5.0 NUMERICAL INTEGRATION

5.1 INTRODUCTION

It was decided to model the dynamics of the three outer satellites of Saturn (Titan, Hyperion and Iapetus) using a numerical integration. This method has been employed by Sinclair and Taylor who have fitted an integration to astrometric observations made during the period 1967 to 1983.

In this chapter we describe the fitting of a numerical integration to visual observations of the three satellites over the period 1874 to 1947. Such visual observations have not been analysed using numerical integration before and this work represents a significant development in the study of natural satellite dynamics. Apart from being a valuable exercise in its own right, it is an important preliminary to the goal of linking pre-1947 visual observations with post-1967 photographic observations in a global solution involving a numerical integration of the satellite orbits over a period of 120 years.

The reasons for adopting numerical integration as a dynamical model are as follows :

(1) As long as we include all significant gravitational effects in the force model then a numerical integration provides a dynamically consistent representation of a satellite system. In this sense it has an advantage over analytic theories whose series must be truncated after a relatively small number of terms if they are to be convenient to use and easy to develop.

(2) The number of free parameters in a numerical integration can be restricted to the dynamically consistent minimum set : a position vector and velocity vector for each satellite plus the masses of the disturbing bodies and the form factors of the primary. In this way, we may avoid the introduction of pseudo-arbitrary parameters which are adopted in some analytic theories and which allow the least-squares fitting process to give artificially small root-mean-square residuals.

(3) The analytic theory of Hyperion is a problem of great complexity. Newcomb has placed it second only to the lunar theory in terms of the difficulty of its formulation. It has been studied by Newcomb, Woltjer and Message and is currently the subject of work by Message, Taylor and Sinclair (see for example Taylor (1984)).

Numerical integration of the motion of Hyperion, by contrast, presents no more difficulty than for any other satellite and hence allows observations of Hyperion to be analysed together with those of Titan and

Iapetus. The results of such an analysis will not be prejudiced by the shortcomings of an analytic theory.

(4) Once we have found a set of parameters with which the integration gives the closest fit to the observations, we may regard an integration based upon these parameters as effectively representing the observations. The information which was contained in the observations is now inherently contained in the parameters in conjunction with the chosen integration algorithm.

This means that we can use an ephemeris produced by the integration as a model to improve the analytic theories of the satellites. Comparison of the theories with the numerical integration may give clues which point to deficiencies and omissions from the theories. This approach has been used (Harper et al.) to identify Solar terms in the theory of Iapetus.

5.2 THE NUMERICAL INTEGRATION PROGRAM 'TITAN'

The numerical integration method used to generate ephemerides of the outer satellites of Saturn is an 8th order central-difference Gauss-Jackson scheme. It employs an iterative starting procedure. A predictor cycle followed by a corrector cycle is used in the integration of the equations for the coordinates of the satellites. In the integration of the equations for the partial derivatives of the coordinates with respect to the parameters, a predictor cycle alone is used.

The program which carries out the integration was written by Dr.A.T. Sinclair at the Royal Greenwich Observatory (see acknowledgements) for use on the Observatory's VAX 11-750 computer. It has been modified to enable it to run on the IBM 3083 of the University of Liverpool Computer Laboratory.

5.2.1 THE COORDINATE SYSTEM OF THE NUMERICAL INTEGRATION

We integrate the rectangular coordinates of the satellites in a Saturnicentric coordinate system whose xy-plane is the equator plane of Saturn. This is assumed to be a fixed plane over the time-span of the integration. The x-axis of the system is in the direction of the ascending node of the equator of Saturn upon the Earth's mean equator and equinox of 1950. The transformation between the integration frame and the mean equator and equinox of 1950 may thus be represented by a constant matrix as described in a previous chapter.

The choice of the equator plane of Saturn as the xy-plane of the integration arises from two considerations.

- The calculation of the perturbations due to the oblateness of Saturn are simplified because of symmetry about the xy-plane.

- There are no other similarly preferred planes in the satellite system and so any reference system is equally convenient from the computational point of view.

5.2.2 THE FORCE MODEL

The accelerations acting upon each satellite in the numerical integration are as follows :

(1) The gravitational attraction of Saturn regarded as a point mass.

This obeys a simple inverse square law which may be written as

[1] $$\underline{a}_i^{0} = - GM \frac{(1 + m_i)}{r_i^3} \underline{r}_i$$

where $\underline{r}_i$ = Saturnicentric position vector of the i^{th} satellite

G = gravitational constant

M = mass of Saturn

m_i = mass ratio of the i^{th} satellite to Saturn

$r_i = |\underline{r}_i|$.

(2) The gravitational attraction of the other satellites.

The acceleration upon the i^{th} satellite due to the attraction of the j^{th} satellite may be written

$$[2] \qquad \underline{a}_{ij} = G\,M\,m_j \left\{ \frac{\underline{r}_j - \underline{r}_i}{r_{ij}^{\,3}} - \frac{\underline{r}_j}{r_j^{\,3}} \right\}$$

where m_j = mass ratio of the j^{th} satellite to Saturn

$\underline{r}_j$ = Saturnicentric position vector of the j^{th} satellite

r_{ij} = the distance between the i^{th} and j^{th} satellites.

The first term is the gravitational attraction itself ; the second term is called the 'indirect term' and arises from the fact that the origin of the coordinate system is the centre of mass of Saturn and not the barycentre of the entire system. The second term is the acceleration of the centre of mass of Saturn relative to the barycentre due to the attraction of the j^{th} satellite upon Saturn. This expression is derived in appendix C.

Titan is the most massive satellite in the system and it perturbs Hyperion and Iapetus quite strongly. Indeed, the motion of Hyperion is characterised by the perturbation by Titan. Of the other satellites, Iapetus is the more massive and its perturbations on the motion of Titan and Hyperion are included in the force model. Perturbations by Rhea are also included. The most significant effect of Rhea is upon the secular rates of the nodes and apsides of the outer satellites, especially Titan. In this respect, it augments the perturbations due to the oblateness of Saturn. However, to allow for periodic perturbations by Rhea in the motion of Titan, the orbit of Rhea is assumed to be circular and fixed in the equator plane of Saturn. In the reference frame of the integration, the position of Rhea at time t is given by

$$X_R = 0.0035232 \cos L$$
$$Y_R = 0.0035232 \sin L$$
$$Z_R = 0$$

where $L = 231^\circ.761 + 79^\circ.69004007\ (t - 2411093.0)$

$$\mu_R = 4.4\ 10^{-6}.$$

The elements of Rhea are from Sinclair (1977) and μ_R, the mass ratio Rhea : Saturn is the value used by Sinclair and Taylor (1985) and taken from Tyler et al (1981).

Perturbations by Hyperion upon Titan were not included in the force model as Hyperion is not massive enough to affect the motion of either of the other satellites.

(3) The gravitational attraction of the Sun.

We may regard the sun as a very distant and very massive 'satellite' of Saturn. The acceleration of the i^{th} satellite due to the attraction of the Sun is given by

$$[3] \qquad \underline{a}_{is} = G\,M\,m_s \left\{ \frac{\underline{r}_s - \underline{r}_i}{r_{is}^{\,3}} - \frac{\underline{r}_s}{r_s^{\,3}} \right\}$$

where μ_s = mass ratio Sun : Saturn

$\underline{r}_s$ = Saturnicentric position vector of the Sun

r_{is} = distance between the Sun and the i^{th} satellite.

As in the previous case, the second term is an indirect term representing the acceleration of the centre of mass of Saturn with respect to the barycentre of the Sun-Saturn-satellites system due to the gravitational attraction of the Sun. The derivation is given in appendix C.

The heliocentric coordinates of Saturn used in the calculation of the solar perturbations are derived from the ephemeris of Saturn in Astron. Pap. Amer. Eph. volume 12. The coordinates are transformed from the ephemeris reference frame to that of the integration and then turned into sets of Chebyshev coefficients covering successive 400-day intervals. The use of Chebyshev series to calculate the coordinates of the Sun relative to Saturn during the integration is intended to speed up the process of interpolation whilst maintaining accuracy in the interpolated coordinates.

(4) The oblateness of Saturn.

The disturbing function of Saturn due to its oblateness may be written as (cf. Herrick (1972), chapter 18)

[4] $$R_e = - GM/r \sum_{n=2}^{\infty} (a_e/r)^n J_n P_n(w)$$

where a_e = equatorial radius of Saturn

r = distance from the centre of Saturn

J_n = n^{th} zonal harmonic coefficient of the potential field

w = z/r = latitude of the point above the equatorial plane of Saturn

and $P_n(w)$ is the Legendre polynomial of degree n.

The acceleration upon a satellite at this point due to the oblateness of Saturn is

[5] $$\underline{a}_e = \nabla R_e$$

Since the coordinate system has been chosen so that Saturn is symmetric about the xy-plane, all odd-numbered harmonic coefficients (J_3, J_5 etc.) disappear. Moreover, the ratio $(a_e/r)^n$ and the harmonic coefficients J_n rapidly become smaller for large values of n and so in practise we may neglect harmonics beyond J_4. Thus R_e contains only contributions from the n=2 and n=4 terms in the expression given above. We may write

[6] $$R_e = -(GM/r) \times \{(a_e/r)^2 J_2 P_2(w) + (a_e/r)^4 J_4 P_4(w)\}.$$

The components of the acceleration upon the satellite are

$$\ddot{x} = \partial R_e/\partial x$$
[7] $$\ddot{y} = \partial R_e/\partial y$$
$$\ddot{z} = \partial R_e/\partial z$$

and we obtain the following expressions (cf. Sinclair and Taylor (1985)) for the acceleration. The identities used to simplify the Legendre polynomials are derived in appendix D.

[8] $$\underline{a}_e = A\,\underline{r} + B\hat{\underline{k}}$$

where $A = (GM/r^3)\ \{J_2\ (a_e/r)^2\ P'_3(w) + J_4\ (a_e/r)^4\ P'_5(w)\}$

$B = -(GM/r^2)\ \{J_2\ (a_e/r)^2\ P'_2(w) + J_4\ (a_e/r)^4\ P'_4(w)\}$

$\hat{\underline{k}}$ = the unit vector in the z direction

and $P'_n(w)$ denotes the first derivative of the Legendre polynomial with respect to its argument.

5.2.3 PARAMETERS OF THE NUMERICAL INTEGRATION MODEL

A number of parameters of the numerical integration are considered to be free parameters : their values may be altered from one integration run to another, usually as a result of a least-squares correction process involving comparison of the integration with data or with analytic theories of the motions of the satellites. These parameters are of two types.

1. Elements of satellite orbits. There are six arbitrary parameters in the motion of any satellite ; these may take the form of classical orbital elements, but in a numerical integration it is more convenient to use the position and velocity vectors of the satellites at some fixed epoch. For each satellite there are three position components and three velocity components. These are equivalent to six orbital elements ; indeed, it is a simple matter to convert classical elements to position and velocity vectors and vice versa (Herrick (1971) gives a particularly lucid account of the equivalence of the two different types of parameter set and the transformation between them).

