

ASTEROID IMPACT MONITORING

Andrea Milani

*Dipartimento di Matematica, Università di Pisa,
Largo Pontecorvo 5, 56127 Pisa, Italy*

(Received: June 18, 2006; Accepted: XXX, 2006)

SUMMARY: Some asteroids and comets with Earth-crossing orbit may impact our planet, thus we need to be able to identify the cases which could have a dangerous close approach within a century. This must be done as soon as such an asteroid is discovered, allowing for follow up observations which might contradict the impact possibility, and in the worst case to organize mitigation, possibly including deflection. The mathematical problem of predicting possible impacts, even with very low probabilities, has been solved by our group in the last few years. This paper presents the basic theory of these impact prediction, and discusses how they are practically used in the impact monitoring systems now operational, in particular the CLOMON2 robot of the Universities of Pisa and Valladolid.

Key words. Asteroids – impacts – orbit determination

1. THE PROBLEM OF IMPACT WARNING

When an asteroid/comet has just been discovered, its orbit is weakly constrained by the available astrometric observations and it might be the case that an orbit with impact on the Earth in the near future (the next 80 ~ 100 years) cannot be excluded. Of course if additional observations are obtained, the uncertainty of the orbit decreases and the impact may become incompatible with the available information. Thus, if we are aware that an impact is possible, it is enough to spread this information to the astronomers to convince them to attempt additional observations of the same object. On the contrary if this piece of information is not available, or is made available too late, when the asteroid has been lost (having become too dim and/or with too uncertain ephemerides to allow for targeted recovery), then the impact risk will remain until the same asteroid is accidentally recovered. This accidental recovery might be too late, e.g., minutes before the impact, for any mitigation action.

The above situation, full of contradictions, can be entangled in only one way: all the asteroids, immediately after being discovered and well before they could be lost, need to be “scanned” for possible impacts in the near future. If impacts are possible, this information must be broadcast to the astronomers as soon as possible. This is the goal of *impact monitoring*. It is somewhat surprising that this was not really possible until late 1999, when the first impact monitoring system, the CLOMON software robot of the University of Pisa, became operational. For many years, even after the risk of impacts of asteroids and comets on our planet had been identified and its probability estimated, even while dedicated surveys were scanning the sky to discover as many *Near Earth Asteroids* (NEA) and comets as possible, the algorithms to scan a given, known asteroid for possible impacts were not effective enough. By using the linear theory (see Section 2.1) of impact prediction it was indeed possible to identify impact possibilities with comparatively high probability, of the order of $10^{-3} \sim 10^{-4}$. On the other hand, if the possible event was the impact of an asteroid with diameter exceeding 1 km, which would result in an explosion

with a yield in excess of 20,000 Megatons, even a probability of the order of $10^{-6} \sim 10^{-7}$ cannot be considered negligible, and to omit to follow up such dangerous asteroid could be considered a crime. On the contrary, unfounded announcements of possible impacts, as in the case of 1997 XF₁₁ in March 1998, can undermine the credibility of the scientific community involved and thus make it more difficult to obtain the resources necessary in a serious case.

For the first time in 1999 our group was able to issue a warning of a possible impact for asteroid 1999 AN₁₀ (Milani et al. 1999). Although the impact probability announced for the year 2039 was $\simeq 10^{-9}$, so small that it would not need to be cause of concern for the public, it was the proof that the mathematical problem had been solved¹. The new methods introduced for the 1999 AN₁₀ case led to the establishment of the impact monitoring system CLOMON later in the same year.

In 2002 the first generation impact monitoring system CLOMON was replaced by the second generation CLOMON2 in Pisa (duplicated also at the University of Valladolid) and SENTRY at NASA's Jet Propulsion Laboratory. These two independent systems, whose output is carefully compared to increase reliability, now guarantee that the potentially dangerous objects are identified very early (within hours from the dissemination of the astrometric data) and followed up until the observations succeed in contradicting the possibility of an impact. Note that during the time span over which these observations are obtained the announcement that some asteroid has the possibility of impacting must be in full view of the public, and in practice it is posted on a web page². This is essential to communicate the need of observations to the astronomical community and also reassures the public that no information on impact risk is withheld.

In case the impact possibility remains for a long time, as it is currently the case for asteroids (99942) Apophis and 2004 VD₁₇ which have been on the *risk pages* of CLOMON2 and SENTRY for more than one year, it is reasonable to begin planning for the mitigation actions which may become necessary if the later observations were to confirm, rather than contradict, the impact. Although the probability of an impact is small for both cases, we must have a technologically possible method to deflect these two asteroids which can be used if necessary; otherwise, the practical utility of the surveys and of the impact monitoring itself would be cast into doubt.

The purpose of this paper is to give a general outline of the mathematical theory and of the computational methods used in the impact monitoring systems, in particular in CLOMON2.

2. TARGET PLANES

The geometry of the encounters with a planet can be described in terms of a target plane, a plane in 3-D space through the center of the target planet, e.g., the Earth, orthogonal to the direction of the relative velocity of the approaching small body. In this context, an impact can be described as an orbit containing a target plane point inside the planet cross section on the same plane.

There are two ways to give a rigorous definition of target plane. The simplest definition is the *Modified Target Plane (MTP)*: it is obtained (Milani and Valsecchi, 1999) by considering the time \bar{t} at which the small body orbit has a relative minimum of the distance from the planet's center of mass (CoM). Let \mathbf{d} and \mathbf{v} be the planetocentric position and velocity vectors of the asteroid at the time \bar{t} ; because the distance is minimum, $\mathbf{d} \cdot \mathbf{v} = 0$. The MTP is the plane, containing $\mathbf{0}$ (the CoM) and normal to \mathbf{v} . This plane contains the point \mathbf{d} , which represents the trace of the close approach on the MTP. A complete description of the close approach is obtained by assigning two coordinates on the MTP, two angles defining the orientation of the MTP, the size of the velocity $v = |\mathbf{v}|$ and the time \bar{t} . The cross section of the planet on the MTP is a disk centered at $\mathbf{0}$ and with the radius R of the planet; if the minimum distance d at time \bar{t} is less than R there is an impact³.

