Geodesy

Petr Vaníček

University of New Brunswick

GLOSSARY

Coordinates These are the numbers that define positions in a specific coordinate system. For a coordinate system to be usable (to allow the determination of coordinates) in the real (earth) space, its position and the orientation of its Cartesian axes in the real (earth) space must be known.

Coordinate system In three-dimensional Euclidean space, which we use in geodesy for solving most of the problems, we need either the Cartesian or a curvilinear coordinate system, or both, to be able to work with positions. The Cartesian system is defined by an orthogonal triad of coordinate axes; a curvilinear system is related to its associated generic Cartesian system through some mathematical prescription.

Ellipsoid/spheroid Unless specified otherwise, we understand by this term the geometrical body created by revolving an ellipse around it minor axis, consequently known as an ellipsoid of revolution. By spheroid, we understand a spherelike body, which, of course, includes an ellipsoid as well.

Errors (uncertainties) Inevitable, usually small errors of either a random or systematic nature, which cause uncertainty in every measurement and, consequently, an uncertainty in quantities derived from these observations.

GPS Global Positioning System based on the use of a flock of dedicated satellites.

Gravity anomaly The difference between actual gravity and model gravity, e.g., the normal gravity, where the two gravity values are related to two different points on the same vertical line. These two points have the same value of gravity potential: the actual gravity potential and the model gravity potential, respectively.

Normal gravity field An ellipsoidal model of the real gravity field.

Positioning (static and kinematic) This term is used in geodesy as a synonym for the "determination of positions" of different objects, either stationary or moving.

Satellite techniques Techniques for carrying out different tasks that use satellites and/or satellite-borne instrumentation.

Tides The phenomenon of the earth deformation, including its liquid parts, under the influence of solar and lunar gravitational variations.

WHAT IS GEODESY?

Geodesy is a science, the oldest earth (geo-) science, in fact. It was born of fear and curiosity, driven by a desire to predict natural happenings and calls for the understanding of these happenings. The classical definition, according to one of the "fathers of geodesy" reads: "Geodesy is the science of measuring and portraying the earth's surface" (Helmert, 1880, p. 3). Nowadays, we understand the scope of geodesy to be somewhat wider. It is captured by the following definition (Vaníček and Krakiwsky, 1986, p. 45): "Geodesy is the discipline that deals with the measurement and representation of the earth, including its gravity field, in a three-dimensional time varying space." Note that the contemporary definition includes the study of the earth gravity field (see Section III), as well as studies of temporal changes in positions and in the gravity field (see Section IV).

I. INTRODUCTION

A. Brief History of Geodesy

Little documentation of the geodetic accomplishments of the oldest civilizations, the Sumerian, the Egyptian, the Chinese, and the Indian, has survived. The first firmly documented ideas about geodesy go back to Thales of Miletus (ca. 625–547 BC), Anaximander of Miletus (ca. 611–545 BC), and the school of Pythagoras (ca. 580–500 BC). The Greek students of geodesy included Aristotle (384–322 BC), Eratosthenes (276–194 BC)—the first reasonably accurate determination of the size of the earth, but not taken seriously until 17 centuries later—and Ptolemy (ca. 75–151 AD).

In the Middle Ages, the lack of knowledge of the real size of the earth led Toscanelli (1397–1482) to his famous misinterpretation of the world (Fig. 1), which allegedly lured Columbus to his first voyage west. Soon after, the golden age of exploration got under way and with it the use of position determination by astronomical means. The real extent of the world was revealed to have been close to Eratosthenes's prediction, and people started looking for further quantitative improvements of their conceptual model of the earth. This led to new measurements on the surface of the earth by a Dutchman Snellius (in the 1610s) and a Frenchman Picard (in the 1670s) and the first improvement on Eratosthenes's results. Interested readers can find fascinating details about the oldest geodetic events in Berthon and Robinson (1991).

At about the same time, the notion of the earth's gravity started forming up through the efforts of a Dutchman Stevin (1548–1620), Italians Galileo (1564–1642) and Borelli (1608–1679), an Englishman Horrox (1619–1641), and culminating in Newton's (1642–1727) theory of gravitation. Newton's theory predicted that the earth's globe should be slightly oblate due to the spinning of the earth around its polar axis. A Frenchman Cassini (1625–1712) disputed this prediction; consequently, the French Academy of Science organized two expeditions to Peru and to Lapland under the leadership of Bouguer and

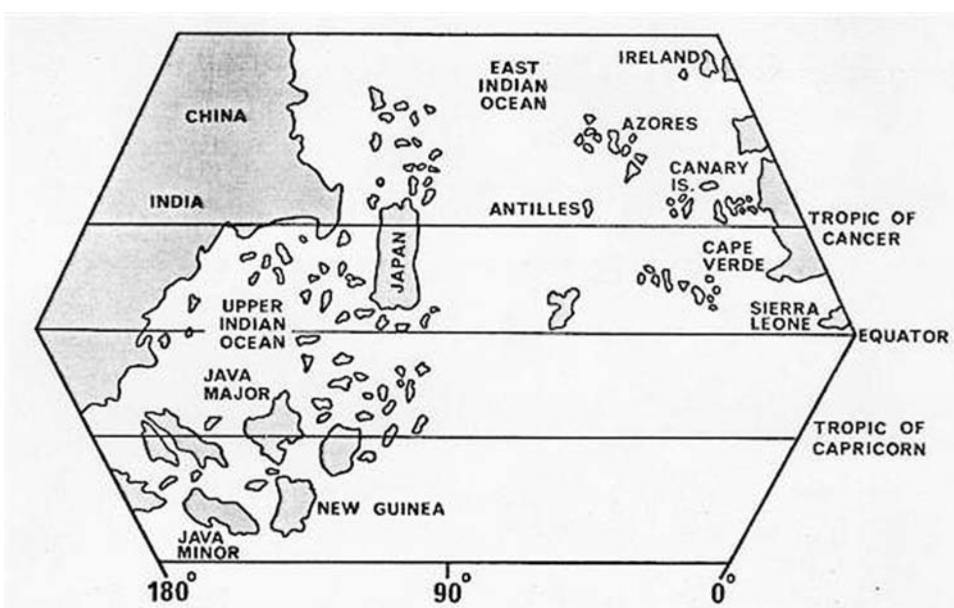

FIGURE 1 Toscanelli's view of the Western Hemisphere.

Maupertuis to measure two meridian arcs. The results confirmed the validity of Newton's prediction. In addition, these measurements gave us the first definition of a meter, as one ten-millionth part of the earth's quadrant.

For 200 years, from about mid-18th century on, geodesy saw an unprecedented growth in its application. Position determination by terrestrial and astronomical means was needed for making maps, and this service, which was naturally provided by geodesists and the image of a geodesist as being only a provider of positions, survives in some quarters till today. In the meantime, the evolution of geodesy as a science continued with contributions by Lagrange (1736–1813), Laplace (1749–1827), Fourier (1768–1830), Gauss (1777–1855), claimed by some geodesists to have been the real founder of geodetic science, Bessel (1784–1846), Coriolis (1792–1843), Stokes (1819–1903), Poincaré (1854–1912), and Albert Einstein. For a description of these contributions, see Vaníček and Krakiwsky (1986, Section 1.3).

B. Geodesy and Other Disciplines and Sciences

We have already mentioned that for more than 200 years geodesy—strictly speaking, only one part of geodesy, i.e., positioning—was applied in mapping in the disguise known on this continent as "control surveying." Positioning finds applications also in the realm of hydrography, boundary demarcation, engineering projects, urban management, environmental management, geography, and planetology. At least one other part of geodesy, geo-kinematic, finds applications also in ecology.

Geodesy has a symbiotic relation with some other sciences. While geodesy supplies geometrical information about the earth, the other geosciences supply physical knowledge needed in geodesy for modeling. Geophysics is the first to come to mind: the collaboration between geophysicists and geodesists is quite wide and covers many facets of both sciences. As a result, the boundary between the two sciences became quite blurred even in the minds of many geoscientists. For example, to some, the study of the global gravity field fits better under geophysics rather than geodesy, while the study of the local gravity field may belong to the branch of geophysics known as exploration geophysics. Other sciences have similar but somewhat weaker relations with geodesy: space science, astronomy (historical ties), oceanography, atmospheric sciences, and geology.

As all exact sciences, geodesy makes heavy use of mathematics, physics, and, of late, computer science. These form the theoretical foundations of geodetic science and, thus, play a somewhat different role vis-à-vis geodesy. In Fig. 2, we have attempted to display the three levels of relations in a cartoon form.

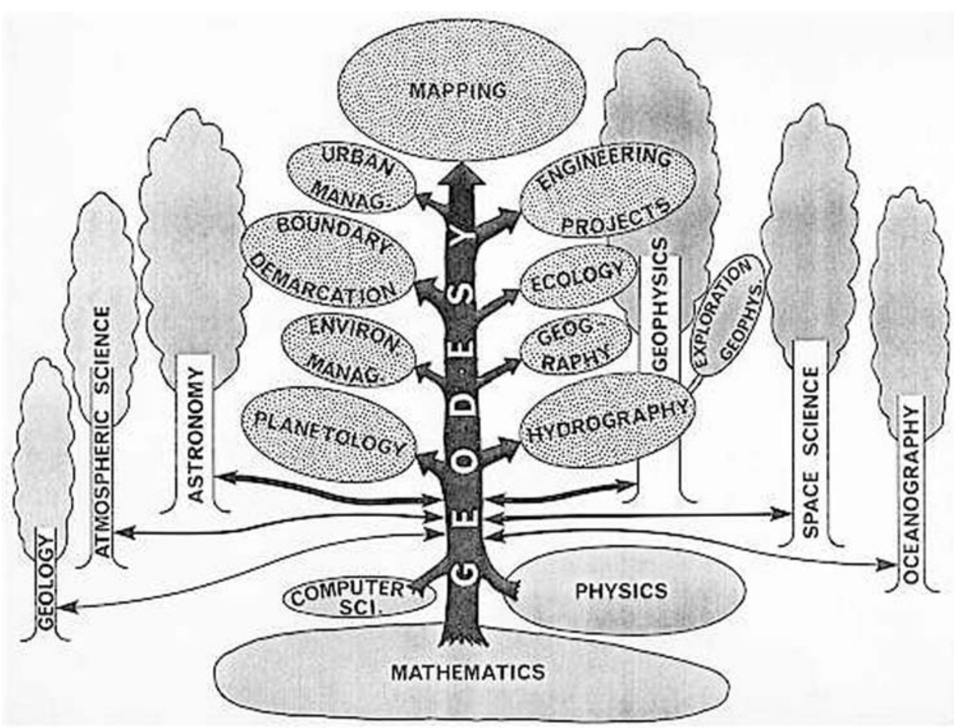

FIGURE 2 Geodesy and other disciplines.

C. Profession and Practice of Geodesy

Geodesy, as most other professions, spans activities ranging from purely theoretical to very applied. The global nature of geodesy dictates that theoretical work be done mostly at universities or government institutions. Few private institutes find it economically feasible to do geodetic research. On the other hand, it is quite usual to combine geodetic theory with practice within one establishment. Much of geodetic research is done under the disguise of space science, geophysics, oceanography, etc.

Of great importance to geodetic theory is international scientific communication. The international organization looking after geodetic needs is the International Association of Geodesy (IAG), the first association of the more encompassing International Union of Geodesy and Geophysics (IUGG) which was set up later in the first third of 20th century. Since its inception, the IAG has been responsible for putting forward numerous important recommendations and proposals to its member countries. It is also operating several international service outfits such as the International Gravimetric Bureau (BGI), the International Earth Rotation Service (IERS), Bureau Internationale des Poids et Mesures—Time Section (BIPM), the International GPS Service (IGS), etc. The interested reader would be well advised to check the current services on the IAG web page.

Geodetic practice is frequently subjugated to mapping needs of individual countries, often only military mapping needs. This results in other components of geodetic work being done under the auspices of other professional institutions. Geodesists practicing positioning are often lumped together with surveyors. They find a limited international forum for exchanging ideas and experience in the International Federation of Surveyors (FIG), a member of the Union of International Engineering Organizations (UIEO).

The educational requirements in geodesy would typically be a graduate degree in geodesy, mathematics, physics, geophysics, etc. for a theoretical geodesist and an undergraduate degree in geodesy, surveying engineering (or geomatics, as it is being called today), survey science, or a similar program for an applied geodesist. Survey technicians, with a surveying (geomatics) diploma from a college or a technological school, would be much in demand for field data collection and routine data manipulations.

II. POSITIONING

A. Coordinate Systems Used in Geodesy

Geodesy is interested in positioning points on the surface of the earth. For this task a well-defined coordinate system is needed. Many coordinate systems are being used in geodesy, some concentric with the earth (geocentric systems), some not. Also, both Cartesian and curvilinear coordinates are used. There are also coordinate systems needed specifically in astronomical and satellite positioning, which are not appropriate to describe positions of terrestrial points in.

Let us discuss the latter coordinate systems first. They are of two distinct varieties: the apparent places and the orbital. The apparent places (AP) and its close relative, the right ascension (RA) coordinate systems, are the ones in which (angular) coordinates of stars are published. The orbital coordinate systems (OR) are designed to be used in describing satellite positions and velocities. The relations between these systems and with the systems introduced below will be discussed in Section II.F. Interested readers can learn about these coordinate systems in Vaníček and Krakiwsky (1986, Chap. 15).

The *geocentric systems* have their z-axis aligned either with the instantaneous spin axis (cf., Section IV.B) of the earth (*instantaneous terrestrial system*) or with a hypothetical spin axis adopted by a convention (*conventional terrestrial systems*). The geocentric systems became useful only quite recently, with the advent of satellite positioning. The nongeocentric systems are used either for local work (observations), in which case their origin would be located at a point on the surface of the earth (topocentric systems called *local astronomic* and *local geodetic*), or for a regional/continental work in the past. These latter nongeocentric (near-geocentric) systems were and are used in lieu of geocentric systems, when these were not yet realizable, and are known as the *geodetic systems*; their origin is usually as close to the center of mass of the earth as the geodesists of yesteryear could make it. They miss the center of mass by anything between a few meters and a few kilometers, and there are some 150 of them in existence around the world.

Both the geocentric and geodetic coordinate systems are used together with *reference ellipsoids* (ellipsoids of revolution or biaxial ellipsoids), also called in some older literature "spheroids." (The modern usage of the term spheroid is for closed, spherelike surfaces, which are more complicated than biaxial ellipsoids.) These reference ellipsoids are taken to be concentric with their coordinate system, geocentric or near geocentric, with the axis of revolution coinciding with the z-axis of the coordinate system. The basic idea behind using the reference ellipsoids is that they fit the real shape of the earth, as described by the geoid (see Section III.B for details) rather well and can thus be regarded as representative, yet simple, expression of the shape of the earth.

The reference ellipsoids are the horizontal surfaces to which the geodetic latitude and longitude are referred, hence, the name. But to serve in this role, an ellipsoid (together with the associated Cartesian coordinate system) must be fixed with respect to the earth. Such an ellipsoid (fixed with respect to the earth) is often called a *horizontal datum*. In North America we had the North American Datum of 1927, known as NAD 27 (U.S. Department of Commerce, 1973) which was replaced by the geocentric North American Datum of 1983, referred to as NAD 83 (Boal and Henderson, 1988; Schwarz, 1989).