   Since we are dealing with three satellites, their orbital elements provide us with eighteen free parameters.

2. Mass and form-factor parameters. In their comparison of astrometric observations of the outer satellites with a numerical integration, Sinclair and Taylor (1985) also treated the masses of Titan, Iapetus and Saturn as parameters to be determined by least-squares iterative fitting. In addition the dynamic form-factors of Saturn, J_2 and J_4 were included as free parameters.

   The set of free parameters now numbers 23.

- The position vector and velocity vector of Titan.

- The position vector and velocity vector of Hyperion.
- The position vector and velocity vector of Iapetus.
- The mass ratio Titan/Saturn.
- The mass ratio Iapetus/Saturn.
- The mass ratio Saturn/Sun.
- The dynamic form factors J_2, J_4.

These are the free parameters in the current work.

5.2.4 PARTIAL DERIVATIVES OF THE COORDINATES

In addition to the coordinates of the satellites, we require the partial derivatives of the coordinates with respect to each of the free parameters of the integration model. These partial derivatives are used in the determination of the parameters by fitting the integration to observations or to analytic theories.

The partial derivatives are calculated by numerical integration in the same way as the coordinates. From the force model we may deduce expressions for the second time derivatives of the partial derivatives, i.e.

$$\frac{d^2}{dt^2}\left(\frac{\partial X_i}{\partial q_k}\right)$$

where X_i is any of the 9 satellite coordinates

q_k is any of the 23 free parameters.

The integration scheme thus incorporates 9 coordinates and 9×23 = 207 partial derivatives. The partial derivatives are not required to the same accuracy as the coordinates and so to save computing time, they are integrated using a predictor step alone.

If we write

[9] $$F_i = d^2X_i/dt^2$$

that is, F_i is the acceleration of the coordinate X_i given by the force model, then we have

[10] $$\frac{d^2}{dt^2}\left(\frac{\partial X_i}{\partial q_k}\right) = \frac{\partial F_i}{\partial q_k} + \sum_{j=1}^{18} \frac{\partial F_i}{\partial X_j}\,\frac{\partial X_j}{\partial q_k}.$$

The first term on the right-hand side is the explicit derivative of F_i with respect to the parameter q_k. In the case where q_k is a component of an initial position or velocity vector, this will be zero since the acceleration is not an explicit function of the initial position and velocity components. However, there are explicit derivatives with respect to the free parameters which correspond to masses and form-factors since these parameters do appear directly in the force model.

The second term on the right-hand side represents the implicit dependence of F_i upon the parameter q_k via the coordinates. It contributes to all the partial derivatives.

The starting values of the partial derivatives in the integration are

[11] $$\frac{\partial X_i}{\partial q_i} = 1$$

where q_i is the initial value of the coordinate X_i and

[12] $$\frac{d}{dt}\left(\frac{\partial X_j}{\partial q_j}\right) = 1$$

where q_j is the initial value of the velocity component dX_j/dt.

All other starting values of derivatives are zero.

5.3 ITERATIVE FITTING METHOD

The parameters of the numerical integration model were determined by comparing visual observations of the three satellites with the predicted positions from the integration. The differences between the observed and computed positions were combined with the partial derivatives described in previous sections to produce an equation of condition for each observation which expresses the (small) corrections to be made to the parameters in terms of the observed-minus-computed residual. Repeated application of this method yields parameter sets which give successively better fits to the observations. This is the basis of iterative least-squares differential correction, the technique which is used in this work to fit the numerical integration to the observations. The stages in the process may be enumerated.

(1) We must obtain a starting set of parameters that will give a reasonably close fit to the observations. The differential correction process relies upon the first-order approximation

$$\Delta p = \frac{\partial p}{\partial e_1} \Delta e_1 + \frac{\partial p}{\partial e_2} \Delta e_2 + \ldots + \frac{\partial p}{\partial e_N} \Delta e_N$$

where p is the observed quantity and Δp is its observed-minus-computed residual, and Δe_1, Δe_2, ..., Δe_N are the small corrections to the parameters e_j.

The starting values must thus be obtained from a model which already matches the observational data quite closely. The analytic theories of the satellites are chosen for this purpose, specifically the theories which Sinclair and Taylor have fitted to astrometric observations covering the period 1967 to 1982. As an initial approximation, the Saturnicentric rectangular coordinates of each satellite are calculated from the analytic theories at intervals corresponding to 10° arcs for several dates around the chosen epoch of the integration. The velocity components are calculated using a 7-point Lagrange interpolation formula differentiated once with respect to the interpolation argument. The initial values of the mass parameters and the dynamic form factors of Saturn are taken from the work of Sinclair and Taylor (1985).

Numerical integration is then used to calculate the positions of the satellites at two thousand random dates in a twenty-year interval centred on the adopted zero epoch of the integration. The Saturnicentric rectangular coordinates of the satellites are compared with the coordinates given by the analytic theories for the same dates. Equations of condition are constructed for each of the nine coordinates ξ (three from each satellite) incorporating the difference

$$\Delta\xi = \xi_{\text{Analytic theory}} - \xi_{\text{Numerical integration}}$$

expressed in terms of the corrections to the parameters of the integration. Solution of the equations of condition yields an improved set of parameters for the integration. This process of fitting the integration to the analytic theories is repeated until convergence is obtained i.e.

the corrections to the parameters become very small. At this point, the positions of the satellites given by the numerical integration closely match those given by the analytic theories and are thus close to the real satellite system.

A set of random dates is used in this process in order to avoid a 'sampling effect' whereby significant short-period terms in the analytic theories are masked because their period matches the sampling period. A different set of random dates is generated for each iteration.

The mass parameters and form factors are not allowed to vary at this stage because solution for these parameters in addition to position and velocity components by comparison with the theories would give only the values implicit in the theories and these were not judged to be better than the initial values adopted for the integration.

We may now begin the task of fitting the integration to the observations.

(2) The Saturnicentric rectangular coordinates of the three satellites are calculated by numerical integration at intervals of 0.25 days for a period sufficient to cover all the observations to be used. In practise the integration is carried out for forty years either side of the zero epoch in two runs : one from 1910 to 1950 and the other from 1910 to 1870. The coordinates of the satellites at the moment of each observation are obtained by interpolation while the integration is in progress. A

fourth-order interpolation formula is employed, both for the coordinates and for the partial derivatives of the coordinates with respect to the parameters.

(3) The coordinates of the satellites are transformed from the reference frame of the integration, based upon the equator plane of Saturn, to the true equator and equinox at the date of the observation. The partial derivatives are also transformed. This transformation consists of two rotations : the first is from the reference frame of the integration to the mean equator and equinox of B1950 and is a fixed transformation ; the second is from the mean equator and equinox of B1950 to the true equator and equinox of date. This is different for each observation.

We now have the Saturnicentric coordinates of each satellite referred to the true equator and equinox of date. We add the topocentric position vector of the centre of Saturn to obtain the topocentric positions of the satellites. From these, we may deduce the Right Ascension and Declination of each satellite and of the centre of Saturn's disk, and hence the position angle and separation of any of the four objects with respect to any other. We calculate only the datum required by the observation, together with the partial derivatives of the datum with respect to the parameters of the integration. Thus we may construct an equation of condition for each observation by combining the partial derivatives with the residual, i.e. the difference <Observed datum> minus <Computed datum>.

(4) Each observation yields an equation of condition. These are collected and combined to give a set of normal equations according to the theory of least-squares. An equation of condition is added to the normal equations only if its O-C residual is less than a specified rejection limit. This rejection limit was initially set at a nominal figure of 2.0 arc-seconds and this was later adopted as the standard limit as the root-mean-square residuals converged to a value of approximately 0.6 arc-seconds. Thus the rejection limit was 3 times the RMS residual, corresponding to rejection of 0.3% of errors which are normally distributed.

The normal equations may now be solved to give the corrections to the parameters. As part of the solution process, the standard errors of the parameters and their cross-correlations may be determined. It is also possible to fix one or more of the parameters if it is evident that there are strong correlations. This was indeed found to occur : for example, the mass of Saturn was highly correlated with the semi-major axes of the satellite orbits and so it was decided to keep a fixed value for Saturn's mass.

It was also found that the correction to Saturn's J_4 form factor was several orders of magnitude larger than any reasonable value of the parameter itself, and so this was also kept constant.

(5) The O-C residuals are analysed according to satellite pair and datum type (position angle or separation). The root-mean-square, mean and standard deviation are calculated and a histogram drawn to show the dis-

tribution of residuals. This is a useful indication of the progress of the iterations : clearly, we want the residuals to show a Gaussian distribution about zero, with a very narrow peak. If the iteration scheme is not working then this will be shown by the histogram. Example histograms are given in a later section where the results are discussed.

When the corrections to the parameters are small compared to the standard errors then the iterative process is complete. Further iterations will not significantly improve the fit of the integration to the observations. However, when the corrections are larger than the standard errors, the process begins again at stage (2)

5.4 RESULTS. (1) PROBLEMS WITH NON-CONVERGENCE IN THE FIRST PHASE

In practise, the least-squares fitting process was not as straightforward as the previous section might suggest. There was some difficulty in obtaining convergence in successive iterations whilst fitting the numerical integration to the analytic theories during stage (1) of the iterative procedure described previously. We may denote the difference between the position of each satellite given by the numerical integration and that given by the analytic theory as

$$\xi = | \underline{r}_{\text{Theory}} - \underline{r}_{\text{Integration}} | .$$

At each iteration, a graph of the 2000 ξ values plotted against time for each satellite showed a clear systematic trend.

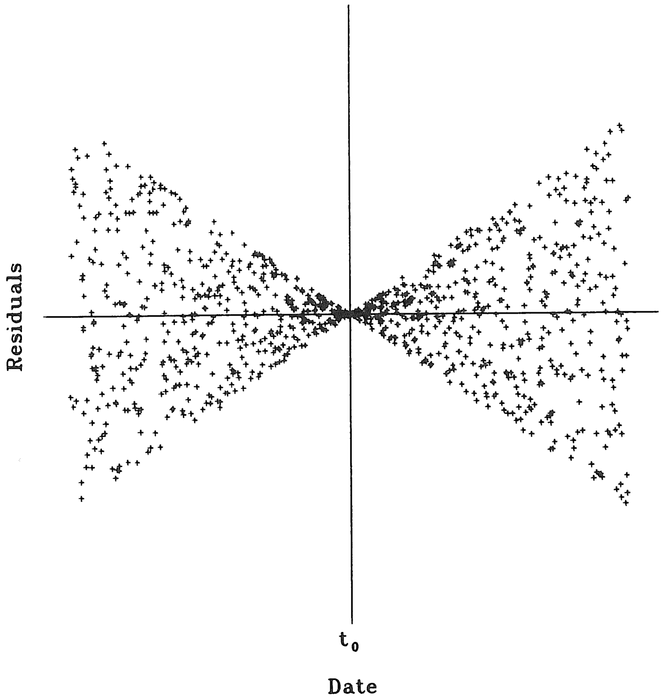


Figure 11. Integration-minus-theory residuals (schematic)

The maximum value of ξ increases linearly as a function of $|t - t_o|$, that is, in proportion to the distance from the initial epoch of the integration. This is the point at which we seek to determine starting values for the position and velocity of each satellite. The effect was most marked for Hyperion. When Hyperion was removed from the iterative process, convergence was achieved quite easily for Titan and Iapetus alone and satisfactory starting conditions were obtained for these satellites.