The other definition, called in the literature either just *Target Plane (TP)* or *b-plane*, uses the same vectors \mathbf{d} and \mathbf{v} describing the state at the closest approach time \bar{t} to compute a planetocentric 2-body approximation of the orbit. If, as it is generally the case, such planetocentric 2-body orbit is hyperbolic, then the TP is the plane orthogonal to the asymptote of the hyperbola, that is to the vector \mathbf{u} which is the limit for $t \rightarrow -\infty$ of the planetocentric velocity along the hyperbolic trajectory; the size $u = |\mathbf{u}|$ is the velocity of escape. The point \mathbf{b} representing the close approach on the TP is the intersection of the asymptote with the TP; its size is the impact parameter b , which is in fact larger than the minimum distance d by a factor

$$\beta = \frac{b}{d} = \sqrt{\frac{v^2 d}{v^2 d - 2GM}}$$

where GM is the gravitational universal constant multiplied by the planet mass. A complete description of the close approaching orbit is obtained by assigning two coordinates ξ, ζ on the TP, two angles θ, ϕ defining the orientation of the TP, the size of the velocity of escape $u = |\mathbf{u}|$ and the time \bar{t} (Greenberg et al., 1988). Note that on the TP the impact cross

¹The impact probability was later found to be even higher for impacts in 2044 and 2046; a few months later, the *precovery* of 1999 AN₁₀ in old plates going back to 1955 allowed to contradict whatever possibility of impact in the 21st century.

²<http://newton.dm.unipi.it/neodys> for CLOMON2 and <http://neo.jpl.nasa.gov> for SENTRY.

³This is true in the approximation in which the planet surface is a sphere; the oblateness of the planet is generally irrelevant for the question of the possibility of an impact, although it may matter when predicting the point of impact.

section is a disk of radius

$$B = R \sqrt{1 + \frac{2GM}{Ru^2}}$$

larger than the physical radius of the planet by a factor accounting for gravitational focusing.

The transformation between the two planes is complicated, because the velocity \mathbf{v} at the close approach is rotated by an angle $\gamma/2$ around the axis of the planetocentric angular momentum. The angle γ measures the total deflection from the incoming to the outgoing asymptote and can be computed by

$$\sin(\gamma/2) = \frac{GM}{v^2 d - GM}.$$

The transformation of coordinates rotating and rescaling the MTP unto the TP is not canonical, thus it is impossible to use the Hamiltonian formalism in the context of the TP (Tommei 2006a; Tommei 2006b). Moreover, the choice of the coordinates on the two planes can be done in different ways, and this has also to be accounted in the transformation.

From an abstract point of view, it does not matter in which way we select a representative vector for a given close approach, provided it is a smooth function of the orbit initial conditions: any other coordinate system can be obtained by a smooth coordinate transformation. However, some coordinate systems are more equal than others, because the propagation of the uncertainty is easy to do in a linear approximation, by using the differential of the transformations, and a coordinate change with large higher order derivatives introduces strong limitations in the applicability of the linearization. Thus there is a significant advantage in using the TP with respect to the MTP, because gravitational focusing introduces a deformation which is more nonlinear where gravity is stronger, that is near collisions.

2.1 Linear Predictions on Target Planes

For a given asteroid, and a set of orbital elements $\mathbf{x} \in \mathbb{R}^6$ at epoch t_0 , there is a unique orbit which can be accurately propagated for some time span⁴ after t_0 . For each close approach to the Earth, occurring within this time span, there is at least one point $\mathbf{y} \in \mathbb{R}^2$ which is the *trace* of this orbit on the target plane⁵. To avoid useless geometric complications, we consider *close* only approaches with a distance from the planet CoM not exceeding some value d_{max} ; practical values for d_{max} range between 0.05 and 0.2 Astronomical Units (AU), thus the target planes are in fact disks with a finite radius.

The classical procedure uses as *nominal* initial conditions at t_0 the solution \mathbf{x}^* of a *least squares fit*. The details of the classical procedure to compute least squares orbits are not within the scope of this paper, just to assign the notation we call $\boldsymbol{\xi} \in \mathbb{R}^m$ the vector of the residuals (observed minus predicted) for all the available observations, both optical and radar, with $m > 6$. Then we use a *target function* $Q = \boldsymbol{\xi}^T W \boldsymbol{\xi}/m$ where W is a symmetrical positive definite weight matrix, accounting for the uneven accuracy of observations from different observatories⁶. Also correlations and biases can be accounted for (Carpino et al., 2003).

The *normal equations* $C_{\mathbf{x}} \Delta \mathbf{x} = K$ for the (linearized) correction $\Delta \mathbf{x}$ to the initial conditions \mathbf{x} are defined by the *design matrix* $B = \partial \boldsymbol{\xi} / \partial \mathbf{x}$

$$C_{\mathbf{x}} = B^T W B \quad , \quad K = -B^T W \boldsymbol{\xi}$$

and are solved by $\Delta \mathbf{x} = \Gamma_{\mathbf{x}} K$, using the *covariance matrix* $\Gamma_{\mathbf{x}} = C_{\mathbf{x}}^{-1}$. The right hand side K is proportional to the gradient of the target function Q . The corrections $\Delta \mathbf{x}$ are applied in an iterative procedure of *differential corrections* until convergence to \mathbf{x}^* corresponding to a (possibly local) minimum Q^* of Q . The nominal solution \mathbf{x}^* should not be understood as “the true solution”, but just as a representative central point in a *region of confidence* of possible solutions. To describe the confidence region, note that $2C/m$ is an approximation, applicable for small residuals $\boldsymbol{\xi}$, to the matrix of second partial derivatives of Q . Thus the target function can be approximated in a neighborhood of \mathbf{x}^* by the expansion

$$Q(\mathbf{x}) = Q^* + \Delta Q = Q^* + \frac{1}{m} (\mathbf{x} - \mathbf{x}^*)^T C_{\mathbf{x}} (\mathbf{x} - \mathbf{x}^*) + \dots \quad (1)$$

where the dots stand for higher order terms (Milani 1999). If these are neglected, the *penalty* ΔQ can be approximated by a quadratic form in the \mathbf{x} variables, and the *confidence region* where its value is small can be approximated by a *confidence ellipsoid* described by the inequality

$$m \Delta Q(\mathbf{x}) \simeq (\mathbf{x} - \mathbf{x}^*)^T C_{\mathbf{x}} (\mathbf{x} - \mathbf{x}^*) \leq \sigma^2 \quad (2)$$

where σ^2 is a confidence parameter, related to the χ^2 test parameter of statistics.

