGM2008 (as in OSU89A/B), a model of the DOT was not estimated simultaneously with the potential coefficients, in contrast to the strategy used in the development of EGM96. In the following sections we describe in some detail the “building blocks” that were used to estimate the EGM2008 very high-degree model.

6.6.2.1 The ITG-GRACE03S Satellite-Only Model

The satellite-only model that was used in the development of EGM2008 is designated ITG-GRACE03S (Mayer-Gürr 2007). This model was developed at the Institute of Theoretical Geodesy of the University of Bonn, in Germany. The ITG-GRACE03S model was based on 57 months of GRACE Satellite-to-Satellite Tracking (SST) data. No other data were used in the development of ITG-GRACE03S. A short-arc analysis approach was used for the development of ITG-GRACE03S, as described by Mayer-Gürr et al. (2007).

ITG-GRACE03S is complete to spherical harmonic degree and order 180, and was accompanied by its full error variance-covariance matrix. Due to the global coverage resulting from the near-polar orbits of the two GRACE spacecraft, and due to the uniform accuracy of the GRACE data, this error variance-covariance matrix is diagonally-dominant. Numerical experiments indicated that a block-diagonal approximation of this matrix according to the BD1 scheme (see (6.22) would be sufficient to maintain the essential characteristics of the errors associated with the ITG-GRACE03S model, without any appreciable loss of accuracy in the development of the combination gravitational solution. This simplified tremendously the numerical implementation of the combination solution, as we discuss in Sect. 6.6.2.3.

The ITG-GRACE03S model was developed and was made available in terms of spherical harmonic coefficients. However, as we discuss next, the analysis of the terrestrial data, and the combination solution that led to the development of EGM2008, were implemented in terms of ellipsoidal harmonics (see also Jekeli 1988 for details). Therefore, in a first step, both the ITG-GRACE03S coefficient model and its associated error variance-covariance matrix were transformed from spherical to ellipsoidal harmonics, using the transformation formulas developed by Jekeli (1988) and implemented by Gleason (1988). The transformation from spherical to ellipsoidal harmonic coefficients is given in Gleason (1988, Eq. 2.8). The reverse (ellipsoidal to spherical) transformation is given in Gleason (1988, Eq. 2.10). Both are linear transformations that preserve the harmonic order but not the harmonic degree. It is very important to recognize that both transformations preserve the structure of the BD1 block-diagonal pattern of the error variance-covariance matrix. This aspect of the transformations is of critical importance to the computational efficiency of both the least-squares adjustment necessary to derive the combination solution, and of the subsequent error propagation associated with the final combined solution. In the development of EGM2008, we first transformed the full error variance-covariance matrix of ITG-GRACE03S from the spherical to the ellipsoidal harmonic representation, and then approximated the resulting matrix to one conforming to the BD1 block-diagonal pattern.

The outcome from this "pre-processing" of the ITG-GRACE03S information is a set of ellipsoidal harmonic coefficients of this satellite-only model, $C_{nm}^{S,e}$, complete to degree and order 180, accompanied by the BD1 approximation of its error variance-covariance matrix, $\mathrm{Cov}(C_{nm}^{S,e})$.

6.6.2.2 The Block-Diagonal (BD) Least-Squares "Terrestrial" Coefficient Estimates

The second "building block" of EGM2008 is the gravitational information obtained from the analysis of a complete global equi-angular $5' \times 5'$ grid of area-mean gravity anomalies. These anomalies have been corrected for ellipsoidal corrections, and have been analytically continued to the surface of the reference ellipsoid. They are denoted by $\overline{\Delta g}_{ij}^{e}$ and represent exactly the same type of observations (albeit at a different grid size) as the $\overline{\Delta g}_{ij}^{e}$ of section "The Numerical Quadrature (NQ) Technique", where we discussed the development of EGM96. In the notation of (6.10), considering the small latitudinal extent of the $5' \times 5'$ cell, the small and regular latitudinal variation of r^e within the cell can be safely ignored (see also Rapp and Pavlis 1990, p. 21,887), so that we may approximate:

$$\overline{(r\Delta g)}_{ij}^{e} \approx r_i^e \cdot \overline{\Delta g}_{ij}^{e}, \tag{6.28}$$

where r_i^e is the geocentric distance to the center of the (i, j)th cell. The product $r_i^e \cdot \overline{\Delta g}_{ij}^{e}$, defined over the surface of the reference ellipsoid, can be expanded into surface ellipsoidal harmonic functions (Heiskanen and Moritz 1967, Sect. 1-20), as:

$$r_i^e \cdot \overline{\Delta g}_{ij}^{e} = \frac{1}{\Delta\sigma_i} \frac{GM}{a} \sum_{n=2}^{M} (n-1) \sum_{m=-n}^{n} C_{nm}^{T,e} \cdot IY_{nm}^{ij}, \tag{6.29}$$

where δ is the reduced co-latitude (Heiskanen and Moritz 1967, Sect. 1-19) and:

$$\Delta\sigma_i = \Delta\lambda \int_{\delta_i}^{\delta_{i+1}} \sin\delta d\delta = \Delta\lambda \cdot (\cos\delta_i - \cos\delta_{i+1}), \tag{6.30}$$

$$IY_{nm}^{ij} = \int_{\delta_i}^{\delta_{i+1}} \overline{P}_{n|m|}(\cos\Delta) \sin\delta d\delta \cdot \int_{\lambda_j}^{\lambda_{j+1}} \left\{ \begin{array}{l} \cos m\lambda \\ \sin|m|\lambda \end{array} \right\} d\lambda \begin{array}{l} \text{if } m \geq 0 \\ \text{if } m < 0 \end{array}. \tag{6.31}$$

