

DYNAMICAL LIMITS ON DARK MASS IN THE OUTER SOLAR SYSTEM

DAVID W. HOGG

Canadian Institute For Theoretical Astrophysics, University of Toronto, McLennan Labs, 60 St. George Street, Toronto, Ontario M5S 1A1, Canada
and
Department of Physics, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139

GERALD D. QUINLAN AND SCOTT TREMAINE

Canadian Institute For Theoretical Astrophysics, University of Toronto, McLennan Labs, 60 St. George Street, Toronto, Ontario M5S 1A1, Canada
Received 26 December 1990; revised 12 February 1991

ABSTRACT

Dark matter in the outer solar system, such as a hypothetical Planet X or a Kuiper belt of material outside Neptune's orbit, may be detected by its dynamical influence on the orbits of the outer planets and comets. The study of simplified model solar systems with known observational errors offers insights into the best strategies to use in dynamical searches for dark mass, the significance of apparent anomalies, and the minimum detectable amounts of dark mass. We use this approach to establish the following results. (i) The inability of modern ephemerides to predict the longitude of Neptune accurately is a consequence of Neptune's long orbital period, and provides no evidence for the existence of dark mass. (ii) Searching for anomalous residuals when fitting observational data to models of the solar system that do not include dark mass is an inefficient way to detect dark mass; a more powerful test (by about a factor of 3–4) is to examine the improvement in fit when the observations are fitted to models that include dark mass. (iii) It is much harder to locate Planet X than to detect its existence: in order to predict the location of Planet X to within 1° its mass must exceed about 10 times the minimum detectable mass. (iv) If Planet X is massive and close enough to exert a detectable dynamical influence on the outer planets, then it is more likely than not that it would have been detected already in the *IRAS* survey. (v) Residuals in the orbits of comets such as Halley's are far more powerful probes for the Kuiper belt than are residuals in the orbits of the outer planets.

1. INTRODUCTION

Residuals in the longitude of Uranus of up to $60''$ led to the prediction of the existence and location of Neptune by Adams and Le Verrier in 1845. The resulting discovery of Neptune within about 1° of Le Verrier's predicted position was one of the great triumphs of celestial mechanics. Since that time astronomers have continued to search for evidence of undiscovered objects ("dark mass") in the outer solar system, by analyzing residuals in the positions of the outer planets, comets, and spacecraft. Pluto was discovered in 1930 during a systematic search that was guided by theoretical predictions of a ninth planet, although the mass of Pluto is now known to be too small for it to have any detectable effect on the other planets.

The study of anomalies in the orbital residuals of the outer planets and comets is plagued by the heterogeneity of the database, which includes observations made over many decades at dozens of observatories, with time dependent and poorly known random and systematic errors. This suggests that an important preliminary step is to analyze simplified model solar systems, both with and without dark mass, in which the distribution of observational errors is known. We shall use this approach to address the following questions:

(i) A number of prediction ephemerides for Neptune have been computed since its discovery, the most recent being the JPL ephemeris DE202. In general, it is found that systematic differences between the predictions and the observations appear within ten years or so of the completion of the ephemeris. The differences are much larger for Neptune than for the other planets. Do these discrepancies signal the existence of dark mass?

(ii) What mass must a distant undiscovered planet have in order to produce detectable anomalies in the planetary orbits? What mass must it have for us to be able to predict its

angular position to within about 1° , as Adams and Le Verrier did for Neptune?

(iii) The planetary system may be surrounded by a disk of material, presumably composed of residual planetesimals that did not accrete to form a planet (the "Kuiper belt"). What limits on the mass of the Kuiper belt can be set from observations of the orbits of comets and the outer planets?

A general review of theoretical and observational constraints on dark matter in the solar system (not including the results of the experiments reported here) is given by Tremaine (1990).

2. NUMERICAL METHODS

We use a simplified model solar system that contains only the four giant planets Jupiter to Neptune. The justification for our neglect of the inner solar system is that ephemerides are far more accurate for the inner planets (Mercury to Mars) than for the outer ones, because of radar and spacecraft ranging observations, and also the effects of distant dark mass on the inner planets will be much smaller than the effects on the outer planets. The initial conditions and planetary masses are taken from Cohen *et al.* (1973, hereafter referred to as CHO). The observer is assumed to be located at the Sun, and measures the latitude β and longitude λ of each of the planets relative to the CHO x - y plane (which is also the invariable plane¹). From the "exact" (E) latitudes and longitudes we generate "observed" (O) latitudes and

¹ The CHO invariable plane differs slightly from the ecliptic and hence our latitudes and longitudes differ from the ecliptic latitude and longitude by up to 2° or so; however, this difference is of no consequence for any of the results of this paper.

longitudes, by adding independent² random errors chosen from a Gaussian distribution with dispersion σ_0 in each coordinate (more precisely, the dispersion in longitude is $\sigma_0/\cos\beta$ to ensure that the errors are isotropic). We assume that N observations are taken at equal time intervals Δt over the period from t_1 to $t_2 = t_1 + (N-1)\Delta t$. We refer to the program that generates the observations as the “simulator.”

The parameters t_1 , t_2 , and Δt are chosen so that the database of simulated observations resembles the database of optical observations used in constructing recent JPL ephemerides such as DE102 (Newhall *et al.* 1983; Standish 1986). The DE102 database contains approximately 1050 observations of each of the four giant planets, spanning the time interval 1911–1971, with rms errors σ_{DE102} in each angular coordinate between $0.37''$ and $0.54''$. We choose $t_1 = 1910$ and $t_2 = 1990$, a slightly longer time span than used by DE102, so that our simulations will more closely resemble newer ephemerides. The initial conditions at time t_1 are found by integrating backwards from the CHO initial conditions, which are given at JD2430000.5 (about 1941.0). The DE102 database contains an average of about $n = 1050/70 = 15$ observations at each opposition; to reduce the computational labor, we assume that all of these can be combined into a single “seasonal” observation. Thus we choose $\Delta t = 1$ yr.

If the actual observational errors were random, independent, and unbiased, the appropriate value for the dispersion σ_0 in the seasonal observation used in our models would be $\sigma_{\text{DE102}}/\sqrt{n}$, which is between $0.10''$ and $0.14''$. However, it is likely that the observational errors include a substantial and time-varying systematic component, in which case our assumption that the errors are drawn from a random Gaussian distribution is unrealistic. Since we cannot model the systematic errors realistically, we shall continue to assume that the errors are random, and account crudely for systematic errors of unknown amplitude by treating the seasonal dispersion σ_0 as a free parameter. Fortunately, simple scaling laws allow us to present all of our results in a form that can be scaled to any choice of σ_0 .

The observations generated by the simulator program are given as input to another program, the “fitter,” which performs a nonlinear least-squares fit to find the initial conditions that generated the observations. The fitter is assumed to know the correct planetary masses, since these were mostly determined by independent methods (Kepler’s law for satellite orbits) in the past and are now known very accurately from spacecraft observations. Following the advice of Oesterwinter & Cohen (1972) we fit for the orbital elements (at time t_1) rather than for the positions and velocities (the least-squares fitting converges more reliably when done this way). We choose as elements the semimajor axis a , the true longitude $f + \omega + \Omega$, $h = e \sin(\omega + \Omega)$, $k = e \cos(\omega + \Omega)$, $p = \sin(i/2) \sin(\Omega)$, and $q = \sin(i/2) \cos(\Omega)$. In these expressions f is the true anomaly, ω the argument of perihelion, Ω the longitude of the ascending node, e the eccentricity, and i the inclination, all measured relative to the CHO invariable plane (latitude $\beta = 0$).

The fitter finds the best solution for the orbital elements of

the planets by minimizing the quantity χ^2 defined below. It also returns a new set of “computed” (C) latitudes and longitudes, which are found by starting the integration with the best-fit orbital elements. Subtracting the observed angular positions from the computed angular positions yields the residuals (O – C) and a value of χ^2 ,

$$\chi^2 = \sum_{i=1}^{4N} \frac{1}{\sigma_0^2} [(\beta_{\text{Oi}} - \beta_{\text{Ci}})^2 + \cos^2(\beta_{\text{Oi}})(\lambda_{\text{Oi}} - \lambda_{\text{Ci}})^2], \quad (1)$$

which measures the goodness of the fit. The nonlinear least-squares fitting is done by the MINPACK subroutine LMDIF developed at the Argonne National Laboratory.

3. PREDICTING THE ORBITS OF URANUS AND NEPTUNE

Postdiscovery observations of Neptune can be fit by a single ephemeris, but roughly ten years after an ephemeris has been constructed Neptune begins to deviate systematically from the ephemeris predictions, mainly in right ascension (Seidemann, *et al.* 1985; Seidemann & Harrington 1988). The discrepancy has appeared in several ephemerides constructed in the past, and is also present in the JPL ephemeris DE200, which is the basis for the major national almanacs and was constructed in 1980. Comparison to the more recent and improved ephemeris DE202 (Standish 1990a) shows that predictions of Neptune’s position over the 20 year interval 1980–2000 are in error by $1''$ in right ascension but only $0.2''$ in declination. A similar but smaller effect is present for Uranus (error $0.35''$ in right ascension, $0.1''$ in declination).

We have investigated this behavior using the programs described in Sec. 2. Using the simulator program we generated a set of observed positions of the four giant planets over the interval $(t_1, t_2) = (1910, 1990)$, at intervals $\Delta t = 1$ yr. Using the fitter program we then found the best-fit solution for the orbital elements of the four planets at time t_1 . No forces from dark mass were included in either the simulator or the fitter. After the fit was complete, we integrated the exact (E) orbits and the computed or best-fit (C) orbits for an additional 20 yr, and measured the latitude and longitude differences (E – C) between the two. The differences are proportional to the seasonal observational error σ_0 . The results, $(\lambda_{\text{E}} - \lambda_{\text{C}})/\sigma_0$ and $(\beta_{\text{E}} - \beta_{\text{C}})/\sigma_0$, are shown in Fig. 1 for about 400 realizations of the random observational errors (note that the results in Figs. 1 and 2 and in Table 1 are not the residuals, which would be O – C).