The horizontal geodetic coordinates, *latitude* φ and *longitude* λ, together with the *geodetic height* h (called by some authors by ellipsoidal height, a logical *nonsequitur*, as we shall see later), make the basic triplet of curvilinear coordinates widely used in geodesy. They are related to their associated Cartesian coordinates x, y, and z by the following simple expressions:

$$\begin{aligned} x &= (N+h)\cos\varphi\cos\lambda \\ y &= (N+h)\cos\varphi\sin\lambda \\ z &= (Nb^2/a^2+h)\sin\varphi, \end{aligned} \tag{1}$$

where N is the local radius of curvature of the reference ellipsoid in the east–west direction,

$$N = a^2(a^2\cos^2\varphi + b^2\sin^2\varphi)^{-1/2}, \tag{2}$$

where a is the major semi-axis and b is the minor semi-axis of the reference ellipsoid. We note that the geodetic heights are not used in practice; for practical heights, see Section II.B. It should be noted that the horizontal geodetic coordinates are the ones that make the basis for all maps, charts, legal land and marine boundaries, marine and land navigation, etc. The transformations between these horizontal coordinates and the two-dimensional Cartesian coordinates x, y on the maps are called *cartographic mappings*.

Terrestrial (geocentric) coordinate systems are used in satellite positioning. While the instantaneous terrestrial (IT) system is well suited to describe instantaneous positions in, the conventional terrestrial (CT) systems are useful for describing positions for archiving. The conventional terrestrial system recommended by IAG is the International Terrestrial Reference System (ITRS), which is "fixed to the earth" via several permanent stations whose horizontal tectonic velocities are monitored and recorded.

gets associated with the time of fixing by time tagging. The "realization" of the ITRS by means of coordinates of some selected points is called the International Terrestrial Reference Frame (ITRF). Transformation parameters needed for transforming coordinates from one epoch to the next are produced by the International Earth Rotation Service (IERS) in Paris, so one can keep track of the time evolution of the positions. For more detail the reader is referred to the web site of the IERS or to a popular article by Boucher and Altamini (1996).

B. Point Positioning

It is not possible to determine either three-dimensional (3D) or two-dimensional (2D) (horizontal) positions of isolated points on the earth's surface by terrestrial means. For point positioning we must be looking at celestial objects, meaning that we must be using either optical techniques to observe stars [geodetic astronomy, see Mueller (1969)] or electronic/optical techniques to observe earth's artificial satellites (satellite positioning, cf., Section V.B). Geodetic astronomy is now considered more or less obsolete, because the astronomically determined positions are not very accurate (due to large effects of unpredictable atmospheric refraction) and also because they are strongly affected by the earth's gravity field (cf., Section III.D). Satellite positioning has proved to be much more practical and more accurate.

On the other hand, it is possible to determine heights of some isolated points through terrestrial means by tying these points to the sea level. Practical heights in geodesy, known as *orthometric heights* and denoted by H^{O}, or simply by H, are referred to the geoid, which is an equipotential surface of the earth's gravity field (for details, see Section III.B) approximated by the mean sea level (MSL) surface to an accuracy of within ± 1.5 m. The difference between the two surfaces arises from the fact that seawater is not homogeneous and because of a variety of dynamical effects on the seawater. The height of the MSL above the geoid is called the *sea surface topography* (SST). It is a very difficult quantity to obtain from any measurements; consequently, it is not yet known very accurately. We note that the orthometric height H is indeed different from the geodetic height h discussed in Section II.A: the relation between the two kinds of heights is shown in Fig. 3, where the quantity N, the height of the geoid above the reference ellipsoid, is usually called the *geoidal height* (geoid

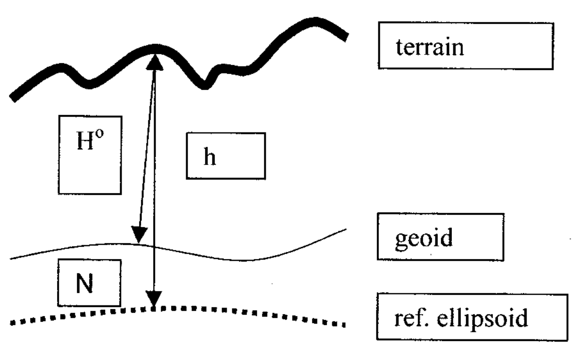

FIGURE 3 Orthometric and geodetic heights.

undulation) (cf., Section III.B). Thus, the knowledge of the geoid is necessary for transforming the geodetic to orthometric heights and vice versa. We note that the acceptance of the standard geodetic term of "geoidal height" (height of the geoid above the reference ellipsoid) makes the expression "ellipsoidal height" for (geodetic) height of anything above the reference ellipsoid, a logical *non-sequitur* as pointed out above.

We have seen above that the geodetic height is a purely geometrical quantity, the length of the normal to the reference ellipsoid between the ellipsoid and the point of interest. The orthometric height, on the other hand, is defined as the length of the plumbline (a line that is always normal to the equipotential surface of the gravity field) between the geoid and the point of interest and, as such, is intimately related to the gravity field of the earth. (As the plumbline is only slightly curved, the length of the plumbline is practically the same as the length of the normal to the geoid between the geoid and the point of interest. Hence, the equation $h \cong H + N$ is valid everywhere to better than a few millimeters.) The defining equation for the orthometric height of point A (given by its position vector $\mathbf{r}_A$) is

$$H^O(\mathbf{r}_A) = H(\mathbf{r}_A) = [W_0 - W(\mathbf{r}_A)]/\mathrm{mean}(g_A), \quad (3)$$

where W_0 stands for the constant gravity potential on the geoid, $W(\mathbf{r}_A)$ is the gravity potential at point A, and $\mathrm{mean}(g_A)$ is the mean value of gravity (for detailed treatment of these quantities, see Sections III.A and III.B) between A and the geoid—these. From these equations it can be easily gleaned that orthometric heights are indeed referred to the geoid (defined as $W_0 = 0$). The $\mathrm{mean}(g)$ cannot be measured and has to be estimated from gravity observed at A, $g(\mathbf{r}_A)$, assuming a reasonable value for the vertical gradient of gravity within the earth. Helmert (1880) hypothesized the value of 0.0848 mGal m^{-1} suggested independently by Poincaré and Prey to be valid for the region between the geoid and the earth's surface (see Section III.C), to write

$$\mathrm{mean}(g_A) \cong g(\mathbf{r}_A) + 0.0848\, H(\mathbf{r}_A)/2\, [\mathrm{mGal}]. \quad (4)$$

For the definition of units of gravity, Gal, see Section III.A. Helmert's (approximate) orthometric heights are used for mapping and for technical work almost everywhere. They may be in error by up to a few decimeters in the mountains. Equipotential surfaces at different heights are not parallel to the equipotential surface at height 0, i.e., the geoid. Thus, orthometric heights of points on the same equipotential surface $W = \mathrm{const.} \neq W_0$ are generally not the same and, for example, the level of a lake appears to be sloping. To avoid this, and to allow the physical laws to be taken into proper account, another system of height is used: *dynamic heights*. The dynamic height of point A is defined as

$$H^D(\mathbf{r}_A) = [W_0 - W(\mathbf{r}_A)]/\gamma_{\mathrm{ref}}, \quad (5)$$

where γ_{ref} is a selected (reference) value of gravity, constant for the area of interest. We note that points on the same equipotential surface have the same dynamic height; that dynamic heights are referred to the geoid but they must be regarded as having a scale that changes from point to point.

We must also mention the third most widely used height system, the *normal heights*. These heights are defined by

$$H^N(\mathbf{r}_A) = H^*(\mathbf{r}_A) = [W_0 - W(\mathbf{r}_A)]/\mathrm{mean}(\gamma_A), \quad (6)$$

where $\mathrm{mean}(\gamma_A)$ is the value of the model gravity called "normal" (for a detailed explanation, see Section III.A) at a height of $H^N(\mathbf{r}_A)/2$ above the reference ellipsoid along the normal to the ellipsoid (Molodenskij, Eremeev, and Yurkina, 1960). We refer to this value as mean because it is evaluated at a point halfway between the reference ellipsoid and the locus of $H^N(\mathbf{r}_A)$, referred to the reference ellipsoid, which (locus) surface is called the *telluroid*. For practical purposes, normal heights of terrain points A are referred to a different surface, called *quasi-geoid* (cf., Section III.G), which, according to Molodenskij, can be computed from gravity measurements in a similar way to the computation of the geoid.

C. Relative Positioning

Relative positioning, meaning positioning of a point with respect to an existing point or points, is the preferred mode of positioning in geodesy. If there is intervisibility between the points, terrestrial techniques can be used. For satellite relative positioning, the intervisibility is not a requirement, as long as the selected satellites are visible from the two points in question. The accuracy of such relative positions is usually significantly higher than the accuracy of single point positions.

The classical terrestrial techniques for 2D relative positioning make use of angular (horizontal) and distance measurements, which always involve two or three points. These techniques are thus differential in nature. The computations of the relative 2D positions are carried out either on the horizontal datum (reference ellipsoid), in terms of latitude difference $\Delta\varphi$ and longitude difference $\Delta\lambda$, or on a map, in terms of Cartesian map coordinate differences Δx and Δy. In either case, the observed angles, azimuths, and distances have to be first transformed (reduced) from the earth's surface, where they are acquired, to the reference ellipsoid, where they are either used in the computations or transformed further onto the selected mapping plane. We shall not explain these reductions here; rather, we would advise the interested reader to consult one of the classical geodetic textbooks (e.g., Zakatov, 1953; Bomford, 1971).

To determine the relative position of one point with respect to another on the reference ellipsoid is not a simple proposition, since the computations have to be carried out on a curved surface and Euclidean geometry no longer applies. Links between points can no longer be straight lines in the Euclidean sense; they have to be defined as geodesics (the shortest possible lines) on the reference ellipsoid. Consequently, closed form mathematical expressions for the computations do not exist, and use has to be made of various series approximations. Many such approximations had been worked out, which are valid for short, medium, or long geodesics. For 200 years, coordinate computations on the ellipsoid were considered to be the backbone of (classical) geodesy, a litmus test for aspiring geodesists. Once again, we shall have to desist from explaining the involved concepts here as there is no room for them in this small article. Interested readers are referred once more to the textbooks cited above.

Sometimes, preference is given to carrying out the relative position computations on the mapping plane, rather than on the reference ellipsoid. To this end, a suitable cartographic mapping is first selected, normally this would be the conformal mapping used for the national/state geodetic work. This selection carries with it the appropriate mathematical mapping formulae and distortions associated with the selected mapping (Lee, 1976). The observed angles ω, azimuths α, and distances S (that had been first reduced to the reference ellipsoid) are then reduced further (distorted) onto the selected mapping plane where (2D) Euclidean geometry can be applied. This is shown schematically in Fig. 4. Once these reductions have been carried out, the computation of the (relative) position of the unknown point B with respect to point A already known on the mapping plane is then rather trivial:

$$x_{\mathrm{B}} = x_{\mathrm{A}} + \Delta x_{\mathrm{AB}}, \qquad y_{\mathrm{B}} = y_{\mathrm{A}} + \Delta y_{\mathrm{AB}}. \tag{7}$$

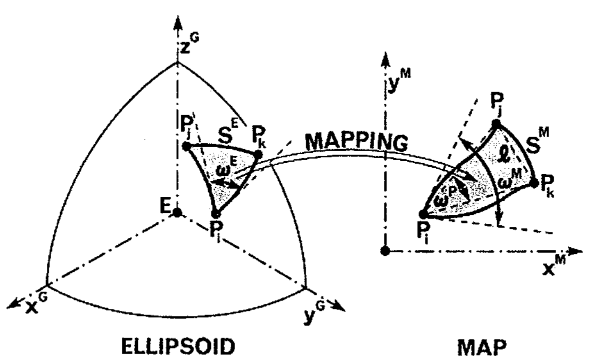

FIGURE 4 Mapping of ellipsoid onto a mapping plane.

Relative vertical positioning is based on somewhat more transparent concepts. The process used for determining the height difference between two points is called *geodetic levelling* (Bomford, 1971). Once the levelled height difference is obtained from field observations, one has to add to it a small correction based on gravity values along the way to convert it to either the orthometric, the dynamic, or the normal height difference. Geodetic levelling is probably the most accurate geodetic relative positioning technique. To determine the geodetic height difference between two points, all we have to do is to measure the vertical angle and the distance between the points. Some care has to be taken that the vertical angle is reckoned from a plane perpendicular to the ellipsoidal normal at the point of measurement.

Modern extraterrestrial (satellite and radio astronomical) techniques are inherently three dimensional. Simultaneous observations at two points yield 3D coordinate differences that can be added directly to the coordinates of the known point A on the earth's surface to get the sought coordinates of the unknown point B (on the earth's surface). Denoting the triplet of Cartesian coordinates (x, y, z) in any coordinate system by $\mathbf{r}$ and the triplet of coordinate differences $(\Delta x, \Delta y, \Delta z)$ by $\Delta\mathbf{r}$, the 3D position of point B is given simply by

$$\mathbf{r}_{\mathrm{B}} = \mathbf{r}_{\mathrm{B}} + \Delta\mathbf{r}_{\mathrm{AB}}, \tag{8}$$

where $\Delta\mathbf{r}_{\mathrm{AB}}$ comes from the observations.

We shall discuss in Section V.B how the "base vector" $\Delta\mathbf{r}_{\mathrm{AB}}$ is derived from satellite observations. Let us just mention here that $\Delta\mathbf{r}_{\mathrm{AB}}$ can be obtained also by other techniques, such as radio astronomy, inertial positioning, or simply from terrestrial observations of horizontal and vertical angles and distances. Let us show here the principle of the interesting radio astronomic technique for the determination of the base vector, known in geodesy as *Very Long Baseline Interferometry* (VLBI). Figure 5 shows schematically the pair of radio telescopes (steerable antennas, A and B) following the same quasar whose celestial position is known (meaning that $\mathbf{e}_s$ is known). The time delay τ can be measured very accurately and the base vector $\Delta\mathbf{r}_{\mathrm{AB}}$ can be evaluated from the following equation:

$$\tau = c^{-1}\,\mathbf{e}_s \Delta\mathbf{r}_{\mathrm{AB}}, \tag{9}$$

where c is the speed of light. At least three such equations are needed for three different quasars to solve for $\Delta\mathbf{r}_{\mathrm{AB}}$.

Normally, thousands of such equations are available from dedicated observational campaigns. The most important contribution of VLBI to geodesy (and astronomy) is that it works with directions (to quasars) which can be considered as the best approximations of directions in an inertial space.