The quantity $r^e \Delta g^e$ represents a harmonic function, and, under the approximation of (6.28), so does the quantity $r_i^e \cdot \overline{\Delta g}_{ij}^{e}$. This allows one to relate the *ellipsoidal* harmonic coefficients $C_{nm}^{T,e}$ of (6.29), to the corresponding *spherical* harmonic coefficients $C_{nm}^{T,s}$, using the exact transformations derived by Jekeli (1988) and implemented and verified by Gleason (1988). Note that our $C_{nm}^{T,s}$ and $C_{nm}^{T,e}$ coefficients are related to the corresponding $\overline{g}_{n,m}^{s}$ and $\overline{\overline{g}}_{n,m}^{e}$ coefficients of Gleason (ibid.) by:

$$\left\{ \begin{array}{c} \overline{g}_{n,m}^{s} \\ \overline{\overline{g}}_{n,m}^{e} \end{array} \right\} = \frac{GM}{a^2}(n-1) \cdot \left\{ \begin{array}{c} C_{nm}^{T,s} \\ C_{nm}^{T,e} \end{array} \right\}. \tag{6.32}$$

Based on (6.29), one forms a system of observation equations that can be written in vector format as:

$$\mathbf{v} = \mathbf{A} \cdot \hat{\mathbf{x}} - \mathbf{L_b}, \tag{6.33}$$

where $\mathbf{L_b}$ is the vector of observations $\overline{\Delta g}_{ij}^{e}$, $\mathbf{v}$ is the vector of corresponding residuals, $\mathbf{A}$ is the design matrix whose elements are formed based on (6.29), and $\hat{\mathbf{x}}$ represents the vector of estimated coefficients $C_{nm}^{T,e}$. The least-squares solution, $\hat{\mathbf{x}}$, which minimizes the quadratic form $\mathbf{v^T P\, v}$, is given by Uotila (1986):

$$\left.\begin{array}{ll} \hat{\mathbf{x}} = \mathbf{N^{-1} U} & \text{(a)} \\ \mathbf{N} = \mathbf{A^T P\, A} & \text{(b)} \\ \mathbf{U} = \mathbf{A^T P\, L_b} & \text{(c)} \end{array}\right\}, \tag{6.34}$$

where $\mathbf{P}$ is the weight matrix associated with the observations $\overline{\Delta g}_{ij}^{e}$. $\mathbf{P}$ was assumed diagonal, with elements equal to the reciprocal of the error variance associated with each individual gravity anomaly observation, i.e.:

$$\mathbf{P} = \sigma_0^2 \cdot \begin{pmatrix} \frac{1}{\sigma_1^2} & & 0 \\ & \ddots & \\ 0 & & \frac{1}{\sigma_K^2} \end{pmatrix}, \tag{6.35}$$

where K is the total number of observations, and σ_0^2 is the a priori variance of unit weight. For the complete, global equi-angular $5' \times 5'$ grid of area-mean gravity anomalies used here, $K = 2160 \times 4320 = 9331200$. The assumption that the gravity anomaly errors are uncorrelated is made out of necessity, rather than desire. It is extremely difficult to estimate the error correlations between the gravity anomalies with any degree of accuracy. It is also practically impossible to handle numerically a full (symmetric) weight matrix of dimension 9,331,200. Even the estimation of realistic error variances for the gravity anomalies is itself a very challenging task. These error variances affect critically the error variance-covariance matrix of the "terrestrial" coefficients $C_{nm}^{T,e}$, which is given by:

$$\mathrm{Cov}(C_{nm}^{T,e}) = \sigma_0^2 \cdot \mathbf{N^{-1}}. \tag{6.36}$$

The gravity anomaly error variances should be such that they represent realistic estimates of the accuracy of the data, not their precision. They should produce an error variance-covariance matrix $\mathrm{Cov}(C_{nm}^{T,e})$ that would permit the combination of $C_{nm}^{T,e}$ with the satellite coefficients $C_{nm}^{S,e}$ using well "calibrated" *relative* weights. The gravity anomaly error estimates, as well as the error estimates associated with the satellite information, should also be realistic in an *absolute* sense. Otherwise, the a posteriori error estimates associated with the combination solution will not reflect adequately the real accuracy of the combined model.

With the complete, global equi-angular $5' \times 5'$ grid of area-mean gravity anomalies as input, the expansion of (6.29) was extended to maximum degree and order $M = 2159$, in ellipsoidal harmonics. This is the maximum degree for which the system of (6.29) still maintains full rank (Colombo 1981a). The "terrestrial" normal equations were approximated by their BD1 counterpart. Although the weights associated with the gravity anomalies do not strictly comply with the requirements for a BD1 sparsity pattern (see section "The Block-Diagonal (BD) Least-Squares Adjustment Technique"), the geographic variation of these weights do not produce significant departures from such a pattern. This is mainly due to the uniformity of the errors of altimetry-derived gravity anomalies, which cover approximately 70% of the Earth's total area.

It should also be emphasized here that the residuals appearing in (6.33) represent a measure of "goodness of fit" and are not necessarily representative of the errors of the gravity anomaly data (Pavlis 1988). In fact, if the gravity anomaly data were limited in spectral content and contained contributions *only* from (a subset of) the solved-for harmonics appearing in (6.29), these residuals would have been identically zero (to the level of the numerical noise).

