Asteroid Close Approaches: Analysis and Potential Impact Detection

Andrea Milani
Università di Pisa

Steven R. Chesley
Jet Propulsion Laboratory

Paul W. Chodas
Jet Propulsion Laboratory

Giovanni B. Valsecchi
Istituto di Astrofisica Spaziale

Recently, several new tools and techniques have been developed to allow for robust detection and prediction of planetary encounters and potential impacts by near-Earth asteroids (NEAs). We review the recent history of impact-prediction theory and cover the classical linear techniques for analyzing encounters, which consists of precise orbit determination and propagation followed by target-plane analysis. When the linear approximation is unreliable, there are various suitable approaches for detecting and analyzing very low-probability encounters dominated by strongly nonlinear dynamics. We also describe an analytic approach that can provide valuable insight into the mechanisms responsible for most encounters. This theory is the foundation of impact monitoring systems, including those currently operational and those being developed.

1. INTRODUCTION

The last 10 years have seen tremendous progress in our ability to assess the risk of an asteroid or comet colliding with Earth. The catalyst for much of the increased interest in these near-Earth objects (NEOs) was a request by the U.S. Congress in 1990 that NASA undertake two workshop studies, one to study ways of increasing the discovery rate of these objects and another to study the technologies and options for deflecting or destroying an NEO should it be found to pose a danger to life on Earth. The report from the first of these workshops proposed an international NEO survey program called Spaceguard, borrowing the name from a similar project in Arthur C. Clarke's science fiction novel *Rendezvous with Rama* (*Morrison,* 1992). The report noted the need for development of new procedures and software to "assess the uncertainty [in Earth-object distance] for any future close approaches."

The exploding state of knowledge of the impact hazard problem in the early 1990s was well captured by an earlier monograph in this series, *Hazards Due to Comets and Asteroids* (*Gehrels et al.,* 1994). In that volume, *Bowell and Muinonen* (1994) suggest the use of the minimum orbital intersection distance (MOID) for close encounter analyses, and define the class of potentially hazardous asteroids (PHAs) as those asteroids having a MOID with respect to Earth of <0.05 AU. In the same volume, *Chodas and Yeomans* (1994) describe a system for predicting future asteroid and comet close approaches, including the computation of close-approach uncertainties and impact probabilities via linear methods. The impact probability computation via linear methods in the impact plane was introduced in the context of asteroid and comet collisions by *Chodas* (1993). The method saw an immediate application when Comet Shoemaker-Levy 9 was found to be on a collision course with Jupiter. Within a few days of the impact announcement, the impact probability was calculated to be 64% (*Yeomans and Chodas,* 1993), and it reached 95% only a week later. Somewhat different linear methods for computing collision probabilities by monitoring the distance between Earth and the asteroid uncertainty ellipsoid are described by *Muinonen and Bowell* (1993). *Chodas and Yeomans* (1996) performed an early nonlinear analysis of orbital uncertainties in an investigation of the prebreakup orbital history of Comet Shoemaker-Levy 9.

In March 1998, the problem of computing impact probabilities received a great amount of press attention because of a prediction by Brian Marsden of the Smithsonian Astrophysical Observatory that the sizable asteroid 1997 XF_{11} would make an extremely close approach to Earth in 2028, coupled with a widely misunderstood statement that a collision was "not entirely out of the question" (*Marsden,* 1999). A linear analysis of the impact probability was immediately performed by Paul W. Chodas and Donald K. Yeomans of the Jet Propulsion Laboratory (JPL), and the chance of collision in 2028 was found to be essentially zero.

When prediscovery observations from 1990 were found the next day and included in the orbital solution, they only served to confirm that there was no possibility of collision in 2028 (*Chodas and Yeomans,* 1999b; *Muinonen,* 1999).

Three months after the 1997 XF_{11} story hit the news, Marsden opened a new area of investigation by suggesting that prior to the discovery of the 1990 observations, 1997 XF_{11} had in fact a small possibility of collision in the decade or so after 2028 because the deep close approach in 2028 could alter the asteroid's orbital period to bring it back to Earth several orbits later. Linear methods were inadequate to analyze these later collision possibilities because the 2028 close approach introduced a strong nonlinearity into 1997 XF_{11}'s predicted motion. Primarily to investigate the post-2028 impact scenarios of 1997 XF_{11}, two groups independently and simultaneously developed new nonlinear methods for analysis of collision possibilities much further into the future than had been possible with the linear methods. *Chodas and Yeomans* (1999a) applied a Monte Carlo technique to sample the linear six-dimensional confidence region at the epoch of observation and then numerically integrate over the timespan of interest using the nonlinear equations. *Milani et al.* (1999, 2000b) applied the multiple-solutions approach to sample the central line of variations of the nonlinear confidence region at epoch and then numerically integrate over the timespan of interest in a similar fashion. The two methods yielded very similar results when applied to the hypothetical 1997 XF_{11} case without the prediscovery observations, with each obtaining an impact probability on the order of 10^{-5} for the year 2040 (*Chodas and Yeomans,* 1999c; *Milani et al.,* 2000b).

In early 1999, Milani et al. applied these same techniques to the case of asteroid 1999 AN_{10} while developing the theory of resonant and nonresonant returns. They also identified for the first time an impacting solution consistent with a complete set of NEA observations, although the collision probability was very small (*Milani et al.,* 1999). This result was confirmed by independent investigators. As more observations of 1999 AN_{10} became available, both the Pisa and the JPL group identified two much more likely impacting scenarios for the years 2044 and 2046, with probabilities on the order of 10^{-6}, the highest seen to that date. Observations later found on archival plates drove these probabilities to essentially zero.

In mid-1999, a second asteroid, 1998 OX_4, was determined (based on all available observations) to have some collision potential (*Milani et al.,* 2000a). Unfortunately this asteroid had become lost, because of its faintness and its location in the Milky Way, despite concerted efforts of numerous observers around the time of discovery, and there was no practical way to obtain additional observations to refine the collision hazard posed by this object. *Milani et al.* (2000a) proposed a method, based on the concept of a negative observation, for ruling out the known collision possibilities without requiring the recovery of 1998 OX_4; their approach has been successfully implemented (Andrea Boattini, personal communication, 2001). In their reassessment of the case, *Muinonen et al.* (2001) found additional possibilities of collision, generally confirming the results of *Milani et al.* (2000a). However, this case highlighted the need for *automatic collision monitoring,* since it is highly likely that, had the collision potential been recognized at the time of the discovery, the resources needed to observe the newly discovered PHA and refine its orbit would have been forthcoming. In response, the CLOMON monitoring system (*Chesley and Milani,* 2000) was developed at the University of Pisa, and it has been used to detect numerous impact possibilities among recently discovered asteroids. Importantly, all these threatening events (apart from some involving very small bodies) were eventually eliminated as a result of additional observations, which in some cases came as a direct result of CLOMON impact detection. CLOMON was, however, a prototype system with an uncertain level of completeness. At the time of this writing, more advanced monitoring systems, based on principles described in this chapter, are under independent development at the University of Pisa and at JPL.

This chapter is organized to give an overview of the ideas and theories needed to reliably detect and analyze potentially threatening Earth-asteroid encounters. The next section describes the most important tool used in encounter analysis, the target plane. In section 3, we describe the linear theory of orbit determination and encounter analysis and also discuss the limitations of the theory. In sections 4 and 5 we describe the various nonlinear sampling and analysis methods available. Section 6 presents an analytic approach that can be used to understand the chaotic nature of the problem in terms of resonant returns and keyholes. Finally, section 7 discusses a few of the more significant open problems that remain.

2. PLANETARY ENCOUNTERS

2.1. Target Planes

The primary method of studying planetary encounters, whether the encounters are with spacecraft or minor planets, continues to rely on the concept of a target plane. Here we use this term generically to encompass any of the various planes and coordinate systems that may be used to study a specific encounter of an asteroid with Earth. (For simplicity we limit this development to asteroid encounters with Earth. The extension to comet encounters, where nongravitational forces must be considered, or to encounters with other planets is straightforward.) In this sense, a target plane is simply a geocentric plane oriented normal to the asteroid's geocentric velocity vector. The point of intersection of the asteroid trajectory with the target plane provides considerable insight into the nature of an encounter, especially when uncertainty is carefully taken into account.

While the fundamental concept is rather straightforward, there are several issues we must address that can unfortunately obscure this simplicity. In general, either of two distinct planes and several coordinate systems can be used. The two available target planes pass under various names through the published literature, which is occasionally con-

fusing, but for the purposes of this development we set forth the following usage:

Target plane or b-plane. The b-plane is the classical target plane that has been used in astrodynamics since the 1960s (e.g., *Kizner,* 1959) and by *Greenberg et al.* (1988) in the framework of Öpik's theory of close encounters (*Öpik,* 1976). It is oriented normal to the incoming asymptote of the osculating geocentric hyperbola, or, equivalently, it is oriented normal to the unperturbed relative velocity $\mathbf{v}_\infty$. The b-plane is named in reference to the so-called impact parameter b, which is the distance from the geocenter to the intercept of the asymptote on this plane, i.e., the minimum encounter distance along the unperturbed trajectory.

Modified target plane (MTP). The MTP is modified in the sense that it is oriented normal to the geocentric velocity at the point of closest approach along the actual trajectory. Rather than noting the position of the intersection of the unperturbed orbit, as is done with the b-plane, we mark the intersection of the perturbed asteroid trajectory with the target plane when using the MTP. The term "MTP" was introduced by *Milani and Valsecchi* (1999), although the concept was introduced earlier (e.g., *Chodas and Yeomans,* 1994).