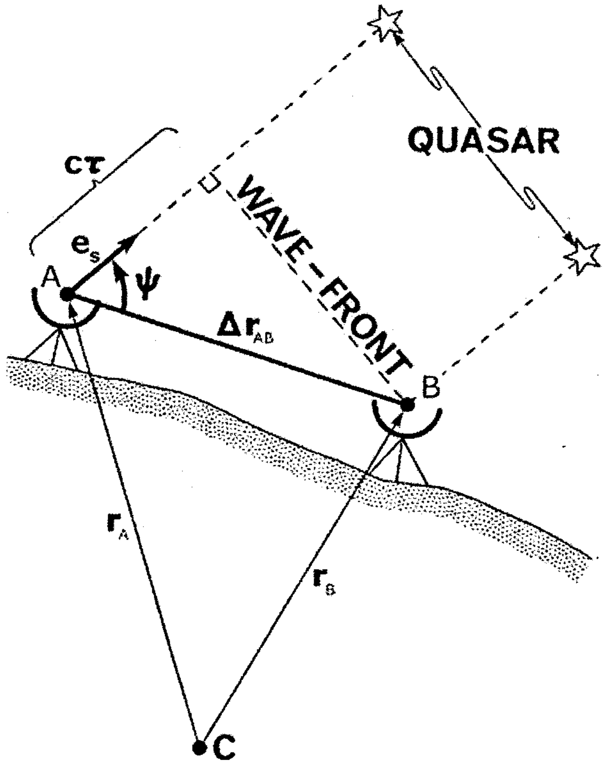

FIGURE 5 Radioastronomical interferometry.

D. Geodetic Networks

In geodesy we prefer to position several points simultaneously because when doing so we can collect redundant information that can be used to check the correctness of the whole positioning process. Also, from the redundancy, one can infer the internal consistency of the positioning process and estimate the accuracy of so determined positions (cf., Section II.E). Thus, the classical geodetic way of positioning points has been in the mode of *geodetic networks*, where a whole set of points is treated simultaneously. This approach is, of course, particularly suitable for the terrestrial techniques that are differential in nature, but the basic rationale is equally valid even for modern positioning techniques. After the observations have been made in the field, the positions of network points are estimated using optimal estimation techniques that minimize the quadratic norm of observation residuals from systems of (sometimes hundreds of thousands) overdetermined (observation) equations. A whole body of mathematical and statistical techniques dealing with network design and position estimation (*network adjustment*) has been developed; the interested reader may consult Grafarend and Sansò (1985), Hirvonen (1971), and Mikhail (1976) for details.

The 2D (horizontal) and 1D (vertical) geodetic networks of national extent, sometimes called national control networks, have been the main tool for positioning needed in mapping, boundary demarcation, and other geodetic applications. For illustration, the Canadian national geodetic levelling network is shown in Fig. 6. We note that national networks are usually interconnected to create continental networks that are sometimes adjusted together—as is the case in North America—to make the networks more homogeneous. Local networks in one, two, and three dimensions have been used for construction purposes. In classical geodetic practice, the most widely encountered networks are horizontal, while 3D networks are very rare.

Vertical (height, levelling) networks are probably the best example of how differential positioning is used together with the knowledge of point heights in carrying the height information from the seashore inland. The heights of selected shore benchmarks are first derived from the observations of the sea level (cf., Section II.B), carried out by means of *tide gauges* (also known in older literature as mareographs) by means of short levelling lines. These basic benchmarks are then linked to the network by longer levelling lines that connect together a whole multitude of land benchmarks (cf., Fig. 6).

Modern satellite networks are inherently three-dimensional. Because the intervisibility is not a requirement for relative satellite positioning, satellite networks can and do contain much longer links and can be much larger in geographical extent. Nowadays, global geodetic networks are constructed and used for different applications.

E. Treatment of Errors in Positions

All positions, determined in whatever way, have errors, both systematic and random. This is due to the fact that every observation is subject to an error; some of these errors are smaller, some are larger. Also, the mathematical models from which the positions are computed are not always completely known or properly described. Thus, when we speak about positions in geodesy, we always mention the accuracy/error that accompanies it. How are these errors expressed?

Random errors are described by following quadratic form:

$$\xi^{\mathrm{T}}\mathbf{C}^{-1}\xi = C_{\alpha}, \tag{10}$$

where **C** is the *covariance matrix* of the position (a three by three matrix composed of variances and covariances of the three coordinates which comes as a by-product of the network adjustment) and C_α is a factor that depends on the probability density function involved in the position estimation and on the desired probability level α. This

FIGURE 6 The Canadian national geodetic levelling network (Source: www.nrcan.gc.ca. Copyright: Her Majesty the Queen in Right of Canada, Natural Resources Canada, Geodetic Survey Division. All rights reserved.)

quadratic form can be interpreted as an equation of an ellipsoid, called a confidence region in statistics or an *error ellipsoid* in geodetic practice. The understanding is that if we know the covariance matrix **C** and select a probability level α we are comfortable with, then the vector difference ξ between the estimated position $\mathbf{r}^*$ and the real position $\mathbf{r}$ is, with probability α, within the confines of the error ellipsoid.

The interpretation of the error ellipsoid is a bit tricky. The error ellipsoid described above is called *absolute*, and one may expect that errors (and thus also accuracy) thus measured refer to the coordinate system in which the positions are determined. They do not! They actually refer to the point (points) given to the network adjustment (cf., Section II.D) for fixing the position of the adjusted point configuration. This point (points) is sometimes called the "datum" for the adjustment, and we can say that the absolute confidence regions are really relative with respect to the "adjustment datum." As such, they have a natural tendency to grow in size with the growing distance of the point of interest from the adjustment datum. This behavior curtails somewhat the usefulness of these measures.

Hence, in some applications, *relative* error ellipsoids (confidence regions) are sought. These measure errors (accuracy) of one position, A, with respect to another position, B, and thus refer always to pairs of points. A relative confidence region is defined by an expression identical to Eq. (10), except that the covariance matrix used, $\mathbf{C}_{\Delta \mathrm{AB}}$, is that of the three coordinate differences $\Delta\mathbf{r}_{\mathrm{AB}}$ rather than the three coordinates $\mathbf{r}$. This covariance matrix is evaluated from the following expression:

$$\mathbf{C}_{\Delta \mathrm{AB}} = \mathbf{C}_{\mathrm{A}} + \mathbf{C}_{\mathrm{B}} - \mathbf{C}_{\mathrm{AB}} - \mathbf{C}_{\mathrm{BA}}, \qquad (11)$$

where $\mathbf{C}_{\mathrm{A}}$ and $\mathbf{C}_{\mathrm{B}}$ are the covariance matrices of the two points, A and B, and $\mathbf{C}_{\mathrm{AB}} = \mathbf{C}_{\mathrm{BA}}^{\mathrm{T}}$ is the cross-covariance matrix of the two points. The cross-covariance matrix comes also as a by-product of the network adjustment. It should be noted that the cross-covariances (cross-correlations) between the two points play a very significant role here: when the cross-correlations are strong, the relative confidence region is small and vice versa.

When we deal with 2D instead of 3D coordinates, the confidence regions (absolute and relative) also become two dimensional. Instead of having error ellipsoids, we have error ellipses—see Fig. 7 that shows both absolute and relative error ellipses as well as errors in the distance S, σ_{S} and azimuth α, $S\sigma_{\alpha}$, computed (estimated) from the positions of A and B. In the 1D case (heighting), confidence regions degenerate to line segments. The Dilution of Precision (DOP) indicators used in GPS (cf., Section V.B)

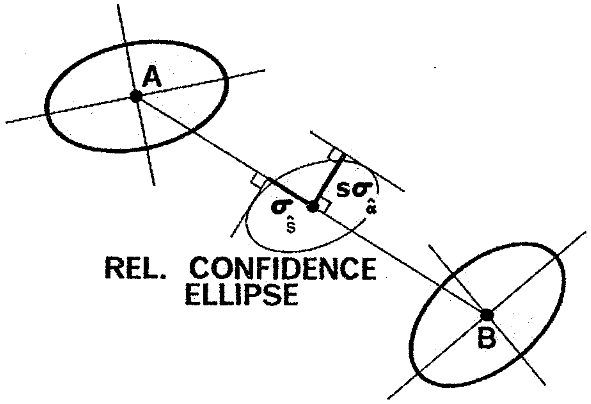

FIGURE 7 Absolute and relative error ellipses.

are related to the idea of (somewhat simplified) confidence regions.

Once we know the desired confidence region(s) in one coordinate system, we can derive the confidence region in any other coordinate system. The transformation works on the covariance matrix used in the defining expression (10) and is given by

$$\mathbf{C}^{(2)} = \mathbf{T}\mathbf{C}^{(1)}\mathbf{T}^{\mathrm{T}}, \tag{12}$$

where $\mathbf{T}$ is the Jacobian of transformation from the first to the second coordinate systems evaluated for the point of interest, i.e., $\mathbf{T} = \mathbf{T}(\mathbf{r})$.

Systematic errors are much more difficult to deal with. Evaluating them requires an intimate knowledge of their sources, and these are not always known. The preferred way of dealing with systematic errors is to prevent them from occurring in the first place. If they do occur, then an attempt is made to eliminate them as much as possible.

There are other important issues that we should discuss here in connection with position errors. These include concepts of blunder elimination, reliability, geometrical strength of point configurations, and more. Unfortunately, there is no room to get into these concepts here, and the interested reader may wish to consult Vaníček and Krakiwsky (1986) or some other geodetic textbook.

F. Coordinate Transformations

A distinction should be made between (abstract) "coordinate system transformations" and "coordinate transformations": coordinate systems do not have any errors associated with them while coordinates do. The transformation between two Cartesian coordinate systems [first (1) and second (2)] can be written in terms of hypothetical positions $\mathbf{r}^{(1)}$ and $\mathbf{r}^{(2)}$ as

$$\mathbf{r}^{(2)} = \mathbf{R}(\varepsilon_x, \varepsilon_y, \varepsilon_z,)\mathbf{r}^{(1)} + \mathbf{t}_0^{(2)}, \tag{13}$$

where $\mathbf{R}(\varepsilon_x, \varepsilon_y, \varepsilon_z,)$ is the "rotation matrix," which after application to a vector rotates the vector by the three *misalignment angles* $\varepsilon_x, \varepsilon_y, \varepsilon_z$, around the coordinate axes, and $\mathbf{t}_0^{(2)}$ is the position vector of the origin of the first system reckoned in the second system, called the *translation vector*.

The transformation between coordinates must take into account the errors in both coordinates/coordinate sets (in the first and second coordinate system), particularly the systematic errors. A transformation of coordinates thus consists of two distinct components: the transformation between the corresponding coordinate systems as described above, plus a model for the difference between the errors in the two coordinate sets. The standard illustration of such a model is the inclusion of the scale factor, which accounts for the difference in linear scales of the two coordinate sets. In practice, when dealing with coordinate sets from more extensive areas such as states or countries, these models are much more elaborate, as they have to model the differences in the deformations caused by errors in the two configurations. These models differ from country to country. For unknown reasons, some people prefer not to distinguish between the two kinds of transformations.

Figure 8 shows a commutative diagram for transformations between most of the coordinate systems used in geodesy. The quantities in rectangles are the transformation parameters, the misalignment angles, and translation components. For a full understanding of the involved transformations, the reader is advised to consult Vaníček and Krakiwsky (1986, Chap. 15).

Let us just mention that sometimes we are not interested in transforming positions (position vectors, triplets of coordinates), but small (differential) changes $\delta\mathbf{r}$ in positions $\mathbf{r}$ as we saw in Eq. (12). In this case, we are not concerned with translations between the coordinate systems, only misalignments are of interest. Instead of using the rotation matrix, we may use the Jacobian of transformation, getting

$$\delta\mathbf{r}^{(2)}(\mathbf{r}) = \mathbf{T}(\mathbf{r})\,\delta\mathbf{r}^{(1)}(\mathbf{r}). \tag{14}$$

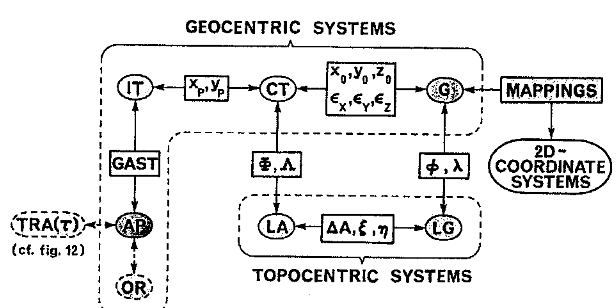

FIGURE 8 Commutative diagram of transformations between coordinate systems.

The final topic we want to discuss in this section is one that we are often faced with in practice: given two corresponding sets of positions (of the same points) in two different coordinate systems we wish to determine the transformation parameters for the two systems. This is done by means of Eq. (13), where $\varepsilon_x, \varepsilon_y, \varepsilon_z, \mathbf{t}_1^{(2)}$ become the six unknown transformation parameters, while $\mathbf{r}^{(1)}, \mathbf{r}^{(2)}$ are the known quantities. The set of known positions has to consist of at least three noncollinear points so we get at least six equations for determining the six unknowns. We must stress that the set of position vectors $\mathbf{r}^{(1)}, \mathbf{r}^{(2)}$ has to be first corrected for the distortions caused by the errors in both sets of coordinates. These distortions are added back on if we become interested in coordinate transformation.

G. Kinematic Positioning and Navigation

As we have seen so far, classical geodetic positioning deals with stationary points (objects). In recent times, however, geodetic positioning has found its role also in positioning moving objects such as ships, aircraft, and cars. This application became known as *kinematic positioning*, and it is understood as being the real-time positioning part of navigation. Formally, the task of kinematic positioning can be expressed as

$$\mathbf{r}(t) = \mathbf{r}(t_0) + \int_{t_0}^{t} \mathbf{v}(t)\, dt, \tag{15}$$

where t stands for time and $\mathbf{v}(t)$ is the observed change in position in time, i.e., velocity (vector) of the moving object. The velocity vector can be measured on the moving vehicle in relation to the surrounding space or in relation to an inertial coordinate system by an inertial positioning system. We note that, in many applications, the attitude (roll and pitch) of the vehicle is also of interest.

Alternatively, optical astronomy or point satellite positioning produce directly the string of positions, $\mathbf{r}(t_1)$, $\mathbf{r}(t_2), \ldots, \mathbf{r}(t_n)$, that describe the required *trajectory* of the vehicle, without the necessity of integrating over velocities. Similarly, a relative positioning technique, such as the hyperbolic radio system Loran-C (or Hi-Fix, Decca, Omega in the past), produces a string of position changes, $\Delta\mathbf{r}(t_0, t_1)$, $\Delta\mathbf{r}(t_1, t_2), \ldots, \Delta\mathbf{r}(t_{n-1}, t_n)$, which once again define the trajectory. We note that these techniques are called hyperbolic because positions or position differences are determined from intersections of hyperbolae, which, in turn, are the loci of constant distance differences from the land-located radiotransmitters. Relative satellite positioning is also being used for kinematic positioning, as we shall see later in Section V.B.

For a navigator, it is not enough to know his position $\mathbf{r}(t_n)$ at the time t_n. The navigator must also have the position estimates for the future, i.e., for the times $t_n, t_{n+1}, \ldots,$ to be able to navigate safely, he has to have the *predicted positions*. Thus, the kinematic positioning described above has to be combined with a *navigation algorithm*, a predictive filter which predicts positions in the future based on the observed position in the past, at times $t_1, t_2, \ldots, t_n$. Particularly popular seem to be different kinds of Kalman's filters, which contain a feature allowing one to describe the dynamic characteristics of the vehicle navigating in a specified environment. Kalman's filters do have a problem with environments that behave in an unpredictable way, such as an agitated sea. We note that some of the navigation algorithms accept input from two or more kinematic position systems and combine the information in an optimal way.

In some applications, it is desirable to have a post-mission record of trajectories available for future retracing of these trajectories. These post-mission trajectories can be made more accurate than the real-time trajectories (which, in turn, are of course more accurate than the predicted trajectories). Most navigation algorithms have the facility of *post-mission smoothing* the real-time trajectories by using all the data collected during the mission.

