Methods and geophysical applications of satellite geodesy

K Lambeck

Rep. Prog. Phys. 1979 42 547–628
Printed in Great Britain © 1979

The Institute of Physics
Methods and geophysical applications of satellite geodesy

KURT LAMBECK
Research School of Earth Sciences, Australian National University, Canberra, Australia

Abstract

Precise geodetic measurements of satellite motions provide information on the forces acting on the satellite and enable the positions of the tracking stations to be estimated in a geocentric reference frame. Both yield geophysically useful results. Of the forces acting on the satellite, the Earth's gravity field is of very considerable interest, providing information on the planet's lateral density variations and, less directly, on the stress state and dynamic mantle processes. Tidal forces on the satellite can also be deduced and these provide estimates of global elastic and anelastic parameters of the Earth. The precise position measurements permit a determination of the Earth's rotational behaviour which gives a global measure of mass and momentum shifts in the oceans and atmosphere and within the Earth itself. These measurements, when refined, will also provide a measure of instantaneous plate tectonic motions. Satellite tracking methods, geodetic procedures and results are described in this review, with the emphasis placed on the geophysical significance of the results. The discussion is limited mainly to the solid Earth. Anticipated or possible future developments in satellite tracking, in geodetic theories and in satellite instrumentation are discussed within a framework of the geophysical rationale. Some of the objectives of satellite geodesy can also be attained by space techniques other than the use of artificial satellites. These include long-baseline interferometric observations of natural stellar or artificial lunar sources and laser range observations to the Moon. These methods are largely compatible with the satellite techniques rather than being directly competitive. The use of the satellite techniques for the Moon and planets is also discussed, again with the emphasis placed on the geophysical consequences of these measurements.

This review was received in April 1978.

Note added in proof. Since this review was written several surveys of, and proposals for, the use of satellites for geophysics have been published. These include: (i) Geodesy: Trends and Prospects 1978 (Washington, DC: Nat. Acad. Sci.), (ii) Applications of Space Technology to Crustal Dynamics and Earthquake Research 1978 (Washington, DC: NASA), (iii) Space Oceanography, Navigation and Geodynamics 1978 (Paris: ESA), (iv) Geodesy, Solid Earth and Ocean Physics, Res. Conf. 9 1979 (Columbus, Ohio: Ohio State University), (v) Artificial Satellites for Geodesy and Geodynamics 1979 (Athens: National Technical University).
## Contents

1. Introduction ........................................... 549
2. Satellite orbit mechanics .............................. 550
   2.1. A first-order theory of satellite motion .......... 550
   2.2. The spectrum of gravity perturbations .......... 555
   2.3. Further orbital perturbations .................. 558
   2.4. Refinements in orbital theories ................ 561
3. Observing Earth satellites ........................... 561
   3.1. Of satellites and observations ................ 561
   3.2. Laser ranging to satellites ................... 564
   3.3. Laser ranging to the Moon ..................... 566
   3.4. Long-baseline radio interferometry .............. 567
4. Estimating geodetic parameters ...................... 569
   4.1. Dynamic methods ................................ 569
   4.2. Geometric methods ................................ 572
   4.3. Surface gravity .................................. 573
   4.4. Results .......................................... 575
5. Further advances ..................................... 577
   5.1. Analyses ......................................... 577
   5.2. Drag-free satellites ............................ 579
   5.3. Satellite-to-satellite tracking ................ 580
   5.4. Satellite altimetry ............................ 581
   5.5. Satellite gradiometry .......................... 583
   5.6. A comparison of the alternatives ............... 584
6. Geophysical discussion .............................. 586
   6.1. Introduction ..................................... 586
   6.2. Earth's gravity field ............................ 588
   6.3. Tectonic motions ................................ 595
   6.4. Tides ............................................ 600
   6.5. Rotation ......................................... 612
7. Lunar and planetary problems ....................... 620
   7.1. Introduction ..................................... 620
   7.2. Lunar gravity field ............................. 620
   7.3. Lunar rotation ................................... 621
8. Conclusions .......................................... 623
References ............................................. 625
1. Introduction

In this review we are concerned with the application of geodetic techniques, with the emphasis on recent space methods, to some geophysical problems related to the structure of the Earth’s interior. Application of these methods to the Moon and other planets are also discussed. The objectives of geodesy can be grouped broadly into those concerned with immediate practical applications and those concerned with scientific or geophysical matters. By the former, I imply the establishment of geodetic control for mapping or navigation, for example, and this would include the ‘geopotential maps’ required for guiding satellites around our planet. The determination of the geodetic control on the Earth’s surface is usually viewed as a static problem, the geodetic markers defining the coordinate system remaining fixed unless upset by earthquakes or bulldozers. But as measuring accuracies improve, the Earth can no longer be considered as a rigid body. When subjected to forces it deforms, relative positions of the geodetic markers change and the gravitational field is modified. Such deformations are considered as mere nuisances to most geodesists but they are of great interest to geophysicists since they provide constraints on either the forces acting on and within the Earth or on the rheological properties of the planet. It is this geophysical aspect that is emphasised in this review.

The deformations may be a consequence of several factors. Lunar and solar gravitational forces, changing due to the periodic variation of the Earth–Moon–Sun geometry, raise tides on the solid Earth’s surface. Changes in the direction of the rotation axis, relative to the planet’s surface, introduce a variable centrifugal force and lead to deformation and further modification of the Earth’s rotational motion. Variable surface loads of ice, water or atmosphere also deform the planet but in less regular ways. Observations of the surface deformations and the rotation provide a measure of the Earth’s rheological response to more-or-less well-known forces. Inside the Earth, motions associated with mantle convection occur whose surface manifestations are seen in earthquakes and in the less dramatic but more pervasive plate tectonic motions. Observations of these surface motions may provide some insight into the mantle processes.

The Earth’s gravity field, measured by geodesists in order to determine the ‘shape’ of the planet, provides Earth scientists with one of their more useful geophysical observations in that any ‘anomalies’ in the gravity field are indicative of an anomalous density distribution within the Earth. These provide further insight into the mantle rheology and in the mechanisms causing plate tectonics.

A complete mathematical and physical treatment of these deformations requires a detailed knowledge of the physics of the Earth’s interior, a knowledge that—despite enormous progress in recent years—is still far from complete. We do not have direct access to the Earth’s interior; we can only postulate its constitution by observing its response to forces—for example, seismic waves, tidal, surface loading or rotational perturbations—and by chemical and physical considerations. Unfortunately, the interpretation of the observational data is seldom unambiguous and often contradictory. We observe at the Earth’s surface a variety of phenomena—gravity, heat flow, travel times and frequencies of seismic waves, and magnetic-field parameters—from which we wish to deduce properties of the Earth’s interior and to construct realistic models.
that best fit these data and the geochemical and petrological evidence. Wholly satisfactory Earth models still do not exist, but what has become increasingly evident is that the geodetic techniques can make important contributions to the construction of these models. In particular, the use of space technology to supplement the geodetic techniques has led to important improvements in the observational geophysical evidence.

Many geophysical observations, especially those based on geodetic measurements, describe properties of the Earth as it is today and do not provide much insight into the past evolution of this planet. This comes more from thermal and chemical considerations. But geophysical observations of the other terrestrial planets, including the Moon, may provide further insight into both the present structure and the past evolution of the Earth. These observations include gravity fields, tidal deformations and rotational characteristics, using principles much the same as the space techniques applied to the Earth.

2. Satellite orbit mechanics

2.1. A first-order theory of satellite motion

The main force acting on a satellite orbiting the Earth is the gravity field. Consider an Earth-fixed coordinate system $x$ with the $x_3$ axis aligned approximately parallel to the rotation axis, $x_1$ directed towards Greenwich and $x_2 = \pi/2$ to the east of $x_1$. The geocentric position of the satellite in this system can be defined by the geocentric distance $r$, latitude $\phi$ and longitude $\lambda$. The gravitational potential of the Earth in this system is conveniently expressed by a spherical harmonic expansion as:

$$U(r, \phi, \lambda) = \frac{GM}{r} \left[ 1 + \sum_{l=2}^{\infty} \sum_{m=0}^{l} \left( \frac{a_0}{r} \right)^l \left( C_{lm} \cos m\lambda + S_{lm} \sin m\lambda \right) P_{lm}(\sin \phi) \right].$$

(2.1)

$G$ is the gravitational constant, $M$ is the mass and $a_0$ is the mean equatorial radius of the Earth, $C_{lm}$, $S_{lm}$ are the Stokes coefficients that describe the Earth’s density distribution according to (see, for example, Heiskanen and Moritz 1967):

$$C_{lm} = \frac{1}{M a_0^l (l+m)!} \int_M r'^l P_{lm}(\sin \phi) \left( \frac{\cos m\lambda}{\sin m\lambda} \right) \, dM$$

$$S_{lm} = \frac{1}{M a_0^l (l+m)!} \int_M r'^l P_{lm}(\sin \phi) \left( \frac{\sin m\lambda}{\cos m\lambda} \right) \, dM$$

where $\delta_{0m} = 1$ if $m = 0$ and $\delta_{0m} = 0$ if $m \neq 0$. $dM$ is the mass element at $r'$, $\phi'$, $\lambda'$. The Legendre polynomials $P_{lm}(\sin \phi)$ are un-normalised, such that over a sphere of unit radius:

$$\int_0^1 P_{lm}(\sin \phi) \left( \frac{\sin m\lambda}{\cos m\lambda} \right)^2 \, ds = \frac{4\pi}{2l+1} \frac{(l+m)!}{(l-m)! (2-\delta_{0m})}$$

While the un-normalised harmonics are useful in many theoretical derivations, they are less suited to numerical calculations and fully normalised harmonics, defined so that the above integral is equal to $4\pi$, are often used. Such harmonics and the associated Stokes coefficients are denoted by bars. If the $x$ system is geocentric the summation over the degree $l$ ranges from 2 to infinity with the order $m$ ranging...
from 0 to $l$. Also, if $x_3$ lies close to a principal axis, $C_{31}$ and $S_{21}$ are very small. The Stokes coefficients closely relate to the inertia tensor of the Earth. For $l=2$, the important terms of the second-order inertia tensor $I_{ii}$ are:

\[
C_{20} = -\frac{I_{33} - \frac{1}{3}(I_{11} + I_{22})}{M_0} z^2
\]
\[
C_{22} = -\frac{I_{11} - I_{22}}{4M_0} z^2.
\]

For the Earth $C_{20} \approx 10^{-3}$ while higher-degree coefficients appear to decrease in magnitude according to a rule first noted by Kaula in 1963:

\[
\epsilon(C_{lm}, S_{lm}) \approx 10^{-5}/l^2.
\]

For Mars, the decrease is somewhat more rapid for the low-degree harmonics (Lambbeck 1978) while for the Moon it is somewhat less (Ferrari 1977).

The acceleration of the spacecraft subject to the potential (2.1) is determined by:

\[
\ddot{x} = \nabla U.
\]

The first approximation to $U$:

\[
U_0 = GM_0/r,
\]
results in the well-known Keplerian motion of the orbiter about its primary. This motion is conveniently expressed through the geometric Kepler elements defined in figure 1. Consider an inertial frame $X$ with, for convenience, $X_3 \parallel x_3$. In this
system x rotates diurnally with the angle between $X_1$ and $x_1$ denoted at any time by $\theta(t) = \theta$. The semi-major axis $a$ and the eccentricity $e$ define the shape and form of the orbit. The angle $\Omega$ in figure 1 is the angle between $X_1$ and the line of nodes, or the intersection of the orbital plane with the equatorial plane. It is referred to as the longitude of the ascending node. The angle $I$ specifies the inclination of the orbit onto the equator. The argument of latitude, $\omega$, is the angle between the line of nodes and the perigee of the orbit. Together, these three angles define the orientation in space of the orbital plane. For Keplerian motion, the five elements $a, e, I, \omega, \Omega$ remain unaltered and only the position of the orbit changes with time. This position can be defined in several ways, by the true anomaly $f$ (figure 1) or by the mean anomaly $M^* (t)$, the angle which the satellite would describe if it moved along a circular orbit at a uniform rate $n$, i.e. $M^*(t) = n(t - t_p)$ where $t_p$ is the time of perigee passage. The mean motion $n(=M^*)$ follows from Kepler’s third law $n^2a^3 = GM$. The true anomaly $f$ relates to $M^*$ via an intermediate parameter, the eccentric anomaly $E$, according to:

$$\tan \left( \frac{1}{2} f \right) = \left( \frac{1 + e}{1 - e} \right)^{1/2} \tan \left( \frac{1}{2} E \right)$$

(2.5(a))

and by Kepler’s equation:

$$E - e \sin E = M^*.$$  

(2.5(b))

The Keplerian motion sketched out above corresponds to a satellite moving around a spherically symmetric planet and, apart from permitting a determination of $GM$, it would be of little geophysical interest. For the actual motion the non-central part of the potential, $\Delta U = U - U_0$, results in perturbations in an acceleration $\delta \mathbf{a} = \nabla \Delta U$ superimposed upon the Keplerian motion which continues to serve as a useful first approximation. At any instant the position and velocity of the satellite is specified by six Keplerian elements, but at each successive instant the non-central field causes the elements to change. These changes are conveniently expressed with the aid of the Lagrange planetary equations of motion, in the form (see, for example, any text on celestial mechanics such as Brouwer and Clemence (1961)):

$$\frac{d\kappa_j}{dt} = B_i + A_{ij} \frac{\partial \Delta U}{\partial \kappa_j} \quad i, j = 1, \ldots, 6$$

(2.6(a))

where $\kappa_i$ refer to the six elements $a, e, I, \omega, \Omega, M^*$. All $B_i$ are zero except for $B_6 = n$. The $A_{ij}$ are given by:

$$A_{1,6} = 2/na$$

$$A_{2,4} = -(1 - e^2)^{1/2}/na^2e$$

$$A_{2,6} = (1 - e^2)/na^2e$$

$$A_{3,4} = -\cos I A_{3,5}$$

$$A_{3,5} = -[na^2(1 - e^2)^{1/2} \sin I]^{-1}$$

(2.6(b))

with $A_{ij} = -A_{ji}$. All other $A_{ij}$ are zero.
Before substituting $\Delta U$ into (2.6), the $r, \phi, \lambda$ in which $\Delta U$ is expressed have to be transformed into Keplerian elements and the angle $\theta$. This transformation is carried out in two parts: a purely geometric one from $r, \phi, \lambda$ to $r, \omega + f, \Omega - \theta, I$ and a second, circular to elliptic transformation of $r$ and $f$ to $a, e, M^*$. Kaula’s (1966) development is followed. The result is:

$$
\Delta U = \sum_{l=2}^{\infty} \sum_{m=0}^{l} \sum_{p=0}^{l} \sum_{q=-\infty}^{\infty} \Delta U_{lmpq}
$$

$$
\Delta U_{lmpq} = \frac{GM}{a} \left( \frac{a_e}{a} \right)^{l} F_{lmpq}(I) G_{lmpq}(e) \mathcal{J}_{lmpq}(\omega, M^*, \Omega, \theta).
$$

(2.7)

$F_{lmpq}(I)$ is a polynomial in $\sin I$ and the summation over $p$ is finite. $G_{lmpq}(e)$ is a polynomial in $e$ and has the property that $\partial G_{lmpq}(e)/\partial e = l \delta_l$. The summation of $q$ ranges from $-\infty$ to $+\infty$ but as orbital eccentricities of most close Earth satellites are small, the summation need be carried out over a few terms only, e.g. $q = 0, \pm 1, \pm 2$. The $F_{lmpq}(I)$ and $G_{lmpq}(e)$ are tabulated in Kaula (1966) for $l \leq 4$. The function $\mathcal{J}_{lmpq}$ is defined as:

$$
\mathcal{J}_{lmpq} = \begin{cases} 
C_{lm} & l - m \text{ even} \\
-S_{lm} & l - m \text{ odd}
\end{cases} \cos \gamma_{lmpq} + \begin{cases} 
S_{lm} & l - m \text { odd} \\
-C_{lm} & l - m \text { even}
\end{cases} \sin \gamma_{lmpq}
$$

(2.8)

with:

$$
\gamma_{lmpq} = (l - 2p)\omega + (l - 2p + q)M^* + m(\Omega - \theta).
$$

Substituting $\Delta U$ as expressed by (2.7) into the Lagrange equations gives the equations of motion in terms of the Keplerian elements. Integration of these equations is complicated by the fact that the elements $\kappa_l$ appearing on the right-hand side of these equations will be time-dependent. This is partly resolved by drawing upon certain properties of the formulation (2.7) and of the geopotential.

The frequency of the fluctuations in the gravitational force acting on the satellite are determined mainly by the factor $(l - 2p + q)M^*$ in the argument $\gamma_{lmpq}$ since $M^*$ remains the principal time-dependent element. Thus the most important perturbations in the satellite motion are those for which $l - 2p + q = 0$ since their periods are longer than one revolution of the satellite motion. Terms in $\Delta U$ containing the mean anomaly are less important after integration of the motion over intervals equal or greater than the orbital period.

Consider the most important term $C_{20}$ in $\Delta U$. For $q = 0$, the perturbing potential is:

$$
\Delta U_{2000} = \frac{GM}{a} \left( \frac{a_e}{a} \right)^{2} F_{2000}(e) G_{2000}(e) C_{20} \cos [(2 - 2p)\omega + M^*].
$$

For $(2 - 2p + q) = 0$ and $q = 0$, $p = 1$ and the main contribution to the perturbed motion comes from:

$$
\Delta U_{2010} = \frac{GM}{a} C_{20} \left( \frac{a_e}{a} \right)^{2} F_{201}(I) G_{210}(e).
$$

Substituting this into the Lagrange equations and with $F_{201}(I) = \frac{3}{4} \sin^2 i - \frac{1}{4}$,
The semi-major axis, eccentricity and inclination of the orbit undergo no secular change as could have been anticipated since the energy of the orbital motion is $G/2a$. The angular momentum of the orbital motion is $na^2(1-e^2)^{1/2}$ and the component of the angular momentum parallel to the rotation axis is $na^2(1-e^2)^{1/2} \cos I$. In a conservative system these quantities do not undergo any secular change. In contrast, the elements $\omega$ and $\Omega$ do experience a secular change at a rate that is proportional to $C_{20}$. Also, $n$ is modified from what it would be for the purely central motion. Integrating (2.9) shows that for an oblate spheroid, whose non-spherical potential is defined only by $C_{20}$, the motion of the satellite around the Earth can be considered as an ellipse of constant shape, size and inclination that precesses in its orbital plane at a rate $d\omega/dt$, and whose orbital plane itself precesses at a rate of $d\Omega/dt$ about the equator. For typical satellites with $a \approx 1.01a_0, e = 0.01$: 

$$d\omega/dt \approx 3.6 \left(5 \cos^2 I - 1 \right) \text{ degrees } d^{-1}$$

and:

$$d\Omega/dt \approx 6.7 \cos I \text{ degrees } d^{-1}.$$ 

These are the principal perturbations in the motion and this is confirmed by observations of close Earth satellites. Other, smaller, perturbations arise from (i) the terms in $\Delta U_{2pq}$ with $q \neq 0$ and $l - 2p + q \neq 0$, and (ii) the higher degree and order terms in the geopotential (2.1). These principal perturbations, together with the time-dependent $\theta$, all occur in the argument $\gamma_{lmpq}$. Then, to determine the perturbations due to any term $\Delta U_{lmpq}$, the equations of motion can be integrated by assuming that $a, e, I$ remain constant and that all the time-dependent parameters occur in $\gamma$. That is:

$$\Delta a_{lmpq} = \beta_a a F_{imp}(l)G_{ipq}(e)(l - 2p + q)\varphi_{lmpq} / \gamma_{lmpq}$$

$$\Delta e_{lmpq} = \frac{\beta_i}{e} F_{imp}(l)G_{ipq}(e)(1 - e^2)^{1/2}[(1 - e^2)^{1/2}(l - 2p + q) - (l - 2p)]\varphi_{lmpq} / \gamma_{lmpq}$$

$$\Delta I_{lmpq} = \frac{\beta_i F_{imp}(l)}{(1 - e^2)^{1/2} \sin I} G_{ipq}(e)[(l - 2p) \cos I - m] \varphi_{lmpq} / \gamma_{lmpq}$$

$$\Delta \omega_{lmpq} = \beta_i[(1 - e^2)^{1/2}e^{-1}F_{imp}(l)\partial G_{ipq}(e)/\partial e]$$

$$- \cot (1 - e^2)^{-1/2}(\partial F_{imp}(l)/\partial I)G_{ipq}(e) \varphi_{lmpq} / \gamma_{lmpq}$$

$$\Delta \Omega_{lmpq} = \frac{\beta_i(\partial F_{imp}(l)/\partial I)}{(1 - e^2)^{1/2} \sin I} G_{ipq}(e) \varphi_{lmpq} / \gamma_{lmpq}$$

$$\Delta M_{lmpq} = \beta_i[(1 - e^2)^{1/2}e^{-1}(\partial G_{ipq}(e)/\partial e)]$$

$$+ 2(l + 1)G_{ipq}(e)F_{imp}(l)\varphi_{lmpq} / \gamma_{lmpq}$$

$$G_{210}(e) = (1 - e^2)^{-3/2}.$$
with:
\[
\beta_l = \frac{GM}{na^3} \left( \frac{a_e}{a} \right)^l.
\]

In these expressions:
\[
\mathcal{M} = \int \mathcal{M} \, dy.
\]

To complete the expression for \( M^* \) we require an estimate of \( n \, dt \) (equation (2.6(a))). For \( l - 2p + q \neq 0 \), \( \Delta t_{mpq} \neq 0 \) and, with Kepler's third law:
\[
\Delta t_{mpq} = -\frac{3}{2} \frac{n}{a} \Delta a_{mpq}.
\]
Hence:
\[
\int n \, dt = -\frac{3}{2} \frac{GM}{a} \left( \frac{a_e}{a} \right)^l F_{imp}(I) G_{pq}(e) \left( \frac{l - 2p + q}{\gamma_{mpq}} \right) \mathcal{M}_{mpq}
\]
and with (2.10) the total perturbation in \( M^* \) is:
\[
\Delta M_{mpq} = \beta_l \left[ -\frac{(1 - e^2)^{1/2}}{e} \frac{\partial G_{pq}(e)}{\partial e} + 2(l + 1)G_{pq}(e) \right] F_{imp}(I) \mathcal{M}_{mpq}^{*} / \gamma_{mpq}.
\]

Equations (2.10) represent a first-order theory for the satellite motion in the case where \( C_{20} \) is the dominant term in the gravitational potential. This is so for Earth and Mars but not for Venus and the Moon and in these latter cases the theory becomes more complex.

2.2. The spectrum of gravity perturbations

For the spherical harmonic expansion (2.1) of the potential to be useful the series must converge relatively quickly so that the main features of the field can be readily identified. For the orbit mechanics it is important that the influence of the Stokes coefficients on the orbit converges rapidly so that the algebra remains tolerable. The latter is ensured by a number of factors. First, the factor \( (a_e/a)^l \) decreases with increasing \( l \), albeit slowly for close satellites. Second, the Stokes coefficients appear to decrease slowly with increasing degree according to (2.3). For orbits with \( a \simeq 1 \cdot 10^6 \) the convergence due to these two factors is quite fast. For example, for \( l = 4 \), \( (a_e/a)^4C_4, m \simeq 4 \times 10^{-7} \); for \( l = 12 \), \( (a_e/a)^{12}C_{12}, m \simeq 4 \times 10^{-9} \); for \( l = 20 \), \( (a_e/a)^{20}C_{20}, m \simeq 4 \times 10^{-9} \) and for \( l = 32 \), \( (a_e/a)^{32}C_{32}, m \simeq 4 \times 10^{-10} \). The order of the harmonics also controls the amplitude of the perturbations through the denominator of equations (2.10). Since \( \phi, \Omega < \theta \) the frequency of the perturbations is given by:
\[
(l - 2p + q)n - m\theta
\]
and the dominant perturbations occur when \( l - 2p + q = 0 \) for those satellites for which \( n > m\theta \). Then, the larger the order the higher the frequency of the predominant perturbations and the smaller the amplitude of the perturbation. For \( n < m\theta \) the dominant perturbations will occur for \( l - 2p + q = 1 \). The special case when \( (l - 2p + q)n = m\theta \) is discussed below.

2.2.1. Zonal harmonics. The dominant orbit perturbations are the secular changes
(2.9) in \( \omega, \Omega \) and \( M^* \) due to \( C_{20} \). Consider now the harmonic \( C_{40} \). The main perturbations for small eccentricity orbits occur for \( q = 0 \) and \( l - 2p + q = 0 \), or \( lmpq = 4020 \). Then:

\[
\Delta U_{4020} = \frac{GM}{a} \left( \frac{a_2}{a} \right)^4 F_{4020}(I)G_{420}(\epsilon)C_{40}.
\]

As for the main perturbations due to \( C_{20} \), \( a, e, I \) are not perturbed but \( \omega, \Omega \) and \( M^* \) are subject to secular rates proportional to \( C_{40} \). Likewise, the other even zonal harmonics give rise to secular perturbations in \( \omega, \Omega \) and \( M^* \) and the observed secular rate is the combined effect of a number of terms. Because of the factor \( (a_e/a)^2C_{10} \), the higher-degree terms generally contribute less but for some satellites, tracked with 10–20 m precision, zonal harmonics up to degree 36 are required to completely describe the secular rates (Gaposchkin 1973). The even zonal harmonics also introduce long-period perturbations. Consider the harmonic \( C_{4, 0} \). For \( l - 2p + q = 0 \) the combination of indices \( lmpq = 4032 \) and \( 401 - 2 \) result in perturbations with frequency \( \pm 2\omega \) where \( \omega \) is given to a first approximation by (2.9). As \( q = \pm 2 \) these perturbations are proportional to \( e^2 \) and are only significant for eccentric orbits. Other even zonal harmonics, except for \( C_{20} \), also introduce perturbations with this frequency. The exception is due to the fact that for the relevant indices \( G_{402}(\epsilon) \) vanishes.

The odd zonal harmonics do not give rise to secular perturbations in the elements since there is no combination of indices for which both \( l - 2p + q \) and \( l - 2p \) vanish. Major perturbations do occur. For example, for \( C_{3, 0} \), \( l - 2p + q = 0 \) if \( q = \pm 1 \). Then, with either \( lmpq = 3021 \) or \( 301 - 1 \) the frequencies of the perturbations are \( \pm \omega \) and their amplitudes are proportional to \( e \). For \( q = \pm 3 \) and \( lmpq = 3033 \) or \( 300 - 3 \) perturbations occur with frequencies \( \pm 3\omega \) but these are smaller by a factor of \( e^2 \) than those of frequency \( \omega \).

Thus, for a satellite orbiting around a rotational symmetric planet, the Keplerian elements \( \omega, \Omega, M^* \) are subject to secular variations due to the combined action of the even zonal changes. Upon this, long-period perturbations are superimposed, with frequencies of \( \omega, 2\omega, 3\omega, \ldots \), that are governed by the odd and even zonal harmonics. The eccentricity and inclination are also periodically perturbed and only the semi-major axis is not subject to secular or long-period fluctuations since, upon substituting the potential (2.7) into the first of the Lagrange equations (2.6), a factor \( (l - 2p + q) \) appears in the numerator. Superimposed upon these effects are short-period perturbations described by those indices for which \( l - 2p + q \neq 0 \). From observations of different elements of satellites with different orbit characteristics, the effects of the individual zonal harmonics of the gravity field can be separated and the coefficients estimated (Kozai 1969, King-Hele and Cook 1974).