As the nominal solution \mathbf{x}^* is surrounded by a 6-dimensional confidence region of acceptable solutions, the trace point $\mathbf{y}^* = \mathbf{g}(\mathbf{x}^*)$ determined by the propagated nominal orbit on the target plane of some encounter is surrounded by a 2-dimensional confidence region. To compute an approximation,

⁴In the current impact monitoring systems, the orbits are generally propagated for 80 to 100 years. Only for some orbits, determined in an especially accurate way, it is meaningful to propagate for longer time spans.

⁵It is possible that a close approach has multiple local minima of the distance to the planet CoM, in which case there are multiple target plane trace points. Reducing d_{max} can often eliminate such complications.

⁶Weighting is also necessary to combine dimensionally different quantities like the angles of the optical observations and the range/range rate of the radar. Radar observations, with relative accuracies in the range $\simeq 10^{-9}$, must have more weight, but the relative weighting with respect to the optical data is a critical problem.

we use the differential of the map $\mathbf{g}(\mathbf{x})$ providing the target plane trace (Milani and Valsecchi 1999). Indeed, the trace point is obtained by finding the time $t_c(\mathbf{x})$ of the crossing of the target plane for each orbit with initial conditions \mathbf{x} (in a neighborhood of \mathbf{x}^*). By using Cartesian geocentric coordinates ξ, η, ζ such that $\eta = 0$ is the target plane, the equation $\eta(t, \mathbf{x}) = 0$ implicitly defines the crossing time $t(\mathbf{x})$ as a differentiable function, thus $\xi(t(\mathbf{x}), \mathbf{x})$ and $\zeta(t(\mathbf{x}), \mathbf{x})$ are differentiable too. Using the differential $D\mathbf{g}(\mathbf{x}^*) = \partial(\xi, \zeta)/\partial\mathbf{x}$ we can compute the covariance and normal matrix of the \mathbf{y} prediction by the linear covariance propagation formula

$$\Gamma_{\mathbf{y}} = D\mathbf{g} \Gamma_{\mathbf{x}} (D\mathbf{g})^T, \quad C_{\mathbf{y}} = \Gamma_{\mathbf{y}}^{-1}$$

defining the *confidence ellipse* on the target plane

$$(\mathbf{y} - \mathbf{y}^*)^T C_{\mathbf{y}} (\mathbf{y} - \mathbf{y}^*) \leq \sigma^2 \quad (3)$$

with the same confidence parameter σ^2 . This formalism is always applicable because the trace function is differentiable, how good is the quadratic approximation of eq. (3) to the full confidence region is another matter. However, if this approximation is adequate, then the *possibility of an impact* can be studied by looking for intersections of the confidence ellipse of eq. (3) with the impact cross section.

By using a Gaussian formalism, which associates to the family of concentric ellipsoids of eq. (2) a probability density function constant on each ellipsoid, it is possible to define a probability density also on the target plane. In the linear approximation corresponding to the differential $D\mathbf{g}(\mathbf{x}^*)$, the probability density of \mathbf{y} is also Gaussian and has constant values on the ellipses of constant σ of eq. (3). Then it is possible to compute a probability integral on the impact cross section, which gives an estimate of the impact probability.

The formalism above is well known for the applications to the navigation of interplanetary spacecraft, a case in which the assumptions of small confidence regions and therefore the applicability of linearization are well founded, simply because spacecraft are tracked as much as necessary to maintain their orbit determination in a linear regime. To estimate the probability of impact of asteroids is much more difficult, due to nonlinearity.

2.2 The Sources of Nonlinearity

There are three main reasons why the impact predictions are nonlinear, thus the target plane confidence regions are poorly approximated by ellipses.

First, when an asteroid/comet has been recently discovered, its orbit is weakly constrained by the available observations, and some of the eigenvalues of the covariance matrix $\Gamma_{\mathbf{x}}$ are large, and some of the axes of the confidence ellipsoids of eq. (2) are long. Then the approximation neglecting the higher order terms in eq. (1) is poor, and the shape of the confidence region containing values of \mathbf{x} compatible

with the observations is very different from an ellipsoid: the trace of this region on the target plane is very different from an ellipse, and the probability density of \mathbf{y} is very different from a Gaussian.

Second, the propagation of the initial conditions $\mathbf{x}(t_0)$ to a time t close to the epoch of the close approach, many years later, is also nonlinear. The formula for linear propagation of the covariance matrix still applies: if A is the 6×6 *state transition matrix*, the matrix of partial derivatives of the elements $\mathbf{x}(t)$ with respect to the elements $\mathbf{x}(t_0)$, then the covariance and normal matrices at time t are

$$\Gamma_{\mathbf{x}(t)} = A \Gamma_{\mathbf{x}(t_0)} A^T, \quad C_{\mathbf{x}(t)} = \Gamma_{\mathbf{x}(t)}^{-1}. \quad (4)$$

In many cases, the confidence ellipsoid defined by $C_{\mathbf{x}(t)}$ is a very poor approximation even when the one defined by $C_{\mathbf{x}(t_0)}$ is a good one. This because, as time passes, the longest axis of the confidence ellipsoid becomes longer and longer. In a 2-body propagation, the largest eigenvalue of the covariance matrix grows quadratically with time; the corresponding eigenspace is approximately aligned with the orbital velocity. Thus the longest axis of the confidence ellipsoid grows linearly with time and is close to the along track direction: when this length is a non negligible fraction of the length of the orbital ellipse the nonlinear confidence region should bend to follow the curvature of the ellipse. In a full N-body propagation, the longest axis of the confidence ellipsoid can grow exponentially with time, with a characteristic *Lyapounov time* of the order of the time interval between two close approaches, which could be as short as 3-4 years.

This effect is amplified because the close approaches can in practice be recorded by their target plane trace only provided the minimum distance d is below some value d_{max} . Then the nominal orbit could have, in this sense, no close approach and no target plane trace at all, in which case the confidence ellipse does not even exist. Even when the nominal trace point exists, the confidence ellipse may extend well outside the disk of radius d_{max} and the real confidence region may contain multiple connected components on the target plane.

The third reason is the gravitational focusing, taking place on the MTP specifically for encounters very close to the Earth CoM, e.g., for the impacts. This effect is generally small, but it affects precisely the orbits we are most interested in studying. Thus it is convenient to remove this source of nonlinearity by choosing the TP as representative plane.

