Impact hazard monitoring: theory and implementation

D. Farnocchia
Jet Propulsion Laboratory, California Institute of Technology
Pasadena, CA 91109, USA
email: Davide.Farnocchia@jpl.nasa.gov

Abstract. We review the most standard impact monitoring techniques. Linear methods are the fastest approach but their applicability regime is limited because of the chaotic dynamics of near-Earth asteroids. Among nonlinear methods, Monte Carlo algorithms are the most reliable ones but also most computationally intensive and so unpractical for routine impact monitoring. In the last 15 years, the Line of Variations method has been the most successful technique thanks to its computational efficiency and capability of detecting low probability events deep in the nonlinear regime. We also present some more recent techniques developed to deal with the new challenges arising in the impact hazard assessment problem. In particular, we describe keyhole maps as a tool to go beyond strongly scattering encounters and how to account for nongravitational perturbations, especially the Yarkovsky effect, when their contribution is the main source of prediction uncertainty. Finally, we discuss systematic ranging to deal with the short-term hazard assessment problem for newly discovered asteroids, when only a short observed arc is available thus leading to severe degeneracies in the orbit estimation process.

Keywords. radiation mechanisms: thermal; celestial mechanics; astrometry; ephemerides; minor planets, asteroids

1. Introduction

Impact monitoring is a crucial component of the mitigation or elimination of the hazard posed by asteroid impacts. Once an asteroid is discovered, it is important to timely identify the risk posed by future Earth encounters so that astronomers can collect additional observations to better characterize the orbit and hopefully rule out the possible impacts. The detection of a possible impact hazard should take place as early as possible to maximize the time available for deflection or other mitigation efforts.

In March 1998 B. G. Marsden, who at the time was the director of the Minor Planet Center, issued IAU Circular 6837 (Marsden 1998), accompanied with a press release. The latter claimed that asteroid 1997 XF11 was going to make a close Earth encounter in 2028, when “... the chance of an actual collision ... (would be) ... small but ... not entirely out of the question ...”. Because of the estimated size of 1997 XF11, a kilometer-size object, an impact would be catastrophic. At that time the Minor Planet Center had a policy of holding onto the data for a time period of about a month before making them available to other users. This policy would have made it impossible for other orbit computing centers to independently assess the 2028 hazard. However, due to the clamor that immediately accompanied the issue, the data were made available to selected groups with a much shorter delay. Then, independent calculations performed at JPL and the University of Helsinki showed that the original claim by Marsden was essentially wrong at that there never was a significant chance of a collision in 2028 (Yeomans et al. 1998).

The 1997 XF11 case had the merit of showing that the asteroid community was still unprepared to deal with the asteroid impact hazard assessment problem. In particular,
it was clear that astrometric data needed to be timely disseminated and that impact prediction calculations had to be independently validated to make sure that the results were transparent and accurate. The situation has much improved since then: the Minor Planet Center distributes astrometric observations of near-Earth asteroids on a daily basis and within hours two independent impact monitoring systems, JPL’s Sentry† and University of Pisa’s NEODyS‡, compute whether or not impacts are possible for new discoveries as well as for asteroids with new observations.

2. The $b$-plane

One of the most common and convenient frameworks to perform the hazard assessment is the so-called $b$-plane (Opik 1951; Kizner 1961; Opik 1976; Valsecchi et al. 2003). Whenever there is a close encounter between an asteroid and the Earth, the $b$-plane is defined as the plane with the Earth at the origin and normal to the asymptotic relative incoming velocity of the asteroid $v_\infty$. The position $b$ of the inbound asymptote determines whether or not there is a collision between the asteroid and the Earth. Grazing impacts occur when $|b| = b_\oplus$, where

$$b_\oplus = R_\oplus \sqrt{1 + \frac{2GM_\oplus}{R_\oplus^2 |v_\infty|^2}},$$

where $GM_\oplus$ and $R_\oplus$ are the dynamical parameter and radius of the Earth, respectively. Note that $b_\oplus$ is scaled from $R_\oplus$ to account for the gravitational focusing of the Earth during the encounter. Impacting trajectories correspond to $|b| < b_\oplus$, which is often referred to as impact cross section.

The coordinates $(\xi, \zeta)$ on the $b$-plane are oriented in such a way that the projection of the heliocentric velocity of the Earth $v_\oplus$ on the $b$-plane is oriented as $-\zeta$ and $\xi = v_\infty \times \zeta / |v_\infty|$. Thus, the $\zeta$ coordinate is related to the along-track position of the asteroid and so reflects how early or late the asteroid is for the Minimum Orbital Intersection Distance (MOID, Gronchi 2005), which is approximated by $|\xi|$ (for more details see Valsecchi et al. 2003). Figure 1 gives a schematic representation of the $b$-plane.

