Monthly Notices
of the
ROYAL ASTRONOMICAL SOCIETY

MNRAS **467,** 4817–4840 (2017) doi:10.1093/mnras/stx402

Statistics of impacts among orbiting bodies: a Monte Carlo approach

Aldo Dell'Oro★

INAF, Osservatorio Astrofisico di Arcetri, Largo E. Fermi 5, I-50125, Firenze, Italy

Accepted 2017 February 14. Received 2017 January 2; in original form 2016 May 5

ABSTRACT

In this paper, we describe a method to investigate the statistics of collisions among bodies orbiting around a common central mass but not interacting with each other, like minor bodies of the Solar system. This method can be used to derive the frequency of collisions and the distribution of any dynamical parameter related to the collision circumstances. It is based on deriving the statistics of impacts from a random sampling of the orbital element space of the bodies under investigation. The fundamental approach is not new, but the mathematical framework is completely original and the final procedure is very simple and easily implementable. The motivation behind this work is to overcome all limitations due to the assumptions about the dynamical behaviour of orbits, which the methods developed so far are based on. We show all the theoretical details of the method and its practical usage, including the determination of error. We show a set of examples demonstrating the satisfactory agreement with independent approaches, when they are available. The limits and drawbacks of the methodology are also highlighted.

Key words: minor planets, asteroids: general – methods: numerical.

1 INTRODUCTION

Mutual, catastrophic or non-catastrophic, collisions among minor bodies of the Solar system are one of the most important processes influencing their evolution. Some of the fundamental properties of the Main Belt asteroids, like the distribution of their sizes and spins or the existence and structure of asteroid families, are the product of a complex interplay among collisions and gravitational and non-gravitational processes. For this reason, a deep knowledge of the statistics of impacts in those populations is a fundamental prerequisite for understanding their origin, evolution and present properties, as well as for the study of any process in the Solar system involving collisions like cratering, space weathering, surface refreshing and so on.

Several works have focussed on the statistics of collisions among minor bodies of the Solar system, and in particular among asteroids. The term 'statistics of collisions' refers to the computation of impact rates and the derivation of distributions of any parameter related to the conditions at the moment of impact, mainly the impact velocity. The classical approach to tackling this problem is based on the pioneering works of Öpik (1951) and Wetherill (1967). Starting from the same dynamical assumptions, Kessler (1981) developed an independent analytical formulation, confirming the results of Wetherill (1967). The approach of Wetherill (1967) was then refined by Greenberg (1982) and Bottke & Greenberg (1993), and finally by Bottke et al. (1994), becoming the most widely used method for computing the probability of impact among asteroids and of the distribution of their impact velocity. The method consists of an analytic description of the geometric condition of the crossing of asteroid orbits, and the final expression of the impact rate is an integral that is computed numerically. The Öpik–Wetherill approach is based on some canonical assumptions about the dynamical evolution of the asteroid orbits. While the semi-major axes a, eccentricities e and inclinations I of the osculating orbits are supposed fixed, the rates of variation of the longitudes of the ascending nodes Ω and of the arguments of the pericentres ω are constant, or in other words, the osculating orbits circulate uniformly. The motion of asteroids along their orbits is fully described by Kepler's second law. Such hypotheses are valid as a first approximation for the large majority of Main Belt asteroids, and they are suitable for a good estimation of their mean lifetimes and impact velocity distributions.

A first attempt to overcome the limitations of the canonical conditions was done by Dell'Oro & Paolicchi (1997). Later, Dell'Oro & Paolicchi (1998) proposed a new method where the impact rate is expressed again in terms of an integral to be evaluated numerically, but the formula was based on a completely different and more general analytical approach with respect to Bottke & Greenberg (1993). Unlike the Öpik–Wetherill method, the Dell'Oro & Paolicchi (1998) approach does not need any a priori hypotheses about the angular elements Ω, ω and the mean anomaly M, and their method was designed for cases characterized by any kind of distribution of Ω, ω and M, including the possibility of mutual correlation. The general formula for the computation of the impact rate requires as input (apart from the orbital elements a, e and I of both target and projectile) an analytic function of six variables describing the joint

★ E-mail: delloro@arcetri.inaf.it

probability distribution of the elements Ω, ω and f (true anomaly) of the target and projectile. In the mathematical formalism of Dell'Oro & Paolicchi (1998), the canonical conditions are reformulated in terms of uniform distributions of node longitudes Ω, the pericentre arguments ω and mean anomalies M of the osculating orbits, and the assumption that no correlation exists among those three angles. The validity of the method of Dell'Oro & Paolicchi (1998) is even more general because it does not even need the hypothesis of constant a, e and I elements, which instead can be characterized by any statistical distribution. The only really necessary condition in the method of Dell'Oro & Paolicchi (1998) is that none of the osculating elements a, e and I are correlated with any of the angular elements Ω, ω and M.

The method of Dell'Oro & Paolicchi (1998) has been used to study the statistics of collisions among different populations of minor bodies, with canonical and non-canonical behaviour, like Main Belt asteroids taking into account the non-uniform distribution of the lines of apses (Dell'Oro et al. 1998), or for bodies in resonant orbits like Trojans and Plutinos (Dell'Oro et al. 1998, 2001, 2013). It has also been used to compute the rate of impacts among the members of asteroid families during the early stage of the dynamical evolution of their orbit after the formation of the family (Dell'Oro et al. 2002). Concerning the problem of the role of the non-uniform distribution of lines of apses of Main Belt asteroids, see the work of JeongAhn & Malhotra (2015) regarding the impact flux on planet Mars.

On the other hand, in the Solar system, dynamical situations exist for which there is a correlation between one (or more) of the elements a, e and I and one (or more) of the elements Ω, ω and M. Examples are orbits with large forced eccentricities and/or inclinations as expected by the Laplace–Lagrange linear secular theory or orbits in Kozai resonance (Murray & Dermott 1999). Some interesting efforts to compute the probability of collisions taking into account the effect of secular perturbations were done by Vokrouhlický, Pokorny & Nesvorny (2012) and Pokorny & Vokrouhlicky (2013), who derived analytical formulas in which the correlations between e and $\tilde{\omega}$ and between I and Ω are taken into account. However, the theory of Pokorny & Vokrouhlicky (2013) is not fully general yet, as it has some limitations on the orbit of the target.

Alternative approaches are based on a numerical integration of the asteroids' motion. If such a computation is dense and extended enough, it is possible to register the temporal sequence of close approaches and extrapolate their rate to the scale of the mutual distances of the order of the physical size of the asteroids (Marzari, Scholl & Farinella 1996; Dahlgren 1998).

Another family of numerical techniques is based on a Monte Carlo approach with a random sampling of the phase space. The first serious implementation of a Monte Carlo approach was proposed by Vedder (1996), who showed how the rate of impacts can be expressed as a function of the parameters of a particular statistical distribution, which in turn can be estimated by sampling the phase space randomly. This method was used with a certain degree of success for collisions among the Main Belt asteroids (Vedder 1998), even if, in some details, the results differ from those obtained by Bottke et al. (1994). More recently, Rickman et al. (2014) developed and tested two different methods for computing the impact probability between a planet and a projectile with given semi-major axis, eccentricity and inclination. One is like the super-sizing method introduced by Horner & Jones (2008) using the concept of Hill's sphere, and the other is based on computing the minimum orbit intersection distance (Wiźniowski & Rickman 2013).

In this paper, we present a new way to use the random samplings of the phase space to obtain the statistics of collisions among orbiting bodies. The method differs from the approaches of both Vedder (1996) and Rickman et al. (2014), even if it shares with them some fundamental ideas typical of any Monte Carlo strategy. The method is based on a new general and self-consistent mathematical formalism to derive the temporal rate of close approaches between particles not interacting gravitationally with each other. This method does not assume any a priori, even if general, statistics of the minimum distances among the particles as in Vedder (1996), and it does not make use of the concept of the Hill's sphere as in Rickman et al. (2014), revising the super-sizing approach from a more general point of view. Moreover, the final expressions that allow us to derive the statistics of collisions from the series of samples of the phase space are extremely simple and straightforwardly implementable. This work was originally triggered by the problem of computing impact rates among asteroids moving on orbits characterized by high forced eccentricity and inclination. Nevertheless, this topic is only partially discussed in this paper. Here, we prefer to focus on the fundamentals of the methodology and its validation, leaving future papers to special cases of scientific interest.

