25 June 1976, Volume 192, Number 4246 SCIENCE

# The Shape of the Earth

Desmond King-Hele

What is the shape of the earth? This is one of the most ancient of scientific questions, but it is a question that is very much alive today. Indeed this is an excellent moment to discuss the figure of the earth, because new techniques introduced in the past few years, such as laser tracking of earth satellites carrying special corner-cube reflectors, are achieving accuracies of 10 cm and may perhaps ultimately reach 1 cm. Geodesy, the science of measuring the earth, has for many years been the Cinderella of the earth sciences, but now it is poised to act as a spearhead of advance in other realms of geophysics.

The first questionings about the shape of the earth probably go back to a very early stage of human evolution, perhaps a million years ago, when our primitive predecessors stood and looked far over the African plain or out over the sea. They stood and wondered: no doubt most of them wondered wrong and thought the earth was flat, but perhaps there were some who thought of the earth as a sphere. As time went on, the idea of a spherical earth probably arose in many cultures, and it was quite popular among the ancient Greeks in the 6th century B.C., particularly among the followers of Pythagoras, who believed that the sphere was the perfect shape, observed that the sun and moon were spherical, and concluded that the earth also enjoyed this perfection. The idea was a good one, and was taken up by influential Greek philosophers such as Plato and Aristotle.

The finest achievement of ancient times in geodesy came during the 3rd century B.C. when Eratosthenes of Alexandria measured the circumference of the earth and possibly came within 1 percent of the correct value, which is just over 40,000 km. Eratosthenes noticed that the sun was overhead at Aswan at midsummer, because it shone right down a well, and was 7.2° or one-fiftieth of a circle away from the vertical at Alexandria, which was almost due north. So the earth's circumference was 50 times the distance from Aswan to Alexandria, and it is said that he calculated this distance by multiplying the average speed of a camel by the time it took on the journey. This was really a fine achievement, a correct concept being followed by the correct observations, and with great economy of effort if he did use camels instead of stretching tape measures across the desert. It was also incidentally the first example of space geodesy, with the sun as the reference object in space (*1*).

In the next 1500 years the ideas of Aristotle prevailed in Christian Europe, and it was quite fortunate for geodesy (though not for astronomy) that Aristotle had a geocentric bias; so the idea of a spherical earth pervades Shakespeare's imagery, and in *A Midsummer Night's Dream* he even goes in for artificial satellite techniques, making Puck orbit the earth in 40 minutes (*2*): this is in error by a factor of 2.2, but geophysicists are used to quite large errors for quantities on the boundary of knowledge.

The next step forward was to flatten the earth slightly around the poles, and this was among the many brilliant advances made by Sir Isaac Newton. In his *Principia* (*3*) Newton imagined one tube of water, or canal, as he called it, running from the north pole to the earth's center, and another from the center to a point on the equator. Water does not run from equator to pole over the surface, so the two canals must "balance": gravity is slightly reduced in the equatorial canal, because of the centrifugal effect of the earth's rotation, so the equatorial canal needs to be slightly longer. Newton estimated the flattening as 1 part in 230. This value is rather too large. The correct value is 1 part in 298.25, the polar diameter being about 43 km (in 12,700 km) shorter than the equatorial diameter. But for nearly a century Newton's value remained the best and most soundly based estimate.

In the early 18th century people were quite fascinated by the problem of the earth's shape, and the best practical way to determine it was to extend the method of Eratosthenes, to find the differing lengths of a degree of latitude at different latitudes. The difference is quite small, 1° of latitude having a length of 111.7 km at the poles and 110.6 km at the equator; but it was enough to be measured.

The French led the way in this work. The first attempts, by the Cassinis, were inaccurate, and they suggested that the earth was drawn out at the poles like a lemon or an egg. But the results of the French expeditions which started out to Lapland and Peru in 1736 were good enough to "flatten both the poles and the Cassinis": they gave values of the flattening between 1 part in 179 and 1 part in 266, which straddled Newton's value of 1 part in 230 (*4*).

Between 1750 and 1950 the value for the flattening was gradually improved, until by 1957 the value generally accepted was 1 part in 297.1. The methods used were to measure arc lengths as the French expeditions did, to record gravity in many parts of the world, to analyze the motion of the moon, and to trace the precessional movements of the earth's axis. But in these 200 years of slow improvement, no general features of the earth's shape, other than the flattening, were securely established, although of course many areas of the earth were accurately mapped.

The author is deputy chief scientific officer in the Space Department of the Royal Aircraft Establishment, Farnborough, Hants, England.

## Satellite Geodesy Begins

The launching of the first artificial satellites in 1957 offered two possible new methods for studying the shape of the earth. One method is to treat the satellites as points in the sky for geometrical triangulation, and the other method is to analyze the effect of the earth's gravitational attraction on satellite orbits, so as to determine how gravity varies over the earth, with great accuracy.

Before describing the results, mainly from this second method in the early years, I should make it clear that in discussing the shape of the earth I shall be concerned not with the mountains and valleys, but with the shape of the mean sea-level surface, continued under the land in a logical fashion. This is the surface usually called the geoid, and it is completely defined by determining the variations of the earth's gravitational attraction with latitude and longitude, and then adding in the acceleration produced by the earth's rotation. In mathematical terms, the geoid is just an equipotential surface of the earth's gravitational potential plus the rotational potential (5). Physically, the geoid is the basic reference shape on which the earth's topography—"height above sea level"—is superposed.

The first application of satellite geodesy was very simple. One effect of the earth's flattening on a satellite orbit is to make the plane of the orbit rotate about the earth's axis in the direction opposite to the satellite's motion, while leaving the orbit inclined at the same angle to the equator. Figure 1 shows the earth as seen from a distant point in space: if the earth were spherical, the orbit would look the same day after day, apart from very small lunisolar perturbations. But a low orbit about the real earth rotates its plane quite rapidly; it swings to the west at a rate of about 8° per day for a near-equatorial orbit, or at 4° per day for an orbit inclined at 60° to the equator. The orbital plane rotates so rapidly that its rate of rotation can be accurately measured by allowing the rotation to build up for several months. If the plane rotates through 500°, and is measured correct to 0.01°, or about 1 km in distance, the rate of rotation can be found accurate to 1 part in 50,000.

