PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY A MATHEMATICAL, PHYSICAL & ENGINEERING SCIENCES

A Brief Survey of Satellite Orbit Determination

R. H. Merson

Phil. Trans. R. Soc. Lond. A 1967 **262**, 71-78
doi: 10.1098/rsta.1967.0032

II. ORBIT DETERMINATION

A brief survey of satellite orbit determination

By R. H. Merson

Royal Aircraft Establishment, Farnborough, Hants

The accurate determination of satellite orbits depends on an adequate accumulation of observations, a sound dynamical theory and a fairly sophisticated sequence of numerical computations. The particular patterns of observation, theory and computation are considered in relation to the objectives of orbit determination. Factors to be taken into account are the type, accuracy and spread of observations; perturbations of the orbit due to air drag, attraction of the Earth, Moon, and Sun, and solar radiation pressure; and the speed and cost of available computers. These factors, together with the overall objectives, determine the main features of the computation; whether to use special or general perturbation techniques, what length of orbit arc to use, what parameters to determine and how to present the results.

1. Introduction

The motions of the heavenly bodies have interested mankind since the beginning of time and have governed many aspects of daily life through the ages. From the point of view of modern space research, however, the science of orbital astronomy started with the work of Tycho Brahe, Kepler and Newton in the sixteenth and seventeenth centuries (see Pannekoek 1961). Tycho Brahe made observations of the positions of Sun, Moon and planets relative to the star background. Kepler, using Tycho's observations of Mars, discovered the elliptical nature of its orbit and the general laws of motion of the planets around the Sun. He was the first to determine the orbital elements of the planets. Eighty years later Newton, combining the laws of Galileo for falling bodies with those of Kepler for celestial bodies, formulated the law of gravitation, from which all subsequent celestial mechanics has developed.

These three men, Tycho Brahe, Kepler and Newton, typify the three divisions into which the study of orbits naturally falls: observation, computation and dynamical theory. In the case of man-made satellites and space probes the general inter-relation between these three divisions is somewhat as follows. Observations of the object are made soon after it is launched and, using some theoretical model of the motion, a preliminary orbit is calculated. From this orbit predictions are made of future positions of the object, enabling further observations to be made. The orbit is then improved, perhaps in several stages, until finally a definitive orbit is obtained which fits all the available observations as well as possible. From studies of the orbits of many objects deductions have been made about the physical phenomena which affect them and these in turn have led to improved theories and better models.

There are several different reasons for determining orbits, and these will be outlined in §2. In some cases the specific requirements have led to the establishment of networks of observing stations, or sensors as they are now generally called. These networks have since been used for observing a variety of objects for different purposes. In other cases observing stations, particularly those involving very little equipment, have been set up without any specific scientific objective in mind. Most of these are now coordinated in some degree and it is probably true to say that the majority of observations made to-day are made with some end in mind.

2. Objectives of orbit determination

The objectives of orbit determination can be split into two groups, which may be called operational and non-operational. Operational objectives are concerned with the launching and flight of a space vehicle and with the requirements of on-board instrumentation. Non-operational objectives are concerned with the accumulation of scientific data not directly related to the launching objectives. The division into the two groups is not completely clear-cut and there are a few borderline cases.

2·1. *Operational objectives*

2·1·1. *Orbit achievement.* The mere knowledge that a satellite has gone into a required orbit is of primary importance, and determination of the orbit reflects back on the accuracy and performance of the launching vehicle.

2·1·2. *Guidance and control.* For space probes particularly, where trajectory changes are usually required, the orbit is continually determined in order to know when and what changes to make.

2·1·3. *Position of on-board instrumentation.* When on-board instruments are measuring properties of the upper atmosphere or beyond, the position of the instrument or perhaps some component of position such as the height above the Earth is required, usually with the best possible accuracy.

2·1·4. *Tracking and telemetry read-out.* The scheduling of observing stations for tracking and telemetry read-out requires a continuous knowledge of the orbit, though not to any great accuracy.

2·2. *Non-operational objectives*

2·2·1. *General surveillance.* Both for scientific and military defence purposes it is necessary to keep track of all satellites above a certain minimum size. From the scientific point of view it is necessary to be able to distinguish one satellite from another and from the military point of view it is required to detect and classify new objects which may be of military significance.