The rms latitude errors for all four planets are quite similar, about $0.16\sigma_0$, and the longitude errors for Jupiter and Saturn are not much larger. However, the rms error in the predicted longitude for Uranus is roughly 4 times as large as the rms latitude error, and the rms longitude error for Neptune is almost 10 times as large as the rms latitude error (Table 1). This behavior is similar to the behavior of DE200 if it is used to predict the orbits of the giant planets over the 20 yr period 1980–2000 (Standish 1990a). As shown in Table 1, the latitude errors for the four planets have an rms value of $0.07''$, but the longitude errors for the more distant planets are much larger, exceeding $0.3''$ for Uranus and $0.9''$ for Neptune, respectively, 5 and 13 times the rms latitude error. If we compare the rms latitude errors from the simulation and the data we arrive at a value $\sigma_0 = 0.45''$ for the seasonal dispersion. This is probably an overestimate since (i) DE200 is based on data spanning 70 yr rather than our assumed value of 80 yr, (ii) the DE200 errors are determined by comparison to DE202 rather than an exact ephemeris.

² The assumption of independence is not strictly correct: transit circle observations should have independent errors in right ascension and declination rather than ecliptic latitude and longitude.

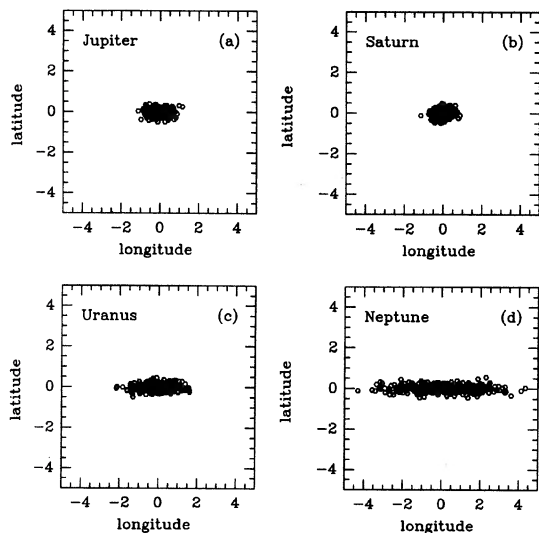


FIG. 1. Differences between the computed and exact latitudes and longitudes of the four giant planets after a 20 yr extrapolation of ephemerides based on 80 yr of observations (1910–1990). The plots show the results from 400 sets of observations with different realizations of the random observational errors. The latitude and longitude differences are measured in units of σ_0 , the seasonal dispersion.

meris, and (iii) dark matter may be contributing to the observed latitude errors. Nevertheless, the best-fit seasonal dispersion is almost certainly larger than the value $0.10''$ – $0.14''$ that would be expected if there were no systematic observational errors.

Given the many idealizations in our model (one observation per year, all observations of equal weight, Gaussian random errors, no systematic errors, exact planetary masses, etc.), the simulated ephemeris is remarkably similar to DE200 in its poor predictive power for the longitudes of Uranus and Neptune. Thus it is likely that the inability of DE200 and other ephemerides to predict Neptune's orbit accurately over 10–20 yr intervals is a natural and inevitable consequence of the limited time span (less than half of one orbital period) over which we have accurate observations of Neptune's orbit.

We have repeated this experiment using more than 80 yr of data, to see how the errors decrease with time. The results, shown in Fig. 2, suggest that the anomalous behavior of Ura-

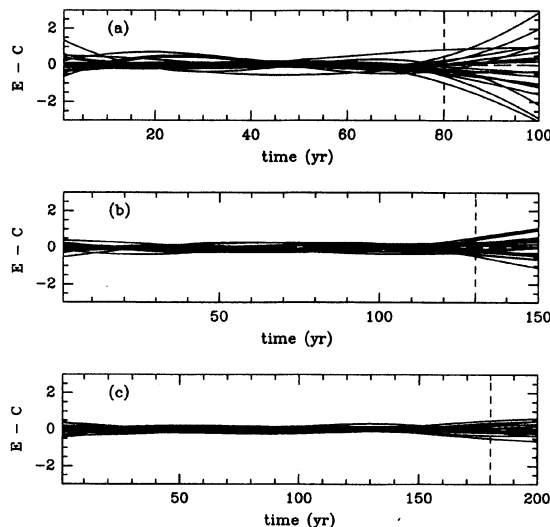


FIG. 2. Differences between the computed and exact longitude of Neptune in a 20 yr extrapolation of ephemerides based on observations spanning (a) 80 yr, (b) 130 yr, and (c) 180 yr. The plots show the results from 20 different sets of observations. The dashed line marks the beginning of the extrapolation. The longitude differences are measured in units of σ_0 .

nus and Neptune will disappear once we have accurate data spanning one full period of Neptune's orbit.

4. LIMITS ON THE MASS OF A TENTH PLANET

4.1 Limits from Orbital Residuals

The most popular explanation for the anomalies in the orbital residuals of Uranus and Neptune is an undiscovered Planet X that orbits beyond Pluto. Some recent predictions of the properties of Planet X are listed in Table 2.

Our goal in this section is not to predict the mass or location of Planet X, a task that we feel is premature given our limited understanding of the origin and significance of the anomalies in the outer planet orbits. Instead we shall try to establish the mass that Planet X must have in order to be detectable in our simulated solar systems. There are several possible definitions of "detectable" in this context:

(i) We can fit the observations to a theoretical model that includes only the known planets and the Sun. If the measurement errors are random, unbiased, and uncorrelated, with

TABLE 1. Extrapolation errors.

Planet	Simulations		DE200	
	$\Delta\beta/\sigma_0$	$\Delta\lambda/\sigma_0$	$ \Delta\beta $	$ \Delta\lambda $
Jupiter	0.16	0.36	0.09"	0.08"
Saturn	0.17	0.33	0.12"	0.15"
Uranus	0.16	0.67	0.00"	0.34"
Neptune	0.16	1.52	0.01"	0.92"

Notes to TABLE 1

The errors ($E - C$) in latitude β and longitude λ for the simulations are measured 20 yr after the completion of an ephemeris based on 80 yr of data; they are the rms values derived from 400 sets of observations. The errors in DE200 are found by comparing its extrapolation from the year 1980–2000 with the newer and improved ephemeris DE202 (Standish 1990a; we have converted the errors in right ascension and declination given by Standish to errors in latitude and longitude).

TABLE 2. Recent Planet X predictions.

	$\mathcal{M}_X$	a_X	e_X	i_X	r_X	$10^6 \mathcal{M}_X / r_X^3$
Harrington (1988)	4	101	0.41	32°	94	4.8
Powell (1988)	0.9	40	0.26	7°	29–50	7–34
Gomes (1989)	0.49	44	0.05	0°	37–46	5.0–6.7

Notes to TABLE 2

The masses and orbital elements are typical numbers taken from the indicated papers; they are not necessarily definite predictions. The mass $\mathcal{M}_X$ is in units of $\mathcal{M}_\oplus$, the semimajor axis a_X and radius r_X are in AU. A range of radii $[a_X(1 - e_X), a_X(1 + e_X)]$ is given if the semimajor axis is small enough that the planet travels a substantial fraction of its orbit during the interval of observation 1910–1990. In this case our approximation that the potential can be represented by a stationary quadrupole becomes inaccurate, so the parameter $\mathcal{M}_X / r_X^3$ provides only a crude estimate of whether the planet is detectable.

known dispersion, then Planet X is detectable—or, at least, some unmodeled effect is present—if the minimum value of χ^2 is significantly larger than expected. We call this “detection by anomalous residuals.”

(ii) We can fit the observations to a theoretical model that includes Planet X, by minimizing χ^2 . Then we say that Planet X is detectable if the minimum value of χ^2 is significantly lower than the minimum value of χ^2 for theoretical models in which Planet X is not present. We shall call this process “detection by modeling.”

(iii) If we minimize χ^2 using a theoretical model that includes Planet X, the best-fit model yields a predicted position for Planet X. In a practical sense, Planet X is detectable only if the predicted position differs from the actual position by less than some small angle $\Delta\theta$, which we shall take to be 1° (both for historical reasons—this was the accuracy of Le Verrier’s prediction for Neptune’s position—and because this is a manageable area to scan on deep photographic plates). We shall call this process “localization” rather than detection.

In general, detection by modeling is more sensitive than detection by anomalous residuals (i.e., modeling can detect a smaller mass), while localization is the least sensitive (i.e., it requires the largest mass).

The search that led to the discovery of Pluto (Tombaugh 1961) covered 75% of the sky but found no other planets. Thus unless Planet X is on a highly inclined orbit and happened to be near the ecliptic pole at the time of Tombaugh’s search, it must be substantially fainter than Pluto. On the other hand, if it is to perturb significantly the orbits of the outer planets, its gravitational force must be substantially larger than Pluto’s. Since the optical flux from a planet at a large distance r_X varies as r_X^{-4} , while the gravitational force in a heliocentric frame varies as r_X^{-3} , these constraints can most easily be reconciled if Planet X is much more distant from the Sun than Pluto. (We consider the observational constraints on Planet X in more detail in Section 4.2).

If we assume that Planet X is distant, we can approximate its potential (in heliocentric coordinates) by the quadrupole component of an expansion in spherical harmonics,

$$\Phi_X(\mathbf{r}) = (G\mathcal{M}_X/2r_X^3) [r^2 - 3(\mathbf{r}\cdot\hat{\mathbf{r}}_X)^2], \quad (2)$$

where $\mathcal{M}_X$ and $\mathbf{r}_X$ are the mass and position vector of Planet X and $\hat{\mathbf{r}}_X = \mathbf{r}_X/r_X$. The fractional error in this expression is $O(r/r_X)$. Furthermore, since Planet X is distant its orbital period is so large that we can ignore its orbital motion over an 80 yr time span and treat $\hat{\mathbf{r}}_X$ as a fixed vector. Thus, the seven parameters that describe a nearby Planet X (six orbital elements and the mass) are reduced to three (the latitude and longitude describing the direction of $\hat{\mathbf{r}}_X$, and the tidal parameter $\mathcal{M}_X/r_X^3$). Because the potential (2) is invariant under the transformation $\hat{\mathbf{r}}_X \rightarrow -\hat{\mathbf{r}}_X$, the direction $\hat{\mathbf{r}}_X$ can be restricted to half of the sky, which we choose to be the longitude range 0° – 180° . Finally, since the orbits of the four giant planets lie close to the x - y plane, the effect of a Planet X with a positive latitude β should be nearly the same as the effect of a Planet X with a negative latitude $-\beta$, so we can restrict our attention to latitudes in the range 0° – 90° .