6.6.2.3 The Least-Squares Combination Solution

The least-squares combination solution coefficient set, $C_{nm}^{C,e}$, is determined essentially as the weighted average of the satellite-only estimate, $C_{nm}^{S,e}$, and of the "terrestrial" estimate, $C_{nm}^{T,e}$, each of these two *independent* estimates being weighted by the inverse of its respective error variance-covariance matrix, according to:

$$\begin{aligned} C_{nm}^{C,e} = {} & \left\{ \left[\mathrm{Cov}(C_{nm}^{S,e})\right]^{-1} + \left[\mathrm{Cov}(C_{nm}^{T,e})\right]^{-1} \right\}^{-1} \cdot \\ & \left\{ \left[\mathrm{Cov}(C_{nm}^{S,e})\right]^{-1} \cdot C_{nm}^{S,e} + \left[\mathrm{Cov}(C_{nm}^{T,e})\right]^{-1} \cdot C_{nm}^{T,e} \right\} . \end{aligned} \tag{6.37}$$

The error variance-covariance matrix of $C_{nm}^{C,e}$, $\mathrm{Cov}(C_{nm}^{C,e})$, is given by:

$$\mathrm{Cov}(C_{nm}^{C,e}) = \left\{ \left[\mathrm{Cov}(C_{nm}^{S,e})\right]^{-1} + \left[\mathrm{Cov}(C_{nm}^{T,e})\right]^{-1} \right\}^{-1} . \tag{6.38}$$

It is important to recognize here that the BD1 approximation of both the satellite-only and the "terrestrial" error variance-covariance matrices permits the evaluation of the combination solution in a fashion that is extremely fast and numerically economic. This is because the linear system representing the entire combination solution is comprised of uncorrelated BD1-type blocks that can be inverted independently of each other. For the solution to degree and order 2,159, the largest (symmetric) matrix that needs to be stored and inverted is of the order of 1,080, a task that is trivial for the currently available computers.

Evaluation of (6.29), using the combined solution coefficients, $C_{nm}^{C,e}$, in the place of $C_{nm}^{T,e}$, yields the set of adjusted area-mean gravity anomalies $\widehat{\Delta g}_{ij}^{e}$, as:

$$r_i^e \cdot \widehat{\Delta g}_{ij}^{e} = \frac{1}{\Delta\sigma_i} \frac{GM}{a} \sum_{n=2}^{M} (n-1) \sum_{m=-n}^{n} C_{nm}^{C,e} \cdot IY_{nm}^{ij}. \tag{6.39}$$

The residual gravity anomalies v_{ij} resulting from the least-squares adjustment that produced the combination solution are then computed as the difference between these adjusted anomalies and the original (input) values:

$$v_{ij} = \widehat{\Delta g}_{ij}^{e} - \overline{\Delta g}_{ij}^{e}. \tag{6.40}$$

These residual anomalies are due to any existing differences between the satellite-only and the "terrestrial" estimates of the gravity anomalies. The values of these residual anomalies are affected directly by the weights used in the combination solution for the satellite-only estimate relative to its "terrestrial" counterpart.

A final step towards the estimation of the combination high-degree solution is the transformation of the *ellipsoidal* harmonic coefficients $C_{nm}^{C,e}$, and of their associated error variance-covariance matrix $\mathrm{Cov}(C_{nm}^{C,e})$, to their *spherical* counterparts, $C_{nm}^{C,s}$ and $\mathrm{Cov}(C_{nm}^{C,s})$. This is performed again using the ellipsoidal-to-spherical transformation formulas of Jekeli (1988) and Gleason (1988). Due to the fact that this transformation preserves the order but not the degree, an ellipsoidal harmonic expansion complete to degree and order 2,159, as in the case of EGM2008, produces a corresponding spherical harmonic coefficient set that extends up to degree 2,190. The "extra" coefficients are linear combinations of the lower degree coefficients (Jekeli 1988). Such "extra" coefficients are of negligible effect for expansions to degree 360 or so (e.g., EGM96), but cannot be omitted in expansion that extend to degree 2,159 (e.g., EGM2008). In such very high-degree expansions, omission of these "extra" coefficients will result in unacceptable modeling errors, especially over high latitude areas (see also Holmes and Pavlis 2007 for details).

6.6.2.4 Error Propagation

Users of a high-resolution global gravitational model require geographically specific estimates of the error associated with various gravitational functionals (e.g., gravity anomalies, height anomalies, deflections of the vertical) computed from the model. These estimates are composed of the commission and the omission error implied by a specific model. The commission (or propagated) error is due to the fact that a model that is based on actual observations can never be error-free since the data supporting its development can never be error-free. The omission (or truncation) error is due to the fact that a model can only have finite resolution; therefore it will always omit a portion of the Earth's true gravity field spectrum, which extends to infinity. Rigorous computation of the commission error implied by any model requires the complete error variance-covariance matrix of its defining parameters. In principle, given this matrix, one can compute the commission error of various model-derived functionals, using error propagation. The error variance-covariance matrix of a spherical harmonic model complete to degree and order 2,159 has dimension ~4.7 million. The computation of such a matrix is beyond the existing computing technology. Even for expansions to degree and order 360, like EGM96, which involve approximately 130,000 parameters, the formation of the normal equation matrix, its inversion, and the subsequent error propagation using the resulting error variance-covariance matrix is a formidable computational task. For EGM96 (Lemoine et al. 1998), such error propagation was only possible for the portion of the model extending to degree and order 70. For EGM2008, which extends to degree 2,159 in ellipsoidal harmonics, the alternative error propagation technique that was developed and implemented by Pavlis and Saleh (2005) was used. This technique is capable of producing geographically specific estimates of a model's commission error, *without* the need to form, invert, and propagate large matrices. Instead, this technique uses integral formulas with band-limited kernels and requires as input the error variances of the gravity anomaly data that are used in the development of the gravitational model.

The main idea behind the technique of Pavlis and Saleh (2005) is the realization that in combination solutions like EGM96 and EGM2008, the satellite-only information influences the combined model only up to a relatively low degree, which is the maximum degree of the satellite-only solution. Up to this maximum degree, the combined solution is the outcome of a least-squares adjustment. However, the higher degree and order portion of the combined gravitational model (beyond the range of influence of the satellite information), is determined *solely* from a complete, global grid of area-mean gravity anomaly data. Therefore, beyond the maximum degree and order of the available satellite-only solution, there is little need to form complete normal matrices, since no "adjustment" takes place within this degree range. The merged (terrestrial plus altimetry-derived) area-mean gravity anomalies are the only data whose signal and error content determine the model's signal and error properties over this degree range. This fact enables high-degree error propagation, *with* geographic specificity, through the use of integral formulas with band-limited kernels, *without* the need to form, invert, and propagate extremely large matrices. We illustrate next the technique introduced by Pavlis and Saleh (2005), using geoid undulations as an example of a model-derived quantity.