**2.2.2. Resonances.** The first-order theory breaks down in the event that the denominators in (2.9) approach zero. Since \( n \gg \omega, \Omega \) the frequency of the perturbations is determined mainly by (2.11) and resonance occurs when:

\[
n = \frac{m\theta}{l - 2p + q}
\]

when the mean motion of the satellite is equal to the order of the tesseral harmonic, \( l - 2p + q = 1 \), or equal to \( 2m \), \( (l - 2p + q) = 2 \), etc. Physically, the resonance occurs
when successive ground paths of the satellite are exactly separated by an interval equal to the wavelength of the geopotential harmonic. After a certain number of revolutions of the satellite around the Earth, the ground track sequence repeats itself exactly and the satellite's motion is perturbed in an identical manner, enhancing the earlier perturbations. The perturbations become of arbitrarily long period and large amplitude and the first-order theory breaks down. No general theory exists at present for handling the resonances of the close Earth satellites. Special theories, for a single resonant term, have been developed and used in the study of geosynchronous orbits (Morando 1963, Allan 1973), but theories for closer satellites are complicated by the fact that the satellite will be resonant with many harmonics of order \( m \) and that interactions with other perturbations, those due to the zonal harmonics and air drag in particular, have been taken into account. For this reason only near-resonant cases, where \( \dot{y}_{lmpq} \) becomes small rather than zero, are used for studies of the gravity field so that the linear theories remain adequate (e.g. Gaposchkin 1973). Most close Earth satellites orbit the Earth from about 10 \((a\approx 9100 \text{ km})\) to 15 \((a\approx 6900 \text{ km})\) revolutions per day and most of the resonances will be for the harmonics of order 10–15 (see, for example, King-Hele et al 1974, Wagner 1977). The higher the order \( m \), the greater the mean motion \( n \) and the lower the required orbit altitude for resonance to occur, but air drag sets a lower limit to the altitude at which satellites can be usefully applied to gravity field studies and only resonances with harmonics of order up to 15 have been studied. To determine lower-order harmonics requires higher orbits and the satellite will be sensitive to only a few coefficients of different degree unless the resonance is sharp. The satellite Midas 4, for example, with a perigee altitude of about 3500 km is in shallow resonance with harmonics of order 8, and is sensitive only to the degree 8 coefficients. Hence only the \( C_8, s, S_8, s \) can be estimated. Occasionally, resonances with \( m=2n \) are observed (Anderle and Smith 1968). Table 1 summarises the characteristics of some near resonances encountered in geodetically useful satellites.

### 2.2.3. Tesseral harmonics.

Orbital perturbations due to the tesseral harmonics \((m\neq 0)\) exhibit a complex spectrum. The main effects are due to the low-order

<table>
<thead>
<tr>
<th>Satellite</th>
<th>Inclination ((\text{deg}))</th>
<th>Semi-major axis ((\text{km}))</th>
<th>Resonant with order ( m )</th>
<th>Resonant period ((\text{d}))</th>
</tr>
</thead>
<tbody>
<tr>
<td>Midas 4 ((6102801))</td>
<td>95</td>
<td>10005</td>
<td>9</td>
<td>2.90</td>
</tr>
<tr>
<td>GEOS 1 ((6508901))</td>
<td>59</td>
<td>8074</td>
<td>12</td>
<td>7.2</td>
</tr>
<tr>
<td>DID ((6701401))</td>
<td>39</td>
<td>7337</td>
<td>13</td>
<td>9.4</td>
</tr>
<tr>
<td>DIC ((6701101))</td>
<td>40</td>
<td>7366</td>
<td>13</td>
<td>1.6</td>
</tr>
<tr>
<td>BEB ((6406401))</td>
<td>80</td>
<td>7362</td>
<td>14</td>
<td>2.9</td>
</tr>
</tbody>
</table>

Table 1. Summary of some shallow resonances occurring in satellite orbits due to the gravity field coefficients of order \( m \).
terms and group around frequencies of $m\theta$. Others group about frequencies that are integral multiples of $n$. As for the zonal harmonics, a number of Stokes coefficients yield perturbations with the same set of frequencies. For example, the harmonics 4, 1; 6, 1; 8, 1; ... all give rise to principal perturbations with frequencies $\Omega - \dot{\theta}$ but whose amplitudes tend to decrease because of the $(a/eatal)C_{lm}$ factor. Separation of the coefficients becomes possible only if the motions of satellites in very different orbits are analysed. Figure 2 indicates a typical perturbation spectrum. Figure 3 illustrates schematically the terms in the geopotential that can be estimated from the analysis of orbital perturbations. Harmonics of order 11–12 are usually estimated from the shallow resonance effects up to high degree, although resonances are possible for all orders.

Low-order Stokes coefficient can be reliably estimated as they result in perturbations with periods that exceed a few hours, but the associated higher-degree coefficients become increasingly difficult to estimate because of the $(a/eatal)C_{lm}$ factor. The lines A in figure 3 indicate approximate limits of the sensitivity of satellites to the potential for 20 m accuracy in tracking, and the shaded region between these lines indicates those coefficients that cannot be estimated. The lines B correspond to 10 m accuracy tracking and lines C to 1 m accuracy tracking. Other techniques, either spatial or terrestrial, will be required to estimate the coefficients in the hatched region.

2.3. Further orbital perturbations

The Earth’s gravitational field is not the only force acting on the satellite that

![Figure 2. Amplitude spectrum of perturbations in the mean anomaly of the satellite DID due to non-resonant tesseral harmonics.](image)
Figure 3. Schematic representation of the non-zonal Stokes coefficients that cannot be determined from satellite orbit perturbation analyses. The coefficients with degree and order \( l, m \) lying in the shaded area between the lines A and A' cannot be determined if the tracking accuracy is lower than 10 m. Those coefficients below A can be determined from short-period perturbation analyses while those above A' follow from shallow resonance studies. The curves BB' represent the limits for 1 m accuracy tracking and CC' for 20 cm accuracy.

perturbs its motion. Table 2 summarises some typical magnitudes of additional effects. Direct lunar and solar attractions are important. The lunar gravitational potential at the satellite position \( \mathbf{r} \) is:

\[
U_M = GM_M \left( \frac{1}{r_M - \mathbf{r}} - \frac{\mathbf{r} \cdot \mathbf{r}_M}{r_M^3} \right)
\]  

(2.12)

where \( r_M \) is the geocentric position of the Moon. The orbital perturbations follow by transforming \( \mathbf{r} \) into the Keplerian satellite elements similar to those used to obtain equation (2.7) and by making a comparable transformation of \( r_M \) into lunar elements. Substituting this transformed potential into the Lagrange equations, and integrating them, gives the sought-after perturbations in the satellite motion. The motions of the Sun and Moon are sufficiently well known for the calculation of these perturbations not to be very troublesome. The lunar and solar gravitational forces also act on the Earth, deform it, and gives rise to a tidal force that acts on the satellite. These changes can be adequately modelled once the parameters defining the response of the planet are known (see §6.4). In precise orbit determinations the solid, ocean and atmospheric tides must be considered. A related orbital perturbation comes from the fact that in the inertial reference frame the Earth's gravity field is not constant. In particular, precession and nutation shift the orientation of the Earth's bulge and \( C_{90} \) will be time-dependent (Lambeck 1973, Kozai and Kinoshita 1973).
Table 2. Characteristic orders of magnitude of perturbations in the motion of an Earth satellite due to different forces acting on the spacecraft. The data are mainly from Gaposchkin (1973).

<table>
<thead>
<tr>
<th>Force</th>
<th>Short period, &lt;1 d (m)</th>
<th>Intermediate period, 1-30 d (m)</th>
<th>Long period, &gt;30 d (m)</th>
<th>Secular</th>
</tr>
</thead>
<tbody>
<tr>
<td>$C_{z,0}$</td>
<td>$2 \cdot 5 \times 10^3$</td>
<td>$10^{-6}$</td>
<td>Yes</td>
<td></td>
</tr>
<tr>
<td>$C_{(2n+1),0}$</td>
<td>50</td>
<td>$10^{-8}$</td>
<td>No</td>
<td></td>
</tr>
<tr>
<td>$C_{2n,0}$</td>
<td>50</td>
<td></td>
<td>Yes</td>
<td></td>
</tr>
<tr>
<td>Tesseral harmonics</td>
<td>200</td>
<td>$2 \times 10^3$ resonance</td>
<td>$2 \times 10^3$ resonances</td>
<td>No</td>
</tr>
<tr>
<td>Solar attraction</td>
<td>$\sim 1$</td>
<td>800</td>
<td>Yes</td>
<td></td>
</tr>
<tr>
<td>Lunar attraction</td>
<td>$\sim 1$</td>
<td>120</td>
<td>Yes</td>
<td></td>
</tr>
<tr>
<td>Air drag</td>
<td>$&lt; 1$</td>
<td></td>
<td>Yes</td>
<td></td>
</tr>
<tr>
<td>Radiation pressure</td>
<td>$&lt; 1$</td>
<td></td>
<td>Yes</td>
<td></td>
</tr>
<tr>
<td>Body tides</td>
<td>20</td>
<td>100</td>
<td>No</td>
<td></td>
</tr>
<tr>
<td>Ocean tides</td>
<td>2</td>
<td>10</td>
<td>No</td>
<td></td>
</tr>
</tbody>
</table>

Surface forces on the satellites are often important. They include atmospheric drag and solar radiation pressure. To compute the drag forces, the atmospheric density at the satellite must be known. This is now determined with some confidence from the satellite motion itself. This is possible because the nature of the drag perturbations differs from the geopotential perturbations. The drag force is (Sterne 1959):

$$F_d = -\frac{1}{2} C_D \frac{A}{m} \rho v^2$$

where the drag coefficient $C_D$ is a function of the satellite shape and the nature of reflection of the air molecules. For spherical satellites $C_D \approx 2$. $A$ is the cross-sectional area and $m$ is the mass of the satellite, $\rho$ is the atmospheric density and $v$ is the satellite velocity relative to the atmosphere. For a geostationary atmosphere $F_d$ is tangential to the orbit. To compute the effect of this force on the satellite motion it is convenient to use an alternate form of the equations of motion (2.6), in which the force function $\Delta U$ is expressed in orthogonal components with one component tangential to the orbit (see, for example, Brouwer and Clemence 1961). The semi-major axis and eccentricity change secularly at a rate proportional to $C_D(A/m)\rho$ as well as containing periodic perturbations. Thus, as there are no gravitational secular effects on $a$ and $e$ any observed secular change can be attributed to the drag forces and the product $C_D(A/m)\rho$ estimated. Corrections are then made for the periodic drag perturbations in $a$, $e$, $\omega$, and $M^\ast$. As the orbit is contracting at a rate $\dot{a}$, the orbital mean motion is also changing at a rate $\dot{\Omega} = -\frac{3}{2}(n/a)\dot{a}$ and $M^\ast$ contains a quadratic variation with time. This is usually the largest observable consequence of the drag forces. For low perigee orbits or for long orbital arcs this outlined theory may be inadequate and in this case empirical parameters may be introduced to account for the drag. From the analyses of the drag perturbations the atmospheric density itself can be deduced (Jacchia 1977, Barlier et al 1978). King-Hele (1964) has investigated the case when the upper atmosphere rotates relative to the Earth (see also Walker 1975, King-Hele and Walker 1976). More direct measurements of the drag forces are now possible with on-board accelerometers (see §5.2).
The radiation forces acting on the satellite include the direct solar radiation and the radiation reflected from the Earth. The computation of the perturbations requires a knowledge of the solar constant, the nature of the reflection of the radiation falling onto the satellite, the albedo of the Earth, the position of the Earth's shadow, and the spacecraft's orientation. The shadow effect and the Earth's variable albedo makes any analytical development awkward and numerical integration methods or numerical harmonic developments are usually employed. Lala (1972), Lala and Schnal (1969) and Schnal (1975) discuss aspects of the orbital theory. Short of having a precise and convenient theory plus a detailed knowledge of the physical parameters, the best solution to solving the surface force problem is to use satellites with small surface area to mass ratios, and with spherically symmetric reflective properties. Two such satellites have been launched specifically for geodetic studies, LAGEOS by the National Aeronautics and Space Administration, and STARLETTE by the French Space Agency.

2.4. Refinements in orbital theories

The analytical approach outlined above offers great insight into the nature of the orbital perturbations but does have some drawbacks when a very precise theory is sought. The outlined theory provides only a first-order linear approximation to the equations of motion. This is inadequate for the oblateness perturbations since $C_{20}$ is about 1000 times greater than the other Stokes coefficients and, if the first-order linear perturbations are adequate for all coefficients except $C_{20}$, the consequences of $C_{20}$ should be described as non-linear perturbations to order $(C_{20})^2$. Linear perturbations of the elements on the right-hand side of (2.4) due to $C_{20}$ should also be included in the theory. Numerous such high-order perturbation solutions have been proposed and Gaposchkin (1973) discusses the more successful ones. The linear theory, plus the second-order oblateness theory, are sufficient to describe the motions of most close Earth satellites with a precision of about 1 m, or 1 in $10^7$, but if a theory accurate to 1 in $10^8$ is sought, commensurable with the highest tracking accuracy now available, further refinements, involving much algebra, are required. In particular, the non-linear interactions between $C_{20}$ and the other zonal harmonics must be included as well as the interactions of the zonal terms with the resonances, lunar and solar perturbations, drag and radiation pressure and the interactions between these other perturbations. Much of the tedious algebra can be carried out on computers but a precise theory, approaching 10 cm, is not yet available. In particular the theory for the long-period perturbations remains incomplete. Berger (1975) has developed a theory complete to order $(C_{3,0})^3$, $C_{2,0}C_{l,0}$, and $C_{l,0}C_{l',0}/(C_{2,0})^3$. A comparison with numerical integration indicates a precision of 10–15 cm for representative geodetic satellites (Berger and Walch 1977).

3. Observing Earth satellites

3.1. Of satellites and observations

Since the launching of the first satellite in October 1957 enormous progress has been made in the ability to precisely track spacecraft. Laser tracking of satellites and the Moon now routinely reaches precisions of a few tens of centimetres compared with the first visual observations of the Sputnik satellites, which were only accurate.
to a few kilometres. Radio observations were even less precise; with the low frequencies used for the radio tracking of Sputnik 1, ionospheric refraction effects were so large that the signal could be recorded long after the satellite disappeared over the observer's horizon. Observations of satellites fall into three categories: optical, electronic or radar. The first are either visual or camera observations. Visual observations proved to be useful in the early stages of the space programme for keeping track of non-instrumented satellites. King-Hele (1966) has discussed these observations and their use in studies of atmospheric densities at altitudes above about 200 km.

Camera tracking—by observing the satellite against a star background to give a direction to the spacecraft relative to the stars—has been one of the simplest and most widely used methods for keeping track of a multitude of objects. Normally, these observations rely on sunlight reflected from the spacecraft to render them visible and tracking is limited to that small part of the orbit when the satellite is illuminated by the Sun and the observer is in darkness, weather permitting. Some satellites, the geodetic satellites GEOS 1 and 2 for example, carried their own light source in the form of high-intensity, short-duration flashes.

Radio tracking includes interferometric, range and velocity observations. They avoid the visibility limitations but require satellites that either emit continuous signals or that can be activated from the ground. Particularly in the first few years of the Space Age there were an insufficient number of such objects in orbit and many geodetic and atmospheric studies relied heavily on the camera tracking of passive satellites. Radar tracking avoids the drawbacks of both optical and radio tracking and is the most self-sufficient form of tracking. A system operated by the US Navy, the Naval Space Surveillance System, and consisting of a 'fence' of receiving stations across the United States has tracked objects as high as 18,000 km. Its relatively low accuracy, however, makes the observations of only limited scientific interest although they have been used in studies of the resonant gravity harmonics. Optical radar tracking, by illuminating retro-reflectors carried on a satellite with a laser beam, is one of the most precise methods for tracking a satellite. Special reflector-carrying satellites are required and there are now some 11 in orbit. Two of these, STARLETTE and LAGEOS, have been specifically designed for high-precision laser tracking. Otherwise the satellites are entirely passive. Day-time measurements are possible although cloud cover still reduces the time available for observation.

Routine tracking of the early satellites launched by the US National Aeronautics and Space Administration was carried out with the network of 'Minitrack' radio interferometer stations operating near 140 MHz. Directional accuracies were of the order of 20 arcsec although ionospheric propagation perturbations were often much greater than this. Nevertheless the data collected with this network were sufficiently precise to determine, even in 1958, the Earth's oblateness and third- and fourth-degree zonal harmonics of the Earth's gravity field with greater accuracies than were possible from surface gravity observations (O'Keefe et al 1959). The network has also provided observations that contributed to the determination of resonant gravity harmonics and to upper atmospheric densities.

The Baker-Nunn camera network of the Smithsonian Astrophysical Observatory was operational soon after the launching of the first US satellite. By photographing the satellite against a star background, directional accuracies of about 2 arcsec are achieved. Because of the large 50 cm aperture, and the camera's ability to follow
the satellite along its orbit, very small objects down to twelfth magnitude could be recorded. The many thousands of observations collected by this network have provided one of the most important data sets for geodetic and atmospheric studies. Early scientific analyses of these observations by Y Kozai and I G Izsak indicated that the Earth's gravity field was considerably more variable than previously anticipated by many geodesists. Subsequent studies at the Smithsonian culminated in the gravity field models of 1966 (Lundquist and Veis 1966) and 1969 (Gaposchkin and Lambeck 1971) in which the basic data base consisted of the Baker–Nunn observations. At the same time these data led to important improvements in the atmospheric density models through the study of drag perturbations (e.g. Jacchia 1967, 1977). The Baker–Nunn camera network has now been largely replaced by a smaller network of laser tracking stations.

Another camera tracking network that achieved some success is the Wild BC-4 network operated by the United States Coast and Geodetic Survey throughout the 1960s for establishing a geodetic control for North America and for the world by simultaneous observations of the large balloon satellites ECHO 1 and 2 and PAGEOS (Schmidt 1974). Cameras operated at many individual observatories, such as the Hewitt camera of the Royal Radar Establishment, provided useful data when used in conjunction with the larger Smithsonian network of Baker–Nunn cameras.

Precise Doppler tracking of satellites was introduced with the United States Navy's navigation satellite programme. Initiated in 1960 with the launching of the first satellite NAV 1B in this programme, it has continued successfully with ever-increasing accuracy. Some of the satellites used today have been operating successfully for ten years. The early Doppler systems gave positional accuracies of several tens of metres while the more recent satellites, together with improved tracking equipment, give precisions of a few metres in position. By transmitting a continuous signal at two distinctly different and high frequencies, of the order of 160 MHz and 324 MHz, corrections for ionospheric refraction can be made with sufficient accuracy. The Navy navigation satellite network consists of a large number of permanent tracking stations, some 20 in 1973, in addition to mobile stations. This large number of stations, together with the all-weather capability of the tracking system, means that frequent observations of the satellite are possible. It is this ability to provide uniform orbital tracking coverage that has been mainly responsible for the success of the system. The early outstanding achievement of the analysis of the Doppler data was the solution of the gravity field published by Guier and Newton (1965). Many of the subsequent studies of the gravity field have, unfortunately, not been published. The system has also provided accurate station positions quickly, a precision of a few metres within a few days now being routinely achieved with portable satellite receivers (Anderle 1974). Another success of this system has been the determination of polar motion continuously for about 10 yr with an accuracy and resolution that is at least as good as that obtained by traditional astronomical observations (Anderle and Beuglass 1970, Anderle, 1975).

In recent years the US Navy has developed the Navy Technology satellites, previously known by the name of Timation, which can be tracked by electronic ranging with a precision of a few metres. These satellites are in very high orbits, of the order of 20 000–14 000 km, and are therefore visible from any one station for long periods at a time and they permit a rapid determination of station position with a precision of a few metres. The satellites also carry cube corner reflectors, permitting a direct comparison of laser and radio range measurements, but because
of their high altitudes few of the present satellite ranging systems can actually obtain echoes with sufficient energy to give accurate range measurements.

Laser range observation was initiated in 1964 with the launching of the Explorer 22 (BE-B) satellite carrying cube corner reflectors. By 1968 six reflector-carrying satellites were in orbit, including two (DID and DIC) launched by the Centre National d'Etudes Spatiales in 1967. Early laser tracking systems were developed at the Smithsonian Astrophysical Observatory, at the Goddard Space Flight Center and at the Centre National d'Etudes Spatiales (Plotkin 1965, Bivas and Moreal-Courtois 1966, Anderson et al 1966). Observations accurate to about 5 m or better were made on an irregular basis during the latter part of the 1960s and the data were generally insufficient to contribute significantly to the geodetic results. In the gravity field model of Gaposchkin and Lambeck (1971), for example, about 2200 laser observations to 6 satellites from 4 stations were used, compared with 57 000 Baker-Nunn observations to 21 satellites from 24 stations. Over the last few years the laser tracking systems have been significantly improved and they have largely replaced the camera systems (Pearlman et al 1977).

3.2. Laser ranging to satellites

The principle of laser ranging consists of the measurement of the time it takes for a short pulse to travel the path from the station to the satellite and back. In the present satellite ranging systems the pulse is usually generated by a Q-switched ruby laser and the beam is reflected back to the station by an array of cube corner reflectors mounted on the spacecraft. A travel time accuracy of 1 ns gives a one-way distance accurate to 15 cm. In the tracking of close Earth satellites, the number of photons received at the receiver is usually quite large and this implies that the pulse length need not be as short as 1 ns in order to obtain nanosecond accuracy, provided that the pulse retains its shape. In the laser system presently employed by the Smithsonian Astrophysical Observatory, the pulse length is of the order of 25 ns although this will be reduced to about 6 ns in the future (Pearlman et al 1977).

The accuracy of the travel time measurements depends critically upon the calibration of the delay times within the transmitter and receiver. This is usually determined by ranging over a known distance to a nearby target. The calibration can be expected to be a function of the signal strength and it is important that the output and return calibration signals are comparable with the actual signals obtained when ranging to satellites. The usual procedure is to calibrate the laser before and after each satellite pass. The Smithsonian lasers have a calibration stability of the order of a nanosecond (Pearlman et al 1977). Distortion of the pulse shape introduced by the transmitter and photoreceiver noise will also contribute a few nanoseconds to the error in the travel time observations. The pulse deformation errors can be reduced by using a pulse-processing system which permits the travel time to be referred to pulse centres rather than to threshold points. A range gate is usually employed to protect against a triggering of the receiver by sky background and electronic noise.

The Smithsonian laser system operates at a repetition rate of about one pulse per 8 s and the average scatter of the travel time residuals during a single satellite pass, relative to a mean orbit, is of the order of a few nanoseconds (Pearlman et al 1977). The Goddard Space Flight Center lasers have a repetition rate of about
1 pulse per second and also exhibit travel time residuals of about 2 ns. Systematic errors may exceed these levels and they can probably be determined only by comparing results from co-located instruments.

A number of other factors contribute to the overall accuracy. Atmospheric refraction, epoch timing and retro-reflector characteristics must be considered. The atmospheric refraction correction does not appear to present serious problems at the few centimetre accuracy level (Hopfield 1969).

Timing of the epoch of observation must be of an accuracy that is comparable with the range measurements. The average velocity of a satellite is of the order of 7 km s\(^{-1}\) so that absolute timing of about 10 \(\mu\)s is required to fix the satellite in its orbit with, say, 10 cm accuracy. Time at the stations can be kept to this accuracy by crystal oscillators that are controlled, through frequency and phase comparisons, against stable VLF transmissions. Absolute time at each station is controlled by periodic comparisons with portable clocks. The US Navy Navigation Technology satellites can now be used to maintain absolute time at the stations to within 1 \(\mu\)s.

The satellite range measurements consist of the distance between an electronic centre of the transmitter-receiver and some point on the satellite which will not correspond to the centre of mass of the satellite. A reflector array correction is therefore required and this depends upon the array characteristics and upon the angle of incidence of the laser pulse. The former is known from the satellite design characteristics while the latter depends on the manner of spacecraft stabilisation. Such corrections are given in Gaposchkin (1973). For the Explorer satellites this correction is good to about 10 cm but it is an order of magnitude better for the special laser reflector satellites such as STARLETTE and LAGEOS. The former was launched in 1975 by the Centre National d'Etudes Spatiales in a 800 km altitude, 50° inclination orbit while LAGEOS was launched in a 5900 km, 110° inclination orbit by the National Aeronautics and Space Administration in 1976.

STARLETTE consists of a 24 cm diameter, small area-to-mass ratio sphere, covered by 60 uniformly distributed cube corner reflectors. The uniform reflectivity means that the measurements can be reduced with high accuracy to the spacecraft centre and that air drag and radiation pressure perturbations are independent of the spacecraft orientation. LAGEOS is a 411 kg, 60 cm diameter sphere, covered with 426 reflectors.

Laser tracking of satellites is carried out by several organisations. The Smithsonian Astrophysical Observatory has four stations, in Arizona, Peru, Brazil and Australia. Two lasers, operated by cooperating agencies in Japan and Greece, contribute to the Smithsonian network. These stations yield range data that are accurate to 1 m or better and routinely observe the laser-reflector-carrying satellites. Modifications to the stations are in progress to improve the accuracies to about 20 cm.

The Goddard Space Flight Center operates a number of mobile laser systems within the United States with accuracies of the order of 20–30 cm (Smith et al 1977). In addition, new laser systems have been constructed in France, the Netherlands and West Germany, all with design accuracies of a few tens of centimetres. Some laser range instruments with 1 m accuracies have been built at the Ondrejov Observatory in Czechoslovakia (Sehnal 1973, Lala and Prokes 1976) and used in Eastern Europe. All of these lasers use Q-switched ruby lasers except for the West German instrument which is a mode-locked neodymium YAG laser marketed by the Sylvania Company of California.
3.3. Laser ranging to the Moon

The possibility of ranging to the Moon with a laser was first demonstrated by Smullin and Fiocco in 1962 although the possibility of doing this with pulsed searchlight illumination from the Earth had already been investigated in the late 1950s. Smullin and Fiocco (1962) used a long pulse to record the reflection from the Moon's surface but while the result was of little scientific interest it did lead to the suggestion that retro-reflectors be placed on the Moon (Alley et al 1965). The scientific benefits from such an exercise were already recognised at this time as being of importance in lunar and terrestrial physics and in the study of gravitation and relativity. In the Russian literature similar suggestions appeared (Kokurin et al 1966). In 1966 NASA endorsed a programme that led to the emplacement, in July 1969, of cube corner reflectors on the Moon as part of the Apollo 11 Lunar Surface Experiments Package. The subsequent Apollo 14 and 15 missions placed additional reflectors on the Moon as did two Soviet missions, Lunakhod 1 and 2. The Apollo 11 and 14 reflector arrays consist of a 46 cm square array of 100 solid fused silica reflectors of 3.8 cm in diameter. Apollo 15 carried a 300 reflector array. The Lunakhods carried French-made reflectors but only the Lunakhod 2 reflector has been seen.

The first laser returns were obtained in August and September of 1969 from a hastily modified McDonald Observatory in Texas. These initial returns were estimated to have an accuracy of 15 ns for the two-way travel time, or about 2.5 m in the one-way distance. Measurements taken in October 1969 were precise to about 50 cm in the one-way distance while data collected routinely since late 1971 have a typical accuracy of between 10 and 20 cm. (Bender et al (1973) discuss the history of the lunar laser ranging programme.) Early laser observations were reported from the Crimean Observatory in the USSR and from the Pic du Midi Observatory in France but little useful data have come from these stations since. New stations have been built in Australia† and Hawaii and other stations are under construction in Japan and France.

All the scientific results emanating from lunar laser ranging so far have been based on the McDonald Observatory data. These results include improved lunar libration parameters and initial results for the Earth's rotation. Williams (1977) reviews the present achievements.