Since the analytic theory of Hyperion is such a complex problem, the short-period perturbations by Titan have been omitted from the computer program developed by Sinclair and Taylor to represent its motion ; these terms are generated separately by numerical integration and may be used as a look-up table in conjunction with the purely analytic part of the theory. However, these terms were not available for the time interval covered by the present study (1870 to 1947) and so the analytic theory of Hyperion used here is not complete. It is probable that this omission is the cause of the non-convergence ; we are trying to fit a dynamically consistent model, viz. the numerical integration, to one which is not consistent. The fitting process attempts to compensate by over-correcting the elements of Hyperion in the integration in successive iterations, and the result is divergence of the integration model from the analytic theory.

Titan is adversely affected by this because its parameters are linked to those of Hyperion by virtue of its large perturbations upon Hyperion. In mathematical terms, we may observe that equations of condition arising from the coordinates of Hyperion contain partial derivatives with respect to the parameters of Titan. Hence Titan's parameters may be subject to large and spurious corrections when Hyperion's position is allowed to diverge.

The solution adopted for this problem was to omit Hyperion from the iterative process of fitting the numerical integration to the analytic theories. Starting parameters were obtained for Titan and Iapetus by iteration, while those for Hyperion were calculated from the analytic theory without iteration. If these parameters were subsequently found to be inadequate then Hyperion could be omitted from comparison of the integration with observations.

Another possible cause for the divergence may be considered : the analytic theories employed as a reference model have been fitted to modern (post-1967) observations. The orbital elements, and in particular the longitude of the satellite at the epoch λ_o, are thus the best values for the period 1967 to 1983. In fitting the integration to these theories at the epoch 1910, we may be extrapolating the theories beyond their range of validity. This will be evident if the mean motions adopted in the theories are slightly in error. The mean motions used are values determined by Struve and Woltjer on the basis of pre-1935 observations covering a rather short time, 30 years at most. In the present work, the theories have been extended backwards some 70 years from the epoch of the data to which they have been fitted and so any error in the mean motion will be magnified as an error in the longitude of the satellite. The solution to this problem is to fit the theories to all the available data to obtain a better value of the mean motion, but this is a considerable exercise in its own right and it was not thought to be appropriate to undertake it in this context.

5.5 LEAST-SQUARES CORRECTION THEORY

The least-squares correction process is based upon the minimisation of the sum of the squares of the observed-minus-computed (O-C) residuals which are formed when the dynamical theory is compared with observations. Each O-C is the right hand side of an equation of condition

[13] $$\frac{\partial p^i}{\partial e_1} \varepsilon_1 + \quad \dots \quad + \frac{\partial p^i}{\partial e_N} \varepsilon_N = p_o^{\ i} - p_c^{\ i} = \Delta p^i$$

where $\varepsilon_j = \Delta e_j = (e_j)_{true} - (e_j)_{conjectured}$.

We wish, therefore, to minimise the function

[14] $$\rho = \quad (\Delta p^i)^2$$

with respect to each of the corrections $\varepsilon_1, \dots, \varepsilon_N$. This means that the derivatives of ρ with respect to each ε_j must be zero. That is,

[15] $$\partial p / \partial \varepsilon_j = 0 \qquad (j = 1, \dots, N).$$

This provides us with the set of N equations called the normal equations :

[16] $$a_{k1}\varepsilon_1 + \quad \dots \quad + a_{kN}\varepsilon_N = b_k$$

where

[17] $$a_{jk} = \sum_{i=1}^{N} \frac{\partial p^i}{\partial e_j} \frac{\partial p^i}{\partial e_k}$$

[18] $$b_j = \sum_{i=1}^{N} \frac{\partial p^i}{\partial e_j} \Delta p^i .$$

From these equations we note that the normal matrix (i.e. the matrix of the a_{jk}'s) is symmetric and positive definite. This implies that $a_{jk} = a_{kj}$ and $a_{jj} > 0$. The corrections ε_k may be determined by inverting the normal equations, which can be written in matrix notation as

[19] $$a\, \underline{\varepsilon} = \underline{b}.$$

Clearly,

[20] $$\underline{\varepsilon} = a^{-1}\underline{b} = s\, \underline{b}$$

where s is the inverse of the normal matrix.

5.5.1 STANDARD ERRORS

In a real physical system described by a parameterised model such as a numerical integration, there exists a set of parameters $\{ \underline{e}^{\,0}\}$ which best describes the system. This set is, as it were, the 'true' set of parameters and it is the set which would be determined if we possessed observations of the system which contained no errors.

In practise, of course, observations are prone to errors of several kinds. The most common (and most amenable to analysis) are random in nature and are due to any number of unforeseen and unpredictable effects on the part of the observer, his measuring apparatus and the conditions in which the observations are made. Such random errors are characterised by the fact that they form a Gaussian (normal) distribution when analysed in large numbers. Thus each observation consists of two parts : the 'true' value of the observed quantity in the sense used above, which we denote as y_i^0, and an error η_i taken from a normal distribution. The value observed in practise is y_i and we may write

[21] $$y_i = y_i^0 + \eta_i.$$

The crux of the problem facing us is this : we do not know the values of the η_i's.

The second type of error is not random. It is systematic in the sense that it affects the all observations in a similar way. It may be additive so that we have

[22] $$y_i = y_i^0 + \eta_i + \xi$$

where ξ is a constant, or it may be multiplicative so that

[23] $$y_i = \beta(y_i^0 + \eta_i)$$

where β is a constant, or it may be a combination of both types. In general, a systematic error has its cause in some feature of the experimental method which can be explained in physical terms without reference to random behaviour. As an example, consider a micrometer used to measure angular separation between satellites. If, whilst measuring separation between a satellite and a bright planet, the observer repeatedly overshoots the far limb of the planetary disk with the movable wire and does not realise the fact, then there will be a trend for the measures to be too large by a small amount whose value has a non-zero mean. Of course, careful observational technique is designed to minimise such effects but there may remain some systematic errors : a micrometer whose screw-thread has been incorrectly calibrated along part of its length may give rise to systematic multiplicative errors i.e. errors of scale. Moreover, such errors may be dependent upon the conditions under which the micrometer is used : ambient temperature changes causing different parts of the device to expand at different rates, to cite one example.

Systematic errors may manifest themselves in the O-C residuals. If this is the case, we can (and must) take steps to eliminate them by tracking down their physical cause.

It may, however, be more likely that the parameters of the model will adjust to incorporate these errors during the least-squares fitting process. Continuing the example of scale errors in separation measures : if the angular separations within a satellite system are systematically measured too large or too small by a given factor then a least-squares procedure based upon osculating orbital elements may respond by converging to a solution where the semi-major axes of the satellites are all too large or too small by a similar factor. If the mass of the primary is also a free parameter, it may also be in error by a corresponding factor.

We must, in the first instance, assume that any errors in the observations are random. After we determine the corrections to the parameters of the model using the least-squares method, we may then calculate the likely errors in the parameters due to errors in the observations upon which the solution is based. We quote the parameters in the form

$$e \pm \delta e$$

where e is the calculated value, and we are confident that the 'true' value lies in the interval $e-\delta e$ to $e+\delta e$ with a probability p. That is,

$$e - \delta e \leq e^0 \leq e + \delta e \qquad [24]$$

with probability p.

Following Brouwer and Clemence (1961, p.226) and Jeffreys (1939, p.61) we quote the standard error of each parameter, denoted by σ, which is the value of δe corresponding to $p = 0.683$. That is to say, it is 68.3% certain that the 'true' value of a parameter will lie within one standard error of its value determined from data whose errors are distributed in a random fashion following a normal distribution.

We may calculate the standard errors of the parameters as follows : after solving the normal equations to determine the corrections to the parameters, these corrections are substituted into the original equations of condition and the residuals calculated. For example, if a particular equation of condition is

$$\alpha_{i1}\varepsilon_1 + \quad \dots \quad + \alpha_{iN}\varepsilon_N = \Delta p^i$$

(where we have written α_{ij} for $\partial p^i/\partial e_j$)

then the residual we require from the equation is

[25] $$\nu_i = \alpha_{i1}\varepsilon_1 + \dots + \alpha_{iN}\varepsilon_N - \Delta p^i$$

where $\varepsilon_1, \dots, \varepsilon_N$ are the values of the unknowns (corrections to the parameters) obtained by solving the normal equations.

We then form the standard errors in the residuals E from

[26] $$E = \{\sum_i \nu_i^2/(m-n)\}$$

where m is the number of equations of condition and n is the number of free parameters.

The standard error of the correction ε_j to the parameter e_j, and hence to the parameter itself, is then

$$[27] \qquad \sigma_j = E \sqrt{S_{jj}}$$

where S_{jj} is the j^{th} diagonal component of the inverse of the normal matrix.

If we require an error estimate with a different level of probability, we multiply the standard error by a factor k such that $p = \mathrm{erf}(k/\sqrt{2})$. The probable error corresponds to a probability $p = \frac{1}{2}$ and we find that $k = 0.6745$ (cf. Brouwer and Clemence (1961) p.226). Thus the 'true' value of the parameter e_j is as likely to lie within the range $e_j - 0.6745\sigma_j$ to $e_j + 0.6745\sigma_j$ as it is to lie outside it. We present below a table of probabilities as a function of multiples of the standard error.

k	p x 100%
0.5	38.3%
0.6745	50.0%
1.0	68.3%
2.0	95.4%
3.0	99.73%
4.0	99.9937%
5.0	99.999943%

5.5.2 CORRELATIONS BETWEEN PARAMETERS

It is implicitly assumed in the least-squares correction process that the free parameters are independent of one another. That is to say, small corrections can be applied to any of the parameters independently and there are no relationships of the form

[28] $$\varepsilon_i + \alpha\varepsilon_j = \beta$$

where ε_i, ε_j are the corrections to parameters e_i, e_j and α, β are constants.

In practise, we often find such relationships and these may be explained in terms of the physics of the problem. Consider, for example, the case where the major semi axis of a satellite orbit and the mass of the primary are both free parameters in, say, a numerical integration model. These are related, together with the mean motion of the satellite, by Kepler's third law

[29] $$n^2a^3 = k^2M$$

where k^2 is the gravitational constant and we neglect the mass of the satellite itself. If we rewrite this equation in terms of small changes then we have

[30] $$2\frac{\Delta n}{n} + 3\frac{\Delta a}{a} = \frac{\Delta M}{M}.$$

Now the mean motion of the satellite will be well-determined by a series of observations spanning several decades, an interval equivalent to hundreds of orbital periods of the satellite. Thus the value of Δn is far better determined than either Δa (which depends, among other things, upon the calibration of the micrometer screw used to make separation measures) or ΔM. This is precisely the situation which gives rise to correlations : a linear combination of two unknowns is better-determined than either of the individual unknowns. Δa and ΔM are not independent since the solution will adjust them to minimise the change to Δn whose value is, as it were, a 'known constant' of the system. Thus we may expect to find correlations between the mass of the primary and the major semi axes of each of the satellites.