The main difference between the two types of target planes arises from the fact that the deflection of the asteroid by the planet or gravitational focusing is directly indicated with the use of the MTP, while the gravitational focusing is in some sense hidden when the b-plane is used. For encounters taking place at high relative velocities, or at large distances, the deflection will be very small and the distinction between the two planes becomes negligible.

One difficulty that arises from the visibility of gravitational focusing on the MTP is that nearby trajectories are deflected by different amounts. When the deflection is significant, as is typically the case for very deep encounters or those with low relative velocities, this leads to substantial nonlinearity in the mapping from the asteroid's preencounter state onto the impact plane. Moreover, the orientation of the plane itself can change significantly with modest changes to the asteroid's trajectory, and this poses another important source of nonlinearity. Of course, for very slow encounters that lead to temporary capture of the asteroid, an asymptote does not exist, and therefore the b-plane approach fails; in such cases the MTP must be used.

This allows us to rate the utility of the b-plane and MTP according to the deflection caused by the encounter. For very low deflections, the two approaches are essentially indistinguishable, while for moderate deflections the preferred plane depends on the circumstances and purposes at hand. For relatively large deflections, the greater nonlinearity of the MTP causes the b-plane to be generally superior, but for very low relative velocities (leading to capture in elliptical orbits) only the MTP is available.

2.2. Target Plane Coordinates

Conventionally, the coordinate origin is at the geocenter, but the orientation of the coordinate system on the target plane is arbitrary. The reference system has been variously fixed by aligning the axes so that one of the nominal target plane coordinates is zero (i.e., the asteroid is placed on an axis) or by aligning a coordinate axis with either the projection of Earth's polar axis (e.g., Figs. 1 and 2) or the projection of Earth's heliocentric velocity (e.g., Fig. 3).

One of the most important objectives in a target-plane analysis is to determine whether a collision is possible and, if not, to decide how deep the encounter will be. When using the MTP we plot the position of the asteroid at the point of closest approach; hence this information is immediately available. On the other hand, with the b-plane we obtain the minimum distance of the unperturbed trajectory at its closest approach point, which is the impact parameter b. The impact parameter alone does not reveal whether the perturbed trajectory will intersect the figure of Earth, but this information can be extracted by scaling Earth's radius $r_\oplus$ according to

$$b_\oplus = r_\oplus \sqrt{1 + \frac{v_e^2}{v_\infty^2}}$$

where v_e is Earth's escape velocity

$$v_e^2 = \frac{2GM_\oplus}{r_\oplus}$$

With this we can say that a given trajectory impacts if $b < b_\oplus$, and not otherwise. An alternative to simply scaling the figure of Earth on the b-plane is to scale b while leaving the figure of Earth at $r_\oplus$. For any single orbit this is obviously equivalent, but when computing the coordinates of slightly different asteroids, each with a slightly different $\mathbf{v}_\infty$, the scaling is not uniform (*Chodas and Yeomans,* 1999c).

A convenient target-plane reference system (ξ, η, ζ) is obtained by aligning the negative ζ-axis with the projection of Earth's heliocentric velocity $\mathbf{v}_\oplus$, the positive η-axis with the geocentric velocity (i.e., normal to the b-plane), and the positive ξ-axis in such a way that the reference system is positively oriented. With this frame of reference, the target-plane coordinates (ξ, ζ) indicate the cross-track and along-track miss distances, respectively. In other words, ζ is the distance by which the asteroid is early or late for the minimum possible distance encounter. The associated early/late timing of the target-plane crossing ($\eta = 0$) is $\Delta t = \zeta/(v_\oplus \sin\theta)$, where θ is the angle between $\mathbf{v}_\infty$ and $\mathbf{v}_\oplus$. On the b-plane the ξ coordinate is the minimum distance that can be obtained by varying the timing of the encounter. This distance, known as the minimum orbital intersection distance (MOID), is equivalent to the minimum separation between the osculating ellipses, without regard to the location of the objects on their orbits. We note that this interpretation of the target-plane coordinates is valid only in the linear approximation and can break down for distant encounters (e.g., beyond several lunar distances).

Whatever target plane or coordinate frame is used, the idea that an asteroid can avoid an impact, either through being off time or by having an orbit that does not even intersect Earth, is important. To have an impact the object must

have a small enough MOID and be on time for the collision. With this in mind, we can characterize an encounter quite well given only the MOID and Δt. The MOID, which is strictly a function of the osculating elements of the asteroid and Earth, can be computed in various ways, with the numerical approach of *Sitarski* (1968) being the most widely used. A recently developed semianalytic approach (*Gronchi,* 2001) can be more robust, with better handling of the cases of multiple local minima of the distance between the two orbits (so-called local MOIDs).

3. LINEAR ENCOUNTER THEORY

3.1. Least-Squares Orbit Determination

The details of the orbit determination process are not within the scope of this presentation; readers should consult *Bowell et al.* (2002). However, we do need to touch on a few of the key ideas of estimation theory to establish our notation and terminology. First, we describe the initial conditions at some epoch t_0 for the asteroid under consideration as the six-dimensional vector X of orbital elements. The observation residuals $\rho = \rho(X)$ are a function of the six orbital elements, and they form an m-dimensional vector where $m > 6$ is the number of (scalar) observations.

The weighted least-squares method of orbit determination seeks to minimize the weighted RMS of the residuals ρ, so we define the cost function as

$$Q = \frac{1}{m}\rho^T W \rho$$

where W is a square, symmetric (though not necessarily diagonal) matrix that should reflect the *a priori* RMS and correlations of the observation errors. We denote the dependence of the residuals on the elements by

$$B = \frac{\partial \rho}{\partial X}(X)$$

where B is an $m \times 6$ matrix. Then we can compute the derivative of the cost function as

$$\frac{\partial Q}{\partial X} = \frac{2}{m}\rho^T W B$$

The stationary points of the cost function Q are solutions of the system of nonlinear equations $\partial Q/\partial X = 0$, which are usually solved by some iterative procedure. The most popular is a variant of Newton's method, known in this context as differential corrections, with each iteration making the correction $X - \Delta X \rightarrow X$, where

$$\Delta X = (B^T W B)^{-1} B^T W \rho$$

This converges (normally) to the best-fitting or nominal solution X^*, where $\Delta X = 0$. We use the usual terminology for the normal matrix $C_X = B^T W B$, and covariance matrix $\Gamma_X = C_X^{-1}$. We note that these matrixes can be computed not only for the nominal solution but also for nearby values of X, a point that will be put to use in sections 4.3 and 5.3 in handling nonlinear situations.

3.2. Orbital Confidence Region

The expansion of the cost function at a point $X = X^* + \Delta X$ in a neighborhood of X^* is

$$Q(X) = Q(X^*) + \frac{1}{m}\Delta X^T C_X \Delta X + \ldots = Q(X^*) + \Delta Q(X)$$

where the dots indicate higher-order terms in ΔX plus a term containing the second derivative of ρ (see *Milani,* 1999). A confidence region is a region where the solutions are not too far from the nominal, as measured by the penalty ΔQ. We shall indicate with $Z_X(\sigma)$ the confidence ellipsoid defined by the inequality

$$m\Delta Q(X) \simeq \Delta X^T C_X \Delta X \le \sigma^2$$

If the confidence ellipsoid $Z_X(\sigma)$, defined by the normal matrix C_X, is small enough (either because the constraint is sufficiently strong, i.e., the eigenvalues of C_X are large, or because the choice of σ is sufficiently small) then the higher-order terms in ΔQ(X) are also small and $Z_X(\sigma)$ is a good approximation of the region in the space of orbital elements where the penalty $\Delta Q < \sigma^2/m$. In such cases we say that the linear theory of estimation holds; however, when the confidence ellipsoid is larger, it can be an unreliable representation of the region containing alternative solutions still compatible with the observations.

The confidence region is generally viewed as a cloud of possible orbits centered on the nominal solution, where density is greatest, and surrounded by a diffuse boundary. This is classically represented by the multivariate Gaussian probability density, which reflects the likelihood that a given volume of the confidence region will contain the true solution; this Gaussian distribution has X^* as the mean and Γ_X as the matrix containing the variances and the covariances of the distributions of the elements (hence its name). But the use of the Gaussian probability density to describe orbital uncertainty assumes that the observational errors are Gaussian, and that W accurately reflects the observational uncertainty. Indeed, the extent of the confidence region is directly dependent on the choice of W, and therefore proper observation weighting is crucial for reliably determining the orbital confidence region. However, given the fact that the error statistics for asteroid astrometry are rarely characterized, and moreover are often not even Gaussian, it should be clear that the selection of W is very problematic. This implies that a careful probabilistic interpretation of the orbital confidence region is elusive [see *Carpino et al.* (2001) for an expanded discussion of this issue].

3.3. Linear Mapping to the Target Plane

Ultimately we wish to use the orbital confidence region to infer how close an asteroid could pass during a given close approach and, if collision is possible, to determine the probability of collision. To do this we must map the orbital uncertainty onto the target plane.