† http://neo.jpl.nasa.gov/risks  
‡ http://newton.dm.unipi.it/neodys

Figure 1. Left: $b$-plane with respect to the asteroid’s trajectory relative to the Earth. Right: $b$-plane coordinates and interpretation. The solid line circle is for the actual size of the Earth, the dashed circle is the impact cross section.
3. Linear impact monitoring

Given a perfect knowledge of an asteroid’s orbit it is possible to establish whether or not there is going to be an impact on Earth. However, orbits are only known within an uncertainty region, whose extent depends on the amount and quality of the observational data (e.g., see Farnocchia et al. 2015b).

The method to be used to perform the hazard assessment depends on the validity of linear approximations when describing the orbital uncertainty. The linear method can be used when the uncertainty of the orbit determined by fitting the astrometric observations and its mapping to the time of the Earth encounter and the corresponding $b$-plane are well approximated by ellipsoids. When these approximations are valid, the uncertainty on the $b$-plane is expressed by an ellipse. Therefore, we have a normal probability distribution $f_b$ defined by a mean $\bar{b}$ corresponding to the nominal $b$-plane prediction and a covariance matrix $\Gamma_b$. The impact probability $IP$ can be computed as integral over the Earth cross section:

$$IP = \int_{|b|<b_0} f_b \, db.$$

This numerical calculation is fast, however it is important to make sure that the linearity assumptions are met. While an observational arc of several days is generally sufficient to attain linearity in the orbit determination process, orbital propagation is likely to introduce nonlinearities. In fact, the position uncertainty tends to stretch along the orbit, which is curved, and so when the uncertainty becomes large enough ellipsoids in Cartesian space are not an adequate representation. Moreover, near-Earth asteroids experience planetary encounters and the differential gravitational pull of the planets may scatter neighboring trajectories. As a rule of thumb the linear method for impact monitoring is reliable for $IP > 0.001$, below this threshold we are generally far from the nominal orbit and so the linear approximation can easily prove inaccurate.

4. Monte Carlo approach

The idea of Monte Carlo methods is to randomly generate a large number $N$ of orbital samples, propagate them in the future, and count the number of impacts $n_{imp}$. The impact probability is simply computed as $IP = n_{imp}/N$. The left panel of Fig. 2 shows the example of asteroid (99942) Apophis for the potential impact of April 2029 based on the observations available in December 2004 when the impact probability was $\sim 3\%$. *† The 2029 impact for Apophis was later ruled out (Chesley 2006; Giorgini et al. 2008; Farnocchia et al. 2013b; Vokrouhlický et al. 2015).*

When the orbit determination is linear and the least squares orbit $\vec{x}$ has an ellipsoidal uncertainty expressed by a covariance $\Gamma_x$, the orbital samples can be randomly generated from the normal distribution with mean $\bar{x}$ and covariance $\Gamma_x$. In case the orbit determination is somewhat nonlinear but it is still possible to find a converging least squares solution, one can iteratively randomly perturb the observational dataset according to the observation uncertainties and add the corresponding least squares solution to the sequence of random orbital samples.

Monte Carlo methods are the most robust and conceptually simple approach but also the most computationally intensive. For instance, to achieve a resolution of $10^{-6}$ in impact probability, at least $10^6$ samples are required. Actually, to deal with Poisson statistics uncertainties, one needs $\sim 4\times10^6$ samples or so. Propagating this large number of samples to scan for possible impacts in the next hundred year would take few days on a single core.

thus making Monte Carlo methods unpractical for routine impact monitoring. However, one can always resort to a Monte Carlo approach for critical cases or when other methods encounter degeneracies and fail.

5. Line of Variations

As shown by the left panel of Fig. 2, when propagating in time orbital uncertainties tend to stretch along the orbit, which on the $b$-plane is approximately represented by the $\zeta$ direction. This feature suggests the possibility of a one-dimensional sampling of the uncertainty region, which is the underlying idea for the Line of Variations (LOV) method (Milani et al. 2005b).

Given the orbital uncertainty resulting from the orbital fit to the astrometry and its corresponding covariance $\Gamma_x$, the weak direction is defined as the unit eigenvector $V$ corresponding to the largest eigenvalue of $\Gamma_x$. It is worth pointing out that the weak direction depends on the coordinates, e.g., Cartesian or Cometary, and the units used for the orbital elements (for more details see Milani et al. 2005a).