We may detect and quantify the correlations in a least-squares correction process by inspection of the normal matrix and its inverse. We define the correlation coefficient between ε_i and ε_j to be

[31] $$c_{ij} = S_{ij} / \sqrt{(S_{ii} \times S_{jj})}$$

where S denotes the inverse of the normal matrix. The correlation coefficient lies between -1 and +1 and it is normally close to zero i.e. the off-diagonal components of S are much smaller than those in the diagonal. However, when c_{ij} is nearly ± 1 then ε_i and ε_j are closely correlated and it may prove difficult to obtain a solution for both of them simultaneously. We may show that the quantity

[32] $$\kappa_{ij} = a_{ij} / \sqrt{(a_{ii} \times a_{jj})}$$

also approaches ±1 when a pair of parameters are correlated. Suppose we have a set of equations of condition

$$\alpha_{i1}\varepsilon_1 + \alpha_{i2}\varepsilon_2 + \ldots + \alpha_{iN}\varepsilon_N = \beta_i$$

for i =1 to M. Suppose also that ε_1 and ε_2 are linearly dependent so that for some constant γ, $\alpha_{i2} = \gamma\alpha_{i1}$(or very nearly) for all equations of condition. Consider the components a_{11}, a_{12} and a_{22} in the normal matrix. We see that

[33] $$a_{12} = \gamma a_{11}$$

[34] $$a_{22} = \gamma^2 a_{11}$$

hence

[35] $$\kappa_{12} = \gamma a_{11}/(a_{11} \times \gamma^2 a_{11}) = \operatorname{sgn} a_{11} = \pm 1.$$

It is interesting to note that γ may be found from

[36] $$\gamma = a_{12}/a_{11} = a_{22}/a_{12}.$$

Brouwer and Clemence (1961, p.231) present a method whereby the normal equations are re-written in order to solve them for the combination $\varepsilon_1 + \gamma\varepsilon_2$. Alternatively, we may choose to remove one of the parameters from the system, generally because we have a better determination of it from another source. This may be achieved quite easily by operating upon the normal equations. Suppose we wish to hold the value of e_k constant.

This means that we must force the normal equations to yield $\varepsilon_k = 0$. We set all the components of row k and column k of the normal matrix to zero <u>except</u> for the diagonal component, which we set to unity.

[37] $$a_{ik} = a_{ki} = 0 \qquad \text{for } i=1 \text{ to } N \text{ except } i=k$$

[38] $$a_{kk} = 1$$

and we set the k^{th} component of the right-hand-side vector $\underline{b}$ to zero.

[39] $$b_k = 0$$

This does not alter the normal matrix with respect to the other free parameters.

5.5.3 CONVERGENCE OF THE ITERATIVE PROCESS

We may regard the iterative correction process upon the parameters of our model as complete when certain criteria are met. Clearly, we wish successive sets of parameters to give an increasingly close fit of the model to the observational data. We may quantify this in terms of the root-mean-square (RMS) residuals of the data. The RMS residuals should decrease in size, settling to a constant value as convergence is attained. The number of data included in the RMS residuals must not decrease : as the closeness of the fit improves, we expect more observations to be included within the chosen rejection limit.

During each least-squares correction cycle, we may also consider the size of the corrections to be applied to the parameters. Corrections should decrease in size in successive iterations. For each of the free parameters e_j, we compare the correction ε_j with the standard error σ_j. When the correction is much smaller than the standard error, we may regard the correction as negligible. Thus when, say, $|\varepsilon_j|/\sigma_j << 0.01$ for all of the free parameters, the correction process is complete. Further iterations will not increase the accuracy with which the parameters can be determined.

5.6 RESULTS (2) : SOLUTION FOR POSITION AND VELOCITY

The first trial involved fitting the numerical integration to visual observations of Titan, Hyperion and Iapetus by solving for the position and velocity vectors of the three satellites and the masses and J parameters. The initial values of the masses and J parameters were those of Sinclair and Taylor (1985) while the initial values of the position and velocity were those obtained by fitting the integration to the analytical theories as described in section 5.4

The initial parameters gave quite a good fit to the observations.

Pair	RMS	Obsns	(%)
Saturn-Titan	1".069	299/321	93
Saturn-Hyperion	1".958	905/938	96
Saturn-Iapetus	0".853	315/321	96
Titan-Hyperion	1".576	688/766	90
Titan-Iapetus	0".696	820/832	99

All observations with an O-C residual of less than 5".0 are included. The fit for Hyperion is not as good as that for Titan and Iapetus because its initial parameters were not determined by iterative fitting to the theory. Nevertheless, 97% of all data are included in the table above.

When the normal equations were constructed from the residuals, a number of significant correlations were found. Among these, the mass of Saturn and its J_2 form factor were both strongly correlated with the initial position and velocity components of the satellites. In addition, the corrections to be applied to J_4 and the mass of Iapetus were larger than any physically reasonable value : the corrected mass of Iapetus would have been negative while the corrected value of J_4 would have exceeded J_2. Accordingly, it was decided to fix the masses and J's and a solution was made for the position and velocity components of the satellites only.

There were again a number of correlations. The x and y position components correlated strongly with the x and y velocity components for each satellite.

Titan x_o with y'_o

y_o with x'_o

Hyperion	x_o with y_o
	x_o with x'_o
	x_o with y'_o
	y_o with y'_o
	x'_o with y'_o
Iapetus	y_o with x'_o

The numerical integration was repeated using the corrected parameters. The O-C residuals resulting from the comparison of this integration with the observations were as follows.

Pair	RMS	Obsns	(%)
Saturn-Titan	2".413	36/321	11
Saturn-Hyperion	2".430	123/938	13
Saturn-Iapetus	0".827	315/321	98
Titan-Hyperion	2".468	351/706	50
Titan-Iapetus	2".397	594/832	71

The rejection limit, as before, is 5".0. The number of observations falling within this limit is much reduced and the RMS residuals are larger, with the exception of Saturn-Iapetus data. Clearly, the least-squares correction process has failed : the corrections applied to the parameters of Titan and Hyperion contain serious errors which are due to the large correlations.

The correlations between initial position and velocity components may be explained by postulating that some function of the position and velocity of each satellite is better-determined by the observations than any of the six individual components. If we consider the motion of a satellite in a nearly circular orbit in the xy-plane (this is approximately true for all the satellites in the numerical integration) then we may write

$$x = a\cos(nt + \varepsilon) + O(e)$$
$$y = a\sin(nt + \varepsilon) + O(e)$$

and

$$x' = -na\sin(nt + \varepsilon) + O(e)$$
$$y' = +na\cos(nt + \varepsilon) + O(e).$$

Neglecting the eccentricity of the orbit we may write

$$x' \simeq -n\,y$$
$$y' \simeq +n\,x.$$

This will be true at each point in the orbit and so the initial position and velocity components will be related thus

$$x'_o \simeq -n\,y_o$$
$$y'_o \simeq +n\,x_o.$$

It is significant that the strongest correlations in the least-squares fitting process are between x_o and y'_o and between y_o and x'_o. This leads us to suspect that relationships such as those above are indeed affecting the initial position and velocity components.

It would be more instructive to employ as the unknowns a set of parameters with more direct physical significance than position and velocity components. Therefore we turn our attention to the osculating elements of the orbits.

5.7 USE OF OSCULATING ELEMENTS AS AUXILIARY PARAMETERS

The osculating elements of the orbit of a satellite at any instant are the elements of the elliptic two-body orbit in which the unperturbed position and velocity of the satellite are the same as the actual position and velocity of the satellite at the epoch. There exist standard formulae for calculating osculating elements from a given set of position and velocity components, and vice versa. The most lucid exposition is that of Herrick (1971, chapter 4).

We wish to replace the position and velocity components of each satellite by the osculating elements as 18 of the parameters to be determined by comparison of the integration with the observations. To do this, we must change the iterative scheme, which now becomes :

1. Execute the numerical integration with the current set of position and velocity parameters and compare this with the observations to obtain the O-C residuals and equations of condition in terms of cor-

rections to the position and velocity parameters. This is exactly the same as before.

2. Convert the position and velocity components of each satellite at the zero epoch of the integration into osculating elements, and transform the equations of condition so that they express the O-C residuals in terms of corrections to the orbital elements. This transformation is carried out by use of partial derivatives of the position and velocity components with respect to the osculating elements. If we write the three position and velocity components of each satellite as the 18 components of a state vector ξ and the osculating elements as the 18 components of a vector $\underline{e}$ then

$$[40] \quad \Delta\xi_i = \sum_{j=1}^{18} \frac{\partial\xi_i}{\partial e_j} \Delta e_j .$$

This allows us to replace each $\Delta\xi_i$ in the equations of condition by a linear combination of the Δe_j's. Each of the equations is transformed from

$$[41] \quad \frac{\partial p}{\partial\xi_1} \Delta\xi_1 + \quad \dots \quad + \frac{\partial p}{\partial\xi_{18}} \Delta\xi_{18} + \quad \frac{\partial p}{\partial J_2} \Delta J_2 + \dots = \Delta p$$

to

$$[42] \quad \frac{\partial p}{\partial e_1} \Delta e_1 + \quad \dots \quad + \frac{\partial p}{\partial e_{18}} \Delta e_{18} + \quad \frac{\partial p}{\partial J_2} \Delta J_2 + \dots = \Delta p.$$

We note that the derivatives with respect to the masses and J factors are unchanged by the transformation since these parameters are the same whether we use position and velocity or osculating elements. We may write the general rule for the transformation of the coefficients of the equations as

$$[43] \quad \frac{\partial p}{\partial e_k} = \sum_{j=1}^{18} \frac{\partial p}{\partial \xi_j} \frac{\partial \xi_j}{\partial e_k}.$$

The transformation coefficients $\partial \xi_j / \partial e_k$ are mostly zero : we need only calculate the derivatives of the position and velocity components of each satellite with respect to its own osculating elements. The required derivatives may be readily obtained from the classical formulae of two-body motion. Herrick (1972, section 15B) gives an account of the calculation of these derivatives using a method based upon the radial, transverse and binormal vectors of the orbit of the satellite.

3. Construct a set of normal equations from the new equations of condition, applying a suitable rejection limit (in this case, 2".0) to the O-C residuals as before.

4. Solve the normal equations to determine the corrections to the osculating elements (and the masses and J parameters if required).