2.3 Minimum Orbital Intersection Distance

A convenient reference system for the geocentric position on the TP (ξ, η, ζ) is obtained by aligning the negative ζ -axis with the projection of the Earth's heliocentric velocity v_{\oplus} , the positive η -axis with the geocentric velocity (i.e., normal to the TP), and the positive ξ -axis in such a way that the reference system is positively oriented. With this frame of reference the TP 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 “miss time” of the target plane crossing ($\eta = 0$) is $\Delta t = \zeta / (v_{\oplus} \sin \theta)$, where θ is the angle between \mathbf{u} and v_{\oplus} : a positive ζ implies that the asteroid is “early” at the date with the Earth, $\zeta < 0$ means the asteroid has been late.

On the b -plane the ξ coordinate is the minimum distance that can be obtained by varying the timing of the encounter. This distance is closely related to the length known as the *Minimum Orbital Intersection Distance (MOID)*, that is the minimum separation between the osculating ellipses as curves in 3-dimensional space, without regard to the phase on each of the two. Note that the approximation of the MOID with the ξ coordinate is valid only in the linear approximation and can break down for distant encounters (e.g., beyond several lunar distances).

To compute the MOID exactly it is possible to write a system of two equations in the two anomalies parameterizing the points on the two orbits. The solutions of this system are couples of values of the anomalies corresponding to points with stationary distance. By a suitable transformation, using either the true of the eccentric anomaly, the system of equations can be expressed in a polynomial form, and it can be solved, e.g., with the resultant method, giving a single polynomial equation of degree 20. The real roots of this equation, which cannot be more than 16, correspond to stationary points for the distance, some of which can be local minima. There are known examples with up to 12 stationary points and up to 4 local minima (Gronchi, 2002).

The most typical case of double minimum is the one of asteroids with high inclination and the two nodal points (intersections of the osculating ellipse with the ecliptic plane) both near the Earth orbit, e.g., 1999 AN₁₀. Then there are two local minima of the distance, both occurring for points on the two orbits near the line of nodes (joining the two nodal points), a maximum point and 3 saddle points⁷. The implication is that the ξ coordinate may refer to each one of the two local minima, possibly with very different values: for 1999 AN₁₀ a very close approach is possible near the ascending node, only a moderately close approach (beyond the lunar distance) is possible near the descending node.

The role of the MOID in impact monitoring is to select, among the large number (thousands, even tens of thousands) of close approaches possible for a given asteroid, the ones which could be very close. If the TP coordinates have a small value of ξ and a large value of ζ , then the encounter has not been close, but another orbit with slightly different orbital phase might get in time to the date with the Earth at the local MOID point, if the value of ζ has a large enough uncertainty. In a linear approximation, applicable to very close encounters, the confidence ellipse has a major axis almost parallel to the ζ axis and a minor axis almost parallel to the ξ axis, that is expressing the uncertainty of the local MOID value.

When the orbit, as determined at the initial

time t_0 , has comparatively large uncertainties, the uncertainty of the MOID value should be computed in a more accurate way. Indeed it is possible that the MOID of the nominal orbit \mathbf{x}^* is large and still a very small the value for the MOID, even 0, is compatible with the observations.

There are mathematical difficulties in the estimate of the uncertainty of the local MOID value d_M because in fact $d_M(\mathbf{x})$ is not a single valued function and anyway is not differentiable when the value is 0 (because of the square root needed to obtain the distance from the smooth distance squared). This problem has been solved by defining a *signed local MOID* which is generally a smooth function: the sign is assigned by using the orientation of the triple of vectors formed by the vector joining the position of the planet and the asteroid at the local MOID point and by the tangent vectors of the two orbits. It can be shown that this modified local MOID is a smooth function of the orbital elements, even for 0 distance, outside a small set of singularities, thus it is possible to use the differential to provide a linear approximation to estimate a confidence interval. (Gronchi and Tommei 2006). In this way it is possible to identify the cases in which very close approaches are possible although the nominal orbit has a large MOID.

3. VIRTUAL ASTEROIDS

When an asteroid has only recently been discovered, or anyway has been observed only for a short time span, we do not know “the orbit” of the real object, but rather we can describe our (lack of) knowledge by thinking of a swarm of *Virtual Asteroids (VA)*, with slightly different orbits all compatible with the observations, that is belonging to the confidence region. The reality of the asteroid is shared among the virtual ones, 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 VA 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, which would be inappropriate over the entire confidence region. Note that the nominal orbit is just one of the VA, and is not extraordinary in this context.

The reason for using a swarm of VA is that they are a set of orbits, which can be propagated one by one, representing the totality of orbits compatible with the observations much better than the nominal solution alone. Moreover, by propagating together with the orbits the corresponding state transition matrices, we can use a linear approximation in a neighborhood of each VA. Every differentiable function can be locally approximated by its differential: of course it is not easy to decide how many such points are needed to keep up with strong nonlinearities. On the other hand, the N-body problem not

⁷See figures showing the location of the stationary point and the level lines of the distance on the NEODYs web site at <http://newton.dm.unipi.it/cgi-bin/needys/neoibo?objects:1999AN10;statpts;gif>

being integrable, there is no way to compute globally the totality of orbits corresponding to the confidence region; only a finite set of orbits can be numerically propagated, and the number of such orbits cannot be huge, otherwise the CPU time needed for the computation might become comparable to the time span from the discovery to the possible collision.

Thus the critical issue is how to sample the confidence region in an efficient way, that is with few orbits⁸ but selected in such a way that they are as representative as possible of the different possible orbits. There are two classes of sampling methods used in the selection of VA: the random, or Monte Carlo (MC) methods, and the geometric sampling methods, in which the sampling takes place on the intersection of a geometric object, a differentiable manifold, with the confidence region.

The MC methods directly use the probabilistic interpretation of the least squares principle. Since the orbit determination process yields a probabilistic distribution 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 (Chodas and Yeomans, 1996). This can be implemented in different ways, by using a random number generator to sample an assumed probability density either in the space of the elements \mathbf{x} , or in the space of all residuals $\boldsymbol{\xi}$, or in a subset of the residuals (*statistical ranging*, Virtanen et al., 2001).