Section 2 provides a general but self-consistent description of the method, avoiding any technicality or mathematical proof of the formulas used in the text. A series of appendices focuses on all the technical details of the paper. In particular, Appendices A, B, C and D contain the theoretical foundations of the method and the derivation of the formulas used in Section 2, while Appendix E focuses on the analysis and estimation of the numerical errors of the method. In Section 3, we show the practical use of the method for the important case of systems fulfilling the canonical Öpik–Wetherill dynamical assumptions, and the results are compared to those derived with independent methods. In Section 4, we show how to implement our method to take into account secular perturbations. In particular, two examples with moderate and large forced eccentricities are discussed. Finally, Appendix F contains the results of further validation tests together with a list of revised values of the probabilities of collision and mean impact velocity for selected targets that can be impacted by Main Belt asteroids.

2 DESCRIPTION OF THE ALGORITHM

The method presented in this paper belongs to the family of Monte Carlo approaches consisting of an exploration of the phase space of the system by a random sampling to extract information about the statistics of impacts of the particles composing the system. The main difference with other Monte Carlo approaches consists of not requiring any a priori choice of the variables to be sampled and a different way to extract from the list of the samples statistical information about close encounters.

The fundamental prerequisite of the method is the availability of a suitable sampling model, which can be any numerical algorithm able to generate a random series of dynamical configurations of the system and, in particular, the relative distances and velocities of the particles. The sampling model depends strictly on the dynamical properties of the system and it must be constructed from time to time accordingly. If only the overall statistical properties of the system are concerned, a sampling model is a sort of surrogate of a direct numerical integration able to simulate the detailed temporal evolution of the system. Taking into account a system composed of only two particles, its evolution is fully described by the two functions $\mathbf{r}_1(t)$ and $\mathbf{r}_2(t)$, where $\mathbf{r}_1$ and $\mathbf{r}_2$ are the position vectors of the two particles and t is the time. In principle, knowing $\mathbf{r}_1(t)$ and $\mathbf{r}_2(t)$,

it is possible to count the number of times the two particles approach within a given distance R. So, using an ideal numerical integrator able to simulate the evolution of the system for a sufficiently long time, it is possible to compute the mean number of close approaches per unit of time. How long the integration time should be depends on the dynamical characteristics of the system. In any case, reliable statistics of the close encounters of the particles requires a full simulation of all the possible kinematic configurations of the system. To do this, a sort of dynamical stability of the system is required. If the mechanical energy of our example system was positive the two particles would be unbound and would move on hyperbolic orbits. Even if a close approach within R is possible, it occurs only once and it is not possible to define any frequency of close encounters. The two particles should move inside a finite region of space from which they cannot exit, but this hypothesis alone would not be sufficient. We require that the probability distribution of all possible dynamical configurations of the system does not depend on time. In this context, for the dynamical configuration of our system, we intend positions $\mathbf{r}_1$, $\mathbf{r}_2$ and velocities $\mathbf{v}_1 = \dot{\mathbf{r}}_1$, $\mathbf{v}_2 = \dot{\mathbf{r}}_2$ of the two particles at a given time t.

Therefore, if $\psi(\mathbf{r}_1, \mathbf{v}_1, \mathbf{r}_2, \mathbf{v}_2)\mathrm{d}\Gamma$ is the probability of finding the two particles with positions and velocities inside the infinitesimal volume $\mathrm{d}\Gamma$ of the phase space around the values ($\mathbf{r}_1$, $\mathbf{v}_1$, $\mathbf{r}_2$, $\mathbf{v}_2$), we assume that the probability density ψ does not depend on time. A sampling model is a numerical algorithm able to generate a list of values of ($\mathbf{r}_1$, $\mathbf{v}_1$, $\mathbf{r}_2$, $\mathbf{v}_2$) the parent distribution of which is ψ. The temporal stability of the function ψ is the basic condition of validity of all the analytical methods developed to compute the probability of collisions among minor bodies, although not explicitly given in these terms. The dynamical stability of the system is rather expressed using other variables different from positions and velocities. Indeed, the development of an analytical method as well as the construction of a sampling model rely on the availability of a dynamical model of evolution of the system describing its essential features. Examples of sampling models will be described further when practical cases are discussed.

The random sampling of the phase space is the essential part of any Monte Carlo approach developed so far. What makes the methodology presented in this work different from the previous approaches is the way we use the samplings to compute the frequency of close encounters. It is important to stress that in this work we do not take into account the case of interacting bodies, or, in other words, we assume that during the close encounters or in any other phase of their evolution the particles do not attract each other. So, no gravitational focusing is included in the computations. The method can be used to study collisions among minor bodies the masses of which are negligible, but probably it could have more or fewer limitations if used for the study of the statistics of close encounters and collisions between minor bodies and planets.

A close encounter within R occurs when the relative distance $r(t) = |\mathbf{r}_2(t) - \mathbf{r}_1(t)|$ between the two particles has a local minimum r_m and $r_\mathrm{m} < R$. Arbitrarily, we name one of the particles the target and the other the impactor, and we chose a reference system such that the target is at rest at the origin. In this reference system, around the close encounter epoch, the impactor spends a time Δt inside a sphere of radius R centred around the target. Δt is the duration of the close encounter. Following the hypothesis of non-interacting particles, we assume that during the interval Δt, the *relative* motion of the impactor can be approximated as a rectilinear trajectory with constant velocity. This hypothesis is not true in general, but we expect that it is better and better fulfilled for smaller and smaller values of R. It is important to stress that in this work we are not interested in evaluating the frequency of any kind of close encounter in general, but only of those occurring at a distance comparable to the physical dimensions of the bodies involved. If $\phi(R)$ is the mean number per unit time of close encounters within a distance R, our goal is to study the values of $\phi(R)$ for $R \to 0$, ignoring the cases with large R. In any case, the function $\phi(R)$ is a priori computed for a wide range of values of R, always assuming rectilinear relative motion, without proving the accuracy of this assumption, and then the relevant information is extracted a posteriori from the trend of $\phi(R)$, as explained later.

The final assumption of the algorithm is about the distribution of relative minimum distances r_m of close encounters within R. Each time the impactor crosses the sphere of radius R, the minimum approach distance is different. We assume that the probability that the minimum approach distance is less than r_m is proportional to r_m^2. This follows from a pure geometric cross-section description of the close encounter probability. Also this hypothesis is not true for all values of R, and in particular for large R, while we expect that it is satisfactorily fulfilled for small R.

The two hypotheses about the motion of the impactor inside the sphere of radius R are used to compute the mean duration $\overline{\Delta t}$ of a close encounter with a given value of the relative velocity $v = |\mathbf{v}_2 - \mathbf{v}_1|$, which we assume constant at least during the crossing of the sphere. As shown in Appendix A, the average value of the transit duration is $\overline{\Delta t} = (4/3)(R/v)$.

The importance of the parameter $\overline{\Delta t}$ is due to its connection with the probability of finding the two particles at a distance smaller than R and the temporal rate of the close encounters. In fact, if we followed the evolution of the system for a period of time T and we counted the number of close encounters $N(R)$ within a distance R that occurred, by definition the mean number of events per unit time would be $\nu(R) = N(R)/T$. On the other hand, we may also write $\nu(R) = p(R)/\tau(R)$, where $p(R) = S(R)/T$, $\tau(R) = S(R)/N(R)$ and $S(R)$ is the sum of the durations Δt of the $N(R)$ close encounters. Clearly, $p(R)$ is the probability of finding the two particles at a distance smaller than R *observing them at any instant of time randomly chosen*, while $\tau(R)$ is the average duration of the close encounters within R, or, in other words, $\tau(R)$ is the average of $\overline{\Delta t}$ taking into account the distribution of v.

The probability $p(R)$ is a typical quantity that can be evaluated by a sampling model. If we generate, using the sampling model, a sufficiently large number n of dynamical configurations, we can compute for each of them the relative distance r_k and the relative velocity v_k of the two particles, where k is the index of the kth generated configuration, running between 1 and n. If $n(R)$ is the number of configurations such that $r_k < R$, then $p(R) = n(R)/n$. As shown in Appendix B, the temporal rate of the close encounters within R can be computed simply as:

$$\phi(R) = \frac{3}{4}\frac{1}{R}\frac{1}{n}\sum_{k:r_k<R} v_k, \tag{1}$$

where the sum includes *only* the $n(R)$ sample relative velocities v_k for which the corresponding sample relative distances r_k are smaller than R, while n is the *total* number of sample configurations (r_k, v_k). From equation (1), each sample configuration contributes to the temporal rate with a different term proportional to v_k. This is in accordance with the improvement of the Öpik–Wetherill method introduced by Bottke et al. (1994), where each different orbital geometry is properly weighted according to its specific probability. In our method and formalism, this weight appears as the value of the relative velocity.