5. Apply the corrections to the elements and convert them to instantaneous position and velocity components for each satellite using standard two-body elliptic formulae. The position and velocity vec-

tors thus obtained are the current parameters for a new iterative cycle beginning again at (1)

The osculating elements used in this scheme are referred to the coordinate system of the integration. Thus the inclination of each satellite's orbit is with respect to the equator plane of Saturn. Likewise, the node, apse and mean longitude are measured in the equator plane of Saturn from the x-axis of the integration coordinate system to the ascending node of the orbit upon the equator plane, and thence in the orbit plane.

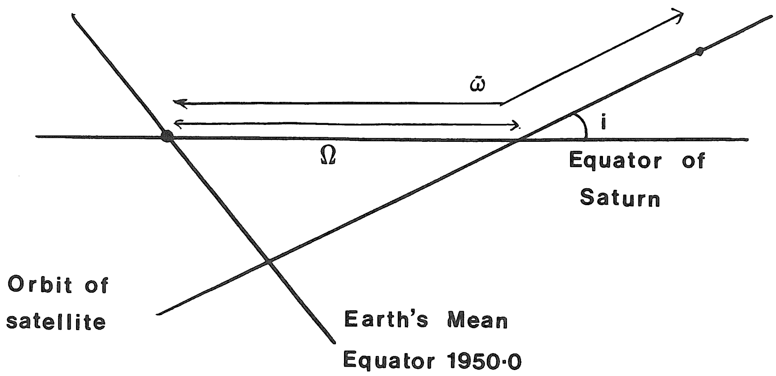


Figure 12. Angular osculating elements

5.8 RESULTS : (3) SOLUTION FOR OSCULATING ELEMENTS

A number of trials were carried out using various sets of free parameters. They are represented in diagram form below.

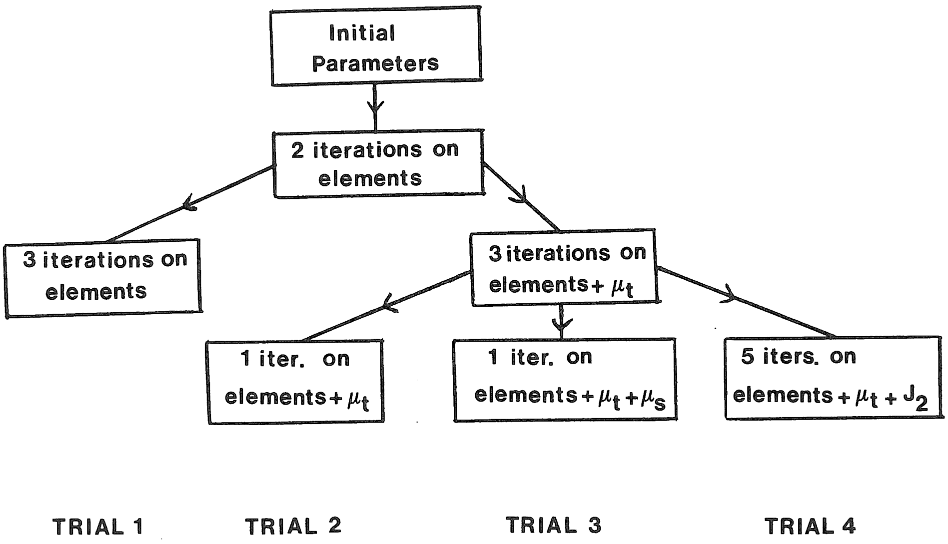


Figure 13. Solutions for osculating elements

In this context, 'element' signifies the set of 18 parameters comprising the six osculating elements for each of the three satellites. μ_T and μ_S are the mass ratios Titan/Saturn and Saturn/Sun respectively and J_2 is the dynamical form-factor of Saturn.

As an example, consider trial 3. Starting with the initial parameter set (obtained by fitting the integration to the analytic theories) we carry out the correction process upon the elements twice, then include μ_T for three correction cycles and finally include μ_s in one cycle. Thus trial 3 consists of six iterations incorporating an increasing number of free parameters. This approach allows us to monitor the behaviour of the least-squares process as new parameters are added : when convergence is attained with a given parameter set, we may increase the set to include further parameters of physical interest.

Trial 1 Five iterations on the osculating elements alone

The correction process converged quite rapidly and the RMS residuals after the 5th iteration are given below. The rejection limit for individual O-C residuals is 2".0. Results are given for both position angle and separation measures. Note that the position angle residuals are in fact $s\Delta p$ and thus they may be compared directly with the separation residuals.

Objects observed	Datum	Observations	RMS	Mean
Saturn - Titan	P	160 of 176	0".460	-0".105
Saturn - Titan	S	137 of 145	0".509	0".157
Saturn - Iapetus	P	154 of 162	0".747	0".064
Saturn - Iapetus	S	153 of 159	0".716	0".462
Saturn - Hyperion	P	458 of 473	0".743	-0".167
Saturn - Hyperion	S	433 of 465	0".871	0".514
Titan - Iapetus	P	411 of 417	0".385	0".009
Titan - Iapetus	S	407 of 415	0".357	0".186
Titan - Hyperion	P	344 of 355	0".452	-0".046
Titan - Hyperion	S	340 of 351	0".530	0".107
Iapetus - Hyperion	P	13 of 13	0".352	0".140
Iapetus - Hyperion	S	13 of 13	0".245	0".055

A total of 3023 out of 3144 (96.2%) data fall within the rejection limit.

The RMS residuals compare well with those obtained by Sinclair and Taylor (1985) who fitted a numerical integration to photographic (astrometric) observations covering the period 1967 to 1982. They obtained

Titan - Hyperion 0".33

Titan - Iapetus 0".22.

These figures are to be compared with the values obtained by combining the sΔp and Δs residuals above, which yield

Titan - Hyperion 0".49

Titan - Iapetus 0".37.

The residuals from visual observations are larger than those from photographic data by about 50%. This can probably be explained by the lower accuracy of the visual observations i.e. the random errors of observation are larger for the older data.

We must recall that the solution has been obtained by fitting to Saturn - satellite data as well as inter-satellite data, while Sinclair and Taylor used only inter-satellite measures. The RMS residuals for measures relative to Saturn are significantly larger than those for inter-satellite measures. This is due in part to the observational errors introduced in estimating the centre of the disk of Saturn when making separation and position angle measures relative to the planet. Inclusion of Saturn - satellite measures may therefore be expected to degrade a solution for the parameters of any dynamical model.

It may be more realistic to compare the residuals of Sinclair and Taylor with those obtained by fitting the integration to inter-satellite measures alone. Of the 3144 observations included in this study, 1580 are relative to Saturn while 1564 are inter-satellite measures : by excluding the Saturn - satellite measures we reduce the data set by 50% and advance the date of the earliest observation by 10 years, reducing the time span by 20%. The question of whether to employ Saturn - satellite data is thus a matter of compromise. A weighting scheme is probably the optimum solution but this introduces problems of its own and is not included in this work. It is deferred to the section on suggestions for further work.

Trial 2 Determination of the mass of Titan

This trial consisted of six iterations. During the first two, only the osculating elements were allowed to vary and so these iterations are the same as those of trial 1. Then the mass ratio Titan/Saturn was included in the set of free parameters and a further four iterations were carried out. Convergence was obtained (in the sense described in section 5.7) and the RMS residuals at a rejection limit of 2".0 were

Objects observed	Datum	Observations	RMS	Mean
Saturn - Titan	P	160 of 176	0".459	-0".103
Saturn - Titan	S	137 of 145	0".507	0".159
Saturn - Iapetus	P	154 of 162	0".747	0".064
Saturn - Iapetus	S	153 of 159	0".715	0".463
Saturn - Hyperion	P	457 of 473	0".731	-0".176
Saturn - Hyperion	S	435 of 465	0".878	0".523
Titan - Iapetus	P	411 of 417	0".385	0".009
Titan - Iapetus	S	407 of 415	0".357	0".185
Titan - Hyperion	P	345 of 355	0".464	-0".040
Titan - Hyperion	S	340 of 351	0".532	0".117
Iapetus - Hyperion	P	13 of 13	0".334	0".103
Iapetus - Hyperion	S	13 of 13	0".251	0".073

The mass of Titan was determined to be

$$\mu_T = (2.36659 \pm 0.00027) \times 10^{-4}.$$

The final RMS residuals are almost exactly the same as those in trial 1. Three more observations fall within the 2".0 limit than in trial 1 though one Saturn-Hyperion measure is lost. Thus, statistically speaking, this solution is based upon a data set of the same size as that used in trial 1.

<u>Trial 3</u> Determination of the mass of Saturn

This trial proceeded in the same way as trial 2 except that in the 6th correction cycle, the mass ratio Saturn/Sun (μ_s) was also allowed to vary. The free parameters were thus the osculating elements of Titan, Hyperion and Iapetus, the mass ratio Titan/Saturn and the mass ratio Saturn/Sun.

Upon carrying out the least-squares correction process on this parameter set, a number of significant correlations were found. The most notable were those between the mass of Saturn and the major semi axes of each of the satellites (as explained in section 5.5.2). To illustrate the effect of such correlations upon the solution, the corrections calculated by the least-squares method were applied to the parameters and a numerical integration was executed with the new parameters. This was compared with the observations and a set of RMS residuals calculated. The rejection limit in the table below is 2".0.

Objects observed	Datum	Observations	RMS	Mean
Saturn - Titan	P	5 of 176	1".026	-0.514
Saturn - Titan	S	4 of 145	1".352	-0.321
Saturn - Hyperion	P	6 of 473	1".366	-0.395
Saturn - Hyperion	S	6 of 465	1".354	0.369
Saturn - Iapetus	P	No observations within the limit		
Saturn - Iapetus	S	No observations within the limit		
Titan - Iapetus	P	7 of 417	1".256	0.044
Titan - Iapetus	S	4 of 415	1".186	-0.398
Titan - Hyperion	P	8 of 355	1".022	0.085
Titan - Hyperion	S	5 of 351	1".307	-0.165

Virtually no observations fall within the limit, and of course the RMS residuals do not possess any statistical significance. Clearly, the correlations implicit in the normal equations cause the corrections to some of the parameters to be in error. Accordingly, the value of the mass-ratio Saturn/Sun was held fixed in the subsequent trial.

Trial 4 Determination of Saturn's J_2

This trial was carried out as follows, starting with the initial parameter set.

1. Two iterations solving for osculating elements only.

2. Three iterations solving for elements plus μ_T.

3. Five iterations solving for elements plus μ_T plus J_2.

Convergence was obtained rather slowly following the introduction of J_2 as a free parameter, since several other parameters (notably major semi axes) underwent further changes to accommodate the new corrected value of J_2. This is a consequence of the form of the oblateness perturbations where J_2 is factored by a_e^2/r^3 (where a_e is the equatorial radius of Saturn and r is the radius vector of the satellite). Since r is proportional to the major semi axis, this factor is effectively a_e^2/a^3. Thus the major semi axis appears in the oblateness forces at a rather high power and a change in its value will also affect the value of J_2. Thus we might expect correlations between J_2 and the major semi axes of the satellite, and these do appear in the least-squares scheme. However, they do not cause erroneous corrections to be made to the parameters concerned.