III. EARTH'S GRAVITY FIELD

A. Origin of the Earth's Gravity Field

In geodesy, we are interested in studying the gravity field in the macroscopic sense where the quantum behavior of gravity does not have to be taken into account. Also, in terrestrial gravity work, we deal with velocities that are very much smaller than the speed of light. Thus, we can safely use Newtonian physics and may begin by recalling mass attraction force $\mathbf{f}$ defined by Newton's integral

$$\mathbf{f}(\mathbf{r}) = \mathbf{a}(\mathbf{r})m = \left(G \int_{\boldsymbol{B}} \rho(\mathbf{r}')(\mathbf{r}' - \mathbf{r})|\mathbf{r}' - \mathbf{r}|^{-3}\, dV \right) m, \tag{16}$$

where $\mathbf{r}$ and $\mathbf{r}'$ are position vectors of the point of interest and the dummy point of the integration; $\boldsymbol{B}$ is the attracting massive body of density ρ, i.e., the earth; V stands for volume; G is Newton's gravitational constant; m is the mass of the particle located at $\mathbf{r}$; and $\mathbf{a}(\mathbf{r})$ is the acceleration associated with the particle located at $\mathbf{r}$ (see Fig. 9). We can speak about the acceleration $\mathbf{a}(\mathbf{r})$, called *gravitation*, even when there is no mass particle present at $\mathbf{r}$, but we cannot measure it (only an acceleration of a mass can be measured). This is the idea behind the definition of the *gravitational field* of body $\boldsymbol{B}$, the earth; this field is defined at all points $\mathbf{r}$. The physical units of gravitation are those of an acceleration, i.e., m s^{-2}; in practice, units of cm s^{-2},

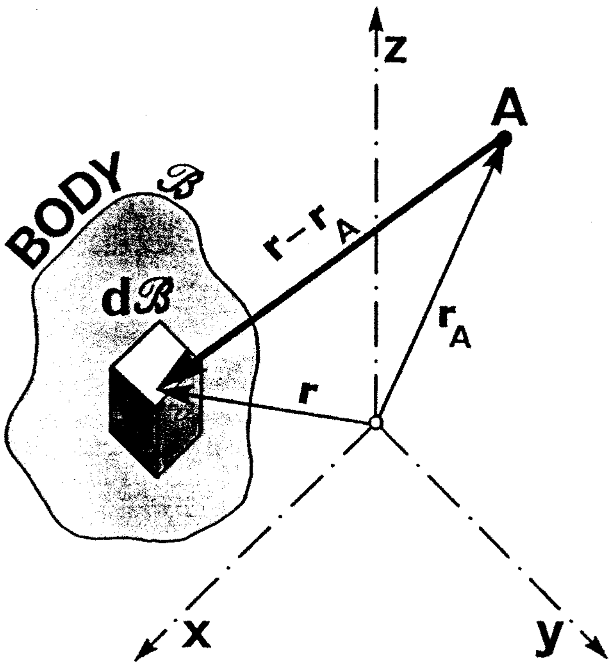

FIGURE 9 Mass attraction.

called "Gal" [to commemorate Galileo's (c.f., Section I.A) contribution to geodesy], are often used.

Newton's gravitational constant G represents the ratio between mass acting in the "attracted capacity" and the same mass acting in the "attracting capacity." From Eq. (16) we can deduce the physical units of G, which are 10 $\text{kg}^{-1}\ \text{m}^3\ \text{s}^{-2}$. The value of G has to be determined experimentally. The most accurate measurements are obtained from tracking deep space probes that move in the gravitational field of the earth. If a deep space probe is sufficiently far from the earth (and the attractions of the other celestial bodies are eliminated mathematically), then the physical dimensions of the probe become negligible. At the same time, the earth can be regarded with sufficient accuracy as a sphere with a laterally homogeneous density distribution. Under these circumstances, the gravitational field of the earth becomes radial, i.e., it will look as if it were generated by a particle of mass M equal to the total mass of the earth:

$$M = \int_B \rho(\mathbf{r}')\, dV. \tag{17}$$

When a "geocentric" coordinate system is used in the computations, the probe's acceleration becomes

$$\mathbf{a}(\mathbf{r}) = -GM\mathbf{r}|\mathbf{r}|^{-3}. \tag{18}$$

Thus, the gravitational constant G, or more accurately GM, called the *geocentric constant*, can be obtained from purely geometrical measurements of the deep space probe positions $\mathbf{r}(t)$. These positions, in turn, are determined from measurements of the propagation of electromagnetic waves and as such depend very intimately on the accepted value of the speed of light c. The value of GM is now thought to be $(3{,}986{,}004.418 \pm 0.008) * 10^8\ \text{m}^3\ \text{s}^{-2}$ (Ries *et al.*, 1992), which must be regarded as directly dependent on the accepted value of c. Dividing the geocentric constant by the mass of the earth $[(5.974 \pm 0.001) * 10^{24}\ \text{kg}]$, one obtains the value for G as $(6.672 \pm 0.001) * 10^{-11}\ \text{kg}^{-1}\ \text{m}^3\ \text{s}^{-2}$.

The earth spins around its instantaneous spin axis at a more or less constant angular velocity of once per "sidereal day"—sidereal time scale is taken with respect to fixed stars, which is different from the solar time scale, taken with respect to the sun. This spin gives rise to a centrifugal force that acts on each and every particle within or bound with the earth. A particle, or a body, which is not bound with the earth, such as an earth satellite, is not subject to the centrifugal force. This force is given by the following equation:

$$\mathbf{F}(\mathbf{r}) = \omega^2\, \mathbf{p}(\mathbf{r})m, \tag{19}$$

where $\mathbf{p}(\mathbf{r})$ is the projection of $\mathbf{r}$ onto the equatorial plane, ω is the siderial angular velocity of 1 revolution per day $(7.292115 * 10^{-5}\ \text{rad s}^{-1})$, and m is the mass of the particle subjected to the force. Note that the particles on the spin axis of the earth experience no centrifugal force as the projection $\mathbf{p}(\mathbf{r})$ of their radius vector equals to $\mathbf{0}$. We introduce the *centrifugal acceleration* $\mathbf{a}_c(\mathbf{r})$ at point $\mathbf{r}$ as $\omega^2\, \mathbf{p}(\mathbf{r})$ and speak of the centrifugal acceleration field much in the same way we speak of the gravitational field (acceleration) $\mathbf{a}(\mathbf{r})$.

The earth (B) gravitation is denoted by $\mathbf{g}_g$ (subscripted g for gravitation) rather than $\mathbf{a}$, and its centrifugal acceleration is denoted by $\mathbf{g}_c$ (c for centrifugal) rather than $\mathbf{a}_c$. When studying the fields $\mathbf{g}_g$ and $\mathbf{g}_c$ acting at points bound with the earth (spinning with the earth), we normally lump these two fields together and speak of the earth's *gravity field* $\mathbf{g}$:

$$\mathbf{g}(\mathbf{r}) = \mathbf{g}_g(\mathbf{r}) + \mathbf{g}_c(\mathbf{r}). \tag{20}$$

A stationary test mass m located at any of these points will sense the total gravity vector $\mathbf{g}$ (acceleration).

If the (test) mass moves with respect to the earth, then another (virtual) force affects the mass: the Coriolis force, responsible, for instance, for the geostrophic motion encountered in air or water movement. In the studies of the earth's gravity field, Coriolis' force is not considered. Similarly, temporal variation of the gravity field, due to variations in density distribution and in earth's rotation speed, which are small compared to the magnitude of the field itself, is mostly not considered either. It is studied separately within the field of geo-kinematics (Section IV).

B. Gravity Potential

When we move a mass m in the gravity field $\mathbf{g}(\mathbf{r})$ from location $\mathbf{r}_1$ to location $\mathbf{r}_2$, to overcome the force $\mathbf{g}(\mathbf{r})\ m$ of the field, we have to do some work w. This work is expressed by the following line integral:

$$w = -\int_{r_1}^{r_2} \mathbf{g}(\mathbf{r})' m\, d\mathbf{r}'. \tag{21}$$

Note that the physical units of work w are kg m^2 s^{-2}. Fortunately, for the gravitational field the amount of work does not depend on what trajectory is followed when moving the particle from $\mathbf{r}_1$ to $\mathbf{r}_2$. This property can be expressed as

$$\oint_C \mathbf{g}(\mathbf{r}')\, d\mathbf{r}' m = \oint_C \mathbf{g}(\mathbf{r}')\, d\mathbf{r}' = 0, \tag{22}$$

where the line integral is now taken along an arbitrary closed curve C. The physical meaning of Eq. (22) is when we move a particle in the gravitational field along an arbitrary closed trajectory, we do not expend any work.

This property must be true also when the closed trajectory (curve) C is infinitesimally short. This means that the gravitational field must be an irrotational vector field: its *vorticity* is equal to 0 everywhere:

$$\nabla \times \mathbf{g}(\mathbf{r}) = \mathbf{0}. \tag{23}$$

A field which behaves in this way is also known as a *potential field*, meaning that there exists a scalar function, called *potential*, of which the vector field in question is a *gradient*. Denoting this potential by $W(\mathbf{r})$, we can thus write

$$\nabla W(\mathbf{r}) = \mathbf{g}(\mathbf{r}). \tag{24}$$

To get some insight into the physical meaning of the potential W, whose physical units are m^2 s^{-2}, we relate it to the work w defined in Eq. (21). It can be shown that the amount of work expended when moving a mass m from $\mathbf{r}_1$ to $\mathbf{r}_2$, along an arbitrary trajectory, is equal to

$$w = [W(\mathbf{r}_2) - W(\mathbf{r}_1)]m. \tag{25}$$

In addition to the two differential equations (Eqs. 23 and 24) governing the behavior of the gravity field, there is a third equation describing the field's *divergence*,

$$\nabla \cdot \mathbf{g}(\mathbf{r}) = -4\pi G\rho(\mathbf{r}) + 2\omega^2. \tag{26}$$

These three *field equations* describe fully the differential behavior of the earth's gravity field. We note that the first term on the right-hand side of Eq. (26) corresponds to the gravitational potential W_g, whose gradient is the gravitational vector $\boldsymbol{g}_g$, while the second term corresponds to the centrifugal potential W_c that gives rise to the centrifugal acceleration vector $\boldsymbol{g}_c$. The negative sign of the first term indicates that, at the point $\mathbf{r}$, there is a sink rather than a source of the gravity field, which should be somewhat obvious from the direction of the vectors of the field. Since the ∇ operator is linear, we can write

$$W(\mathbf{r}) = W_g(\mathbf{r}) + W_c(\mathbf{r}). \tag{27}$$

A potential field is a scalar field that is simple to describe and to work with, and it has become the basic descriptor of the earth's gravity field in geodesy (cf., the article "Global Gravity" in this volume). Once one has an adequate knowledge of the gravity potential, one can derive all the other characteristics of the earth's gravity field, $\mathbf{g}$ by Eq. (24), W_g by Eq. (27), etc., mathematically. Interestingly, the Newton integral in Eq. (16) can be also rewritten for the gravitational potential W_g, rather than the acceleration, to give

$$W_g(\mathbf{r}) = G \int_B \rho(\mathbf{r}')|\mathbf{r}' - \mathbf{r}|^{-1}\, dV, \tag{28}$$

which is one of the most often used equations in gravity field studies. Let us, for completeness, spell out also the equation for the centrifugal potential:

$$W_c(\mathbf{r}) = \tfrac{1}{2}\,\omega^2 p^2(\mathbf{r}), \tag{29}$$

[cf., Eq. (19)].

A surface on which the gravity potential value is constant is called an *equipotential surface*. As the value of the potential varies continuously, we may recognize infinitely many equipotential surfaces defined by the following prescription:

$$W(\mathbf{r}) = \text{const.} \tag{30}$$

These equipotential surfaces are convex everywhere above the earth and never cross each other anywhere. By definition, the equipotential surfaces are horizontal everywhere and are thus sometimes called the level surfaces.

One of these infinitely many equipotential surfaces is the *geoid*, one of the most important surfaces used in geodesy, the equipotential surface defined by a specific value W_0 and thought of as approximating the MSL the best (cf., Section II.B) in some sense. We shall have more to say about the two requirements in Section IV.D. At the time of writing, the best value of W_0 is thought to be 62,636,855.8 $\pm$ 0.5 m^2 s^{-2} (Burša *et al.*, 1997). A global picture of the geo id is shown in Fig. 5 in the article "Global Gravity" in this volume, where the geoidal height N (cf., Section II.B), i.e., the geoid-ellipsoid separation, is plotted. Note that the departure of the geoid from the mean earth ellipsoid (for the definition see below) is at most about 100 m in absolute value.

When studying the earth's gravity field, we often need and use an idealized model of the field. The use of such a model allows us to express the actual gravity field as a sum of the selected model field and the remainder of the actual field. When the model field is selected to resemble closely the actual field itself, the remainder, called an anomaly or disturbance, is much smaller than the actual field itself. This is very advantageous because working with significantly smaller values requires less rigorous mathematical treatment to arrive at the same accuracy of the final results. This procedure resembles the "linearization" procedure used in mathematics and is often referred to as such. Two such models are used in geodesy: spherical (radial field) and ellipsoidal (also called normal or Somigliana-Pizzetti's) models. The former model is used mainly in satellite orbit analysis and prediction (cf., Section V.C), while the normal model is used in terrestrial investigations.

The normal gravity field is generated by a massive body called the *mean earth ellipsoid* adopted by a convention. The most recent such convention, proposed by the IAG in 1980 (IAG, 1980) and called Geodetic Reference System of 1980 (GRS 80), specifies the mean earth ellipsoid as having the major semi-axis "a" 6,378,137 m long and the flattening "f" of 1/298.25. A flattening of an ellipsoid is defined as

$$f = (a - b)/a, \tag{31}$$

where "b" is the minor semi-axis. This ellipsoid departs from a mean earth sphere by slightly more than 10 km; the difference of $a - b$ is about 22 km. We must note here that the flattening f is closely related to the second degree coefficient $C_{2,0}$ discussed in the article "Global Gravity."

This massive ellipsoid is defined as rotating with the earth with the same angular velocity ω, its potential is defined to be constant and equal to W_0 on the surface of the ellipsoid, and its mass is the same as that (M) of the earth. Interestingly, these prescriptions are enough to evaluate the *normal potential* $U(\mathbf{r})$ everywhere above the ellipsoid so that the mass density distribution within the ellipsoid does not have to be specified. The departure of the actual gravity potential from the normal model is called *disturbing potential* $T(\mathbf{r})$:

$$T(\mathbf{r}) = W(\mathbf{r}) - U(\mathbf{r}). \tag{32}$$

The earth's gravity potential field is described in a global form as a truncated series of spherical harmonics up to order and degree 360 or even higher. Many such series have been prepared by different institutions in the United States and in Europe. Neither regional nor local representations of the potential are used in practice; only the geoid, the gravity anomalies, and the deflections of the vertical (see the next three sections) are needed on a regional/local basis.