The statistics of the close encounters does not consider only the frequency of the events, but also the distribution of any parameter related to the close encounter. From this point of view, we assume that the value of the parameter does not change during the close encounter, i.e. on a time-scale corresponding to the duration of the close encounter. This approximation is expected to become better and better fulfilled for the relative velocity as R becomes smaller and smaller. The parameter q under investigation may be a quantity that can be computed from the positions and velocities of the particles, like the relative velocity itself and its orientation in space, or it may be an intermediate quantity in the computation of positions and velocities, like the parameters describing the orientation of the orbit or the location along the orbit. Together with the sample relative distance and velocity (r_k, v_k), the sampling model should provide also the corresponding sample value q_k of the parameter.

In Appendix C it is shown that the fraction of close encounters occurring per unit of time within R such that the parameter belongs to an interval of values $\mathcal{Q}$ is:

$$\Psi(R, \mathcal{Q}) = \frac{\sum_{k:r_k<R, q_k\in\mathcal{Q}} v_k}{\sum_{k:r_k<R} v_k}, \tag{2}$$

where the sum in the denominator includes all the sample velocities for which $r_k < R$, while the sum in the numerator includes *only* the sample velocities such that $r_k < R$ *and* $q_k \in \mathcal{Q}$. Moreover, the expected value of the parameter q is given by:

$$\langle q(R)\rangle = \frac{\sum_{k:r_k<R} q_k v_k}{\sum_{k:r_k<R} v_k}, \tag{3}$$

where the symbol $\langle q(R)\rangle$ emphasizes that the mean value of q refers to only the close encounters within a distance R, and the sums on the numerator and denominator include all the sample velocities v_k such that $r_k < R$, while q_k is the corresponding value of the parameter for each pair (see Appendix C for details). In a few words, the expected value of a parameter q for the close approaches within R is simply the weighted average of the sample values q_k by the corresponding relative velocity. One of the most important close encounter parameters is the relative velocity, the mean value of which is then correctly evaluated by:

$$\langle v(R)\rangle = \frac{\sum_{k:r_k<R} v_k^2}{\sum_{k:r_k<R} v_k}. \tag{4}$$

Equations (1)–(4) are the core tools of our algorithm, and they show in a very simple way how to obtain the statistics of close encounters within a given distance R from the series of samples (r_k, v_k, q_k). Those equations are the same whatever sampling model is used, because it is not important which procedure is followed to generate the series of random samples, provided that the distribution matches the density distribution ψ of the dynamical configuration of the system.

The computation of the quantities $\phi(R)$, $\Psi(R, \mathcal{Q})$ and $\langle q(R)\rangle$ by equations (1)–(3) is not a deterministic process since it is based on a Monte Carlo sampling. All the sums ($\sum_{k:r_k<R} v_k$), ($\sum_{k:r_k<R, q_k\in\mathcal{Q}} v_k$) and ($\sum_{k:r_k<R} q_k v_k$) are random variates, the values of which depend on the particular randomly generated sequence of the samplings (r_k, v_k, q_k). Repeating the computation with a new independent series of samplings, the values of the quantities provided by equations (1)–(3) fluctuate more or less around the mean values. Only for $n \to \infty$ do the formulas converge to the exact theoretical values. Both the numbers of terms of those sums and the values of each term are random variables. In particular, in sums ($\sum_{k:r_k<R} v_k$) and ($\sum_{k:r_k<R} q_k v_k$), the number of terms is $n(R)$, the number of samples for which the relative distance r_k is less than R. Typically, $n(R)$ is a Poisson deviate, and it can be null if no dynamical configuration has yet been found with the required property.

If $\overline{v}$ is the mean value of the relative velocity generated by the sampling model, $\sum_{k:r_k<R} v_k \sim n(R)\overline{v}$. However, $n(R)$ is a random variable with mean and variance equal to $np(R)$. Moreover, we expect that, for R small enough, the probability $p(R)$ of finding the two particles at a distance from each other shorter than R is proportional to R^3. So, in conclusion, the standard deviation of $\phi(R)$, computed starting from different and independent series of samplings, is $\sigma_\phi \propto R^{1/2} n^{-1/2}$. It is necessary to increase the total number of samples by a factor 100 (that is the computational time) to reduce the error of $\phi(R)$ by a factor 10. This is a common drawback of every Monte Carlo approach, which is an intrinsic numerical inefficiency.

On the other hand, an important advantage of this Monte Carlo approach is that equations (1)–(4) can be used even if the system is composed of more than two particles. Let us imagine that our system is composed of N different particles. For each pair of particles, identified by the indices i and j (with $i \neq j$), we can compute separately the rate of close encounters $\phi_{ij}(R)$, the distribution $\Psi_{ij}(R, \mathcal{Q})$ of the parameter q and its expected value $\langle q(R)\rangle_{ij}$, using separated series of samplings (r_{ijk}, v_{ijk}, q_{ijk}). On the other hand, we may be interested in the averages of the above quantities over the ensemble of all possible pairs of particles, given by:

$$\overline{\phi(R)} = \frac{\sum_{i\neq j} \phi_{ij}(R)}{M}, \tag{5}$$

$$\overline{\Psi(R, \mathcal{Q})} = \frac{\sum_{i\neq j} \Psi_{ij}(R, \mathcal{Q})}{M} \tag{6}$$

and

$$\overline{\langle q(R)\rangle} = \frac{\sum_{i\neq j} \langle q(R)\rangle_{ij}}{M}, \tag{7}$$

where $M = N(N-1)$ is the number of all possible pairs. The same averages are defined in the case that we are interested in, that is the statistics of close encounters between a fixed particle (target) and a swarm of another N particles (impactors), provided that $M = N$, that is the number of all the possible pairs of target and impactor. In any case, as is shown in Appendix D, all three averaged quantities $\overline{\phi(R)}$, $\overline{\Psi(R, \mathcal{Q})}$ and $\overline{\langle q(R)\rangle}$ can be evaluated directly by equations (1)–(4) using a *unique* series of random samplings (r_k, v_k, q_k), provided that at each new generation of the system configuration the two particles are *randomly* chosen among those constituting the system. The basic requirement is that at each generation, the probability of drawing lots for a particular couple of particles is the same for all possible couples.

The major advantage of this approach is numerical. As explained before, the numerical uncertainty in computing equations (1)–(4) comes from the random fluctuations inherent in the random generation of the samples (r_k, v_k, q_k). The variance of such fluctuations scales as $1/n$, where n is the total number of samples. If we consider that *each* pair of particles has n samples, the variance of the average rate $\overline{\phi(R)}$ scales as $1/(nM)$, where M is the number of terms in the average (the number of possible pairs). This means that if we do not

need to know the single $\phi_{ij}(R)$ to large precision, we can reduce the number of samples n for each pair by a factor M to curb the total number of samples generated to compute $\overline{\phi(R)}$. The advantage is that, given the level of variance we need for $\overline{\phi(R)}$, the computational time does not depend on the number of particles composing the system. The extreme case consists of systems with a continuous distribution of orbits (particles), the orbital elements of which are generated directly by the sampling model (see the example in Section 3).

The last step of our method consists in extrapolating the parameters of the statistics for $R \to 0$. All quantities $\phi(R)$, $\Psi(R, \mathcal{Q})$ and $\langle q(R) \rangle$, referring to either a single pair of particles or averaged over different couples of particles, are all functions of the close encounter range R. As pointed out before, the only interval of values of R we are interested in is that corresponding to the physical size of the particles under investigation. On the other hand, it is almost never possible to compute $\phi(R)$, and the other related quantities, directly for values of R small enough. The precision in the computation of $\phi(R)$ depends on the number $n(R)$ of sample configurations with $r_k < R$ that we have generated. The larger $n(R)$ is, the smaller the uncertainty of $\phi(R)$ is. However, the probability $p(R)$ of generating such sample configurations decreases very fast with R and typically $p(R) \propto R^3$; therefore, the computational time needed to generate the same number $n(R)$ of sample configurations for different values of R is proportional to R^{-3}. For example, for asteroids, we would like to compute $\phi(R)$ for values of R of the order of 1–100 km. As we see later, for Main Belt asteroids with typical size of the orbits of the order of 2÷3 au, the probability of generating sample configurations with a relative distance less than 1–100 km is so small as to require a prohibitive computational effort. On the other hand, we have some conceivable expectations about the trend of $\phi(R)$ for $R \to 0$. Ignoring any mutual gravitational effect, $\phi(R)$ should be proportional to R^2 for R small enough, a trend simply due to the pure geometrical cross-section of the close encounter.