We began each of our numerical experiments by choosing a direction $\hat{\mathbf{r}}_X$ for Planet X. We examined 31 different directions on the sky—because of the symmetries these were equivalent to 98 positions on the whole celestial sphere—with latitudes 0° , 30° , 60° , and 90° and longitudes in the range 0° – 180° . For each direction we chose a number of values of

the tidal parameter $\mathcal{M}_X/r_X^3$ that measures the strength of the perturbation from Planet X. For each value of the tidal parameter we used the simulator to generate observed (O) latitudes and longitudes for the four giant planets over the interval $(t_1, t_2) = (1910, 1990)$, at intervals $\Delta t = 1$ yr, and with fixed seasonal dispersion σ_0 . The accelerations of the planets included the influence of the potential (2). We then passed the observations to the fitter program, which determined the best-fit orbital elements of the four planets at time t_1 .

We first investigated whether Planet X could be detected by anomalous residuals, by instructing the fitter program to calculate the best-fit initial conditions for the orbits on the assumption that no Planet X was present (i.e., $\mathcal{M}_X = 0$). From these initial conditions we find the computed (C) latitudes and longitudes, the residuals $(O - C)$, and χ^2/ν , which measures the goodness of the fit. Here ν is the number of degrees of freedom, equal to the number of observations $2 \times 4 \times N = 640$ (latitude and longitude are independent observations) minus the number of fitted parameters (in this case, 24: six orbital elements for each of four planets).

An example of the results is shown in Fig. 3. The dashed lines marked 10%, 1%, and 0.1% in the figure are the values of χ^2/ν that would be exceeded by the indicated fraction of independent realizations of the random observational errors, if $\mathcal{M}_X$ were really zero. We fit the χ^2/ν values found from our experiments by a parabola

$$\chi^2/\nu = 1 + \alpha_1 (\mathcal{M}_X/\sigma_0 r_X^3)^2, \quad (3)$$

since the expectation value of χ^2/ν is unity when $\mathcal{M}_X = 0$, and should increase as $(\mathcal{M}_X/\sigma_0 r_X^3)^2$. Here α_1 is a free parameter that is determined by the fitting procedure. The intersection of this parabola with the three dashed lines then

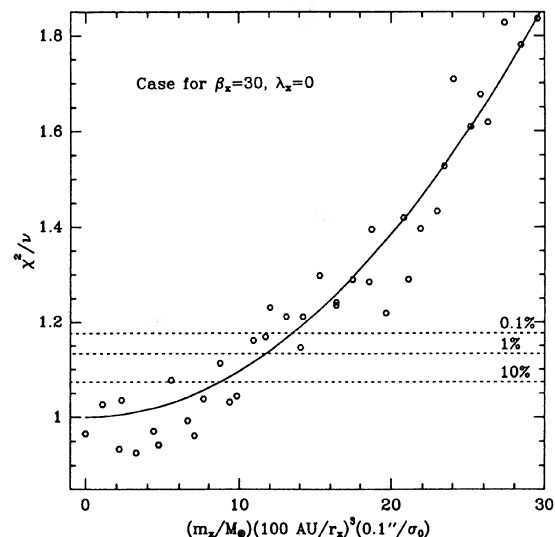


FIG. 3. χ^2 per degree of freedom found by fitting 80 yr of observations to the orbits of the four giant planets. The orbits were perturbed by a planet of mass $\mathcal{M}_X$ at a distance r_X and fixed latitude $\beta_X = 30^\circ$ and longitude $\lambda_X = 0^\circ$; the fit was done assuming $\mathcal{M}_X = 0$. The plot shows the results from 39 different choices of $\mathcal{M}_X$; the solid line is a parabolic fit to the results. If $\mathcal{M}_X$ were really zero, we would expect the value of χ^2 to lie above the dashed line labeled 10% in 10% of the realizations; the lines marked 1% and 0.1% are defined in an analogous way.

tells us the values of $\mathcal{M}_X/r_X^3$ that would degrade the residuals significantly at the 90%, 99%, and 99.9% confidence levels, for a typical realization of the random observational errors. We shall state that Planet X is detectable by anomalous residuals if the parabola (3) lies above the 1% line in Fig. 3, corresponding to a 99% confidence level.

We used a similar procedure to investigate whether Planet X could be detected by modeling. We instructed the fitter to find the best-fit orbits twice: first, on the assumption that no Planet X was present, and second, fitting for $\mathcal{M}_X/r_X^3$, λ_X , and β_X as well as the orbital parameters of the giant planets. We then computed $\Delta\chi^2$, the amount by which χ^2 was lowered in the fit with the three extra free parameters. If the mass of Planet X is really zero then the difference $\Delta\chi^2$ should be distributed as a chi-squared distribution with 3 degrees of freedom.

An example of the results is shown in Fig. 4. The dashed lines marked 10%, 1%, and 0.1% in the figure are the values of $\Delta\chi^2$ that would be exceeded by the indicated fraction of independent realizations of the random observational errors, if $\mathcal{M}_X$ were really zero. We fit the $\Delta\chi^2$ values found from our experiments by a parabola

$$\Delta\chi^2 = 3 + \alpha_2 (\mathcal{M}_X/\sigma_0 r_X^3)^2, \quad (4)$$

since the expectation value of $\Delta\chi^2$ is 3 when $\mathcal{M}_X = 0$, and should increase as $(\mathcal{M}_X/\sigma_0 r_X^3)^2$. Here α_2 is a free parameter that is determined by the fitting procedure. The intersection of this parabola with the three dashed lines then tell us the values of $\mathcal{M}_X/r_X^3$ that would be detectable by this fitting procedure at the 90%, 99%, and 99.9% confidence levels, for a typical realization of the random observational errors.

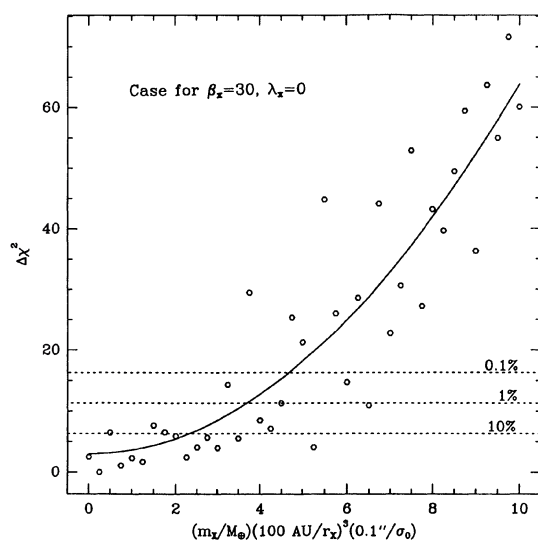


FIG. 4. Reduction in χ^2 achieved by including Planet X in the fitting program. The orbits of the four giant planets were perturbed by a planet of mass $\mathcal{M}_X$ at a distance r_X and fixed latitude $\beta_X = 30^\circ$ and longitude $\lambda_X = 0^\circ$. The positions of the planets were observed at 1 yr intervals over an 80 yr time span. The fit was first done assuming $\mathcal{M}_X = \lambda_X = \beta_X = 0$ (as in Fig. 3), and was then repeated with $\mathcal{M}_X/r_X^3$, λ_X , and β_X as adjustable parameters; $\Delta\chi^2$ is the amount by which χ^2 was lowered. The solid line is a parabolic fit to the results. The dashed lines labeled 10%, 1%, and 0.1% mark the values of $\Delta\chi^2$ that would be exceeded by the indicated fraction of the realizations if $\mathcal{M}_X$ were really 0.

We shall state Planet X is detectable by modeling if the parabola (4) lies above the 1% line in Fig. 4, corresponding to a 99% confidence level.

The results of these experiments are summarized in Fig. 5, which shows the detectability limit as a function of the direction to Planet X. The figure shows that a distant planet in the ecliptic should be detectable by anomalous residuals if

$$\frac{\mathcal{M}_X}{\mathcal{M}_\oplus} \gtrsim (20-30) \left(\frac{\sigma_0}{0.1''} \right) \left(\frac{r_X}{100 \text{ AU}} \right)^3, \quad (5)$$

where the range in the coefficient reflects the dependence on the longitude of Planet X. The planet would be detectable by modeling if

$$\frac{\mathcal{M}_X}{\mathcal{M}_\oplus} \gtrsim (6-10) \left(\frac{\sigma_0}{0.1''} \right) \left(\frac{r_X}{100 \text{ AU}} \right)^3. \quad (6)$$

Note that a planet at a higher latitude (30° or 60°) is easier to detect. However, a planet at the ecliptic pole $\beta = 90^\circ$ is much harder to detect, with a detectability limit of about $\mathcal{M}_X = 200\mathcal{M}_\oplus$ (anomalous residuals) or $\mathcal{M}_X = 70\mathcal{M}_\oplus$ (modeling) at $r_X = 100$ AU. The reason is that in this case $\mathbf{r} \cdot \hat{\mathbf{r}}_X$ is nearly zero, so the perturbing potential (2) is spherically symmetric. The principal effects of a spherical potential are to change the relation between semimajor axis and mean motion (which is hard to detect without ranging information), and to induce precession of the apse (which is hard to detect since the planetary eccentricities are low). Thus the nonaxisymmetric perturbations are by far the easiest to detect.