The McDonald Observatory is a general purpose observatory and can be used for lunar laser ranging for only a few hours each day. The observing period is usually limited to observations when the Moon is at its highest point in the sky and about three hours before and about three hours after this point is reached. Each of these three runs consists of 50–300 laser measurements. The laser system is a Q-switched ruby laser with a few nanoseconds pulse length and a pulse repetition rate of a pulse every 3 s. The beam divergence of the transmitted beam is of the order of 10⁻⁶ sr. The effective aperture of the receiver exceeds 1 m². The return signal, of energy 10⁻¹⁹ J, contains very few photons. This means that the detection and timing techniques tend to be different for lunar laser systems than for satellite lasers. At the McDonald Observatory the travel time measurements are precise to 0.1 ns and the overall accuracy is of the order of a few nanoseconds. Comparisons of the measurement against orbital theories of the lunar motion indicate that, for a six-year period, the rms residuals are of the order of 40 cm (Williams 1977). This

† The Australian station at Orroral Valley, near Canberra, has successfully ranged to the Moon since late 1978.
is in part due to unmodelled or inadequately modelled errors in the orbital theory and confirms that the instrumental errors are unlikely to exceed 20 cm.

The laser ranging observatory constructed on the island of Mauii, Hawaii, is a mode-locked neodymium YAG laser system emitting pulses of 200 ps with an energy of 0.25 J at 5320 Å and at a rate of 3 s⁻¹. The system is intended to range to the Moon and to high reflector-carrying satellites such as LAGEOS and the Navigation Technology satellites with a precision of 0.5 ns. Photomultiplier jitter appears to be one of the main obstacles to higher accuracy although it can be statistically reduced if the pulse repetition rate is high.

Ramsden (1977) reviews some of the progress that can be expected in laser ranging. Expected improvements include:

(i) Improved optical efficiency of the receivers by increasing their reflectivity at the laser wavelengths.
(ii) Higher laser output using new host materials from which larger rod sizes can be manufactured.
(iii) Reduction in the detector noise.
(iv) Improvements in the pointing accuracy so that the transmitted beam width can be reduced and, for the same reason, improvements in the prediction accuracy of the satellite positions.
(v) Reduction in the laser pulse length to a few tens of picoseconds.
(vi) Use of two wavelength measurements to enable a correction to be made for the atmospheric refraction since the refractivity is frequency-dependent.
(vii) Use of continuous CO₂ lasers to measure the relative velocity of the target by the Doppler shift in the frequency of the return signal compared with the transmitted signal. The frequency difference is measured by heterodyne detection. The accuracy of these measurements depends mainly on the frequency stability of the laser over the observing time, a stability of 1 Hz corresponding to an accuracy of 0.01 mm s⁻¹.

3.4. Long-baseline radio interferometry

The technique of long-baseline interferometry is one of using widely separated radio antennae in the interferometric mode. Signals received from a radio source at the two ends of the baseline are compared and the difference in the times of arrival are estimated. The radiation source could be a signal emitted from a satellite or from a transmitter placed on the Moon, as has been done in the Apollo Lunar Lander programme, or it could be a natural stellar source. Ideal stellar sources are strong, broad-band, point sources, sufficiently far beyond our Galaxy for them to have negligible proper motions. A number of quasars satisfy these requirements. For the stellar sources, the incoming wavefront can be considered to be plane and the time delay, for a baseline of length D, is:

\[ \tau = D \cos \psi / \lambda \]  

(3.1)

where \( \psi \) is the angle of incidence of the wavefront relative to the baseline. Consider a source at right ascension \( \lambda_S \) and declination \( \delta_S \). The baseline orientation can be described by the right ascension \( \lambda_B \) and declination \( \delta_B \) of its projection onto the celestial sphere and (3.1) becomes:

\[ \tau = \frac{D}{c} [\sin \delta_S \sin \delta_B + \cos \delta_S \cos \delta_B \cos (\lambda_S - \lambda_B)]. \]
Alternatively, the baseline can be specified by the spherical coordinates $R_i$, $\phi_i$, $\lambda_i$ $(i=1, 2)$ of the two antennae. Then, for simplicity, we assume a spherical Earth of radius $R$:

$$\sin \delta_b = \frac{R}{D} (\sin \phi_2 - \sin \phi_1)$$

$$\cos \lambda_b = \frac{R}{D \cos \delta_b} \left[ \cos \phi_2 \cos (\lambda_2 + \theta) - \cos \phi_1 \cos (\lambda_1 + \theta) \cos (\lambda_2 - \lambda_1) \right]$$

and (3.1) becomes:

$$\tau = \frac{R}{c} \left[ \sin \delta_b (\sin \phi_2 - \sin \phi_1) + \cos \delta_b \left[ \cos \phi_2 \cos (\lambda_1 + \theta - L_b) - \cos \phi_1 \cos (\lambda_2 - \lambda_1 - L_b) \right] \right]$$

where $\theta$ is the sidereal angle of the Earth at time $t$. The time delay $\tau$ therefore varies diurnally with an amplitude and phase that depend upon the stellar and baseline coordinates. Observation of this delay at different times and for different sources permit the determination of these coordinates.

From (3.1) the precision of the baseline determination is inversely proportional to the baseline length since the error in determining $\tau$ is essentially independent of the delay. Baselines of up to $100$ km can be achieved by establishing a radio link between the two antennae but propagation anomalies limit the useful extension of the baseline beyond this. However, a direct radio link is not necessary now that stable atomic frequency and time standards are available. The received signal at each antenna is recorded onto tape together with the frequency standard information that ensures that the two receivers are operating at the same frequency. The time standards are required to ensure that the clocks remain constant over the period it takes to make a single measurement. The magnetic tapes are then brought together and processed in a correlator. A maximum correlation is found when one record is offset by an amount equal to the time required for the wave to travel the difference in distance between the source and the two antennae. The first measurements were reported by Broten et al (1967), Bare et al (1967) and Moran et al (1967). While these papers emphasised the radio astronomy aspects, the geophysical implications of these measurements were quickly recognised by Gold (1967) and MacDonald (1967). Counselman (1976) reviews the method and recent progress.

The length of the baseline is now limited only by the condition that the source be mutually visible from the two ends. Thus baselines between California and Australia have been used by the Jet Propulsion Laboratories. Measurement errors in the delay time $\tau$ are of several types. Delays introduced by the atmosphere are important. The ionosphere can make a significant contribution but, as the index of refraction is frequency-dependent, observations at two distinctly different frequencies (for example, S and X band) can remove this. Tropospheric refraction is more important. As with laser measurements, the total effect of the dry atmosphere is about $2-2$ m for a vertical ray path but the water vapour content is much more important at radio than at optical wavelengths. Pressure, temperature and humidity measurements at the receivers can reduce the uncertainty of the correction to about $20$ cm for a zenith direction. For better results it is necessary to measure the water vapour content along the path. This can be done with a microwave radiometer.
directed along the line of sight to the radio source. When operated near the water vapour line of 22 GHz, the total water vapour content along the ray path is measured and the delay correction can be made with an accuracy equivalent to about 2 cm in distance (Staelin 1969).

Figure 4. Results for the observed nominal distance between two radio telescopes—Goldstone, California and Haystack, Massachusetts—measured at different times by long-baseline interferometry (from Shapiro et al. 1974).

Instrumental errors resulting from instabilities in the frequency standards, calibration uncertainties, variations in the geometry of the antenna dish, and miscellaneous electronic sources are believed not to exceed about 5 cm and this is borne out by repeated measurements across the continental United States between the 64 m diameter antenna of the Deep Space Network in Goldstone, California and the 37 m diameter antenna of the Haystack Observatory in Massachusetts. The results by Shapiro et al. (1974) are illustrated in figure 4. Long-baseline interferometer observations using artificial lunar sources are discussed by King et al. (1976).

4. Estimating geodetic parameters

4.1. Dynamic methods

In §2 I outlined what is essentially a problem of celestial mechanics, the description of the motion of a close Earth satellite subject to a variety of forces. The geodetic problem is the inverse of this: the estimation of certain unknown parameters describ-
ing the forces from the observed motions of a satellite. Consider a tracking station \( P \) at \( x_p \). Relative to an inertial frame this position is \( X_p \). Assume that the topocentric position of the satellite \( \rho \) is observed and that the geocentric position \( X_g \) of the satellite is given by the orbital theory. These three quantities are related by the simple equation:

\[
X_p + \rho = X_g
\]  

(4.1)

where \( X_g = X_g(\kappa_i) \) can be related to the Keplerian elements \( \kappa_i \). Specifically, from the geometry illustrated in figure 1:

\[
X_g = \begin{pmatrix}
\cos \Omega & -\sin \Omega & 0 \\
\sin \Omega & \cos \Omega & 0 \\
0 & 0 & 1
\end{pmatrix}
\begin{pmatrix}
1 & 0 & 0 \\
0 & \cos I & -\sin I \\
0 & -\sin I & \cos I
\end{pmatrix}
\begin{pmatrix}
\cos \omega & -\sin \omega & 0 \\
\sin \omega & \cos \omega & 0 \\
0 & 0 & 1
\end{pmatrix}
\begin{pmatrix}
r \cos f \\
r \sin f \\
0
\end{pmatrix}
\]  

(4.2)

where the true anomaly \( f \) is related to the mean anomaly \( M^* \) by the equations (2.5). The Keplerian elements are time-dependent and with (2.6) \( X_g \) can then be expressed as:

\[
X_g(t) = X(a_0, e_0, I_0, \omega_0, M^*_0, C_{tm}, S_{tm}, \beta_i, t - t_0)
\]  

(4.3)

where the subscripts 0 indicate the values of the elements at time \( t_0 \) and where \( \beta_i \) refers to parameters that may be introduced to quantify further forces acting on the spacecraft.

Conventionally, the station position is referred to Earth-fixed axes \( x \). The inertial system \( X \) is defined relative to the stars. It can be defined in several ways but we will consider an \( X_3 \) axis normal to the mean ecliptic plane at a given epoch and \( X_1 \) in this plane and directed towards the mean equinox. The motion of the Earth's rotation axis relative to this frame is described by Euler's dynamic and kinematic equations for rotating bodies and is referred to as precession and nutation. Precession refers to the secular motion of the plane of the equator about the ecliptic at a rate of about 50 arcsec yr\(^{-1}\), and nutation is the sum of a variety of periodic motions of the spin axis superimposed upon this secular movement. The relation between \( X \) and a set of axes \( X' \) parallel to the rotation axis but not rotating with the Earth can be expressed as:

\[
X = R(\kappa, \nu, \omega)R(\Delta \mu, \Delta \nu, \Delta \epsilon)X'
\]

where \( R(\Delta \mu, \Delta \nu, \Delta \epsilon) \) describes the nutation and \( R(\kappa, \nu, \omega) \) describes the precession. Together they rotate the instantaneous sidereal system \( X'' \) to the mean system \( X \) at an epoch \( T_0 \). The rotation axis is not fixed relative to the Earth, nor is the rotation of the Earth uniform and the complete transformation from \( X \) to \( x \) is:

\[
X_p = R(\kappa, \nu, \omega)R(\Delta \mu, \Delta \nu, \Delta \epsilon)R(\theta)R(m_1, m_2)X_p
\]

where \( R(m_1, m_2) \) rotates the \( x \) axes parallel to the rotation axis and \( R(\theta) \) rotates this instantaneous system, rotating with the Earth, to \( X' \). These two rotations are dis-
Methods and geophysical applications of satellite geodesy

cussed in more detail in §6.5. To describe the position of \( P \) completely we must know (i) the motion of the Earth in space (the precession and nutation), (ii) the motion of the rotation axis with respect to the Earth’s surface, and (iii) the position of \( P \) in the global terrestrial framework. The Eulerian theory of precession and nutation, plus astronomical observations to determine the constants, provides the first; observations of polar motion and the length of day provide the second. Conventional geodetic observations provide the position of the tracking station relative to a regional or continental geodetic system which itself can only be related to a geocentric system by satellite observations as discussed below. Thus:

\[
X_p = X(x_p, m_1, m_2, \theta, \Delta \mu, \Delta \nu, \Delta \epsilon, \kappa, \nu, \omega).
\] (4.4)

Other choices of inertial reference system are possible and often it is convenient to use a quasi-inertial system. As discussed briefly in §2.3, the Earth’s gravity field in the system \( X \) is time-dependent since the direction of the rotation axis and thereby the rotational bulge is not fixed in space and for each instant the gravity field should be re-evaluated in the \( X \) system. The use of certain quasi-inertial systems avoids the need to do this yet does not introduce significant perturbations into the description of the satellite motion. Camera observations of the satellite positions against the star background should be transformed into the same inertial system as used to describe the motions. Veis (1963) discusses the relations between the various reference frames in some detail.

With (4.3) and (4.4) equation (4.1) loses much of its simplicity and it is usually solved iteratively. In a first approximation one could assume that \( X_p \) and the parameters describing the forces are known and then solve for the elements \( a_0, e_0, I_0, \omega_0, Q_0, M_0 \) at epoch \( t_0 \). In a second iteration one could attempt to solve for a subset of the force parameters \( C_{lm}, S_{lm}, \beta_i \) as well as improve upon the values of the elements. The first step would usually be carried out individually for a series of observations for a particular satellite. The length of such a series, or orbital arc, will depend on the accuracy of the analytical theory or the integration process and on the objectives of the analysis. The second step would be carried out for a number of orbital arcs of one or more satellites. In subsequent iterations one would introduce further unknown, or inadequately known, parameters in the vector \( R_p \) (equation (4.4)), such as the station positions \( x_p \) or the rotation parameters. Clearly, the actual estimation process is a major effort involving a great deal of computing work. In the solution by Gaposchkin and Lambeck (1971), for example, 21 satellite orbits were analysed for a total of 114 orbital arcs and more than 60 000 observations. Each orbit was described by as many as 12 or 13 unknowns, the six Keplerian elements and coefficients to describe non-gravitational forces. The final solution was for these elements, some 300 gravity field coefficients and 40 station positions.

Experience indicates that the solutions converge. This is due to several factors. (i) The different forces give rise to orbital perturbations with quite different spectra, permitting an adequate separation of the parameters. This was seen in the discussion of the gravity field perturbations. (ii) Numerous satellites in different orbits are used. In particular, orbits with different inclinations are important to separate the gravity perturbations because of the inclination function \( F_{lm}(I) \) in the perturbation equations (2.10). Orbits of different semi-major axes and eccentricities are also useful in this respect. (iii) A large global tracking network is required to ensure that the distribution of observations along the orbit is both dense and uniform. This enables the short-period perturbations due to \( C_{lm}, S_{lm} \) to be separated from
each other and from the discrepancies arising in equation (4.1) from errors in $X_P$.

(iv) Satellites in near-resonant orbits impose constraints on some of the potential coefficients. (v) Independent constraints, from other geodetic observations, are also important. One type, constraining relative station positions, is provided by the geometric method discussed below. Surface gravity data have also been used to constrain the solutions for the $C_{lm}$s, $S_{lm}$s, and the tracking of planetary spacecraft provides a precise determination of the first term in the geopotential, $GM$ and of relative station positions. Conventional geodetic observations can be used to constrain the relative positions of stations related to the same geodetic datum provided that the precision of the latter is equal to or exceeds that attained by the satellite method. All of these data sources were used in the solution by Gaposchkin and Lambeck and in the more recent solutions (e.g. Gaposchkin 1974).

4.2. Geometric methods

The geometric method of satellite geodesy does not require any information on the satellite orbit, the spacecraft being used merely as a well-elevated target that is observed simultaneously from a number of stations. A few examples illustrate the method. Two cameras, widely separated, observing the satellite at the same instant against star backgrounds provide two directions in space which define a plane on which the two stations lie. An additional set of observations defines a second plane whose intersection with the first gives the direction of the vector joining the two stations (figure 5). Observations from numerous stations to different satellite

![Figure 5. Illustrations of the geometric method of position determination from observations of satellites. (a) illustrates the case of simultaneous direction observations from two stations $P_A$ and $P_B$ to the satellites at $S_1$ and $S_2$. (b) illustrates the case of simultaneous distance measurements from four stations to the satellite positions $S_i$. In the first case the direction of $P_A$ to $P_B$ is determined, and in the second case the position of $P_D$ relative to the positions $P_A$, $P_B$, $P_C$ is determined.](image)
Methods and geophysical applications of satellite geodesy

positions enables the establishment of a network of interstation vectors. This is similar to terrestrial geodetic triangulation networks except that now it is a fully three-dimensional calculation, whereas classical geodetic triangulations are usually two-dimensional. Furthermore, the station separations may be very large (of inter-continental dimensions), being governed by the satellite height. No information on the scale of the network is provided by the optical observations alone. Global geometric networks have been established from the Baker-Nunn camera network and from the BC-4 camera network (Schmidt 1974).

A variant of the geometric method is to simultaneously observe distances from a number of stations to the satellite. Assume, for example, that a satellite is observed simultaneously from three stations whose positions are known. These observations determine the satellite position as the intersection of three spheres, centred at the respective stations and with radii equal to the observed ranges. A simultaneous observation from a fourth unknown position locates this station on a sphere centred at the satellite (figure 5(b)). Repeating the observations, when the satellite is at different locations, determines the position of the fourth station relative to the other three. This approach has been used extensively in a United States Army system called SECOR (sequential collation of range) during the 1960s but does not appear to have contributed significantly to the estimation of global geodetic parameters. Laser ranging to the satellite could provide very precise measurements of relative station positions but the condition that the satellite be mutually visible from four stations at any given time usually results in unsatisfactory geometry, particularly as atmospheric absorption and refraction limits observation to altitudes above about 30° elevation. Combinations of simultaneous range and distance measurements are also possible but as the laser range measurements are now much more precise than the optical data, there is little benefit gained by doing this. In global studies, the geometric method alone appears to offer few advantages over the dynamic method and its main contribution has been, by imposing conditions on relative station positions, to strengthen the dynamical solutions for station coordinates (e.g. Gaposchkin and Lambeck 1971, Gaposchkin 1974). Only in regional problems, such as measuring crustal deformations across major fault systems, may it be useful to use a number of laser tracking systems for simultaneous range observations.

4.3. Surface gravity

The traditional way of estimating the Earth's external gravity field is from surface gravity measurements or from geodetic observations of the geometric form of equipotential surfaces. The equipotential surface most widely used by geodesists is the geoid, \( W_0 \), the level surface approximating the ocean surface. Observations of gravity, \( g \), made on the Earth's surface at \( P' \) are reduced to \( P \) on the geoid and compared with theoretical gravity values to give gravity anomalies. The theoretical gravity values, \( \gamma \), are based on an ellipsoidal Earth model with the same mass and equatorial radius as the Earth and whose surface is of the same potential as the geoid. The potential corresponding to this model is \( V \). If \( P \) is projected onto this reference surface at \( Q \), the gravity anomaly is defined as:

\[
\Delta g = g_P - \gamma_Q. \tag{4.5}
\]

\( \Delta g \) is the free-air gravity anomaly if, in its reduction from the surface to the geoid, only the height difference is considered and the attraction from any intermediate
topography is ignored. The geoid height of \( P \) above \( Q \) is denoted by \( N \). The anomalous potential at \( P \) is defined as:

\[
\Delta W = W_0 - V_P = W_0 - V_Q + (\partial V/\partial r)N = \gamma N
\]

or

\[
N = \Delta W / \gamma \tag{4.6}
\]

which relates the geoid height \( N \) to the anomalous potential \( \Delta W \). With (4.5) the gravity anomaly becomes (Heiskanen and Moritz 1967):

\[
\Delta g = -\frac{\partial \Delta W}{\partial r} + \frac{1}{\gamma} \frac{\partial \gamma}{\partial r} \Delta W
\]

\[
= -\frac{\partial \Delta W}{\partial r} - \frac{2}{R} \Delta W. \tag{4.7}
\]

The potential \( W \), described in a reference system that rotates with the Earth, contains the gravitational potential \( U \), equation (2.1) and the potential of the centrifugal force. The gravitational attraction \( V \) of the reference body, also including the centrifugal force, can be expanded into a spherical harmonic expansion containing only even zonal terms. Then:

\[
\Delta W = \frac{GM}{r} \sum_l \sum_m \left( \frac{R_e}{r} \right)^l (C_{lm}^* \cos m\lambda + S_{lm} \sin m\lambda)P_{lm}(\sin \lambda) \tag{4.8}
\]

where \( C_{lm}^* = C_{lm} \) (observed) - \( C_{lm} \) (reference body). Practically, only \( C_{2,0} \) is modified. Then with (4.6) and (4.7):

\[
N = a_0 \sum_l \sum_m (C_{lm}^* \cos m\lambda + S_{lm} \sin m\lambda)P_{lm}(\sin \phi) \tag{4.9}
\]

and

\[
\Delta g = \gamma \sum_l \sum_m (l - 1) (C_{lm}^* \cos m\lambda + S_{lm} \sin m\lambda)P_{lm}(\sin \phi) \tag{4.10}
\]

with

\[
C_{lm}^* = \frac{(2 - \delta_{0m})(2l + 1)(l - m)!}{4\pi(l - 1)(l + m)!} \int_s \Delta g P_{lm}(\sin \phi) \left( \frac{\cos m\lambda}{\sin m\lambda} \right) ds \tag{4.11}
\]

where the integral is over the Earth's surface. To evaluate the Stokes coefficients from the integrals (4.11) requires that gravity be known over the entire surface of the Earth. Usually, one first computes the area mean from the individual measurements. If blocks of \( 1^\circ \times 1^\circ \) size are adopted, some 19 000 of a total of 64 800 would have one or more gravity measurements within it. About 21\% of the oceanic \( 1^\circ \times 1^\circ \) blocks contain one or more observations and 46\% of the continental blocks contain some data. For many of these blocks the data are of questionable reliability. In the second step one is obliged to 'predict', in a more or less arbitrary way, what gravity should be in the unsurveyed areas (Orlin 1966). Finally one computes the Stokes coefficients using the integrals (4.11). For practical purposes these coefficients should be comparable with those deduced from the satellite orbit analyses. However, because of the incomplete coverage, the surface data are most useful when introduced into the satellite solutions as additional constraint between the Stokes coefficients rather than for estimating them directly.
4.4. Results

Estimation of gravity field coefficients and station positions is carried out at a number of centres: the Center for Astrophysics of the Harvard and Smithsonian Observatories, the Geodynamics Branch of the Goddard Space Flight Center, the Groupe de Recherche de Geodesie Spatiale of the Centre National d'Etudes Spatiales, and the Naval Weapons Laboratories.

Recent models from the first centre (Gaposchkin 1974, 1977) have continued using an analytical approach to describing the satellite motion. Gaposchkin's (1974) solution includes Baker-Nunn camera and laser range observations from the Smithsonian and other tracking networks, including the BC-4 geometric triangulation data of the National Ocean and Atmospheric Administration. Surface gravity data, deep-space probe information from the Jet Propulsion Laboratories, and surface triangulation results were included in the solution for geopotential harmonics through degree and order 18. More recently, Gaposchkin (1977) has analysed the laser tracking data only and combined this with surface gravity data to estimate the harmonics through to degree and order 24. The solutions by the Goddard Space Flight Center (e.g. Smith et al 1976) use similar satellite data to that discussed above in addition to some Doppler data on four satellites and radar observations on one satellite. Surface gravity data have also been incorporated. Numerical integration techniques are used to describe the satellite motion. Geopotential coefficients complete to degree and order 16 and station locations have been determined. The results published by the Centre National d'Etudes Spatiales are also based on numerical solutions of the equations of motion (Balmino et al 1976). The solutions by the Naval Weapons Laboratory use only Doppler observations of the US navigation satellites and the GEOS geodetic satellites. No recent results for the gravity field have been published in the open literature although station coordinate solutions have been discussed (Anderle 1974).

The accuracies of the solutions are difficult to assess and published results of standard deviations have nearly always been over-optimistic. There are few external standards against which they can be tested. One test is to compare solutions, based on satellite observations only, against the terrestrial gravity data using methods developed by Kaula (1966). A practical limitation of this approach is that the global solutions are now nearly always iterative to ensure a satisfactory separation of potential coefficients and station coordinates and purely satellite solutions are seldom computed. Comparisons of the combined solutions against gravity are therefore only of limited value. Other tests employed include the computation of orbits and comparison with observations not used in the solution. This may be a satisfactory test for the lower-degree coefficients if no orbital arcs of the satellite have been included in the solution but it does not test for the higher-degree coefficients to which the satellite is relatively insensitive. In the iterative solutions the various subsets of data are usually compared with each other and with the previous iteration results before combining them and starting the next iteration. In this way any major discrepancies are readily detected and resolved by either removing the troublesome data or by adding new data to improve the separation of parameters. This does have obvious liabilities and makes a comparison of the solutions by different institutions rather important. The published solutions mentioned above are independent in that in one an analytical approach has been adopted and that in the others different integration procedures and programs have been used. The data base is generally
similar in all models but there are some important differences. Gaposchkin's (1977) solution, for example, uses only the laser tracking data while Smith et al. (1976) use Doppler in addition to laser and optical data.

Gaposchkin (1974, 1977) estimates that the accuracy of the station coordinates is of the order of 2 or 4 m for the better determined stations, a value borne out by the intercomparisons of dynamic and geometrical satellite geodesy results, terrestrial triangulation and coordinates determined from the deep-space probe missions and lunar laser ranging but a comparison of the coordinate solutions by Gaposchkin (1977), Balmino et al. (1976) and Smith et al. (1976) suggests that there are systematic discrepancies that may exceed 5 m.

All gravity field models discussed above yield similar results for the form of the geoid; a form described by King-Hele (1975) as '... the western hemisphere has been taken over by a goat which is deep in discussion with a man from the East, a highbrow whose cranium dominates Asia' (see figure 6). There are, however, important differences in the detail. Some of these are illustrated in figure 7 for some selected geoid profiles computed from the above gravity field solutions. Differences of as much as 10 m occur with wavelengths typically of 20°. Comparisons with recently acquired GEOS 3 altimetry data do provide important tests since this provides information on the detail in the geoid relative to the lower degree, and better known, harmonics describing the satellite motion. Comparisons of the gravity anomaly maps are usually less satisfactory since, comparing (4.9) and (4.10), the shorter wavelength and less certain information receives more emphasis in $\Delta g$ than in $N$.

The gravity field models may be evaluated in terms of their discrete, non-dimensional, power spectra:

$$V_t^2(\Delta U) = \sum_m (\tilde{C}_{tm}^2 + \tilde{S}_{tm}^2)$$  \hspace{2cm} (4.12)

† In the projection used in figure 6, they are no longer on speaking terms and have turned their backs on each other.
Methods and geophysical applications of satellite geodesy

Figure 7. Geoid profiles according to the solutions by Balmino et al, Gaposchkin and Smith et al along an approximate circle from Northern Europe to the lesser Antilles. Also shown is the geoid height as observed by the GEOS 3 radar altimeter along the same profile (based on data provided by G Balmino).

where the $\tilde{C}_{lm}$ and $\tilde{S}_{lm}$ are fully normalised Stokes coefficients. $V_l^2(\Delta U)$ is also referred to as the potential degree variance. Figure 8 illustrates these spectra for the three models. In general, there is comparable power contained in each degree variance up to degree 16, but beyond this considerable variation occurs indicating that the noise contribution to $V_l^2(\Delta U)$ becomes dominant. No serious attempts have been made to obtain a true measure of the noise spectrum but if we consider the differences in the Stokes coefficients obtained in the three solutions as indicative of noise, then these difference spectra exceed the signal at degrees about 16 and upwards.