On the other hand, such an expectation is the basis of the definition of the *intrinsic* probability of collision P_i introduced by Wetherill (1967), and widely used in the study of statistics of collisions among asteroids. The quantity P_i is defined for a pair of bodies with intersecting orbits, and it depends only on the orbital parameters, regardless of the physical sizes of the two bodies. It is defined as the mean number of close approaches per unit of time within 1 km, and it is usually expressed in units of km^{-2} yr^{-1}. In other terms, $P_i = \phi(R)$ for $R = 1$ km. Following Wetherill (1967), the probability of collision, during an interval of time T, between two bodies with radii R_t and R_p, moving in orbits for which P_i is known, is given by $(R_t + R_p)^2 P_i T$, if the radii are expressed in kilometres and T in years.

However, note that, from a strictly mathematical point of view, the assumption that $\phi(R) \propto R^2$ may not be true in some special cases. For two orbits always perfectly coplanar, for example, if both their inclinations are null, we lose the three-dimensionality of the system, the rate of close approaches cannot be proportional to the usual geometrical cross-section but it is proportional to R, for all values of R. Apart from this special case, the typical situation is that $\phi(R)$ is proportional to R for large values of R, and $\phi(R) \propto R^2$ for small R. The exact meaning of large and small depends on the geometry of the orbits and, in particular, on their inclinations. Let us take into account orbits with typical semi-major axis a, eccentricity e and inclination I. The maximum distance from the ecliptic plane that the particle can reach is typically about $a(1 + e)\sin I$. For values of R much larger than $a(1 + e)\sin I$, the system appears almost bi-dimensional and so $\phi(R) \propto R$, but for $R \ll a(1 + e)\sin I$ the three-dimensionality of the system is fully recovered and $\phi(R) \propto R^2$. The threshold dividing the two regimes depends on I.

In any case, the general strategy is to search for a value R_ϕ such that for $R < R_\phi$, the rate $\phi(R) \propto R^2$. In an equivalent manner, we look for a value R_ϕ such that for $R < R_\phi$ then $\phi(R)/R^2$ is constant. In such an interval of values of R, the ratio $\phi(R)/R^2$ is by definition equal to P_i. For the distribution and the mean value of a close encounter parameter q computed by means of equations (2) and (3), we expect that they will tend asymptotically to a finite value for $R \to 0$. In other terms, we look for a value R_q such that for $R < R_q$ the mean value $\langle q(R) \rangle$ is constant. Note that, in general, the two thresholds R_ϕ and R_q, if they exist, could not be the same. The limit U_m for $R \to 0$ of the mean value of the close encounter relative velocity $\langle v(R) \rangle$ is our estimation of the mean impact velocity.

In conclusion, our algorithm can be summarized as follows. A series of values R_i of the close encounter range is defined. Generally, the values R_i are logarithmically scaled covering a wide interval of orders of magnitude of the relative distance between the particles. The computation has a recursive loop. Each time the loop restarts, a new dynamical configuration is generated. If the index k identifies the kth configuration, for each value R_i the sums $S_{i,v} = (\sum_{k:r_k<R_i} v_k)$ and $S_{i,v^2} = (\sum_{k:r_k<R_i} v_k^2)$ are defined. Moreover, if we are interested in the statistics of a particular parameter q, the additional sums $S_{i,\mathcal{Q}_j} = (\sum_{k:r_k<R_i, q_k \in \mathcal{Q}_j} v_k)$ and $S_{i,qv} = (\sum_{k:r_k<R_i} q_k v_k)$ are defined, where $\mathcal{Q}_j$ is a suitable set of bins of values of the parameter q used to build up the distribution of q. Anyway, all sums S are initially set equal to zero. The loop consists of the following four steps:

(L1) A random dynamical configuration is generated using the sampling model. The result is the relative distance r_k, the corresponding value of the relative velocity v_k and the value q_k of the parameter we are interested in. The sampling model, or in other words the algorithm used to generate r_k, v_k and q_k, depends on the particular dynamical system under investigation and it is developed separately.

(L2) The total number of generated configurations n is incremented by 1, and the values of the sums $S_{i,v}$, S_{i,v^2} and $S_{i,qv}$ are incremented respectively by v_k, v_k^2 and $q_k v_k$ *if* $r_k < R_i$, while the sums $S_{i,\mathcal{Q}_j}$ are incremented by v_k if $r_k < R_i$ *and* $q_k \in \mathcal{Q}_j$.

(L3) The values of $\phi(R_i)$, $\Psi(R_i, \mathcal{Q}_j)$, $\langle q(R_i) \rangle$ and $\langle v(R_i) \rangle$ are recomputed with equations (1)–(4) using the new values of the sums S. Moreover, the errors in evaluating $\phi(R)$, $\langle q(R_i) \rangle$ and $\langle v(R_i) \rangle$ due to the random fluctuations in the generation of r_k, v_k and q_k are estimated (see Appendix E for details).

(L4) The series of values $\phi(R_i)$, $\langle q(R_i) \rangle$ and $\langle v(R_i) \rangle$ are analysed to identify, if it exists, a value R_ϕ such that for all $R_i < R_\phi$ then $\phi(R_i) \propto R_i^2$, or in other terms, if $\phi(R_i)/R_i^2$ are compatible with a constant value for $R_i < R_\phi$, taking into account the random fluctuations. A similar check is done also for $\langle v(R_i) \rangle$ and $\langle q(R_i) \rangle$, looking for values R_v and R_q such that $\langle v(R_i) \rangle$ and $\langle q(R_i) \rangle$ become constant for $R < R_v$ and $R < R_q$, respectively, compatible with the estimated random fluctuations. If the result of these checks is positive, the loop is interrupted, otherwise it restarts from the step (L1).

The final result of the processing is an estimation of the asymptotic value of the ratio $\phi(R)/R^2$ for $R \to 0$, as our evaluation of P_i, and the limit values of $\langle v(R) \rangle$ and $\langle q(R) \rangle$, respectively, as estimates of U_m and of the mean of q at impact. The algorithm fails when the expected limit trends are not identified with acceptable accuracy. This can happen if the total number of samplings n is not sufficient, and the algorithm requires a much larger number of samplings with,

consequently, a much longer computational time. In some cases, depending on the properties of the dynamical system, the necessary computing time is prohibitive.

3 CANONICAL STATISTICS

The first example of the use of the method described in Section 2 that we want to discuss in this paper regards the statistics of collisions among asteroids assuming the same dynamical hypotheses that Öpik–Wetherill theory is based on (Öpik 1951; Wetherill 1967). Such assumptions are:

(i) The semi-major axes a, eccentricities e and inclinations I of the osculating orbits are fixed.

(ii) The rate of variation of the longitudes of the nodes Ω and arguments of the pericentres ω of the osculating orbits are constant, or, in other words, the osculating orbits circulate uniformly.

(iii) The motion of the asteroids along their osculating orbits is fully described by Kepler's second law.

The condition (iii) states nothing more than that the mean anomalies M of the bodies uniformly change with time. Therefore, the conditions (ii) and (iii) entail that all the possible values between 0 and 2π of the three angles Ω, ω and M are equally probable, if we observe the bodies at any instant of time randomly chosen. Moreover, there is no correlation among the three angles of the same orbit or between the angles of different orbits. Let us consider a system composed of only two particles, identified by the indices i and j. The values of the orbital elements (a_i, e_i, I_i) and (a_j, e_j, I_j) are fixed once and for all. The sampling model correctly describing the dynamical properties of the system consists of the following steps:

(C1) The values of the angles Ω_i, ω_i, M_i of the ith particle, and the angles Ω_j, ω_j, M_j of the jth particle are randomly and independently chosen in the interval $[0, 2\pi]$.

(C2) The vector position $\mathbf{r}_i$ and vector velocity $\mathbf{v}_i$ of the ith particle are computed from the orbital elements a_i, e_i, I_i, Ω_i, ω_i and M_i. The vector position $\mathbf{r}_j$ and vector velocity $\mathbf{v}_j$ of the jth particle are computed from the orbital elements a_j, e_j, I_j, Ω_j, ω_j and M_j.