Detection by modeling is about three times as sensitive as detection by anomalous residuals according to Eqs. (5) and (6). The explanation for this is that the anomalous-residuals test is not sensitive to patterns in the residuals (since it depends only on the sum of the squared residuals), while the modeling test is. This difference between the two tests in-

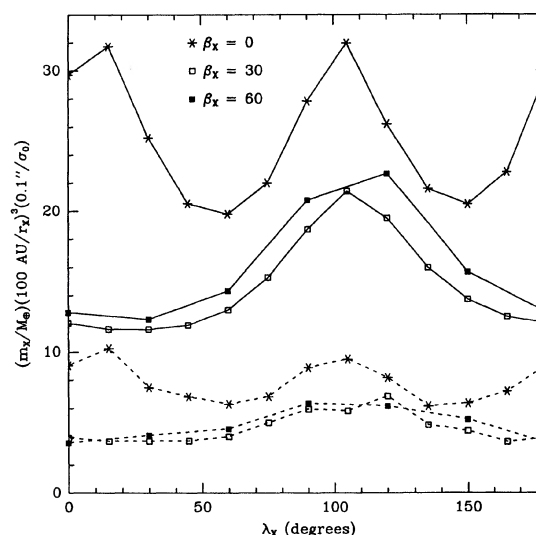


FIG. 5. Limits on the mass $\mathcal{M}_X$ of a planet at a distance r_X , latitude β_X , and longitude λ_X that can be detected at the 99% confidence level when fitting 80 yr of observations of the four giant planets. The solid lines show the limits found from the requirement that the residuals not be too large (see Fig. 3); the dashed lines show the limits found from the improvements in χ^2 when the Planet X parameters are included in the fit (see Fig. 4).

creases as the time span of the data increases; for example, with a 200 yr time span detection by modeling is nearly four times as sensitive as detection by anomalous residuals.

The detectability limits in Eqs. (5) and (6) are overoptimistic, since we have neglected systematic observational errors. If we account crudely for systematic errors by increasing the seasonal dispersion σ_0 to $0.45''$, as we estimated in the previous section, then all of the mass limits would increase by a factor of 4–5 over their values for $\sigma_0 = 0.1''$. However, neither test is reliable if systematic errors are present, since they are based on the assumption that the errors are random.

Because of the unknown character of the systematic errors, the only certain way to detect Planet X is to localize it by least-squares fits and then find it by an optical or infrared search. Each panel in Fig. 6 shows the error in the fitted position of Planet X for a given latitude and mass and a random selection of longitudes, once again using 80 yr of data and seasonal dispersion $\sigma_0 = 0.1''$. For example, if Planet X lies in the ecliptic, we require $\mathcal{M}_X \gtrsim 100 \mathcal{M}_\oplus$ at $r_X = 100$ AU for most realizations to have position errors $\lesssim 1''$; for comparison the mass detectable by modeling is about $10 \mathcal{M}_\oplus$ and the mass detectable by anomalous residuals is about $30 \mathcal{M}_\oplus$. At higher latitudes the differences are even larger.

Our results can be compared with published predictions of the parameters of Planet X (Table 2). The planets predicted by Powell (1988) and Gomes (1989) have relatively small semimajor axes, and hence our approximations (that only the quadrupole potential is important, and that the planet does not move during the observational period) are of doubtful validity; nevertheless, the tidal parameter $\mathcal{M}_X/r_X^3$ should still provide a crude estimate of the strength of the perturbation. The planets predicted by Harrington (1988) and Gomes (1989) are small enough that they would barely be detectable according to Fig. 5 if $\sigma_0 \gtrsim 0.1''$. One difference between the Planet X predictions in Table 2 and our investi-

gations is that they used data over a longer time span (back to 1790), whereas we have taken the more conservative view than only data since 1910 or so should be used. We have carried out fits using longer time spans $t_2 - t_1$ (up to 600 yr), assuming the same dispersion σ_0 for all the data (although in practice, of course, σ_0 cannot be treated as constant even over the past 100 yr). For example, increasing the time span from 80 to 150 yr decreases the minimum detectable mass in the ecliptic by about a factor of 4 (detection by anomalous residuals) or 4.5 (detection by modeling).

We can also compare our results with those of Seidelmann (1971) and Rawlins & Hammerton (1973). Seidelmann integrated the motions of the outer planets along with three hypothetical planets, called P1, P2, and S1 (based on Pickering's predictions of transneptunian planets). He found that planet P1 would perturb the outer planets in an obvious way, but that the perturbations from planets P2 and S1 would be too small to be detectable. This is crudely consistent with our detectability limit (6), since planet P1 has $\mathcal{M}_X/r_X^3 \approx 59 \mathcal{M}_\oplus / (100 \text{ AU})^3$ while P2 and S1 have $\mathcal{M}_X/r_X^3 \approx 1 \mathcal{M}_\oplus / (100 \text{ AU})^3$. Rawlins and Hammerton used classical perturbation theory to determine the effect on the orbits of Neptune and Uranus of a hypothetical tenth planet on a circular orbit in the plane of the ecliptic. They considered planets with various starting longitudes and with semimajor axes in the range 40–600 AU, and performed a least-squares fit (by comparing the perturbations with the actual residuals) to determine the mass of Planet X as well as corrections to the orbital elements of the disturbed planet. They found solutions that substantially reduced the residuals with $\mathcal{M}_X/r_X^3 \approx (19\text{--}26) \mathcal{M}_\oplus / (100 \text{ AU})^3$, which is above the limit in Eq. (6) if $\sigma_0 \lesssim 0.2''$. They also use a longer time span of observations than used in our simulations.

4.2 Other Limits

4.2.1 The Goldreich–Ward limit

A massive Planet X on an inclined orbit would exert a torque on the planets that could either cause the plane of the ecliptic to precess away from the solar equator, or cause Neptune's orbital plane to precess away from the ecliptic (Goldreich & Ward 1972). The absence of either effect sets limits on the mass of Planet X if its orbit is inclined; these limits are particularly important because Tombaugh's (1961) photographic search was less complete at high ecliptic latitudes, and because the dynamical constraints from the previous section are weak near the ecliptic pole. The following quantitative estimate of the limits on a Planet X in an inclined orbit is very similar to estimates given by Goldreich and Ward, although the derivation is somewhat different.

Suppose that Planet X is on an inclined, circular orbit of radius r_X . If $\mathcal{M}_i$ and a_i are the masses and semimajor axes of the known planets, then Planet X will contain most of the angular momentum in the solar system if $\mathcal{M}_X r_X^{1/2} \gg \sum_i \mathcal{M}_i a_i^{1/2}$, that is, if

$$\frac{\mathcal{M}_X}{\mathcal{M}_\oplus} \gg 120 \left(\frac{100 \text{ AU}}{r_X} \right)^{1/2}. \quad (7)$$

Mutual torques among the known planets, and between the known planets and Planet X, lead to precession of the planetary orbital angular momentum vectors.

We first consider the case in which mutual torques among the known planets are strong enough that their orbits precess as a unit under the influence of Planet X, that is, the inclina-

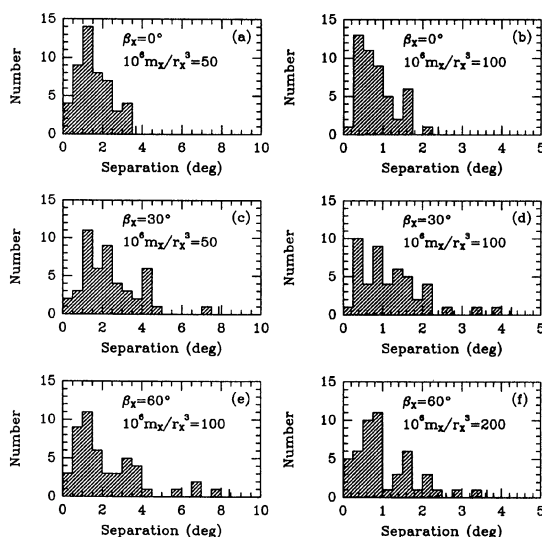


FIG. 6. Angular separation between true and fitted positions of Planet X (using data with $\sigma_0 = 0.1''$ spanning 80 yr). Each histogram shows the results of about 50 experiments with randomly chosen values of the longitude λ_X ($\mathcal{M}_X$ is in units of $\mathcal{M}_\oplus$; r_X is in AU).

tions of the orbits of the known planets relative to the ecliptic will remain small, even though the plane of the ecliptic rotates. If inequality (7) is satisfied, the normal $\hat{n}$ to the plane of the known planets will precess around the normal $\hat{n}_X$ to the orbit of Planet X. The precession rate is (e.g., Goldreich 1965)

$$\frac{d\hat{n}}{dt} = \frac{3G\mathcal{M}_X}{4r_X^3} \frac{\sum_i \mathcal{M}_i a_i^2}{\sum_i \mathcal{M}_i (G\mathcal{M}_\odot a_i)^{1/2}} (\hat{n} \times \hat{n}_X) (\hat{n} \cdot \hat{n}_X). \quad (8)$$

The torque on the Sun from the planets is negligible, so the direction of the solar spin axis remains unchanged during this precession. At present, the angle between the ecliptic pole and the solar spin axis is only $7^\circ \approx 0.12$ radian. Unless this near alignment is a coincidence, the precession of the ecliptic over the history of the solar system must have been small, say $t_0 |d\hat{n}/dt| < 0.1$ radian ($t_0 = 4.5 \times 10^9$ yr is the age of the solar system). To allow for the possibility of other torques, or an initial misalignment that was partially cancelled by precession, we shall take the more conservative limit $t_0 |d\hat{n}/dt| < 1$ radian. Taking a typical value $|(\hat{n} \times \hat{n}_X) (\hat{n} \cdot \hat{n}_X)| = 0.25$, Eq. (8) yields

$$\frac{\mathcal{M}_X}{\mathcal{M}_\oplus} < 1.9 \left(\frac{r_X}{100 \text{ AU}} \right)^3. \quad (9)$$

This limit applies only if inequality (7) is satisfied; combining the two inequalities we conclude that upper limits on the mass of an inclined Planet X are given by

$$\frac{\mathcal{M}_X}{\mathcal{M}_\oplus} < 1.9 \left(\frac{r_X}{100 \text{ AU}} \right)^3, \quad \text{for } r_X \gtrsim 320 \text{ AU}; \quad (10)$$

$$\frac{\mathcal{M}_X}{\mathcal{M}_\oplus} < 120 \left(\frac{100 \text{ AU}}{r_X} \right)^{1/2}, \quad \text{for } r_X \lesssim 320 \text{ AU}. \quad (11)$$

These constraints apply only if $|(\hat{n} \times \hat{n}_X) (\hat{n} \cdot \hat{n}_X)| \gtrsim 0.25$, that is, if the inclination of the orbit of Planet X is between 15° and 75° or between 105° and 165° . If the inclination is closer to 0° or 90° the constraints (10) and (11) are weaker; on the other hand, the constraint is conservative because we have allowed 1 radian of precession while the actual alignment of the ecliptic pole and solar spin is within 0.12 radian. If $r_X \gtrsim 5 \times 10^3$ AU the Galactic tide will perturb the orbit of Planet X and ensure that it has a substantial inclination over most of the age of the Solar System (Tremaine 1990).