When the computational resources are not an issue and the probabilistic error models are reliable, the MC methods are more rigorous and complete, thus they are often used for checking the results once a case of possible impact has been identified. If by impact monitoring we mean checking all newly discovered, or anyway re-observed, asteroids for the possibility of a future impact, then computational complexity is the main concern and the MC methods are too slow, at least for the current computer hardware. Then the geometric sampling methods have to be used. In this paper we will concentrate on the one-dimensional sampling methods, in which the geometric object is a smooth line sampled by a regular sequence of intervals. More complex methods, such as 2-dimensional ones using a 2-manifold and sampled with a Delaunay triangulation, have been proposed and are being studied (Tommei 2005) but are not yet being used in operational impact monitoring.

3.1 The LOV as Geometric Sampling

As discussed in Section 2.2, some years after the epoch of initial conditions the confidence region becomes stretched in the along track direction. Since the goal of impact monitoring is to warn of possible impact with a long warning time (typically decades), the best way to sample the confidence region is by defining a curve which intuitively can be the "spine"

of such an elongated confidence region.

To give a formal definition, let us start again from the linear approximation, that is from the normal equation $C_{\mathbf{x}} \Delta \mathbf{x} = K$. The matrix $C_{\mathbf{x}}$ has a smallest eigenvalue λ_1 , corresponding to the longest axis $\sigma_1 = 1/\sqrt{\lambda_1}$ of the confidence ellipsoid of eq. (2) for the parameter $\sigma = 1$. Let \mathbf{v}_1 be the eigenvector corresponding to λ_1 : it represents the *weak direction* for orbit determination.

The *Line Of Variation* (LOV) is the curve in the space of initial conditions \mathbf{x} on which K is parallel to \mathbf{v}_1 , that is, the gradient of Q is along the weak direction (Milani et al., 2005a). In the linear approximation the LOV coincides with the long axis of the confidence ellipsoid, but in general it is a curved line. An explicit analytical expression to compute the LOV is not available, but there is an efficient iterative procedure to compute a point on the LOV starting from an arbitrary first guess \mathbf{x}_0 . The algorithm is a variant of differential corrections in which, at each step, the correction $\Delta \mathbf{x}$ is reduced by removing the component along \mathbf{v}_1

$$\Delta \mathbf{x} = \Gamma_{\mathbf{x}} K - (\mathbf{v}_1^T \Gamma_{\mathbf{x}} K) \mathbf{v}_1 .$$

At convergence this iterative *constrained differential corrections* method gives a point on the LOV. To sample the LOV it is possible to use a step-like propagation, starting from the nominal \mathbf{x}^* which belongs to the LOV, moving along the weak direction for a short step $\mathbf{x}' = \mathbf{x}^* + h \sigma_1 \mathbf{v}_1$, with small h , then performing constrained differential corrections until converging to the LOV point \mathbf{x} ; to it we assign $\sigma = h$ as parameter (although this is an approximation of the unknown parameterization of the LOV as a smooth line, exact only in the linear case). With this procedure it is possible to compute a set of LOV points centered at the nominal and regularly spaced in the parameter σ .

The problem with the LOV definition is that it is not independent from the choice of coordinates in the space of initial conditions \mathbf{x} . Under a coordinate change with partial derivatives matrix A , the normal and covariance matrix change as in eq. (4), the eigenvalues and eigenvectors are not invariant: e.g., the LOV for Keplerian elements and the one for Cartesian coordinates are not the same. Thus we have to choose the coordinate system which gives the LOV more representative of the set of orbits filling the confidence region, and this depends upon the purpose for which we have built the sampling. For impact monitoring, at least when we are interested in predictions for times much later than the initial conditions, the most important changes in the orbit elements are the ones in semimajor axis, and the metric should be chosen accordingly.

3.2 The LOV trace on the Target Planes

Once the LOV sampling has been computed, we have a set of VA \mathbf{x}_i for $1 \leq i \leq 2k + 1$, corresponding to values $(i - k - 1) \cdot h$ of the parameter

⁸In the impact monitoring practice with current computer hardware, this means between a few thousands and a few tens of thousands of orbits for each asteroid.

σ . By propagating each of the VA orbits for a given time span (80 ~ 100 years) we record for each VA all the close approaches to the Earth within the distance d_{max} . Each close approach is represented by at least one trace point (ξ, ζ) on the TP; the confidence ellipse is also computed, with a major semiaxis of length s (stretching) and a minor one w (width).

Up to this point the procedure is the same, whatever the sampling method. However, the LOV sampling is not just a set of points, we can exploit the fact that they sample a smooth line: the trace of the LOV on the TP is also a smooth line. Let us suppose two consecutive VAs, \mathbf{x}_i and \mathbf{x}_{i+1} , have TP trace points \mathbf{y}_i and \mathbf{y}_{i+1} straddling the Earth impact cross section, such that the trace point \mathbf{y}_i is “early”, that is $\zeta_i > 0$, while \mathbf{y}_{i+1} is late, $\zeta_{i+1} < 0$. Then there is one point $\mathbf{x}_{i+\delta}$ on the LOV (as a continuous curve) corresponding to the parameter $\sigma = (i - k - 1 + \delta)h$ with $0 < \delta < 1$ such that $\zeta_{i+\delta} = 0$: this must necessarily occur provided the trace of the segment of the LOV between \mathbf{y}_i and \mathbf{y}_{i+1} lies entirely within the distance d_{max} from the Earth CoM. This is the first instance of the *principle of simplest geometry* we will further discuss in the next Section: cases with extreme nonlinearities violating the continuity condition implicitly used above for the function $\zeta(\sigma)$ are possible, but this is less frequent than the simple case in which the segment joining \mathbf{y}_i to \mathbf{y}_{i+1} is not much curved and behaves qualitatively like a straight line.

The point $\mathbf{x}_{i+\delta}$ on the LOV, which was not among the original set of VAs, can be computed as follows. First an approximate value of the real parameter δ is computed by using a *regula falsi* step, in such a way that it approximates the point giving a local minimum of the distance to the Earth on the TP. Then a step of length δh is performed with the same formula $\mathbf{x}' = \mathbf{x}_i + \delta h \sigma_1 \mathbf{v}_1$ used for the original VA sampling, finally a constrained differential iterative procedure is used to correct to a LOV point which is taken as $\mathbf{x}_{i+\delta}$. If the corresponding TP trace $\mathbf{y}_{i+\delta}$ is not the minimum distance point along the LOV trace the procedure is iterated. At convergence we obtain a LOV point of (local) minimum of the close approach distance; if the continuity condition is satisfied, this regula falsi iteration has guaranteed convergence. If this TP point is inside the Earth impact cross section, then $\mathbf{x}_{i+\delta}$ there is a *Virtual Impactor* (VI), that is a connected set of initial conditions leading to an impact (at about the same date). If the point $\mathbf{y}_{i+\delta}$ is outside the impact cross section, but the width w of the confidence ellipse computed by linearizing at $\mathbf{y}_{i+\delta}$ is large enough, then there are intersections of the confidence ellipse and the impact cross section and there is anyway a VI, although the corresponding initial conditions do not belong to the LOV.