The nonlinear function F maps a given orbit X to a point Y on the target plane: $Y = F(X)$. This mapping consists of a propagation from an epoch near the observations t_0 to the time of the encounter t_1, followed by a projection onto the target plane. The Jacobian DF linearly maps orbits near the nominal to nearby points on the target plane

$$\Delta Y = DF(X^*)\,\Delta X$$

The matrix DF is practically computed by propagating numerically the orbit together with the variational equations (providing the state transition matrix), then projecting on the cross section defined by the target plane. In the linear approximation — i.e., when the confidence region is small and F is not too nonlinear — the confidence ellipsoid $Z_X(\sigma)$ in the space of orbital elements maps onto the target plane as an ellipse $Z_Y(\sigma)$ defined by

$$\Delta Y^T\, C_Y\, \Delta Y \le \sigma^2$$

where C_Y is the normal matrix describing the uncertainty on the target plane. As is well known from the theory of Gaussian probability distributions (*Jazwinski,* 1970), the relationship between the covariance matrices of the variables X and Y can be shown by

$$C_Y^{-1} = \Gamma_Y = DF\ \Gamma_X\ DF^T$$

3.4. Linear Target Plane Analysis

Continuing under the assumption of linearity, we can characterize the target-plane ellipse Z_Y given only the nominal target-plane coordinates Y, which marks the center of the ellipse, and the associated covariance matrix Γ_Y, which indicates the size and orientation of the ellipse. The square roots of the eigenvalues of Γ_Y are the semimajor and semiminor axes of Z_Y, and the corresponding eigenvectors indicate the orientation of the associated ellipse axes.

At this point we define several terms and variables associated with the target-plane analysis that are important for interpretation. First, d denotes the distance from the asteroid to the geocenter on the target plane, i.e., $d = \|Y\|$ (on the b-plane we have d = b). Also, α denotes the angle between Y and the major axis of Z_Y. The minimum distance from the major axis to the geocenter is $d \sin\alpha$. The semimajor and semiminor axes of Z_Y are termed the stretching Λ and the semiwidth w, respectively, for reasons that we describe immediately.

When an encounter occurs several years or decades in the future, as is typical for the cases considered to date, the orbital uncertainty is strongly dominated by uncertainty in the anomaly. In such cases Z_Y is very long and slender and thus in the vicinity of Earth the ellipse can often be treated as a strip of constant width, which allows a convenient interpretation of the above parameters. Specifically, under this "strip approximation," when moving along the major axis, or spine, of Z_Y we are essentially changing only the anomaly of the asteroid; thus the minimum distance to the geocenter is the minimum encounter distance possible for any variation in the timing of the encounter. But this is precisely the definition of the MOID, which leads to the result that $d \sin\alpha$ is a good approximation for the MOID in such situations. It follows that the semiwidth w is essentially the uncertainty in the MOID. Furthermore, under the strip approximation, the semimajor axis of Z_Y will be closely aligned with the projection of Earth's heliocentric velocity on the target plane; i.e., it is aligned with the ζ-axis, if the reference system on the target plane was chosen in this way. Its length, which is the stretching Λ, indicates the amount by which the original ellipsoid Z_X has been stretched by the propagation F, as well as the timing uncertainty of the encounter $\Delta t = \Lambda/(v_\oplus \sin\theta)$. In some cases, this strip approximation can be inappropriate, generally because either the encounter is close to the time of the observations t_0, leading to a less-eccentric ellipse with errors not dominated by uncertainty in the time of the encounter t_1, or because the mapping F is locally highly nonlinear and interrupted returns (see section 5.4) are present on the target plane. In the former case, a strictly linear analysis of the encounter will typically be suitable, while in the latter case more sophisticated methods are needed to detect and analyze the encounter.

To compute the impact probability, one simply integrates the target-plane probability density p_{TP} over the cross-sectional area of Earth (*Chodas and Yeomans,* 1994). Within the Gaussian formalism, the bivariate probability density p_{TP} can be computed as the product of two univariate probability densities

$$p_{TP}(\sigma_1,\sigma_2) = p(\sigma_1)p(\sigma_2)$$

In fact, the probability density $p(\sigma)$ is often assumed to be Gaussian, but can also be modeled as a uniform density over the interval $|\sigma| \le 3$, so that $p(\sigma) = 1/6$ within this interval and zero elsewhere; in the latter case the above formula is an approximation. Note that the value 3 for the limits on σ is somewhat arbitrary and based on experience; a rigorous statistical limit cannot be computed without going through a rigorous weighting procedure as described in section 3.2 and in *Carpino et al.* (2000). Although the real error distribution of asteroid astrometry is clearly not uniform, it has much stronger tails than a Gaussian distribution with formally derived standard deviations. In practice, the true probability distribution lies somewhere between the two, and either distribution will often suffice to obtain an order-of-magnitude estimate of the impact probability.

The recently proposed Palermo Technical Scale can be used to infer the significance of a potential impact event,

based on the impact probability, size of the object, and time until the potential impact (*Chesley et al.,* 2001). However, when the impact probability is very low, say $\ll 10^{-9}$, it is practically indistinguishable from zero; it is therefore reasonable to say that an impact is not possible. In a strictly theoretical sense, if the probability is computed within a Gaussian formalism, it is never zero, although the values can be meaninglessly small.

3.5. Limitations of Linear Theory

The linear theory described so far has little use when searching for low-probability impact solutions. It should serve as a point of reference for applying the nonlinear approaches described in the following sections, but it is very important to use this valuable tool in a critical way, being aware of its limitations.

Nonlinearity arises from each stage in the computation procedure:

1. The least-squares fit for the elements X at epoch t_0 can be poorly constrained and so the ellipsoid $Z_X(\sigma)$ has some very long axes and is a poor approximation of the true confidence region. This form of nonlinearity is often manifested by confidence regions that appear as curved ellipsoids, or "bananoids."

2. The integral flow is nonlinear, and the longer the time of propagation the more nonlinearity accumulates. This means that even if the confidence ellipsoid were adequate at epoch t_0, the linear mapping of the covariance to a much later time t_1 might be very different from the exact propagation of the orbits in the original confidence ellipsoid, which will often appear folded, leading to interrupted returns as described by resonant return theory (sections 5.4 and 6).

3. The projection from the space of elements X onto the target plane Y is nonlinear, partly as a result of the transformation from orbital elements to Cartesian coordinates. Additionally, we note that for encounters with substantial timing uncertainty, the target-plane confidence ellipse will have a very long major axis, and the true confidence region will be actually curved by an amount corresponding to the curvature of Earth's heliocentric orbit (i.e., the true confidence region deviates from the confidence ellipse to the same extent that Earth's path deviates from a line). This problem is most severe for encounters with very large timing uncertainty or very low relative velocity.

Besides the more obvious limitation of linear theory, that the confidence region is not ellipsoidal, there is also the important problem that only the encounters experienced by the nominal orbit are even detected. And yet it is not unusual for the asteroid uncertainty to grow to a significant fraction of the entire orbit, or even wrap all the way around the orbit one or more times by the time of a potential collision, at which time the nominal orbit may be far from Earth, even on the opposite side of the Sun, and thus the linear analysis would not indicate any hazard at that time. Indeed, threatening encounters cannot even be reliably detected unless the probability of impact rises above ~10^{-4}, assuming that only encounters of the nominal orbit passing within 0.1 AU are considered.

We know that the curvature of the true confidence region, with respect to its linear approximation, is essentially the curvature of Earth's heliocentric orbit. Thus the use of the linear approximation cannot be accurate to 10^{-4} AU when used over a distance larger than $\sqrt{10^{-4}} = 10^{-2}$ AU. Thus, as a rule of thumb for Earth encounters, if $d \cos \alpha$ is not more than several lunar distances, then the local curvature of the confidence region will have only a small effect and the value of $d \sin \alpha$ will be a reliable estimate of the MOID to within a few Earth radii. This implies that even encounters detected along the nominal orbit cannot be reliably recognized as threatening using linear theories until the impact probability exceeds ~10^{-3}. For all these reasons, the linear analysis of the nominal trajectory is often inadequate or even erroneous, and nonlinear sampling and analysis techniques are required.

4. VIRTUAL ASTEROIDS

4.1. Discretization of the Confidence Region

When an asteroid is discovered, we cannot immediately know "the orbit" of the real object; rather, we can describe our knowledge by thinking of a swarm of virtual asteroids (VAs) with slightly different orbits all compatible with the observations. The reality of the asteroid is shared among the VAs, in the sense that only one of them is real but we do not know which one. Since the confidence region contains a continuum of orbits, each virtual asteroid is in turn representative of a small region; i.e., its orbit is also uncertain, but to a much smaller degree. This smaller uncertainty enables us to use for each VA some local algorithms, such as linearization, that would be inappropriate over the entire confidence region. Note that the nominal orbit is just one of the VAs, and is not extraordinary in this context.

The problem is then how to sample the confidence region with a limited number of VAs (in practice, a number ranging between a thousand and several tens of thousands) in such a way that, if an impact is indeed compatible with the available information, a virtual impactor (VI) is found among the selected virtual asteroids (*Milani et al.,* 2000a). Again, a VI is not an isolated collision orbit, but a representative of a small connected set in the confidence region in the initial conditions space, formed by points leading to a collision with a given planet at approximately the same time. Two main approaches have been devised to scan for VIs: the Monte Carlo methods and the line-of-variations methods.

4.2. Monte Carlo Sampling

The Monte Carlo (MC) methods directly use the probabilistic interpretation of the least-squares principle. Since the orbit-determination process yields a probabilistic dis-

tribution in the space of orbital elements, the distribution can be randomly sampled to obtain a set of equally probable virtual asteroids. They will be more dense near the nominal solution, where the probability density is maximum, and progressively less dense as the RMS of the residuals increases.

The random sampling needed to begin an MC test can be done in a simple way in the linear case, where the probability density in the orbital-elements space is well approximated by a Gaussian distribution with the mean value and covariance established by the nominal solution. In this case, a standard random number generator, providing a unit variance Gaussian distribution for a single variable, can be used to generate a random sampling according to the Gaussian distribution in the elements space (*Chodas and Yeomans,* 1999c).