5. Further advances

5.1. Analyses

The methods of satellite geodesy discussed above are based on the analysis of the satellite motion in the non-central gravity field. Tracking accuracies have, however, reached such levels that they are now comparable to the magnitude of the perturbations and any further improvements will give only limited gains in the knowledge of the gravity field. Some progress can be made with (i) more satellites of the STARLETTE type in low orbits with different inclinations and eccentricities, and (ii) more laser stations, geographically well distributed, to provide uniform and
Figure 8. Geopotential power spectra, defined by equation (4.12), for the three solutions by Balmino et al (○), Gaposchkin (△) and Smith et al (□). The spectra of the differences in the Stokes coefficients of these three solutions are also indicated. ▽, △-○; ◆, □-△; ●, □-○.

Dense tracking coverage. Both of these factors will contribute to a better separation of the gravity field coefficients and to a greater stability in the solutions for the coefficients from degrees about 10 to 20. There are no immediate plans to launch additional satellites of this type or to develop laser tracking stations beyond those mentioned in §3.

In view of this limitation future studies of orbital perturbations using precise Earth-based tracking data will probably emphasise aspects other than the static gravity field. The tidal deformations of the Earth, both the solid Earth and oceans, cause perceptible perturbations in the motion of satellites with frequencies quite distinct from those caused by the Earth's gravity field (§6.4). Kinematic studies of the motion of the rotation axis will become important as will the studies of relative motions, of tectonic origin, between the tracking stations (§6.2). These studies require precise orbital theories and a precise determination of the gravity field insofar as it perturbs the satellite motion.

Numerical integration methods appear to be adequate to describe the motion of geodetic satellites with a precision of a few tens of centimetres over periods of several days, provided that the parameters defining the forces acting on the spacecraft are known. Balmino (1975) discusses various improvements that can be made in the integration of the equations of motion of close Earth satellites. Theories can also be expected to be pushed to higher order by using computers for doing much of the algebra, as is currently done for selected aspects of satellite theory and lunar theory. But a complete and very precise analytical theory, taking into account all the interactions between the perturbations due to the various forces acting on the
Methods and geophysical applications of satellite geodesy

Methods for geophysical applications of satellite geodesy 579

satellite, appears unlikely to provide a practical solution for estimating geodetic parameters. For the study of long-period phenomena a combination of numerical and analytical approaches may be useful. Such methods use the analytical theory to describe the short-period perturbations and integrate numerically for the long-period behaviour. This avoids the need for a complete analytical long-period theory and permits an integration with large step sizes without requiring high-order numerical integration algorithms.

If further information on the gravity field is sought, and from a geophysical viewpoint this is most desirable, an entirely new approach to satellite geodesy is required. Alternatively there should be increased activity in collecting, compiling and revising surface gravity data but in view of the relative inaccessibility of many parts of the world this is not likely to be a practical alternative. Of the satellite methods, the use of drag-free systems, radar altimeters and satellite-to-satellite tracking have already been proven. Another possibility is satellite gradiometry. Each of these methods is discussed briefly below and an attempt is made to evaluate their likely impact on our knowledge of the Earth's gravity field.

With these new techniques the gravity field will be determined with a spatial resolution as small as 100 km or less. Thus an expansion complete to degree and order 180 is required. This makes the representation extremely unwieldy since the field at any point will be the sum of \((l+1)^2\) terms. As discussed above, that part of the field that perturbs the satellite is conveniently expressed with the aid of spherical harmonics but an alternative representation for the remainder may be desirable, particularly as some of the new techniques may not give global coverage. Radar altimeters, for example, provide geoid information only over oceans. A number of alternative representations have been discussed. Giacaglia and Lundquist (1972) have proposed a representation by sampling functions each of which has the value of unity at one point and zero at the remaining sampling points. Other geopotential representations include a surface layer approach (Koch 1968) or by a number of buried masses or discs but their advantages over spherical harmonics for the Earth are not obvious. As the new systems measure directly the potential or a derivative of it, the most convenient representation will probably be in the form of residuals between the observed field and some reference model.

5.2. Drag-free satellites

The sensitivity of the satellite to the gravity field can be increased by lowering the orbit but this will be at the expense of increasing the atmospheric drag forces. One solution to this problem is the drag-free technique, originally proposed by B O Lange in 1964. It consists of a small proof mass within a cavity located at the centre of mass of the spacecraft. The proof mass, sheltered from the surface forces, follows a gravitational orbit while sensors measure the motion of the shell relative to it. This information can be used to activate control thrusters that guide the satellite as a whole along the gravitational orbit. Such an experiment, called DISCOS (disturbance compensation system), has been developed jointly at the Stanford and Johns Hopkins Universities and launched for the US Navy in September 1972, in the satellite TRIAD 1 (Johns Hopkins and Stanford Universities 1973). The Doppler tracking of the spacecraft indicates that the surface force cancellation is accurate to \(10^{-10} - 10^{-11}\) g, and that any residual non-gravitational forces on the satellite are less than \(10^{-10}\) g. Such spacecraft at low altitudes could provide signi-
ificant improvements in the gravity field knowledge. The magnitude of the orbital perturbations are proportional to \((R/a)(C_{1m} \text{ or } S_{1m})\) or, with (2.3), to \((R/a)^{11/2}\) and a satellite, which at 1000 km is barely perturbed by twelfth degree coefficients, will be equally perturbed at 300 km by seventeenth degree coefficients, all other factors being equal. The effective minimum height of the spacecraft will be governed largely by the amount of thruster propellant required to compensate for the exponentially increasing drag force, or by minimum orbiter lifetime required. For example, a drag force system with a lifetime of 1 yr at 240 km altitude will have a lifetime of only some six months at 200 km altitude. The NASA Space Shuttle, to be operative throughout the 1980s, could provide a useful means of launching a drag-free satellite in an orbit as low as 200 or even 150 km and retrieve it towards the end of its gravitational orbit for re-use in subsequent shuttle missions.

The French Space Agency (CNES) in 1975 launched in its D5B satellite, carrying a sensor which measures the forces on the proof mass relative to the shell without attempting to modify the satellite's motion. Instead, the gravitational orbit is reconstructed from the on-board accelerometer measurements. Accuracies of \(10^{-10}\) g have been achieved.

### 5.3. Satellite-to-satellite tracking

Additional information can be obtained from the orbital analyses if the satellite is tracked in a continuous manner so that the high-frequency perturbations, which contain much of the information on the high degree and order harmonics, can be observed. This becomes increasingly important the lower the satellite altitude, since the visibility of the satellite from the ground is more and more reduced. The tracking of one satellite from another can alleviate this problem. The simplest concept is to track a low orbiter, sensitive to the gravity forces, from a high satellite whose orbit, much less perturbed, can be readily and precisely determined from the ground-based observations. The relative positions of the two spacecraft can be measured using Doppler or electronic distance measurements. Ionospheric refraction effects are largely eliminated by multiple frequency tracking. Accuracies of the range-rate of \(0.1\) mm s\(^{-1}\), for averaging times of about 10 s, appear feasible and would indicate variations in the height of equipotential surfaces to within about 10 cm. Ideally, the lower satellite would be equipped with a surface force compensating device. The high satellite could be in a geostationary orbit although this does not give global coverage unless at least three such equatorial satellites, separated by 120° in longitude, are available. Preliminary experiments have been carried out using the ATS-6 geostationary satellite and GEOS 3 (Sjogren and Wimberley 1977). In this system a signal is transmitted from the ground station to ATS-6, retransmitted from ATS-6 to GEOS 3 and back, via the satellite, to the observer. The rate of change of the total range is observed but, as ATS-6 is geostationary, this is very nearly equal to the range-rate between the two satellites. Differentiation of these range-rate measurements gives the line-of-sight accelerations of GEOS 3 relative to ATS-6 and this is roughly proportional to the gravitational acceleration at the lower spacecraft altitude. The technique is very similar to that used for measuring the Moon's gravity field (Muller and Sjogren 1968, Sjogren 1977). The data collected appear to be insufficient in quantity and in coverage while the GEOS 3 satellite is too high to be sensitive to the higher harmonics and this particular system has not yet made an important contribution to the knowledge of the Earth's gravity field.
Another version of the satellite-to-satellite tracking concept is to use two satellites in similar, near-circular, orbits separated by some 100–300 km. Neglecting surface forces, the total energy of a satellite remains constant along its orbit, any variation of the potential energy being balanced by a change in kinetic energy. Hence the relative velocity of the two spacecraft, deduced from Doppler measurements, provides a measure of the difference in gravitational potential between the two positions. Both satellites will be affected by the long-wavelength variations in the gravity field in approximately the same way, and their relative velocities are insensitive to this part of the field. But the relative velocities would be sensitive to the short-wavelength spatial variations in the field. Surface forces cannot be neglected and the satellite orbits will evolve differently with time. Compensation of these forces, using the drag-free system discussed previously, can alleviate this particular problem. In any event, gradual changes in the orbits are of some interest in that it will make the measurements progressively sensitive to different wavelengths in the gravity field. Maneuverable satellites could be envisaged in which the two spacecraft, initially in similar orbits, are allowed to slowly drift apart and, when the separation becomes excessive so as to reduce the times of mutual visibility, they can be brought together once more. Wolff (1969), Schwarz (1970) and Morrison (1970) have investigated this concept and a possible system has been studied in some detail at CNES. Schwarz has compared the high–low versions and concludes that the latter is marginally more sensitive to the gravity field. Some preliminary results for the low–low method have been obtained from Doppler tracking between the Apollo and Soyuz spacecraft in 1975 (Vonbun 1977). Considerably more precise systems, using optical interferometer measurements, have been proposed in relation with the Space Shuttle programme.

5.4. Satellite altimetry

With the present knowledge of the Earth’s gravity field, the orbits of satellites at some 1000 km altitude can be determined with an accuracy approaching the tracking accuracy if the distribution of observations along the orbit is relatively dense. Then, if the height of the spacecraft above the Earth’s surface can be measured with a radar altimeter, the geometric form of this surface can be determined relative to the orbit. Particularly useful will be the spacecraft heights over the oceans since this surface approximates an equipotential. The geocentric position of the satellite at the time of the radar measurement, less the observed height, provides a measure of the geometric shape of the ocean surface and the geoid height. Radar altimeter measurements could be made over the entire Earth’s surface—observations over land would provide an average terrain elevation. Such measurements have been made for the Moon with the laser altimeter on board Apollo 14 (Kaula et al 1974) but the interpretation of the broader beam radar echoes for the Earth orbiters is not without ambiguity. Also the elevations do not relate to the potential in the way the ocean surface does. The accuracy with which the ocean surface can be determined depends upon the spatial resolution and precision of measurements as well as on the precision with which the orbit can be computed from ground tracking observations and the present gravity field and other force models. A first radar altimeter was flown on Skylab and provided useful but limited data. The altimeter on board GEOS 3, launched in 1975, has provided a much more comprehensive data set. The altitude is sampled at one second intervals, resulting in a spatial
resolution at the ocean surface of about 7 km, and this is about equal to the size of the footprint of the radar. The precision of the radar measurements is of the order of 1 m. Thus variations in the form of the ocean surface of about 1 m over distances of a few tens of kilometres can be detected. Such observations are of considerable interest in studying a number of marine geophysics problems whose associated gravity anomalies cannot be detected by orbit perturbation analyses and which previously could only be investigated from shipboard gravity measurements. Examples include the ocean trenches. Despite the large amplitudes of the short-wavelength gravity anomalies over these features, they are not detected in the satellite accelerations because their wavelength is short but the geometry of the ocean surface over these features is readily seen (figure 9). Other examples include ocean ridges and sea mounts and it is anticipated that the altimeter observations will provide important data for studies of the ocean crust and lithosphere (Poehls et al. 1977).

The radar altimeter observations provide the form of the ocean surface, a surface which will not always coincide with the mean geoid. Differences occur due to tides, winds and currents, temperature, pressure and salinity variations. They can exceed a metre in some cases (table 3). Hence, for refined geoid studies, information on these oceanographic factors is required. Alternatively, by sampling selected regions at different times, time-dependent fluctuations in the shape of the ocean surface may be detected and these would be of oceanographic interest if the precision of the altimeter and the frequency of sampling are both high. An oceanographic satellite, SEASAT A, is to be launched later in 1978† and its radar altimeter has a design precision of some 10–20 cm. This is compatible with the amplitudes of many of the oceanographic anomalies. If the GEOS 3 radar is operational at the same time

![Image of GEOS 3 radar altimeter profile over the Marianas Trench in the Eastern Pacific.](image)

Figure 9. GEOS 3 radar altimeter profile over the Marianas Trench in the Eastern Pacific.

† SEASAT A ceased operation in October 1978 after a short but successful life.
Table 3. Order of magnitudes of differences between sea level and the geoid. These are global estimates. Local variations may reach much larger values.

<table>
<thead>
<tr>
<th>Phenomena</th>
<th>Amplitude (m)</th>
<th>Period</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ocean tides (relative to ocean surface)</td>
<td>~0.5–1.5</td>
<td>12 h, 24 h, plus smaller long-period tides</td>
</tr>
<tr>
<td>Variation in sea floor relative to centre of mass of Earth</td>
<td>~0.3</td>
<td>12 h, 24 h</td>
</tr>
<tr>
<td>Variable atmospheric loading</td>
<td>~0.1–0.2</td>
<td>Irregular, with strong annual component</td>
</tr>
<tr>
<td>Ocean currents</td>
<td>~0.1–0.2</td>
<td>Constant and seasonal components</td>
</tr>
<tr>
<td>Salinity</td>
<td>~0.1–0.2</td>
<td></td>
</tr>
<tr>
<td>Temperature</td>
<td>~0.1–0.2</td>
<td></td>
</tr>
<tr>
<td>Wind stress</td>
<td>~0.1–0.2</td>
<td></td>
</tr>
</tbody>
</table>

as SEASAT A, the sampling frequency over any one region will be increased and this may permit studies to be made of higher frequency sea-surface fluctuations such as tides.

5.5. Satellite gradiometry

The measurement of the gravity gradients of the external field was attempted as early as 1880 by Eötvös, and was further developed during the early part of the twentieth century for geophysical exploration work. The development of more portable and precise gravimeters, however, soon replaced the cumbersome and time-demanding gradiometers. More recently, it has been proposed to measure the gradients of the Earth’s gravity field by satellite-borne gradiometers (Bell et al 1965, Forward 1972) and a number of analyses have indicated that, if gradients could be measured at satellite heights with an accuracy of 0.01 EU (1 Eötvös unit = 10^{-9} \text{ gal cm}^{-1} = 10^{-9} \text{ s}^{-2}), the gradiometer is highly competitive with other methods of gravity field determination (Forward 1973). Gravity gradiometers have also been proposed for mapping the Moon’s gravity field (Forward 1976) and for estimating the mass distribution within the asteroid belt (Forward 1971). Satellite gradiometer designs have been studied by the Bell Aerospace Company, the Hughes Research Laboratories and by the Charles Stark Draper Laboratory. The Hughes system, discussed by Forward, consists of two sets of proof masses mounted on pairs of cross arms, which rotate about an axis through the intersection of, and normal to, the cross arms at frequency \( \omega \). In the presence of the gravity field, the end masses are differentially accelerated and the relative torques are measured by torsion-sensitive sensors. The torques vary sinusoidally with frequency 2\( \omega \) and with an amplitude that is proportional to the gravity gradient. The Bell Aerosystem instrument employs the rotating cross-arm scheme with four accelerometers mounted at the end points to sense directly the gravity gradient at twice the spin frequency. To measure all elements of the gravity gradient tensor, three orthogonally mounted instruments are required but usually only the vertical gradient components will be significant. For the instruments to be useful for satellite geodesy, a precision of better than 0.05 EU, with an integration time of a few tens of seconds, is required. Present studies indicate that flight instruments capable of measuring to 0.1 EU
can be developed but to attain an order of magnitude improvement will present one of the more challenging technical problems associated with improving our knowledge of the gravity field.

5.6. A comparison of the alternatives

In the search for increased accuracy and resolution of our knowledge of the Earth's gravity field, space technology offers a number of alternative approaches outlined above. Some of these have already been tested in orbit but their full potential has not yet been developed. Satellite gradiometry is still in the developmental stages and no instrument has been tested in orbit. In accepting the need to improve our knowledge of the Earth's gravity field a number of recent studies have compared the new methods in attempts to decide whether to pursue one technology or another (e.g. Kaula 1970). To make an order of magnitude comparison between the different systems we use the spherical harmonic expansion (2.1) of the gravity field. This is for convenience only and does not imply that this may be the most useful representation of the high-frequency content of the geopotential. The empirical rule (2.3) for the order of magnitude estimate of the Stokes coefficients is adopted as it appears to be valid at least up to degree about 40.

Following Kaula, we compare the various systems by computing the appropriate discrete power spectra of the measured quantity. Variations in the height of a geopotential surface at the satellite are given by:

\[ N_r = \frac{\Delta U}{g_r} = r \sum_l \sum_m \left( \frac{a_l}{r} \right)^l (C_{lm} \cos m\lambda + S_{lm} \sin m\lambda) P_{lm}(\sin \phi) \]  

(5.1)

and the spectrum of \( N_r \) at \( r \) is:

\[ V_l^2(N_r) = r^2 \left( \frac{a_e}{r} \right)^2 \frac{2l + 1}{l^4} \frac{10^{-10}}{10} \simeq 2r^3 \left( \frac{a_e}{r} \right)^{2l} \frac{10^{-10}}{l^3} \]  

(5.2)

for large \( l \). From ground-based observations of the satellite position one only observes those harmonics which result in \( V_l^2(N_r) \) greater than the measurement noise. With the altimeter, geopotential surface height variations are measured directly at the geoid and:

\[ V_l^2(N_R) = a_e^2 \frac{2l + 1}{l^4} 10^{-10} \simeq 2a_e^2 \frac{10^{-10}}{l^3} \]  

(5.3)

With satellite-to-satellite tracking by measuring the Doppler effect, one measures mainly a component in the differential velocity of two satellites and this is proportional to the potential difference between the two satellites, to their velocity and to the damping factor \((a_e/r)^l\). The power spectrum of the velocity differences will be:

\[ V_l^2(\Delta v_r) \simeq v^2 \left( \frac{a_e}{r} \right)^2 \frac{2l}{l^4} (2l + 1) \frac{10^{-10}}{l^4} \]

where the satellite velocity \( v \) is equal to \((GM/r)^{1/2}\). Hence:

\[ V_l^2(\Delta v_r) \simeq \left( \frac{GM}{r} \right) \left( \frac{a_e}{r} \right)^2 \frac{2l}{l^4} (2l + 1) \frac{10^{-10}}{l^4} \simeq 2 \frac{GM}{r} \left( \frac{a_e}{r} \right)^2 \frac{10^{-10}}{l^3} \]  

(5.4)
The gravity perturbation at the satellite height is:

$$\delta g = -\frac{\partial \Delta U}{\partial r} = \sum_l \frac{GM}{r^2} (l+1) (\frac{a_n}{r})^l \left( C_{lm} \cos m\lambda + S_{lm} \sin m\lambda \right) P_{lm}(\sin \phi)$$

and the power spectrum of the gradient of $g_r$ is:

$$V_l^2 (\text{grad } g_r) = (\frac{g_r}{r})^2 (l+1)^2 (l+2)^2 (\frac{a_n}{r})^{2l+4} \frac{(2l+1)}{l^4} 10^{-10}$$

The tracking of satellites from the ground or from very high satellites provides a measure of both the integrated or long-period perturbations and the high-frequency perturbations. It is the latter that are given by the spectrum (5.2) and a complete comparison also requires a discussion of the long-period perturbations. The spectrum (5.2) describes approximately the magnitude of the perturbations, apart from resonances that can be expected for the coefficients for which $m\theta > n$. Low–low satellite tracking, radar altimetry and gradiometer measurements directly sample the potential, geoid height or gravity gradient and there are no integrated effects. The mean square amplitudes of the relevant quantities are illustrated in figure 10. These can be compared with the error spectra of the various techniques. Probably

![Figure 10](image-url)

**Figure 10.** Comparison of the root mean square orbital perturbation due to potential coefficients of degree $l$. 1 represents the spectrum of geoid undulations as measured by an altimeter. 2–4 represent the spectra of the potential difference between satellites at altitudes of 250, 500 and 1000 km. 5–7 represent the spectra of the gravity gradient at three altitudes: 150, 250 and 500 km.
the simplest approach is to assume that, for the techniques that directly sample the field, measurements averaged over a few tens of seconds are uncorrelated. For 10 s, the distance covered by the satellite is only about 70 km and the error spectrum can be expected to be white over the wavelengths considered here. Tracking satellites at 1000 km altitude with Earth-based instrumentation accurate to 1 m permit observation of the high-frequency perturbations due to harmonics up to degree about 16. The same perturbations can be observed if this satellite is tracked by Doppler from a high satellite with an accuracy of about 0.1 mm s⁻¹, the advantage of the latter system being that the tracking coverage will be more uniform and the perturbation effects of the individual coefficients are more readily separated. A satellite altimeter with an accuracy of 1 m gives a small improvement in resolution, harmonics up to degree 20 giving a signal that lies above the noise. The results roughly summarise the present situation with the GEOS 3 satellite. For a 50 cm altimeter system the signal is above the noise up to degree 30 and represents a significant improvement, while for 10 cm accuracy, as proposed for the SEASAT satellite, the resolution extends to degree 100. These resolutions will only be available for the oceans and the observed result will be the geometric form of the ocean surface, not the geoid. One goal of physical oceanography is to be able to distinguish between these two surfaces and this requires that the geopotential be measured by another technique. If this technique is to be satellite–satellite tracking, a combination of high tracking accuracy and low satellite altitude is required to provide results that are compatible with the 10 cm altimeter. For example, measurements accurate to $3 \times 10^{-2}$ mm s⁻¹ for a 280 km altitude satellite or 0.01 mm s⁻¹ for a 150 km satellite are required. In both cases the low-flying satellite must be equipped with surface-force compensation devices. Low-flying gradiometers appear to be particularly sensitive to the high degree coefficients in the geopotential, the power spectrum (5.5) decreasing less rapidly than that of the other system. At 150 km altitude a gradiometer accurate to 0.02 EU would provide a resolution to degree 100.

6. Geophysical discussion

6.1. Introduction

While celestial mechanicians seek more erudite theories of satellite motion and geodesists use the motion to attain their geodetic goals of measuring the Earth, geophysicists use the information gained from the forces acting on the satellite to further their knowledge of the Earth's interior. Examples readily come to mind. Anomalous, i.e., non-hydrostatic, gravity implies lateral density variations within the Earth. These can be discussed in relation to the stress state of the Earth and mantle convection, and compared with tectonic, seismic, magnetic and surface heat flow observations. Slow motions of the stations may be indicative of global surface deformations which in turn relate to plate tectonics and mantle processes. Tidal and rotational deformations give constraints on elastic and anelastic properties of the Earth. At present, satellite observations provide—or promise to provide—information on the Earth's gravity field, on tides, on its rotational motion and on tectonic displacements. Each of these areas of geophysics is discussed in some detail and an attempt is made to evaluate the impact of the satellite observations on these sciences.

Important progress in geophysics made in the last decade has centred around the plate tectonics hypothesis. Seismological and other geophysical evidence points to
a layering in the upper mantle that, for present purposes, can be characterised by the rheological properties. The first layer, of about 70–100 km thickness, is referred to as the lithosphere and includes the crust. It is considered to be relatively rigid, cool, and responds mainly elastically to applied forces. The lithosphere overlies a high temperature, less viscous layer. This layer, the asthenosphere, cannot support shear stresses for any length of time. The lower limit of this layer is not always defined in the same way. We will take it to include the phase transition near 350–400 km depth that marks a transition of iron magnesium silicates, olivine (about 80% Mg$_2$SiO$_4$ + 20% Fe$_2$SiO$_4$), to denser forms, spinel, that are stable at high pressure (Ringwood and Major 1970). It extends down to a second major discontinuity near 650 km. The latter is apparently due to the separation of spinel into its constituent oxides (Ming and Bassett 1975). The lower mantle starts below this transition zone. Ringwood (1975) discusses in detail the petrology of the mantle.

The Earth’s seismicity is restricted mainly to certain zones corresponding to ocean trenches, ocean ridges and mountain chains (figure 11). In particular, deep earthquakes—from depths greater than about 150 km—are rarely found outside the ocean trench areas, the deepest earthquakes occurring near 600–700 km. This observation suggests that the Earth may be subdivided into units which are relatively aseismic and which join up along the ocean ridges, the trenches, mountain chains and large faults, defined by the seismicity distribution (Isacks et al. 1968). Magnetic anomalies over the oceans indicate that the oceanic crust increases in age outwards from the ocean ridges, and that sea-floor spreading occurs; basaltic material rises up to the surface along the ridges, cools and moves outwards as new material follows it (Vine and Mathews 1963, Vine 1966). The ingredients of (i) upper mantle rheology, (ii) seismicity distribution and (iii) the sea-floor spreading concept, form the basis of the plate tectonics hypothesis (figure 12). The lithosphere is assumed to be decoupled from the asthenosphere; material moves up from the mantle along the ocean ridges; the essentially rigid, aseismic, lithospheric plates are driven apart.

![Figure 11. Distribution of epicentres of earthquakes 1961–1967 with depths from 0–700 km (from Stacey 1977).](image-url)
and subducted into the mantle at the ocean trenches; return flow occurs at some depth in the asthenosphere to close the cycle. This kinematic picture of plate tectonics is supported by a variety of geological and geophysical observations (LePichon et al 1973) but as to what causes it, what mechanism drives the plates from the ocean ridges to be eventually subducted back into the mantle, remains unclear. Thermal convection must play a dominant role but details remain obscure. Satellite geodesy may make several contributions to understanding the processes involved. Observations of plate motions may give some insight into the nature of the motion over short time intervals. The gravity field may provide constraints on the convective processes and the tidal and rotational studies may give some constraints on global anelastic properties.

6.2. Earth's gravity field

The determination of the Earth's gravity field from satellite observations has been discussed in §4.1. Contributions from surface gravity observations have been discussed in §4.3. If the Earth is in hydrostatic equilibrium, its figure under self-gravitation and uniform rotation is a spheroid whose form is given by the solution of Clairaut's equation (Jeffreys 1962). The solution requires that the Earth's density gradient be known but, under the not very stringent condition that the polar moment of inertia $I_{33} > 0.32 Ma^2$, it need not be known with great accuracy. For the Earth $I_{33} \approx 0.33 Ma^2$. For the Moon $I_{33} \approx 0.39 Ma^2$ while for Mars $I_{33} \approx 0.37 Ma^2$. Thus the Moon is relatively homogeneous in density while the Martian density gradient lies between those of Earth and Moon. The gravitational potential of the equilibrium Earth model can be described by a spherical harmonic expansion containing even zonal harmonics, with only the second degree term being important. Figure 13 illustrates the difference between the observed gravity anomalies $\Delta g$ (equation (4.10)) and the hydrostatic field, based on the solution by Lerch et al (1977), gravity anomalies of the order of 40–80 mgal, with wavelengths greater than about 1000 km occur everywhere on the Earth's surface.