By computing the probability density function with a suitable Gaussian approximation centered at $\mathbf{x}_{i+\delta}$ it is possible to estimate the probability integral on the impact cross section, that is the *Impact probability* (IP) associated with the given VI. These computations are always approximate, nevertheless they are better than the estimates done with MC type sampling when the IP is low, because the MC estimates based upon the number of impacting VA

suffer from the uncertainties of small number statistics. In the most extreme cases, a MC sampling is likely not to provide any impacting VA if the number of VAs is less than $1/IP$. On the contrary, the geometric sampling methods described here can detect VIs with IP of the order of $10^{-7} \sim 10^{-8}$ (and sometimes even less) starting from a few thousands VAs on the LOV. The issue of completeness in the searches for VIs is more complex and needs to be discussed in the context of the geometric theory of the next Section.

4. GEOMETRY OF TARGET PLANE TRAILS

To understand the properties of the TP trace of the LOV we need to use the finite sample formed by the trace points \mathbf{y}_i as markers of a geometric structure. To do this, after computing all the close approaches to the Earth for all the VAs \mathbf{x}_i , with $i = 1, 2k + 1$, we sort them by time of the closest approach. Then the recorded close approaches appear to cluster around a discrete set of encounter times, which can be interpreted as passages of the Earth through the point corresponding to the (local) MOID while the asteroid is neither very late nor very early at its MOID point. Each of these clusters of close approaches form a *shower*, and a shower is represented as a set of trace points on the TP.

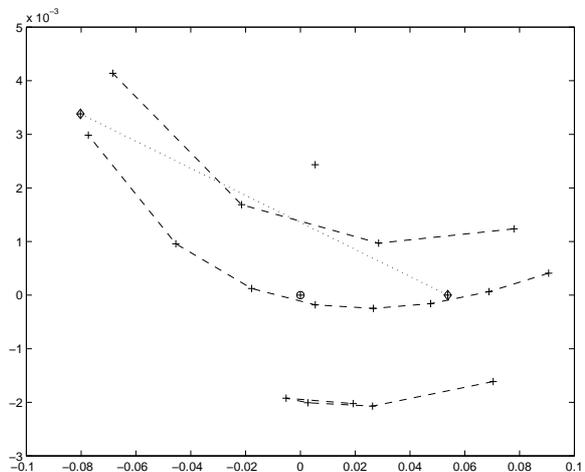


Fig. 1. A single shower with five trails for the asteroid 1998 OX₄ in January 2046. The Earth is depicted by the “⊕”. The two diamonds that are connected by the dotted line actually correspond to a trail with collision, emphasizing that care must be taken when interpolating between solutions. Note the axes are not to scale. These trails have been computed by using only the observations obtained in 1998; the asteroid was recovered in 2002 and the new observations have shown that there is no risk of impact.

In some cases, corresponding to comparatively slow encounters, the situation can be somewhat more complicated, see Figure 6 of Milani et al. (2005b),

but let us assume this decomposition of the set of close approaches in showers has been obtained. Next we decompose each shower into contiguous LOV segments; this is easily obtained by sorting the shower according to the index i . A subset of a shower with contiguous indexes i is a *trail*. In some cases a trail is a *singleton*, formed by just one of the selected VAs. Figure 1 shows the trace on the TP for a shower containing 5 trails, including one singleton and a “doubleton” with just two TP points.

4.1 The principle of simplest geometry

We can conjecture that a trail with $h \leq i \leq k$ corresponds to a continuous set, a segment of the LOV, with a corresponding curve segment of TP trace points joining \mathbf{y}_h to \mathbf{y}_k . Because of the finite sampling, we cannot prove that this must be the case. This hypothesis could be verified by densifying the LOV sampling: if some of the new VA miss the TP, that is do not have a close approach (within d_{max}) around the same date, we cannot exploit the differentiable structure of the LOV. However, if such a TP segment of differentiable curve exists, we can draw very important conclusions.

Let us take as an example the “doubleton” of Figure 1. To interpolate linearly between the two extremes \mathbf{y}_i and \mathbf{y}_{i+1} , as suggested by the dotted line in the Figure, is obviously a very poor approximation: the trails with more TP points do suggest a significant curvature of the TP trace curves. We can use an additional piece of information: the map from the LOV to the TP is differentiable, thus there is a tangent vector to the TP trace curve at the point \mathbf{y}_i . If the angle between this vector and the direction to the origin is $< \pi/2$, this implies that the close approach distance is decreasing for increasing values of the LOV parameter σ at the value σ_i corresponding to \mathbf{x}_i . The same computation can be done at \mathbf{y}_{i+1} , and the close approach distance is found to increase with σ at the value σ_{i+1} . If the TP trace segment joining continuously \mathbf{y}_i to \mathbf{y}_{i+1} exists, then there is for some value of σ in the interval (σ_i, σ_{i+1}) a local minimum of the close approach distance.

In conclusion, if we make the assumption that the behavior of the TP curve is simple, more exactly as simple as it is compatible with the existing decomposition of the shower into trails, we expect to have at least one local minimum of the close approach distance for each trail. This is why we adopt the *principle of simplest geometry*, by which the curve does not exit the TP disk of radius d_{max} : then there need to be at least one minimum of the close approach distance. We can define constructive algorithms for the determination of at least one minimum. Note that it is also an assumption that such minimum is unique for each trail. In the case of the “doubleton” of Figure 1, the minimum is indeed such that there is a VI, but this result cannot be obtained by using a simple linear approximation.

In fact, the shape of the TP trace curve between two given VAs can be much more complex than the simplest geometry, the convergence of the minimum-finding algorithms and/or the uniqueness of the local minimum may fail. However, if the TP

trace curve has extreme nonlinearities over a very short segment of the LOV, that implies the stretching is large and thus the IP is low. This is a robust qualitative argument, that is the VI we can find by using this method have larger IP that the ones we can miss, but unfortunately we have not yet been able to transform it into a quantitative argument, that is in an estimate of the maximum IP of the missed VIs. Thus we do not yet have an analytical estimate of the maximum IP for which we can guarantee completeness in the search for VI.