In the nonlinear case, the probability density in the orbital-elements space is not given by an explicit analytical formula and, although it theoretically exists, cannot be directly computed. Thus the random sampling needs to take place in the space of observations, where the probability density is assumed to be Gaussian according to some assumed error model, although establishing such a model is a complex problem (*Carpino et al.,* 2001). The orbit determination process is then repeated, and the nominal solution for each set of modified observations is taken as a VA. Of course, this nonlinear MC procedure is computationally more expensive and should be used only when a linear MC is inappropriate. In practice, however, the cases that require the nonlinear MC treatment have very short observational arcs (typically much less than 30 days), and the time interval over which to search for impactors is orders of magnitude longer. Thus the computational load of the orbit determination is much less than the propagation, and the overhead of the nonlinear MC method (with respect to the linear MC) is not important.

4.3. Line of Variations Sampling

Whereas the MC samples consist of a scattering of unrelated points covering the entire six-dimensional confidence region, there are important computational advantages to sampling a reduced subspace (e.g., a one-dimensional continuous line) that is hopefully representative of the entire confidence region. For this purpose we can use the line of variations (LOV), which is the line of weakness of the orbit-determination solution. In the linear case, when the confidence region is well represented by the confidence ellipsoid, the LOV is simply the major axis of the ellipsoid Z_X, defined by the eigendirection V_1 associated with the largest eigenvalue λ_1 of the nominal covariance matrix $\Gamma(X^*)$, and its semilength is $\sigma_1 = \sqrt{\lambda_1}$. In the nonlinear case, for each solution X in the confidence region we can compute the covariance matrix $\Gamma(X)$ and the corresponding unit eigenvector $V_1(X)$ satisfying some orientation rule, e.g., the component of V_1 along the axis of the semimajor axis is positive. Then $\sigma_1(X)V_1(X)$ is a well-defined, smooth vector field, and the LOV is the unique solution of the ordinary differential equation

$$\frac{dX}{d\sigma} = \sigma_1(X)V_1(X)$$

with initial conditions at the nominal solution X^* for $\sigma = 0$. A numerical method for solving this equation is described by *Milani* (1999); more accurate methods have been developed for this and other applications.

It is natural to sample the LOV with VAs spaced uniformly in the cumulative probability so that each VA, representing a slice of the confidence region, has the same probability of representing the true orbit. One approach is to assume a Gaussian distribution along the LOV, which leads to short steps in the independent variable σ near zero and progressively larger steps as one moves away from the nominal solution. An alternate and more convenient approximation assumes a uniform probability density up to some value (e.g., for $-3 \le \sigma \le 3$); under this hypothesis the sampling will be uniform in σ.

4.4. Discussion

Our ultimate objective in this development is to investigate the entirety of the orbital confidence region to determine if there is any meaningful possibility of future collision. To this end, after computing a set of virtual asteroids, one must propagate each of them forward to the end of some timespan of interest, which in most cases extends 50–100 yr into the future. During the propagations any planetary encounters are noted for later analysis.

The most straightforward means of searching for collisions is to note the VAs that impact directly. With this very simple approach, if some of the VAs have close approaches within the radius of the planet they qualify as Virtual Impactors, and a simple estimate of the probability of impact for that asteroid over the timespan of interest would simply be the ratio of VIs to VAs. The LOV method of generating VAs is ill suited to this technique because such VAs cannot impact unless the LOV directly intersects Earth on the target plane, even though nearby off-LOV solutions may impact. However, the Monte Carlo-derived VAs do sample the entire volume of the orbital confidence region, and thus they can reveal all potential collisions with a reliability described by Poisson statistics. But this points to the main limitation of such a simple method: A VI can be reliably detected only if the impact probability is of the order of the inverse of the number of VAs used. For example, to achieve a complete scan of the confidence region down to probabilities of impact of ~10^{-6} would require the propagation of a few million orbits, presently requiring a week or two of CPU time (for just one NEA). Thus this approach is currently too inefficient to be generally practical. For this reason it is necessary to apply some form of local analysis in the neighborhood of the close approaches detected for each VA.

The LOV approach is especially favorable for local analysis because one can exploit the fact that the LOV VAs be-

long to a continuous one-dimensional set. Thus it is possible to interpolate between the target plane points belonging to two consecutive VAs. As an example, if one VA passes from the nodal point before Earth, and the next consecutive one passes behind Earth, by continuity there must be a point on the LOV, corresponding to an intermediate value of σ, such that the encounter takes place at the MOID. Therefore the MOID of the VA orbit, at a time shortly before the encounter, can be used as a criterion to identify any potentially threatening encounters. In such cases some local analysis is needed to search for VI regions significantly smaller than the VA sampling scale, thus permitting searches for VIs with a probability significantly lower than the inverse of the number of VAs. Monte Carlo solutions are less well suited to interpolation because they are not ordered and are not constrained to a subspace. However, as was explained in section 3.4, the trace of the confidence region on the target plane is often a very narrow strip (e.g., see Figs. 1 and 2). In such cases this LOV interpolation can also be done using MC-derived VAs, but one must take care to ensure that the interpolation is based on dynamically similar VAs, as this is not generally clear *a priori*.

Even though interpolation along the LOV is generally very effective, we cannot be confident that the LOV approach will detect all potential collisions. If some VIs lie well off the LOV and are separated from it by some strong nonlinearity (e.g., the MOID having been changed with respect to the value along the LOV by some encounter at the other node or with another planet), then the VAs selected along the LOV may fail to indicate some potentially threatening encounters. The MC approach would, in principle, reveal such encounters. In the end, we reach the conclusion that both methods have important advantages, and a robust search procedure should use both methods in a coherent hybrid scheme to ensure efficient and reliable detection of all VIs above a given probability.

5. NONLINEAR TARGET-PLANE ANALYSIS

5.1. All Possible Encounters

Whether the confidence region in initial orbital elements is sampled with a Monte Carlo method, yielding a cloud of VAs, or with a Line Of Variations method, yielding a one-dimensional string of VAs, the samples are all numerically integrated forward to some time horizon (typically 50–100 yr), and all close approaches to perturbing bodies within some detection threshold distance are recorded. At each close approach, the time, distance, geocentric position, and velocity coordinates are noted. The close-approach threshold distance used for Earth is typically ~0.1–0.2 AU.

As noted in the previous section, a simple search for VIs among the sample VA orbits is not adequate for finding all potential collisions because this approach is likely to miss low-probability impacts. Consider, for example, a lost asteroid. Its confidence region quickly spreads along the entire orbit, which for a typical NEA has a length of ~10 AU. Since the diameter of Earth is less than 10^{-4} AU, and the number of samples is typically ~10^4 or less, it is unlikely that one of the virtual asteroids will impact directly, even if the orbits intersect and an impact is possible. There is clearly a need to use a method of local analysis in the neighborhood of a given VA to search for nearby VIs.

When ordered by time, the close approaches of all the VAs are seen to cluster around a series of encounter times — for example, when Earth passes close to the asteroid's nodal crossing point with the asteroid nearby. We call the subset of VAs that approach Earth around a given encounter time a *virtual shower,* since it is reminiscent of a stream of meteoroid particles encountering Earth. It is possible that the entire set of VAs form the shower if, for example, the encounter time is close to the initial time t_0. Often, however, the confidence region is so spread out along the orbit that only a small subset of VAs pass within the close-approach detection threshold.

A shower can be broken down into separate trails, which are subsets of closely related VAs, each subset following a qualitatively different dynamical path to the encounter. Often, a shower contains several different trails (see Table 1 in *Milani et al.,* 1999). For example, an asteroid that is lost by more than one full revolution can produce a shower containing trails in which VAs have performed different numbers of revolutions around the Sun since time t_0. Another situation occurs when VAs go through a close approach that changes their orbital periods in such a way as to bring them back for another close approach years later. Such *resonant returns* typically occur in pairs characterized by different stretching (see section 6).

When the individual approaches of a shower are plotted in the target plane, the separate trails typically reveal themselves through differing characteristic MOID values (Fig. 1a). Even if the initial confidence region is sampled with a Monte Carlo method, and the VAs are therefore off the line of variations, the confidence regions of the trails are usually highly stretched, with widths much narrower than the differences in MOIDs. As a result, it is often possible to separate the VAs of a shower into distinct trails simply by sorting on MOID (*Chodas and Yeomans,* 1999c). This method will not work, however, if the MOIDs of the trails are too similar. A more reliable method of separating the VAs of a shower into separate trails is to sort and index them according to their semimajor axis, the variation of which leads to the stretching in the target plane. This indexing can be immediately obtained with the LOV method and requires only minimal computation for MC samples. The trails can then be detected as sequences of n consecutive solutions X_i, $i = k, k + 1, \ldots, k + (n - 1)$ within a given shower (Fig. 1b).