This method was applied to Sputnik 2 in 1958, and the rate of rotation was found to be about 0.7 percent less than expected, corresponding to a flattening of 1 part in just over 298, rather than 297.1 as was previously believed (6). The value now established is 1 part in 298.25,

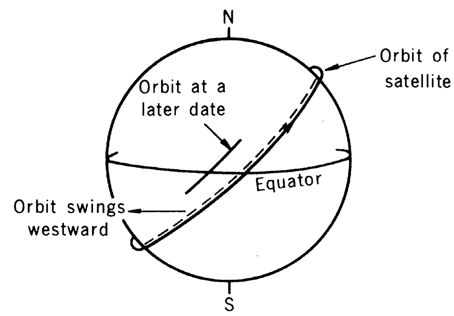


Fig. 1. The gravitational pull of the earth's equatorial bulge makes the orbital plane of an eastbound satellite swing westward.

so that the equatorial diameter exceeds the polar diameter by 42.77 km, which is about 170 m less than the value favored before the launch of satellites. Although 170 m is not much for practical purposes, it was of course important for geodesy, and the new value of the flattening showed that the actual earth differed in shape significantly from the shape that would be taken up by a purely liquid body, which would give 1 part in 299.7.

This was only the beginning of the use of orbit analysis for determining the earth's shape. In proceeding further, the simplest assumption to make is that the earth is symmetrical about its axis and has no variations with longitude, so that its shape is completely specified by a slice through the poles. The most helpful way of analyzing this shape is to assume that it is made up of an infinite number of harmonics, the second harmonic defining the flattening, the third harmonic often being called pear-shaped, the fourth harmonic being square-shaped, and so on, as shown in Fig. 2. These are the shapes that would be obtained by slicing the earth through the poles if only one particular harmonic existed. In fact all the harmonics exist, and the procedure is to try to calculate them separately and then put them all together to give the final shape.

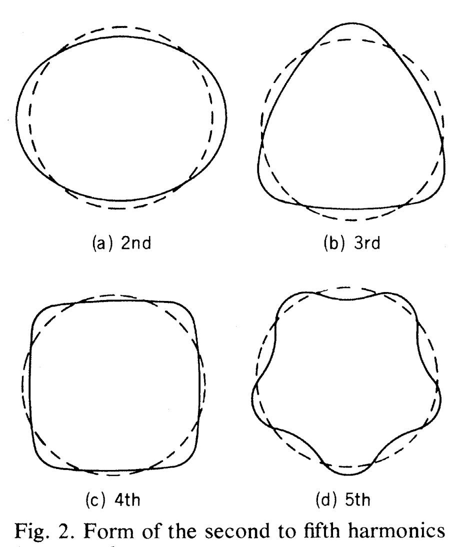


Fig. 2. Form of the second to fifth harmonics (not to scale).

Mathematically, the longitude-averaged gravitational potential $\overline{U}$ at a point distant $r$ from the earth's center is written as

$$\overline{U} = \frac{\mu}{r}\left[1 - \sum_{n=2}^{\infty} J_n \left(\frac{R}{r}\right)^n P_n(\cos\theta)\right]$$

where $\theta$ is the colatitude, $\mu$ is the gravitational constant for the earth (398,601 $km^3$ $sec^{-2}$), $R$ is the earth's equatorial radius (6378.14 km), $P_n(\cos\theta)$ is the Legendre polynomial of degree $n$ and argument $\cos\theta$, and the $J_n$ are constant coefficients which have to be determined. The first harmonic ($n = 1$) is missing, as the origin is at the earth's center of mass.

The constant $J_2$ expressing the main oblateness effect is by far the largest of the coefficients, being of order $10^{-3}$. All the others, $J_3$, $J_4$, $J_5$, . . . , are of order $10^{-6}$, and tend to decrease as their degree $n$ increases. Evaluating an infinite set of coefficients would require superhuman powers, and in practice the series is truncated, or sets of selected coefficients are evaluated.

Besides making the orbital plane rotate, the $J_2$ term produces a steady angular motion of the perigee, the point where the satellite comes nearest to the earth's center. For an orbit close to the earth the perigee moves at several degrees per day.

The most important perturbations to satellite orbits, after the major effects caused by the flattening, are those due to the third and other odd harmonics. If the earth were symmetrical about the equator, the perigee would remain at a fixed distance from the earth's center, apart from small lunisolar and other perturbations. But the asymmetry in the gravitational field produced by harmonics of degree 3, 5, 7, . . . makes the perigee distance change appreciably as the perigee swings from north to south of the equator. For a satellite in an orbit inclined at 45° to the equator, the perigee distance becomes about 10 km greater when the perigee is at its farthest south than when it is farthest north. In 1959 this effect was measured on the U.S. satellite Vanguard 1, and the measurements showed there was a tendency toward a pear shape with the stem at the north, a tendency amounting to about 30 m, as it was then thought (7).

Since 1959 it has become possible to evaluate a large number of harmonics which are independent of longitude—the zonal harmonics. Some recent sets of values are given in Table 1, and the profile of the geoid as given by those coefficients is shown in Fig. 3 relative to a spheroid of flattening 1 part in 298.25, which is shown as a broken line. The vertical scale is greatly exaggerated, and the surface is not really concave at the south pole. This profile is an average over all longitudes, not an actual physical slice through the poles; the profile should be accurate to within 1 m. The north pole is 19 m above the symmetrical figure, and the south pole is 26 m below. So if you bored a hole through the ice at the north pole and fell into the sea, you would be 45 m further from the equator than an equally foolish explorer who went to the south pole and burrowed down to sea level.

## Observations

So far I have said nothing about the observations used in determining the orbits on which these results were based. The measurements of the rotation of the orbital plane of Sputnik 2 relied on observations from kinetheodolites, backed up by visual observations made by volunteer observers in their back gardens (*10*). The kinetheodolites were originally intended for tracking missiles on trials ranges, and the volunteer observers represented an equally impromptu response. However, the kinetheodolites offered an accuracy of about 30 seconds of arc and the visual observers, after a reasonable spell of practice, were able to achieve accuracies of about 2 minutes of arc. The orbit of Vanguard 1 was determined by using the National Aeronautics and Space Administration (NASA) Minitrack radio interferometer observations, which are accurate to between 1 and 2 minutes of arc. These observations are effectively accurate to about 200 m for a satellite fairly close to the earth. The reason for the success of the analysis is that the orbital perturbations are much larger than the irregularities in the geoid. The tendency toward a pear shape amounting to 45 m gives rise to an oscillation of 10 km in the orbit, which is greater by a factor of about 200. So observations accurate to 200 m can potentially give results on the geoid height accurate to 1 m, and excellent values for the zonal harmonic coefficients can be found by using observations of only moderate accuracy.