Now consider the case where the planetary orbits do not precess as a unit. The planet most weakly coupled to the rest of the known solar system is Neptune; a crude criterion is that Neptune's orbit pole will follow the orbit poles of the other known planets if its precession frequency ν_s due to the known planets exceeds its precession frequency ν_X due to Planet X. If the mutual inclinations are not too large, the precession frequencies are

$$\nu_s = \frac{3G\sum' \mathcal{M}_i a_i^2}{4(G\mathcal{M}_\odot)^{1/2} a_N^{7/2}} = (4.6 \times 10^5 \text{ yr})^{-1}, \quad (12)$$

$$\begin{aligned} \nu_X &= \frac{3G\mathcal{M}_X a_N^{3/2}}{4(G\mathcal{M}_\odot)^{1/2} r_X^3} \\ &= (4.2 \times 10^8 \text{ yr})^{-1} \frac{\mathcal{M}_X}{\mathcal{M}_\oplus} \left(\frac{100 \text{ AU}}{r_X} \right)^3, \end{aligned} \quad (13)$$

where the subscript N denotes Neptune and a prime denotes a sum over all planets except Neptune. The constraint

$\nu_s > \nu_X$ then yields

$$\frac{\mathcal{M}_X}{\mathcal{M}_\oplus} < 900 \left(\frac{r_X}{100 \text{ AU}} \right)^3. \quad (14)$$

Again, this is a conservative constraint in that the present inclination of Neptune is only about 0.01 radian; if this is not a coincidence and Planet X has a substantial inclination then ν_s must be much greater than ν_X .

The limits (10), (11), and (14) are shown in Fig. 7 as a zigzag dashed line. We also show the lines $\mathcal{M}_X/\mathcal{M}_\oplus = \epsilon(r_X/100 \text{ AU})^3$ for $\epsilon = 5, 10$, and 20 . The line for $\epsilon = 5$ corresponds to the smallest Planet X mass that has been used to model the orbital residuals of the outer planets (Table 2); the lines for $\epsilon = 5$ and 10 bracket our estimate of the minimum detectable mass [Eq. (6)] if the seasonal dispersion is $0.1''$; the lines $\epsilon = 10$ and 20 bracket

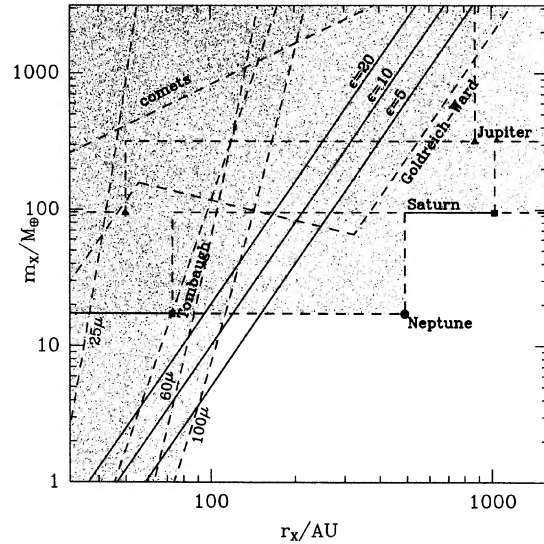


FIG. 7. Limits on the distance r_X and mass $\mathcal{M}_X$ of Planet X. The dashed line labeled "Goldreich-Ward" marks the boundary of the region ruled out by the Goldreich-Ward limit [Eqs. (10), (11), and (14)] if the orbit is inclined. The dashed line labeled "Tombaugh" marks the boundary of the region ruled out by Tombaugh's search, assuming that the planet has visual albedo $p \geq 0.02$ and density $\rho \leq 2 \text{ g cm}^{-3}$, and that the limiting visual magnitude of the search was 15.2. The dashed line labeled "comets" denotes the region ruled out because Planet X does not significantly perturb the energies of new comets [Eq. (16)]. The dashed lines labeled with wavelengths in microns denote the regions ruled out because the reemitted solar radiation would exceed the *IRAS* flux limit in the given wavelength band (assuming $\rho = 2 \text{ g cm}^{-3}$). The dashed staircases denote the distances at which the intrinsic luminosities of Jupiter, Saturn, and Neptune would exceed the *IRAS* flux limits at wavelengths of $12 \mu\text{m}$ (triangles), $25 \mu\text{m}$ (squares), and $60 \mu\text{m}$ (circles); the limits at $100 \mu\text{m}$ are almost the same as the limits at $60 \mu\text{m}$. Neither the *IRAS* constraints nor Tombaugh's constraint are inflexible since neither survey was complete (see text). The density of shading is proportional to the number of constraints violated, counting 1 each for the comet constraint, the Goldreich-Ward constraint, the Tombaugh constraint, the *IRAS* constraint from reemitted solar radiation, and the *IRAS* constraint from intrinsic luminosity. The heavy solid lines denote estimates of the mass and distance at which Planet X would exert a significant dynamical influence on the orbits of the outer planets, $\mathcal{M}_X/\mathcal{M}_\oplus = \epsilon(r_X/100 \text{ AU})^3$, for $\epsilon = 5, 10$, and 20 [compare Eq. (6)].

the minimum detectable mass in the more realistic—but still optimistic—case that the seasonal dispersion is $0.2''$. The figure shows that if Planet X is inclined and its mass exceeds about $100 M_{\oplus}$, the Goldreich–Ward limit provides a more sensitive constraint than that found by modeling the orbits of the outer planets. Any dynamically significant undiscovered planet of mass $100 M_{\oplus}$ or greater must lie close to the ecliptic to escape the Goldreich–Ward limit.

4.2.2 Tombaugh's limit

The search that led to the discovery of Pluto still provides a valuable limit on the brightness of any undiscovered planet (Tombaugh 1961). Tombaugh's search covered most of the sky, although the search was less sensitive at high ecliptic latitudes. Near the ecliptic, Tombaugh's estimated limiting magnitude was 16, with "a good chance" that a planet with magnitude < 17 would have been discovered.

To convert Tombaugh's limiting magnitude into a constraint on the mass and distance of Planet X, we assume visual albedo $p = 0.02$ (since small bodies in the outer solar system are usually dark), density $\rho = 2 \text{ g cm}^{-3}$ (higher than the known densities of most satellites in the outer solar system), and visual limiting magnitude 15.2 (Tombaugh's estimate of 16 with a correction of 0.8 from photographic to visual). We find

$$\frac{M_X}{M_{\oplus}} < 3.4 \left(\frac{r_X}{100 \text{ AU}} \right)^6, \quad (15)$$

which is also plotted on Fig. 7. The limit is a conservative one, both because Tombaugh probably could detect planets fainter than photographic magnitude 16, and because our assumed albedo is low. In particular, the rock-ice cores of the giant planets all have masses less than $30 M_{\oplus}$; thus it is plausible that any undiscovered planet with mass $\gtrsim 30 M_{\oplus}$ would have a gaseous envelope and hence would have a much higher albedo, similar to those of the giant planets ($p = 0.4\text{--}0.6$).

4.2.3 IRAS limits

The *Infrared Astronomical Satellite* (IRAS) survey (Chester 1986; Beichman *et al.* 1988) provides the most sensitive limits on Planet X from direct observations; unfortunately, we shall see that the interpretation of the IRAS results is complicated by selection effects, so that firm limits on the parameters of Planet X cannot be derived.

The IRAS survey was carried out in four wavelength bands, centered at 12, 25, 60, and $100 \mu\text{m}$. The flux limits in each band are roughly 0.4, 0.5, 0.6, and 1–2 Jy, respectively (we shall take 1 Jy as the flux limit at $100 \mu\text{m}$ in the discussions below).

There are two distinct contributions to the infrared flux from Planet X. First, incident solar radiation is absorbed and reemitted as thermal radiation. Assuming that the planet is a black body and has no other source of heat its temperature will be $T = 280 \text{ K} (r_X/\text{AU})^{-1/2}$, and assuming a mass M_X and mean density ρ (once again we take $\rho = 2 \text{ g cm}^{-3}$) we can compute the maximum distance r_X at which the reemitted flux from the planet would exceed the IRAS flux limit in each wavelength band. These limits are plotted in Fig. 7.

A second contribution to the flux comes from the intrinsic luminosity of the planet. The giant planets Jupiter, Saturn, and Neptune all have substantial intrinsic luminosities, arising from secular cooling from a hot initial state or (in the

case of Saturn) gravitational settling of helium. The lack of detectable intrinsic luminosity from Uranus probably reflects its relatively low mass and small distance from the Sun: heat transport to the surface is inhibited by solar heating. Rather than attempt to estimate the intrinsic luminosity of a planet of arbitrary mass, we have simply shown on Fig. 7 the distances at which the intrinsic emission from Jupiter, Saturn, and Neptune would exceed the IRAS flux limits.³ The limits depicted are conservative because we have considered only three planets; to be more realistic we should draw a smooth curve that connects the outer corners of the excluded region and extends to masses below Neptune's.

Unfortunately it is not certain that Planet X would be detected by IRAS if its flux exceeded these limits, for several reasons. A small fraction of the sky (4%) was not covered by the survey; moreover, in about 4% of the sky near the Galactic plane the flux limit was higher because of confusion, especially near $100 \mu\text{m}$. A more serious complication is that the detectability depends on the proper motion (the following discussion is based on Chester 1990). If the proper motion is very low, the planet would be indistinguishable from a stationary source, and hence might lie in the catalog undetected unless it happened to be at high Galactic latitude (where most sources have been identified). On the other hand, if the proper motion is too large the source might be rejected as spurious.