C. Magnitude of Gravity

The gravity vector $\mathbf{g}(\mathbf{r})$ introduced in Section III.A can be regarded as consisting of a magnitude (length) and a direction. The magnitude of $\mathbf{g}$, denoted by g, is referred to as *gravity*, which is a scalar quantity measured in units of acceleration. It changes from place to place on the surface of the earth in response to latitude, height, and the underground mass density variations. The largest is the latitude variation, due to the oblateness of the earth and due to the change in centrifugal acceleration with latitude, with amounts to about 5.3 cm s^{-2}, i.e., about 0.5% of the total value of gravity. At the poles, gravity is the strongest, about 9.833 m s^{-2} (983.3 Gal); at the equator it is at its weakest, about 978.0 Gal. The height variation, due to varying distance from the attracting body, the earth, amounts to 0.3086 mGal m^{-1} when we are above the earth and to around 0.0848 mGal m^{-1} (we have seen this gradient already in Section II.B) when we are in the uppermost layer of the earth such as the topography. The variations due to mass density variations are somewhat smaller. We note that all these variations in gravity are responsible for the variation in weight: for instance, a mass of 1 kg at the pole weighs 9.833 kg m^1 s^{-2}, while on the equator it weighs only 9.780 kg m^1 s^{-2}.

Gravity can be measured by means of a test mass, by simply measuring either the acceleration of the test mass in free fall or the force needed to keep it in place. Instruments that use the first approach, pendulums and "free-fall devices," can measure the total value of gravity (absolute instruments), while the instruments based on the second approach, called "gravimeters," are used to measure gravity changes from place to place or changes in time (relative instruments). The most popular field instruments are the gravimeters (of many different designs) which can be made easily portable and easily operated. *Gravimetric surveys* conducted for the geophysical exploration purpose, which is the main user of detailed gravity data, employ portable gravimeters exclusively. The accuracy, in terms of standard deviations, of most of the data obtained in the field is of the order of 0.05 mGal.

To facilitate the use of gravimeters as relative instruments, countries have developed *gravimetric networks* consisting of points at which gravity had been determined through a national effort. The idea of gravimetric networks is parallel to the geodetic (positioning) networks we have seen in Section II.D, and the network adjustment process is much the same as the one used for the geodetic networks. A national gravimetric network starts with national gravity reference point(s) established in participating countries through an international effort; the last such effort, organized by IAG (cf., Section I.C), was the International Gravity Standardization Net 1971 (IGSN 71) (IAG, 1974).

Gravity data as observed on the earth's surface are of little direct use in exploration geophysics. To become useful, they have to be stripped of

1. The height effect, by reducing the observed gravity to the geoid, using an appropriate vertical gradient of gravity $\partial g/\partial H$
2. The dominating latitudinal effect, by subtracting from them the corresponding magnitude of normal gravity γ [the magnitude of the gradient of U, cf., Eq. (24)] reckoned on the mean earth ellipsoid (see Section III.B), i.e., at points (r_e, φ, λ)

The resulting values Δg are called *gravity anomalies* and are thought of as corresponding to locations (r_g, φ, λ) on the geoid. For geophysical interpretation, gravity anomalies are thus defined as (for the definition used in geodesy, see Section III.E)

$$\Delta g(r_g, \varphi, \lambda) = g(r_t, \varphi, \lambda) - \partial g/\partial H\, H(\varphi, \lambda) - \gamma(r_e, \varphi, \lambda), \tag{33}$$

where $g(r_t, \varphi, \lambda)$ are the gravity values observed at points (r_t, φ, λ) on the earth's surface and $H(\varphi, \lambda)$ are the orthometric heights of the observed gravity points. These orthometric heights are usually determined together with the observed gravity. Normal gravity on the mean earth ellipsoid is part of the normal model accepted by convention as discussed in the previous section. The GRS 80 specifies the normal gravity on the mean earth ellipsoid by the following formula:

$$\gamma(r_e, \varphi) = 978.0327(1 + 0.0052790414 \sin^2 \varphi + 0.0000232718 \sin^4 \varphi + 0.0000001262 \sin^6 \varphi)\text{Gal}. \tag{34}$$

Gravity anomalies, like the disturbing potential in Section III.B, are thought of as showing only the anomalous part of gravity, i.e., the spatial variations in gravity caused by subsurface mass density variations. Depending on what value is used for the vertical gradient of gravity $\partial g/\partial H$, we get different kinds of gravity anomalies: using $\partial g/\partial H = -0.3086$ mGal m^{-1}, we get the *free-air gravity anomaly*; using $\partial g/\partial H = \frac{1}{2}(-0.3086 - 0.0848)$ mGal m^{-1}, we get the (simple) *Bouguer gravity anomaly*. Other kinds of anomalies exist, but they are not very popular in other than specific theoretical undertakings.

Observed gravity data, different kinds of point gravity anomalies, anomalies averaged over certain geographical cells, and other gravity-related data are nowadays available in either a digital form or in the form of maps. These can be obtained from various national and international agencies upon request. Figure 10 shows the map of free-air gravity anomalies in Canada.

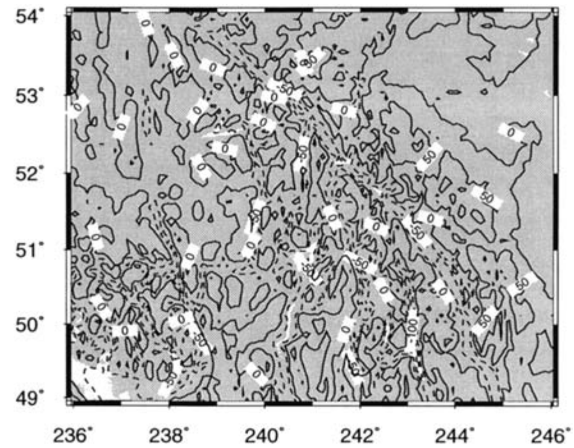

FIGURE 10 Map of free-air gravity anomalies in Canadian Rocky Mountains.

D. Direction of Gravity

Like the magnitude of the gravity vector $\mathbf{g}$ discussed in the previous section, its direction is also of interest. As it requires two angles to specify the direction, the direction of gravity is a little more difficult to deal with than the magnitude. As has been the case with gravity anomalies, it is convenient to use the normal gravity model here as well. When subtracting the direction of normal gravity from the direction of actual gravity, we end up with a small angle, probably smaller than 1 or 2 arcmin anywhere on earth. This smaller angle θ, called the *deflection of the vertical*, is easier to work with than the arbitrarily large angles used for describing the direction of $\mathbf{g}$. We thus have

$$\theta(\mathbf{r}) = \sphericalangle[\mathbf{g}(\mathbf{r}), \boldsymbol{\gamma}(\mathbf{r})], \tag{35}$$

where, in parallel with Eq. (24), $\boldsymbol{\gamma}(\mathbf{r})$ is evaluated as the gradient of the normal potential U:

$$\boldsymbol{\gamma}(\mathbf{r}) = \nabla U(\mathbf{r}). \tag{36}$$

We may again think of the deflection of the vertical as being only just an effect of a disturbance of the actual field, compared to the normal field.

Gravity vectors, being gradients of their respective potentials, are always perpendicular to the level surfaces, be they actual gravity vectors or normal gravity vectors. Thus, the direction of $\mathbf{g}(\mathbf{r})$ is the real vertical (a line perpendicular to the horizontal surface) at $\mathbf{r}$ and the direction of $\boldsymbol{\gamma}(\mathbf{r})$ is the normal vertical at $\mathbf{r}$: the deflection of the vertical is really the angle between the actual and normal vertical directions. We note that the actual vertical direction is always tangential to the actual *plumbline*, known in physics also as the line of force of the earth's gravity field. At the geoid, for $\mathbf{r} = \mathbf{r}_g$, the direction of $\boldsymbol{\gamma}(\mathbf{r}_g)$ is to a high degree of accuracy the same as the direction of the normal to the mean earth ellipsoid (being exactly the same on the mean earth ellipsoid).

If the mean earth ellipsoid is chosen also as a reference ellipsoid, then the angles that describe the direction of the normal to the ellipsoid are the geodetic latitude φ and longitude $\boldsymbol{\lambda}$ (cf., Section II.A). Similarly, the direction of the plumbline at any point $\mathbf{r}$ is defined by *astronomical latitude* $\boldsymbol{\Phi}$ and *astronomical longitude* $\boldsymbol{\Lambda}$. The astronomical coordinates $\boldsymbol{\Phi}$ and $\boldsymbol{\Lambda}$ can be obtained, to a limited accuracy, from optical astronomical measurements, while the geodetic coordinates are obtained by any of the positioning techniques described in Section II. Because θ is a spatial angle, it is customary in geodesy to describe it by two components, the meridian ξ and the prime vertical η components. The former is the projection of θ onto the local meridian plane, and the latter is the projection onto the local prime vertical plane (plane perpendicular to the horizontal and meridian planes).

There are two kinds of deflection of vertical used in geodesy: those taken at the surface of the earth, at points $\mathbf{r}_\mathrm{t} = (r_t, \varphi, \lambda)$, called *surface deflections* and those taken at the geoid level, at points $\mathbf{r}_\mathrm{g} = (r_g, \varphi, \lambda)$, called *geoid deflections*. Surface deflections are generally significantly larger than the geoid deflections, as they are affected not only by the internal distribution of masses but also by the topographical masses. The two kinds of deflections can be transformed to each other. To do so, we have to evaluate the curvature of the plumbline (in both perpendicular directions) and the curvature of the normal vertical. The former can be quite sizeable—up to a few tens of arcseconds—and is very difficult to evaluate. The latter is curved only in the meridian direction (the normal field being rotationally symmetrical), and even that curvature is rather small, reaching a maximum of about 1 arcsec.

The classical way of obtaining the deflections of the vertical is through the differencing of the astronomical and geodetic coordinates as follows:

$$\xi = \boldsymbol{\Phi} - \varphi, \eta = (\boldsymbol{\Lambda} - \lambda)\cos\varphi. \tag{37}$$

These equations also define the signs of the deflection components. In North America, however, the sign of η is sometimes reversed. We emphasize here that the geodetic coordinates have to refer to the geocentric reference ellipsoid/mean earth ellipsoid. Both geodetic and astronomical coordinates must refer to the same point, either on the geoid or on the surface of the earth. In Section II.B, we mentioned that the astronomical determination of point positions $(\boldsymbol{\Phi}, \boldsymbol{\Lambda})$ is not used in practice any more because of the large effect of the earth's gravity field. Here, we see the reason spelled out in Eqs. (37): considering the astronomically determined position $(\boldsymbol{\Phi}, \boldsymbol{\Lambda})$ to be an approximation of the geodetic position (φ, λ) invokes an error of $(\xi, \eta/\cos\varphi)$ that can reach several kilometers on the surface of the earth. The deflections of the vertical can be determined also from other measurements, which we will show in the next section.

E. Transformations between Field Parameters

Let us begin with the transformation of the geoidal height to the deflection of the vertical [i.e., $N \to (\xi, \eta)$], which is of a purely geometrical nature and fairly simple. When the deflections are of the "geoid" kind, they can be interpreted simply as showing the slope of the geoid with respect to the geocentric reference ellipsoid at the deflection point. This being the case, geoidal height differences can be constructed from the deflections $((\xi, \eta) \to \Delta N)$ in the following fashion. We take two adjacent deflection points and project their deflections onto a vertical plane going through the two points. These projected deflections represent the projected slopes of the geoid in the vertical plane; their average multiplied by the distance between the two points gives us an estimate of the difference in geoidal heights ΔN at the two points. Pairs of deflection points can be then strung together to produce the geoid profiles along selected strings of deflection points. This technique is known as *Helmert's levelling*. We note that if the deflections refer to a geodetic datum (rather than to a geocentric reference ellipsoid), this technique gives us geoidal height differences referred to the same geodetic datum. Some older geoid models were produced using this technique.

Another very useful relation (transformation) relates the geoid height N to the disturbing potential $T (T \to N, N \to T)$. It was first formulated by a German physicist H. Bruns (1878), and it reads

$$N = T/\gamma. \tag{38}$$

The equation is accurate to a few millimeters; it is now referred to as *Bruns's formula*.

In Section III.C we introduced gravity anomaly Δg of different kinds (defined on the geoid), as they are normally used in geophysics. In geodesy we need a different gravity anomaly, one that is defined for any location $\mathbf{r}$ rather than being tied to the geoid. Such gravity anomaly is defined by the following exact equation:

$$\Delta g(\mathbf{r}) = -\partial T/\partial h|_{\mathbf{r}=(\mathrm{r},\varphi,\lambda)} + \gamma(\mathbf{r})^{-1}\partial\gamma/\partial h|_{\mathbf{r}=(\mathrm{r},\varphi,\lambda)} T(r - Z, \varphi, \lambda), \tag{39}$$

where Z is the displacement between the actual equipotential surface $W = \text{const.}$ passing through $\mathbf{r}$ and the corresponding (i.e., having the same potential) normal equipotential surface $U = \text{const.}$ This differential equation of first order, sometimes called *fundamental gravimetric equation*, can be regarded as the transformation from $T(\mathbf{r})$ to $\Delta g(\mathbf{r})(T \rightarrow \Delta g)$ and is used as such in the studies of the earth's gravity field. The relation between this gravity anomaly and the ones discussed above is somewhat tenuous.

Perhaps the most important transformation is that of gravity anomaly Δg, it being the cheapest data, to disturbing potential $T(\Delta g \rightarrow T)$, from which other quantities can be derived using the transformations above. This transformation turns out to be rather complicated: it comes as a solution of a scalar boundary value problem and it will be discussed in the following two sections. We devote two sections to this problem because it is regarded as central to the studies of earth's gravity field. Other transformations between different parameters and quantities related to the gravity field exist, and the interested reader is advised to consult any textbook on geodesy; the classical textbook by Heiskanen and Moritz (1967) is particularly useful.

F. Stokes's Geodetic Boundary Value Problem

The scalar *geodetic boundary value problem* was formulated first by Stokes (1849). The formulation is based on the partial differential equation valid for the gravity potential W [derived by substituting Eq. (24) into Eq. (26)],

$$\nabla^2 W(\mathbf{r}) = -4\pi G\rho(\mathbf{r}) + 2\omega^2. \tag{40}$$

This is a nonhomogeneous elliptical equation of second order, known under the name of *Poisson equation*, that embodies all the field equations (see Section III.A) of the earth gravity field. Stokes applied this to the disturbing potential T (see Section III.B) outside the earth to get

$$\nabla^2 T(\mathbf{r}) = 0. \tag{41}$$

This is so because T does not have the centrifugal component and the mass density $\rho(\mathbf{r})$ is equal to 0 outside the earth. (This would be true only if the earth's atmosphere did not exist; as it does exist, it has to be dealt with. This is done in a corrective fashion, as we shall see below.) This homogeneous form of Poisson equation is known as *Laplace equation*. A function (T, in our case) that satisfies the Laplace equation in a region (outside the earth, in our case) is known as being *harmonic* in that region.

Further, Stokes has chosen the geoid to be the boundary for his boundary value problem because it is a smooth enough surface for the solution to exist (in the space outside the geoid). This of course violates the requirement of harmonicity of T by the presence of topography (and the atmosphere). Helmert (1880) suggested to avoid this problem by transforming the formulation into a space where T is harmonic outside the geoid. The actual disturbing potential T is transformed to a disturbing potential T^{h}, harmonic outside the geoid, by subtracting from it the potential caused by topography (and the atmosphere) and adding to it the potential caused by topography (and the atmosphere) condensed on the geoid (or some other surface below the geoid). Then the Laplace equation

$$\nabla^2 T^{\mathrm{h}}(\mathbf{r}) = 0 \tag{42}$$

is satisfied everywhere outside the geoid. This became known as the *Stokes-Helmert formulation*.