5.2. Filtering Trails

The points of a trail are just sample points of a continuous line more or less following the LOV, and none need approach very close to the minimum possible distance for that trail. To decide if an impact is possible, we need to de-

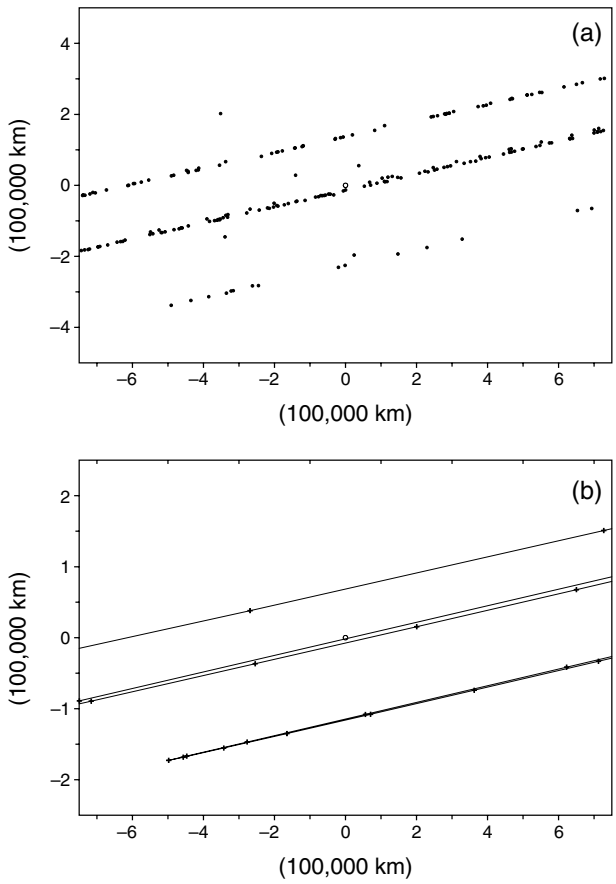

Fig. 1. The January 2046 shower from the VAs of 1998 OX_4 plotted on the b-plane with VAs derived by **(a)** MC and **(b)** LOV methods. Earth is represented to scale by the small open circle in the center of each plot. Since one of the trails clearly intersects Earth, a collision is possible. Note that the axes in this figure are not the (ξ, ζ) axes discussed in section 2.2, and the trails are not therefore aligned with the vertical axis.

termine precisely how close this trail approaches Earth. Since this new problem is computationally intensive, it is essential to apply some filter to discard immediately the trails that clearly cannot collide. Two methods are available for this: selection by MOID and linear analysis.

For each VA, we can compute the MOID, reduced by the amount of gravitational focusing (as described in section 2.2 for $b_\oplus$) to obtain the perturbed minimum approach distance $MOID_{GF}$. To determine whether an orbit will impact, the MOID should be computed based on the elements near the time of the encounter but not during the encounter (because the effects of the encounter itself will corrupt the MOID). However, lunar perturbations on Earth cause high-frequency fluctuations in the MOID, with amplitudes on the order of one Earth radius. This effect can be minimized by computing the MOID one lunar period before the encounter, but in general if $MOID_{GF} < 2r_\oplus$ then a collision for that particular trail cannot safely be ruled out, while larger values of $MOID_{GF}$ do indicate that no impact can take place due solely to timing uncertainty. But the MOID itself is uncertain due to the dispersion of the orbital elements at the time of the encounter (*Bonanno,* 2000), so in the absence of MOID uncertainty information we must use a generous safety margin, i.e., only ruling out collision when $MOID_{GF}$ is greater than, say the lunar distance.

As an alternative, we can perform a linear target-plane analysis for the closest-approaching VA found in each trail and discard the cases in which an impact is well outside of the confidence ellipse. Both methods have advantages and disadvantages. If the confidence ellipse on the target plane is wider than our safety margin, then the MOID of the VA under consideration is not representative of the values the MOID could assume for initial conditions inside the confidence region but far from the LOV. On the other hand, if the VA considered has only a shallow approach (e.g., further than 0.02 AU as discussed in section 3.5), the results of the linear analysis can be unreliable. The best solution is to use information from both methods, e.g., using the value of the MOID and allowing for the possibility of a decrease of its value by the 3σ width of the confidence ellipse.

Once a particular trail has been found to allow a very low minimum distance, i.e., too close to rule out a collision, then we need to carefully determine the actual value of this minimum. We first consider straightforward cases of trails with simple geometry and then turn to complex cases with strong nonlinearities.

5.3. Straightforward Encounters

When the VAs of a simple trail are projected onto the target plane, the ζ-coordinates of the intercept points form a monotonic sequence proceeding from one side of Earth to the other (i.e., spanning zero). To find the VA which makes the closest possible approach of a trail, several methods can be used.

One technique is a variant of Newton's method, which starts from the solution X_j with the closest approach among the VAs for the trail and applies corrections to obtain a new VA with an even closer approach. First, we compute the direction of the long axis of the target-plane confidence ellipse for X_j; we then find the point Y_{min} along this line that is closest to Earth in the target plane. The desired correction in the target plane is therefore $\Delta Y = Y_{min} - F(X_j)$. We then need to determine a change ΔX in the initial orbital elements about X_j that satisfies the linear mapping

$$DF\ \Delta X = \Delta Y$$

Since DF maps from a six-dimensional space to a two-dimensional space, there is no unique solution, but with the further constraint that the solution minimize the sum of the squares of residuals, a unique solution for ΔX can be found (*Milani et al.,* 2000a). With the new solution $X_j + \Delta X$ in hand, we numerically integrate to the encounter time and compute the new close approach distance. Since the direction of the correction ΔY is recomputed at each step, any

curvature introduced by gravitational focusing is accounted for. The iterations are stopped when the close approach distances stop decreasing. This approach has proven very effective in most cases, but it is sometimes prone to divergence, especially in strongly nonlinear cases, as will be described in section 5.4.

Another method for finding the VA producing the minimum possible close approach along the LOV uses the *regula falsi* interpolation method. Starting from two consecutive VAs having target-plane ζ-coordinates with opposite signs, Y_j and Y_{j+1}, we compute the point Y_{min} closest to Earth along the line between them. We then obtain a new VA by interpolating to the corresponding point along the LOV between X_j and X_{j+1} in the space of initial orbital elements and iterate the procedure. This approach does not require the computation of the confidence ellipse in the target plane or the use of the linear mapping DF, but it does require that the trail comprise at least two VAs. This is in contrast to Newton's method above, which is more complex but can proceed from only a single point on the target plane.

Once the minimum possible distance D has been found for initial conditions along the LOV, we need to take into account the width of the confidence region on the target plane. The trace of the LOV on the target plane is a kind of spine of the confidence region, and the width can be estimated by using the minor axis w of the confidence ellipse (see section 3.4). For example, since w corresponds to 1σ, the minimum distance to the confidence region corresponding to σ = 3 is approximately D – 3w, assuming that D > 3w. The knowledge of the distance between the LOV and Earth, measured in sigmas, can be essential for evaluating the probability of impact, as detailed in section 5.5. This approximation corresponds to linearization in the neighborhood of the LOV point corresponding to the close approach at distance D, and this is adequate if D and w are small enough.

Both Newton's and the interpolation methods move along the line of variation to determine the minimum encounter distance. There is also a method developed by *Sitarski* (1999) that moves in the shortest direction to find the minimum cost Q for collision, and with this information the impact hazard can be eliminated, or at least the impact probability bounded. This method does not, however, permit a careful assessment of the impact probability since the probability density in the vicinity of the VI is unknown. Another method has been proposed by *Muinonen* (1999) to find the minimum value of the penalty ΔQ compatible with a collision, without restricting the search to the LOV. A mixed method, involving both one-dimensional and multidimensional explorations of the confidence region, has been developed by *Muinonen et al.* (2001). Additional discussion on this point is found in section 3.3 of *Bowell et al.* (2002).

The upper trails indicated in Fig. 1 are typical straightforward trails. Another excellent example of this type was the 2028 close approach of 1997 XF_{11} (based upon the 1997–1998 observations only). This case has a special feature: The confidence region is long (millions of kilometers), but the nominal solution results in a quite close approach in 2028. Thus linearization around the nominal is good enough to draw the conclusion that an impact in 2028 is not possible (Fig. 2a), despite the fact that the departure between the confidence region and the linear ellipse is evident (hundreds of kilometers) near the ends of the 3σ target-plane ellipse (Fig. 2b); this departure is greater on the MTP than on the b-plane (*Milani and Valsecchi,* 1999; *Chodas and Yeomans,* 1999c).

5.4. Complex Encounters

Some trails do not follow the simple behavior described above. When the VAs for these complex trails are projected into the target plane, the sequence of intercept points approaches Earth along the LOV, slows down its approach,

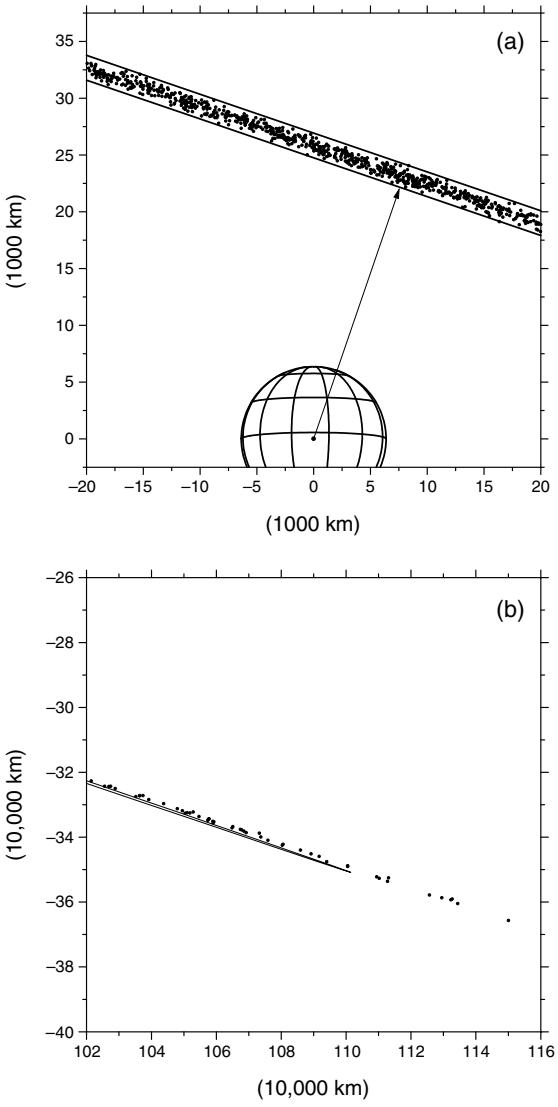

Fig. 2. The linear (3σ ellipse) and nonlinear (Monte Carlo points) confidence regions on the 2028 b-plane for the 88-d arc of 1997 XF_{11}. **(a)** Confidence region near nominal solution; **(b)** confidence region near 3σ. The full region extends over some 2,000,000 km, but is only 2000 km wide.

and then turns back and recedes in the direction from which it approached (see Fig. 7 in *Milani et al.,* 2000b). We call this behavior an interrupted return, and it arises as a result of previous close approaches, as explained in section 6. The bottom trail depicted in Fig. 1 is an interrupted return, as is clear from Fig. 1b where the trail both enters and exits the figure on the right. If the orientation of the LOV is mapped on the target plane for each VA then this reversal in direction can easily be detected.