(C3) The relative distance $r = |\mathbf{r}_i - \mathbf{r}_j|$ and the relative velocity $v = |\mathbf{v}_i - \mathbf{v}_j|$ are computed.

The steps (C1)–(C3) are the procedure to be used in the step (L1) of the algorithm loop of Section 2, if canonical conditions (i)–(iii) are assumed. If we are dealing with a group of asteroids composed of more than two bodies, a step (C0) must be included just before the step (C1), in which two asteroids are randomly selected provided only they are distinct ($i \neq j$). In this case, each new generation of the values r and v starts from new sets of the orbital elements (a_i, e_i, I_i) and (a_j, e_j, I_j).

Canonical statistics of collisions among asteroids is an already deeply investigated scientific case, and well-established computational tools have been developed, in particular by Greenberg (1982) and Bottke et al. (1994). The more general approach of Dell'Oro & Paolicchi (1998) proved to be equivalent, if canonical hypotheses are assumed, to Greenberg (1982) and Bottke et al. (1994), providing the same results for the same test cases. Appendix F describes a series of validation tests of our Monte Carlo algorithm, comparing the results of this method with those provided by Bottke et al. (1994) and Dell'Oro & Paolicchi (1998). Here we discuss only some example cases to show the use of our method.

The first example is a pair of bodies with orbital elements $a = 3.42$ au, $e = 0.578$ and $I = 0.435$ rad, and $a = 1.59$ au, $e = 0.056$ and $I = 0.466$ rad. This case was studied by Bottke et al. (1994) and Dell'Oro & Paolicchi (1998). The intrinsic probability of collision computed using the methods of Bottke et al. (1994) and Dell'Oro & Paolicchi (1998) is $P_i = 5.70 \times 10^{-18}$ km^{-2} yr^{-1}, while the mean impact velocity and the standard deviation of the impact velocity are, respectively, $U_\mathrm{m} = 14.94$ km s^{-1} and $\sigma_U = 5.95$ km s^{-1}. We computed the quantities $\phi(R)$ and $\langle v(R)\rangle$ with equations (1) and (4) using $n = 10^{12}$ samplings. The results are shown in Fig. 1. A well-identified trend $\phi(R) \propto R^2$ holds for values of R below 10^7 km, about 4 per cent of the semi-major axis of the smaller orbit. Also $\langle v(R)\rangle$ and the standard deviation of the relative velocity $\sigma_v(R)$ become constant for $R < 10^7$ km. The standard deviation of the frequency distribution of the relative velocity for the close encounter within a distance R is computed as $\sigma_v^2(R) = \langle v^2(R)\rangle - \langle v(R)\rangle^2$, where, according to equation (3), the expected value of the square of the relative velocity is

$$\langle v^2(R)\rangle = \frac{\sum\limits_{k:r_k<R} v_k^3}{\sum\limits_{k:r_k<R} v_k}. \tag{8}$$

The asymptotic values of $\phi(R)/R^2$, $\langle v(R)\rangle$ and $\sigma_v(R)$ for $R \to 0$ are fully compatible with the parameters P_i, U_m and σ_U computed independently, and represented by long-dashed lines in the plots.

Below 10^5 km, the random fluctuations of this particular sampling become significant. The length of the error bars is equal to the expected standard deviation of the vertical shift of the respective points from their average positions that we would obtain with an infinite number of samples or repeating the sampling a large number of times. Fig. 2 gives an idea of the extent of the random fluctuations of the value of $\phi(R)/R^2$ computed using equation (1). The plots show $\phi(R)/R^2$ versus R obtained using four sample series (r_k, v_k) independently generated but with the same number of samples. It is important to stress that the fluctuations of the values $\phi(R)$, $\langle v(R)\rangle$ and $\sigma_v(R)$ for different values of R are not independent but correlated with each other. In fact, the parameter $\phi(R)$ is obtained using the values of v_k such that $r_k < R$, which are also included among the samples used to compute $\phi(R')$ with $R' > R$. In this way, systematic trends in the fluctuations of the values $\phi(R)$ can occur (like that shown in Fig. 2 bottom right), but they have nothing to do with the mean trend of the function $\phi(R)$ versus R.

Fig. 3 shows some frequency distributions of the relative velocity computed using equation (2). In this case, the parameter q is the relative velocity itself. The interval of values of v from 0 to 30 km s^{-1} has been divided into 1000 bins, $\mathcal{V}_i = [v_i, v_i + \Delta v]$, each $\Delta v = 30$ m s^{-1} wide. So, from equation (2), the fraction of events occurring with a relative velocity belonging to each bin is

$$\Psi(R, \mathcal{V}_i) = \frac{\sum\limits_{k:r_k<R, v_k\in[v_i, v_i+\Delta v]} v_k}{\sum\limits_{k:r_k<R} v_k}. \tag{9}$$

In the plots in Fig. 3, the values $\Psi(R, \mathcal{V}_i)/\Delta v$ versus v_i are represented as black dots. Four different cases are taken into account corresponding to the distribution of the relative velocities for the close encounter occurring within distances $R = 3 \times 10^7$ km, $R = 5 \times 10^6$ km, $R = 2 \times 10^6$ km and $R = 1 \times 10^6$ km. The plots show how, on decreasing R, the distributions $\Psi(R, \mathcal{V}_i)$ converge to the distribution of the impact velocity predicted by Bottke et al. (1994) and drawn in each plot as a thin line. Obviously, the random noise increases for smaller R. The total number of samples (r_k, v_k) used in equation (9) was $n = 10^{12}$. The probability of