During the 1960's much more accurate observational techniques became available, particularly photographic methods.

Table 1. Recent sets of zonal harmonic coefficients.

| Wagner (*8*) | | King-Hele and Cook (*9*) | |
|---|---|---|---|
| Coefficient | Value ($\times 10^9$) | Coefficient | Value ($\times 10^9$) |
| $J_2$ | 1,082,635 ± 11 | $J_3$ | −2,531 ± 7 |
| $J_4$ | −1,600 ± 12 | $J_5$ | −246 ± 9 |
| $J_6$ | 530 ± 26 | $J_7$ | −326 ± 11 |
| $J_8$ | −200 ± 29 | $J_9$ | −94 ± 12 |
| $J_{10}$ | −224 ± 45 | $J_{11}$ | 159 ± 16 |
| $J_{12}$ | −208 ± 17 | $J_{13}$ | −131 ± 22 |
| $J_{14}$ | 166 ± 25 | $J_{15}$ | −26 ± 24 |
| $J_{16}$ | 3 ± 55 | $J_{17}$ | −258 ± 19 |
| $J_{18}$ | −86 ± 56 | | |
| $J_{20}$ | −85 ± 61 | | |

Preeminent among these for its contribution to satellite geodesy was—and still is—the Baker-Nunn tracking camera of the Smithsonian Astrophysical Observatory, which has an aperture of 500 mm, an accuracy of 2 seconds of arc, and can track satellites down to magnitude 12, more than 100 times fainter than the faintest star visible to the naked eye. There were 12 of these cameras at work in the 1960's, and they have provided over a million observations (*11*). There were many other cameras in operation, the most accurate being the 600-mm Hewitt camera in Britain, with an accuracy of 1 second of arc. Other cameras in continual use included the AFU-75 in the Soviet Union and Eastern Europe, the Antares camera in France, the Wild BC-4 used by what was then the U.S. Coast and Geodetic Survey, and many others (*12*). Most of these cameras remain in operation today, though in somewhat less intensive use.

Another quite different technique of observation that reached a high degree of perfection during the 1960's was the radio Doppler tracking of satellites which carried very stable radio transmitters, developed by the U.S. Navy for their navigation system. Doppler methods were also used for studying the moon's gravitational field.

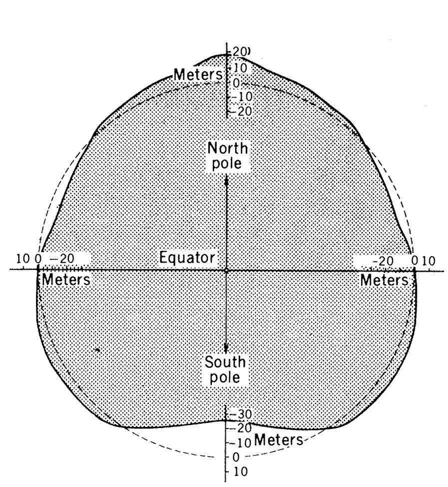


Fig. 3. Height of the meridional geoid section (solid line) relative to a spheroid of flattening 1 : 298.25 (broken line) with the harmonics of Table 1.

## The Geoid in the 1960's

Both the photographic and the Doppler techniques of observation offered accuracies of 5 to 10 m in position, and allowed great advances to be made in the study of the shape of the geoid during the 1960's. The observations could be used either dynamically, to obtain the gravitational field from orbital perturbations; or geometrically, to find station positions from simultaneous observations; or in a combination of these two techniques of analysis.

To take the orbital method first, the improved accuracy provided the chance to study the variations of the gravitational field with longitude. The orbital perturbations caused by the variations of the gravitational field with longitude are usually very much smaller than those caused by the variations with latitude, because the spinning of the earth tends to blur any sharp effects dependent on longitude. The satellite tends to sample all longitudes equally. The largest perturbations due to variations of the gravitational field with longitude are only a few hundred meters, and their frequency is usually near 24 hours or 12 hours, so that observations have to be concentrated closely in time. To define the variations of gravity with longitude, the potential has to be expressed in a double sum of tesseral harmonics, as they are called, which are dependent on both latitude and longitude; the potential is written

$$U = \overline{U} + \frac{\mu}{r} \sum_{\ell=2}^{\infty} \sum_{m=1}^{\ell} \left(\frac{R}{r}\right)^{\ell} P_{\ell}^{m}(\cos\theta) \times [C_{\ell m}\cos m\lambda + S_{\ell m}\sin m\lambda]$$

where $\overline{U}$ is the longitude-averaged potential previously defined, $P_{\ell}^{m}(\cos\theta)$ is the associated Legendre function of order $m$ and degree $\ell$, the $C_{\ell m}$ and $S_{\ell m}$ are constants to be determined, and $\lambda$ denotes longitude, positive to the east. Now we have a double infinite set of coefficients to evaluate in order to find the complete gravitational field. This is even more difficult to deal with, and in practice the matrix of coefficients is truncated, often at about the 16th order and degree, leaving only about 250 coefficients to be evaluated.

The alternative geometrical method for determining a geodetic network is simple in principle, but not so easy in practice. All you need to do is to observe a satellite simultaneously from several different stations, derive a large number of equations between the satellite position coordinates and the station coordinates, and then eliminate the satellite coordinates to obtain the station positions. It is not really necessary to know where the satellite is, or to know that the stations are on the earth's surface, although in practice it is useful to include this information among the equations to be solved. The most extensive network determined by purely geometric methods has been established by the cameras of the U.S. Coast and Geodetic Survey, which gave 45 worldwide station positions with a mean error of about 4 m (*13*).

The Doppler techniques were vigorously pursued in the 1960's by the U.S. Navy scientists, and excellent maps of the geoid were obtained by dynamic methods from these observations, as well as measurements of the polar motion, from the late 1960's onward (*14*).