2·2·2. *Air density.* Satellites passing through the upper atmosphere suffer loss of energy due to atmospheric drag, and by studying the decay of a satellite it is possible to estimate the air density in the vicinity of the perigee. From the orbits of many satellites a general picture of the neutral atmosphere between about 200 and 1000 km has been built up.

2·2·3. *Gravitational constants.* The major force on a space object close to Earth is the central attraction of the Earth, the physical constant involved is the product GM, the product of the universal gravitational constant and the Earth's mass. GM has been determined to about 1 part in 400000 by the method of radio Doppler tracking of lunar probes.

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY A MATHEMATICAL, PHYSICAL & ENGINEERING SCIENCES

Apart from atmospheric drag, the major perturbing force on a near-Earth satellite is due to the departure from spherical symmetry of the Earth's gravitational field. If the potential is expanded in a series of spherical harmonics, the coefficients of these terms can be related to the amplitudes of perturbations in the satellite's motion.

From the determination of the orbits of lunar satellites it will be possible in the near future to obtain more accurate knowledge of the mass and gravitational field of the Moon than is at present available.

2·2·4. *Navigation, geodetic data.* If the orbit of a satellite is accurately known, and observations of some components of its relative position and/or velocity are made from a station on the Earth then it is possible to determine the location of the observing station. With an active satellite system, that is, satellites that are transmitting signals, the way is open for a navigation system that is independent of atmospheric conditions. With passive satellites, or with special satellites emitting light flashes, and observed by very accurate cameras, it is possible to strengthen the geodetic ties between various places and continents on the Earth. In general, the positions of the observing stations are determined at the same time as the satellite orbit, using the same observations. Simultaneous observations from two or more places can also be used, but since this 'direct' method does not involve the determination of the orbit, it does not concern us here.

Satellites have, of course, been launched specifically for navigation and geodetic purposes and the classification in these cases should perhaps be 'operational'.

2·2·5. *Sensor accuracy.* The size of the random component of error of a sensor can be estimated by studying the equipment itself. Biases, however, can only be determined by studying the residuals of observations with respect to some computed orbit.

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY A MATHEMATICAL, PHYSICAL & ENGINEERING SCIENCES

3. Orbit computation

3·1. *Preliminary orbits*

The determination of a preliminary or initial orbit from a small number, possibly a non-redundant number, of observations is not of very general interest.

Some operational groups will determine preliminary orbits for their own use, but virtually the whole of the available information on such orbits is provided by the American surveillance networks (Spadats, Spasur), so that the majority of workers do not have to bother with them. A good account of methods used for determining preliminary orbits has been given by Escobal (1965).

3·2. *Definitive orbits*

The general procedure for all definitive orbit computations is to set up some dynamical model of the orbit, and use the observations to improve the parameters of the model by the process of differential correction. The model can either be a set of differential equations representing the motion, or a set of functions of time representing the changes in fundamental parameters of the motion. In the former case the orbit is generated by numerical integration of the differential equations of motion, whereas in the latter case the model functions are obtained by analytical integration of the equations of motion. In the terminology of celestial mechanics, numerical integration procedures are called 'special perturbation methods' and the analytical procedures are called 'general perturbation methods'.

For both 'integration' and 'analytical' procedures, as I prefer to call them, there are basically two ways in which the observations can be introduced. One way is to take all observations during some restricted period of time, and, by differential correction of the parameters of the model, minimize the sum of squares of observation residuals. The second way is to introduce the observations one at a time in chronological order and use a numerical filtering technique to estimate the system variables and other quantities.

In all methods it is necessary to compute the partial derivatives of observations with respect to the model parameters, i.e. the basic quantities defining the orbit, and again there is a choice of procedure. At one extreme the parameters are varied by small amounts one at a time and the effect on the observations is obtained, at the other extreme analytical partial derivatives, based on a simplified model, are used. There is also an intermediate possibility, whereby the derivatives are generated by an integration procedure incorporating only the dominant orbit perturbations.

In the following sections some of the advantages and disadvantages of the various procedures will be considered in relation to possible objectives, and examples will be cited.

4. Analytical methods

As far as is known, practical orbit determination of artificial space objects by analytical methods has been confined to Earth satellites. Although many theoretical studies of space probe missions and interplanetary missions have been carried out using the 'zone of influence' technique, operational orbits of lunar and other space probes have been obtained by an integration procedure. This is a fundamental necessity at the present time since analytical methods cannot handle the transfer from one major attracting body to another with sufficient accuracy.