The RMS residuals obtained after the final iteration are as follows. The rejection limit on individual O-C's is 2".0.

Objects observed	Datum	Observations	RMS	Mean
Saturn - Titan	P	160 of 176	0".462	-0".101
Saturn - Titan	S	137 of 145	0".513	0".160
Saturn - Iapetus	P	154 of 162	0".750	0".059
Saturn - Iapetus	S	153 of 159	0".712	0".461
Saturn - Hyperion	P	457 of 473	0".732	-0".169
Saturn - Hyperion	S	435 of 465	0".876	0".521
Titan - Iapetus	P	411 of 417	0".383	0".006
Titan - Iapetus	S	407 of 415	0".355	0".188
Titan - Hyperion	P	344 of 355	0".454	-0".043
Titan - Hyperion	S	341 of 351	0".539	0".121
Iapetus - Hyperion	P	13 of 13	0".330	0".073
Iapetus - Hyperion	S	13 of 13	0".235	0".064

Again, these RMS residuals are almost identical to those obtained in trials 1 and 2, with the same number of data falling within the 2".0 limit. Statistically, therefore, the parameters are based upon effectively the same data set as in trials 1 and 2. The values obtained for μ_T and J_2 are

$$\mu_T = (2.36651 \pm 0.00028)\ 10^{-4}$$

$$J_2 = 0.01779 \pm 0.00043$$

using a value of 60000 km (4.0107×10^{-4} AU) for the equatorial radius of Saturn.

5.9 DISCUSSION OF RESULTS

Values of the mass of Titan

We present below the values determined for the mass of Titan in this work and in two other recent papers. Sinclair and Taylor's (1985) value was obtained by fitting a numerical integration to photographic observations over the period 1967 to 1982. Tyler et al (1981) derived their value from analysis of radio tracking of the Voyager 1 spacecraft. We also include values calculated by Message based upon comparison of his theory of the motion of Hyperion with Woltjer's opposition mean data. Value (a) is a weighted average of values determined from individual terms of the theory whilst value (b) is a least-squares solution.

Source	$\mu_T \times 10^{-4}$
Trial 2	2.36659 ± 0.00027
Trial 4	2.36651 ± 0.00028
Sinclair and Taylor (1985)	2.36777 ± 0.00055
Tyler et al (1981)	2.3664 ± 0.0008
Message (a)	2.3648 ± 0.0055
Message (b)	2.3677 ± 0.0004

There is good agreement between our values and that of Tyler et al., though (as with J_2) they do differ by several standard errors from Sinclair and Taylor's determination. It is probably not sufficient to invoke an argument based on scale errors in visual observations to explain this discrepancy. The system is most sensitive to the mass of Titan via its perturbations upon Hyperion and thus a theoretical analysis of the dependence of the system upon Titan's mass will inevitably involve the theory of the motion of Hyperion. We do not propose to perform such an analysis in this work, recalling that one of the principal reasons for adopting numerical integration as a model was to avoid the complications of the theory of the motion of Hyperion !

Values of J_2

The value of J_2 includes the gravitational effect of the rings, the secular effect of the four inner satellites and any error in the mass adopted for Rhea. For comparison, the equivalent value obtained by Sinclair and Taylor is

$$J_2 = 0.01675 \pm 0.00089.$$

They write this combined or 'lumped' value as $\bar{J}_2$ and relate it to the true value by

[44] $$\bar{J}_2 = J_2 + 0.000061 + 52.6\ \Delta\mu_R$$

where the constant represents the secular contribution of the satellites Mimas, Enceladus, Tethys and Dione and of the rings, and $\Delta\mu_R$ is the correction required to the mass of Rhea. They assume that $\Delta\mu_R$ is zero and thus the true value of J_2 is obtained by subtracting 0.000061 from the 'lumped' value derived directly from observations. Thus for the true values we have

This work	0.01773 ± 0.00043
Sinclair and Taylor (1985)	0.01669 ± 0.00089
IAU (1976) recommended	0.01645.

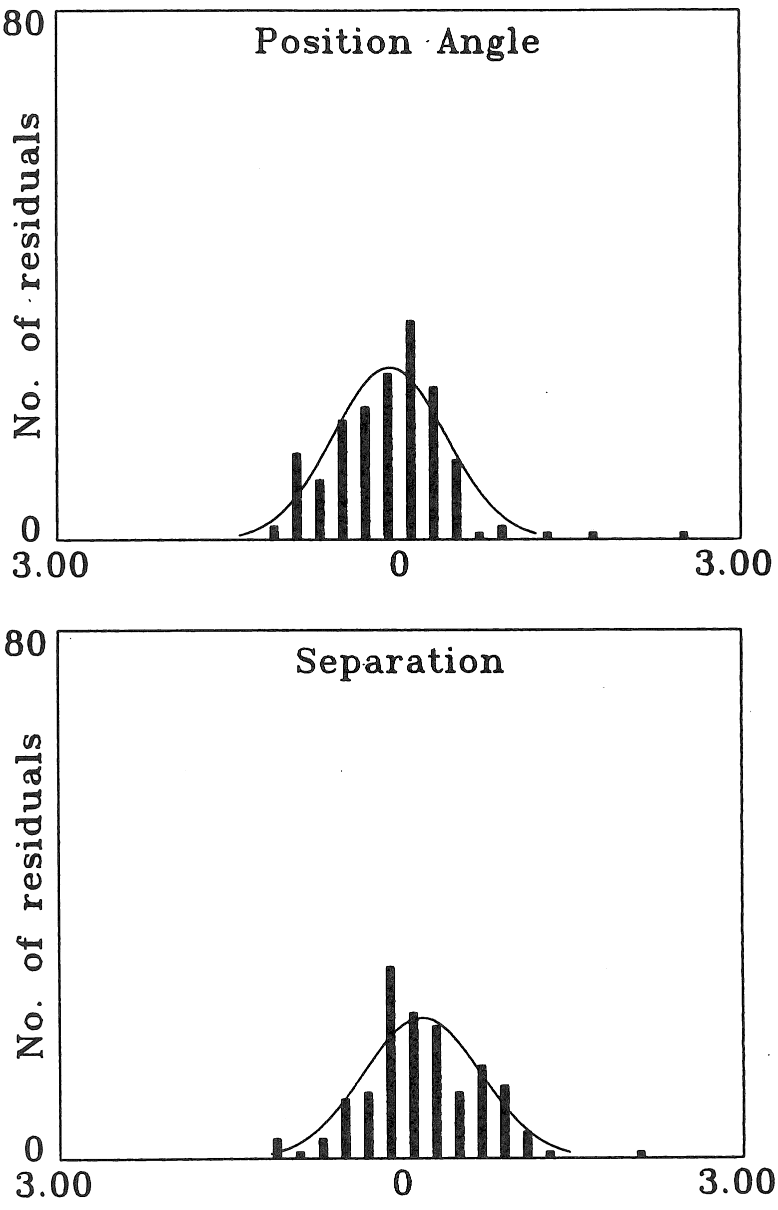
Our value is consistent with that of Sinclair and Taylor. They differ by 0.00104, rather more than one standard error of Sinclair and Taylor. As noted before, the oblateness perturbations are dependent upon the size of the satellite orbit in addition to J_2. Neglecting the eccentricity of the orbit, the principal factor in the second harmonic of the oblateness disturbing function includes J_2/a^3. A small change in the major semi axis will alter this factor quite considerably, hence changing the magnitude of the oblateness perturbations in the model and affecting the determination of J_2. This is the reason why the major semi axes are correlated with J_2. Visual observations often contain an error of scale which means that the values of the major semi axis of the orbit determined from such observations may include an error in the form of a multiplicative factor. For a particular observatory (i.e. a particular micrometer) this factor will be roughly the same for all the observed satellites and so the quantity J_2/a^3 will be in error by the same multiplicative factor for each of them. We may expect the observed value of J_2 to be a little too large or too small accordingly.

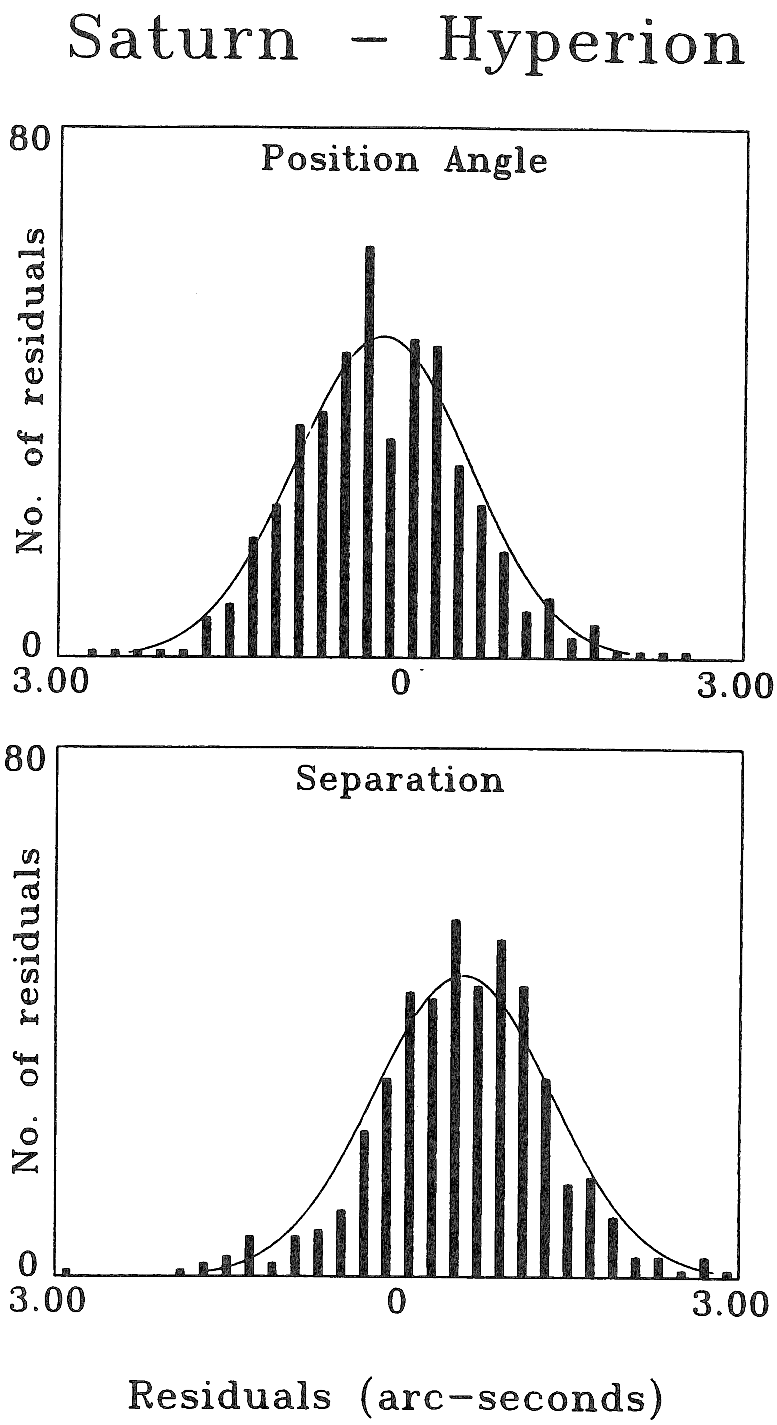
In the current work, the value of J_2 obtained from the integration is perhaps some 6% larger than most other determinations, which are closer to 0.165. This discrepancy is probably due to a scale error of the kind described.