Another application of the principle of simplest geometry can be appreciated from the trail with 5 TP points appearing near the bottom of Figure 1. Also in this case the first and last TP point of the trail, \mathbf{y}_i and \mathbf{y}_{i+4} , correspond to increasing and decreasing close approach distance, respectively. Thus, if the TP trace curve segment joins \mathbf{y}_i and \mathbf{y}_{i+4} without exceeding the distance d_{max} , then it must have at least one point with minimum close approach distance. However, the behavior of the TP trace curve cannot be approximated by a straight line and the map between the confidence region for the initial conditions \mathbf{x} and the TP points \mathbf{y} is essentially nonlinear, because there is a *fold* line where the differential of the map $\mathbf{x} \mapsto \mathbf{y}$ is degenerate (Milani et al., 2005b). The close approach distance decreases to a minimum, then increases again, but because the TP trace curve “turns back”. This phenomenon is called *resonant return*.

A *return* is a trail with the additional condition that the VA forming it have experienced another close approach between the times of the available observations and the times of the close approaches belonging to the trail. Among the returns there are those such that the intermediate close approach is to the same planet and has occurred near the same local MOID point of the close approaches belonging to the return (e.g., both near the ascending node for a high inclination orbit). That implies the Earth is near the same value of mean anomaly in the intermediate and in the return encounter, i.e., the close approach occurs at about the same date in different years. Also the asteroid needs to be near the same anomaly to be close to the MOID point along its orbit: the time span Δt between the two encounters needs to be close to an integer multiple of the Earth’s orbital period and to an integer multiple of the asteroid period, thus the two orbits must be nearly resonant. This is the motivation for the name “resonant return” (Milani et al. 1999).

We have developed an analytical theory, based upon Öpik formalism for planetary encounters, describing in a qualitative (also in an approximate quantitative way) the shape of the TP trace curves associated to returns, in particular for resonant returns (Valsecchi et al. 2003). This is possible because there is a comparatively simple analytical formula approximating the change in the asteroid semimajor axis resulting from the intermediate encounter, as a function of the coordinates (u, θ, ξ, ζ) . Without going into long details, it is enough to point out that each intermediate encounter can generate as many as four “turning back points” in the TP trace curves of successive encounters. Non-resonant re-

turns to the same planet (e.g., after a close approach at the other node) and even encounters with another planet can also generate turning points. Thus the behavior of the bottom trail of Figure 1 is by no means exceptional, rather generic, and the principle of simplest geometry cannot be used to exclude it. We conclude that the algorithms to find local minima of the close approach distance for each trail must be able to cope with this case, thus they must not assume that the $\mathbf{x} \mapsto \mathbf{y}$ map is well approximated locally by a linear map.

4.2 The algorithms to find the minimum close approach distance

We have shown in the previous section that, by assuming the principle of the simplest geometry, each value of the LOV parameter σ belonging to a segment (σ_h, σ_k) corresponds to a point of the TP trace differentiable curve $\mathbf{y}(\sigma)$ joining the TP points \mathbf{y}_h and \mathbf{y}_k . For each of these σ we can compute the the distance squared to the CoM of the Earth on the TP $b^2(\sigma) = \mathbf{y}(\sigma) \cdot \mathbf{y}(\sigma)$ and its derivative $f(\sigma) = \partial b^2 / \partial \sigma$ with respect to the LOV parameter σ . Then we can scan the set of TP points y_i , $h \leq i \leq k$ to find the couples of consecutive indexes $i, i + 1$ such that the signs of this derivative are discordant $f(\sigma_i) < 0$ and $f(\sigma_{i+1}) > 0$ in such a way that an elementary theorem of calculus ensures there is at least one value of $\sigma_0 \in (\sigma_i, \sigma_{i+1})$ such that $f(\sigma_0) = 0$. Moreover, the classical algorithm of *regula falsi* can be used to find one such value of σ .

The difference with the procedure outlined in Section 3.2 is in that no assumption needs to be done on the direction and curvature of the TP trace curve. Indeed, in the cases of resonant returns, the TP trace curve may never cross the $\zeta = 0$ line, because it “turns back” before crossing it. Thus the minimum distance may be much larger than the local MOID. e.g., the two cases in Figure 1 of the “doubleton” and of the resonant return can be handled without problems. In other cases the curve may turn back *after* crossing the $\zeta = 0$ line, in which case we expect a double minimum of the close approach distance. These different cases need to be handled with an adaptive algorithm, capable of identifying the simplest geometry of the TP trace curve compatible with the available sampling and to take the necessary action, that is selecting additional sampling points to be used as initial conditions for iterative procedures to reach all local minima (Milani et al., 2005b).

4.3 Reliability and completeness of impact monitoring

Whenever one of the TP points \mathbf{y}_0 of local minimum (in the close approach distance along the LOV) is within the Earth’s impact cross section, a VI is found, and we have a *representative* of the VI, that is an explicitly computed set of initial conditions \mathbf{x}_0 such that they are compatible with the observations and lead to a collision at a given date. Whether a simpler algorithm, such as the one described in Sec-

tion 3.2, or the more robust algorithm of Section 4.2, has been used does not matter: once found, the VI representative is a proof that the collision can occur, and the problem is how to associate an IP to the VI. Linearization at \mathbf{y}_0 of some Gaussian probability density is the only algorithm efficient enough to be used in operational impact monitoring, although targeted investigations with method similar to Monte Carlo are possible and are used in especially difficult cases, occurring when there is significant nonlinearity in a neighborhood of the VI representative.

A more subtle case occurs when \mathbf{y}_0 is outside the impact cross section, but the TP confidence ellipse computed by linearizing at \mathbf{x}_0 does contain collisions. In this case an explicit representative of the VI is not available; the linearization can be of questionable accuracy, especially when the width w of the TP confidence ellipse is large. Both including and excluding these cases from the list of VIs is unsafe.