To obtain some order of magnitude ideas of the size of the density anomalies $dp$ required to produce this anomalous field and of its consequences, we assume that they originate in a layer of thickness $dr$. The anomalous surface density layer $\sigma$
on a sphere of radius $r$ is:

$$\sigma = d\rho dr = (\tilde{\rho}_{lm'} \cos m\lambda + \tilde{\rho}_{lm''} \sin m\lambda)\tilde{P}_{lm} (\sin \phi) \, dr$$

(6.1)

whose external potential is, if $dr \ll a_0$:

$$\Delta U = 4\pi G \rho \sum_{lm} \left( \frac{a_e}{r} \right)^{l+1} \frac{dr}{2l+1} (\tilde{\rho}_{lm'} \cos m\lambda + \tilde{\rho}_{lm''} \sin m\lambda)\tilde{P}_{lm} (\sin \phi).$$

(6.2)

Equating this with the observed potential (2.1) gives:

$$\frac{\tilde{\rho}_{lm'}}{\tilde{\rho}_{lm''}} dr = \frac{\tilde{R}}{3} (2l + 1) \left( \frac{R}{r} \right)^{l+2} \left( \frac{C_{lm}}{S_{lm'}} \right).$$

(6.3)

If the density anomalies occur over several layers of radius $r_j$, thickness $r_j$ and anomalous density $d\rho_{(j)}$:

$$\frac{C_{lm'}}{S_{lm'}} = \frac{3}{\rho \tilde{R}} 2l + 1 \sum_j \left( \frac{r_j}{\tilde{R}} \right)^{l+2} dr_j \left( \frac{\tilde{\rho}_{(j)lm'}}{\tilde{\rho}_{(j)lm''}} \right).$$

(6.4)

The power spectrum of surface density (6.1), analogous to (4.12), is:

$$V_s^2(\sigma) = \sum_m (\tilde{\rho}_{lm'}^2 + \tilde{\rho}_{lm''}^2) \, dr^2$$

and with (6.3) and (2.3):

$$V_s^2(\sigma) = \left( \frac{\tilde{R}}{3} \right)^2 \frac{(2l + 1)^3}{l^4} 10^{-10} \left( \frac{R}{r} \right)^{2l+4}.$$

(6.5)

Plausible sources of the density anomalies are numerous. We only consider for the moment (i) the asthenosphere, (ii) the principal phase transition zones and (iii) the surface topography.

For the first case $r \approx 6.2 \times 10^8$ cm and $dr \approx 2 \times 10^7$ cm. Then:

$$V_s^2(\rho) \approx 3.4 \times 10^{-7} \frac{(2l + 1)^3}{l^4} 10^{-10} (1.03)^{2l+4}.$$
The total RMS density anomalies in this layer would be of the order of \( \sqrt{\sum V_l^2(\rho)} \), or about 0.005 g cm\(^{-3} \) for \( l=2-20 \). If such anomalies are a consequence of lateral temperature variations—the thermal expansion coefficient of the upper mantle is of the order of \( 3 \times 10^{-5} \) °C\(^{-1} \)—a modest RMS temperature fluctuation of some 50°C is sufficient. If the density anomalies are a consequence of partial fusion, fluctuations of some 2% in the degree of partial fusion are required.

In the second case \( r \approx 6 \times 10^8 \) cm and the density contrast across the transition zone near 400 km depth is \( \Delta \rho \approx 0.3 \) g cm\(^{-3} \). A lateral variation in the temperature at this transition zone results in a migration in the depth of the zone since the pressure–temperature condition for the olivine to spinel transition is approximated by:

\[
P (\text{kb}ar) = aT (\text{°C}) + b
\]

where \( a \approx 0.062 \) kbar °C\(^{-1} \). If this variable surface layer is the cause of the gravity anomalies, the power spectrum follows from (6.5) as:

\[
V_l^2(\rho) \approx \frac{\tilde{\rho}R}{3 \Delta \rho} \frac{(2l+1)^3}{l^4} 10^{-10} \left( \frac{R}{r} \right)^{4l+2}.
\]

For \( 2 \leq l \leq 20 \), the RMS topography of this layer required to explain the observed gravity field is about 3 km. This implies lateral temperature fluctuations of the order of 100°C. These two examples illustrate several points that emphasise the difficulty in interpreting the gravity field. First, the density anomalies required to explain the gravity field are small and their consequences on seismic velocities are often smaller than can be observed by present seismological techniques. The implied temperature fluctuations are also small and unlikely to result in recognisable heat flow fluctuations at the Earth's surface. Second, a number of mechanisms causing density anomalies need to be considered. If lateral temperature anomalies occur, density fluctuations may result from thermal expansion, variable partial melting or from a migration of the phase transition zones. The migration of the boundary with increasing temperature may be downward or upward depending on whether the phase transition is exothermic or endothermic. McQueen and Stacey (1976) have suggested that irregularities in the two transition zones near 400 and 650 km contribute together to the observed gravity field. Temperature-dependent chemical differentiation may also occur. Third, the lateral variations of few other geophysical quantities, reflecting mantle conditions, are known with the same detail as gravity. Only the Earth's topography is known in more detail but, for the wavelengths with which we are concerned here, the relation between gravity and topography is not very important (Lambeck 1976). This is unlike Mars where there is a close connection between gravity and topography for the long-wavelength harmonics (Phillips and Saunders 1975, Lambeck 1978c). Some seismic observations indicate lateral variations but until now the emphasis is on estimating differences between, for example, average ocean and continental mantle, rather than mapping the actual mantle structure. Heat flow measurements show significant variations from one area to the next but this can be attributed largely to near-surface crustal variations. There are, in consequence, few data against which the density models, deduced from the surface gravity data, can be directly tested.

The most obvious contribution to the Earth's gravity field is the topography. If we expand the heights of the Earth’s surface, measured relative to the geoid, in
spherical harmonics as:

\[ h = (\bar{h}_{l'm'} \cos m\lambda + \bar{h}_{l'm''} \sin m\lambda) \bar{P}_{lm}(\sin \phi) \]

the potential outside the Earth due to this layer is given by (6.2) with \( d\Omega d\rho \) replaced by \( h_\rho \) where \( \rho_e \) is the average density of the crust. The non-dimensional power spectrum of this potential is:

\[ V_1^2(\Delta U_h) = \left( \frac{3 \rho_e}{\bar{\rho} R} \right)^2 \frac{1}{(2l+1)^2} \sum_m [(\bar{h}_{l'm'})^2 + (\bar{h}_{l'm''})^2]. \] (6.6)

For the Earth, the actual topography is replaced by an equivalent topography in which oceans, large lakes and ice sheets are replaced by an equivalent rock layer (Balmino et al 1976). Comparing this computed spectrum with the observed spectrum \( V_1^2(\Delta U) \) (equation (4.10)) gives (figure 14):

\[ V_1^2(\Delta U_h) \gg V_1^2(\Delta U). \]

This is a consequence of the near-isostatic compensation of topography in which elevated areas are compensated by relatively near-surface mass deficiencies and vice versa.

Observations of this phenomenon go back to the geodetic observations of G Everest in India. Mathematical formulations of the compensation were proposed almost simultaneously by Pratt in 1854 and Airy in 1855. In the latter model, the crust is assumed to be of uniform density but of variable thickness and overlying

![Figure 14](image-url)  

**Figure 14.** Observed power spectrum of the geopotential compared with that due to the attraction of the topography and that due to topography and its compensation.
a denser mantle. The thickness of the crust is a function of the elevation of the topography such that at a constant depth, below the deepest part of the crust, the pressure is everywhere constant. In the Pratt model, the crust is assumed to be of constant thickness relative to the geoid but of variable density such that pressure at the base of the layer is constant. A comparison of these models with observations indicates that both mechanisms occur. For old stable continents and possibly for ocean basins, the Pratt model prevails with the depth of compensation corresponding roughly to the base of the lithosphere, about 100–200 km deep under the continents. In tectonically active areas, such as the European Alps, the Airy model explains well the observations. Variants of both models have been proposed. Heiskanen and Vening-Meinesz (1958) discuss the isostatic theory and observations in detail and, on a regional scale, isostasy appears to prevail (Woollard 1972). Deviations from isostasy are described by an isostatic gravity anomaly. This is the free-air gravity anomaly corrected for the attraction by the topography and the compensating mass. Such deviations have, in the past, usually been attributed to the fact that the Earth's crust can support some stresses associated with uncompensated topography.

On a spherical Earth, the Airy model predicts a compensating layer of density $\rho_m - \rho_c$ ($\rho_m$ is the mantle density and $\rho_c$ is the crustal density) with thickness $H$ relative to a mean thickness $D$ given by:

$$H = \sum_i \sum_m \left( \frac{R - D}{R} \right)^i \left( \bar{h}_{lm} \cos m\lambda + \tilde{h}_{lm} \sin m\lambda \right) \tilde{P}_{lm} (\sin \phi).$$

The power spectrum of the potential due to this layer and the topography is:

$$V^2_l(\Delta U) = \left( \frac{3\rho_c}{3\eta R} \right)^2 \left( \frac{1}{2l+1} \right)^2 \left[ 1 - \left( \frac{R - D}{R} \right)^2 \right] \frac{1}{2} V^2_l(h).$$

For $D \approx 40$ km, as suggested by many regional studies, $V^2_l(\Delta U) \ll V^2_l(\Delta U)$, at least for harmonics up to degree 30 (figure 14). This difference could mean one of several things. First, the Airy model chosen could be inadequate. This does not appear to be the case since a very similar result is obtained using a Pratt model with compensation depth equal to $2D$. For the long wavelengths under consideration here, the mechanism by which mass is distributed in the vertical column is relatively unimportant. Second, the depth of compensation could be greater than the above value. This appears unlikely from the regional studies of isostasy and in any case the depth is unlikely to be below the base of the lithosphere. Much greater depths are required in order that $V^2_l(\Delta U) \approx V^2_l(\Delta U)$. A third possibility is that the compensation of the topography is not complete. This can be investigated by computing the stresses set up in the crust by the topography and its partial compensation and it is generally recognised that the values required are excessive. A fourth possibility is that the mantle, below the depth of compensation, is not in hydrostatic equilibrium as implied in the isostatic models. This is the most satisfactory explanation. That lateral density anomalies must occur deep in the mantle in order to explain some of the larger gravity anomalies was already recognised by Jeffreys (1942) but until the satellite results became available and accepted it was still widely believed that any isostatic anomalies were likely to be small. In view of the fact that $V^2_l(\Delta U) \ll V^2_l(\Delta U)$ the free-air gravity anomaly map (figure 13) is almost identical to the
Methods and geophysical applications of satellite geodesy

isostatic anomaly map and it reflects density anomalies below the crust that are not directly related to the topography.

While the satellite results provide important evidence that the Earth is not in hydrostatic equilibrium, they do not indicate how these anomalies are supported in the mantle. There are two conflicting arguments. One is that the mantle possesses sufficient strength to be able to support indefinitely the density anomalies. This is argued by Jeffreys (1970) and it implies that the gravity field, as observed today, has remained constant since the formation of the planet. The more current, alternative, interpretation is that the density anomalies are associated with mantle convection. In this case the gravity field will have evolved with time. A way of deciding between these two alternatives, independently of other geophysical evidence for convection, is to compute the stresses required to support the density anomalies and to decide whether these can be supported by Earth materials under pressures and temperatures that are representative of the mantle. Early calculations were performed by Jeffreys (1942) and the most complete calculation is that by Kaula (1963). In the latter study Kaula assumes that the planet is initially in hydrostatic equilibrium and is then loaded by topography and internal density anomalies distributed so that the model gravity field equals the observed field. To obtain a unique solution Kaula minimises the shear strain energy over the Earth. This condition probably results in minimal estimates of the shear stresses in the mantle. For harmonics complete to degree and order four, Kaula finds maximum shear stresses of the order of 250 bar with a major contribution, about 160 bar, coming from the difference between the observed and hydrostatic values of $C_{20}$. Higher-degree harmonics in the gravity field increase this value. From Jeffreys' analyses and from the nature of the geopotential power spectrum, this increase follows the $1/l^2$ rule and maximum shear stresses of up to 400 bar may be expected. Reliable estimates of the yield point of the mantle at relevant pressures and temperatures are not available but the general consensus seems to be that, below the lithosphere, creep will occur at shear stresses well below 100 bar (e.g. Weertman 1970). Other forces associated with flow in the mantle are therefore required to explain the observed gravity field. The non-hydrostatic gravity field also implies that there is free potential energy stored within the Earth and, if motion is kinematically possible and no further energy is generated elsewhere, there will be a rearrangement of mass towards distributions of lower energy. The anomalous gravity field can therefore be expected to bear some relation to motions within the Earth and, in a very general way, should reflect convective processes.

If any part of the mantle can be described by a single equation of state, the equilibrium condition implies that equipotential surfaces are also surfaces of constant density and temperature. Lateral density variations, whether by thermal expansion, partial fusion, mineralogical phase changes or chemical differentiation, can be attributed as a consequence of horizontal temperature gradients or to compositional variations within the Earth. Thus the gravity may also be expected to relate to other geophysical observations that are perturbed by these temperature- and pressure-dependent phenomena. Any relation will not be simple and will not be without ambiguity, due to the variety of inter-related processes involved. Some of these were seen in the discussion of the contribution of the low velocity layer and phase transition zones to surface gravity. A further example is provided by an ocean-continent model in which the lithosphere is thicker under the latter. Free-air gravity over the continent will be positive even if isostasy prevails. If the base of the
lithosphere can be assumed to be a geotherm, temperatures under the continent will be lower than at the same depth under the oceans. This makes a further contribution to the positive gravity over land. Should partial fusion occur, it will occur at shallower depths under the oceans than under the continents, should it occur under the latter at all. This adds a further positive gravity anomaly over the continent. Phase transition zones under the continent will migrate upwards in response to the lateral temperature differences if they are exothermic and also add to the gravity anomaly. The total anomalies rapidly become excessive, even if the lateral temperature difference is a modest 100°C and there must be other ways of reducing them. Lighter mantle materials under continents than under oceans, as proposed by Ringwood (1966) for other reasons, alleviate some of this difficulty as would an endothermic phase transition.

In most dynamic situations the problem is further complicated by the interaction between the convective processes and the boundary layers. For example, if the lithosphere were rigid, convection underneath it will result in negative anomalies over the upwelling current. But if the boundary deforms, the gravity anomaly will be a consequence of the mass deficit in the rising column and of the deformed layer. The two contributions will be in opposition and it is not always evident which of the two dominates (see, for example, Kaula 1972, McKenzie 1977). Thus gravity anomalies over ocean ridges, created by thermal expansion of the lithosphere, will consist of a negative part due to the lower density of the material and a positive part due to the ridge topography being closer to the ocean surface. In this particular case the latter dominates and gravity anomalies over the ridges tend to be marginally positive (Lambeck 1972, Sclater et al. 1975).

Despite such complications there is evidence which suggests that the gravity anomalies relate to dynamic mantle processes. Active plate margins, in particular the subduction zones, tend to be associated with positive anomalies. Thus the Pacific Ocean is surrounded by positive anomalies, extending from the tip of South America up to Alaska and down the western side to New Zealand (figure 1.3). Also, the complex southern margin of the Eurasian plate tends to be associated with positive anomalies as do major parts of the ocean ridges, notably the North Atlantic ridge, the East Pacific Rise and parts of the Indian Ocean ridge system. Negative anomalies tend to occur within the interiors of the plates. Ocean basins, in particular, are associated with negative anomalies. Kaula (1972) has interpreted the gravity field in terms of flow in the asthenosphere and the response of the lithosphere to this flow. While the validity of this conclusion is generally accepted a more quantitative interpretation still does not exist. In particular, the convection problem does not appear to be sufficiently well understood for the observed gravity field to provide useful constraints on convection models. But gravity anomalies may provide some useful information on particular aspects of convection. Thermal models of ocean ridges must give gravity anomalies that are in agreement with those observed. Subduction zone models can also be controlled by gravity data. In particular, the models must explain both the large negative but short-wavelength anomaly observed over the trench and the positive and much longer wavelength anomaly seen in the satellite solutions (figures 9 and 13). Radar altimeter observations of the geoid are proving to be particularly useful in providing homogeneous and precise data over many of the world's trenches that previously had been investigated less systematically with marine gravity observations. Elastic and viscous properties of the lithosphere, particularly the ocean lithosphere loaded by volcanic islands, can also be profitably
studied with the satellite altimeter, as can the changeover from continental to oceanic lithosphere along some of the continental margins. In all cases rather detailed problems are investigated but their solution will provide useful constraints on global convection models. It is probably in this direction that high resolution and precise gravity observations, coupled with other geophysical observations, will contribute most to a further understanding of the Earth's interior.

Apart from the inherent ambiguity in potential inversion procedures, a further reason for the inability to deduce mantle flow patterns from gravity is that few other pertinent geophysical observations are available that permit a testing of the proposed models. Surface heat flow, as indicated earlier, varies regionally but this is due mainly to changes in the radiogenic contents of crustal rocks. The crustal contribution is removed, using empirically derived relations (see, for example, Roy et al 1972), and the so-reduced heat flow is considered to represent a measure of the heat flowing out of the upper mantle. The corrections are large: on average the continental crust may contribute some 50% to the observed flow. It is this heat flow that can be used as a constraint in thermonutectonic models and be expected to correlate with gravity. Unfortunately the crustal corrections are generally inadequately known and the reduced heat flow is unreliable. Recent global compilations of reduced heat flow show considerable similarity with the distribution of oceans and continents and, within the continents, with the distribution of tectonically active areas and old shields (Chapman and Pollack 1975, Pollack and Chapman 1977). Whether this correlation is due to actual flow from the mantle or to inadequate corrections that are dependent on the ocean–continent distribution remains uncertain.

Several seismic observations indicate lateral variations that can be expected to correlate with the gravity anomalies. Average velocity–depth curves have been established for both compressional and shear waves. Thus, in a given area, the expected times of first arrival of seismic waves generated by an earthquake can be computed and compared with the observed times. Differences, $\Delta t$, between predicted and observed travel times will be due to uncertainties in the time and location of the earthquake, and due to departures of the actual velocities from average values used in the model. Equations of state relating seismic velocities to density are poorly known for the mantle but empirical evidence suggests a linear relation between the seismic velocity of compressional waves and density of the form:

$$V_p = ap + b$$

where $a \approx 3$ km s$^{-1}$ g$^{-1}$ cm$^8$. If correct, we would expect to see a correlation between early arrivals and positive gravity. Such correlation is seen in a number of parts of the world such as North America and Australia.

Studies of surface waves also indicate that there are important lateral upper mantle variations (e.g. Knopoff 1972, Leeds et al. 1974) but no attempts appear to have been made to relate these to gravity observations.

6.3. Tectonic motions

6.3.1. Rates of motions. Our present notions about the plate tectonic motions come from (i) the matching of geological features across faults, (ii) the paleomagnetic record, (iii) analysis of fault plane solutions of seismic events along the plate boundaries, and (iv) geodetic measurements along the boundaries. The extent of the mismatch of geological features gives a gross estimate of cumulative movements...
over about $10^7$ yr or more but the results are qualitative and subject to distortion by erosion and other processes along the plate boundaries. The recent paleomagnetic data indicate average motions over time intervals of the order of $10^6$ yr. The first attempts at establishing the global pattern of plate tectonics from the magnetic evidence were by Morgan (1968) and LePichon (1968). A more recent reconstruction has been made by Minster et al (1974). Beyond about 10–15 million years ago the resolution of the data becomes less reliable. The success of the plate tectonics model in predicting motions suggests that, when viewed over such long integrating times, the plates can be considered as rigid entities and that the plate deformation is restricted to their margins. Seismic observations of plate motions consist of summing the seismic moments of earthquakes along the plate boundaries (Brune 1968). When integrated over a sufficiently long period—about 100 yr—these motions are often in agreement with those deduced from the paleomagnetic evidence but averaged over much longer intervals. Discrepancies between seismic slip and plate motions do occur, however, in many parts of the world. Between India and Eurasia the disparity between the two estimates of plate motion is by a factor of about 3 (Chen and Molnar 1977). Near Japan it is by a factor of about 5. Along the Marianas Trench the plate motions are estimated as nearly 10 cm yr$^{-1}$ but no large seismic deformations appear to have occurred at all during the last few centuries (Kanamori 1977). This suggests that aseismic deformation occurs that is not associated with large energy release at the frequencies to which seismometers are normally sensitive. Seismometers sensitive to much lower frequency or zero frequency displacements are required. This becomes the domain of geodesy: geodesy becomes low-frequency seismology.

Long intervals of geodetic observations along a few active plate margins exist. The best example is along the San Andreas and associated faults in California. The observations consist of repeated surveys of triangulation networks over the years following the great San Francisco earthquake of 1906. In the last decade these surveys have been complemented by geodimeter traverses (Whitten 1970). These measurements indicate that in some segments the motion along the fault is relatively continuous while in others the fault surface appears to be locked and strain is accumulating to be released at some time in the future by rupture along the fault. On the strength of these measurements, the average motion between the Pacific and North American plate is estimated to be of the order of 3 cm yr$^{-1}$ (Savage and Burford 1973). Geodetic triangulation measurements taken over some 50 yr and some 10 yr of geodimeter measurements indicate very comparable results and compare with 5.5 cm yr$^{-1}$ predicted by plate tectonic models (Minster et al 1974). The discrepancy may be due to slower average plate motions during the past half-century than during the past several million years or it may be due to the fact that motion between the two plates is accommodated over a much wider area than just the San Andreas and related faults. Geodetic measurements in the Pamirs to the north of the Indian–Eurasian plate boundary are also indicative of motion being absorbed over large distances (Pevnev et al 1975). In the global plate tectonic models the boundaries are assumed to be simple linear features, but when inspected more closely they are quite complex. The African plate, for example, is believed to be moving northwards into Eurasia at a rate of one or two cm yr$^{-1}$. At the Mediterranean end of the junction oceanic crust still lies between the two plates but at the Arabian end the collision is between two thick continental crusts. The consequence of this is that the movement between the two plates is absorbed by east–west motion.
of small plates wedged between the eastern parts of the two super-plates, small plates moving at rates of as much as 10 cm yr\(^{-1}\). LePichon et al (1973) give several other examples where the plate margins are complicated by the presence of small plates moving in directions quite different from the global motions.

All the above-mentioned observations measure only the relative motions between the adjacent plates. An indication of absolute motions would be of considerable interest in understanding the mechanisms driving the plates. The determination of absolute motions begs the question—absolute with respect to what framework? Many geophysicists have preferred mantle convection models that restrict flow to the asthenosphere, although one suspects that this is for mathematical or conceptual limitations only. In this case ‘absolute’ motion is considered to be the motion of the lithosphere relative to the deep mantle. One then searches for markers on the surface that remain fixed relative to the lower mantle and with respect to which the surface motions can be measured. A favourite set of markers are the volcanic island chains such as the Hawaiian and Emperor chains in the Pacific. These volcanoes are thought to be surface reflections of so-called ‘hot spots’ or ‘mantle plumes’ in which deep mantle materials are brought through the asthenosphere to the surface (Morgan 1971). Then, as the plates move over these hot spots, sequences of volcanoes of increasing age are left behind on the plate. Comparisons of these tracks with relative plate motions—based on the magnetic observations—provide an idea of the stability of these points and a measure of absolute motion (Minster et al 1974). Some independent geodetic observations of the absolute motions is clearly desirable to try and test these ideas. Also a technique is required that provides a precise determination of relative motions of points away from the plate boundary to supplement near-fault geodimeter observations.

Together these geodetic measurements will aid in determining to what extent the plates deform internally. Is there evidence for stress propagating slowly across the plates or along the boundaries so that seismic activity in different parts of the world may be related, as suggested by Chinnery and Landers (1975)? The geodetic measurements, together with seismic moment estimates of plate motion, will indicate the extent and nature of aseismic slip and lead to a greater understanding of the earthquake source mechanism. Finally, a detailed picture of the way plates move may contribute to the understanding of the mechanisms driving the plates. Satellite and other space techniques will offer a number of possibilities to measure these motions.

6.3.2. **Satellite methods for measuring plate motions.** The geometric method of satellite geodesy, outlined in §4.2, offers a simple way of measuring relative positions of points separated by distances greater than a few hundred kilometres. The attainable precision of these relative positions is comparable to the precision of the measurements themselves, or a few centimetres in the case of laser range observations. This places the goal of measuring instantaneous plate motions within the realm of reality once a sufficient number of precise laser tracking systems becomes operational. The method, being based on simultaneous observations of the satellite, does require a high concentration of tracking stations in any one area under investigation, since one will want to be sure that any internal deformations of any plate are also measured and not confused with relative motions between the plates. This requirement will be in conflict with many other objectives of satellite geodesy which require a more uniform and global distribution of tracking stations. An interesting variant of the
geometric solution has recently been proposed in connection with the Space Shuttle in which the laser is mounted on the spacecraft and ranges to a series of reflectors on the ground in quick succession.

Lunar laser range observations from stations on different plates, such as the McDonald Observatory on the American plate, the Hawaii Observatory on the Pacific plate and the Orroral station on the Australian plate, will also permit a determination of relative motions once all stations reach the centimetre measuring accuracy. With just a few stations, one on each plate, it will not be possible to separate drift of the plate as a whole from local or regional deformations and it will be necessary to introduce mobile stations that observe either satellites or the Moon in conjunction with the principal stations. Long-baseline interferometry is also a potentially powerful method for measuring plate motions as shown by the very good precision obtained in the measurements across the United States (Coates et al 1975). Current plans are to attempt to measure the motion of one of the Hawaiian islands relative to three stations on the North American plate. As the expected relative motion is of the order of 8 cm yr$^{-1}$ (Minster et al 1974) this experiment should provide a good test of the long-baseline interferometry method.

An alternative approach is to use dynamical methods in which the reference frame is provided by the orbit itself. The success of this approach requires that we can model all forces that displace the satellite by more than a few centimetres, a demanding task that is still to be achieved even for the stable orbit of the LAGEOS satellite. In reality, the influence of many orbital perturbations on the station position determination will be quasi-random and much reduced if observations are made over long time intervals. It is this characteristic that in the past has permitted station positions to be computed with accuracies that are comparable to the tracking system accuracies, even though the orbital theories may not always have been of an equal precision. With a network of permanent tracking stations the orbit of the stable satellite will be maintained with high accuracy. Observations from mobile stations then permit the positions of other points to be determined relative to the permanent tracking stations. This raises the question of whether absolute plate motions can be measured in this way. The station positions are determined relative to the orbits which, if all forces are known, can be related to an inertial reference frame. Thus one has to model all perturbations, including the long-period ones, with high accuracy. More troublesome is the fact that the station motion, apart from plate tectonic motions, needs to be known relative to this inertial frame. Thus, if there is any uncertainty in the secular or long-period variations in the rotation of the Earth this could manifest itself as an easterly or westerly drift of all stations which would have nothing to do with plate tectonics. Likewise polar wander cannot be separated from plate motions and the dynamic method also gives relative motions only.

Anderle (1975) has analysed some 10 yr of Doppler observations from the TRANET navigation network in an attempt to determine whether or not some stations have undergone noticeable displacements. In most cases the computed station positions indicate oscillatory and abrupt motions, most probably a consequence of residual orbital perturbations, of revisions of the gravity field and station coordinate solutions and of modifications in orbital theories, rather than of tectonic upheavals.

6.3.3. Vertical motions. In the plate tectonic hypothesis only little attention is given to the vertical motions occurring over the Earth’s surface. This is in contrast with
older geophysical ideas that vertical motions have been dominant throughout the entire geological history of the Earth and that other motions are only of secondary importance. It is perhaps this neglect of the vertical motions in the plate tectonic model that is the cause of opposition to it (for example, Belousov 1972). Recent attempts at discussing the vertical motions within the plate tectonic framework are made by Menard (1973). Vertical motions of the Earth's crust have a variety of origins and occur worldwide. Continental interiors have been subject to slow motions due to the redistribution of surface loads associated with erosive processes and, in some cases, to more rapid motions associated with the deposition and removal of extensive ice sheets. The latter type of movement in North America is at present about 5 mm yr⁻¹. Vertical motions along plate margins comprising ocean ridges and subduction zones have already been mentioned. Along the Himalayas, vertical motions of some 0·5 mm yr⁻¹ occur as the Indian plate is pushed into the Eurasian plate. Similar rates of uplift are observed in the European Alps (see, for example, Schaeer et al 1975). Ocean basins are subject to vertical motions due to the continual cooling of the lithosphere and the loading by sediments. Observations of vertical motion give information on the elastic and anelastic properties of the crust and mantle and possibly of the variations of these parameters with age or other geological factors. These motions, together with horizontal motions, give additional insight into the nature of the mechanism driving the plate. Evidence for the movements lie mainly in the geological records: in the dates and positions of raised or sunken beaches, or in the thickness of sediment layers.