I have to choose one geoid map to summarize the achievements of the 1960's, and my choice is the Smithsonian Standard Earth II, published in 1970 (Fig. 4). This map represents the culmination of many years' work with the Baker-Nunn cameras and also includes data from other cameras and from gravimetric surveys, plus a few of the first measurements from laser tracking of satellites. More than 100,000 photographic observations were used, and about 200,000 simultaneous equations were solved for about 400 unknowns, including nearly 300 harmonic coefficients and the station coordinates. This is a combined solution which utilizes both geometric and dynamic methods (*15*). Figure 4 shows contours of the geoid at 10-m intervals, relative to a reference spheroid of flattening 1 part in 298.25. The greatest departure of the actual geoid from this regular reference spheroid is a depression about 110 m deep south of India. Not far away, near New Guinea, is the highest hump, about 80 m high. Since the reference equator is circular, this means that if you swam along the equator from south of India to north of New Guinea, with diversions to avoid the land (and the sharks), your distance from the earth's center would increase by nearly 190 m, although of

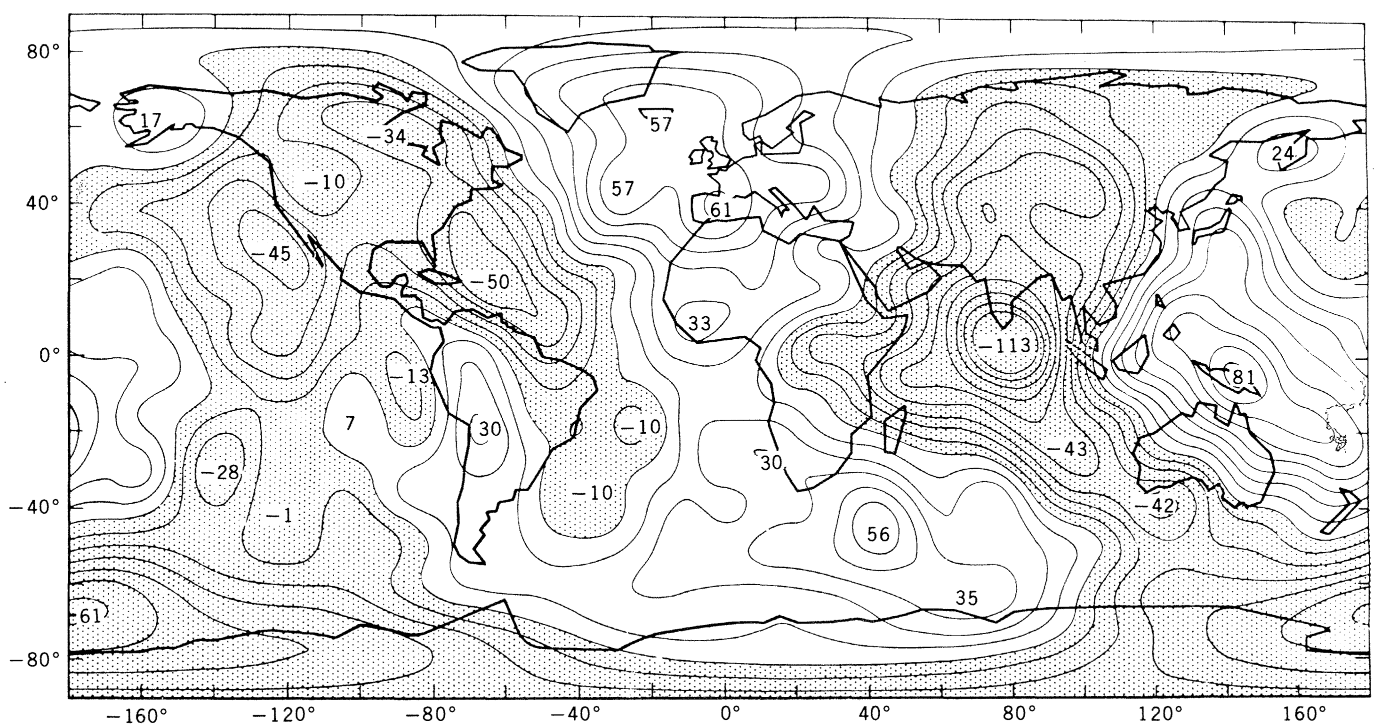

Fig. 4. Geoid of Smithsonian Standard Earth II. Contours are at 10-m intervals, relative to a spheroid of flattening 1 : 298.25.

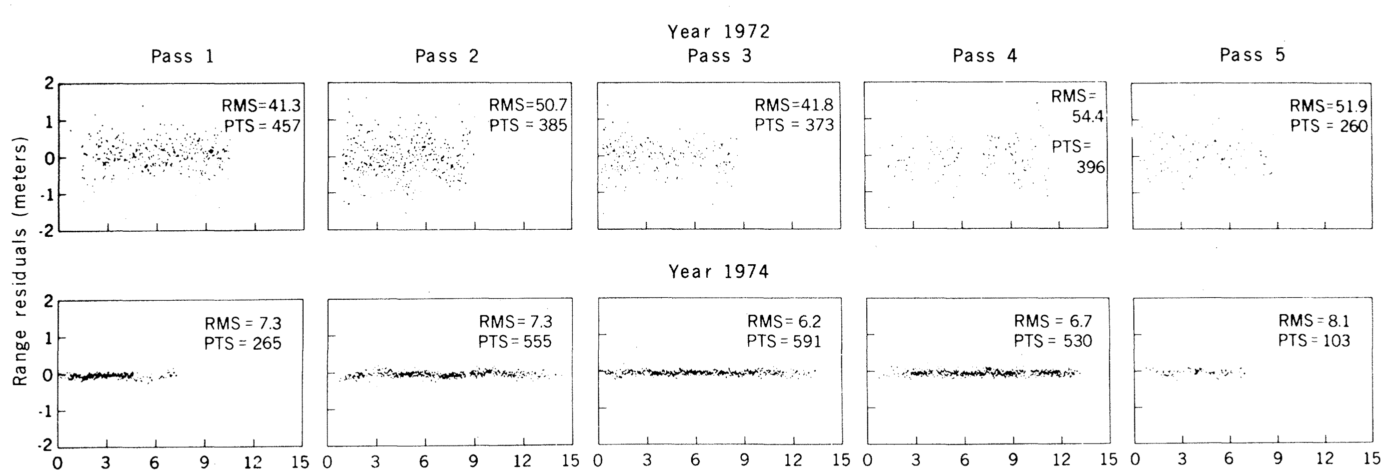

Fig. 5. Residuals from Goddard Space Flight Center laser observations of Beacon C (1965-32A) in 1972 and in 1974.

course you would never go uphill, because you are on the mean sea-level surface, which is level by definition. The other two major humps on the geoid, each about 60 m high, are centered near Britain and south of Madagascar, while there are major depressions between 45 and 60 m deep southeast of New Zealand and off California and Florida. A good way to remember this complex map is to focus on the dark areas, which are the regions below the reference spheroid, and if you look carefully you can see that the western hemisphere is dominated by a creature that looks rather like a goat; this animal is confronting a man from the east who has a strong chin, rather like Popeye, and covers most of Asia. Anyone with a fallible memory will welcome this simple way of visualizing the geoid.