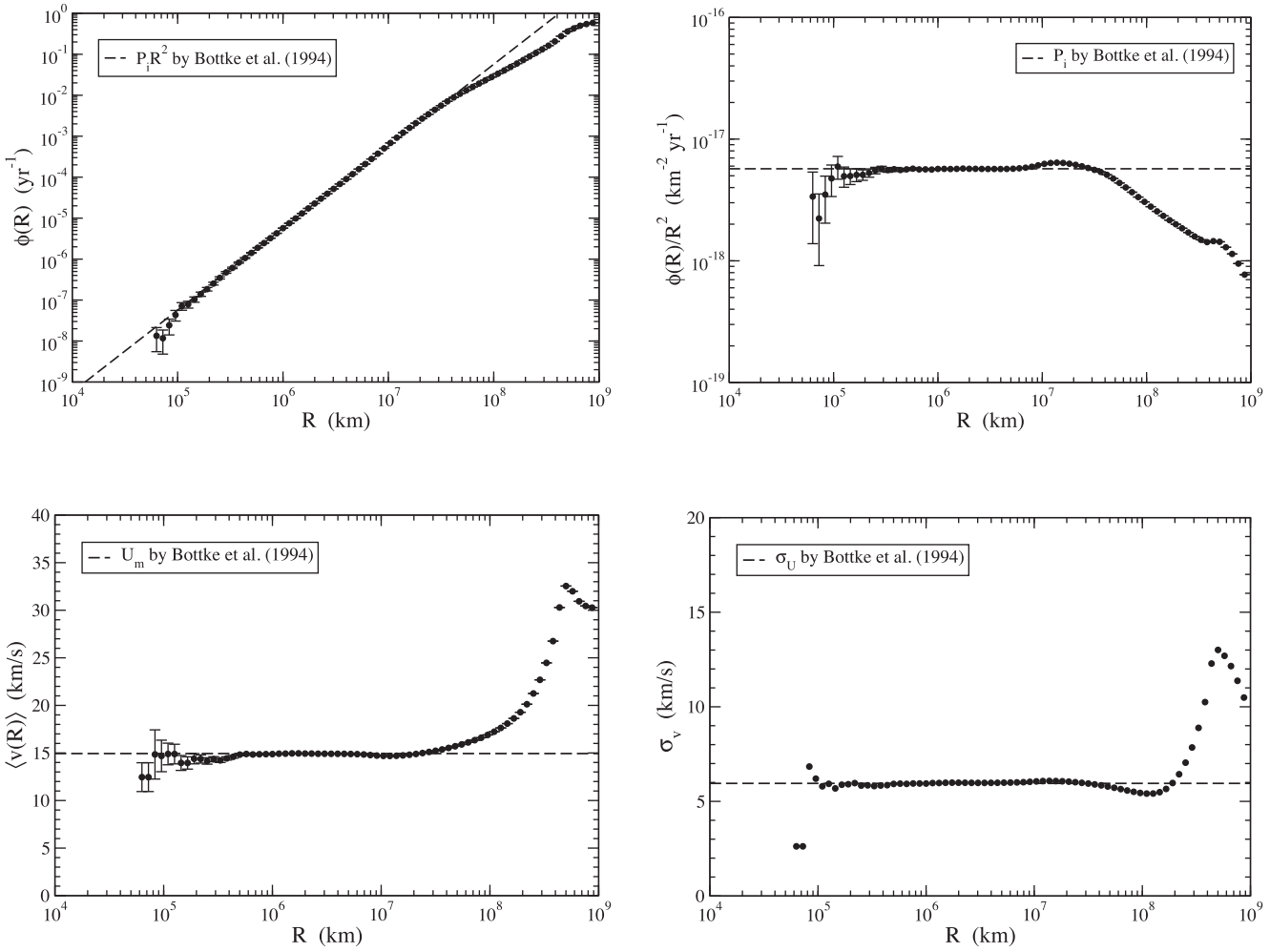

Figure 1. Canonical statistics of close encounters between two orbits with elements $a = 3.42$ au, $e = 0.578$ and $I = 0.435$ rad, and $a = 1.59$ au, $e = 0.056$ and $I = 0.466$ rad. Top left: Frequency of close encounters $\phi(R)$ computed using equation (1) versus R. Top right: Ratio $\phi(R)/R^2$ versus R. Bottom left: Mean close encounter velocity $\langle v(R)\rangle$ computed using equation (4) versus R. Bottom right: Standard deviation of the close encounter velocity versus R. The long-dashed lines represent the intrinsic probability of collision P_i, the mean impact velocity $U_{\rm m}$ and the standard deviation of the impact velocity σ_U computed using the method of Bottke et al. (1994).

generating a dynamical configuration with $r_k < R = 10^6$ km is about $p(R) = 1.9 \times 10^{-8}$, and the corresponding number of samples was about 1.9×10^4. So, on average about 20 samples fall into each of the 1000 velocity bins $\mathcal{V}_i$, with a standard deviation of the order of 20 per cent. On the other hand, all the features of the impact velocity distribution are correctly reproduced on average, including the peaks corresponding to the encounter close to the pericentres and apocentres of the two orbits.

In the second example, there are two orbits with orbital elements much more similar than in the previous case to highlight the expected behaviour for $R \to 0$, in particular for $\phi(R)$. The two orbits have orbital elements $a_0 = 2.5$ au, $e_0 = 0$ and $I_0 = 0$ deg, and $a_0 = 2.5$, $e = 0.001$ and $I = 0.05$ deg. The smaller differences between the orbital elements allow us to explore values of R much smaller than in the previous example without excessive computational effort. Fig. 4 shows the result of this exercise. The methods of Bottke et al. (1994) and Dell'Oro & Paolicchi (1998) give as the intrinsic probability of collision $P_i = 8.76 \times 10^{-16}$ km^{-2} yr^{-1} (long-dashed line in Fig. 4 at left), in good agreement with the values of $\phi(R)/R^2$ provided by our algorithm. In particular, the trend $\phi(R) \propto R^2$ is clearly found for R below $\sim 5 \times 10^5$ km. The relative velocity $\langle v(R)\rangle$ rapidly converges to the limit value of 25 m s^{-1}, corresponding to the only possible value of the impact velocity. Indeed, the distribution of the impact velocity is a Dirac function, equal to zero except for the value of 25 m s^{-1}. This is because only two crossing geometries are possible. The two orbits have the same semi-major axis a_0. Since the first one is circular, if they have a point in common it must be at a distance from the Sun equal to a_0. This entails that the true anomaly f of the crossing point on the eccentric orbit must fulfil the condition $\cos f = -e$. At any point of the circular orbit, the transverse component of the orbital velocity is $v_0 = \sqrt{GM/a_0}$, while the radial component is null. For the eccentric orbit, at the crossing point, the modulus of the orbital velocity is again equal to v_0, with radial component $v_{\rm R} = \pm e v_0$ and transverse component $v_{\rm T} = \sqrt{1 - e^2}\, v_0$. The inclination of the circular orbit being null, the crossing point is one of the nodes (ascending or descending) of the eccentric orbit. This means that the direction of the transverse component of the velocity on the eccentric orbit forms an angle I with respect to the transverse component of the circular orbit. In conclusion, the relative velocity U at the crossing

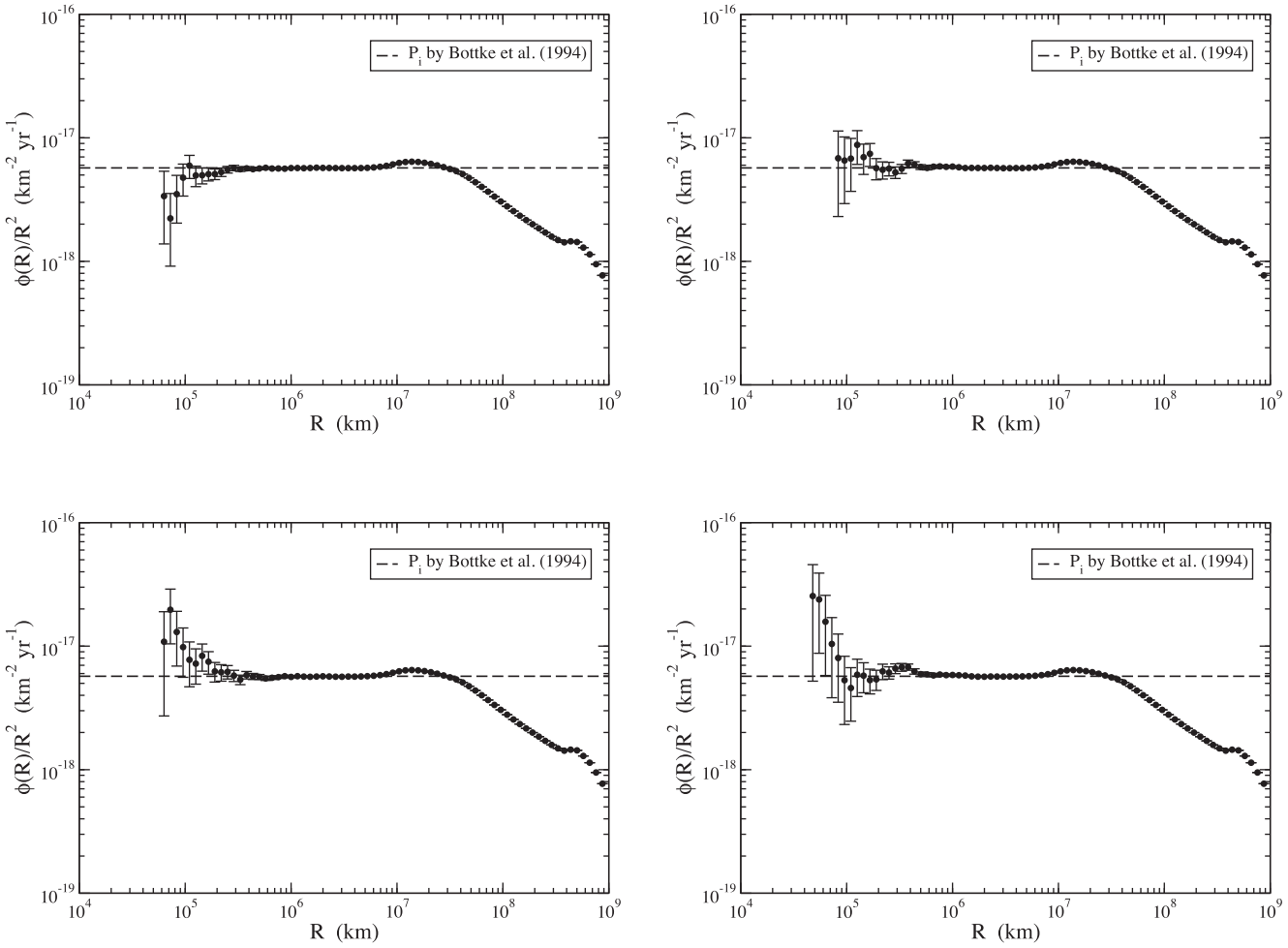

Figure 2. The plot $\phi(R)/R^2$ versus R of Fig. 1 at top right (top left in this figure) is compared with three other similar plots obtained with different random sampling series (top right, bottom left and right).

point is

$$U^2 = (v_{\rm T}\cos I - v_0)^2 + v_{\rm R}^2 + v_{\rm T}^2 \sin^2 I = 2v_0^2(1 - \cos I\sqrt{1-e^2}). \quad (10)$$

From equation (10), then $U = 25$ m s^{-1}.