The boundary values on the geoid are constructed from gravity observed on the earth's surface in a series of steps. First, gravity anomalies on the surface are evaluated from Eq. (33) using the free-air gradient These are transformed to Helmert's anomalies Δg^{h}, defined by Eq. (39) for $T = T^{\mathrm{h}}$, by applying a transformation parallel to the one for the disturbing potentials as described above. By adding some fairly small corrections, Helmert's anomalies are transformed to the following expression (Vaníček *et al.*, 1999):

$$2r^{-1}T^{\mathrm{h}}(r - Z, \varphi, \lambda) + \partial T^{\mathrm{h}}/\partial \mathrm{r}|_{\mathrm{r}} = -\Delta g^{\mathrm{h}*}(\mathbf{r}). \tag{43}$$

As T^{h} is harmonic above the geoid, this linear combination, multiplied by r, is also harmonic above the geoid. As such it can be "continued downward" to the geoid by using the standard Poisson integral.

Given the Laplace equation (42), the boundary values on the geoid, and the fact that $T^{\mathrm{h}}(\mathbf{r})$ disappears as $r \rightarrow \infty$, Stokes (1849) derived the following integral solution to his boundary value problem:

$$T^{\mathrm{h}}(\mathbf{r}_{\mathrm{g}}) = T^{\mathrm{h}}(r_g, \Omega)R/(4\pi) \int_G \Delta g^{\mathrm{h}*}(r_g, \Omega')\, S(\Psi)\, d\Omega', \tag{44}$$

where Ω, Ω' are the geocentric spatial angles of positions $\mathbf{r}$, $\mathbf{r}'$, Ψ is the spatial angle between $\mathbf{r}$ and $\mathbf{r}'$; S is the Stokes integration kernel in its spherical (approximate) form

$$\begin{aligned} S(\Psi) \cong\ & 1 + \sin^{-1}(\Psi/2) - 6\sin(\Psi/2) - 5\cos\Psi \\ & - 3\cos\Psi \ln[\sin(\Psi/2) + \sin^2(\Psi/2)]; \end{aligned} \tag{45}$$

and the integration is carried out over the geoid. We note that, if desired, the disturbing potential is easily transformed to geoidal height by means of the Bruns formula (38).

In the final step, the solution $T^{\mathrm{h}}(\mathbf{r}_{\mathrm{g}})$ is transformed to $T(\mathbf{r}_{\mathrm{g}})$ by adding to it the potential of topography (and the atmosphere) and subtracting the potential of topography (and the atmosphere) condensed to the geoid. This can be regarded as a back transformation from the "Helmert harmonic space" back to the real space. We have to mention that the fore and back transformation between the two spaces requires knowledge of topography (and the atmosphere), both of height and of density. The latter represents the most serious accuracy limitation of Stokes's solution: the uncertainty in topographical density may cause an error up to 1 to 2 dm in the geoid in high mountains.

Let us add that recently it became very popular to use a higher than second order (Somigliana-Pizzetti's) reference field in Stokes's formulation. For this purpose, a global field (cf., the article, "Global Gravity"), preferably of a pure satellite origin, is selected and a residual disturbing potential on, or geoidal height above, a *reference spheroid* defined by such a field (cf., Fig. 5 in the article "Global Gravity") is then produced. This approach may be termed a *generalized Stokes formulation* (Vaníček and Sjöberg, 1991), and it is attractive because it alleviates the negative impact of the existing nonhomogeneous terrestrial gravity coverage by attenuating the effect of distant data in the Stokes integral (44). For illustration, so computed a geoid for a part of North America is shown in Fig. 11. It should be also mentioned that the evaluation of Stokes' integral is often sought in terms of Fast Fourier Transform.

G. Molodenskij's Geodetic Boundary Value Problem

In the mid-20th century, Russian physicist M. S. Molodenskij formulated a different scalar boundary value problem to solve for the disturbing potential outside the earth (Molodenskij, Eremeev, and Yurkine 1960). His criticism of Stokes' approach was that the geoid is an equipotential surface internal to the earth and as such requires detailed knowledge of internal (topographical) earth mass density, which we will never have. He then proceeded to replace Stokes's choice of the boundary (geoid) by the earth's surface and to solve for $T(\mathbf{r})$ outside the earth.

At the earth's surface, the Poisson equation changes dramatically. The first term on its right-hand side, equal to $-4\pi G\rho(\mathbf{r})$, changes from 0 to a value of approximately $2.24 * 10^{-6}$ s^{-2} (more than three orders of magnitude larger than the value of the second term). The latter value is obtained using the density ρ of the most common rock, granite. Conventionally, the value of the first term right on the earth's surface is defined as $-4k(\mathbf{r}_t)\pi G\rho(\mathbf{r}_t)$, where the function $k(\mathbf{r}_t)$ has a value between 0 and 1 depending on the shape of the earth's surface; it equals 1/2 for a flat surface, close to 0 for a "needle-like" topographical feature, and close to 1 for a "well-like" feature. In Molodenskij's solution, the Poisson equation has to be integrated over the earth's surface and the above variations of the right-hand side cause problems, particularly on steep surfaces. It is still uncertain just how accurate a solution can be obtained with Molodenskij's approach; it looks as if bypassing the topographical density may have introduced another problem caused by the real shape of topographical surface.

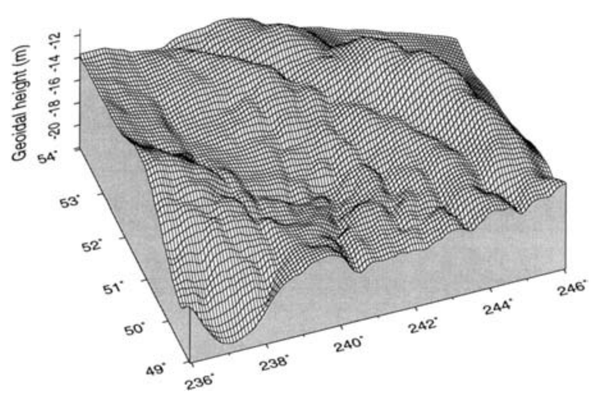

FIGURE 11 Detailed geoid for an area in the Canadian Rocky Mountains (computed at the University of New Brunswick).

For technical reasons, the integration is not carried out on the earth's surface but on a surface which differs from the earth's surface by about as much as the geoidal height N; this surface is the telluroid encountered already in Section II.B. The solution for T on the earth's surface (more accurately on the telluroid) is given by the following integral equation:

$$\begin{aligned} T(\mathbf{r}_t) &- R/(2\pi)\int_{tell}[\partial/\partial n'|\mathbf{r}_t-\mathbf{r}_t'|^{-1}-|\mathbf{r}_t-\mathbf{r}_t'|^{-1} \\ &\times \cos\beta/\gamma\,\partial\gamma/\partial H^{\mathrm{N}}]T(\mathbf{r}_t')\,d\Omega' \\ &= R/(2\pi)\int_{tell}[\Delta g(\mathbf{r}_t')-\gamma][\xi'\tan\beta_1+\eta'\tan\beta_2] \\ &\times|\mathbf{r}_t-\mathbf{r}_t'|^{-1}\cos\beta\,d\Omega', \end{aligned} \tag{46}$$

where n' is the outer normal to the telluroid; β is the maximum slope of the telluroid (terrain); β_1, β_2 are the north–south and east–west terrain slopes; and ξ', η' are deflection components on the earth's surface. This integral equation is too complicated to be solved directly and simplifications must be introduced. The solution is then sought in terms of successive iterations, the first of which has an identical shape to the Stokes integral (44). Subsequent iterations can be thought of as supplying appropriate corrective terms (related to topography) to the basic Stokes solution.

In fact, the difference between the telluroid and the earth's surface, called the *height anomaly* ζ, is what can be determined directly from Molodenskij's integral using a surface density function. It can be interpreted as a "geoidal height" in Molodenskij's sense as it defines the Molodenskij "geoid" introduced in Section II.B (called quasi-geoid, to distinguish it from the real geoid). The difference between the geoid and quasi-geoid may reach up to a few meters in mountainous regions, but it disappears at sea (Pick, Pícha, and Vyskočil, 1973). It can be seen from

Section II.B that the difference may be evaluated from orthometric and normal heights (referred to the geoid, and quasi-geoid, respectively):

$$\zeta - N = H^{\mathrm{O}} - H^{\mathrm{N}}, \tag{47}$$

subject to the error in the orthometric height. Approximately, the difference is also equal to $-\Delta g^{\mathrm{Bouguer}} H^{\mathrm{O}}/\gamma$.

H. Global and Local Modeling of the Field

Often, it is useful to describe the different parameters of the earth's gravity field by a series of spherical or ellipsoidal harmonic functions (cf., the article "Global Gravity" in this volume). This description is often referred to as the *spectral form*, and it is really the only practical global description of the field parameters. The spectral form, however, is useful also in showing the spectral behavior of the individual parameters. We learn, for instance, that the series for T, N, or ζ converge to 0 much faster than the series for Δg, ζ, η do: we say that the T, N, or ζ fields are smoother than the Δg, ζ, η fields. This means that a truncation of the harmonic series describing one of the smoother fields does not cause as much damage as does a truncation for one of the rougher fields by leaving out the higher "frequency" components of the field. The global description of the smoother fields will be closer to reality.

If higher frequency information is of importance for the area of interest, then it is more appropriate to use a point description of the field. This form of a description is called in geodesy the *spatial form*. Above we have seen only examples of spatial expressions, in the article "Global Gravity" only spectral expressions are used. Spatial expressions involving surface convolution integrals over the whole earth [cf., Eqs. (44) and (46)], can be always transformed into corresponding spectral forms and vice versa. The two kinds of forms can be, of course, also combined as we saw in the case of generalized Stokes's formulation (Section III.F).

IV. GEO-KINEMATICS

A. Geodynamics and Geo-Kinematics

Dynamics is that part of physics that deals with forces (and therefore masses), and motions in response to these forces and geodynamics is that part of geophysics that deals with the dynamics of the earth. In geodesy, the primary interest is the geometry of the motion and of the deformation (really just a special kind of motion) of the earth or its part. The geometrical aspect of dynamics is called kinematics, and therefore, we talk here about *geo-kinematics*. As a matter of fact, the reader might have noticed already in the above paragraphs involving physics how mass was eliminated from the discussions, leaving us with only kinematic descriptions.

Geo-kinematics is one of the obvious fields where cooperation with other sciences, geophysics here, is essential. On the one hand, geometrical information on the deformation of the surface of the earth is of much interest to a geophysicist who studies the forces/stresses responsible for the deformation and the response of the earth to these forces/stresses. On the other hand, it is always helpful to a geodesist to get an insight into the physical processes responsible for the deformation he is trying to monitor.

Geodesists have studied some parts of geo-kinematics, such as those dealing with changes in the earth's rotation, for a long time. Other parts were only more recently incorporated into geodesy because the accuracy of geodetic measurements had not been good enough to see the real-time evolution of deformations occurring on the surface of the earth. Nowadays, geodetic monitoring of crustal motions is probably the fastest developing field of geodesy.

B. Temporal Changes in Coordinate Systems

In Section II.A we encountered a reference to the earth's "spin axis" in the context of geocentric coordinate systems. It is this axis the earth spins around with 366.2564 *sidereal revolutions* (cf., Section III.A), or 365.2564 revolutions with respect to the sun (defining *solar days*) per year. The spin axis of the earth moves with respect to the universe (directions to distant stars, the realization of an inertial coordinate system), undergoing two main motions: one very large, called precession, with a period of about 26,000 years and the other much smaller, called nutation, with the main period of 18.6 years. These motions must be accounted for when doing astronomical measurements of either the optical or radio variety (see Section II.C).

In addition to precession and nutation, the spin axis also undergoes a torque-free nutation, also called a wobble, with respect to the earth. More accurately, the *wobble* should be viewed as the motion of the earth with respect to the instantaneous spin axis. It is governed by the famous Euler's gyroscopic equation:

$$\mathbf{J}\boldsymbol{\omega} + \boldsymbol{\omega} \times \mathbf{J}\boldsymbol{\omega}, \tag{48}$$

where $\mathbf{J}$ is the earth's *tensor of inertia* and $\boldsymbol{\omega}$ is the instantaneous spin *angular velocity vector* whose magnitude ω we have met several times above. Some relations among the diagonal elements of $\mathbf{J}$, known as moments of inertia, can be inferred from astronomical observations giving a solution $\boldsymbol{\omega}$ of this differential equation. Such a solution describes a periodic motion with a period of about 305 sidereal days, called Euler's period.

Observations of the wobble (Fig. 12) have shown that beside the Euler component, there is also an annual periodic component of similar magnitude plus a small drift. The magnitude of the periodic components fluctuates

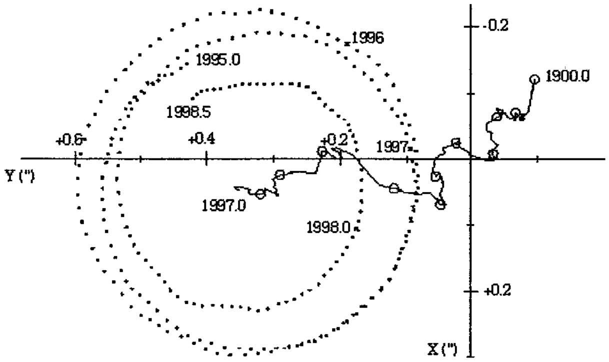

FIGURE 12 Earth pole wobble. (Source: International Earth Rotation Service.)

around 0.1 arcsec. Thus, the wobble causes a displacement of the pole (intersection of instantaneous spin axis with the earth's surface) of several meters. Furthermore, systematic observations show also that the period of the Euler component is actually longer by some 40% than predicted by the Euler equation. This discrepancy is caused by the nonrigidity of the earth, and the actual period, around 435 solar days, is now called Chandler's. The actual motion of the pole is now being observed and monitored by IAG's IERS on a daily basis. It is easy to appreciate that any coordinate system linked to the earth's spin axis (cf., Section II.A) is directly affected by the earth pole wobble which, therefore, has to be accounted for.

As the direction of ω varies, so does its magnitude ω. The variations of spin velocity are also monitored in terms of the *length of the day* (LOD) by the Bureau Internationale des Poids et Mesures—Time Section (BIPM) on a continuous basis. The variations are somewhat irregular and amount to about 0.25 msec/year—the earth's spin is generally slowing down.

There is one more temporal effect on a geodetic coordinate system, that on the datum for vertical positioning, i.e., on the geoid. It should be fairly obvious that if the reference surface for heights (orthometric and dynamic) changes, so do the heights referred to it. In most countries, however, these changes are not taken too seriously. The geoid indicated by the MSL, as described in Section III.B, changes with time both in response to the mass changes within the earth and to the MSL temporal changes. The latter was discussed in Section III.B, the former will be discussed in Section IV.C.