Conventional geodetic measurements give heights relative to sea level and any variation in the heights may be due to a change in the distance of the surface from the Earth's centre of mass, a change in sea level, or both. As uplift occurs, gravity on the surface can be expected to vary as well but at a rate that will depend on the mechanism of uplift. If uplift is due to a density change the gravity variation is a consequence of (i) a change in distance of the surface from the centre of mass, and (ii) a redistribution of mass. Clearly, in these cases a combination of gravity and geodetic levelling measurements are desirable (see Whitcomb (1976) for a review of vertical motion measurements in relation to seismically active areas).

Satellite methods as outlined above provide information on changes in height and horizontal position simultaneously. Geometric satellite geodesy measurements can give relative height changes only. Dynamic methods can give a measure of the motion relative to the centre of mass of the satellite. But as the uplift rates are usually an order of magnitude smaller than horizontal motions it is unlikely that space techniques will give new results in the immediate future. Prospects for monitoring horizontal displacements are certainly more promising.

There are, however, some problems where space techniques may be very helpful. A case in question is the Palmdale bulge, an area of aseismic uplift observed along the San Andreas Fault in Southern California, north of Los Angeles. A maximum vertical motion of 25 cm has been recorded in the interval 1959–1974, with an area of about 200 km by 100 km experiencing lesser uplift (Castle et al 1976). This uplift has apparently not been uniform, much of it having occurred within a few years from 1961–1965. Renewed vertical uplift activity occurred in the area in the early 1970s. Earlier, but less reliable, observations going back to the end of the nineteenth century suggest that several periods of uplift, followed by a collapse of the bulge, may have occurred. Prescott and Savage (1976) discuss the lateral motions in the region. The present centre of activity is near the junction of the
San Andreas Fault with several other faults and at a point where the former has a major east-west deflection. This deflection may present a major obstacle to the movement of the Pacific plate past North America. The uplift of the Palmdale bulge is therefore assumed to be a consequence of strain accumulation along this part of the fault. A continued monitoring of the uplifted area is of importance in assessing the future development of the bulge, whether it will continue to grow or whether maximum uplift has been achieved to be followed by collapse. Periodic geodetic levelling, the technique used until now, is one way of monitoring the motion but, in view of the rapid changes that appear to occur, it may not be the most convenient for detailed studies. The continual monitoring of gravity will also be useful although the interpretation of such measurements in terms of uplift is ambiguous. Differential long-baseline interferometry may be one of the best ways of monitoring the uplift in a continuous manner. With one station away from the uplift area and a second within it, precise height-difference measurements can be made if the two stations are not separated by a large distance so that many observational errors, such as ionospheric refraction, are common to the two ends of the baseline. The short baseline also permits the two antennae to be linked electrically and the two sets of signals can be compared directly without requiring independent frequency controls. An experiment along these lines is being carried out by the Jet Propulsion Laboratories of the California Institute of Technology using the Goldstone Deep Space antennae and a mobile antenna in the Palmdale area.

6.4. Tides

The Earth's tides show themselves in numerous ways, as variations in sea level, in fluctuations in gravity in the Earth's surface, as irregularities in the Earth's rotation or as perturbations in the motion of close Earth satellites. Table 4 summarises some orders of magnitude of these effects. The tidal deformations must be known for a number of satellite studies, for example to reduce the altimeter observations to mean geoid heights, to analyse orbit perturbations for the gravity field, or to study meteorological or oceanographic forces that modify the Earth's rotation. In addition, the tidal phenomena are of some interest in their own right. The discussion on tides is conveniently separated into solid Earth, oceanic and atmospheric tides.

6.4.1. Earth tides. The Earth, not being perfectly rigid, deforms under the solar and lunar gravitational attractions. This results in periodic changes in gravity, in the stress state and in the form of the planet's surface. These deformations are

<table>
<thead>
<tr>
<th>Phenomena</th>
<th>Amplitude</th>
<th>Period</th>
</tr>
</thead>
<tbody>
<tr>
<td>Radial surface deformation</td>
<td>30-40 cm</td>
<td>12 h, 24 h</td>
</tr>
<tr>
<td>Gravity changes</td>
<td>&lt;0.1 mgal</td>
<td>12-24 h</td>
</tr>
<tr>
<td>Change in length of day</td>
<td>0.3 ms</td>
<td>14 d</td>
</tr>
<tr>
<td></td>
<td>0.2 ms</td>
<td>28 d</td>
</tr>
<tr>
<td>Perturbations in satellite orbits</td>
<td>0.3-3'</td>
<td>10-300 d</td>
</tr>
</tbody>
</table>
conveniently defined with the aid of the Love numbers introduced by A E H Love in 1909.

The potential $U$ of the gravitational attraction at $r$ due to a mass $m^*$ at $r^*$ is given by (2.12). This potential can be expanded into a series of Legendre polynomials $P_{l0}(\cos S)$ as:

$$U(r) = \frac{Gm^*}{r^*} \sum_{l=2}^{\infty} \left( \frac{r}{r^*} \right)^l P_{l0}(\cos S)$$

(6.7)

where $S$ is the geocentric angle between $r$ and $r^*$. It is given by:

$$\cos S = \frac{rr^*}{|r^*|}$$

(6.8)

The $r^*$, $\phi^*$, $\lambda^*$ and $r$, $\phi$, $\lambda$ are the spherical coordinates of the mass $m^*$ at $r^*$ and of the position $r$ at which $U$ is evaluated. With (6.8) and the addition theorem of spherical harmonics (see, for example, Jeffreys and Jeffreys 1962 p646), the polynomials can be expressed by:

$$P_{l0}(\cos S) = \sum_{m=-l}^{l} (2-\delta_{0m}) \frac{(l-m)!}{(l+m)!} P_{lm}(\sin \phi) P_{lm}(\sin \phi^*) \cos m(\lambda-\lambda^*)$$

and

$$U(r) = \frac{Gm^*}{r^*} \left( \frac{r}{r^*} \right)^l \sum_{m=-l}^{l} (2-\delta_{0m}) \frac{(l-m)!}{(l+m)!} P_{lm}(\sin \phi) P_{lm}(\sin \phi^*) \cos m(\lambda-\lambda^*)$$

The motion of the attracting mass (Moon or Sun) is more conveniently expressed by orbital elements rather than by the $r^*$, $\phi^*$, $\lambda^*$ and the time derivatives $\dot{r}^*$, $\dot{\phi}^*$, $\dot{\lambda}^*$. A convenient choice, useful for the subsequent discussion of the tidal evolution of the lunar orbit, consists of the instantaneous Keplerian elements $\kappa_i$ of the Moon (or Sun). The appropriate transformation from spherical coordinates to these elements is the same as discussed in §2. The result is:

$$\left( \frac{1}{a^*} \right)^{l+1} P_{lm}(\sin \phi^*) \begin{bmatrix} \cos m\phi^* \\ \sin m\lambda^* \end{bmatrix} = \left( \frac{1}{a^*} \right)^{l+1} \sum_{p=0}^{l} F_{imp}(\kappa_i) \sum_{q=-\infty}^{\infty} G_{mpq}(\epsilon^*) \begin{bmatrix} \cos v_{impq}^* + i \sin v_{impq}^* & 1-m \text{ even} \\ \sin v_{impq}^* - i \cos v_{impq}^* & 1-m \text{ odd} \end{bmatrix}$$

(6.9)

with $v_{impq}^* = (l-2p)\omega^* + (l-2p+q)M^* + m(\Omega^* - \theta)$. As the eccentricities of both the lunar and Earth orbits are small, the summation over the index $q$ needs to be carried out only over a small number of terms, i.e. $q = 0, \pm 1, \pm 2$. Also, because the factor $(R/a^*)^{l+1}(1/60)^l$ is small for the Moon and very much smaller for the Sun, only terms with $l = 2$ will be important. With the transformation (6.9) the potential $U(r)$ becomes:

$$U(r) = \frac{Gm^*}{a^*} \left( \frac{r}{a^*} \right)^2 (2-\delta_{0m}) \frac{(2-m)!}{(2+m)!} P_{2m}(\sin \phi) \sum_{p} F_{2mp}(\kappa_i)$$

$$\times \sum_{q} G_{2pq}(\epsilon^*) \begin{bmatrix} \cos 2m \text{ even} \\ \sin 2m \text{ odd} \end{bmatrix} (v_{2mpq}^* - m\lambda).$$

(6.10)
In a first approximation, the only time-dependent variable is the Earth’s rotation $\theta$ and the potential will exhibit three main periodicities: long periods when $m=0$, nearly diurnal periods when $m=1$, and nearly semi-diurnal periods when $m=2$. In a second approximation, the mean anomalies of the Moon or Sun also vary with time and additional periods occur in the potential for different values of the indices $p$ and $q$. These group about the three fundamental periods. Further periods result due to the small variations in $\omega_*$ and $\Omega_*$ with time due to the attraction of the Earth’s oblateness and of the Sun on the Moon. The total spectrum of the potential is rich indeed, particularly when the total tide-raising potential, that due to the Sun and Moon, is considered. The principal terms in the tidal potential are the semi-diurnal tides with $lmpq=2200$; the lunar tide is usually referred to as $M_2$ following the notation introduced by G.H. Darwin, and the solar tide as $S_2$. The main diurnal tides are $K_1$ ($lmpq=2110$), of combined lunar and solar origin, $O_1$ ($lmpq=2100$) of lunar origin and $P_1$ ($lmpq=2100$) of solar origin.

The tide-generating potential (6.7) can be expressed as:

$$U_i(r) = r^l S_l(\phi, \lambda)$$

where $S_l(\phi, \lambda)$ is a surface harmonic. At $r=R$ this potential is:

$$U_l(R) = \left(\frac{R}{r}\right)^l U_i(r).$$

The response of the Earth to this applied potential is assumed to be linear and observations verify that this is so. The deformations and additional potential $\Delta U_l(R)$ are also harmonic in degree $l$. Thus:

$$\Delta U_l(R) = k_l U_l(R) = k_l \left(\frac{R}{r}\right)^l U_i(r).$$

Continuing this potential outwards, using Dirichlet’s theorem, gives:

$$\Delta U_l(r) = k_l \left(\frac{R}{r}\right)^{2l+1} U_i(r).$$

The radial $d_r$ and tangential $d_t$ deformations are defined as:

$$d_{r,i}(R) = \frac{h_i}{g} U_i(R) e_r = \frac{h_i}{g} \left(\frac{R}{r}\right)^l U_i(r)e_r$$

$$d_{t,i}(R) = \frac{l_i}{g} \nabla U_i(R)e_t$$

(6.12)

where $e_r$ and $e_t$ are unit vectors in the radial and tangential directions. The $h_i$, $k_i$, $l_i$ are the Love numbers of degree $l$ defining the change in potential and the deformations at $r=R$.

Parameters similar to these Love numbers can be introduced to define the Earth’s deformation when subject to a surface load. Consider a surface layer, harmonic in degree $s$. The gravitational potential of this load at the surface is $U_s(R)$. After deformation the additional potential is defined as $k_s U_s(R)$ and the total potential (1 + $k_s$) $U_s$. $k_s$ is called the load Love numbers. Analogous load Love numbers $h_s$, $l_s$ can be introduced to describe the deformation of the surface by this load. Deformations associated with tangential stresses applied to the Earth’s surface can also be defined with the aid of additional Love numbers. Theoretical values for
Table 5. Some Love numbers and load deformation numbers of degree n.

<table>
<thead>
<tr>
<th></th>
<th>$h_n$</th>
<th>$l_n$</th>
<th>$k_n$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Love numbers</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$n=2$</td>
<td>0.61</td>
<td>0.083</td>
<td>0.30</td>
</tr>
<tr>
<td>$n=3$</td>
<td>0.29</td>
<td>0.015</td>
<td>0.094</td>
</tr>
<tr>
<td>Load deformation</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$n=2$</td>
<td>-1.0</td>
<td>0.030</td>
<td>-0.31</td>
</tr>
<tr>
<td>$n=3$</td>
<td>-1.5</td>
<td>0.074</td>
<td>-0.20</td>
</tr>
</tbody>
</table>

These Love numbers can be computed from the seismic models of the Earth’s interior (Longman 1966, Farrell 1972). Table 5 summarises some results.

Substituting (6.9) into (6.10) gives the tidal potential of the Earth as (Kaula 1964):

$$\Delta U(t) = \frac{Gm^*}{a^*} \sum_{m=0}^{\infty} \left( \frac{R}{r} \right)^{l+1} \left( \frac{R}{r} \right)^{l} k_l(2 - \delta_{0m}) \frac{(l-m)!}{(l+m)!} P_l m (\sin \phi)$$

$$\times \sum_{p=0}^{l} F_{\text{imp}}(I^*) \sum_{q=-\infty}^{\infty} G_{\text{imp}}(e^*) \begin{cases} \cos l-m & \text{even} \\ \sin l-m & \text{odd} \end{cases} \left( \psi_{\text{imp}}^* - m\lambda \right). \quad (6.13)$$

This expression assumes an instantaneous or elastic response. The consequence of anelasticity is to delay the response of the Earth by an amount $\Delta t$. The maximum deformation is reached at a time $\Delta t$ after the Moon or Sun has passed through the observer’s meridian. During this interval, the Earth has rotated through an angle $\theta \Delta t$ while the Moon has moved through $n \Delta t$, $n$ being the mean lunar motion. Viewed from space (figure 15), the tidal bulge is ahead of the Moon by approximately $(\theta - n) \Delta t \approx \theta \Delta t$ since $n \ll \theta$. To introduce this delay it is only necessary to modify the argument $\psi_{\text{imp}}^*$ to $\psi_{\text{imp}}^* + \epsilon_{\text{imp}}$ where:

$$\epsilon_{\text{imp}} = (l-2p)\omega^* + (l-2p+q)M^* + m(\Omega^* - \theta) \Delta t$$

$$\approx (l-2p+q)n^* - m\theta \Delta t. \quad (6.14)$$

Figure 15. The Earth’s tidal bulge. The upper figure is for an elastic response and the lower figure corresponds to a delayed response.
For a lagged response $\Delta t$ is negative. $\epsilon_{mpq}$ may be positive or negative depending on the sign of $(l-2p+q)n^*-m\hat{\theta}$.

Observational evidence for the low-degree Love numbers comes from measurements of the Earth's response to external potentials. The total perturbing potential due to the Moon's attraction and subsequent terrestrial deformation is:

$$U_1(r) = [1 + k_1(R/r)^{2l+1}] U_l(r).$$

Gravity at the surface of the Earth changes by:

$$\Delta g = -\frac{\partial U_1'}{\partial r} + \frac{\partial g}{\partial r} d_{r,t},$$

where the first term is the contribution from the perturbing potential and the second term results from the change in distance between the surface and the centre of mass of the Earth. With the above expressions for $U_1'(r)$ and $d_{r,t}$, at $r = R$, and for the principal term $l = 2$:

$$\Delta g_2 = -\frac{2}{R} (1 - \frac{3}{2}k_2 + h_2) U_2(r).$$

Gravity at the Earth's surface varies periodically with the same spectrum as $U_2(r)$ and with amplitudes that are proportional to the linear relation $(1 + h_2 - \frac{3}{2}k_2)$. Typically, these changes in gravity are smaller than 0.1 mgal. If the observed gravity lags the theoretical gravity by an angle $\Gamma_{mpq}$ due to anelasticity, then the delay between the tide-raising potential and the response follows as:

$$\epsilon_{2mpq} = \frac{1 - \frac{3}{2}k_2 + h_2}{-\frac{3}{2}k_2 + h_2} \Gamma_{2mpq}.$$

Similarly, the direction of the vertical relative to the Earth's surface changes according to:

$$\frac{1 + k_2 - h_2}{Rg} \frac{\partial U_2}{\partial \phi},$$

in the meridian, and by:

$$\frac{1 + k_2 - h_2}{Rg \cos \phi} \frac{\partial U_2}{\partial \lambda},$$

along small circles of latitude. These deflections can be measured by sensitive horizontal pendulums. Any delay $\Gamma_{2mpq}$ in the pendulum response, due to anelasticity, relates to $\Gamma_{2mpq}$ according to:

$$\epsilon_{2mpq} = \frac{1 + k_2 - h_2}{k_2 - h_2} \Gamma_{2mpq}.$$

Observations of the tidal deflections of the vertical relative to the Earth's rotation axis and of elastic strains provide further relations between $h_2$, $k_2$ and $l_2$. The subject of Earth tides has been recently reviewed by Slichter (1972), Jobert (1973) and Melchior (1974). Baker and Lennon (1976), Ostrovsky (1976), Melchior et al (1976) and Pertsev (1977) present results for the factors $1 + h_2 - \frac{3}{2}k_2$ and $1 + k_2 - h_2$ from tide observations in Europe and Asia. These factors, and the corresponding lags, show considerable dispersion that is mainly a consequence of the interference of the ocean tide with the body tide. Even observations in the middle of continents...
appear to be perturbed by the ocean tide (Pertsev 1969, Kuo et al 1970). Observations
of the solid Earth phase lags are generally unsatisfactory due to instrumental
problems and the ocean loading contributions. Tide observations in the interior of
the USSR indicate a lag of about 0.5°, part of which can be attributed to ocean loading
even though the stations lie far from the coast. The tide lags for western Europe
can also be attributed to tidal loading and the corrected lags are small (Melchior
and Baker and Lennon (1976) confirm that the solid Earth lag is a small fraction
of a degree.

The angle \( \epsilon_{impq} \) represents the lag between stress and strain and as such is
a measure of the internal friction (or \( Q \)) of the Earth. That is:

\[
\tan \epsilon_{impq} \approx \frac{1}{Q} = \frac{1}{2\pi} \frac{\Delta E}{E}
\]

where \( \Delta E \) is the energy lost during one cycle of the applied force and \( E \) is the peak
elastic energy stored in the cycle. For \( \epsilon \approx 0.2^\circ \), \( Q \approx 300 \). This is the \( Q \) of a shear
wave of degree 2. At the higher seismic frequencies the spheroidal free oscillations
of degree 2 suggest similar values of \( Q \) (Anderson and Hart 1978).

6.4.2. Ocean tides. The subject of ocean tides has been reviewed recently by Cart-
wright (1977) and we consider only certain aspects of it here. The total ocean tide
contains a large number of frequencies corresponding to the combination of indices
\( lmpq \). Any one such component is denoted by the subscript \( \beta \). For each component
the tide is given at any position on the Earth by an amplitude \( \xi_\beta(\phi, \lambda) \) and a phase
\( \chi_\beta(\phi, \lambda) \), both of which vary over the Earth’s surface. Thus:

\[
\xi_\beta(\phi, \lambda; T) = \xi_\beta(\phi, \lambda) \cos [2\pi f_\beta T - \chi_\beta(\phi, \lambda)]. \tag{6.15}
\]

The phase \( \chi_\beta \) is expressed with respect to the Greenwich meridian. \( T \) is with respect
to 0.00 h 1 January 1900 in this definition.

In the discussion of tidal influences on satellite orbits or the Earth’s rotation,
we are concerned only with the very-long-wavelength components in the tide.
This suggests that a spherical harmonic expansion of the ocean tide is appropriate.
That is, \( \xi_\beta \cos \chi_\beta \) and \( \xi_\beta \sin \chi_\beta \) are expanded as follows:

\[
\xi_\beta^\circ \cos \chi_\beta = \sum_{s=1}^{\infty} \sum_{t=0}^{s} (a_{\beta, st'} \cos t\lambda + b_{\beta, st'} \sin t\lambda) P_{st}(\sin \phi)
\]

\[
\xi_\beta^\circ \sin \chi_\beta = \sum_{s=1}^{\infty} \sum_{t=0}^{s} (a_{\beta, st'} \cos t\lambda + b_{\beta, st'} \sin t\lambda) P_{st}(\sin \phi).
\]

On the continents \( \xi_\beta^\circ = 0 \). Substituting these into (6.15) (Lambeck 1977):

\[
\xi_\beta = \sum_{s=1}^{\infty} \sum_{t=0}^{s} D_{\beta, st^\pm} \cos (2\pi f_\beta T - t\lambda - \epsilon_{\beta, st^\pm}) P_{st}(\sin \phi) \tag{6.16}
\]

with

\[
D_{\beta, st^\pm} \cos \epsilon_{\beta, st^\pm} = \frac{1}{2}(a_{\beta, st'} \pm b_{\beta, st'})
\]

\[
D_{\beta, st^\pm} \sin \epsilon_{\beta, st^\pm} = \frac{1}{2}(a_{\beta, st'} \pm b_{\beta, st'}).
\]

The summation of the form \( \sum_{\pm} D^\pm \cos (\alpha \pm \beta - \epsilon^\pm) \) implies
\( D^+ \cos (\alpha + \beta - \epsilon^+) \) +
\( D^- \cos (\alpha - \beta - \epsilon^-) \).
Knowledge of the world's ocean tides is rather sparse. Long records of tides exist along many parts of the shoreline and are extremely valuable for predicting the tides locally but these tides are most often influenced by the coastline geometry and by the shallow coastal sea floor where frictional forces significantly modify the tides from their open sea behaviour. The best observations on the open ocean tides come from island stations and show that the ocean tide does not in general exceed more than a metre. The limited availability of unperturbed tidal stations means that the global tidal patterns cannot be established with reliability from measurements alone and one has to resort to theory for estimating the global tides. The development of bottom pressure gauges for measuring the tides in the open sea has led to important improvements in the knowledge of regional ocean tides, as demonstrated by Munk et al (1970) and Filloux (1971) for tides off the Californian coast, by Luther and Wunsch (1975) for the Central Pacific, and D E Cartwright and colleagues for the North Atlantic. But the information is still too sparse to be useful for global models. Thus the prediction of open ocean tides is largely based on theory. Numerous global $M_2$ tide models have been published in recent years from which the coefficients $D_{M_2,e^{±r}}, \epsilon_{M_2,e^{±r}}$ can be estimated. The results have been discussed by Lambeck (1977). Once known, the potential outside the Earth due to the ocean tide follows as:

$$\Delta U_p(r) = 4\pi G R \rho_w \sum_{l=2}^{\infty} \sum_{m=0}^{l} \sum_{p=0}^{l} \sum_{q=-\infty}^{\infty} \sum_{j=0}^{l} \sum_{g=-\infty}^{\infty} \frac{1+k_{e^{±r}}}{2^s+1} \frac{(R)}{r}^{s+1} \times D_{p,e^{±r}} \cos(2\pi f_p T + t\lambda - \epsilon_{p,e^{±r}}) P_{st}(\sin \phi)$$

where $\rho_w$ is the density of water. The factor $(1+k_{e^{±r}})$ allows for the Earth's elastic yielding under a variable load.

6.4.3. Satellite observations of tide parameters. The tidal potentials perturb the motion of close Earth satellites, perturbations that are of interest in that they permit some of the tidal parameters to be estimated from precise analyses of orbits and because tidal effects on the lunar motion represent a special case. Also these tidal effects should be known in the evaluation of any other orbital perturbations. To study these perturbations the coordinates $r, \phi, \lambda$ in the potentials (6.13) or (6.17) are transformed into the Keplerian elements $\kappa_1$ of the satellite, using the transformation (6.9). The potential now becomes:

$$\Delta U_I(r) = \sum_{l=2}^{\infty} \sum_{m=0}^{l} \sum_{p=0}^{l} \sum_{q=-\infty}^{\infty} \sum_{j=0}^{l} \sum_{g=-\infty}^{\infty} \Delta U_{impqg}^{(a)}$$

$$\Delta U_{impqg}^{(b)} = k_l \left( \frac{R}{a^*} \right)^l \left( \frac{R}{a} \right)^{l+1} \frac{Gm_*}{a^*} (2-\delta_{0m}) \frac{(l-m)!}{(l+m)!} F_{imp}(I^*) F_{imp}(I)$$

$$\times G_{ipq}(e^*) G_{ipq}(e) \cos (\epsilon_{impq} - \epsilon_{impq} + \epsilon_{impq})$$

where the lag $\epsilon_{impq}$ is defined by (6.14).

The orbital accelerations are found by substituting $\Delta U_{impqg}$ into the Lagrange equations (2.6). For the inclination of the satellite orbit, for example:

$$\frac{dI}{dt} \bigg|_{impqg} = \frac{Gm_*[(l-2p) \cos I - m]}{na^2(1-e^2)^{3/2}} \left( \frac{R}{a^*} \right) \left( \frac{R}{a} \right)^{l+1} (2-\delta_{0m}) \frac{(l-m)!}{(l+m)!}$$

$$\times F_{imp}(I) F_{imp}(I^*) G_{ipq}(e^*) G_{ipq}(e) \sin (\epsilon_{impq} - \epsilon_{impq} + \epsilon_{impq}).$$
The integration of this, and similar expressions for the other elements, is carried out either numerically or analytically by assuming that the only time-varying elements are (i) the mean motions of the satellite and the tide-raising body, and (ii) the linear rates in $\Omega$, $\omega$, $\dot{\Omega}$, $\dot{\omega}$ due principally to interactions of the motions with the Earth's flattening and the solar attraction (see, for example, Kaula 1966, Gaposchkin 1973). For example, the results for the variation in semi-major axis and inclination are:

$$\Delta a_{mpq} = \frac{2}{n a} \frac{(l-2j+g)}{\tilde{v}_{mpq} - \tilde{v}_{l,mj}} A_{lmpqj} \cos \gamma_{lmpqj}$$

$$\Delta i_{mpq} = \frac{1}{n a^2 (1-e^2)^{1/2} \sin I} \frac{[\cos I (l-2j)-m]}{\tilde{v}_{mpq}^* - \tilde{v}_{l,mj}} A_{lmpqj} \cos \gamma_{lmpqj}$$

(6.19)

with

$$A_{lmpqj} = k_i \left( \frac{R}{r} \right)^{l+1} \frac{G m^*}{a} (2-\delta m) \frac{(l-m)!}{(l+m)!} F_{lmp}(I^*) F_{lmj}(I) G_{1pq}(e^*) G_{1jg}(e)$$

and

$$\gamma_{lmpqj} = \tilde{v}_{mpq}^* - \tilde{v}_{l,mj}.$$}

These and corresponding perturbations in the other elements of the satellite orbit have periods longer than one day only for those terms for which the combination of indices $l-2j+g$ vanishes. For all other terms not satisfying this condition, the perturbations will be of small amplitude since their frequencies, entering into the divisor of such equations as (6.18), are large. The frequencies of the long-period tidal terms are governed by both the lunar and satellite motions around the Earth and the spectrum will differ quite considerably from that observed at the Earth's surface in, for example, gravity. Also, as the amplitude of the orbital perturbations depends upon frequency, tidal terms observed at the Earth's surface to be of small amplitude may become important in the satellite perturbation spectrum if the lunar and satellite motions become commensurable when $\tilde{v}_{mpq}^* \approx \tilde{v}_{l,mj}$. Thus, by a careful selection of elements for an orbit, different fundamental tidal frequencies can be made to have more or less important effects on the satellite. Table 6 summarises the theoretical perturbations in the inclination of the satellite STARLETTE and GEOS 1. The $S_2$ tide perturbs the satellite motion considerably more than does the $M_2$ tide, while at the Earth's surface it is only about one-half of $M_2$. In the case of GEOS 1, the $K_1$ tide perturbation in inclination is about three times greater than the perturbation due to $M_2$, while on the Earth's surface it represents only about 60% of $M_2$ in amplitude.