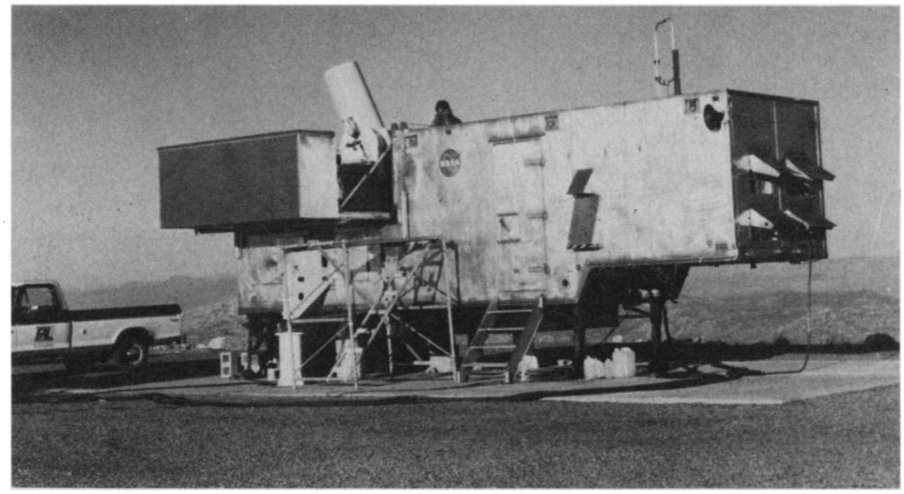

Fig. 6. The Goddard Space Flight Center satellite laser Moblas 2 in operation at Otay Mountain, near San Diego.

## Laser Tracking

Since this geoid map was produced in 1970, several new techniques of measurement have come into successful operation. The first that demands to be mentioned is laser tracking, a subject in which the United States and France have led the way. A laser pulse is sent from a ground station to a satellite carrying special corner reflectors, from which the pulse proceeds back along the same path and is detected at the ground station. The time of arrival of the pulse can be measured accurate to a few nanoseconds, and the distance of the satellite from the station can be measured with an accuracy which is now in the neighborhood of 10 cm. The first satellite with corner reflectors was launched in 1965, and now there are 12 such satellites in orbit which are suitable for laser tracking. Some laser observations were made in the late 1960's, but these were only accurate to a few meters, and were not really much better than the camera observations. But in the last 5 years there has been a dramatic improvement in the accuracy of laser tracking, as is neatly shown in Fig. 5: this gives the residuals for observations of the same satellite, Beacon C (1965-32A), made by lasers operated by the Goddard Space Flight Center in 1972 and in 1974. With the laser tracker used in 1972 the scatter was about 50 cm, but the tracking in 1974 gave root-mean-square residuals of about 7 cm. Further improvements can be expected and, as far as can be foreseen at present, the ultimate accuracy is likely to be about 2 cm or perhaps even 1 cm.

One of the advantages of laser tracking is that the instruments are quite small. The lasers used in the past few years have been typically ruby lasers with 10-nanosecond pulses, a 10-cm-aperture transmitter, and a 50-cm-aperture receiving telescope. Figure 6 shows the complete trailer of the NASA mobile laser System Moblas 2, as used at Otay Mountain, near San Diego, for the San Andreas Fault experiment, which has the aim of measuring the rate of movement of (nominally fixed) stations on opposite sides of the San Andreas Fault, correct to a few millimeters per year.

## The Geoid in the 1970's

The centimetric accuracy of the laser tracking is impressive, but we have to remember that these accuracies cannot be used effectively unless the model for the gravitational field is accurate; otherwise, orbital errors can become quite dominant. So better models of the gravitational field and more accurate geoids are badly needed, and attempts are continually being made to produce improved models, with the laser observations themselves now beginning to contribute notably toward this goal. Figure 7 shows the geoid in the recent GEM 8 model derived at the Goddard Space Flight Center, combined with measurements of surface gravity on a grid of 1° in longitude by 1° in latitude (*16, 17*). The GEM 8 model utilizes a total of 562,000 optical, radio, and laser observations of 27 different satellites, and includes 1600 surface gravity measurements over squares 5° by 5° in latitude and longitude. The data include 66,000 laser observations of satellites. The model includes harmonic coefficients complete to degree and order 25. Comparison of Fig. 7 and Fig. 4 shows that the new map of the geoid is quite similar to the previous map obtained 5 years earlier: this is just as well, because it is the same earth, and any variations of the geoid shape with time are considerably less than the current errors in the models. Figure 7 shows the dip south of India as 106 m deep, compared to 113 m in Fig. 4, and the hump near New Guinea is 86 m high instead of 81 m. The hump near Britain has moved toward the mid-Atlantic ridge south of Iceland, and the depression near Florida has moved its center toward the Puerto Rico trench. One of the greatest changes is the increased height of the hump in southern Peru, 52 m instead of 30 m.

These latest models of the geoid are probably accurate to about 2 m over most of the world, but the values of the individual harmonics of degree higher than about 10 are poorly determined, because the inversion of immense matrices produces some nasty correlations. So special techniques are needed for obtaining high-order harmonics.

One such method that has proved very successful is to look for satellites which happen to experience 15th-order resonance, when the satellite makes 15 revolutions per day, and its orbital period is such that the track over the earth repeats exactly after 15 revolutions. In other words, the track is 24° in longitude farther west on each revolution. Harmonics of order 15 in the geopotential can be regarded as having maximums and minimums every 24° in longitude, and so their effect on a resonant orbit tends to build up steadily from day to day (*18*), to produce a large orbital perturbation, which can be accurately analyzed. Figure 8 shows what happened to the orbital inclination at resonance for the satellite 1971-54A, one of the U.S. military launches, which chanced to pass through resonance in 1974, after decaying slowly for 3 years under the action of air drag. Figure 8 shows that the changes at resonance are well defined and the theoretical curve fits very well. With the aid of these methods it has been possible (*19*) to determine values for the individual coefficients of order 15 and degree 15, 17, . . . 33, which are very poorly defined in the comprehensive models of the geoid. It is also possible to find (though less accurately) values of 30th-order harmonics from analysis of 15th-order resonance. Similar analyses of resonances of order 11, 12, 13, and 14 are beginning to give values for coefficients of these orders, to improve the comprehensive solutions.