Consider the gravity anomaly computed from a combined model as being composed of two separate spectral parts:

$$\widehat{\Delta g} = \widehat{\Delta g}_L + \widehat{\Delta g}_H = \sum_{n=2}^{L} \widehat{\Delta g}_n + \sum_{n=L+1}^{H} \widehat{\Delta g}_n, \tag{6.41}$$

where, L and H stand for *Low*- and *High*-degree, respectively. L denotes the maximum degree of the satellite-only model used to develop a combined solution which extends to degree and order H. The corresponding geoid undulation is then:

$$\widehat{N} = \widehat{N}_L + \widehat{N}_H = \sum_{n=2}^{L} \widehat{N}_n + \sum_{n=L+1}^{H} \widehat{N}_n, \tag{6.42}$$

and can be written as (Heiskanen and Moritz 1967, Eq. 2-163b):

$$\widehat{N} = \frac{R}{4\pi\gamma} \iint_{\sigma} \widehat{\Delta g}\, S(\psi)\, d\sigma. \tag{6.43}$$

The Stokes function $S(\psi)$ can also be decomposed into separate spectral components as (see ibid., Eq. 2-169):

$$\begin{aligned} S(\psi) &= \sum_{n=2}^{\infty} \frac{2n+1}{n-1}\, P_n(t) \\ &= \sum_{n=2}^{L} \frac{2n+1}{n-1} P_n(t) + \sum_{n=L+1}^{H} \frac{2n+1}{n-1} P_n(t) + \sum_{n=H+1}^{\infty} \frac{2n+1}{n-1} P_n(t) \\ &= S_L(\psi) + S_H(\psi) + S_\infty(\psi), \end{aligned} \tag{6.44}$$

where $t = \cos(\psi)$ and $P_n(t)$ is the Legendre polynomial of degree n. Substituting $S(\psi)$ in (6.43) by its three spectral components from (6.45), and considering (6.41), due to the orthogonality of spherical harmonics we have:

$$\widehat{N} = \frac{R}{4\pi\gamma} \iint_{\sigma} \left(\widehat{\Delta g}_L + \widehat{\Delta g}_H + 0\right) \cdot [S_L(\psi) + S_H(\psi) + S_\infty(\psi)]\, d\sigma \quad \Rightarrow$$

$$\widehat{N} = \frac{R}{4\pi\gamma} \iint_{\sigma} \widehat{\Delta g}_L\, S_L(\psi) d\sigma + \frac{R}{4\pi\gamma} \iint_{\sigma} \widehat{\Delta g}_H\, S_H(\psi) d\sigma = \widehat{N}_L + \widehat{N}_H. \tag{6.45}$$

Therefore, a strict, degree-wise separation of spectral components can be achieved by restricting the spectral content of the kernel function accordingly, *as long as* the integration is performed globally. The *High*-degree band-limited version of Stokes's equation:

$$\widehat{N}_H = \frac{R}{4\pi\gamma} \iint_{\sigma} \widehat{\Delta g}_H\, S_H(\psi)\, d\sigma\ , \tag{6.46}$$

implies, for *uncorrelated* errors of $\widehat{\Delta g}_H$, the following error propagation formulas:

$$\left.\begin{aligned} \sigma^2(\widehat{N}_H) &= \left(\frac{R}{4\pi\gamma}\right)^2 \iint_{\sigma} \sigma^2(\widehat{\Delta g}_H)\, S_H^2(\psi)\, d\sigma & (a) \\ \sigma_{12}(\widehat{N}_H) &= \left(\frac{R}{4\pi\gamma}\right)^2 \iint_{\sigma} \sigma^2(\widehat{\Delta g}_H)\, S_H(\psi_1)\, S_H(\psi_2)\, d\sigma & (b) \end{aligned}\right\}. \tag{6.47}$$

Equation 6.47a provides the error variance of the high-degree geoid undulation component, while (6.47b) the error covariance of the same component between two points located at ψ_1 and ψ_2 spherical distance respectively. Discretized versions of (6.47a, b) allow the computation of $\sigma^2(\widehat{N}_H)$ and $\sigma_{12}(\widehat{N}_H)$ from $\sigma^2(\widehat{\Delta g}_H)$ through global convolutions. One can implement (6.47a) using the 1D FFT approach of Haagmans et al. (1993), with H covering the degree range where the merged (terrestrial plus altimetry-derived) Δg define solely the solution. The geoid error covariances from (6.47b) can also be computed using global convolution, although with considerably less efficiency compared to the computation of error variances for points on regular grids. This approach is applicable to any functional f, related to Δg by an integral formula. Pavlis and Saleh (2005) provide the functional relationships required to propagate a model's error onto gravity anomalies, gravity disturbances, geoid undulations, and the components of the deflection of the vertical. Equations like (6.47a, b) employ the spherical approximation, which is considered quite adequate for error propagation work. Apart from this, these equations are rigorous, and their numerical implementation is only subject to discretization errors. Finally, the band limiting of integration kernels removes the singularity at the origin of kernels like Stokes's and Vening Meinesz's, therefore the innermost zone effects require no special treatment.