In the linear regime, i.e., with an exactly ellipsoidal uncertainty, the LOV is a straight line

$$x(\sigma) = \bar{x} + \sigma V, \quad \sigma \in \mathbb{R}$$

where $\bar{x}$ is the least squares solution. At this point one can generate multiple solutions by taking a finite number of $\sigma$ values covering the desired interval, e.g., 5$\sigma$.

More in general the LOV is not a straight line and multiple solutions are generated iteratively by making sure that they correspond to a minimum in the space orthogonal to the weak direction:

- $x_0 = \bar{x}$;
- $x'_{i+1} = x_i + \Delta \sigma V$, which is on the LOV in the linear approximation;
- $x_{i+1}$ is the least squares solution in the space passing through $x'_{i+1}$ and normal to $V$. 

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{figure2.png}
\caption{Map on the April 2029 $b$-plane for asteroid (99942) Apophis based on the observations available in December 2004 when the impact probability was $\sim 3\%$. The left panel is done by using Monte Carlo points, while the right panel is based on the Line of Variations (LOV) technique. The projection of the LOV on the $b$-plane is $\sim 100$ km wide.}
\end{figure}
For more details on the computation of LOV multiple solutions we refer to Milani et al. (2005a).

Once the multiple solutions on the LOV are generated, they can be propagated in the future and mapped to the b-planes of future encounters. The advantage is that, since the LOV is a one-dimensional manifold, it is possible to interpolate between the multiple solutions. The right panel of Fig. 2 shows the Apophis example already discussed in Sec. 4, but LOV multiple solutions are used instead of Monte Carlo samples. Since the width of the uncertainty region is only about 100 km, the impact probability calculation is a one-dimensional problem and can be performed by integrating between $\sim 1.0\sigma$ and $\sim 1.1\sigma$. When the width of the uncertainty is significant, one can always use a linear approximation in the off-LOV direction to integrate over the impact cross section.

As of today, the LOV method has been the most successful impact monitoring technique because it allows the detection of low probability events in the nonlinear regime with a reasonable computational load. A few thousand multiple solutions are generally sufficient to achieve completeness levels of $10^{-7}$ in impact probability (Milani et al. 2005b), and so the hazard assessment can generally be performed in less than a few hours. For these reasons, the LOV method is the engine of the two automated impact monitoring systems Sentry and NEODyS (Milani et al. 2005b). On the other hand, multiple planetary encounters may stretch and fold the LOV thus making its geometry quite complex and making it possible to miss some pathological cases.

6. The Yarkovsky effect

For most asteroids impact prediction uncertainties result from the uncertainty in the astrometric positions used to compute the orbit. On the other hand, for some objects the dominant source of uncertainty are nongravitational perturbations, especially the Yarkovsky effect (Bottke et al. 2006), which mainly acts as a secular variation in semimajor axis. The reason why nongravitational perturbations induce uncertainty is that they are essentially unknown as they depend on the physical properties of the individual asteroid considered. Because of their small magnitude, these perturbations can generally be neglected when modeling an asteroid’s dynamics. However, when an orbit becomes well constrained and when impact predictions are affected by scattering encounters or long-term propagation, the Yarkovsky effect becomes a key consideration in the impact hazard assessment problem. This is the case of asteroids (101955) Bennu (Milani et al. 2009; Chesley et al. 2014), (99942) Apophis (Chesley 2006; Giorgini et al. 2008; Farnocchia et al. 2013b; Vokrouhlický et al. 2015), (29075) 1950 DA (Giorgini et al. 2002; Farnocchia & Chesley 2014), and (410777) 2009 FD (Spoto et al. 2014).

The simplest way to derive the Yarkovsky effect is direct estimation: if the astrometric dataset is good enough one can directly estimate the Yarkovsky related semimajor axis drift $\frac{da}{dt}$ from the orbital fit (Farnocchia et al. 2013a). Chesley et al. (2003) computed the first ever direct detection of the Yarkovsky effect for asteroid (6489) Golevka and since then we have had few tens additional detections (Vokrouhlický et al. 2008; Farnocchia et al. 2014; Nugent et al. 2012; Farnocchia et al. 2013a; Chesley et al. 2015).

When there is no Yarkovsky signal in the astrometry one can use an asteroid’s physical model as well as general properties of the asteroid population to sample the semimajor axis drift $\frac{da}{dt}$ as (Bottke et al. 2006):

$$\frac{da}{dt} \propto \frac{(1 - A) \cos \gamma}{\sqrt{a \rho D}} f(TI, \omega),$$

where $a$ is the semimajor axis, $A$ the Bond albedo, $\rho$ the bulk density, $D$ the diameter,
The obliquity of the spin axis, and \( f \) a nonlinear function of thermal inertia \( TI \) and rotation frequency \( \omega \). As an example, the left panel of Fig. 3 shows the \( da/dt \) distribution for Apophis as obtained by using the known physical characterization (for details see Vokrouhlický et al. 2015).