The standard error of our determination is half that of Sinclair and Taylor. We base our value on 50 years' data while Sinclair and Taylor have only 16 years of photographic observations. We note also that the two integrations have overall RMS residuals of similar size. The value of J_2 is determined from its secular effect on the nodes and inclinations of the satellite orbits and hence we expect better values from data which are spread over a greater time interval. Thus it is plausible that our standard error should be smaller than that of Sinclair and Taylor, though the discrepancy in the actual values of J_2 is disturbing and suggests that our value should be used with caution.

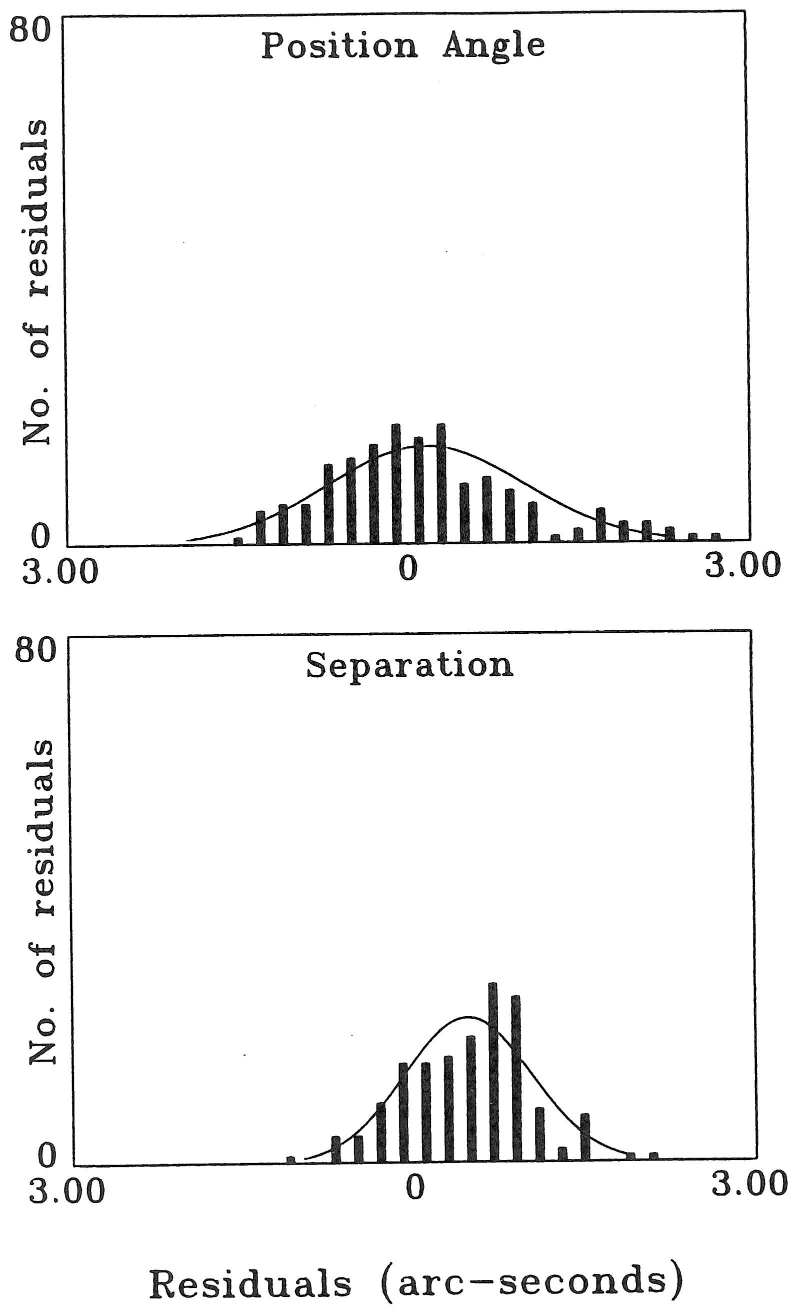
Distribution of the residuals

Saturn – Titan

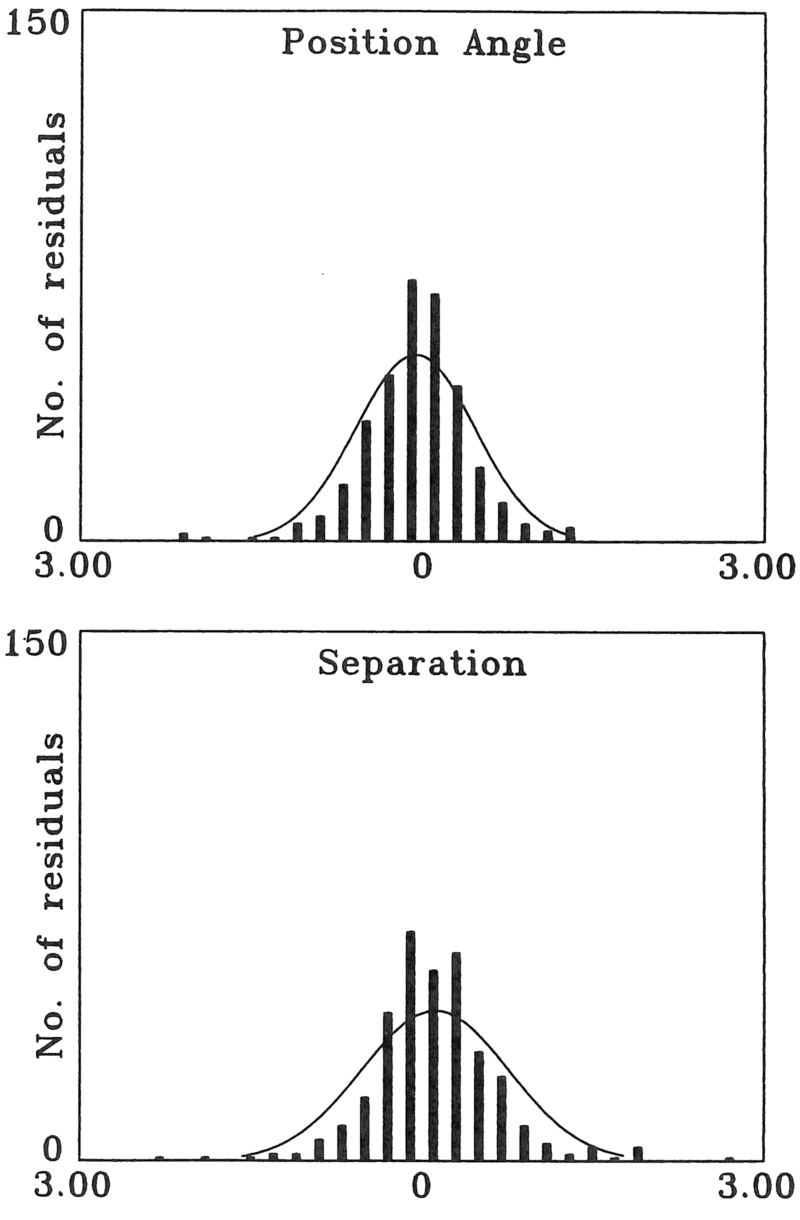



Saturn - Hyperion
80
Position Angle
No. of residuals
0
3.00
0
3.00
80
Separation
No. of residuals
0
3.00
0
3.00
Residuals (arc-seconds)

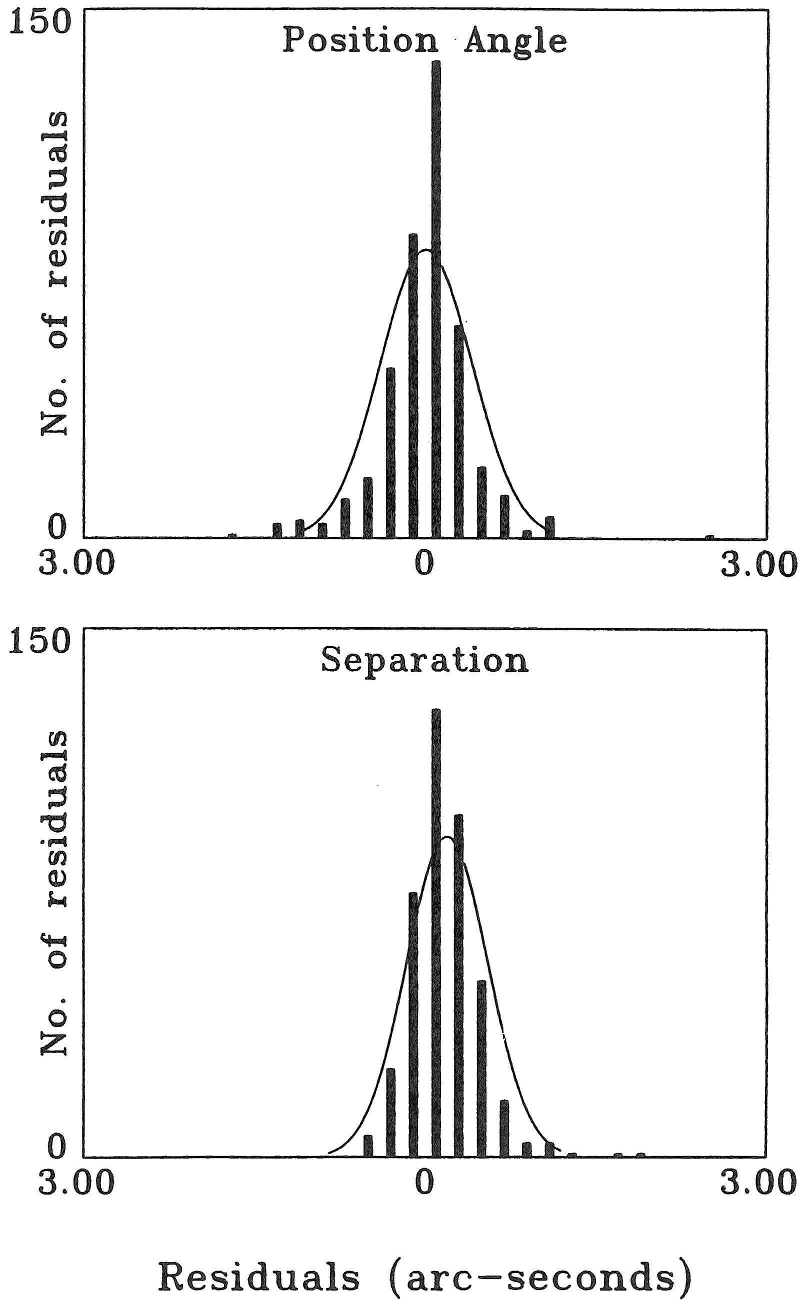
Saturn - Iapetus



Titan – Hyperion



Titan – Iapetus



In the accompanying diagrams we present a number of histograms of O-C residuals from Trial 3 to illustrate the distribution of residuals in all three of the successful (i.e. convergent) trials.