If we assume that the error correlation between $\widehat{\Delta g}_L$ and $\widehat{\Delta g}_H$ is negligible due to orthogonality, then the total error variance of a field functional, f, at the geographic location $(r,\ \varphi,\ \lambda)$, as computed from a specific gravitational model, can be written as:

$$\begin{aligned}\sigma_f^2(r,\varphi,\lambda) \approx\ & \sigma_f^2(r,\varphi,\lambda)\_commission\_L \\ & +\sigma_f^2(r,\varphi,\lambda)\_commission\_H\ . \\ & +\sigma_f^2(r,\varphi,\lambda)\_omission\end{aligned} \tag{6.48}$$

$\sigma_f^2(r,\varphi,\lambda)\_commission\_L$ can be computed by propagation of the complete error variance-covariance matrix resulting from the least-squares adjustment that produced the combined solution, employing, e.g., the 2D FFT approach of Haagmans and Van Gelderen (1991). $\sigma_f^2(r,\varphi,\lambda)\_commission\_H$ can be computed by global convolution based on an integral formula as we illustrated above for the case of geoid undulations. Finally, $\sigma_f^2(r,\varphi,\lambda)\_omission$ may be estimated using, e.g., some local covariance model. This approach does not require one to form, invert, and propagate extremely large matrices. Figure 6.4 shows the propagated error of the geoid undulations computed from the EGM2008 model up to degree and order 2,159. This computation was performed on a global 5′ × 5′ grid. Corresponding computations were also performed for gravity anomalies and for the deflection of the vertical components. In this fashion, the estimation of the propagated error of some specific functional at an arbitrary $(\varphi,\ \lambda)$ location can be easily performed using interpolation, given the pre-computed global 5′ × 5′ grid of the propagated error of the functional in question.

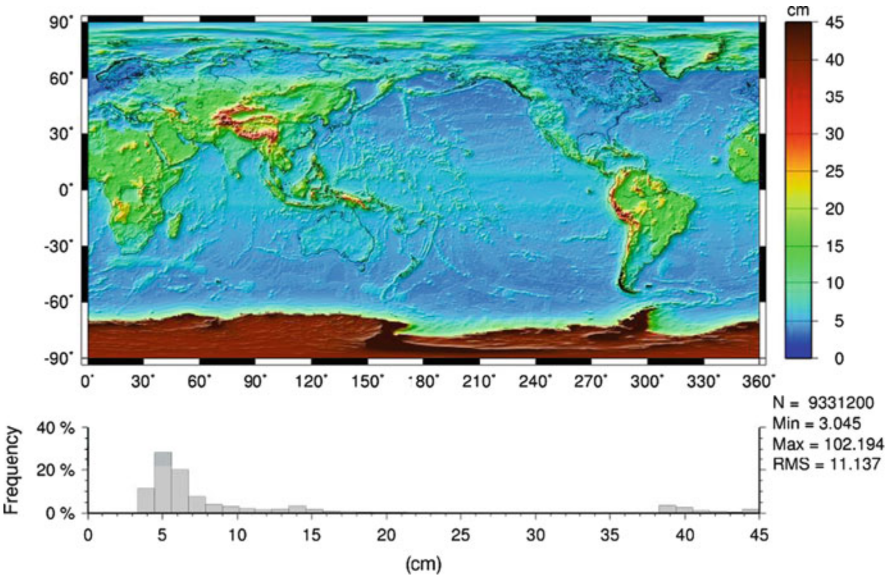

Fig. 6.4 Propagated error estimates (in centimeters) of the geoid undulations computed using the EGM2008 model to degree and order 2,159

6.6.2.5 Accuracy Assessment

The propagated error estimates of any global gravitational model depend strongly on the error estimates that were assigned to the data used in its development. Many times these data error estimates differ significantly from the true accuracy of the data. In addition, certain assumptions that may have been introduced within the development of a model (e.g., the assumption that the errors of the data are uncorrelated) could affect significantly the reliability of the model's propagated error estimates. In contrast, comparisons of model-derived quantities with data *independent* from the model, are in principle capable of revealing the true accuracy of a model. Of course, this requires the independent data that are used to test a model to be of significantly higher accuracy than the model-derived quantities. Such comparisons with independent data serve two general purposes:

(a) Evaluation of the accuracy of a model and inter-comparison of the performance of competing models.
(b) "Calibration" of the propagated error estimates. The comparison between the observed discrepancies between model-derived quantities and independent data, and the propagated errors of the model, allows one to test the veracity of the propagated error estimates, and to "calibrate" these estimates so that they match the observed performance of the model.

Several different data types, withheld from a model's development, are used for these purposes. These independent data have different spectral sensitivities and/or

Table 6.3 GPS/Leveling comparisons over CONUS

Model (Nmax)	Bias removed		Linear trend removed	
	Number passed edit	Weighted std. deviation (cm)	Number passed edit	Weighted std. deviation (cm)
EGM96 (360)	4,096	21.4	4,092	18.2
GGM02C_EGM96	4,169	18.9	4,165	17.6
EIGEN-GL04C (360)	4,167	19.5	4,163	18.1
EGM2008 (360)	4,185	17.6	4,181	16.4
EGM2008 (2,190)	4,201	7.1	4,197	4.8
USGG03 ($1' \rightarrow$10,800)	4,201	9.1	4,197	5.8

occupy different geographic regions. Tests that are usually employed here include satellite orbit determination and comparisons with tracking data, comparisons with geoid undulations obtained from GPS and leveling data (see e.g., Pavlis et al. 1993), comparisons with deflections of the vertical obtained from astrogeodetic techniques (Jekeli 1999b), comparisons employing altimetry and general ocean circulation models, etc. A useful practice, introduced during the development of EGM96, and used also during the development of EGM2008, is to invite a voluntary evaluation working group, independent of a model's developers, that evaluates and provides feedback to the model's developers regarding candidate preliminary solutions, as well as the final outcome from a modeling effort, in a manner as objective as possible. These groups usually work under the auspices of the International Association of Geodesy (IAG) and upon completion of a certain evaluation effort they report their findings in IAG-sponsored publications, which can be accessed freely by the public. In the case of EGM2008, the results from such an evaluation of both a preliminary version of the model (PGM2007A) (Pavlis et al., 2007b), as well as the final version of it, by 25 different international investigating teams are reported in *Newton's Bulletin No. 4*, which is jointly published by the Bureau Gravimétrique International (BGI) and the International Geoid Service (IGeS).