In the case of Earth satellites, on the other hand, accurate orbits can be obtained using analytical methods and, in fact, most satellite orbits are determined in this way, mainly because the computer costs are very much less than in the case of an integration procedure of comparable accuracy.

All analytical methods need an adequate perturbation theory. The depth of the theory required depends on the accuracy of the observations, the length of orbital arc and the size of the perturbations themselves. The case of short, long and medium arcs will now be considered in turn.

4·1. *Short arcs*

The advantage of using very short arcs (5 to 15 min of time) is that luni-solar, solar radiation and even drag perturbations can be neglected. The first orbit determination program at the R.A.E., in November 1957, used observations from a single pass over an interferometer or kinetheodolite station (King-Hele & Merson 1958). No perturbations were included at all, but the results obtained over a period of time were good enough to show that the pre-satellite value of the Earth's flattening was 1 part in 300 too large.

A more sophisticated short-arc method was used by Rossoni, Sconzo & Winfield (1964) with observations of the satellite Anna. The perturbations due to the second, third and fourth zonal harmonics (i.e. the J_2, J_3, J_4 terms) in the Earth's potential were computed by Brouwer's (1959) method, all other perturbations being neglected. Although very good observations were used and good fits were obtained, the results are somewhat doubtful.

The trouble is that small biases in the system, in the observations and in the fixed parameters, cannot be detected and are not smoothed out.

4·2. *Long arcs*

The main advantages of using long arcs (of several months) are that far less observations are required *in toto* than for shorter arcs and that there is a considerable reduction in the overall computer time required. Very little work appears to have been done in this field, although Michielsen (1964) has obtained good fits to optical observations of Discoverer satellites. He determines the geometric parameters e, ω, i, Ω by a differential correction procedure using only the cross-track components of observation residuals. The along-track components are used to adjust the secular perturbations due to the Earth's gravitational field. Long-periodic perturbations are handled by including sinusoidal variations of the geometric parameters in the model, the amplitudes of the variations being 'free', i.e. determined along with the parameters themselves. Michielsen has used his orbits to study the zonal harmonics of the Earth's field, and his method deserves close attention.

It is fairly obvious that long arcs will not be suitable for the direct analysis of very accurate observations aimed at geodetic results, for it is doubtful whether the orbit perturbations can be computed to sufficient accuracy; but there is a distinct possibility that air density studies could be speeded up by developing the long-arc technique.

4·3. *Medium arcs*

In order to avoid the general difficulties of short arcs, namely the lack of sufficient observations and the problem of biases, and the difficulties of computing perturbations for long arcs, most analytical orbit determination programmes use medium arcs in the region of 1 to 8 days. An epoch is chosen near the middle of the arc and estimates of orbital parameters at the epoch are obtained by a differential correction procedure. In defining the model it is usual to use the standard elements of classical celestial mechanics, a, e, i, Ω, ω and M (or the time at some node or perigee). The complete model is specified by showing how to obtain the position x, y, z of the satellite at any time t (and the velocity $\dot{x}$, $\dot{y}$, $\dot{z}$ as well, if rate observations are included) in an Earth-centred inertial frame, in terms of the elements at the epoch. It is at this stage that the various perturbation theories are used. Of theories of the perturbations due to the zonal harmonics of the Earth's gravitational field, the most elegant and for that reason perhaps the most widely used, is that of Kozai (1959). Kozai gives a first-order theory of short-periodic, secular and long-periodic perturbations, omitting only the long-periodic perturbation in mean anomaly and the transformations necessary when rate observations are used. Merson (1966) has recently remedied these omissions in designing the model for a new orbit determination program, PROP, at R.A.E. Kozai's theory is the basis for the D.O.I. (Differential Orbit Improvement) program of the Smithsonian Astrophysical Observatory, which is used for the analysis of the very accurate Baker-Nunn camera observations. It is also used by Spadats (Colorado Springs) in their current routine program for space surveillance, and by Esdac (Darmstadt), who will be responsible for computing the orbits of E.S.R.O. satellites.

The conceptually simplest general perturbation procedure is that of Von Zeipel, on which Brouwer (1959) has based his theory of satellite motion. Brouwer's theory has been incorporated into computer programs at various centres, notably at N.A.S.A.'s Goddard Space Flight Center where the routine analysis of Minitrack observations is carried out.