### New Techniques of Measurement

The analysis of resonant orbits is one dynamical technique of satellite geodesy which has already established that it has a special part to play. There are several others waiting to take the stage, and perhaps the most promising is satellite-to-satellite tracking, which was tried out on the Apollo-Soyuz flight in 1975. If two satellites in the same orbit, a few hundred kilometers apart, accurately determine their relative velocity by Doppler methods, this is a measure of the change in the gravitational field in the region between them. Alternatively, a satellite in a low orbit can be tracked from another which is in a geostationary orbit, and again the relative velocity can be determined. This has been done with the NASA satellites Geos 3 (1975-27A) launched in April 1975 in the low orbit and ATS-6 (1974-39A) in a geostationary orbit as the "fixed" tracking station, no longer on the earth like a normal tracking station, but perched high above a point on the earth's equator. As Geos 3 flies over the gravitationally rough areas, its velocity relative to ATS-6, measured accurate to 1 mm/sec, shows characteristic variations. Satellite-to-satellite tracking is potentially a very good method, but it is only just starting to be used.

Several new geometrical methods also look very promising. Outstanding among these is altimetry from satellites, using a radar altimeter to measure the satellite's height over the ocean. This was suggested many years ago, but was first used in the Skylab satellite in 1973, and now very good results are being obtained from the Geos 3 satellite. Some people did not quite believe the geoid shapes which had been determined dynamically: they cannot really credit the idea that there is a hollow in the Indian Ocean and a hump near New Guinea. Figure 9 shows the altimeter measurements of the sea-level surface profile from Skylab compared with the geoid profile given by the GEM 8 model: clearly, the altimeter measurements do agree in indicating a dip in the Indian Ocean and a hump near New Guinea. In fact the agreement is very good all over the world, to within a few meters. The altimeter gives finer detail, but may be less accurate absolutely, because it includes orbital errors. The new altimeter in Geos 3 is more accurate, and, when the results are fully analyzed, the relative heights should be accurate to better than 1 m.

There are many other techniques of satellite geodesy likely to be important in the future, which deserve a mention. Instruments aboard satellites designed to measure the gradient of gravity accurately as the satellite goes around the earth—called gravity gradiometers—may well prove to be the best method of measuring some of the high-order harmonics. But the instrumentation required is quite complex and expensive. Then there are the Doppler methods, both from the earth and from tracking close earth satellites from synchronous satellites, as Geos 3 is now being tracked by ATS-6. The Doppler methods have been in operation for a long time, but improvements in accuracy are continually being made. Another method that should not be forgotten is the technique of laser ranging to reflectors on the moon: this may seem rather remote from the earth, but the method has certain advantages which promise to give particular earth constants very accurately. There is also the drag-free satellite, using the technique of keeping a small sphere at the center of a large hollow sphere, with the larger sphere being moved by microthrust jets to keep the small sphere at its center and free of all forces other than gravity. This technique has been tested in the Triad satellite: it is excellent but technically difficult and requires elaborate instrumentation (*20*). An approximation to

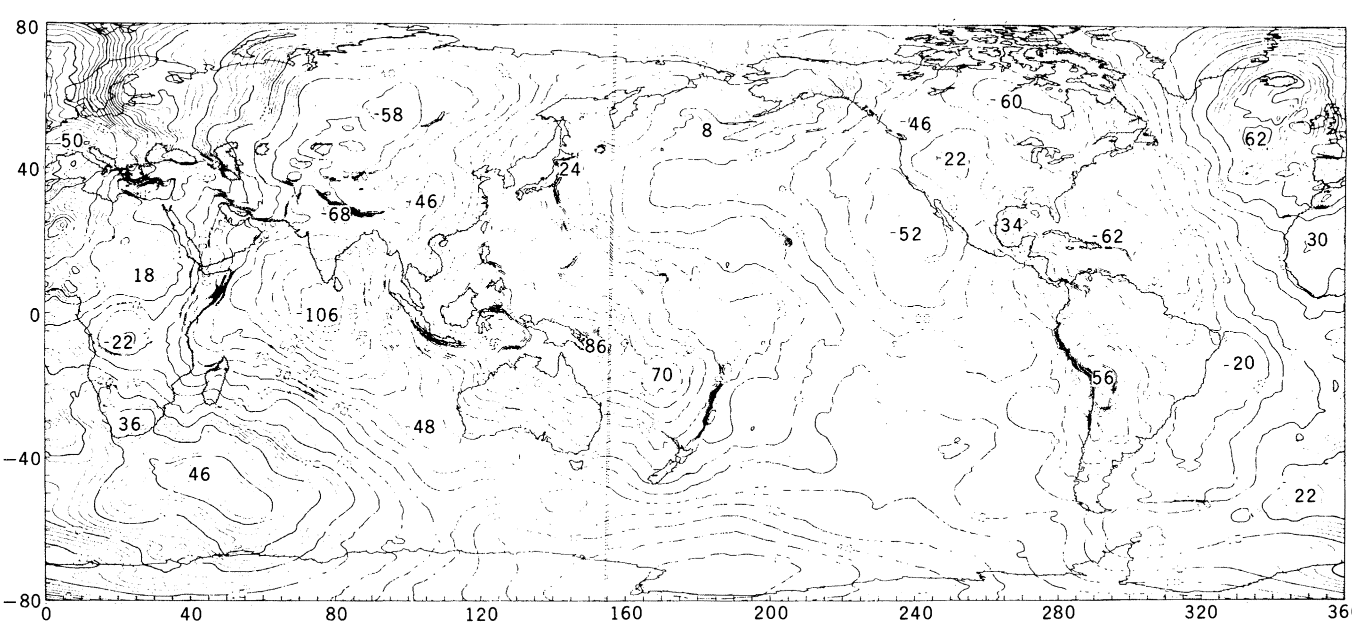


Fig. 7. NASA/Goddard Space Flight Center global detailed gravimetric geoid based on a combination of the GEM 8 earth model and surface gravity data measured on a grid of 1° by 1°. The contour interval is 2 m; earth radius 6378.142 km, flattening 1 : 298.255, and gravitational constant 398,600.9 km$^3$/sec$^2$. [From (*16*)]

this ideal is to use a very heavy, very dense satellite at a great height: this is the aim with the NASA laser geodynamic satellite, Lageos, which was launched on 4 May 1976. Lageos (see cover photograph) is a 411-kg spherical satellite, 0.6 m in diameter, carrying 426 retroreflectors, and is established in a circular orbit at a height of 6000 km. This satellite, tracked by lasers with an accuracy of perhaps 2 cm, may prove to be an important instrument for geodesy for the next 50 years.