Once a semimajor axis drift probability distribution is available, the Yarkovsky effect can be included in the model to provide more reliable predictions. For example, the right panel of Fig. 3 shows the distribution of \( \zeta \) for Apophis on the 2029 \( b \)-plane. The difference between the gravity-only solution and the one that accounts for the Yarkovsky effect is evident and clearly shows the importance of the Yarkovsky effect to attain accurate impact predictions.

To perform the hazard assessment including the Yarkovsky effect there are two options. The first option is to use a Monte Carlo method for both \( da/dt \) and the orbital elements:
- Randomly sample \( da/dt \);
- Compute a best-fitting orbit with Yarkovsky accelerations as defined by \( da/dt \);
- Randomly sample the orbital elements according to the covariance matrix;
- Propagate samples in the future and count the impacts.

The second option is to generalize the Line of Variations technique to a seven-dimensional space (Spoto et al. 2014).

7. Keyholes

Even when a collision is ruled out for an Earth encounter, it is possible that such encounter places the asteroid onto an impact trajectory for a later encounter. Chodas (1999) defined a keyhole as the \( b \)-plane region corresponding to future encounters, e.g., the 2036 keyhole on the 2029 \( b \)-plane for Apophis is formed by the projections on the 2029 \( b \)-plane of the trajectories that impact in 2036.

Though non-resonant returns are possible (Milani et al. 1999), keyholes are generally associated to a resonance: the primary Earth encounter changes the asteroid’s orbital period in such a way that the asteroid completes \( n \) orbit revolutions as the Earth makes \( m \). Valsecchi et al. (2003) shows that the points on the primary \( b \)-plane leading to resonant returns form approximate circles. Therefore, keyholes can be found as the intersection of
Figure 4. Left: 2036 resonant circle for Apophis on the 2029 $b$-plane. The corresponding keyhole is the intersection between the 2029 $b$-plane uncertainty and the circle. Right: Apophis’ keyhole map: for each value of $\zeta$ in 2029 we compute the minimum post-2029 encounter distance. Whenver this distance is less than $1R_\oplus$ we have a possible future impact.

these circles and the orbital uncertainty mapped in the primary $b$-plane. The left panel of Fig. 4 shows the example of the 2036 keyhole for Apophis on the primary 2029 $b$-plane.

Identifying the location of the possible keyholes is very useful for the impact monitoring problem. By densely sampling the LOV we can construct the so-called keyhole map: for each $\zeta$ value on the primary encounter we can compute the minimum distance to the Earth at future encounters. Whenever this distance is less than $1R_\oplus$ we have a possible impact and therefore a keyhole. Scaling from the primary $b$-plane to the one corresponding to the possible impact provides the size of the keyhole. For instance, the 2036 keyhole for Apophis is 600 m wide (Chesley 2006). The right panel of Fig. 4 shows a close-up of the Apophis keyhole map.

A comparison between the keyhole locations and widths and the mapped orbital uncertainty on the primary $b$-plane can be used to perform the hazard assessment, as shown by the right panel of Fig. 3. The impact probability can easily be computed by integrating the $\zeta$ probability density function over the keyhole. This technique has been used for (101955) Bennu (Chesley et al. 2014), (99942) Apophis (Farnocchia et al. 2013b; Vokrouhlický et al. 2015), and (410777) 2009 FD (Spoto et al. 2014).

Keyholes can also be used to inform deflection attempts for hazardous asteroids. In fact, if the keyhole size is small it could be easier to deflect the asteroid before the intervening encounter and move it off the keyhole rather changing the trajectory by several thousand kilometers afterwards. Chodas (2012) performed a statistical study and found that keyhole widths range from less than a kilometer to hundreds of kilometers or even more. While close intervening encounters are likely to scatter nearby trajectories and therefore would enhance deflection attempts, distant encounters between 0.10 and 0.15 au can have the opposite effect of focusing neighboring trajectories.