IRAS observations were arranged so that every source was observed twice within about 1.5 days; this pair of observations is called an "hours-confirming coverage" or HCON. All sources received two HCONs within 1–2 weeks; a third HCON was started roughly 6 months later, but only 72% of the sky was covered before the mission was terminated.

If $r_X \gtrsim 2500 \text{ AU}$ there would be no detectable proper motion and hence Planet X would probably not have been noticed, even if it were in the IRAS Catalog, unless it were at high Galactic latitude. If $400 \text{ AU} \lesssim r_X \lesssim 2500 \text{ AU}$, the proper motion would be detectable, but only if Planet X were in the region covered by three HCONs. If $r_X \lesssim 400 \text{ AU}$ the proper motion might be detectable with only the first two HCONs.

Thus, Planet X should already have been identified in the IRAS survey if its flux exceeds the IRAS flux limits, as shown in Fig. 7, and if in addition it is (i) not in the 4% of the sky left unexamined by IRAS, (ii) not too close to the Galactic plane, (iii) not beyond 2500 AU, and (iv) within the 72% of the sky covered by 3 HCONs, or close enough that its proper motion was detectable with only the first two HCONs. Obviously we cannot rule out the possibility that Planet X might have been missed by IRAS, or might lie unnoticed within the IRAS Catalog, but it is more likely than not that Planet X would have been detected if its flux exceeds the IRAS flux limits.

Given that the excluded regions shown in Fig. 7 are conservative ones, and given that the curves $\epsilon = 10$ and 20 bracket the minimum detectable mass in the optimistic case

³ We have used effective temperatures of 110, 74, and 51 K, respectively; these are lower than the observed effective temperatures of the intrinsic luminosity (124, 95, 56 K) to account for the fact that the planets cool more rapidly at a larger distance from the Sun. We have estimated these corrections assuming that the effective temperature T_e obeys the relation $dT_e/dt = K(T_e^4 - T_s^4)$ where K is a constant and T_s is the effective temperature that would obtain if the solar luminosity were the sole heat source.

that the seasonal dispersion is only $0.2''$ and there are no systematic errors [Eq. (6)], we may conclude that it is unlikely that any Planet X large enough to exert significant perturbations on the known planets would have escaped detection by *IRAS*.

4.2.4 Other arguments

A variety of other arguments providing dynamical constraints on the mass of Planet X are summarized by Tremaine (1990). The most significant of these for a relatively nearby Planet X (say, $r_X \lesssim 10^3$ AU) comes from the observation that the mean inverse semi-major axis of new comets is a_0^{-1} , where $a_0 = 2 \times 10^4$ AU. This is close to the value predicted by theoretical models of the Oort cloud. A Planet X with $r_X \lesssim a_0$ would change the average inverse semimajor axes of incoming comets by $\Delta(1/a) = 2\mathcal{M}_X/(\mathcal{M}_\odot r_X)$. If $\Delta(1/a)$ were as large as a_0^{-1} the current good agreement between theory and observation would be destroyed, yielding a limit (Kirk 1978)

$$\mathcal{M}_X \lesssim \frac{\mathcal{M}_\odot r_X}{2a_0} \quad \text{or} \quad \frac{\mathcal{M}_X}{\mathcal{M}_\oplus} \lesssim 820 \left(\frac{r_X}{100 \text{ AU}} \right), \quad (16)$$

which is plotted in Fig. 7.

5. LIMITS ON THE MASS OF THE KUIPER BELT

The Kuiper belt is a hypothetical disk of dark matter that orbits in the ecliptic beyond Neptune (Kuiper 1951). It is presumably composed of planetesimals from the protoplanetary disk that have not been incorporated into a planet. The Kuiper belt may be the source of the Jupiter-family comets and Chiron (Duncan *et al.* 1988).

In order to have a definite model for the potential from the Kuiper belt, we shall assume that the belt is sufficiently far outside the orbit of Neptune that we can approximate its potential by the quadrupole component of an expansion in spherical harmonics,

$$\Phi_d(x, y, z) = \frac{G\mathcal{M}_d}{4r_d^3} (2z^2 - x^2 - y^2). \quad (17)$$

The belt is then described by one free parameter, the ratio $\mathcal{M}_d/r_d^3$. If $\mathcal{M}_d/r_d^3$ is sufficiently large then the belt will cause detectable perturbations in the motion of the outer planets and comets such as Halley's comet.

5.1 Limits from the Outer Planets

The procedure we used to examine the effect of Planet X on the orbits of the outer planets can also be used to examine the effect of the Kuiper belt. We added to our simulator program the forces exerted by the belt potential (17) for various values of $\mathcal{M}_d/r_d^3$. We took observations at intervals of 1 yr over the interval 1910–1990, added Gaussian random errors with dispersion σ_0 , and then fit for the orbital elements of the outer planets at the starting time.

We first investigated whether the belt could be detected by anomalous residuals, by instructing the fitter program to calculate orbits on the assumption that no belt was present (i.e., $\mathcal{M}_d = 0$). Figure 8 shows χ^2/ν from these fits as a function of the belt parameter $\mathcal{M}_d/r_d^3$. At the 99% confidence level the belt is detectable by anomalous residuals if

$$\frac{\mathcal{M}_d}{\mathcal{M}_\oplus} \gtrsim 420 \left(\frac{\sigma_0}{0.1''} \right) \left(\frac{r_d}{100 \text{ AU}} \right)^3. \quad (18)$$

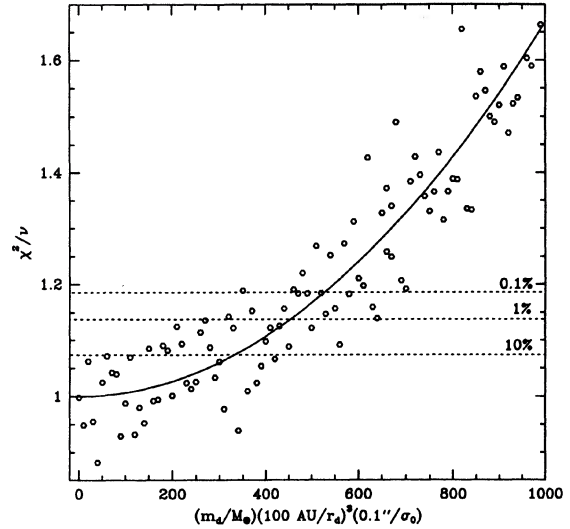


FIG. 8. χ^2 per degree of freedom found by using 80 yr of observations of the four giant planets to fit the orbital elements at the starting time. The observations were influenced by a disk of mass $\mathcal{M}_d$ at a distance r_d ; the fit was done assuming $\mathcal{M}_d = 0$ (see Fig. 3).

To determine whether the belt could be detected by modeling, we computed $\Delta\chi^2$, the difference between the minimum value of χ^2 with $\mathcal{M}_d = 0$ and with the best-fit value of $\mathcal{M}_d/r_d^3$, and fit the results to a parabola

$$\Delta\chi^2 = 1 + \alpha_3 \left(\frac{\mathcal{M}_d}{\sigma_c r_d^3} \right)^2, \quad (19)$$

where the numerical constant on the right is unity because if $\mathcal{M}_d = 0$ the $\Delta\chi^2$ values should be distributed as a chi-squared distribution with one degree of freedom. At the 99% confidence level the belt is detectable by modeling if

$$\frac{\mathcal{M}_d}{\mathcal{M}_\oplus} \gtrsim 120 \left(\frac{\sigma_0}{0.1''} \right) \left(\frac{r_d}{100 \text{ AU}} \right)^3. \quad (20)$$

These limits are comparable with Whipple's (1964) prediction of a disk of mass $10\text{--}20 \mathcal{M}_\oplus$ at $40\text{--}50$ AU to explain the latitude residuals of Neptune, which gives $\mathcal{M}_d/\mathcal{M}_\oplus (100 \text{ AU}/r_d)^3 \simeq 80\text{--}310$. The limits on the belt mass are less sensitive than the corresponding limits on the mass of Planet X because of the absence of nonaxisymmetric perturbations.

5.2 Limits from Halley's Comet

Long-period comets are potentially more sensitive than the outer planets to the effects of a belt of material in the outer solar system, both because their aphelia are larger so that they sample more distant regions, and because their orbits are highly eccentric and inclined so that precession induced by non-Keplerian forces is easier to detect. Hamid *et al.* (1968) studied the secular perturbations of seven comets by a belt outside Neptune and estimated the maximum mass of such a belt to be $0.5 \mathcal{M}_\oplus$ if it were located at 40 AU [$\mathcal{M}_d/r_d^3 = 8 \mathcal{M}_\oplus / (100 \text{ AU})^3$]. The most stringent limit was set by Halley's comet. A possible concern about this technique is that determining accurate orbits for comets with small perihelia requires empirical corrections for non-

gravitational forces, which might mask the gravitational forces from a belt.

To illustrate the limits on the mass of the Kuiper belt that are set by the motion of comets, we followed the motion of Halley's comet in the presence of the belt potential (17). We did not include perturbations of the comet's orbit by the planets, that is, the comet moved in the potential field of the belt and the Sun only. The justification for this approximation is that effects on the comet due to interactions between the belt and the planets should be much smaller than the direct effects of the belt. We measured the geocentric⁴ latitude and longitude of Halley's comet at 8.5 day intervals, from 9 months before to 1 yr after perihelion, during its two most recent perihelion passages (1910 and 1986). The orbital elements for Halley's comet were taken from Yeomans (1986). We added to the exact latitudes and longitudes Gaussian errors, with dispersion $\sigma_c/\cos(\beta)$ for the longitudes and σ_c for the latitudes.