### Future Prospects

In the past, for hundreds of years, the shape of the earth has been a fascinating subject, but a subject of academic interest rather than of practical importance. Of course, cartography and mapping have been practically important, but the presence of that 100-m depression in the Indian Ocean did not really bother anyone very much. Now, with the greatly improved accuracy of laser tracking and the other methods—an accuracy of about 10 cm at present, possibly reducing to 2 cm in 5 years' time—satellite geodesy is going to make a great impact on the other earth sciences.

As one example, consider the effect of exact knowledge of the gravitational field. This will provide a crucial test of theories about the earth's interior; already the map of tectonic plates has resemblances to the geoid maps. In future, patterns of convection currents or density irregularities within the earth will have to be fitted to satisfy the observed gravitational field. Together with the satellite altimeter measurements, the gravitational field determination will give an accurate profile of the geoid surface, including earth tides and ocean tides, which cannot fail to be of great benefit to oceanographers, whose present knowledge of the ocean tides has serious deficiencies.

Another example is a subject I have so far barely mentioned—the polar motion, or the locus of the point where the earth's axis of rotation cuts the surface. For many years it has been known that the pole executes small gyrations with an amplitude of about 10 m with two main periods, annual and 14-monthly, but also with many irregular variations which will be determined as the new centimetric geodesy develops, from laser tracking of earth satellites and from ranging to reflectors on the moon. These measurements, again apparently rather academic at first sight, may have repercussions in unexpected fields: in meteorology, because the variations in the angular momentum of the lower atmosphere directly affect the polar motion, causing annual and shorter-period variations; perhaps in oceanography, because changes in ocean current systems have a similar effect; and surely in seismology, because every earthquake shakes the polar axis slightly. To see the situation as mere cause and effect, as in the previous sentence, is probably too simplistic, however. The twitches in the earth's rotation, expressed by the polar motion and changes in the earth's rate of rotation, may themselves be a source of geophysical events, because they make the earth throw its enormous weight around jerkily, perhaps triggering movements of the tectonic plates. The earth is a self-exciting earthquake generator, and it looks as though satellite geodesy may in the end offer the key to the origins of the excitement. The kinematics of the earth's motion and the dynamics underlying it will have to be fully understood before we can claim to understand the earth itself, which is the aim of geophysics. I believe that satellite geodesy, in its new centimetric clothing, will be decisive in helping us toward a full understanding of the earth we live on.

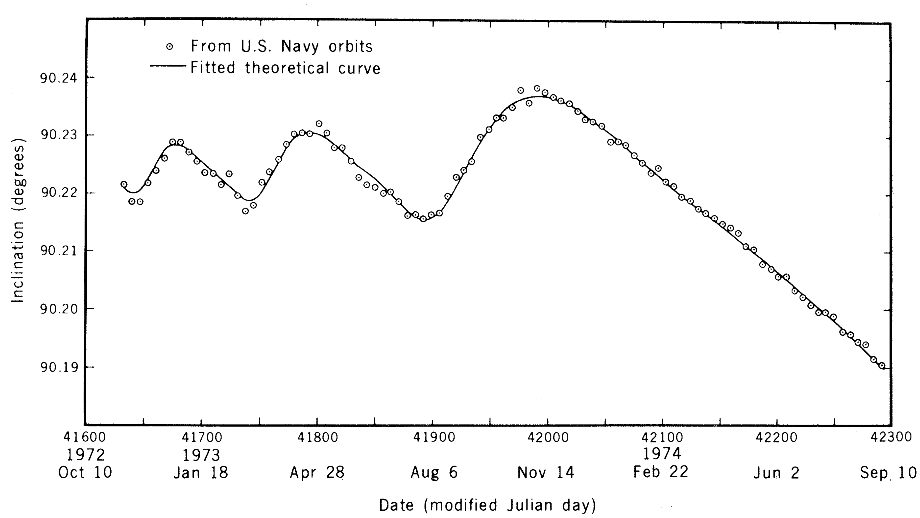


Fig. 8. Values of inclination at resonance, with fitted theoretical curve, for the satellite 1971-54A.

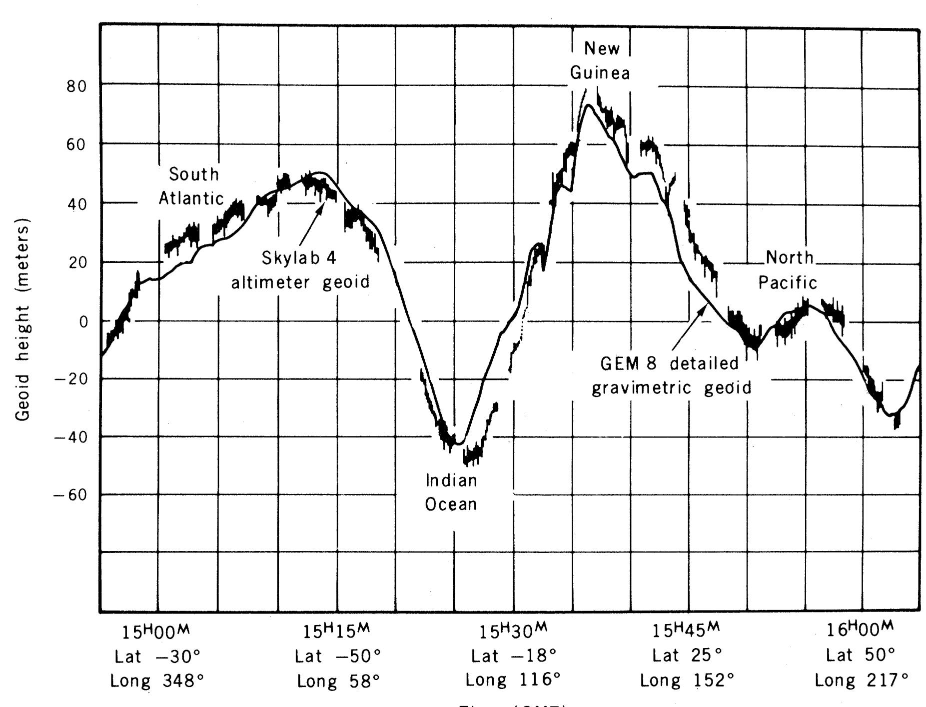


Fig. 9. Comparison of GEM 8 detailed gravimetric geoid with Skylab 4 "around the world" altimeter data (31 January 1974).