As an example of the evaluation of the EGM2008 and other models, Table 6.3 summarizes the results from the comparison of geoid undulations computed from GPS positioning and spirit leveling to model-derived values, over the conterminous United States (CONUS). A (thinned) set consisting of 4201 GPS/Leveling stations was used in this comparison. A ± 2 m editing criterion was applied to the differences between model-derived values and GPS/Leveling estimates. The analysis was done on a State by State basis, and the conversion from height anomalies to geoid undulations (Rapp, 1997b) was applied using a set of spherical harmonic coefficients of the elevation implied by the DTM2006.0 database (see Sect. 6.7), to a degree commensurate to the maximum degree of the gravitational model being tested. It is noteworthy that in this comparison, the EGM2008 model (which was developed based on $5' \times 5'$ area-mean gravity anomalies) performs better than the detailed ($1' \times 1'$) gravimetric geoid (USGG03), computed at the National Geodetic Survey (NGS) of the United States, using the most detailed *point* gravity anomaly data available for the area.

6.7 Data Requirements and Data Availability

The development of a GGM of very high degree and order requires a global set of gravity anomalies defined over a grid whose cell size is commensurate with the maximum degree of the expansion (e.g., $5' \times 5'$ for expansions to degree and order 2,160). One can form such a global grid, by merging gravity anomaly data obtained from terrestrial, ship-borne, air-borne, and satellite altimeter measurements. In addition to these data, elevation information in the form of a global Digital Topographic Model (DTM) is required. The resolution of this DTM should be considerably higher than the resolution of the gravity anomaly data grid to be compiled. We review next the essential aspects of these data requirements and describe the data that were available for the development of the EGM2008 model.

6.7.1 Elevation Data

The pre-processing and analysis of the detailed surface gravity data necessary to support the development of a GGM to harmonic degree and order 2,160, depends critically on the availability of accurate topographic data, at a resolution sufficiently higher than the resolution of the area-mean gravity anomalies, which will be used eventually for the development of the GGM. In Lemoine et al. (1998, Sect. 2.1) *Factor* discusses some of the uses of such topographic data within the context of the development and the subsequent use of a high-resolution GGM. These include the computation of Residual Terrain Model (RTM) effects, the computation of analytical continuation terms (g_1), the computation of Topographic/Isostatic gravitational models that may be used to “fill-in” areas void of other data, and the computation of models necessary to convert height anomalies to geoid undulations (Rapp, 1997b). For these computations to be made consistently, it is necessary to compile first a high-resolution global Digital Topographic Model (DTM), whose data will support the computation of all these terrain-related quantities.

For EGM96 (Lemoine et al. 1998), which was complete to degree and order 360, a global digital topographic database (JGP95E) at $5' \times 5'$ resolution was considered sufficient. JGP95E was formed by merging data from 29 individual sources, and, as acknowledged by its developers, left a lot to be desired in terms of accuracy and global consistency. Since that time, and thanks primarily to the Shuttle Radar Topography Mission (SRTM) (Werner 2001), significant progress has been made on the topographic mapping of the Earth from space. During approximately 11 days in 2000 (February 11–22), the SRTM collected data within latitudes 60°N and 56°S, thus covering approximately 80% of the total landmass of the Earth with elevation data of high, and fairly uniform, accuracy. Rodriguez et al. (2005) discuss in detail the accuracy characteristics of the SRTM elevations. Comparisons with ground control points whose elevations were determined independently using kinematic GPS positioning, indicate that the 90% absolute error of the SRTM elevations ranges from ±6 to ±10 m, depending on the geographic area (ibid., Table 2.1). Additional information regarding the SRTM can be obtained from the web site of the United States' Geological Survey (USGS) (http://srtm.usgs.gov/), and from the web site of NASA's Jet Propulsion Laboratory (http://www2.jpl.nasa.gov/srtm).

In preparation for the development of the EGM2008 model, we compiled DTM2006.0 by overlying the SRTM data over the data of DTM2002 (Saleh and Pavlis 2003). In addition to the SRTM data, DTM2006.0 contains ice elevations derived from ICESat laser altimeter data over Greenland (Ekholm, personal communication 2005) and over Antarctica (DiMarzio, personal communication 2005). Over Antarctica, data from the "BEDMAP" project (http://www.antarctica.ac.uk/aedc/bedmap/) were also used to define ice and water column thickness. Over the ocean, DTM2006.0 contains essentially the same information as DTM2002, which originates in the estimates of bathymetry from altimetry data and ship depth soundings of Smith and Sandwell (1997). DTM2006.0 was compiled in $30'' \times 30''$ resolution (providing height and depth information only), and in $2' \times 2'$ and $5' \times 5'$ resolutions, where lake depth and ice thickness data are also included. DTM2006.0 is identical to DTM2002 in terms of database structure and information content. Pavlis et al. (2007a) provide details about the DTM2006.0 database and its use towards the development and implementation of the EGM2008 model.

6.7.2 Terrestrial Gravity Anomaly Data

For the development of EGM2008, terrestrial gravity anomaly data were compiled in the form of $5' \times 5'$ area-mean values. These values were estimated from point gravity measurements using Least Squares Collocation (LSC) (Moritz 1980), following the general approach described by Kenyon and Pavlis (1996). Ship-borne data were also used (primarily near the coasts), as well as airborne measurements where such measurements were available. Over certain areas, the terrestrial gravity data were limited to a resolution corresponding to $15' \times 15'$ area-mean values. In order to compile a global dataset with as much as possible uniform spectral content, capable of supporting the estimation of potential coefficients to degree 2,160, the spectral content of these gravity anomalies beyond degree 720 (corresponding to the $15' \times 15'$ resolution), was augmented with the gravitational information obtained from a global set of gravity anomalies implied by the Residual Terrain Model (RTM) effect (Forsberg 1984). This approach was initially tested and verified over areas where high quality gravity data are available (USA, Australia), as Pavlis et al. (2007a) discuss in more detail. The gravity anomalies synthesized in this fashion were designated as "fill-in" data.