We first investigated whether the belt could be detected by anomalous residuals. The simulated observations were used by the fitter program to determine the best-fit orbital elements of Halley's comet at the starting time assuming that no belt was present (i.e., $\mathcal{M}_d = 0$). The fitter program also fit for the best values of two parameters describing nongravitational forces, since these are usually fit when determining orbits from real data. The nongravitational force was assumed to have a radial component $A_1 g(r)$, a tangential component (normal to the radial direction, in the orbital plane and in the direction of the comet's motion) $A_2 g(r)$, and a normal component (in the direction of the comet's angular momentum) $A_3 g(r)$, where (Marsden *et al.* 1973)

$$g(r) = [\alpha(r_0/r)^m]/[1 + (r/r_0)^n]^k, \quad (21)$$

with $\alpha = 0.1126$, $m = 2.15$, $n = 5.093$, $k = 4.6142$, and $r_0 = 2.808$ AU. It is customary to fit for the parameters A_1 and A_2 and to set $A_3 = 0$, because there are no comets that show evidence of a nongravitational force in the normal direction (see, e.g., Marsden 1985).

Figure 9 shows χ^2/ν as a function of the parameter $\mathcal{M}_d/r_d^3$. For a typical realization of the observational errors, at the 99% confidence level we can detect a belt mass given by

$$\frac{\mathcal{M}_d}{\mathcal{M}_\oplus} \gtrsim 3 \left(\frac{\sigma_c}{0.2''} \right) \left(\frac{r_d}{100 \text{ AU}} \right)^3. \quad (22)$$

The appropriate value of σ_c can be estimated only crudely. Yeomans (1986) quotes 5254 observations with rms error 1.0" at the 1986 perihelion passage; assuming random errors the corresponding uncertainty in the 75 observations in our model would be $1.0''(75/5254)^{1/2} = 0.12''$. A similar estimate for the 1911 passage yields an error of 0.28". Thus $\sigma_c = 0.2''$ is a plausible estimate if systematic errors are negligible.

To determine whether the belt could be detected by modeling, we computed $\Delta\chi^2$, the difference between the minimum value of χ^2 with $\mathcal{M}_d = 0$ and with the best-fit value of $\mathcal{M}_d/r_d^3$, and fit the results to the parabola (19). At the 99%

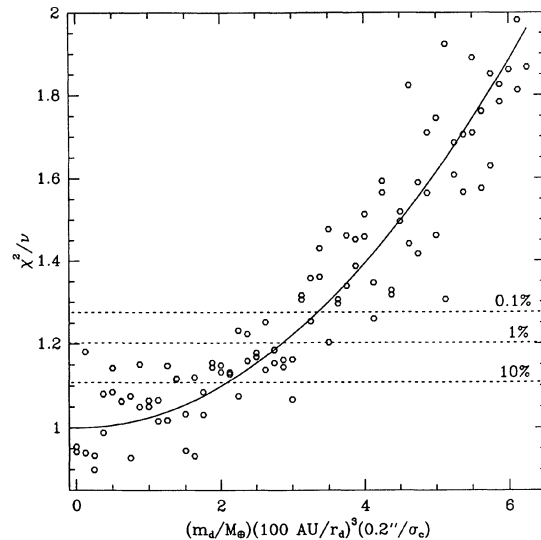


FIG. 9. χ^2 per degree of freedom found by fitting the starting orbital elements of Halley's comet using observations from two perihelion passages. The observations were influenced by a disk of mass $\mathcal{M}_d$ at a distance r_d ; the fit was done assuming $\mathcal{M}_d = 0$ (see Fig. 3).

confidence level the belt is detectable by modeling if

$$\frac{\mathcal{M}_d}{\mathcal{M}_\oplus} \gtrsim 1 \left(\frac{\sigma_c}{0.2''} \right) \left(\frac{r_d}{100 \text{ AU}} \right)^3. \quad (23)$$

The mass limit (23) is about 8 times smaller than the limit found by Hamid *et al.* (1968) if $\sigma_c = 0.2''$. Part of the reason is that their limit was based on observations from the 1835 and 1911 perihelion passages, which leads to much larger errors than the 1911–1986 pair of perihelion passages. The Hamid *et al.* result was based on the difference between the predicted (from 1835) and observed (in 1911) longitude of the ascending node, $\Delta\Omega = 0.0014^\circ$; the analogous difference between the 1911 and 1986 perihelion passages was $\Delta\Omega = -0.00067^\circ$, about a factor of 2 smaller (Yeomans 1986), which suggests that the present limit on the mass of the Kuiper belt is at least a factor of 2 lower than the limit given by Hamid *et al.*

The nongravitational parameters A_1 and A_2 are not effective in masking the gravitational force from the belt. This is understandable, because these parameters describe forces in the plane of the cometary orbit, whereas the belt exerts a force normal to the orbital plane that leads to precession of the ascending node. The nongravitational parameter A_3 , which we have assumed to be zero in these calculations, does lead to precession of the ascending node. We tried some fits in which we allowed $A_3 \neq 0$ and found that a larger belt could indeed be hidden, but only if A_3 is allowed to be larger than A_1 and A_2 , which is inconsistent with the usual physical models for the origin of the nongravitational forces.

5.3 Other Limits

The constraint (16) on the mass of Planet X from the semimajor axes of new comets also applies to the Kuiper belt. Thus

$$\frac{\mathcal{M}_d}{\mathcal{M}_\oplus} \leq 820 \left(\frac{r_d}{100 \text{ AU}} \right). \quad (24)$$

⁴To measure geocentric coordinates we had to follow the motion of the Earth in the simulator as well; the heliocentric observer that we used for studying the outer planets is inappropriate in this case because of the small perihelion distance of Halley's comet.

Limits on the belt mass can also be obtained from tracking interplanetary spacecraft. Boss & Peale (1976) stated that precision tracking of deep space probes is about three orders of magnitude less sensitive to the gravitational acceleration of a belt in the outer solar system than are the secular perturbations of comets. This conclusion was derived by assuming $10^{-6} \text{ cm s}^{-2}$ to be a lower bound to the detectable spacecraft acceleration. Anderson & Standish (1986) quote a much smaller lower bound, $5 \times 10^{-9} \text{ cm s}^{-2}$, and use this to put a limit of $5 \mathcal{M}_{\oplus}$ on the mass of a belt just outside the orbit of Neptune. This is similar to the limit derived by Hamid *et al.* (1968) from the secular perturbations of Halley's comet.

6. OTHER ANOMALIES

In this section we briefly evaluate some of the other evidence for anomalies in the orbital residuals of Uranus and Neptune that might be explained by forces due to dark mass in the outer solar system (see Seidelmann & Harrington 1988 for a review).

There are several predisccovery observations of Neptune. Lalande observed Neptune twice in 1795. His observations differed from the ephemeris prediction in right ascension by about 7–12" (Rawlins 1970; Seidelmann & Harrington 1988). The exact difference depends on which ephemeris is used in the comparison, and is not much larger than the uncertainty in the ephemerides. For example, if the recent JPL ephemerides DE200 and DE202 are integrated backwards to 1795 they give predictions for Neptune's right ascension that differ by 6" (Standish 1991).

Galileo observed Neptune several times in 1612–1613 (Kowal & Drake 1980), although only one of these observations is of astrometric value. Galileo's position differs from the ephemeris (DE102) position by 63". It is possible, however, that the ephemeris position is correct and that Galileo's drawing is not to scale; evidence supporting this point of view is that the direction from the reference star to Neptune in Galileo's drawing is the same as the direction predicted by the ephemeris (see also Taylor 1985). It is rather unlikely that the discrepancy between Galileo's position and the ephemeris could be explained by a perturbation from an undiscovered planet (see, e.g., Rawlins 1981; Standish 1981). Standish (1990b) has discovered another observation of Neptune by Galileo in January 1613, which may also be of astrometric value. This observation appears to agree with the predictions of present-day ephemerides.

Additional possible evidence for anomalies is that the Uranus observational data cannot all be fit with a single ephemeris (Seidelmann *et al.* 1985). Modern observations can be fit successfully only if the observational data from before 1900 are discarded. This difficulty is not evidence for unexplained orbital motions, however, if the pre-1900 observations were subject to some systematic error that has not yet been deduced from the available historical evidence.

Gomes & Ferraz-Mello (1987, 1988) claim that the orbital residuals of Uranus and Neptune contain significant periodicities. The strongest claimed periodicities are in the longitudes and have periods of 78 yr (Neptune) and 50.5 yr (Uranus). These periods are not sharply defined by the analysis, and it is possible that the 50.5 yr period for Uranus corresponds to the synodic period of 45.4 yr with respect to Saturn. The 78 yr period in the Neptune residuals may correspond to half the period of Neptune ($\frac{1}{2} \times 164.8 \text{ yr} = 82.4 \text{ yr}$), which should be the period of the strongest perturbation from a distant planet. However, the claimed periodicities

may be due to time-dependent systematic errors in the observations, since the estimate of their statistical significance is based on the assumption of random errors.

Clearly, none of these phenomena provides strong evidence for unexplained orbital behavior in the outer planets.

7. SUMMARY

The study of simplified model solar systems with known observational errors offers many insights into the search for dark matter in the real solar system. Most of the conclusions described below are based on experiments in which the four giant planets are observed once per year during the period 1910–1990, and the observational errors in latitude and longitude are independent Gaussian random variables with standard deviation σ_0 . The conclusions should apply equally well to observations made n times per year with standard deviation $n^{1/2}\sigma_0$. Our results are expressed in terms of the ratio of σ_0 to an assumed value of 0.1", although this is almost certainly smaller than the effective errors in the real data.

(i) If no dark mass is present, the ability of the best-fit ephemeris to predict latitudes over a 20 yr period is about the same for all the giant planets. The rms error in longitude predictions, however, is about four times worse for Neptune than for Jupiter (see Table 1). Thus the failure of recent ephemerides to predict the longitude of Neptune accurately over the interval 1980–1990 does not provide evidence that dynamical anomalies are present in the outer solar system.