It can be shown that for an interrupted return the derivative of Δt with respect to the σ parameter on the LOV passes through zero, and, not surprisingly, Newton's method can fail completely for these strongly nonlinear cases. Methods to identify such cases, and to perform an alternative local analysis to detect possible degenerate VIs and estimate their impact probability, are being tested. One possible approach is to use an LOV method including controls to detect interrupted returns and revert to an MC method restricted to the neighborhood of the VA near the folding point of the interrupted return. Another possibility is to use the regula falsi method along the LOV, exploiting the fact that minima of the approach distance D as a function of σ do exist, and they can be identified as zeros of the derivative $dD/d\sigma$. We must handle interrupted returns carefully because the point on the target plane at which the trail appears to stop and go back has relatively low stretching and may actually lie inside Earth's disk. This corresponds to the tangent case of section 6 and may result in a probability of impact larger that that of an ordinary trail.

5.5. Probability of Impact

If the minimum possible close approach distance of a trail is less than one Earth radius, then we know that an impact is possible, and our next task is to estimate the probability of impact P_I. In the following, we assume that we have identified the orbit X_I with the minimum possible encounter distance along the LOV as described in section 5.3. For this orbit we have σ_Λ, which is the σ distance from the nominal along the LOV, and we also have the target-plane covariance Γ_Y, from which we obtain the local stretching Λ and semiwidth w.

Now, taking X_I as nominal, we can compute the impact probability $\tilde{P}_I$ according to the linearized procedure described in section 3.4. We note that, since we are at the minimum distance point along the LOV, the assumption of local linearity is valid despite the fact that the location and properties of the trail itself may be dominated by nonlinear effects (see Fig. 2). But since X_I is not actually the nominal orbit, we must correct $\tilde{P}_I$ for the distance from the nominal to obtain

$$P_I = \tilde{P}_I \frac{p(\sigma_\Lambda)}{p(0)}$$

where we have used again the univariate probability density $p(\sigma)$ discussed in section 3.4.

When Λ is large (e.g., greater than the lunar distance), as is typical for highly stretched, far-future encounters, the strip approximation (section 3.4) permits the assumption that the probability density is a constant over the cross-section of Earth, which allows a very simple and convenient computation of the probability. If we can further assume that $w \ll r_\oplus$, then simple one-dimensional methods may be used, since the full width of the confidence strip falls onto Earth. In this case, the geometric chord length ℓ of the intersection of the trail's LOV with Earth's disk is the main parameter of interest, and P_I is computed from the cumulative probability distribution over ℓ

$$P_I = \frac{p(\sigma_\Lambda)}{\Lambda} \ell$$

On the other hand, when $w \gg r_\oplus$, fully two-dimensional methods must be used to compute impact probability; this only requires the additional consideration of the lateral distance to the geocenter σ_w so that the probability density orthogonal to the LOV can be computed. P_I is then calculated from a constant bivariate probability density integrated over the cross-section of Earth

$$P_I = \frac{p(\sigma_\Lambda)\, p(\sigma_\omega)}{\Lambda\omega} \pi r_\oplus^2$$

When $w \simeq r_\oplus$, these approximations are not suitable because $p(\sigma_w)$ cannot be assumed constant over the width of Earth: A two-dimensional probability integral needs to be computed. Even more complicated cases can be handled with suitable probabilistic formalisms, such as the ones of *Muinonen et al.* (2001).

6. RETURNS AND KEYHOLES

6.1. Resonant Returns

Virtual impactors are normally found numerically, given that a realistic physical model of their motion is quite complex; on the other hand, an exploration of the problem using analytical tools can give insight into the approximate location of VIs in elements space. Let us discuss the simple and rather common case of a VI whose impact takes place at a resonant return.

A resonant return occurs when, as a consequence of an encounter with Earth, the asteroid is perturbed into an orbit of period $P' \approx k/h$ yr, with h and k integers; then, after h revolutions of the asteroid and k revolutions of Earth, both bodies are again in the region where the first encounter occurred and a second encounter takes place (*Milani et al.,* 1999).

An analytical theory of resonant returns has been recently developed by *Valsecchi et al.* (2001). This theory treats close encounters with a suitable extension of Öpik's theory (*Öpik,* 1976), and adds a Keplerian heliocentric propagation (modified to account for the evolution of the

MOID; see section 6.2) between encounters, thus establishing a link between the outcome of an encounter and the initial conditions of the following one. The motion during an encounter with Earth is modeled by simply assuming that it takes place on a hyperbola; one of the asymptotes of the hyperbola, directed along the unperturbed geocentric encounter velocity $\mathbf{v}_\infty$, crosses the b-plane at a right angle, and the vector from Earth to the intersection point is **b**.

We can describe $\mathbf{v}_\infty$ in terms of its modulus, v_∞, and two angles, θ and ϕ; θ is the angle between $\mathbf{v}_\infty$ and Earth's heliocentric velocity $\mathbf{v}_\oplus$, while ϕ is the angle between the plane containing $\mathbf{v}_\infty$ and $\mathbf{v}_\oplus$ and the plane containing $\mathbf{v}_\oplus$ and the ecliptic pole.

We recall the (ξ, η, ζ) reference frame defined in section 2.2, where η is normal to the b-plane and ζ is oriented in the direction opposite to that of the projection of $\mathbf{v}_\oplus$ on the b-plane. Then ξ lies in both the b-plane and the plane normal to $\mathbf{v}_\oplus$ (*Greenberg et al.,* 1988). This choice of the b-plane coordinates has the nice property that ξ is simply the local MOID, while ζ is proportional to the time delay with which the asteroid "misses" the closest possible approach to Earth.

Öpik's theory then simply states that the encounter consists of the instantaneous transition, when the small body reaches the b-plane, from the preencounter velocity vector $\mathbf{v}_\infty$ to the postencounter vector $\mathbf{v}'_\infty$, such that $v'_\infty = v_\infty$, and θ' and ϕ' are simple functions of v_∞, θ, ϕ, ξ, and ζ; the angle between $\mathbf{v}_\infty$ and $\mathbf{v}'_\infty$ is the deflection γ, given by

$$\tan\frac{\gamma}{2} = \frac{c}{b}$$

where $c = GM_\oplus/v_\infty^2$. Finally, simple expressions relate a, e, i to v_∞, θ, ϕ (*Carusi et al.,* 1990), and ω, Ω, f to ξ, ζ, t_0 (*Valsecchi et al.,* 2001); t_0 is the time at which the asteroid passes at the node closer to the encounter. It is important to note that for small encounter distance a depends only on v_∞ and θ and, correspondingly, a' depends only on v_∞ and θ'.

We can formulate the condition for a resonant return to occur as a condition on θ'; for a ratio of the periods equal to k/h, we obtain a value of a', say a'_0, given by $a'_0 = \sqrt[3]{k^2/h^2}$, and a corresponding value of θ', say θ'_0.

Given θ, we can easily compute θ'_0 from ξ, ζ:

$$\cos\theta'_0 = \cos\theta\frac{b^2 - c^2}{b^2 + c^2} + \sin\theta\frac{2c\zeta}{b^2 + c^2}$$

thus obtaining the locus of points on the b-plane leading to a given resonant return. We can rearrange it in the form

$$\xi^2 + \zeta^2 - 2D\zeta + D^2 = R^2$$

that is the equation of a circle centered on the ζ-axis (*Valsecchi et al.,* 2000); R is the radius of such a circle and D is the value of the ζ-coordinate of its center, and they can be obtained from

$$R = \left|\frac{c\sin\theta'_0}{\cos\theta'_0 - \cos\theta}\right|$$

$$D = \frac{c\sin\theta}{\cos\theta'_0 - \cos\theta}$$

Actually, the expression for the locus of points on the b-plane leading to a given resonant return contains terms that are of the third order in ξ, ζ, and c (*Valsecchi et al.,* 2001), but for the purposes of the present qualitative discussion these terms can be ignored.