C. Temporal Changes in Positions

The earth is a deformable body and its shape is continuously undergoing changes caused by a host of stresses; thus, the positions of points on the earth's surface change continuously. Some of the stresses that cause the *deformations* are known, some are not, with the best known being the tidal stress (Melchior, 1966). Some loading stresses causing crustal deformation are reasonably well known, such as those caused by filling up water dams, others, such as sedimentation and glaciation, are known only approximately. Tectonic and other stresses can be only inferred from observed deformations (cf., the article, "Tectonophysics" in this encyclopedia). The response of the earth to a stress, i.e., the deformation, varies with the temporal frequency and the spatial extent of the stress and depends on the rheological properties of the whole or just a part of the earth. Some of these properties are now reasonably well known, some are not. The ultimate role of geodesy vis-à-vis these deformations is to take them into account for predicting the temporal variations of positions on the earth's surface. This can be done relatively simply for deformations that can be modeled with a sufficient degree of accuracy (e.g., tidal deformations), but it cannot yet be done for other kinds of deformations, where the physical models are not known with a sufficient degree of certainty. Then the role of geodesy is confined to monitoring the surface movements, kinematics, to provide the input to geophysical investigations.

A few words are now in order about tidal deformations. They are caused by the moon and sun gravitational attraction (see Fig. 13). Tidal potential caused by the next most influential celestial body, Venus, amounts to only 0.036% of the luni-solar potential; in practice only the *luni-solar tidal potential* is considered. The tidal potential $W_t^{☾}$ of the moon, the lunar tidal potential, is given by the following equation:

$$W_t^{☾}(\mathbf{r}) = GM^{☾}|\mathbf{r}^{☾} - \mathbf{r}|^{-1} \sum_{j=2}^{\infty} \mathbf{r}^j |\mathbf{r}^{☾} - \mathbf{r}|^{-j} P_j(\cos \Psi^{☾}), \tag{49}$$

where the symbol ☾ refers to the moon, P_j is the Legendre function of degree j, and Ψ is the geocentric angle

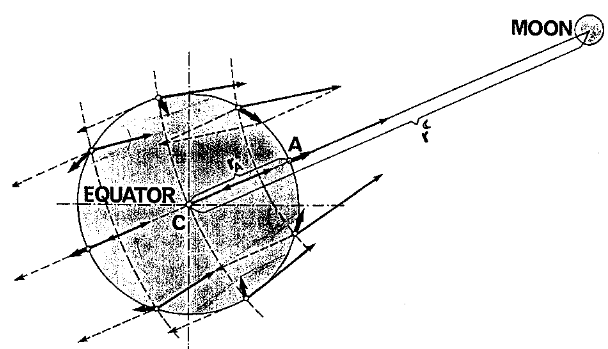

FIGURE 13 The provenance of tidal force due to the moon.

between $\mathbf{r}$ and $\mathbf{r}^{☾}$. The tidal potential of the sun $W_t^{\odot}$ is given by a similar series, and it amounts to about 46% of the lunar tidal potential. To achieve an accuracy of 0.03%, it is enough to take the first two terms from the lunar series and the first term from the solar series.

The temporal behavior of tidal potential is periodic, as dictated by the motions of the moon and the sun (but see Section IV.D for the tidal constant effect). The tidal waves, into which the potential can be decomposed, have periods around 1 day, called diurnal; around 12 h, called semidiurnal; and longer. The tidal deformation as well as all the tidal effects (except for the sea tide, which requires solving a boundary value problem for the Laplace tidal equation, which we are not going to discuss here) can be relatively easily evaluated from the tidal potential. This is because the rheological properties of the earth for global stresses and tidal periods are reasonably well known. Geodetic observations as well as positions are affected by tidal deformations, and these effects are routinely corrected for in levelling, VLBI (cf., Section II.C), satellite positioning (cf., Section V.B), and other precise geodetic works. For illustration, the range of orthometric height tidal variation due to the moon is 36 cm, that due to the sun is another 17 cm. With tidal deformation being of a global character, however, these tidal variations are all but imperceptible locally.

Tectonic stresses are not well known, but horizontal motion of tectonic plates has been inferred from various kinds of observations, including geodetic, with some degree of certainty, and different maps of these motions have been published (cf., the article, "Tectonophysics" in this encyclopedia). The AM0-2 absolute plate motion model was chosen to be an "associated velocity model" in the definition of the ITRF (see Section II.A), which together with direct geodetic determination of horizontal velocities define the temporal evolution of the ITRF and thus the temporal evolution of horizontal positions. From other ongoing earth deformations, the post-glacial rebound is probably the most important globally as it is large enough to affect the flattening [Eq. (31)] of the mean earth ellipsoid as we will see in Section IV.C.

Mapping and monitoring of ongoing motions (deformations) on the surface of the earth are done by repeated position determination. In global monitoring the global techniques of VLBI and satellite positioning are used (see Section V.B). For instance, one of the IAG services, the IGS (cf., Section I.C), has been mandated with monitoring the horizontal velocities of a multitude of permanent tracking stations under its jurisdiction. In regional investigations the standard terrestrial geodetic techniques such as horizontal and vertical profiles, horizontal and levelling network re-observation campaigns are employed. The positions are then determined separately from each campaign with subsequent evaluation of *displacements*. Preferably, the displacements (horizontal or vertical) are estimated directly from the observations collected in all the campaigns. The latter approach allows the inclusion of correlations in the mathematical model for the displacement estimation with more correct estimates ensuing. For illustration, Fig. 14 shows such estimated horizontal displacements from the area of Imperial Valley, CA, computed from standard geodetic observations.

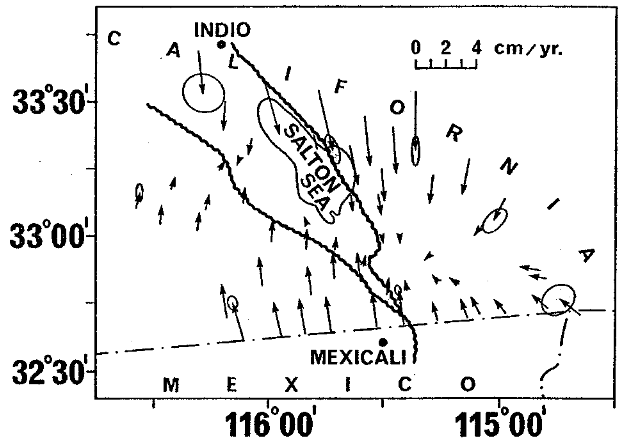

FIGURE 14 Mean annual horizontal displacements in Imperial Valley, CA, computed from data covering the period 1941–1975.

It is not possible to derive absolute displacements from relative positions. Because the repeated horizontal position determination described above is usually of a relative kind, the displacements are indeterminate. It then makes sense to deal only with relative quantification of deformation such as *strain*. Strain is, most generally, described by the displacement gradient matrix $\mathbf{S}$; denoting the 2D displacement vector of a point $\mathbf{r}$ by $\mathbf{v}(\mathbf{r})$ we can write

$$\mathbf{S}(\mathbf{r}) = \nabla' \mathbf{v}^{\mathrm{T}}(\mathbf{r}), \tag{50}$$

where ∇' is the 2D nabla operator. The inverse transformation

$$\mathbf{v}(\mathbf{r}) = \mathbf{S}(\mathbf{r})\,\mathbf{r} + \mathbf{v}_0, \tag{51}$$

where $\mathbf{v}_0$ describes the translational indeterminacy, shows better the role of $\mathbf{S}$, which can be also understood as a Jacobian matrix [cf., Eq. (14)] which transforms from the space of positions (real 2D space) into the space of displacements. The symmetrical part of $\mathbf{S}$ is called the deformation tensor in the mechanics of continuum; the antisymmetrical part of $\mathbf{S}$ describes the rotational deformation. Other strain parameters can be derived from $\mathbf{S}$.

D. Temporal Changes in Gravity Field

Let us begin with the two requirements defining the geoid presented in Section III.B: the constancy of W_0 and the fit of the equipotential surface to the MSL. These two requirements are not compatible when viewed from the point of time evolution or the temporal changes of the earth's gravity field. The MSL grows with time at a rate estimated to be between 1 and 2 mm year^{-1} (the eustatic water level change), which would require systematic lowering of the value of W_0. The mass density distribution within the earth changes with time as well (due to tectonic motions, post-glacial rebound, sedimentation, as discussed above), but its temporal effect on the geoid is clearly different from that of the MSL. This dichotomy has not been addressed by the geodetic community yet.

In the areas of largest documented changes of the mass distribution (those caused by the post-glacial rebound), the northeastern part of North America and Fennoscandia, the maximum earth surface uplift reaches about 1 cm year^{-1}. The corresponding change in gravity value reaches up to 0.006 mGal year^{-1} and the change in the equipotential surfaces $W =$ const. reaches up to 1 mm year^{-1}. The potential coefficient $C_{2,0}$ (cf., Section III.B), or rather its unitless version $J_{2,0}$ that is used most of the time in gravity field studies, as observed by satellites shows a temporal change caused by the rebound. The rebound can be thought of as changing the shape of the geoid, and thus the shape of the mean earth ellipsoid, within the realm of observability.

As mentioned in Section IV.C, the tidal effect on gravity is routinely evaluated and corrected for in precise gravimetric work, where the effect is well above the observational noise level. By correcting the gravity observations for the periodic tidal variations we eliminate the temporal variations, but we do not eliminate the whole tidal effect. In fact, the luni-solar tidal potential given by Eq. (49) has a significant constant component responsible for what is called in geodesy *permanent tide*. The effect of permanent tide is an increased flattening of gravity equipotential surfaces, and thus of the mean earth ellipsoid, by about one part in 10^5. For some geodetic work, the tideless mean earth ellipsoid is better suited than the mean tide ellipsoid, and, consequently, both ellipsoids can be encountered in geodesy.

Temporal variations of gravity are routinely monitored in different parts of the world. These variations (corrected for the effect of underground water fluctuations) represent an excellent indicator that a geodynamical phenomenon, such as tectonic plate motion, sedimentation loading, volcanic activity, etc., is at work in the monitored region. When gravity monitoring is supplemented with vertical motion monitoring, the combined results can be used to infer the physical causes of the monitored phenomenon. Let us mention that the earth pole wobble introduces also observable variations of gravity of the order of about 0.008 mGal. So far, these variations have been of academic interest only.

V. SATELLITE TECHNIQUES

A. Satellite Motion, Functions, and Sensors

Artificial satellites of the earth appeared on the world scene in the late 1950s and were relatively early embraced by geodesists as the obvious potential tool to solve world-wide geodetic problems. In geodetic applications, satellites can be used both in positioning and in gravitational field studies as we have alluded to in the previous three sections. Geodesists have used many different satellites in the past 40 years, ranging from completely passive to highly sophisticated active (transmitting) satellites, from quite small to very large. Passive satellites do not have any sensors on board and their role is basically that of an orbiting target. Active satellites may carry a large assortment of sensors, ranging from accurate clocks through various counters to sophisticated data processors, and transmit the collected data down to the earth either continuously or intermittently.

Satellites orbit the earth following a trajectory which resembles the Keplerian ellipse that describes the motion in radial field (cf., Section III.A); the higher the satellite is, the closer its orbit to the Keplerian ellipse. Lower orbiting satellites are more affected by the irregularities of the earth's gravitational field, and their *orbit* becomes more perturbed compared to the Keplerian orbit. This curious behavior is caused by the inherent property of gravitational field known as the attenuation of shorter wavelengths of the field with height and can be gleaned from Eq. (4) in the article "Global Gravity." The ratio a/r is always smaller than 1 and thus tends to disappear the faster the larger its exponent l which stands for the spatial wave number of the field. We can see that the *attenuation factor* $(a/r)^\ell$ goes to 0 for growing ℓ and growing r; for $r > a$, we have:

$$\lim_{l\to\infty} (a/r)^\ell = 0. \tag{52}$$

We shall see in Section V.C how this behavior is used in studying the gravitational field by means of satellites.

Satellite orbits are classified as high and low orbits, polar orbits (when the orbital plane contains the spin axis of the earth), equatorial orbits (orbital plane coincides with the equatorial plane of the earth), and pro-grade and retro-grade or bits (the direction of satellite motion is either eastward or westward). The lower the orbit is, the faster the satellite circles the earth. At an altitude of about 36,000 km, the orbital velocity matches that of the earth's spin, its orbital period becomes 24 h long, and the satellite moves only in one meridian plane (its motion is neither pro- nor retro-grade). If its orbit is equatorial, the satellite remains in one position above the equator. Such an orbit (satellite) is called *geostationary*.

Satellites are *tracked* from points on the earth or by other satellites using electromagnetic waves of frequencies that can penetrate the ionosphere. These frequencies propagate along a more or less straight line and thus require intervisibility between the satellite and the tracking device (transmitter, receiver, or reflector); they range from microwave to visible (from 30 to 10^9 MHz). The single or double passage of the electromagnetic signal between the satellite and the tracking device is accurately timed and the distance is obtained by multiplying the time of passage by the propagation speed. The propagation speed is close to the speed of light in vacuum, i.e., 299, 792, 460 m s^{-1}, with the departure being due to the delay of the wave passing through the atmosphere and ionosphere.

Tracked satellite orbits are then computed from the measured (observed) distances and known positions of the tracking stations by solving the equations of motion in the earth's gravity field. This can be done quite accurately (to a centimeter or so) for smaller, spherical, homogeneous, and high-flying spacecraft that can be tracked by lasers. Other satellites present more of a problem; consequently, their orbits are less well known. When orbits are extrapolated into the future, this task becomes known as the *orbit prediction*. Orbit computation and prediction are specialized tasks conducted only by larger geodetic institutions.

B. Satellite Positioning

The first satellite used for positioning was a large, light, passive balloon (ECHO I, launched in 1960) whose only role was to serve as a naturally illuminated moving target on the sky. This target was photographed against the star background from several stations on the earth, and directions to the satellite were then derived from the known directions of surrounding stars. From these directions and from a few measured interstation distances, positions of the camera stations were then computed. These positions were not very accurate though because the directions were burdened by large unpredictable refraction errors (cf., Section II.B). It became clear that *range (distance) measurement* would be a better way to go.

The first practical satellite positioning system (TRANSIT) was originally conceived for relatively inaccurate naval navigation and only later was adapted for much more accurate geodetic positioning needs. It was launched in 1963 and was made available for civilian use 4 years later. The system consisted of several active satellites in circular polar orbits of an altitude of 1074 km and an orbital period of 107 min, the positions (ephemeredes given in the OR coordinate system—cf., Section II.A) of which were continuously broadcast by the satellites to the users. Also transmitted by these satellites were two signals at fairly stable frequencies of 150 and 400 MHz controlled by crystal oscillators. The user would then receive both signals (as well as the ephemeris messages) in his specially constructed TRANSIT *satellite receiver* and compare them with internally generated stable signals of the same frequencies. The beat frequencies would then be converted to *range rates* by means of the Doppler equation

$$\lambda_R = \lambda_T(1 + v/c)(1 + v^2/c^2)^{\frac{1}{2}}, \qquad (53)$$

where v is the projection of the range rate onto the receiver-satellite direction; λ_R, λ_T are the wavelengths of the received and the transmitted signals; and c is the speed of light in vacuum. Finally, the range rates and the satellite positions computed from the broadcast ephemeredes would be used to compute the generic position **r** of the receiver (more accurately, the position of receiver's antenna) in the CT coordinate system (cf., Section II.A). More precise satellite positions than those broadcast by the satellites themselves were available from the U.S. naval ground control station some time after the observations have taken place. This control station would also predict the satellite orbits and upload these predicted orbits periodically into the satellite memories.