Most of the sets of residuals show a Gaussian distribution centred about zero. This is what we expect if the data are subject only to random errors and this is a basic premise of the least-squares correction process. It is interesting to note, however, that the mean residuals in separation measures of Saturn-Hyperion and Saturn-Iapetus are rather large : the distribution is still nearly Gaussian but with a significantly non-zero mean value. This suggests that these data contain systematic errors. The mean residuals are

Saturn - Titan	+0".160
Saturn - Hyperion	+0".522
Saturn - Iapetus	+0".462

where all the residuals less than 5".0 have been included.

We therefore seek a source of systematic errors which tends to increase the observed separation of Hyperion and Iapetus with respect to Saturn by about 0".5. A number of possibilities may be considered :

<u>The phase defect of the disk of Saturn as seen from the Earth.</u>

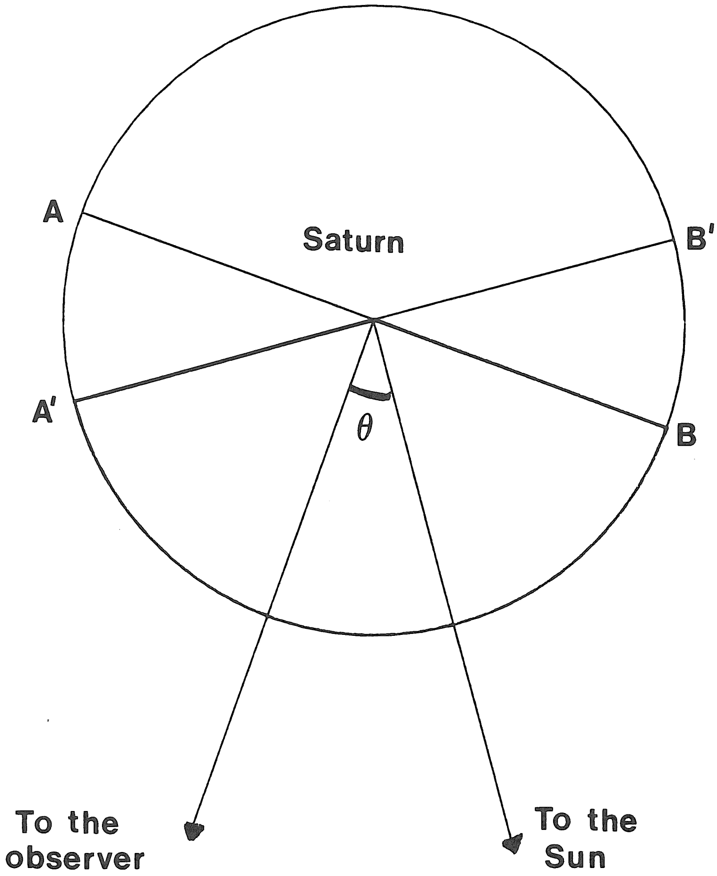


Figure 14. Phase defect of the disk of Saturn

Except at opposition, the illuminated area of Saturn presented to the Earth will not be perfectly circular. One limb will be in shadow (AA' in the accompanying figure) and the visible part of the planet is the segment B to A'. Thus the apparent diameter of the disk is reduced by factor $\sin^2\vartheta/2$ where ϑ is the angle at Saturn subtended by the Sun and the Earth. This has a maximum value of 0.002755 if we assume the Earth and Saturn to move in circular coplanar orbits. Thus the apparent diameter of Saturn (19".4 at mean opposition distance) may be reduced by approximately 0".05 at most by the phase defect. The separation between the planet and a satellite is made with respect to the apparent centre of the planet's disk and this will differ from the centre of the true disk by exactly half the phase defect, $0''.025$. This is too small to satisfy our requirements, and moreover it may act to increase or decrease the measured separation depending on whether the satellite is one the same side of the planet as the phase defect or not. Both cases are equally likely and the net effect on a large number of observations will be zero.

The phase defect may manifest itself in another way which depends upon the reflective properties of the disk of the planet. When the planet is viewed from a direction other than the sub-Solar direction, the centre of illumination of the disk (that is, the point at which the light intensity is greatest) may not coincide with the sub-Solar point. The observer may identify the centre of the disk with the centre of illumination and hence introduce a further error similar to the phase defect. However, as in the case of geometrical phase defect, the average effect upon a large set of observations should be zero. Moreover, it ought to affect all satellites equally unless we invoke some magnitude dependency. I am grateful to Dr. Kaare Aksnes for this suggestion.

The figure of the satellite

Iapetus is known to have an uneven surface albedo : part of the satellite is dark while other parts are brighter. As in the case of Saturn's phase defect, the observer measures separation with respect to the centre of the illuminated part of the satellite and this may not coincide with the centre of mass, introducing an error. However, the apparent diameter of Iapetus at mean opposition distance is $0''.23$, too small to explain the systematic error under consideration. Furthermore, Hyperion is even smaller ($0''.03$ at mean opposition) yet it displays the systematic error even more markedly than Iapetus.

Thickness of the micrometer wire

The wire employed in the micrometer of the Washington 26-inch refractor is stated (Wash. Obs. 1874, Appendix 1) to have a thickness of $0''.251$ in the field of view. Systematic positioning of the wire so that its edge, rather than its centre, coincided with the object under scrutiny may explain part of the large mean separation residuals. However, we can only invoke this argument for the positioning of the wire upon the centre of the disk of Saturn since the same micrometer was used to make the series of inter-satellite measures from 1892 which do not show large mean residuals.

Calibration of the micrometer scale

Newcomb gives the distance corresponding to one revolution of the micrometer screw of the Washington instrument as

$$9''.9480 \pm 0''.0015$$

based upon transit measurements (Wash. Obs. 1874, Appendix 1). The error coresponds to about $\pm 0''.15$ in a measured separation of $1000''$. The largest separation measured in the satellite system of Saturn is some $600''$ and the error in this measure would be $0''.09$.

The error is proportional to the measured separation and so it will cause an error of scale in the entire satellite system. This effect is well known and it results in the major semi-axes of all of the satellites being in error by the same factor. This in turn leads to an erroneous determination of the mass of Saturn when these major semi-axes are used with Kepler's third law.

This cannot, however, be used to explain the large mean separation residuals because the mean residuals are not in proportion to the maximum separation between each satellite and the planet, as we would expect of they were due to multiplicative scale errors.

<u>Unsteadiness of the atmosphere</u>

Turbulence in the atmosphere prevents any celestial object from appearing as a sharply-defined disk or point source. A satellite whose apparent diameter would be 0".5 in the absence of an atmosphere will appear to the observer as a disk perhaps one arc-second across, depending upon local atmospheric conditions. Thus the observer cannot measure the position of the centre of the true disk but only the centre of a larger disk produced by the rapid motion of the object, a motion far faster than the response time of the human eye. This, however, will affect inter-satellite observations in the same way as Saturn-satellite observations and so it cannot be used to explain the anomalously large separation mean residuals for Saturn-satellite observations. I thank Hal Levison for suggesting this to account for residuals in micrometer observations.

<u>Personal errors of observation</u>

Each observer introduces some small systematic error into the data by virtue of his chosen observational technique. It may be significant that most of the Saturn-satellite observations used in this work were made by one observer (Hall). The main exception is the 1874 series of observations which were undertaken by Newcomb. These data, therefore, carry the stamp of Hall's personal error. This may have been a tendency to determine separation measures too large by about half an arc-second for the fainter satellites. Analysis of Hall's observations of the other faint satellites of Saturn during this period could provide further useful information to support this hypothesis, but that must be the subject of future work.

Final parameter sets

In addition to the masses and J factors determined (or adopted) in the three successful trials, the position and velocity vectors of the satellites were also determined at the epoch JED 2418800.5 and they are presented in appendix E.

5.10 CONCLUSIONS AND SUGGESTIONS FOR FURTHER WORK

The principal aim of the work in this chapter was to show that numerical integration can be used successfully in conjunction with visual micrometric observations to model the dynamics of a satellite system. As such, it is a natural extension of the work of Sinclair and Taylor who established the use of numerical integration with photographic astrometric observations. By its very nature, a numerical integration model must be compared to each observation individually : one cannot form opposition mean points as one might when using an analytical theory. This work is heavily dependent on computers, both to perform the integration itself and to carry out the least-squares comparison of the data with the integration. While integration has been widely used in modelling planetary and lunar dynamics since the late 1940s, its application to natural satellite dynamics is a new and developing field. This is due to a large extent to the fact that satellite theories do not generally need to give positions as accurately with respect to the primary in order to yield the same accuracy in the calculated values of the observed positions. It is also important to note that satellite theory has suffered a long period of relative neglect until the advent of space probes to Mars and the outer planets.

The absence of observational data over the period 1930 to 1967 is another problem in the study of the outer satellites of Saturn. With the death of G Struve, observations become scarce, particularly for Hyperion and Iapetus. Having fitted Sinclair's integration to data spanning 1874-1933 (this work) and 1967-1982 (Sinclair and Taylor 1985), the logical next step is to construct a model which is fitted to all the data, covering over a century. This is an ambitious plan and it may prove difficult. In principle, it is a straightforward calculation since both existing models provide partial derivatives with respect to the same parameters of the integration. Equations of condition may be combined to form a set of normal equations regardless of whether the equations of condition are generated by comparison with photographic observations or visual observations.

Among the problems, we may list

1. Significant disparities in the sizes of the random (and systematic) errors in the different types of observation. As we have noted, visual micrometric observations are subject to errors of scale which affect separation measures. Moreover, visual observations relative to Saturn may include systematic errors caused by the method used to make micrometer measurements.

   A system of weighting observations according to their type may be the solution to this problem, though it does not address the question of systematic errors.

2. The mean values of the separation residuals relative to Saturn tend to differ from zero. They should not do this if they were subject only to random errors. An investigation of the cause of this systematic error is necessary in order to give greater validity to the parameters determined in this chapter.

3. The distribution of the data over the 109 years of the proposed global numerical integration is very uneven. It is split into two sets of roughly equal size, about 3000 data in each. One set consists of quite accurate photographic observations over a period of 16 years, while the other is a collection of rather lower accuracy micrometric ob-

servations covering a period of 50 years. There is a 40-year gap between the two sets, where data are all but absent.

4. The resources required to perform such an integration may prove restrictive. Using Sinclair's integration program 'Titan' on an IBM 3083, the two 50-year runs would take a total of about 1200 seconds of CPU time, equivalent to many hours on a VAX machine. The least-squares analysis of the observations must also be added to this. Evidently, careful planning of the logistics of such a project would be essential.