Figure 3a shows the arrangement of the b-plane circles corresponding to resonant returns to close encounters in 2040, 2044, and 2046 for the August 2027 encounter of asteroid 1999 AN_{10}; these correspond, respectively, to the mean-motion resonances 7:13, 10:17, and 11:19. A straight line is also drawn in the plot at $\xi = 6r_\oplus$ that represents a string of VAs having the same orbit and MOID (and thus the same ξ) and spaced in the timing of encounter with Earth (and thus different values of ζ). It is clear that in such an arrangement there are two regions for each resonance that lead to resonant returns, and it is also clear that they are located at the crossings of the straight line with the appropriate circle. Figure 3a also shows the circle associated with the 3:5 mean-motion resonance, corresponding to a resonant encounter in 2032; note that the straight line does not actually touch the circle, implying that no close encounter takes place.

6.2. Keyholes

Chodas (1999) introduced the term "keyhole" to indicate small regions of the b-plane of a specific close encounter such that, if the asteroid passes through one of them, it will hit Earth on a subsequent return. This can be generalized by requiring that the subsequent approach occurs to within a given small distance. An impact keyhole is thus just one of the possible preimages of Earth's cross section on the b-plane and is therefore tied to the specific value for the postencounter semimajor axis that allows the occurrence of the next encounter at the given date. Figure 3b depicts the two keyholes associated with the 7:13 resonance that are also shown in the top half of Fig. 3a.

The b-plane circles corresponding to a given a' address the question of the timing (i.e., the ζ-coordinate) of the subsequent encounter. However, if we use a simple Keplerian propagation between encounters, this would leave unchanged the MOID (the ξ-coordinate) of the next encounter. This would be unrealistic, as in general, the MOID varies between encounters for two main reasons: On a long timescale, secular perturbations (*Gronchi and Milani,* 2001) make it slowly evolve through the so-called Kozai cycle, or ω-cycle, while on a shorter timescale significant quasi-periodic variations are caused by planetary perturbations and, for planets with massive satellites, by the displacement of the planet with respect to the center of mass of the planet-satellite system.

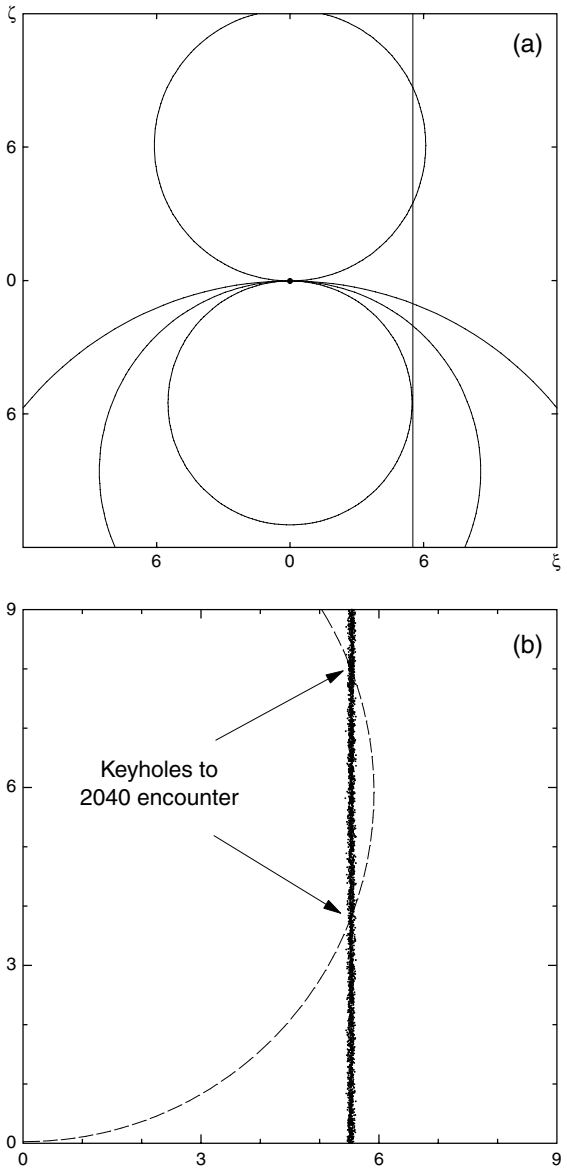

Fig. 3. Circles corresponding to various mean-motion resonances on the b-plane of the August 2027 encounter with Earth of asteroid 1999 AN_{10}. Distances are in Earth radii augmented for gravitational focusing ($b_{\oplus}$; see section 2.2). **(a)** Uppermost circle: 7:13 resonance; then, with the centers along the ζ-axis, from top to bottom: 3:5, 10:17, and 11:19 resonances. The vertical line at $\xi = 5.5\ b_{\oplus}$ represents fictitious asteroids, all with the same orbit as 1999 AN_{10} and spaced in the time of encounter with Earth. **(b)** Explicit depiction of keyholes for the 2040 returns on the 2027 b-plane of 1999 AN_{10}. The stream of Monte Carlo points corresponds to the vertical line in **(a)**. The dashes represent the best-fitting circle passing through the impacting zones.

For the purpose of obtaining the size and shape of an impact keyhole we can, however, just model the secular variation of the MOID as a linear term affecting ξ", the value of ξ at the next encounter

$$\xi'' = \xi' + \frac{d\xi}{dt}(t_0'' - t_0')$$

where t_0' and t_0'' are the times of passage at the node, on the post-first-encounter orbit, that are closest to, respectively, the first and the second encounter. We can compute the time derivative of ξ either from a suitable secular theory for crossing orbits (*Gronchi and Milani,* 2001) or by deducing it from a numerical integration. The result could then be corrected to take into account the short periodic terms, possibly by using the output of a numerical integration. Without such numerical corrections, the theory would reliably predict very close encounters (e.g., within a few thousandths of AU), but not collisions, which require accuracies an order of magnitude better.

The computation of the size and shape of an impact keyhole is described in detail in *Valsecchi et al.* (2001); here we give a qualitative discussion. Let us start from the image of Earth in the b-plane of the subsequent encounter, the one in which the impact should take place. We denote the coordinate axes in this plane as ξ", ζ" and consider the circle centered in the origin and of radius $b_{\oplus}$. The points on the b-plane of the first encounter that are mapped — by the Keplerian propagation plus the MOID drift — into the points of Earth's image circle in the b-plane of the second encounter constitute the Earth preimage we are looking for.

As far as the location is concerned, impact keyholes must lie close to the intersections, in the first encounter b-plane, of the circle corresponding to the suitable resonant return and the vertical line expressing the condition that the MOID be equal to

$$-\frac{d\xi}{dt}(t_0'' - t_0')$$

What about size and shape? To address this question, we can examine the matrix of partial derivatives $\partial(\xi'',\zeta'')/\partial(\xi,\zeta)$. We do this under the assumption, as in *Valsecchi et al.* (2001), that for the encounter of interest $c^2 \ll b^2$, since this condition is not too restrictive in most asteroid encounters with Earth. The matrix of partial derivatives has the following structure

$$\begin{bmatrix} \frac{\partial\xi''}{\partial\xi} & \frac{\partial\xi''}{\partial\zeta} \\ \frac{\partial\zeta''}{\partial\xi} & \frac{\partial\zeta''}{\partial\zeta} \end{bmatrix} \approx$$

$$\begin{bmatrix} 1 + \mathcal{O}\left(\frac{c}{b}\right) & \mathcal{O}\left(\frac{c}{b}\right) \\ \mathcal{O}(1) - 2hf\frac{\xi\zeta}{b^4} & 1 + \mathcal{O}\left(\frac{c}{b}\right) + hf\frac{\xi^2 - \zeta^2}{b^4} \end{bmatrix}$$

where f is a constant not depending on h.

The first row of the matrix shows that the ξ dimension of the preimage of Earth is essentially unchanged; the second row, however, gives a completely different picture. First, both derivatives grow linearly with h, the number of

heliocentric revolutions the asteroid makes between the two encounters; second, the relative size of the two elements of this row depends critically on the values of ξ and ζ. Apart from the case in which $\xi \approx \zeta$, which we will discuss below, $\partial\zeta''/\partial\zeta$ can be rather large; this means that the separation in ζ of two VAs, that grows with $\partial\zeta''/\partial\zeta$ going from the first to the second encounter, can increase by a large amount. If we now go backward in time, from the second to the first encounter, the expansion in ζ becomes a contraction; this contraction affects the preimage of Earth, by squeezing it along ζ. The result of this is that the keyhole has the form of a lunar crescent that closely follows the appropriate b-plane circle, with a "width" in ξ of about $2b_\oplus$ and a maximum thickness of $\sim 2b_\oplus/(\partial\zeta''/\partial\zeta)$.

When $\xi = \zeta$, the third term in $\partial\zeta''/\partial\zeta$ becomes 0, and somewhere in the vicinity the entire derivative becomes very small, <1; this can happen if the absolute value of the MOID at the first encounter is equal to the radius of the appropriate resonant circle. As a consequence of the smallness of $\partial\zeta''/\partial\zeta$, the dimension in ζ of the keyhole can be larger than $b_\oplus$!

The possibility of anomalously large keyholes is a manifestation of a more general phenomenon that takes place whenever a string of VAs is almost tangent to, but does not cross, a resonant return circle, as in the case of the 3:5 resonant circle shown in Fig. 3a. When this is the case, what happens at the time of the resonant return is that the string of VAs "enters" the b-plane of the resonant encounter from the side of either positive or negative ζ but then, since none of the VAs quite reaches the resonant semimajor axis, at some point it has to "turn back." The turning point is obviously a point where $\partial\zeta''/\partial\zeta = 0$. This situation is frequently encountered in numerical integrations and is also known as "interrupted return," as shown in the lower trail of Fig. 1. If this was to take place with the MOID at the second encounter $<b_\oplus$, an anomalously large keyhole would occur.