At most, one TRANSIT satellite would be always "visible" to a terrestrial receiver. Consequently, it was not possible to determine the sequence of positions (trajectory) of a moving receiver with this system; only *position lines* (lines on which the unknown position would lie) were determinable. For determining an accurate position (1 m with broadcast and 0.2 m with precise ephemeredes) of a stationary point, the receiver would have to operate at that point for several days.

Further accuracy improvement was experienced when two or more receivers were used simultaneously at two or more stationary points, and relative positions in terms of interstation vectors $\Delta\mathbf{r}$ were produced. The reason for the increased accuracy was the attenuation of the effect of common errors/biases (atmospheric delays, orbital errors, etc.) through differencing. This relative or differential mode of using the system became very popular and remains the staple mode for geodetic positioning even with the more modern GPS used today.

In the late 1970s, the U.S. military started experimenting with the GPS (originally called NAVSTAR). It should be mentioned that the military have always been vitally interested in positions, instantaneous and otherwise, and so many developments in geodesy are owed to military initiatives. The original idea was somewhat similar to that of the now defunct TRANSIT system (active satellites with oscillators on board that transmit their own ephemeredes) but to have several satellites orbiting the earth so that at least four of them would be always "visible" from any point on the earth. Four is the minimal number needed to get an instantaneous 3D position by measuring simultaneously the four ranges to the visible satellites: three for the three coordinates and one for determining the ever-changing offset between the satellite and receiver oscillators.

Currently, there are 28 active GPS satellites orbiting the earth at an altitude of 20,000 km spaced equidistantly in orbital planes inclined 60 arc-degrees with respect to the equatorial plane. Their orbital period is 12 h. They transmit two highly coherent cross-polarized signals at frequencies of 1227.60 and 1575.42 MHz, generated by atomic oscillators (cesium and rubidium) on board, as well as their own (broadcast) ephemeredes. Two pseudo-random timing sequences of frequencies 1.023 and 10.23 MHz—one called P-code for restricted users only and the other called C/A-code meant for general use—are modulated on the two carriers. The original intent was to use the timing codes for observing the ranges for determining instantaneous positions. For geodetic applications, so determined ranges are too coarse and it is necessary to employ the carriers themselves.

Nowadays, there is a multitude of GPS receivers available off the shelf, ranging from very accurate, bulky, and relatively expensive "*geodetic receivers*" all the way to hand-held and wrist-mounted cheap receivers. The cheapest receivers use the C/A-code ranging (to several satellites) in a point-positioning mode capable of delivering an accuracy of tens of meters. At the other end of the receiver list, the most sophisticated geodetic receivers use both carriers for the ranging in the differential mode. They achieve an accuracy of the interstation vector between a few millimeters for shorter distances and better than $S10^{-7}$, where S stands for the interstation distance, for distances up to a few thousand kilometers.

In addition to the global network of tracking stations maintained by the IGS (cf., Section I.C), there have been networks of continually tracking GPS stations established in many countries and regions; for an illustration, see Fig. 15. The idea is that the tracking stations are used as traditional position control stations and the tracking data are as well used for GPS satellite orbit improvement. The stations also provide "*differential corrections*" for roving GPS users in the vicinity of these stations. These corrections are used to eliminate most of the biases (atmospheric delays, orbital errors, etc.) when added to point positions of roving receivers. As a result of the technological and logistical improvements during the past 20 years, GPS positioning is now cheap, accurate, and used almost

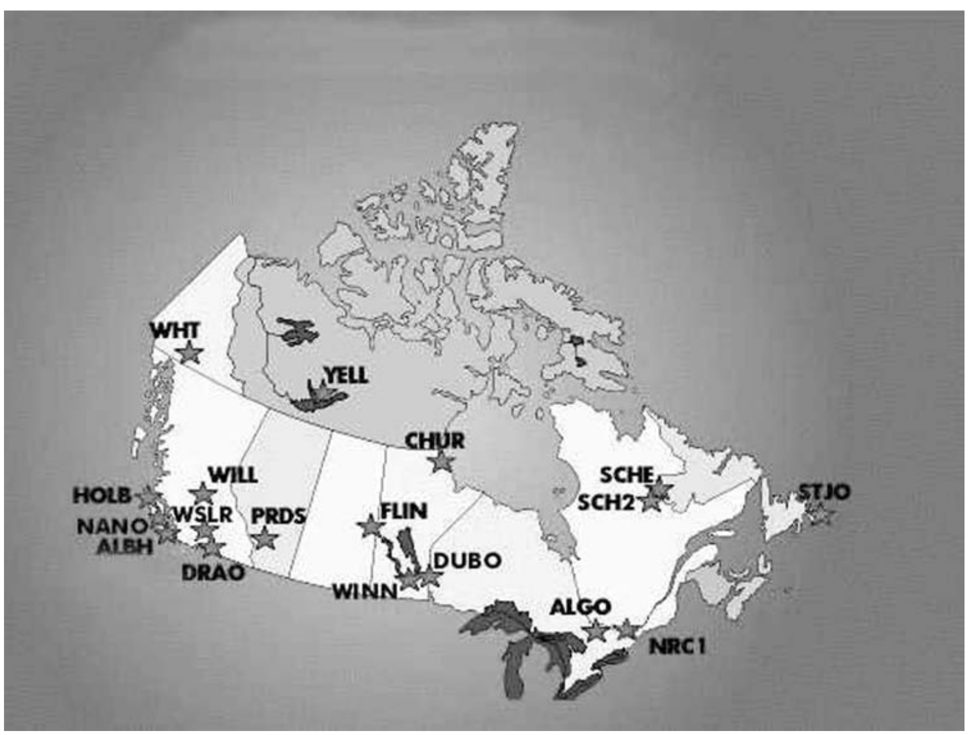

FIGURE 15 Canadian Active Control System. (Source: www.nrcan.gc.ca. Copyright: Her Majesty the Queen in Right of Canada, Natural Resources Canada, Geodetic Survey Division. All rights reserved.)

everywhere for both positioning and precise navigation in preference to classical terrestrial techniques.

The most accurate absolute positions **r** (standard error of 1 cm) are now determined using small, heavy, spherical, high-orbiting, passive satellites equipped with retro-reflectors (LAGEOS 1, LAGEOS 2, STARLETTE, AJISAI, etc.) and laser ranging. The technique became known as *Satellite Laser Ranging* (SLR), and the reason for its phenomenal accuracy is that the orbits of such satellites can be computed very accurately (cf., Section V.A). Also, the ranging is conducted over long periods of time by means of powerful astronomical telescopes and very precise timing devices. Let us just mention that SLR is also used in the relative positioning mode, where it gives very accurate results. The technique is, however, much more expensive than, say, GPS and is thus employed only for scientific investigations

Finally, we have to mention that other satellite-based positioning exist. These are less accurate systems used for nongeodetic applications. Some of them are used solely in commercial application. At least one technique deserves to be pointed out, however, even though it is not a positioning technique per se. This is the synthetic aperture radar interferometry (INSAR). This technique uses collected reflections from a space-borne radar. By sophisticated computer processing of reflections collected during two overflights of the area of interest, the pattern of ground deformation that had occurred between the two flights can be discerned (Massonnet *et al*., 1993). The result is a map of relative local deformations, which may be used, for instance, as a source of information on co-seismic activity. Features as small as a hundred meters across and a decimeter high can be recognized.

C. Gravitational Field Studies by Satellites

The structure of the earth's gravity field was very briefly mentioned in Section III.H, where the field wavelengths were discussed in the context of the spectral description of the global field. A closer look at the field reveals that:

1. The field is overwhelmingly radial (cf., Section III.A) and the first term in the potential series, GM/r, is already a fairly accurate (to about 10^{-3}) description of the field; this is why the radial field is used as a model field (cf., Section III.B) in satellite studies.
2. The largest departure from radiality is described by the second degree term $J_{2,0}$ (cf., Section IV.D), showing the ellipticity of the field, which is about 3 orders of magnitude smaller than the radial part of the field.
3. The remaining wavelength amplitudes are again about 3 orders of magnitude smaller and they further decrease with increasing wave number ℓ. The decrease of amplitude is seen, for instance in Fig. 8 in the article "Global Gravity." In some studies it is possible to use a mathematical expression describing the decrease, such as the experimental Kaula's rule of thumb, approximately valid for ℓ between 2 and 40:

$$\sqrt{\Sigma_{m=2}^{\ell}\left(C_{lm}^2 + S_{lm}^2\right)} \approx R * 10^{-5}/\ell^2, \qquad (54)$$

where C_{lm} and S_{lm} are the potential coefficients (cf., Eq. (4) in "Global Gravity").

As discussed in Section V.A, the earth's gravity field also gets smoother with altitude. Thus, for example, at the altitude of lunar orbit (about 60 times the radius of the earth), the only measurable departure from radiality is due to the earth's ellipticity and even this amounts to less than 3×10^{-7} of the radial component. Contributions of shorter wavelength are 5 orders of magnitude smaller still. Consequently, a low-orbiting satellite has a "bumpier to use a satellite as a gravitation-sensing device, we get more detailed information from low-orbiting spacecraft.

The idea of using satellites to "measure" gravitational field (we note that a satellite cannot sense the total gravity field, cf., Section III.A) stems from the fact that their orbital motion (free fall) is controlled predominantly by the earth's gravitational field. There are other forces acting on an orbiting satellite, such as the attraction of other celestial bodies, air friction, and solar radiation pressure, which have to be accounted for mathematically. Leaving these forces alone, the equations of motion are formulated so that they contain the gravitational field described by potential coefficients C_{lm} and S_{lm}. When the observed orbit does not match the orbit computed from the known potential coefficients, more realistic potential coefficient values can be derived. In order to derive a complete set of more realistic potential coefficients, the procedure has to be formulated for a multitude of different orbits, from low to high, with different inclinations, so that these orbits sample the space above the earth in a homogeneous way. We note that because of the smaller amplitude and faster attenuation of shorter wavelength features, it is possible to use the described *orbital analysis* technique only for the first few tens of degrees ℓ. The article "Global Gravity" shows some numerical results arising from the application of this technique.

Other techniques such as "satellite-to-satellite tracking" and "gradiometry" (see "Global Gravity") are now being used to study the shorter wavelength features of the gravitational field. A very successful technique, "satellite altimetry," a hybrid between a positioning technique and gravitational field study technique (see "Global Gravity") must be also mentioned here. This technique has now been used for some 20 years and has yielded some important results.

SEE ALSO THE FOLLOWING ARTICLES

EARTH SCIENCES, HISTORY OF • EARTH'S MANTLE (GEOPHYSICS) • EXPLORATION GEOPHYSICS • GEOMAGNETISM • GLOBAL GRAVITY MODELING • GRAVITATIONAL WAVE ASTRONOMY • RADIO ASTRONOMY, PLANETARY • REMOTE SENSING FROM SATELLITES • TECTONOPHYSICS

BIBLIOGRAPHY

Berthon, S., and Robinson, A. (1991). "The Shape of the World," George Philip Ltd., London.

Boal, J. D., and Henderson, J. P. (1988). The NAD 83 Project—Status and Background. *In* "Papers for the CISM Seminars on the NAD'83 Redefinition in Canada and the Impact on Users" (J. R. Adams ed.), The Canadian Institute of Surveying and Mapping, Ottawa, Canada.

Bomford, G. (1971). "Geodesy," 3rd ed., Oxford Univ. Press, London.

Boucher, C., and Altamini, Z. (1996). "International Terrestrial Reference Frame." *GPS World* **7**(9), 71–75.

Bruns, H. (1878). "Die Figur der Erde," Publication des Königlichen Preussischen Geodätischen Institutes, Berlin, Germany.

Burša, M., Raděj, K., Šíma, Z., True, S. A., and Vatrt, V. (1997). "Deter-

Satellite Altimetry," *Studia Geophys. Geod*. **41,** 203–216.

Grafarend, E. W., and Sansò, F., ed. (1985). "Optimization and Design of Geodetic Networks," Springer-Verlag, Berlin/New York.

Heiskanen, W. A., and Moritz, H. (1967). "Physical Geodesy," Freeman, San Francisco.

Helmert, F. R. (1880). "Die mathematischen und physikalishen Theorien der höheren Geodäsie," Vol. I, Minerva, G. M. B. H. Reprint, 1962.

Hirvonen, R. A. (1971). "Adjustment by Least Squares in Geodesy and Photogrammetry," Ungar, New York.

International Association of Geodesy (1974). "The International Gravity Standardization Net, 1971," Special Publication No. 4, Paris, France.

International Association of Geodesy (1980). "The geodesist's handbook," *Bull. Géodesique* **54**(3).

Lee, L. P. (1976). "Conformal Projections Based on Elliptic," Cartographica Monograph 16, B. V. Gutsell, Toronto, Canada.

Massonnet, D., Rossi, M., Carmona, C., Adragna, F., Peltzer, G., Feigl, K., and Rabaute, T. (1993), "The displacement field of the Landers earthquake mapped by radar interferometry," *Nature* **364**.

Melchior, P. (1966). "The Earth Tides," Pergamon, Elmsford, NY.

Mikhail, E. M. (1976). "Observations and Least Squares," IEP-A dun-Donnelley Publisher.

Molodenskij, M. S., Eremeev, V. F., and Yurkina, M. I. (1960). "Methods for Study of the External Gravitational Field and Figure of the Earth," Translated from Russian by the Israel Program for Scientific Translations for the Office of Technical Services, U.S. Department of Commerce, Washington, DC., 1962.

Mueller, I. I. (1969). "Spherical and Practical Astronomy as Applied to Geodesy," Ungar, New York.

Pick, M., Pícha, J., and Vyskočil, V. (1973). "Theory of the Earth's Gravity Field," Elsevier, Amsterdam/New York.

Ries, J. C., Eanes, R. J., Shum, C. K., and Watkins, M. M. (1992). "Progress in the determination of the gravitational coefficient of the earth," *Geophys. Res. Lett*. **19**(6), 529–531.

Schwarz, C. R., ed. (1989). "North American Datum of 1983," NOAA Professional Paper NOS 2, National Geodetic Information Center, National Oceanic and Atmospheric Administration, Rockville, MD.

Stokes, G. G. (1849). "On the variation of gravity at the surface of the earth," *Trans. Cambridge Philos. Soc.* **VIII,** 672–695.

U.S. Department of Commerce (1973). "The North American Datum," Publication of the National Ocean Survey of NOAA, Rockville, MD.

Vaníček, P., and Krakiwsky, E. J. (1986). "Geodesy: The Concepts," 2nd ed., North-Holland, Amsterdam.

Vaníček, P., and Sjöberg, L. E. (1991). "Reformulation of Stokes's theory for higher than second-degree reference field and modification of integration kernels," *J. Geophys. Res.* **96**(B4), 6529–6539.

Vaníček, P., Huang, J., Novák, P., Véronneau, M., Pagiatakis, S., Martinec, Z., and Featherstone, W. E. (1999). "Determination of boundary values for the Stokes-Helmert problem," *J. Geod.* **73,** pp. 180–192.

Zakatov, P. S. (1953). "A Course in Higher Geodesy," Translated from Russian by the Israel Program form Scientific Translations for the Office of Technical Services, U.S. Department of Commerce, Washington, DC., 1962.