(ii) The minimum detectable mass $\mathcal{M}_X$ of a Planet X in the ecliptic at distance $r_X \gg 30 \text{ AU}$ —in the sense that models that do not include Planet X yield a χ^2 that is too large at the 99% confidence level in a typical realization—is given by [Eq. (5) and Fig. 5]

$$\frac{\mathcal{M}_X}{\mathcal{M}_{\oplus}} \gtrsim (20-30) \left(\frac{\sigma_0}{0.1''} \right) \left(\frac{r_X}{100 \text{ AU}} \right)^3. \quad (25)$$

The minimum detectable mass, in the sense that significantly better fits are obtained with Planet X than without, is given by [Eq. (6) and Fig. 5]

$$\frac{\mathcal{M}_X}{\mathcal{M}_{\oplus}} \gtrsim (6-10) \left(\frac{\sigma_0}{0.1''} \right) \left(\frac{r_X}{100 \text{ AU}} \right)^3. \quad (26)$$

The detectability limit is somewhat lower if the planet is at ecliptic latitudes in the range 30°–60° (Fig. 5) but is higher by about a factor of 7 if the planet is at the ecliptic pole.

(iii) The minimum detectable mass in Eq. (26) is smaller by a factor of 3 than the limit in Eq. (25); this ratio increases if the time span of data is longer than our assumed value of 80 yr. Thus the absence of anomalous residuals in a fit to a solar system model that does not include Planet X does not preclude the possibility that more sophisticated analysis will produce strong evidence that Planet X is present.

(iv) In order to predict the location of Planet X within about 1°, the same accuracy with which Le Verrier predicted the location of Neptune, the mass of the planet must be at least ten times larger than the limit in Eq. (26).

(v) The minimum detectable mass $\mathcal{M}_d$ of a disk or belt of dark material (Kuiper belt) in the ecliptic at distance $r_d \gg 30 \text{ AU}$ —in the sense that the orbital residuals of the giant planets in models without the belt yield too large a value of χ^2 —is given by [Eq. (18)]

$$\frac{\mathcal{M}_d}{\mathcal{M}_{\oplus}} \gtrsim 420 \left(\frac{\sigma_0}{0.1''} \right) \left(\frac{r_d}{100 \text{ AU}} \right)^3. \quad (27)$$

The minimum detectable mass, in the sense that significantly better fits are obtained with the belt than without, is given by [Eq. (20)]

$$\frac{\mathcal{M}_d}{\mathcal{M}_\oplus} \gtrsim 120 \left(\frac{\sigma_0}{0.1''} \right) \left(\frac{r_d}{100 \text{ AU}} \right)^3. \quad (28)$$

(vi) Limits on the mass of a distant disk from residuals in the orbit of Halley's comet are much more stringent than the limits from the orbits of the outer planets. We have assumed that observations of Halley's comet are made at 8.5 day intervals from 9 months before perihelion to 1 yr after perihelion, with rms errors σ_c in latitude and $\sigma_c/\cos(\beta)$ in longitude. Using data from the most recent two perihelion passages only, and allowing for the effects of nongravitational forces (but with $A_3 = 0$), we find that the minimum detectable mass $\mathcal{M}_d$ of a disk at radius $r_d \gg 30$ AU, in the sense that fits without the belt give too large a value of χ^2 , is given by [Eq. (22)]

$$\frac{\mathcal{M}_d}{\mathcal{M}_\oplus} \gtrsim 3 \left(\frac{\sigma_c}{0.2''} \right) \left(\frac{r_d}{100 \text{ AU}} \right)^3. \quad (29)$$

The minimum detectable mass, in the sense that significantly better fits are obtained with the belt than without, is given by [Eq. (23)]

$$\frac{\mathcal{M}_d}{\mathcal{M}_\oplus} \gtrsim 1 \left(\frac{\sigma_c}{0.2''} \right) \left(\frac{r_d}{100 \text{ AU}} \right)^3. \quad (30)$$

All of these results are based solely on numerical experiments. The great uncertainty in translating the results to realistic limits on dark mass in the solar system is the unknown effect of systematic errors in the observations. The crudest approach is to treat the effects of systematic errors simply by enlarging the assumed random error σ_0 . Errors in the extrapolation of planetary latitudes in the DE200 ephemeris then suggest that the appropriate value of σ_0 is $0.4''$ – $0.5''$. Time-dependent systematic errors can easily masquerade as perturbations from dark mass.

We close with some comments on the current status of the search for Planet X.

(vii) None of the anomalies in the orbital residuals of Ura-

nus and Neptune that we described in Secs. 3 and 6 provide convincing evidence for Planet X. The discrepancy in Galileo's predisccovery observation may arise because his drawing is not to scale, while the discrepancies in Lalande's predisccovery observations are not much larger than the estimated error in the ephemerides. The discrepancies in the pre-1900 Uranus data may be due to unknown systematic errors. The inability of modern ephemerides to predict Neptune's position is, as we have seen, a natural consequence of Neptune's long orbital period. The periodicities in the motion of Uranus and Neptune claimed by Gomes & Ferraz-Mello (1987, 1988) may not be significant if there are time-varying systematic errors.

(viii) Given the inevitable presence of systematic errors, no claim for Planet X will be convincing unless it is sufficiently accurate that a successful search can be mounted. We have seen that to localize the planet within 1° or so requires about ten times the mass that is needed to detect its existence at a statistically significant level. Thus it is most unlikely that the present database can be used to determine the planet's position accurately enough to motivate a thorough search.

(ix) If Planet X is a normal planet (i.e., not a black hole or other exotic object) and is massive enough to exert significant dynamical effects on the outer planets, then it is more likely than not that it would have already been detected in the *IRAS* survey.

Given the lack of any convincing dynamical anomalies that can be explained by Planet X, the null results from the *IRAS* survey, the inadequacy of the present database due to unknowable systematic errors in the historical data, and the lack of any appealing theoretical reasons to believe in Planet X, we believe that there is little motivation for continuing to explore the Planet X hypothesis until improved data on the orbits of the outer planets become available, for example from the *Galileo* or *Cassini* spacecraft or from pulsar timing.

This work was supported by the Natural Sciences and Engineering Research Council of Canada. We thank Tom Chester for patiently explaining many details of the *IRAS* survey to us.

REFERENCES

- Anderson, J. D., and Standish, E. M. 1986, in *The Galaxy and the Solar System*, edited by R. Smoluchowski, J. N. Bahcall, and M. S. Matthews (University of Arizona Press, Tucson), pp. 286–296.
- Beichman, C. A., Neugebauer, G., Habing, H. J., Clegg, P. E., and Chester, T. J., Eds. 1988, *Infrared Astronomical Satellite (IRAS) Catalogs and Atlases* (Explanatory Supplement, NASA, WA), Vol. 1.
- Boss, A. P., and Peale, S. J. 1976, *Icarus*, 27, 119.
- Chester, T. J. 1986, in *Light on Dark Matter*, edited by F. P. Israel (Reidel, Dordrecht), pp. 3–22.
- Chester, T. J. 1990, private communication.
- Cohen, C. J., Hubbard, E. C., and Oesterwinter, C. 1973, *Astron. Papers Am. Ephemeris*, 22, 1.
- Duncan, M., Quinn, T., and Tremaine, S. 1988, *ApJL*, 328, L69.
- Goldreich, P. 1965, *Rev. Geophys.*, 4, 411.
- Gomes, R. S. 1989, *Icarus*, 80, 334.
- Gomes, R. S., and Ferraz-Mello, S. 1987, *A&A*, 185, 327.
- Gomes, R. S., and Ferraz-Mello, S. 1988, *A&A*, 203, 170.
- Hamid, S. E., Marsden, B. G., and Whipple, F. L. 1968, *AJ*, 73, 727.
- Harrington, R. S. 1988, *AJ*, 96, 1476.
- Kirk, J. 1978, *Nat*, 274, 667.
- Kowal, C. T., and Drake, S. 1980, *Nat*, 287, 311.
- Kuiper, G. P. 1951, in *Astrophysics: A Topical Symposium*, edited by J. A. Hynek (McGraw-Hill, New York), pp. 357–424.
- Marsden, B. G. 1985, in *Dynamics of Comets: Their Origin and Evolution*, edited by A. Carusi and G. B. Valsecchi (Reidel, Dordrecht), pp. 343–352.
- Marsden, B. G., Sekanina, Z., and Yeomans, D. 1973, *AJ*, 78, 211.
- Newhall, X. X., Standish, E. M., and Williams, J. G. 1983, *A&A*, 125, 150.
- Oesterwinter, C., and Cohen, C. J. 1972, *Celest. Mech.*, 5, 317.
- Powell, T. C. 1988, *BAAS*, 20, 897.
- Rawlins, D. 1970, *AJ*, 75, 856.
- Rawlins, D. 1981, *Nat*, 290, 164.
- Rawlins, D., and Hammerton, M. 1973, *MNRAS*, 162, 261.
- Seidemann, P. K. 1971, *AJ*, 76, 740.
- Seidemann, P. K., and Harrington, R. S. 1988, *Celest. Mech.*, 43, 55.
- Seidemann, P. K., Santoro, E. J., and Pulkkinen, K. F. 1985, in *Dynamical Astronomy*, edited by V. G. Szebehely and B. Balázs (University of Texas Press, Austin), pp. 55–78.
- Standish, E. M. 1981, *Nat*, 290, 164.
- Standish, E. M. 1986, in *Relativity in Celestial Mechanics and Astrometry*, IAU Symposium No. 114, edited by J. Kovalevsky and V. A. Brumberg (Reidel, Dordrecht) pp. 71–83.
- Standish, E. M. 1990a, *A&A*, 233, 272.
- Standish, E. M. 1990b, unpublished talk given at the 21st meeting of the

AAS Division on Dynamical Astronomy, Austin, Texas

Standish, E. M. 1991, private communication

Taylor, G. E. 1985, JBAA, 95, 116

Tombaugh, C. W. 1961, in Planets and Satellites, edited by G. P. Kuiper and B. Middlehurst (University of Chicago Press, Chicago), pp. 12–30

Tremaine, S. 1990, in Baryonic Dark Matter, edited by D. Lynden-Bell and

G. Gilmore (Kluwer, Dordrecht), pp. 37–65

Whipple, F. L. 1964, AJ, 69, 563

Yeomans, D. K. 1986, in Proceedings of the 20th ESLAB Symposium on the Exploration of Halley's Comet, edited by B. Battrock, E. J. Rolfe, and R. Reinhard (ESA, Heidelberg), Vol. 2, pp. 419–425