Despite the improvements in gravity anomaly resolution, coverage, and accuracy that were realized during the EGM2008 modeling effort, there are still many areas of the globe (most notably Antarctica) where gravity anomaly data are sparse, poor in accuracy, or completely non-existent. In addition, the coverage and quality of the available marine gravity data leave a lot to be desired. Pavlis (1988) demonstrated that long-wavelength errors present in the available marine gravity anomalies are a major contributor to the inconsistencies observed between satellite-only and surface gravity-only solutions. Marine gravity data are important to aid the separation (at least over short wavelengths) between the geoid undulation and the DOT signals, within the altimetry-derived sea surface height measurements. Efforts should therefore continue to try and improve the present status of the marine gravity data availability and quality.

6.7.3 Altimetry-Derived Gravity Anomalies

Two sources of altimetry-derived gravity anomalies were used for the compilation of the global $5' \times 5'$ area-mean gravity anomaly file used for the development of EGM2008. One set was estimated at the Danish National Space Center (DNSC), and was made available near the end of year 2007. This set was (internally) designated DNSC07C. DNSC07C is a predecessor of the DNSC08GRA data set that is described by Andersen et al. (2010b). The other set of altimetry-derived gravity anomalies was estimated in a collaborative effort between the Scripps Institution of Oceanography (SIO) and the National Oceanic and Atmospheric Administration (NOAA). The SIO/NOAA set is a predecessor of the data set described by Sandwell and Smith (2009). Preliminary tests performed during the development of EGM2008 indicated that the DNSC07C dataset performed better than the SIO/NOAA set near the coastlines, with the opposite behavior being observed over open ocean areas. Accordingly, the two altimetry-derived sets were "spliced" together, so that over a zone of ~ 190 km from the coast the DNSC07C set was used, followed by a "transition" zone of ~ 85 km where a weighted mean anomaly value computed from the two estimates (using complementary weights that vary linearly as one moves away from the coast), leading finally to 100% use of the SIO/NOAA values over the open ocean.

6.7.4 The Merged $5' \times 5'$ Area-Mean Gravity Anomaly File

To implement the BD estimation technique discussed in Sect. 6.6.2.2, one has to set up a global, complete file of $5' \times 5'$ area-mean gravity anomalies. Since the estimator does not allow for overlapping (duplicate) data input, one has to select for each $5'$ cell on the ellipsoid, the most accurate anomaly estimate out of multiple data that may be available for that cell (e.g., marine and altimetry-derived values). Rapp and Pavlis (1990) discuss such kind of data selection and merging algorithm. In the development of EGM2008, a similar (but not identical) algorithm was used. This process resulted in a complete global grid (9,331,200 values) of $5' \times 5'$ area-mean gravity anomalies, which were then input to the BD high-degree model estimator. Table 6.4 summarizes the overall statistics of this merged file.

Table 6.4 Statistics of the $5' \times 5'$ area-mean gravity anomalies of the merged file used to develop the EGM2008 model (Units are mGal)

Data source	% area	Minimum	Maximum	RMS	RMS σ
ArcGP	3.0	−192.0	281.8	30.2	3.0
Altimetry	63.2	−361.8	351.1	28.4	3.0
Terrestrial	17.6	−351.9	868.4	41.2	2.8
Fill-in	16.2	−333.0	593.5	46.8	7.6
Non Fill-in	83.8	−361.8	868.4	31.6	2.9
All	100.0	−361.8	868.4	34.5	4.1
(φ, λ)		19.4°, 293.5°	10.8°, 286.3°		

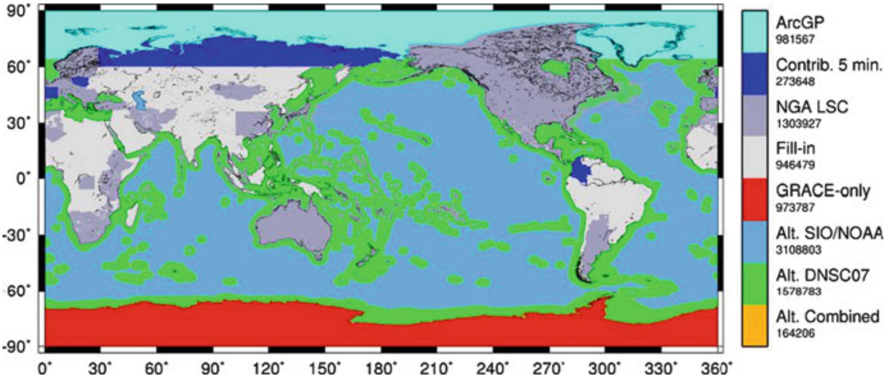

Fig. 6.5 Geographic distribution and source identification of the $5' \times 5'$ area-mean gravity anomalies in the merged file used to develop the EGM2008 model

Some noteworthy aspects of this merged file include the extensive use of $5' \times 5'$ area-mean gravity anomalies from the Arctic Gravity Project (ArcGP) (Kenyon and Forsberg 2008), and the avoidance of use of any Topographic/Isostatic mean anomalies (Pavlis and Rapp 1990). Over Antarctica, the $5' \times 5'$ area-mean gravity anomalies were synthesized purely on the basis of the ITG-GRACE03S (Mayer-Gürr 2007) model. This makes the EGM2008 model completely free of any isostatic hypothesis, at the cost of producing a smooth field over Antarctica (since ITG-GRACE03S is complete only up to degree and order 180). Figure 6.5 shows the geographic distribution and source identification of the $5' \times 5'$ area-mean gravity anomalies in the merged file used to develop the EGM2008 model.