The theories of Kozai and Brouwer, together with the similar theories of Merson (1963 *a*), and Zhongolovich (1960), are essentially equivalent, differing only in their definitions of orbital elements (which lead to slight but important, differences in the actual formulae) and in certain minor details. Because of the differences in definitions it is important that orbital elements derived by one theory should not be used for predictions or other purposes using another theory. The Russian theories (Zhongolovich 1960; Proskurin & Batrakov 1959) have been incorporated in B.E.S.M. computer programs in Moscow and Leningrad.

After the zonal harmonics, the second important perturbation for near-Earth satellites is that due to atmospheric drag. The size of the drag is measured by determining the acceleration of the mean motion (equivalent to the rate of change of orbital period) and this is added to the list of parameters to be adjusted. This parameter can then be used to fix the secular perturbations of the geometric parameters. It was thought for some time that tesseral harmonics in the Earth's gravitational field need not be taken into account in medium-arc determinations. However, in attempting to determine the orbit of Ariel 2 from Minitrack observations using the R.A.E. Pegasus computer program (Merson 1963 *b*) (based on Merson's theory), time residuals appeared which when plotted against sidereal time had a period of one day and amplitude of about 0·2 s. These were found to be due to the effect of the leading tesseral harmonic terms (coefficient $J_{2,n}$), and Gooding (1964) showed how to take account of these by a simple addition to the mean anomaly formula.

The most troublesome perturbations are those due to the attractions of the Sun and Moon and to the Sun's radiation pressure. Formulae of one sort or another are available for the main secular perturbations, but no one as yet appears to have developed criteria for asserting when they need to be used in medium-arc determinations and if so, which terms should be included.

The accumulation of orbital data determined using medium arcs has been the main source for studies of air density and of the Earth's gravitational field. In addition, the observation residuals which have been generated have contributed much to our knowledge of the actual accuracy of sensors.

The accuracy of the orbit determinations depends mainly on the spread of the observations in position and time, the best orbits coming from observations made by networks of stations with a good latitude distribution.

5. Integration methods

The method of numerical integration of the equations of motion has the advantages that it can be used in many cases where the analytical method fails and that the formulation of the equations is relatively easy compared with the rather complex algebra that is required in analytical theories. Because of the relatively high frequency of certain perturbations, step lengths of the order of 1 or 2 min have to be used for near-Earth objects, relaxing to perhaps 15 min when the object is several Earth radii away. This leads to the main disadvantage of integration methods, which is the large computation load, probably an order of magnitude greater than for analytical methods.

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY A MATHEMATICAL, PHYSICAL & ENGINEERING SCIENCES

There are three well-known procedures for generating an orbit by numerical integration, namely, Cowell's method, Encke's method and the method of the variation of parameters.

In Cowell's method the equations of motion are expressed in cartesian coordinates and are directly integrated using a high-order predictor-corrector technique.

In Encke's method some reference orbit is taken, perhaps the normal two-body orbit in the absence of perturbations or perhaps the two-body orbit with certain secular perturbations included (Escobal 1966). The equations that are integrated represent the departure of the actual orbit from the reference orbit. Various integration procedures have been used, including the 4th order Runge–Kutta method and predictor-corrector methods of different types. The numerical difficulty arises of treating the differences between nearly equal quantities and considerable ingenuity has been exercised in its elimination.

In the third method Lagrange's equations for the variation of orbital elements are integrated. This method has been used at the R.A.E. to provide accurate data for checking analytical expressions, but appears to have been little used generally. Although comparable with the Encke method in accuracy, the equations are more complicated and consume more computing time.

As has already been indicated, observations can be introduced in batches as in the normal analytical methods or they can be introduced one at a time.

5·1. *Batch processing*

Observations extending over some period of time, usually not more than two days, are taken, and starting with approximate initial position and velocity components, the equations of motion are integrated and observation residuals obtained. The initial conditions, and possibly other parameters as well, are then adjusted by a differential correction procedure.