The last example consists of a continuous distribution of orbits forming a torus with minimum pericentre distance $q_{\rm min}$ and maximum apocentre distance $Q_{\rm max}$. More precisely, the density distributions of the pericentre and apocentre distances (q, Q) are different from zero and uniform in the region $q \in [q_{\rm min}, Q_{\rm max}]$ and $Q \in [q_{\rm min}, Q_{\rm max}]$, apart from the values where $q < Q$. The inclinations are uniformly distributed in the interval $[0, I_{\rm max}]$. In this case, we no longer deal with orbits with fixed orbital elements, and the sampling model must be changed accordingly. More exactly, it consists of the following steps:

(CT1) The pericentre distance q_A of a particle A is randomly generated between $q_{\rm min}$ and $Q_{\rm max}$, while its apocentre distance Q_A is randomly generated between q_A and $Q_{\rm max}$. Moreover, the pericentre distance q_B of a particle B is randomly generated between $q_{\rm min}$ and $Q_{\rm max}$, while its apocentre distance Q_B is randomly generated between q_B and $Q_{\rm max}$.

(CT2) The semi-major axes and eccentricities of the particles A and B are obtained from:

$$a_A = \frac{Q_A + q_A}{2}, \qquad e_A = \frac{Q_A - q_A}{Q_A + q_A}, \quad (11)$$

and

$$a_B = \frac{Q_B + q_B}{2}, \qquad e_B = \frac{Q_B - q_B}{Q_B + q_B}. \quad (12)$$

(CT3) The inclinations I_A and I_B are randomly generated between 0 and $I_{\rm max}$.

(CT4) The previously defined steps (C1), (C2) and (C3) are executed, where $i = A$ and $j = B$.

Therefore, in this sampling model, every time a new dynamical configuration is generated, *all* orbital elements change, *including* semi-major axes, eccentricities and inclinations. In this case, as explained before, the quantities computed with equations (1)–(4) are the averages values of ϕ, Ψ and $\langle v\rangle$ taking into account all the possible pairs of orbits (a_A, e_A, I_A) and (a_B, e_B, I_B) generated according to the rules (CT1)–(CT3), assuming uniform circulations of the angles Ω, ω and M of both orbits. Fig. 5 shows $\phi(R)$ and $\langle v(R)\rangle$, versus R, for a torus of orbits with $q_{\rm min} = 2.99$ au, $Q_{\rm max} = 3.01$ au and $I_{\rm max} = 0.2$ deg. For comparison, we have computed the mean P_i and

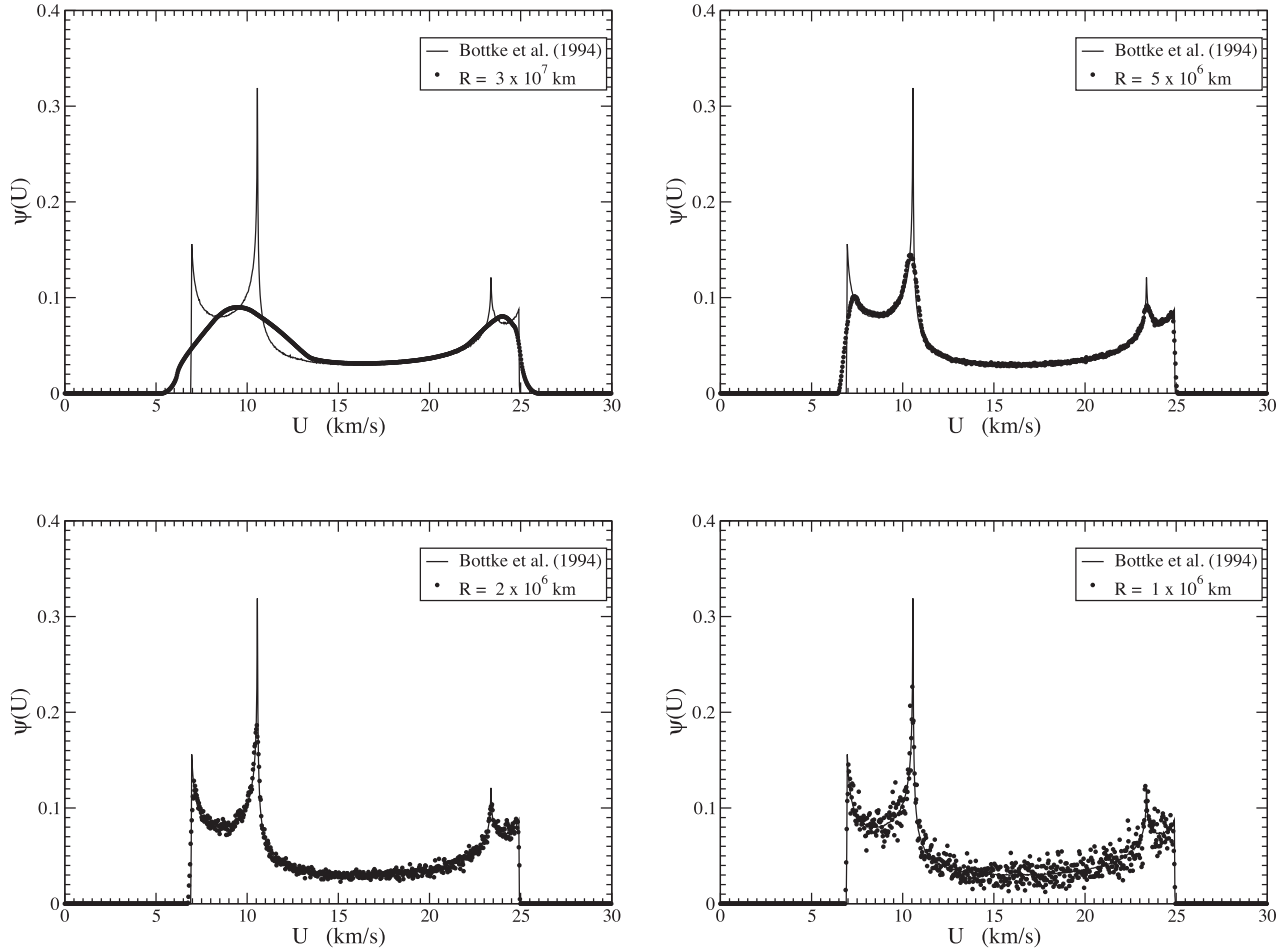

Figure 3. For the same case in Fig. 1, distributions of the close encounter velocity (dots) obtained using equation (2) for the close encounters occurring within different distance limits R (3×10^7 km at top left, 5×10^6 km at top right, 2×10^6 km at bottom left and 10^6 km at bottom right). A set of bins 30 m s^{-1} wide are used as intervals $\mathcal{Q}$ in equation (2). Each plot is compared with the impact velocity distribution obtained by Bottke et al. (1994) (thin line).

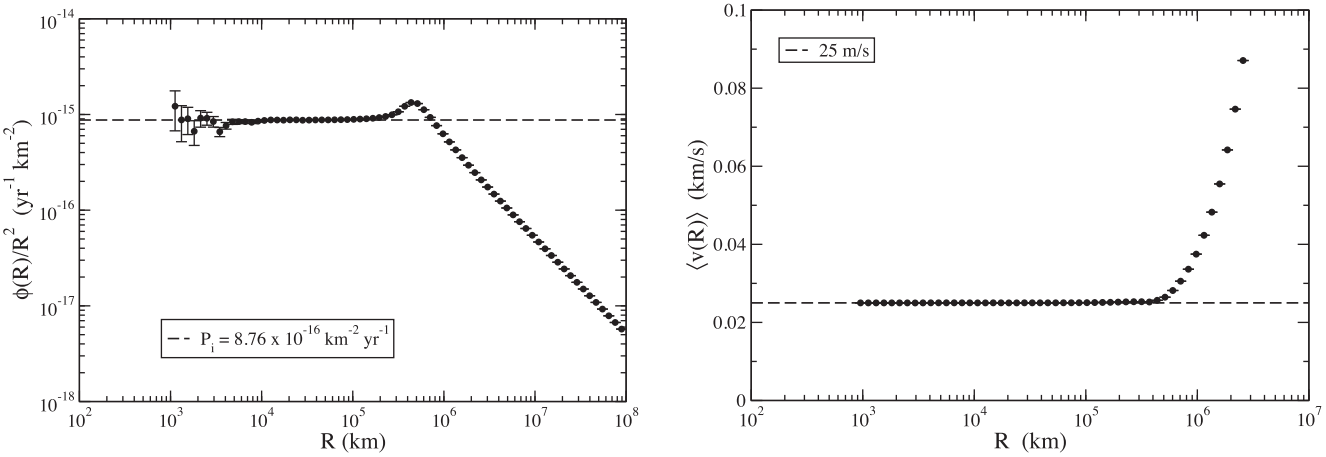

Figure 4. Close encounters between two orbits with elements $a = 2.5$ au, $e = 0$ and $I = 0$ deg, and $a = 2.5$ au, $e = 0.001$ and $I = 0.05$ deg: $\phi(R)/R^2$ versus R (left) and $\langle v(R) \rangle$ versus R (right). The long-dashed line in the plot at left is the expected value of P_i computed using the methods of Bottke et al. (1994) and Dell'Oro & Paolicchi (1998).