**References and Notes**

1. The accuracy achieved by Eratosthenes has been much argued. See I. Fischer, *Q. J. R. Astron. Soc.* **16**, 152 (1975).
2. W. Shakespeare, *A Midsummer Night's Dream*, act 2, scene 1, line 175 (1595).
3. I. Newton, *Principia* (Royal Society, London, 1687), p. 422.
4. For a description of the 18th-century French expeditions and their results, see I. Todhunter, *A History of the Mathematical Theories of Attraction and the Figure of the Earth* (Macmillan, London, 1873; reprinted by Dover, New York, 1962), chaps. 7 and 12.
5. This simple definition of the geoid skates over a number of treacherous complexities. See G. Bomford, *Geodesy* (Oxford Univ. Press, London, ed. 3, 1971).
6. R. H. Merson and D. G. King-Hele, *Nature (London)* **182**, 640 (1958).
7. J. A. O'Keefe, A. Eckels, R. K. Squires, *Science* **129**, 565 (1959).
8. C. A. Wagner, *J. Geophys. Res.* **78**, 3271 (1973).
9. D. G. King-Hele and G. E. Cook, *Planet. Space Sci.* **22**, 645 (1974).
10. D. G. King-Hele, *Observing Earth Satellites* (Macmillan, London, 1966), chaps. 5 and 6.
11. The early years of the Baker-Nunn camera operations have been described by E. N. Hayes, *The Smithsonian's Satellite Tracking Program* (Smithsonian Institution, Washington, D.C., part 1, 1962; part 2, 1964; and part 3, 1965).
12. See A. G. Massevitch and A. M. Losinsky, *Space Sci. Rev.* **11**, 308 (1970).
13. H. H. Schmid, *J. Geophys. Res.* **79**, 5349 (1974).
14. R. J. Anderle, *U.S. Nav. Weapons Lab. Tech. Rep. TR-2952* (1973).
15. E. M. Gaposchkin and K. Lambeck, *Smithson. Astrophys. Obs. Spec. Rep. 315* (1970); shortened version in *J. Geophys. Res.* **76**, 4855 (1971).
16. S. Vincent and J. G. Marsh, paper presented at the International Union of Geodesy and Geophysics Assembly, Grenoble, France, August 1975.
17. C. A. Wagner, F. J. Lerch, J. E. Brownd, J. A. Richardson, *Goddard Space Flight Center Rep. X-921-76-20* (1976).
18. The theory was developed by R. R. Allan, *Planet. Space Sci.* **15**, 53 (1967); *ibid.*, p. 1829; *ibid.* **21**, 205 (1973).
19. See D. G. King-Hele, D. M. C. Walker, R. H. Godding, *ibid.* **23**, 1239 (1975); also, for coefficients of even degree, *ibid.*, p. 229. For an example of 13th-order resonance, see C. A. Wagner, *J. Geophys. Res.* **80**, 3791 (1975).
20. For descriptions of gravity gradiometers, drag-free satellites, and various other techniques, see G. Veis, Ed., *The Use of Artificial Satellites for Geodesy and Geodynamics* (National Technical University, Athens, 1974), especially pp. 135–192.


# The Arctic Mirage and the Early North Atlantic


H. L. Sawatzky and W. H. Lehn


The desert mirage, or *fata morgana*, is an almost universally familiar phenomenon. It occurs when air overlying a warm surface is heated by conduction, which induces temperature and therefore density gradients of a magnitude sufficient to alter the "normal" optical properties of the air. Refraction, and mirroring at significant surfaces of discontinuity, cause the path of light passing from an image to an observer to become not a straight line, but a complex of changes in direction, with the result that the image is optically displaced, vertically and often also laterally, from its actual location. False lakes, floating mountains, inverted images, and the like are commonly experienced examples of the desert mirage.

There is an equivalent but essentially inverse effect, much less frequently experienced, yet nevertheless not rare in the middle and higher latitudes. This is the arctic mirage or, using the Icelandic word for the phenomenon, the *hillingar* effect, in which the observed image is optically displaced, in a vertical direction only, from its real location. We propose that the arctic mirage may have been a significant factor in the development of certain early historical concepts of the nature of the world, and possibly in exploration and discovery.

## Nature of the Arctic Mirage

The arctic mirage occurs when air rests on a pronouncedly colder surface. This situation establishes (i) a higher vapor pressure in the immediately superincumbent air than at the surface; (ii) temperature inversion conditions which impart a high degree of stability; and (iii) a vertical temperature gradient which imparts particular refractive properties to the affected air (*1–3*). Light passing from an image to an observer, within a band of air described by these conditions, is continuously refracted in an arc with the earth on its concave side. The result, over large, relatively featureless surfaces, is to permit the passage of light from an object to an observer over distances greater than the normal distance to the horizon (see Fig. 1) (*4, 5*).

Although warm air advected over a colder surface may cause the phenomenon, the key to the occurrence of the arctic mirage in its clearest expression are large, relatively stationary high-pressure cells. Subsidence within such air masses results in adiabatic heating, at a uniform rate of approximately 1°C per 100 meters. As the air is simultaneously cooled by radiation and by conduction from the colder surface, a temperature inversion, whose depth and magnitude are determined by the relative motionlessness of the air mass and the temperature characteristics of the surface and superincumbent air, comes into being. The higher vapor pressure prevailing in the air inhibits the escape of moisture from the surface, and hence the formation of fog or mist, leaving the almost vapor-free air extremely transparent. The high degree of atmospheric stability that characterizes the situation further facilitates the direct passage of light, while the positive upward temperature gradient produces an exaggerated negative upward density gradient, which substantially enhances the normal refractive capability of the air (see Fig. 2).

Given a sufficiently steep vertical temperature (density) gradient, the refractive capability of the air may become great enough to equal or exceed the curvature of the earth, which to the observer is thus optically rendered flat or concave upward, with the horizon receding toward infinity, or, in practical terms, toward the limits from which light can, by virtue of the transparency of the air, pass from an image to an observer. The threshold condition for achieving a refractive effect matching the curvature of the earth is established by a temperature increase with height of approximately 0.11°C per meter. A higher gradient causes the surface to appear saucer-shaped. Observations of vertical temperature gradients as great as 0.5°C per meter, under conditions of strong inversion, are on record (*6*). However, much weaker inversions, which can have vertical profiles 300 meters or more deep, still have a greater refractive capability than normal air, and may therefore raise ele-


H. L. Sawatzky is an associate professor in the Department of Geography and W. H. Lehn is an associate professor in the Faculty of Engineering at the University of Manitoba, Winnipeg, Canada R3T 2N2.