As stated above, the computation of VIs requires sophisticated numerical modeling of the orbital evolution. In this context, the simplifications made in the theory just described may lead one to think that it has no practical value. However, the main advantage of the theory is the geometrical understanding of the global structure of potentially impacting solutions that it gives, and this is particularly relevant for complex or pathological cases such as interrupted returns. In this sense, it can be considered as a tool for guiding the numerical explorations and for contriving fictitious critical scenarios to be fed to automatic monitoring programs for testing.

7. CONCLUSIONS AND FUTURE WORK

The last three years have seen both an enormous increase in our ability to identify impact possibilities and the establishment of the first efficient impact-monitoring system; this progress has been in part driven by the discovery of objects in orbits with very small, but detectable, Earth-impact regions within their confidence region. The situation is now evolving toward a more systematic study of the problem, with the goal of improving the currently operating monitoring system (CLOMON) and of establishing an additional and independent one at JPL.

As a consequence of the first two years or so of operation of CLOMON, with the capability of detecting collision possibilities for newly discovered objects while they are still observable, it has become routinely possible to concentrate followup observations on potentially colliding NEAs; in fact, most NEAs larger than ~100 m in diameter discovered after the end of 1999, having impact possibilities in the confidence regions of their short-arc orbits, have been followed long enough that their confidence regions could be reduced to the point of excluding any detectable impact.

The analytic theory as it stands is still largely incomplete. Future developments should include (1) a complete treatment of nonresonant returns, i.e., returns taking place at the other node, for asteroids having either both nodes at a heliocentric distance very close to 1 AU, or the other node at a heliocentric distance very close to the orbital radius of another planet; (2) a better understanding of the complex phenomena associated with interrupted returns and anomalously large keyholes; and (3) the possibility of sequences of more than two encounters, with the last encounter of the sequence leading to an impact.

The work now in progress, the results of which should be available soon, includes the following developments and tests: (1) the Sentry system, a fully independent monitoring system at JPL, which will include "second-generation" features such as the ones mentioned below (the second-generation CLOMON-2 system is also under development); (2) improved methods to handle interrupted returns in order to sharply reduce the cases of divergent iterations in the Newton method; (3) mechanisms to implement focused or densified Monte Carlos to handle difficult cases, including interrupted returns; and (4) a new observation weighting scheme that allows the use of modified Gaussian probabilities as a reliable estimate of impact probabilities.

Note added in proof: The second-generation impact monitoring systems are now on line, both the Sentry system at JPL (http://neo.jpl.nasa.gov/risk/) and the CLOMON2 system at the University of Pisa (http://newton.dm.unipi.it/cgi-bin/neodys/neoibo?riskpage:0;main) and duplicated at the University of Valladolid (http://unicorn.eis.uva.es/cgi-bin/neodys/neoibo?riskpage:0;main).

Acknowledgments. This research was conducted at the University of Pisa and at IAS-CNR under a contract with the Italian Space Agency, and at the Jet Propulsion Laboratory, California Institute of Technology, under a contract with the National Aeronautics and Space Administration.

REFERENCES

Bonanno C. (2000) An analytical approximation for the MOID and its consequences. *Astron. Astrophys., 360,* 411–416.

Bowell E. and Muinonen K. (1994) Earth-crossing asteroids and

comets: Groundbased search strategies. In *Hazards Due to Comets and Asteroids* (T. Gehrels, ed.), pp. 149–197. Univ. of Arizona, Tucson.

Bowell E., Virtanen J., Muinonen K., and Boattini A. (2002) Asteroid orbit computation. In *Asteroids III* (W. F. Bottke Jr. et al., eds.), this volume. Univ. of Arizona, Tucson.

Carpino M., Milani A., and Chesley S. R. (2001) Error statistics of asteroid optical astrometric observations. *Icarus,* in press.

Carusi A., Valsecchi G. B., and Greenberg R. (1990) Planetary close encounters: Geometry of approach and post-encounter orbital parameters. *Cel. Mech. Dyn. Astron., 49,* 111–131.

Chesley S. R. and Milani A. (2000) An automatic Earth-asteroid collision monitoring system. *Bull. Am. Astron. Soc., 32,* 0602.

Chesley S. R., Chodas P. W., Milani A., Valsecchi G. B., and Yeomans D. K. (2001) Quantifying the risk posed by potential Earth impacts. *Icarus,* in press.

Chodas P. W. (1993) Estimating the impact probability of a minor planet with the Earth. *Bull. Am. Astron. Soc., 25,* 1236.

Chodas P. W. (1999) Orbit uncertainties, keyholes, and collision probabilities. *Bull. Am. Astron. Soc., 31,* 2804.

Chodas P. W. and Yeomans D. K. (1994) Predicting close approaches of asteroids and comets to earth. In *Hazards Due to Comets and Asteroids* (T. Gehrels, ed.), pp. 241–258. Univ. of Arizona, Tucson.

Chodas P. W. and Yeomans D. K. (1996) The orbital motion and impact circumstances of Comet Shoemaker-Levy 9. In *The Collision of Comet Shoemaker-Levy 9 and Jupiter* (K. S. Noll et al., eds.), pp. 1–30. IAU Colloq. 156.

Chodas P. W. and Yeomans D. K. (1999a) Could asteroid 1997 XF_{11} collide with Earth after 2028? *Bull. Am. Astron. Soc., 31,* 703.

Chodas P. W. and Yeomans D. K. (1999b) Orbit determination and estimation of impact probability for near-Earth objects. Paper AAS 99-002, 21st Annual AAS Guidance and Control Conference, Breckenridge, Colorado.

Chodas P. W. and Yeomans D. K. (1999c) Predicting close approaches and estimating impact probabilities for near-Earth objects. Paper AAS 99–462, AAS/AIAA Astrodynamics Specialists Conference, Girdwood, Alaska.

Gehrels T., Matthews M. S., and Schumann A. M., eds. (1994) *Hazards Due to Comets and Asteroids.* Univ. of Arizona, Tucson. 1300 pp.

Greenberg R., Carusi A., and Valsecchi G. B. (1988) Outcomes of planetary close encounters: A systematic comparison of methodologies. *Icarus, 75,* 1–29.

Gronchi G. F. (2001) On the stationary points of the squared distance between two ellipses with a common focus. *SIAM J. Sci. Comp.,* in press.

Gronchi G. F. and Milani A. (2001) Proper elements for Earth-crossing asteroids. *Icarus, 152,* 58–69.

Jazwinski A. H. (1970) *Stochastic Processes and Filtering Theory.* Academic, New York. 376 pp.

Kizner W. (1959) *A Method of Describing Miss Distances for Lunar and Interplanetary Trajectories.* External Publication No. 674, Jet Propulsion Laboratory, California Institute of Technology.

Marsden B. G. (1999) A discourse on 1997 XF_{11}. *J. British Interplanet. Soc., 52,* 195–202.

Milani A. (1999) The asteroid identification problem I: Recovery of lost asteroids. *Icarus, 137,* 269–292.

Milani A. and Valsecchi G. B. (1999) The asteroid identification problem II: Target plane confidence boundaries. *Icarus, 140,* 408–423.

Milani A., Chesley S. R., and Valsecchi G. B. (1999) Close approaches of asteroid 1999 AN_{10}: Resonant and non-resonant returns. *Astron. Astrophys., 346,* L65–L68.

Milani A., Chesley S. R., Boattini A., and Valsecchi G. B. (2000a) Virtual impactors: Search and destroy. *Icarus, 145,* 12–24.

Milani A., Chesley S. R., and Valsecchi G. B. (2000b) Asteroid close encounters with Earth: Risk assessment. *Planet. Space Sci., 48,* 945–954.

Morrison D., ed. (1992) *The Spaceguard Survey: Report of the NASA International Near-Earth-Object Detection Workshop.* (Available on line at http://impact.arc.nasa.gov/reports spaceguard/)

Muinonen K. (1999) Asteroid and comet encounters with the Earth: Impact hazard and collision probability. In *The Dynamics of Small Bodies in the Solar System, A Major Key to Solar System Studies* (A. E. Roy and B. A. Steves, eds.), p. 127. NATO ASI Proc. 522, Kluwer, Dordrecht.

Muinonen K. and Bowell E. (1993) Asteroid orbit determination using Bayesian probabilities. *Icarus, 104,* 255–279.

Muinonen K., Virtanen J., and Bowell E. (2001) Collision probability for Earth-crossing asteroids using orbital ranging. *Cel. Mech. Dyn. Astron., 81,* 93–101.

Öpik E. J. (1976) *Interplanetary Encounters: Close-Range Gravitational Interactions.* Elsevier, New York. 155 pp.

Sitarski G. (1968) Approaches of the parabolic comets to the outer planets. *Acta Astron., 18,* 171–195.

Sitarski G. (1999) How to find an impact orbit for the Earth-asteroid collision. *Acta Astron., 49,* 421–431.

Valsecchi G. B., Milani A., Gronchi G. F., and Chesley S. R. (2000) The distribution of energy perturbations at planetary close encounters. *Cel. Mech. Dyn. Astron., 78,* 83–91.

Valsecchi G. B., Milani A., Gronchi G. F., and Chesley S. R. (2001) Resonant return to close approach: Analytical theory. *Astron. Astrophys.,* in press.

Yeomans D. K. and Chodas P. (1993) *Periodic Comet Shoemaker-Levy 9 (1993e).* IAU Circular 5807.