**Table 6. Perturbations in orbits of two satellites due to the Earth's tides.**

<table>
<thead>
<tr>
<th>Tide</th>
<th>Period (d)</th>
<th>$\Delta i$ (arcsec)</th>
<th>$\Delta \Omega$ (arcsec)</th>
<th>Period (d)</th>
<th>$\Delta i$ (arcsec)</th>
<th>$\Delta \Omega$ (arcsec)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$M_2$</td>
<td>11</td>
<td>0.19</td>
<td>0.20</td>
<td>12</td>
<td>0.29</td>
<td>—</td>
</tr>
<tr>
<td>$S_2$</td>
<td>36</td>
<td>0.21</td>
<td>0.43</td>
<td>56</td>
<td>0.41</td>
<td>0.47</td>
</tr>
<tr>
<td>$K_2$</td>
<td>46</td>
<td>0.11</td>
<td>0.16</td>
<td>80</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>$O_1$</td>
<td>11</td>
<td>0.09</td>
<td>0.05</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>$P_1$</td>
<td>60</td>
<td>0.18</td>
<td>0.22</td>
<td>85</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>$K_1$</td>
<td>90</td>
<td>0.95</td>
<td>1.47</td>
<td>160</td>
<td>2.0</td>
<td>—</td>
</tr>
</tbody>
</table>
The amplitude of the orbit perturbations are proportional to the Love number $k_2$ and the phases lag the direct attraction of the Sun or Moon on the satellite by an amount $\epsilon \omega_{\text{mpg}}$. If the solid tide potential only were of importance, any difference in the Love number and phase lags from theoretical values would be small indeed but ocean tides also perturb the satellite motions and, if not allowed for in the orbital theory, introduce significant variations in the tidal effective Love numbers with frequency. The ocean tide perturbations are found by applying the transformation (6.9) to the ocean tide potential and substituting the resulting expression into the Lagrangian equations. For the satellite inclination, for example:

$$\frac{dI}{dt}_{\beta,stuw} = \frac{4\pi G R^2 \rho_w}{a} \sum_{s=0}^{\infty} \sum_{t=0}^{\infty} \sum_{v=-\infty}^{\infty} \sum_{w=-\infty}^{\infty} \frac{1 + k_s'}{2s+1} \left(\frac{R}{a}\right)^s$$

$$\times \frac{1}{na^2(1-e^2)^{1/2} \sin I} D_{\beta,stuw} F_{stuw}(I) G_{stuw}(e) [(s-2u) \cos I - t]$$

$$\times \left[ \begin{array}{c} \pm \sin \gamma_{\beta,stuw} \\
\cos \gamma_{\beta,stuw} \end{array} \right]$$

with

$$\gamma_{\beta,stuw} = \omega_{stuw} \pm 2\pi \omega_{\beta} + \epsilon_{\beta,stuw}.$$ 

Integrating these expressions with respect to time, assuming that all time-dependent variables are contained in $\gamma_{\beta,stuw}$, gives for $a$ and $I$, for example:

$$\Delta a_{\beta,stuw} = \frac{2}{na} A_{\beta,stuw} \sin I \frac{s-2u+v}{\gamma_{\beta,stuw}^{\pm}} \left[ \begin{array}{c} \pm \cos \\
\sin \gamma_{\beta,stuw} \end{array} \right]$$

$$\times \left[ \begin{array}{c} \pm \cos
\sin \gamma_{\beta,stuw} \end{array} \right]$$

with

$$A_{\beta,stuw} = \frac{4\pi G R^2 \rho_w}{a} \left(1 + k_s'\right) \left(\frac{R}{a}\right)^s D_{\beta,stuw} F_{stuw}(I) G_{stuw}(e).$$

Long-period perturbations (longer than one day) occur when $\gamma_{\beta,stuw}$ does not contain the sidereal angle $\theta$. Only those coefficients $D_{\beta,stuw}$ of the semi-diurnal tide ($m=2$) and $s, t=2,3, 4,5, 6,7, \ldots$ give rise to the principal long-period terms. Other long-period terms are caused by the coefficients $D_{\beta,stuw}$ with $m=2$ and $s, t=2,3, 5,7, 9,11; \ldots$; but now $\nu = \pm 1$ and the amplitudes of these terms are smaller than those of previous coefficients by a factor $e$. Thus, unless the satellite orbit is very eccentric these perturbations are quite small. Similarly, only coefficients $D_{\beta,stuw}$ with $s, t=2,1; 4,1; 6,1; \ldots$; of the diurnal tides ($m=1$) give rise to long-period perturbations with $\nu=0$. We note that the amplitudes of the perturbations are proportional to $(R/a)^{s+1}$ so that the coefficients with $s>4$ also tend to be small. Furthermore, we note that the perturbations due to $D_{\beta,21}$ or $D_{\beta,22}$ have the same dependence on the orbital elements of the satellite as the perturbation due to the solid tide of the same frequency $f_\beta$ and that the two cannot be separated. The $D_{\beta,stuw}$ occur only in the ocean tide potential. Due to its different inclination function
$F_{stw}(I)$, it can be separated from the leading term in the ocean tide expansion, even though the two have the same frequency, if at least two orbital elements or two different orbits of close Earth satellites are available for analysis. Thus, from the orbital perturbation analyses one cannot obtain a great deal of information on the detailed structure of the ocean tide. Yet this satellite information is of considerable interest in the study of the Moon's motion and in providing an estimate of the total amount of tidal energy dissipated in the oceans (Lambeck 1975, 1977).

6.4.4. Tidal dissipation. The question of tidal dissipation and its consequences on the lunar orbit and Earth's rotation has been much discussed since the work by G H Darwin early in this century. Geophysicists, astronomers and oceanographers have contributed to the understanding and confusion of the subject and most recently satellite geodesists have entered the fray. This is as much a reflection of a fascinating subject as an indication of a problem of some importance in understanding the origin and dynamical evolution of the Moon. The subject has recently been reviewed by Lambeck (1977) and we can only enter into a few salient aspects here that revolve around the contributions that satellite observations can make. Equation (6.18) gives the tide potential acting on a satellite. Such a satellite could be the Moon in which case we are concerned with the perturbations in the Moon's orbit due to the potential of the tide raised on the Earth by the Moon itself. These effects follow directly from equation (6.19) where, for the satellite elements $\kappa_s$, we substitute for the Moon's elements $\kappa_m$. Only secular perturbations will be important and these occur when $\dot{\psi}_{1pq} - \dot{\psi}_{1pq} = 0$, or when $p = j$ and $q = g$. Then, dropping the asterisks denoting lunar or solar quantities:

$$
\begin{align*}
\Delta t_{mpq} &= 2K_{lm}[F_{tmp}(I)]^2([G_{pq}(I)]^2(l - 2p + q) \sin \epsilon_{mpq} \\
\dot{t}_{mpq} &= K_{lm} \frac{(1 - e^2)^{1/2}}{a e} \frac{F_{tmp}(I)}{[G_{pq}(I)]^2} \sin \epsilon_{mpq} \\
&\times [(1 - e^2)^{1/2}(l - 2p + q) - (l - 2p)] \sin \epsilon_{mpq}
\end{align*}
$$

with

$$
K_{lm} = \frac{G \mu_k}{[G(M + m)a]^{1/2}} \left( \frac{R}{a} \right)^{2l-1} \frac{(l - m)!}{(l + m)!} (2 - \delta_{0m}).
$$

For the ocean tide, the procedure is the same, but with the potential (6.17) instead of (6.13). Then:

$$
\begin{align*}
\Delta \dot{\alpha}_{stuv} &= K'_{\alpha, stuv} (s - 2u + v) \left[ \sin \left( \frac{s - t}{2} \right) \right]^{s - t \text{ even}} \cos \frac{s - t}{2} \text{ odd} \\
\dot{\alpha}_{stuv} &= K'_{\alpha, stuv} \frac{(1 - e^2)^{1/2}}{ae} (1 - e^2)^{1/2}[(s - 2u + v) - (s - 2u)] \\
&\times \left[ \sin \left( \frac{s - t}{2} \right) \right]^{s - t \text{ even}} \cos \frac{s - t}{2} \text{ odd} \\
\frac{dI}{dt}_{\alpha, stuv} &= K'_{\alpha, stuv} \left( \frac{1 - e^2}{2} \sin I \right) \left( \frac{1 - e^2}{2} \sin I \right) [s - t \text{ even} \left( \frac{1 - e^2}{2} \sin I \right) \text{ odd} + s - t \text{ odd} \left( \frac{1 - e^2}{2} \sin I \right) \text{ even}]
\end{align*}
$$

(6.22)
with
\[
K'_{s,tuv} = \frac{3GMF_{stw}(I)G_{stw}(e)}{R[G(M + m)a^{1/2}]} \frac{1}{2s + 1} \left( \frac{R}{a} \right)^3 D_{s,tuv}^{4,4}.
\]

If the ocean parameters are known, the secular perturbations in the lunar orbit can be directly evaluated. Only the \( s, t = 2,2 \) (for semi-diurnal tides) will be important as other terms in the ocean expansion lead to a zero net contribution when averaged over a period of the lunar motion.

The angular momentum of the Earth–Moon system consists of three parts:
(i) that associated with the orbital motions of the two bodies about their centre of mass, given by:
\[
H_1 = \frac{Mm_M}{M + m_M} a^2 (1 - e^2)^{1/2} n.
\]
(ii) that associated with the Earth’s spin:
\[
H_2 = C\theta
\]
and (iii) that associated with the Moon’s spin:
\[
H_3 = C_M \theta_M.
\]
Since \( C_M/C \simeq m_M R^2 / M R^2 < 10^{-8} \) and \( \theta_M / \theta \simeq 1/27 \), the last contribution is negligible.

Conservation of angular momentum of the Earth–Moon system requires that:
\[
H_1 \cos I + H_2 = \text{constant}. \tag{6.23}
\]

The rotational energy associated with the Earth’s spin is:
\[
E_1 = \frac{1}{2} C\theta^2
\]
and the energy associated with the orbital motion is:
\[
E_2 = \frac{1}{2} a^2 n^2 \frac{Mm_M}{M + m_M} - \frac{GMm_M}{a} = \frac{GMm_M}{2a}.
\]
This includes potential and kinetic energies. Without dissipation:
\[
E = E_1 + E_2 = \text{constant} = \frac{1}{2} \left( C\theta^2 - \frac{GMm_M}{a} \right). \tag{6.24}
\]
The conservation laws require that there are no secular changes in \( a, e, I \) but if the tidal response of the Earth is not elastic, dissipation of energy occurs. From (6.23) it follows that:
\[
\dot{\theta}_T = \frac{1}{C_M + m_M} \frac{Mm_M}{a^2} \frac{n}{\dot{a}} - \frac{3}{2} \cos I \left( \frac{\dot{n}}{n} + e \cos I \dot{e} + \sin I \frac{dI}{dt} \right) \tag{6.25}
\]
with, from Kepler’s law:
\[
\dot{n} = -\frac{3}{2} \frac{n}{a} \dot{a}. \tag{6.26}
\]
The \( d_e, \dot{e}, \frac{dI}{dt} \) are given by (6.21) or (6.22). The action of tidal dissipation therefore causes secular changes in the lunar orbit that would otherwise not occur, as
well as causing a secular tidal acceleration $\dot{\theta}$ of the Earth. The rate at which the tidal energy is dissipated is, with (6.24):

$$\frac{dE}{dt} = C \dot{\theta} \dot{\theta} - \frac{1}{2} m n a^2 \dot{n}.$$  (6.27)

The above analysis considers only tides raised by the Moon on the Earth. Solar tides raised on the Earth must be added as should tides raised by the Earth on the Moon but the latter are generally considered to be less important (e.g. Kaula 1964).

There are three approaches to estimating the above mentioned quantities. The first is from the analysis of astronomical observations of the Sun's and Moon's motion. The second method is to evaluate the tidal energy dissipation in the world's oceans. The third method is to apply the results for tidal parameters estimated from close Earth satellite perturbations directly to the lunar problem since the parameters that cause short-period perturbations in the satellite orbits also describe the secular evolution of the lunar orbit. Astronomical observations of the accelerations are of several types and have been discussed recently by Morrison and Ward (1975), Muller and Stephenson (1975), Muller (1976) and Lambeck (1977). The evaluation of tidal energy dissipation was first attempted by Jeffreys (1920) and Heiskanen (1921). This approach has generally been considered to be less precise than the first, but recent calculations (Lambeck 1975, 1977) indicate that this is no longer the case. The third method has several advantages: (i) no assumptions need be made as to where dissipation occurs in the Earth, (ii) it enables a separation of the amount of dissipation that occurs in the Earth from that occurring in the Moon, should the latter be significant, by comparing satellite results with astronomical accelerations, and (iii) it enables the accelerations to be estimated separately for each tidal frequency of both lunar and solar tides. Results obtained from the satellite analysis, although only preliminary, are in essential agreement with those obtained by the other two methods (Cazenave et al 1977, Daillet 1977). Table 7 summarises recent astronomical estimates for $\dot{n}$ and $\dot{\theta}$ and represent average values for the last 3000 yr, being based on both telescope observations and studies of ancient eclipse records. Also indicated is the contribution to $\dot{n}$ by the $M_2$ tide only. Satellite and ocean tide model results for $\dot{n}$ are also summarised and the agreement is satisfactory.

The energy dissipated in the global ocean, $M_2$ tide averaged over one cycle, follows from (6.26) with (6.25) and (6.22) as:

$$\left\langle \frac{dE}{dt} \right\rangle = \frac{8 \pi G R^8 m^* \rho_w}{a} \frac{(1 + k_2')}{a} \left( \frac{R_1}{a} \right) F_2 (I) G_{200}(e) D_{M_2, 22^+} \sin \epsilon_{M_2, 22^+}.$$

Table 7. Estimates of the lunar acceleration $\dot{n}$, the Earth's acceleration $\dot{\theta}$ and the non-tidal acceleration $\dot{\theta}_{NT}$ as deduced from astronomical, satellite, tide and paleorotation observations. The satellite and tide results represent present-day values. The astronomical result is the average during the last 3000 yr and the paleorotation observations are averages over the last $500 \times 10^6$ yr (Lambeck 1978a, b).

<table>
<thead>
<tr>
<th>Estimate</th>
<th>$\dot{n}$ (10^{-23} s^{-2})</th>
<th>$\dot{\theta}$ (10^{-22} s^{-2})</th>
<th>$\dot{\theta}_{NT}$ (10^{-22} s^{-2})</th>
</tr>
</thead>
<tbody>
<tr>
<td>Astronomical observations (telescope and eclipses)</td>
<td>$-1.35 \pm 0.10$</td>
<td>$-5.5 \pm 0.5$</td>
<td>$1.6 \pm 0.6$</td>
</tr>
<tr>
<td>Satellite orbit perturbations</td>
<td>$-1.3 \pm 0.25$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ocean tide models</td>
<td>$-1.5 \pm 0.5$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Coral and bivalve fossil records</td>
<td>$-1.2 \pm 0.2$</td>
<td>$-5.3 \pm 0.6$</td>
<td>$1.0 \pm 0.6$</td>
</tr>
</tbody>
</table>
The same result is obtained by evaluating the rate at which the Moon does work on the ocean (Lambeck 1977). Values based on astronomical, tide and satellite results are indicated in table 7 and show satisfactory agreement. The agreement between the astronomical and satellite results is probably better than we have any right to expect in view of the preliminary nature of the latter but it does indicate that we have a powerful new method for estimating the rate of tidal energy dissipation. Improved results can be expected when analyses of longer series of observations of the STARLETTE and GEOS 3 satellites are completed. The agreement between the two estimates also indicates that dissipation of tidal energy within the Moon is likely to be negligible compared with the terrestrial sink. That the three estimates are in good agreement stresses the fact that the primary sink on Earth is the ocean, and that little dissipation occurs within the solid Earth itself. The tide models tend to give estimates that are somewhat higher than the other methods and this is possibly a consequence of these models not taking into account the crustal deformation of the crust by the variable tidal load and the subsequent modification of the tide (see the discussion in Cartwright (1977)).

While there are three ways of estimating the lunar acceleration and the rates of energy dissipation, only the astronomical observations can give a true estimate of the Earth's secular acceleration \( \dot{\theta} \). From the lunar accelerations the tidal acceleration of the Earth \( \dot{\theta}_T \) follows from (6.25) and this part can be deduced from the three types of observations discussed above. But other factors may contribute to the total acceleration by an amount:

\[
\dot{\theta}_{NT} = \dot{\theta} - \dot{\theta}_T.
\]

All indications are that this non-tidal part is small but positive, of the order of \( 1-2 \times 10^{-22} \text{s}^{-2} \) (Lambeck 1978b). Proposed causes for this non-tidal acceleration are numerous and include a slow on-going response of the Earth's mantle to the removal of the Pleistocene ice fields, to long-period electromagnetic interactions between the core and mantle, and to a slow growth of the core. While satellite observations can contribute to improved values of \( \dot{\theta}_T \), only astronomical observations can provide \( \dot{\theta} \) and hence \( \dot{\theta}_{NT} \). New techniques for measuring \( \dot{\theta} \), such as VLBI, will not be of any immediate help either, since the Earth's rotation is a combination of long period and secular trends and several hundred years of observations are required in order to separate the two.

6.5. Rotation

6.5.1. A brief geophysical discussion. When measured with high precision, the Earth's rotation is seen to be highly variable due to a multitude of factors. Forces and deformations in the atmosphere, oceans, crust, mantle and core all perturb the rotation to varying degrees from idealised rigid body motion and a complete study requires one to delve into many aspects of the Earth and planetary sciences. A discussion on the Earth's rotation is conveniently separated into three parts: (i) precession and nutation, (ii) polar motion, and (iii) changes in length of day.

Precession and nutation describe the motion of the Earth's rotation axis relative to inertial space and are a consequence of the lunar and solar gravitational attraction on the Earth's equatorial bulge. Briefly, precession is the regular motion of the Earth's rotation axis about the pole of the ecliptic, a passage that takes about 26 000 yr. The rate of precession is proportional to:

\[
H = \frac{1}{2}(I_{11} - I_{22})/I_{33}
\]
Methods and geophysical applications of satellite geodesy

a constant that, when combined with the Stokes coefficient \( C_{20} \) (equation (2.2)), gives the Earth's polar moment of inertia, or:

\[-C_{20}/H = I_{33}/MR_{e}^2.\]

Periodic oscillations about this mean motion occur and are referred to as the forced nutation of the Earth. The main nutation term is a 19.7 yr oscillation of 9 arcsec amplitude. The standard treatment of this motion in space for a rigid Earth is by Woolard (1953). Observational evidence is discussed by Federov (1963) and in papers edited by Federov et al (1979). The main discrepancy between the theory and the observed motion occurs in the amplitudes of some of the nutation terms and this is a consequence of the Earth's liquid core. Polar motion concerns the movement of the Earth's rotation axis relative to a framework fixed within the Earth. This motion includes a 14-month oscillation, referred to as the Chandler wobble, an annual term due to seasonal rearrangements of mass in the Earth's atmosphere and hydrophere and possibly a secular term and irregular long-period fluctuations of uncertain origins (figure 16(a)). The amplitudes of the annual and Chandler wobbles are of the order of 0.10–0.15 arcsec and the secular drift is believed to be of the order of 0.002–0.003 arcsec yr\(^{-1}\). Changes in the length of day, or changes in the speed of rotation of the Earth about the instantaneous rotation axis, have the richest spectrum of all (figure 16(b)). A secular change, resulting in an increase of the length of day by some 0.001–0.002 s per century, is mainly a consequence of the work done by the Moon in raising the ocean tides (§6.4). Despite the smallness of this acceleration its consequences—when integrated over geological time—are impressive; if the dissipation mechanism has remained constant the Moon would have been very close to the Earth about 1.5 thousand million years ago, there would have been about 1200 d in the year and the length of day would have been about 7 h. Fluctuations on a time scale of 10–100 yr are also clearly evident in the astronomical data, changes in length of day of some 4–5 ms having occurred within 10–30 yr (figure 17). Only the core is sufficiently mobile and contains sufficient mass to explain these fluctuations. How these core motions are transferred to the mantle remains uncertain but electromagnetic forces appear to play an important role. Seasonal changes in the length of day—annual, semi-annual and biennial—are a consequence of variations in the zonal atmospheric circulation, as are many of the higher frequency, less regular perturbations.

The geophysical information obtained from the study of the Earth's rotation include the following.

(i) Estimates of the rate of dissipation of tidal energy in the oceans.
(ii) Limits on the anelastic behaviour of the Earth's mantle.
(iii) Estimates of the net fluctuation in the angular momentum of the atmosphere.
(iv) Indications that long-period tides may not follow an equilibrium theory.
(v) Limits on core properties such as degree of density stratification and viscosity.
(vi) Information on electromagnetic processes operating at the core–mantle boundary and limits on the electrical conductivity of the lower mantle.

Many aspects of the Earth's rotation are discussed in reviews by Munk and MacDonald (1960), Rochester (1973, 1979) and Lambeck (1978a). I limit the present discussion only to an outline of methods of observing rotation changes using space techniques.

Until recently, all observations of the Earth's rotation were made by classical astronomical techniques. Length of day observations are made by comparing the
Polar motion is deduced by measuring fluctuations in the observer's astronomical latitude, the angle between the observer’s vertical and the instantaneous rotation axis. Nutation is observed by measuring the changes in the angles between stars and the mean rotation axis. Such observations, of variable quality, go back several hundred years and all present geophysical knowledge of the forces causing the Earth to depart from rigid body motion is based upon these data. But these methods are reaching the limits in both precision and resolution and new approaches are being explored, spurred on by developments in space science and technology.

Precise tracking of satellites for gravitational studies, laser ranging to the Moon for studying the lunar motion, long-baseline interferometry, observations for de-
Ciphering extra-galactic radio sources, or the precise maneuvering of interplanetary flights all require the precise tracking of the motions of the Earth's rotation axis. At the same time, these new techniques permit the rotational motions to be measured with a precision and resolution that will ultimately represent major improvements over conventional astronomical observations. To provide useful results, these techniques must give (i) improved precision, (ii) improved resolution, (iii) long-term stability and (iv) permit a long-term observing programme. Present observations of polar motion are precise to about 0.01–0.02 arcsec for observations averaged over 5 d. The amount by which the Earth is slow or fast can be measured with an accuracy of about 1.5–2.0 ms for an integration time of about 5 d. Reasonable goals for new systems would be daily values accurate to 0.01 arcsec in pole position and 1 ms in the rotation angle $\theta$.

6.5.2. Satellite methods. The equations used to describe the motion of satellites are usually referred to an inertial reference frame $\mathbf{X}$ while the tracking stations, from which the satellite is observed, rotate with the Earth and are known within a terrestrial frame $\mathbf{x}$. The relation between the two systems is given by equation (4.4). The precession and nutation matrices $\mathcal{P}(\kappa, \nu, \omega)$ and $\mathcal{P}(\Delta \mu, \Delta \nu, \Delta \epsilon)$ are usually taken to be known with adequate accuracy, although it will become necessary to revise the constants appearing in the theory and to understand better the consequences of the non-rigidity of the Earth upon them. The polar motion matrix is:

$$\mathcal{M}(m_1, m_2) = \begin{pmatrix} 1 & 0 & -m_1 \\ 0 & 1 & m_2 \\ m_1 & -m_2 & 1 \end{pmatrix}$$

and the sidereal rotation matrix is:

$$\mathcal{R}(\theta) = \begin{pmatrix} \cos \theta & -\sin \theta & 0 \\ \sin \theta & \cos \theta & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$
instantaneous rotation axis while the second matrix determines the position of the Greenwich meridian at a specified time. If polar motion, for example, is ignored or an imprecise value of \( \theta \) is used, discrepancies result in the transformation from \( x \) to \( X \) and this leads to differences between the observed and computed positions of satellites. That is, the satellite position is reduced to a system \( X' \) which differs from \( X \) by the neglect of, or errors in, \( \Delta(m_1, m_2) \) or an error \( \Delta \theta \) in \( \theta \), while the predicted position refers to the \( X \) system. Thus the orientation of the orbit is defined by the angles \( I, \Omega \) and \( \omega \) relative to \( X \) but the observations yield \( I', \Omega' \) and \( \omega' \) relative to \( X' \). The differences contain information on the rotation angles \( m_1, m_2 \) and \( \Delta \theta \). This is illustrated in figure 18 in which the position of the pole is defined in terms of polar coordinates \( m_0, A \) by:

\[
m_1 = m_0 \cos A \quad m_2 = m_0 \sin A.
\]

Then (Lambeck 1971):

\[
\Delta i = I' - I = m_0 \sin (\Omega - \theta - A) \\
\Delta \omega = \omega' - \omega = m_0 \cos (\Omega - \theta - A) \csc I \\
\Delta \Omega = \Omega' - \Omega = \Delta \theta + m_0 \cos (\Omega - \theta - A) \cot I \tag{6.28}
\]

and the three angular elements exhibit periodic fluctuations with periods nearly equal to 24 h due to polar motion. Errors in the sidereal angle enter directly into the discrepancy in the right ascension of the orbit. Amplitudes of the orbital perturbations are of the order of \( m_0 \). In any satellite orbit calculation, the discrepancies between observed and computed positions are a consequence of numerous factors (§2.3) and the success of the method requires that these, particularly any with periods near 24 h, are precisely known. Examples of such perturbations are the Stokes

\[\text{Figure 18. Effect of errors in polar motion on satellite orbit.}\]
The amplitude of the total polar motion perturbations in the satellite motion is about 5 m and to observe this with high accuracy requires accurate tracking systems and frequent observations during consecutive 24-h periods.

The first results for pole positions from satellite orbit perturbation studies were obtained by Anderle and Beuglass (1970) from the Doppler tracking of US Navy navigation satellites. Since 1969, Doppler pole positions have been computed on a regular basis and results have been obtained a posteriori for the years 1964–1969. Hence a continuous record of some 13 yr is now available (Anderle 1973, 1976). A great advantage of the Doppler method is that the observations are relatively precise, that the observations are independent of weather conditions and that the network consists of a large number of stations (19 in 1973) that permanently track the satellite. Pole positions are determined at intervals of 2 d compared with 5 d by the traditional astronomical services. Results since 1970 are in excellent agreement with the astronomical observations (figure 19). Data obtained at CNES since mid-1976 from a multi-station Doppler tracking network give comparable results.

Laser observations, despite their high accuracy, have been less useful for polar motion studies mainly because of the fewer stations available and of their weather dependence, Smith et al. (1972) measured the apparent fluctuation in the inclination of a satellite orbit from one station. This only gives the component of polar motion in the station meridian. Due to the periodically changing geometry of the orbit relative to the tracking station, pole positions can only be obtained at intermittent intervals.

Satellite methods suffer a drawback, similar to the astronomical observations, in that it is difficult to maintain a homogeneous reference system. The parameters defining the Earth's gravity field and the tracking stations are subject to regular revision as more and more data are accumulated and analysed. Each revision results in discontinuities in the pole path record. The LAGEOS satellite should remedy some of these problems once a regular observing programme is established. Its high altitude make it relatively insensitive to the higher degree terms in the gravity

Figure 19. Comparison of polar motion results for 1972 as observed by the Bureau International de l'Heure (full curve) and R. J. Anderle from satellite orbit perturbations (•). Only the \( m_1 \) component is shown.
field and increases the frequency and duration with which it can be observed from any station. Its low surface area to mass ratio and its regular shape and reflectivity parameters also minimise the perturbations due to air drag and radiation pressure, whether direct or indirect. Station coordinates must be known with high accuracy. The main problem until now has been to acquire a sufficient number of laser tracking stations, geographically well distributed around the world, that routinely and accurately track the satellite.