We have developed, for the CLOMON2 impact monitoring system, a method to confirm possible VIs by an iterative procedure which has shown the capability to converge to a VI representative, in most cases in which such a VI exists. It is based on a modified Newton’s method, first proposed in Milani et al. (2000). If \mathbf{y}_0 is the point on the LOV TP trace with minimum distance from the Earth, corresponding to the initial condition \mathbf{x}_0 , but $|\mathbf{y}_0| > B$, we select a point \mathbf{y}'_1 on the TP with $|\mathbf{y}'_1| = B$, e.g., by moving radially. Then we find the point \mathbf{x}_1 in the confidence region near \mathbf{x}_0 with the minimum penalty among those projecting into \mathbf{y}'_1 on the TP, using the differential of the $\mathbf{x} \mapsto \mathbf{y}$ map at \mathbf{x}_0 . Then the TP trace $\mathbf{g}(\mathbf{x}_1) = \mathbf{y}_1$ is computed, and it is not \mathbf{y}'_1 because of the nonlinearity, but by iterating this procedure convergence to a VI representative is possible. The difficult point is defining a criterion to terminate the above iterative procedure when convergence is not achieved. Such “divergence” should provide a good indication that the intersection of the linear confidence ellipse and the impact cross section was a *spurious VI*; see Milani et al. (2005b) for details.

5. OPERATIONAL IMPACT MONITORING

The impact monitoring software robots CLOMON2 and SENTRY have been operational since early 2002. In the years 2002 and 2003 this setup has “solved” more than 100 cases of asteroids having VI, in the following sense. These VI cases have been reported, the astronomical community has taken action by performing follow up observations until enough information was available to exclude the possibility of impacts in this century. In most cases this follow up would have taken place anyway, but the warnings are far from useless, since in the same time span there have been 9 cases which have been lost while still having VIs; however, these were all small asteroids (< 100 m diameter), indicating that the telescopes used for this targeted follow up might have not been large enough. This good result has also been due to the action of the Spaceguard

Central Node (SCN) in promoting the recovery campaigns for VI cases⁹. Currently the asteroids found to have possible impacts are several per week, and it is expected that this rate will further increase when the next generation asteroid surveys will be operational, starting in 2007.

In the most serious cases¹⁰ the two impact monitoring systems cross-check their results before announcing the existence of a VI: this procedure takes typically just a few hours.

Since late 2004 two cases have been the most significant sources of concern for the personnel of the monitoring systems. (99942) Apophis has been on the “risk pages” of CLOMON2 and SENTRY since December 2004, with an estimated IP peaking at 1/37 on December 27, 2004 and then declining as new and more accurate observations were received. Now the orbit is very well determined, also with radar observations, the IP for the remaining 2036 VI is low but it is difficult to contradict this impact possibility because the optical observations are not accurate enough to improve the orbit. 2004 VD₁₇ has an orbit already observed over multiple oppositions and its uncertainty does not grow very fast with time because there are few and shallow close approaches. There is a VI for 2102, again with not very high probability but difficult to be removed because the orbit is already very good. These two cases have for the first time raised the issue of planning for mitigation action: knowing there is some risk does not solve the problem, if the risk does not go away by follow up observations.

REFERENCES

- Carpino, M., Milani, A. and Chesley S.R.: 2003, *Icarus*, **166**, 248.
- Chesley, S.R., Chodas, P.W., Milani, A., Valsecchi G.B. and Yeomans, D.K.: 2002, *Icarus*, **159**, 423.
- Chodas, P.W. and Yeomans, D.K.: 1996, in *The Collision of Comet Shoemaker-Levy 9 and Jupiter*, K.S. Knoll et al. eds., IAU Coll. 156, pp. 1-30.
- Greenberg, R., Carusi, A. and Valsecchi, G.B.: 1988, *Icarus*, **75**, 1.
- Gronchi, G.F.: 2002, *SIAM Journ. Sci. Comp.*, **24**, 61.
- Gronchi, G.F. and Tommei, G.: 2006, *Cel. Mech. Dyn. Ast.*, submitted.
- Milani, A.: 1999, *Icarus*, **137**, 269.
- Milani, A. and Valsecchi, G.B.: 1999, *Icarus*, **140**, 408.
- Milani, A., Chesley, S.R. and Valsecchi, G.B.: 1999, *Astron. Astrophys.*, **346**, L65.
- Milani, A., Chesley, S.R., Boattini, A. and Valsecchi, G.B.: 2000, *Icarus*, **1545**, 12.
- Milani, A., Sansaturio, M.E., Tommei, G., Arratia, O. and Chesley, S.R.: 2005a, *Astron. Astrophys.*, **431**, 729.
- Milani, A., Chesley, S.R., Sansaturio, M.E., Tommei, G. and Valsecchi, G.B.: 2005b, *Icarus*, **173**, 362.
- Tommei, G. : 2005, in *Dynamics of Populations of Planetary Systems*, Milani, A. and Knežević, Z. eds., Cambridge Univ. Press, pp. 259–264.
- Tommei, G.: 2006a, *Cel. Mech. Dyn. Ast.*, **94**, 173.
- Tommei, G.: 2006b, Ph.D. thesis, University of Pisa.
- Valsecchi, G.B., Milani, A., Gronchi, G.F. and Chesley, S.R.: 2003, *Astron. Astrophys.*, **408**, 1179.
- Virtanen, J., Muinonen, K. and Bowell, E. : 2001: *Icarus*, **154**, 412.

Kontrola i praćenje asteroidnih udara

Neki od asteroida i kometa sa putanjama koje seku putanju Zemlje mogu da udare u našu planetu, pa je neophodno da budemo u mogućnosti da identifikujemo slučajeve opasno bliskih prilaza u narednih 100 godina. Ova kontrola mora da se sprovede čim se takav asteroid otkrije, da bi se obezbedilo praćenje i nova posmatranja koja mogu da otklone mogućnost udara, ili da, u najgorem slučaju, omoguće preduzimanje zaštitnih mera, uključujući i eventualno skretanje asteroida sa sudarne putanje. Matematički problem predviđanja mogućih udara, čak i onih sa vrlo malom verovatnošću, rešila je naša grupa u poslednjih nekoliko godina. U ovom radu prikazujemo osnove teorije predviđanja udara i razmatramo kako se ona koristi u savremenim sistemima za praćenje asteroidnih udara u funkciji u ovom trenutku, posebno CLOMON2 robota Univerziteta u Pizi i Valjadolidu.

⁹The number of asteroids lost while still having VIs has increased in 2005, due to the lack of funding which forced the SCN to interrupt its operations.

¹⁰The priority is assigned to the most serious cases by using a numeric metric, the *Palermo Scale*, taking into account the impact probability, the impact energy and the time to the impact date (Chesley et al., 2002).