Four such major computing programs are known, at N.W.L. (Anderle 1965), A.P.L., J.P.L. (Miller *et al.* 1966) and Spadats, all in the United States and there are probably several more. All these four use Cowell's method for orbit integration,† probably because it is easier to implement than Encke's method, although it requires a shorter step length and is of inferior precision to the latter. The programs differ in their handling of partial derivatives. In a program designed for Earth satellites only (A.P.L.) the simple two-body derivatives are used. Only negligible second-order errors are thereby introduced and the computation is much abbreviated. In the case of programs designed for space probes, however, the variant approach is adopted. In this, small changes in the six initial conditions are made one at a time and the changes in the observations are determined. This means that the differential equations have to be integrated six times to generate the partial derivative matrix. Provided the initial conditions are reasonably close to the true values, however, the partial derivative matrix need only be determined once and need not be recalculated during subsequent iterations of the adjustment procedure. The J.P.L. program (Miller *et al.* 1966) has been used for determining the Earth constant *GM*. The A.P.L. program is used in connexion with the U.S. Navy Satellite Navigation System. Both the A.P.L. and N.W.L. programs have been used for determining zonal and tesseral harmonics of the Earth's gravity field.

† A.P.L., and possibly others, have recently switched to the Encke method.

5·2. *The minimum variance method*

In the minimum variance method the deviations of the position and velocity from a reference trajectory are calculated at each observation time as a linear combination of the previous deviations and the current observation, using the filtering technique of Kalman (1960). This method has been developed extensively at the Goddard Space Flight Center (N.A.S.A. 1962) in connexion with the Apollo manned space flight project. Its main advantage over the least squares procedure is that an up-to-the-minute best estimate of the orbit is obtained, a prime requirement for an operational mission involving course corrections. The information obtained, however, is no more accurate than that obtained by the least squares procedure at much less computing cost so that the minimum variance method should not normally be used for non-operational objectives. The only exception that can be seen at the moment is in the possible introduction of a stochastic model for testing drag fluctuations in the manner proposed by Rauch (1965).

6. Concluding remarks

It has not been possible, in such a short survey, to mention more than a few of the important concepts that underlie the large volume of work that is now being done in orbit determination. Such problems as the singularities of elements in circular and equatorial orbits, the pseudo-resonance at the critical inclination and resonances caused by luni-solar and tesseral harmonic perturbations, amongst others, have had to be omitted. Nor has it been possible to name more than a few of the computer programs that exist. Nevertheless, it is hoped that a framework has been provided which will link together the subsequent papers of this symposium.

References (Merson)

Anderle, R. J. 1965 *Proc. COSPAR-IUTAM-IAU Symp., Paris.*
Brouwer, D. 1959 *Astr. J.* **64**, 378–97.
Escobal, P. R. 1965 *Methods of orbit determination.* New York: John Wiley and Sons.
Escobal, P. R. 1966 *A.I.A.A.J.* **4**, 5.
Gooding, R. H. 1964 *R.A.E. Tech. Memor.* no. Space 41.
Kalman, R. E. 1960 *J. Basic Engng*, **82**, 1.
King-Hele, D. G. & Merson, R. H. 1958 *J. Br. Interplan. Soc.* **16**, 446–71.
Kozai, Y. 1959 *Astron. J.* **64**, 367–77.
Merson, R. H. 1963*a* *R.A.E. Tech Note* no. Space 26.
Merson, R. H. 1963*b* *Dynamics of satellites* (edited by M. Roy), p. 83. Berlin: Springer-Verlag.
Merson, R. H. 1966 *R.A.E. Tech. Rep.* no. 66255.
Michielsen, H. F. 1964 Lockheed Missiles and Space Co. Res. Lab., Palo Alto, Calif., unpublished rep.
Miller, L. W. *et al.* 1966 *Jet Prop. Lab. Calif. Inst. Tech., Tech. Rep.* no. 32-911.
Pannekoek, A. 1961 *A history of astronomy.* London: Allen and Unwin.
Proskurin, V. F. & Batrakov, Yu. V. 1959 *Artificial Earth Satellites*, no. 3.
Rauch, H. E. 1965 *A.I.A.A. J.* **3**, 4.
Rossoni, J. P., Sconzo, P. & Winfield, D. 1964 *Air Force Cambridge Res. Lab. Rep.* AFCRL no. 64-273.
Special Projects Branch N.A.S.A. 1962 *N.A.S.A. Rep.* X-640-62-191.
Zhongolovich, I. D. 1960 *Bull. Inst. Th. Astr.* **7**, 743.