6.8 Use of Global Gravitational Models and of Their By-Products

The estimated coefficients of the high-degree combination solution, $C_{nm}^{C,s}$, allow the user to compute the various functionals of the gravitational potential (e.g., gravity anomalies, height anomalies, deflections of the vertical), on or above the physical surface of the Earth, using harmonic synthesis. A versatile computer program

(HARMONIC_SYNTH), written in FORTRAN, which can be used to perform such harmonic synthesis, in various modes (e.g., for randomly scattered locations, for grids of point and/or area-mean values) was made available by Holmes and Pavlis (2006). This program, accompanied by test input and output files, and associated documentation is freely available from:

http://earth-info.nga.mil/GandG/wgs84/gravitymod/new_egm/new_egm.html

With regards to geoid computations, the user should also pay attention to some important issues related to the Permanent Tide, and the Geodetic Reference System (GRS) to which the computed geoid undulations (and/or height anomalies) refer. For example, in applications involving ellipsoidal heights obtained from space techniques (e.g., GPS), the user should be aware of the fact that the International Earth Rotation and Reference Systems Service (IERS), reports positions with respect to a (conventional) "Tide-Free" (also known as "Non-Tidal") crust. Therefore, in order to maintain consistency, geoid undulations and/or height anomalies involved in computations that use positions derived from space techniques, should be computed in the same Tide-Free system. In contrast, in applications involving satellite altimetry the "Mean Tide" system is used. Therefore, geoid undulations that are to be subtracted from altimetry-derived sea surface heights, in order to estimate the dynamic ocean topography, should also be computed in the Mean Tide system. The definition of the three systems used with regards to the Permanent Tide (Tide-Free, Mean, and Zero), and the relationships between the geoid undulations expressed in different systems is discussed in Lemoine et al. (1998, Chap. 11). This chapter is also available on-line from:

http://cddis.nasa.gov/926/egm96/doc/S11.HTML

In the same chapter, the issue of expressing the geoid undulations and/or height anomalies with respect to a specific GRS is discussed. In the case of EGM2008, the conversion from an "ideal" mean-Earth ellipsoid (whose semi-major axis remains numerically unspecified), in the Tide-Free system, and the WGS84 GRS, involves the application of a zero-degree height anomaly equal to -41 cm. The zero-degree height anomaly (ζ_z). that was computed when the WGS84 EGM96 geoid was released was equal to -53 cm (Lemoine et al. 1998, Chap. 11). The primary reason for the change in the numerical value of ζ_z from the EGM96 days to the current best estimate, is the discovery by Ouan-Zan Zanife (CLS, France) of an error in the Oscillator Drift correction applied to TOPEX altimeter data (Fu and Cazenave 2001, p. 34). The erroneous correction was producing TOPEX sea surface heights, biased by approximately 12–13 cm.

Under:
http://earth-info.nga.mil/GandG/wgs84/gravitymod/egm2008/egm08_wgs84.html
the user can find a modified version of the HARMONIC_SYNTH program, specifically designed to compute geoid undulations at arbitrarily scattered locations, in the Tide-Free system, with respect to the WGS84 GRS. In the same web site, the user can also find pre-computed global grids of these geoid undulations, at both $1' \times 1'$ and $2.5' \times 2.5'$ grid-spacing, as well as a FORTRAN program to interpolate from these grids.

6.9 Temporal Variations

The topic of temporal gravity field variations, although outside the main scope of the present discussion, cannot be omitted. Non-tidal temporal gravity field variations originate from mass redistribution within the entire solid Earth-Ocean-Atmosphere-Hydrosphere-Cryosphere system. Some of these variations have strong seasonal signals (e.g., variations in the atmosphere and hydrosphere) while others are episodic (e.g., redistribution of mass due to seismic activity). Until recently, Satellite Laser Ranging (SLR) data were the only source of information based on which temporal variations in a handful of very low-degree harmonic coefficients of the gravitational field could be determined (see e.g., Cheng et al. 1997). As a result of the success of the GRACE mission, this situation has changed dramatically in recent years. GRACE offers the capability of constant monitoring of gravitational field variations, with a temporal resolution of approximately 1 month, and a spatial resolution of approximately 400 km. This has opened up an entirely new area of geodetic research and of geodetic contributions towards the establishment of an Earth Observing System, especially in view of its importance in areas related to global Climate Change (e.g., polar ice melting). Under:

http://www.csr.utexas.edu/grace/publications/citation.html

the interested reader can find a plethora of publications involving the use of GRACE data to address a wide variety of science topics.

6.10 Outlook

It is becoming increasingly clear these days that the demarcation between global and regional (or local) gravimetric approximation studies is shifting (if not disappearing altogether). The satellite data that have become available from missions like GRACE and GOCE is prompting some geodesists that used to focus their efforts on local gravimetric studies, to consider also global problems. On the other side, the increasing availability of detailed gravimetric data prompts some global modelers to increase the resolution of their models, effectively "stepping" into the spectral regime that was considered traditionally part of the regional or local approximation work. Mathematical innovations that could facilitate the bridging of any existing gap between these two regimes and provide (better) solutions to some of the problems identified before are therefore highly desirable.

Improvements in gravimetric data coverage and quality are still necessary over vast areas, especially in Antarctica, South America, Africa, and parts of Asia. Airborne gravimetric surveys have provided a wealth of data over remote areas that are very difficult to access and survey otherwise (e.g., Greenland). Such data acquisition techniques currently offer the best means of filling-in the existing gravimetric data gaps.

Innovative analysis techniques have been developed and are constantly being refined. These techniques, and the availability of ever more capable computers, have enabled geodesists to process vast amounts of data on a more or less routine basis these days. But the geodesist's "appetite" for increased accuracy and resolution keeps challenging even some of the most capable computers that are available today.

While some of the traditional geodetic problems may have been solved to a satisfactory degree of accuracy (which is indicative of the progress made within the discipline), the important role of geodesy in the monitoring of the evolving Earth System opens up new possibilities for innovative work. The detection and monitoring of minute changes in the gravitational field is quickly becoming a valuable tool for the study of Climate Change. So, while the character of global gravimetric problems may be changing, new challenges arise, and the future of the discipline seems to this author to be limited only by the imagination and innovation of its practitioners.