Satellite methods have not yet provided useful observations of changes in the length of day. From (6.28) only the longitude of the ascending node is perturbed by errors in the sidereal angle but this orbital element is difficult to compute with precision. The zonal harmonics in the gravity field cause secular and long-period perturbations, with air drag and radiation pressure also making important contributions to $\Omega(t)$. Separation of these effects from errors in $\theta$ is not possible in the way it is for the polar motion, where a distinctly periodic signal is introduced into the orbital elements. Even with satellites such as LAGEOS it is improbable that the motion of the node will be predictable with sufficient accuracy over long time intervals. Observations of the node may, however, provide useful information on shorter period and irregular changes in the length of day.

6.5.3. Lunar laser ranging. The determination of the Earth's rotation from lunar laser ranging proceeds as follows. In figure 20(a) the observer at P measures the Earth–Moon distance over a period of several hours so that the Moon is seen to cross the meridian (at time $T_0$). If the Earth's rate of rotation were inaccurately known this will result in a misplacement of P, relative to an inertial system, by an angle $d\theta$ and there will be an error in the computed station–Moon distances of:

$$ds = \frac{S_0 R}{S} \sin (\theta \Delta T) \, d\theta$$

![Figure 20](image-url)  

*Figure 20.* Geometry of lunar laser ranging. (a) illustrates the case where observations are made at times $\Delta T$ before and after the transit time $T_0$ for determining rotation. The Moon's motion is ignored here. (b) illustrates the determination of astronomical latitude from range observations to the Moon when it is in the meridian.
where $\Delta T$ is in the time interval before transit and $S_0$ is the Earth-Moon distance at transit. During the sequence of observations of the Moon's passage across the meridian, the discrepancy $dS$ between the observed and computed distance varies sinusoidally with a frequency $\theta$. No other error or variation, apart from the tidal distortion of the Earth, is expected to possess such a diurnal variation and the observed discrepancies should provide an unambiguous estimate of $\Delta \phi$. Observations of polar motion are more difficult to deduce from the laser range observations. An error $\Delta \phi$ in the station latitude results in a range error $dS$ that varies with the Moon's declination $\delta$ (figure 20(b)) according to:

$$dS = \frac{S_0 R}{S} \sin (\phi - \delta) \Delta \phi.$$ 

As $\delta$ varies slowly, $dS$ will remain nearly constant over a day but it will vary from day to day as the declination varies, with a period of about 27 days. Thus to determine $\Delta \phi$ requires that any 27 day period in the lunar librations and the lunar motion about the Earth are known with accuracy. Also, changes in polar motion with periods much less than 27 days will be difficult to observe.

A disadvantage of the lunar laser is that around the new Moon the reflectors are difficult to find with the narrow laser beams and there is usually a period of a few days when no observations are possible. This disadvantage can be overcome by combining lunar and artificial satellite observations. The former give a direct determination of $\theta$ but with interruptions during which $\theta$ could be interpolated using observations of the node of high-altitude satellites such as LAGEOS. At the same time, the satellite observations provide a more direct determination of latitude changes than does the lunar method. Preliminary lunar laser results for the Earth's rate of rotation, obtained by Stolz et al. (1977) and Harris and Williams (1977), indicate accuracies comparable to those obtained by conventional astronomical observations.

### 6.5.4. Radio interferometry

The radio interferometer observations provide a measure of the baseline orientation according to (6.3), a derivation based on the assumption that the terrestrial frame $x$ is parallel to the inertial frame $X$ except for the diurnal rotation. Any polar motion will introduce apparent changes in the latitudes and longitudes of the receivers according to:

$$\Delta \phi_i = m_1 \cos \lambda_i + m_2 \lambda_i$$

$$\Delta \lambda_i = m_1 \tan \psi_i \sin \lambda_i - m_2 \tan \psi_i \cos \lambda_i.$$ 

The change in the delay due to the polar motion then follows from (6.3) as:

$$\frac{c}{R} \Delta \tau = [\sin \delta_\alpha (\cos \phi_2 \cos \lambda_2 - \cos \phi_1 \cos \lambda_1) + \cos \delta_\alpha (-\sin \phi_2 + \sin \phi_1) \cos (L_\alpha - \theta)]m_1$$

$$+ [\sin \delta_\alpha (\cos \phi_2 \sin \lambda_2 - \cos \phi_1 \sin \lambda_1) + \cos \delta_\alpha (-\sin \phi_2 + \sin \phi_1) \sin (L_\alpha - \theta)]m_2.$$ 

The neglect of polar motion therefore introduces a constant error in the delay as well as a diurnal variation. Daily averaged observations taken at successive days should then reveal a gradual variation in the delay as the pole shifts. Observations from two baselines (a minimum of three stations) are required to separate $m_1$ and $m_2$ but precise positions for either the baseline or stellar sources are not essential provided that, for each baseline, the same radio source is always observed.
Determination of the speed of rotation will be more complicated since, from (3.3), an error in \( \theta \) introduces a delay:

\[
\frac{c}{R} \Delta \tau = \cos \delta \left[ -\cos \phi_2 \sin (\lambda_2 + \theta - L_2) + \cos \phi_1 \sin (\lambda_1 + \theta - L_2) \right] d\theta
\]

that varies diurnally and will be difficult to separate from other diurnal sources such as the above component from polar motion and from tides. The dependence of the diurnal \( d\theta \) signal on the baseline and source coordinates differs from the diurnal polar motion and tide signal and a separation of \( d\theta \) from the other sources is possible by an appropriate choice of baselines.

7. Lunar and planetary problems

7.1. Introduction

The geophysical problems discussed in the preceding section are equally relevant to the Moon and other terrestrial planets. Planetary orbiters indicate that all these bodies have irregular topographies and gravity fields as well as special problems associated with their tides and rotations. A study of the planets, as well as being of intrinsic interest, is also of relevance to understanding the evolution of the Earth as it is widely believed that they represent different steps in a general evolutionary process (Kaula 1975). Only some general remarks on the planetary problems can be made in this review.

Of the terrestrial planets, only for Mars is the gravity field known with some detail and reliability through the Mariner and Viking programmes. For a number of reasons the martian orbiters are not ideal for gravity field investigations in that the orbits are very eccentric—for example, Mariner 9 had a periapsis of 1400 km and apoapsis of about 11 000 km. Nevertheless, valuable information now exists. A recent solution by Gapcynski et al (1977), using Mariner 9 and Viking 1 and 2 Orbiter observations, gives the field complete to degree and order 6. Recently, the orbit of Viking Orbiter 2 has been lowered to about 300 km perigee and considerable improvements in the resolution of the gravity field can now be expected. Discussions on the interpretation of the Martian gravity field are by Phillips et al (1978) and Lambeck (1978c). The encounters with Mercury and Venus only permit approximate limits to be imposed on the second degree harmonics (Anderson 1974).

7.2. Lunar gravity field

The gravity field of the Moon can be determined in a similar way to that developed for the Earth although there are some important differences. First, the spacecraft accelerations and positions are measured from the Earth so that only direct observations of the near side are possible. This could be rectified by using a relay satellite in a high-altitude orbit about the Moon. The low satellite would be tracked relative to the high satellite which in turn is tracked from Earth. This is analogous to the high–low satellite-to-satellite tracking concept discussed in §5. Second, the Moon's gravity field does not possess a single dominant zonal coefficient as does the Earth and any analytical theory for the motion of a lunar orbiter is more complex than the comparable terrestrial problem. Also, because of this, the spectrum of lunar orbital perturbations is not as ordered as it is for the Earth. Third, because
the Moon possesses no atmosphere, the satellites can orbit the Moon at much lower altitudes than in the case of Earth satellites. Thus artificial Moon satellites are often very sensitive to small-wavelength fluctuations in the gravity field.

The first solutions for the lunar gravity field used the fact that the observations of the spacecraft acceleration relative to the distant terrestrial observer are approximately equal to the radial component of gravity. This method was first used with impressive results by Müller and Sjogren (1968) to obtain a gravity map of the near side of the Moon from the Lunar Orbiter 5. The method has subsequently been used to obtain gravity over selected features using low-altitude Apollo and Apollo subsatellite passes (Sjogren 1977). These results must be considered as preliminary only in that (i) a systematic reduction of the measured accelerations down to the surface of the Moon or to a surface of constant height has not always been made, and (ii) the line of sight accelerations correspond to gravity only when the spacecraft is near the observer–centre of Moon axis. In most published results a correction for the non-coincidence of the spacecraft with this axis has not been made and the gravity towards the limb of the Moon becomes distorted. More recently, tracking data from a number of different spacecraft have become available and the orbital theories have been refined to permit the elements of the orbits to be computed and analysed for the lunar Stokes coefficients in a way similar to that used for the Earth. Now the gravity field of both the near and far sides of the Moon can be mapped although, by analysing mean orbital elements, the short-wavelength information, contained in the direct line of sight acceleration observations, is lost. Ferrari (1977) has modelled the gravity field by a spherical harmonic expansion complete to degree and order 16. He gives a detailed evaluation of the accuracy of the Stokes coefficients and only harmonics below degree and order 10–12 appear to be individually significant. Considerably more tracking data are available, both for satellites already used in Ferrari's study and for additional satellites and it is anticipated that this study will only be one more step towards a comprehensive lunar gravity field model. Interpretations of the lunar gravity field have been made by Kaula (1971) and Sjogren (1977).

7.3. Lunar rotation

The Moon's rotational motion about its centre of mass has been much discussed since the formulation of Cassini's laws and Newton's theory in the 1690s. Cassini noted that (i) the period of rotation of the Moon about its polar axis equalled the period of the Moon's motion about the Earth, (ii) the inclination of the lunar equator to the plane of the ecliptic is constant, and (iii) the poles of the Moon's axis of rotation, of the ecliptic and the lunar orbit are coplanar. Such motion is compatible with the configuration in which the dissipation of tidal energy within the Moon is a minimum (Colombo 1967). The physical librations represent small oscillations about this mean rotation due to the terrestrial, and to a lesser extent, the solar attraction on the Moon's irregular shape. The existence of these librations was already recognised by Newton and the theory of the motion has been much discussed by the leading mathematicians in subsequent centuries.

The physical librations are similar to the Earth's precession and nutation and are described by the Eulerian equations of motion of a triaxial ellipsoid subject to external forces (in this case the Earth's gravity) and to the condition of synchronous rotation. The frequencies of these oscillations is a function of the motion of the
Moon about the Earth and the amplitudes of the motion are dependent on the inertia tensor of the Moon. Specifically, the libration amplitudes are proportional to:

\[ \alpha = \frac{I_{33} - I_{22}}{I_{11}}, \quad \beta = \frac{I_{33} - I_{11}}{I_{22}}, \quad \gamma = \frac{I_{22} - I_{11}}{I_{33}} \]

with the constraint:

\[ \alpha - \beta + \gamma = \alpha \beta \gamma \]

where \( I_{11}, I_{22}, I_{33} \) are the moments of inertia about axes directed along the long axis \( O_x \) of the Moon, towards the Earth, along an axis \( O_{x2} \) in the lunar equatorial plane and along the rotation axis \( O_r \).

Observations of the physical librations together with observations of the lunar gravity field coefficient:

\[ C_{20} = -\frac{1}{M} \left[ I_{33} - \frac{1}{2} (I_{11} + I_{22}) \right] \]

give an estimate of the lunar moment:

\[ \frac{I_{33}}{M R^2} = -\frac{C_{20} (1 + \beta)}{2 \beta - \gamma + \beta \gamma} \]

In an accurate theory, librations arising from the attraction on the third and fourth harmonics of the Moon's gravity must be included.

Theories, with precisions comparable to lunar laser ranging measurements, have been developed by Williams et al. (1973), Eckhardt (1973) and Migus (1976). Comparisons of these theories with the laser range observations from the MacDonald Observatory give (Williams 1977):

\[ \beta = (6.313 \pm 0.003) \times 10^{-4} \]
\[ \gamma = (2.274 \pm 0.006) \times 10^{-4}. \]

These values are compatible with those found using the differential long-baseline interferometry method applied to the Moon (King et al 1976). They compare with:

\[ \beta = 6.294 \times 10^{-4} \quad \gamma = 2.310 \times 10^{-4} \]

found from astronomical telescope observations (Koziel 1967). With the recent estimates for \( C_{20} \) and \( C_{2,2} \) given by Ferrari (1977):

\[ C_{20} = -2.046 \times 10^{-4} \]
\[ C_{2,2} = 2.173 \times 10^{-5} \]

and the above laser results for \( \beta \) and \( \gamma \):

\[ C/\frac{M R^2}{C} = 0.392 \pm 0.002. \quad (7.1) \]

This value is somewhat lower than the 0.40 found in the pre-lunar laser ranging days. Models for the lunar interior based on the lower value (7.1) have been reviewed by Ringwood (1977).

Free librations about the equilibrium position of the rotating Moon should
also be considered. These are analogous to the free precession or Chandler wobble of the Earth except that, because of the significant ellipticity of the lunar equator ($I_{11} \neq I_{22}$) and of the synchronous rotation of the Moon, there are now three components. They are (i) a fluctuation in rotation rate, with a period of about 2.9 yr, (ii) a precession of the spin angular momentum axis relative to a frame rotating with the Moon’s orbital plane, and with a theoretical period of about 24 yr, and (iii) a free wobble with a period of about 7.4 yr. The third motion involves a change in the orientation of the spin axis relative to axes fixed in the Moon. Sekiguchi (1970) and Peale (1973a, b) discuss the theory. Together, these three motions introduce six free parameters into the theory of the lunar motion (three amplitudes and three phases, the frequencies being assumed known from the theory). Peale (1975, 1976) has discussed possible excitation mechanisms that could maintain these free modes against damping. Meteorite impacts on the Moon are the main source of excitation of these motions. Estimates of the libration amplitude require assumptions on the size range and frequencies of impacts and upon the mechanism by which rotational energy is dissipated. Thus any estimates will be uncertain but Peale (1976) concludes that the amplitudes of the librations are unlikely to exceed 0.01 arcsec and to be well below the projected noise level of laser range observations. Calame (1977) has made a tentative identification of these free oscillations in some 6 yr of laser range observations but the results are at variance with Peale’s study of the excitation and dissipation mechanisms in that she finds amplitudes well above the levels predicted by Peale.

8. Conclusions

Nearly 20 years of activity in space research have resulted in striking results in satellite geodesy. Most impressive is the improved knowledge of the Earth’s gravity field, results that have been of value in a number of geodynamic studies. Other important results include the ability to determine positions on the Earth’s surface relative to each other and relative to the centre of mass with precisions exceeding 1 part in $10^6$. Few people would deny that the initial goals of satellite geodesy, as set out in the National Geodetic Satellite Program started by NASA in 1965, have been met. At the same time few people will deny that the initial results for the Earth’s tides and rotation, coupled with recent progress in precision tracking and developments in satellite instrumentation, leave no doubts that the best results are still to come.

Further improvements in the gravity field are an essential part of a future programme for two reasons. One is the geophysical interest in this field, the other is that precise and stable orbits are required for many other satellite studies. The oceanographic SEASAT A mission, for example, required orbits over the ocean areas with a precision equal to the measurement accuracy, or a few tens of centimetres. This requirement can be met with precise laser tracking, provided that a sufficient number of stations are available. The former requirement of improved resolution and precision for geophysical purposes can be met in part by ground tracking, but significant future progress will be possible only by introducing new techniques. Satellite-to-satellite microwave Doppler tracking is one possibility although laser Doppler tracking has been studied and may result in very major accuracy improvements. Gradiometer measurements of the gravity field gradient is another possibility. In both programmes the required low-flying satellites must be fitted with surface
force compensation devices. The Space Shuttle may be a convenient vehicle for launching these spacecraft.

Studies of the temporal variations in the gravity field provide estimates of the mass redistribution associated with solid and ocean tides and within the atmosphere. Precision laser tracking of low satellites will probably be the best way of observing these effects since many of these changes result in orbital perturbations that are longer than a day or so, and the direct gravity sensing instruments may not have sufficient long-term stability.

Present astronomical methods of measuring the Earth's rotation are reaching their limits in both precision and in resolution and space techniques, whether they be satellite tracking, lunar ranging or radio interferometry, will make important contributions in the near future. The choice of technique is perhaps not evident. For the determination of the variations in the speed of rotation, lunar laser range observations give the most direct result except that observations are not possible around new Moon. A combination of the lunar range measurements with laser ranging to high satellites in stable orbits, such as LAGEOS, offers a solution with the former providing a measure of long-period variations and the latter providing a measure of short duration (≤ 10 d) changes. Of the laser ranging methods, LAGEOS offers a better target than the Moon as far as polar motion determination is concerned. Long-baseline interferometry measurements will also contribute to the rotation determination.

It is important to emphasise that the three new methods of determining the Earth's rotation are primarily designed for other purposes. Laser ranging to satellites will always be necessary for gravity field studies, lunar laser range observations determine the motions of the Moon, and long-baseline interferometry provides information on the structure of the stellar sources. Also, in all cases, not only irregularities in rotation contribute to discrepancies between theory and observation. Often a 'bootstrap' approach is required to separate the rotation parameters from tidal, crustal motion and gravity perturbations. Finally, these new techniques all have their own peculiar error sources which have not yet been sufficiently tested. Hence there is a need, at least in the near future, for a three-way approach and for detailed intercomparisons before concluding whether one method is superior to any other.

Measurement of precise relative positions will be of importance in studies of plate tectonics once centimetre-level accuracies are attained. Laser ranging to the LAGEOS satellite appears to be one way to do this. Differential long-baseline interferometry offers another approach in areas where rapid motions over relatively short distances are to be observed. Again, no one method can be recommended as being better than any other while there has not yet been an intensive period of testing and intercomparisons of competing techniques. In view of the need to measure both internal deformations and relative motions between plates, mobile stations appear desirable so that a large number of sites can be occupied during any one year. Mobile laser units for satellite tracking have been used for some years by the Goddard Space Flight Center while the Jet Propulsion Laboratory has used a mobile radio antenna in conjunction with a fixed station. A mobile laser unit for ranging to both satellites and the Moon has also been proposed (Bender et al 1977). Such units could be used for measuring the regular slow motions associated with the global tectonic framework as well as for measurements in areas where sudden and rapid deformation is observed or to monitor post-seismic deformations.
Methods and geophysical applications of satellite geodesy

References

Allan RR 1973 Planet. Space Sci. 21 205–25
Alley CO et al 1965 J. Geophys. Res. 70 2267–9
Anderle RJ 1973 Geophys. Surv. 1 147–61
——— 1974 J. Geophys. Res. 79 5319–31
——— 1975 Naval Surface Weapons Center Tech. Rep. 3433 (Dahlgren, Virginia) p50
——— 1976 Naval Surface Weapons Center Tech. Rep. 3464 (Dahlgren, Virginia) p10
Anderle RJ and Beuglass LK 1970 Bull. Geod. 96 125–41
Anderle RJ and Smith SJ 1968 J. Astronaut. Sci. 15 1–4
Anderson DL and Hart RS 1978 J. Geophys. Res. 83 5869–82
Baker TF and Lennon GW 1976 Proc. 7th Int. Symp. on Earth Tides (Budapest: Akademiai Kiado) pp463–78
Balmino G 1975 Satellite Dynamics ed G Giacaglia (Berlin: Springer-Verlag) pp50–97
Bare C et al 1967 Science 157 189–91
Belousov VV 1972 Tectonophys. 13 119–23
Bender PL et al 1973 Science 182 229–38
Bender PL et al 1977 Tectonophys. in press
Berger X 1975 Satellite Dynamics ed G Giacaglia (Berlin: Springer-Verlag) pp35–40
Berger X and Walch J 1977 Manuscripta Geodaetica 2 99–133
Brotien NW et al 1967 Science 156 1592–4
Brune JN 1968 J. Geophys. Res. 73 777–84
Cartwright DE 1968 Phil. Trans. R. Soc. A 263 1–55
Castle RO, Church JP and Elliott MR 1976 Science 192 251–3
Chapman DS and Pollack HN 1975 Earth Planet. Sci. Lett. 28 23–32
Chen WP and Molnar P 1977 J. Geophys. Res. 82 2945–69
Chinnery MA and Landers TE 1975 Nature 258 490–3
Coates RJ et al 1975 Tectonophys. 29 9–18
Colombo G 1967 Measure of the Moon (Dordrecht: Reidel) pp12–22
Daillet S 1977 Thesis University of Toulouse
Eckhardt DH 1973 Moon 6 127–34
Federov YP 1963 Nutation and Forced Motion of the Earth’s Pole (Oxford: Pergamon)
Federov YP, Bender PL and Smith ML (ed) 1979 Nutation and the Earth’s Rotation (Dordrecht: D Reidel)
Ferrari A 1977 J. Geophys. Res. 82 3065–84
Filoux JH 1971 Deep Sea Res. 18 275–84
——— 1976 Moon 16 3–26
Gaposchkin EM 1973 Smithsonian Astrophysical Observatory Special Rep. 353 388
Gold T 1967 Science 157 302–4
Guier WH and Newton RR 1965 J. Geophys. Res. 70 4613–26
Heiskanen W 1921 Ann. Acad. Sci. Fennicae A 18 1–84
Heiskanen WA and Moritz H 1967 Physical Geodesy (New York: Freeman and Co)
Hopfield H 1969 J. Geophys. Res. 74 4489–99
Isacks B, Oliver J and Sykes LR 1968 J. Geophys. Res. 73 5855–99
Jacchia LG 1967 Smithsonian Astrophysical Observatory Special Rep. 245 24
— 1977 Smithsonian Astrophysical Observatory Special Rep. 375 106
Jeffreys H 1920 Phil. Trans. R. Soc. A 221 239–64
— 1962 The Earth (Cambridge: Cambridge University Press) 4th edn
— 1970 The Earth (Cambridge: Cambridge University Press) 5th edn
Kaula WM 1963 J. Geophys. Res. 68 4967–78
— 1964 Rev. Geophys., Space Phys. 2 661–85
— 1966 Introduction to Satellite Geodesy (Mass.: Blaisdell)
— 1971 Phys. Earth Planet. Int. 4 185–92
— 1975 Icarus 26 1–15
King RW, Counselman CC and Shapiro IL 1976 J. Geophys. Res. 81 6251–6
King-Hele DG 1964 Planet. Space Sci. 12 835–53
— 1975 Phil. Trans. R. Soc. A 278 67–109
King-Hele DG and Cook GE 1974 Planet. Space Sci. 22 645–72
Knopoff L 1972 Tectonophys. 13 497–519
Kokurin YL et al 1966 IETP Lett. 3 139–41
Kozai Y 1969 Smithsonian Astrophysical Observatory Special Rep. 295 17
Kuo JT, Jachens RC, Ewing H and White G 1970 Science 168 968–71
Methods and geophysical applications of satellite geodesy

--- 1975 J. Geophys. Res. 80 2917-25
--- 1976 J. Geophys. Res. 81 6333-40
--- 1977 Phil. Trans. R. Soc. A 287 545-94
--- 1978a The Earth’s Variable Rotation: Geophysical Causes and Consequences (Cambridge University Press) in press
--- 1978c J. Geophys. Res. in press
Leeds AR, Knopoff L and Kausel EG 1974 Science 186 141-3
LePichon X 1968 J. Geophys. Res. 73 3661-97
LePichon X, Francheteau J and Bonnin J 1973 Plate Tectonics (Amsterdam: Elsevier)
Longman IM 1966 Geophys. J. R. Astron. Soc. 11 133-7
Lundquist CA and Veis G 1966 Smithsonian Astrophysical Observatory Special Rep. 200 2
MacDonald GJ F 1967 Science 157 304-7
McQueen HWS and Stacey FD 1976 Tectonophysics. 34 T1-T8
Melchior P 1974 Geophys. Survey 1 275-303
Migus A 1976 Moon 15 165
Ming LC and Bassett WA 1975 Science 187 66
Moran J M et al 1967 Science 157 676-7
Morando B 1963 The Use of Artificial Satellites for Geodesy ed G Veis (Amsterdam: North- Holland) pp42-51
Morgan WJ 1968 J. Geophys. Res. 73 1959-82
--- 1971 Nature 230 42-3
Morrison F 1970 J. Geophys. Res. 75 2142
Muller PM 1976 Jet Propulsion Laboratory Rep. SP 43-36 p24
Muller PM and Sjogren WL 1968 Science 161 680-4
Muller PM and Stephenson BR 1975 Growth Rhythms and History of the Earth’s Rotation ed G Rosenberg and S K Runcorn (New York: Wiley) pp459-534
Munk WH and MacDonald GJ F 1960 The Rotation of the Earth (Cambridge: Cambridge University Press) p323
Munk WH, Snodgrass F and Wimbush M 1970 Fluid Dyn. 1 161-235
O’Keeffe JA, Eckels A and Squires K 1959 Astron. J. 64 245-53
Ostrovsky AY 1976 Proc. 7th Int. Symp. on Earth Tides (Budapest: Akademiai Kiado) pp121-6
Peale SJ 1973a Rev. Geophys. Space Phys. 11 767-93
--- 1973b Moon 8 515-31
--- 1975 J. Geophys. Res. 80 4939-46
--- 1976 J. Geophys. Res. 81 1813-27
Pearlman MR, Lanham NW, Lehr CG and Wohlen J 1977 Phil. Trans. R. Soc. A 284 431-42
--- 1977 Ann. Geophys. 33 63-6
Pevnev AK, Gusev TV, Odinev NN and Saprykin GV 1975 Tectonophysics. 29 421-8
Phillips RJ and Saunders RS 1975 J. Geophys. Res. 80 2893-7
Plotkin HH 1965 COSPAR Inf. Bull. 29 19-21
Poehls KA, Kaula WM and Schubert G 1977 Final Results of Geos-3 Project (NASA)
Pollack HN and Chapman DS 1977 Sci. Am. 237 No 2, pp60-76
Prescott WH and Savage JC 1976 J. Geophys. Res. 81 4901-8
Ramsden SA 1977 Phil. Trans. R. Soc. A 284 457-60
— 1977 The Vernadsky Lecture. USSR Acad. Sci.
Ringwood AE and Major A 1970 Phys. Earth Planet. Int. 3 89
Roy RF, Blackwell DD and Decker ER 1972 Nature of the Solid Earth ed E C Robertson
Savage J C and Burford RO 1973 J. Geophys. Res. 78 832-45
Schaer JP, Reimer GM and Wagner GA 1975 Tectonophys. 29 293-300
Schwarz C 1970 Ohio State University Rep. 147 134
Sclater JG, Lawyer LA and Parsons B 1975 J. Geophys. Res. 80 1031-52
Schmidt HH 1974 J. Geophys. Res. 79 5349-76
Sehnal L 1973 Space Res. 13 51
Sekiguchi N 1970 Moon 2 78-93
Shapiro I I et al 1974 Science 186 920-2
Sjogren WL 1977 Phil. Trans. R. Soc. 285 219-25
pp285-320
Smith DE et al 1972 Science 178 405-6
Smith DE et al 1976 J. Geophys. Res. 81 1006-26
Smith DE et al 1977 Phil. Trans. R. Soc. A 284 529-36
Smullin LD and Fiocco G 1962 Proc. IEEE 50 1703
Stacey FD 1977 Physics Of The Earth (New York: Wiley)
Staelin DH 1969 Science 157 3024
Stolz A et al 1977 Science 193 997-9
Veis G 1963 The Use of Artificial Satellites for Geodesy ed G Veis (Amsterdam: North-Holland)
pp201-16
Vine FJ 1966 Science 154 1405-15
Vine FJ and Mathews DM 1963 Nature 199 947-9
Vonbun FO 1977 Phil. Trans. R. Soc. A 284 475-83
Wagner CA 1977 J. Geophys. Res. 82 915-27
Walker DMC 1975 Planet. Space Sci. 23 1573-9
Whitcomb JH 1976 J. Geophys. Res. 81 4937-44
Williams JG, Slade MA, Eckhardt D H and Kaula WM 1973 Moon 8 469-83
Wolff M 1969 J. Geophys. Res. 74 5295-330
pp463-805