8. Systematic ranging

When the observation arc is very short, a few hours or so, there are severe degeneracies in the orbit determination process: it is generally not possible to find a unique least squares orbital solution and orbital uncertainties cannot be described by ellipsoids.
However, it may be important to identify potential impactors even with little observational information. In fact, as of August 2015, the only two asteroids detected before an impact, 2008 TC$_3$ and 2014 AA, were only discovered 20 hours before striking the Earth.$\dagger\ddagger$

Orbit determination techniques based on Bayesian inversion theory (Muinonen & Bowell 1993) are suitable to provide a probability assessment of the orbital element space with short observation arcs. The posterior probability density function is given by:

$$f_{\text{post}}(\mathbf{x}) \propto f_{\text{prior}}(\mathbf{x}) \exp\left(-\frac{\chi^2}{2}\right)$$  \hspace{1cm} (8.1)$$

where $\mathbf{x}$ are the orbital elements, $\chi^2$ is the usual chi-squared of the observation residuals against $\mathbf{x}$, and $f_{\text{prior}}$ is a suitable prior distribution. The choice of $f_{\text{prior}}$ is not trivial and to some extent arbitrary (for some discussion see Farnocchia et al. 2015a): possible options are a uniform distribution in Cartesian state or a distribution based on an asteroid population model (e.g., Grav et al. 2011).

To fill the phase space one can rely on the fact that a short arc of observations provides a constraint of the plane-of-sky position and velocity, while the topocentric range $\rho$ and topocentric range rate $\dot{\rho}$ are unconstrained. Therefore, ranging methods are well suited to generate orbital samples. Virtanen et al. (2001) describe statistical ranging, where the topocentric range is randomly sampled at two observation epochs to compute an orbit. Another technique is systematic ranging, which is extensively described in Chesley (2005) and Farnocchia et al. (2015a) and we briefly recall here.

The idea of systematic ranging is to scan a dense grid in ($\rho$, $\dot{\rho}$) and for each fixed grid point compute a constrained least-squares solution, which has a small uncertainty. Thus, for each grid point we have an orbit that can be propagated forward in time, which allows one to identify regions of the ($\rho$, $\dot{\rho}$) plane corresponding to close approaches and impacts. Moreover, the fit to the astrometric observations results in a probability density in ($\rho$, $\dot{\rho}$) via Eq. (8.1). Figure 5 shows the application of systematic ranging to 2014 AA by using all the seven observations acquired before the impact. The resulting impact probability is > 99%, thus showing that systematic ranging would have been effective at predicting the impact.

---


$\ddagger$ http://www.nasa.gov/jpl/asteroid/first-2014-asteroid-20140102

---

Figure 5. Application of systematic ranging to 2014 AA. The left panel shows the eccentricity as a function of topocentric range and topocentric range rate. The right panel shows the RMS of the astrometric residuals, as well as the region of impacting solutions.
The development of an automated system to perform short-arc hazard assessment on newly discovered asteroids is an ongoing project.

9. Conclusions

The last two decades have witnessed major steps forward in tackling the problem of assessing the asteroid impact hazard. The data used in the orbit calculations and impact predictions are now readily available to users and potential impactors are detected within hours to trigger follow-up efforts. Even so, impact monitoring remains a current challenge for the asteroid community. Surveys like LSST (Jones et al. 2009) and GAIA (Hestroffer 2015) will soon start producing data and so the next few years will see a significant increase in the amount and quality of astrometric data. Significant effort will be devoted at ensuring an efficient and prompt data processing, as well as accurate impact predictions. In particular, this includes the development of an ever more reliable statistical treatment of the astrometry (e.g., Chesley et al. 2010; Farnocchia et al. 2015c) and accurate modeling of asteroid dynamics.

Acknowledgements

The author is grateful to Steven R. Chesley and Paul W. Chodas for numerous helpful discussions and to Giovanni B. Valsecchi for his comments during the review process.

This research was conducted at the Jet Propulsion Laboratory, California Institute of Technology under a contract with NASA.

Copyright 2015 California Institute of Technology.

References

Chesley, S., Farnocchia, D., Pravec, P., & Vokrouhlicky, D. 2015, IAU General Assembly, 22, 48872
Chodas, P. W. 1999, BAAS, 31, 1117
Chodas, P. W. 2012, LPI Contributions, 1667, 6471
Farnocchia, D., Chesley, S. R., Tholen, D. J., & Micheli, M. 2014, Celestial Mechanics and Dynamical Astronomy, 119, 301
Farnocchia, D., Chesley, S. R., & Micheli, M. 2015a, Icarus, 258, 18
Gronchi, G. F. 2005, Celestial Mechanics and Dynamical Astronomy, 93, 295
Hestroffer, D. 2015, IAU General Assembly, 23, 712
Kizner, W. 1961, Planet. Space Sci., 7, 125
Marsden, B. G. 1998, IAU Circ., 6837, 1
Yeomans, D., Chodas, P., Muinonen, K., et al. 1998, IAU Circ., 6879, 3