the mean $U_{\rm m}$ using the methods of Bottke et al. (1994) and Dell'Oro & Paolicchi (1998). We generated a random list of 1000 orbits (a, e, I) following the steps (CT1)–(CT3). Then, for each of the 499 500 possible pairs of orbits, we computed P_i and $U_{\rm m}$ using the formulas of Bottke et al. (1994). Finally, we computed the average values, obtaining $\overline{P_i} = 2.3 \times 10^{-16}$ km^{-2} yr^{-1} and $\overline{U_{\rm m}} = 46$ m s^{-1}. In Fig. 5 at left, the long-dashed line represents the function $\overline{P_i} R^2$, which is in good agreement with the values obtained by our method. The

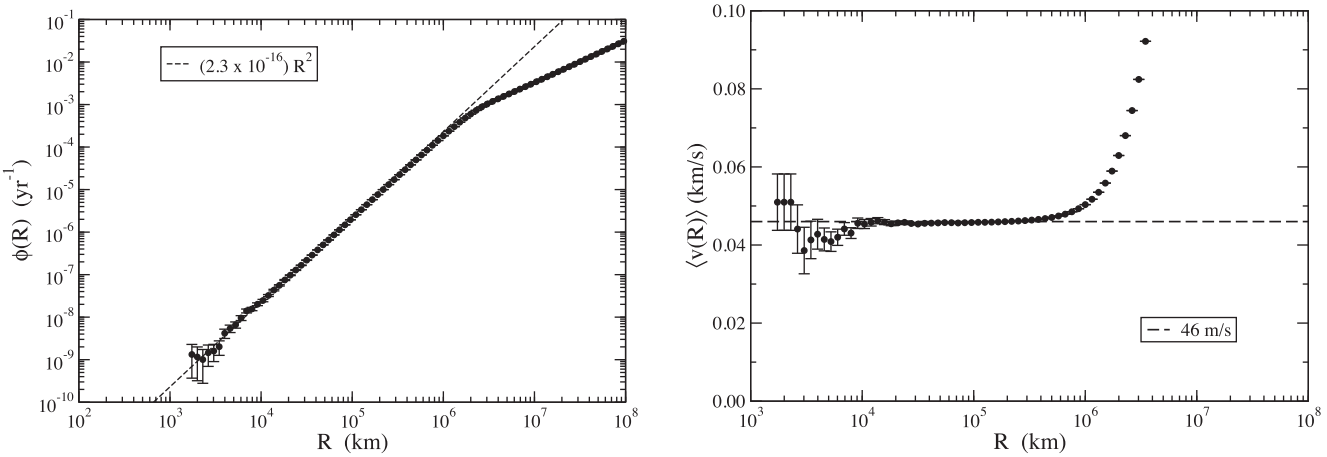

Figure 5. Close encounter rate $\phi(R)$ versus R and mean relative velocity for close encounters within a distance R versus R, computed for a continuous distribution of orbits (see text for details), assuming canonical statistics hypotheses.

same good agreement is also seen between $\overline{U_{\rm m}}$ and the asymptotic value of $\langle v(R)\rangle$ (Fig. 5 at right), at least for $R > 10^4$ km, while below 10^4 km random fluctuations became dominant.

4 SECULAR PERTURBATIONS

In this section, we give an example of the present method in which the dynamical evolution of the orbits does not fulfil the canonical hypotheses of the Öpik–Wetherill statistics. In this case, the problem of computing the probability of collision cannot be solved using either the classical methods based on the approach or the assumption of Öpik–Wetherill, nor the more general techniques introduced by Dell'Oro & Paolicchi (1998), for reasons that will be clear soon.

The most important and general case of this type is the statistics of close approaches among objects the orbits of which evolve with time according to the Laplace–Lagrange linear secular theory (Murray & Dermott 1999), which describes the results of the first-order perturbations. We consider only the simplest case of two bodies A and B orbiting around a central mass m_0. The masses m_A and m_B of the two bodies are negligible with respect to m_0. Moreover, the mass m_A is negligible with respect to m_B. In this condition, the body B does not feel any gravitational perturbation from A, and its orbit is pure Keplerian. In contrast, B is a perturber for A, and the orbit of A is no longer Keplerian. The secular theory tells us how the orbit of A changes with time.

We pass over some technical details of the secular theory, and in particular the distinction between mean and osculating elements, that are not important for the collision statistics. We use the terms orbital elements or instantaneous orbital element referring to the mean elements. Leaving aside the details of the theory, the features of the evolution of the orbital elements that are relevant for the statistics of collisions are the following.

The semi-major axis a of the orbit of A is again a constant of motion and so its value is fixed. The eccentricity varies with time and its value is linked to the value of the longitude of the pericentre. In particular, the evolution of the orbit A is more suitably described by the non-singular elements h and k defined as:

$$h = e\sin\tilde{\omega}, \qquad k = e\cos\tilde{\omega},$$

where e is the eccentricity and $\tilde{\omega} = \omega + \Omega$ is the longitude of the pericentre. In the plane (k, h), the point P representing the orbit moves with constant angular velocity along a circle of radius e_p (Fig. 6 top left). The quantity e_p is the proper eccentricity. The centre of the circle does not coincide, in general, with the origin of the plane (k, h). Its distance from the origin is the forced eccentricity e_f, and its position angle with respect to the k-axis is the forced pericentre longitude $\tilde{\omega}_f$. The distance of the point P from the origin is the instantaneous eccentricity e while its position angle with respect to the k-axis is the instantaneous pericentre longitude $\tilde{\omega}$. As already anticipated, the position angle α of the point P on the circle changes uniformly with time, i.e. $\mathrm{d}\alpha/\mathrm{d}t$ is constant. There is a similar scheme for the evolution of the inclination I. The corresponding non-singular elements are:

$$P = \sin I\sin\Omega, \qquad Q = \sin I\cos\Omega.$$

In the plane (Q, P) (Fig. 6 top right), the point representing the instantaneous orbit moves with constant angular velocity ($\mathrm{d}\beta/\mathrm{d}t = $ const.) along a circle of radius $\sin I_p$ (sine of the proper inclination). The distance between the centre of this circle and the origin is the sine of the force inclination I_f, and the position angle of the centre is the forced longitude of node Ω_f. For the instantaneous orbital elements, the distance of the point P from the origin is the sine of the instantaneous inclination I, and its position angle with respect to the Q-axis is the instantaneous longitude of node Ω.

The forced elements e_f, $\tilde{\omega}_f$, I_f and Ω_f are directly related to the orbital elements of the perturber B, and the semi-major axis of the orbit of A, and so they are fixed. Since the representative points P are constrained to move uniformly along the above mentioned circles, then e and I are no longer constant, and $\mathrm{d}\Omega/\mathrm{d}t$ and $\mathrm{d}\omega/\mathrm{d}t$ are not constant too, violating the fundamental assumptions of the Öpik–Wetherill approach. However, more importantly, in this dynamical system correlations between the values of e and $\tilde{\omega}$, and between the values of I and Ω, exist.

This last feature prevents us from using the method of Dell'Oro & Paolicchi (1998) to compute the collision probabilities between bodies the orbits of which behave as described above, as well as any other method strictly based on Öpik–Wetherill assumptions, like Kessler (1981) or Bottke et al. (1994). Although the method of Dell'Oro & Paolicchi (1998) was developed for a generic distribution of the elements ω and Ω, nevertheless, it requires that no correlation can exist between them and any of the elements a, e and I. Note that the importance of the correlation above for the statistics

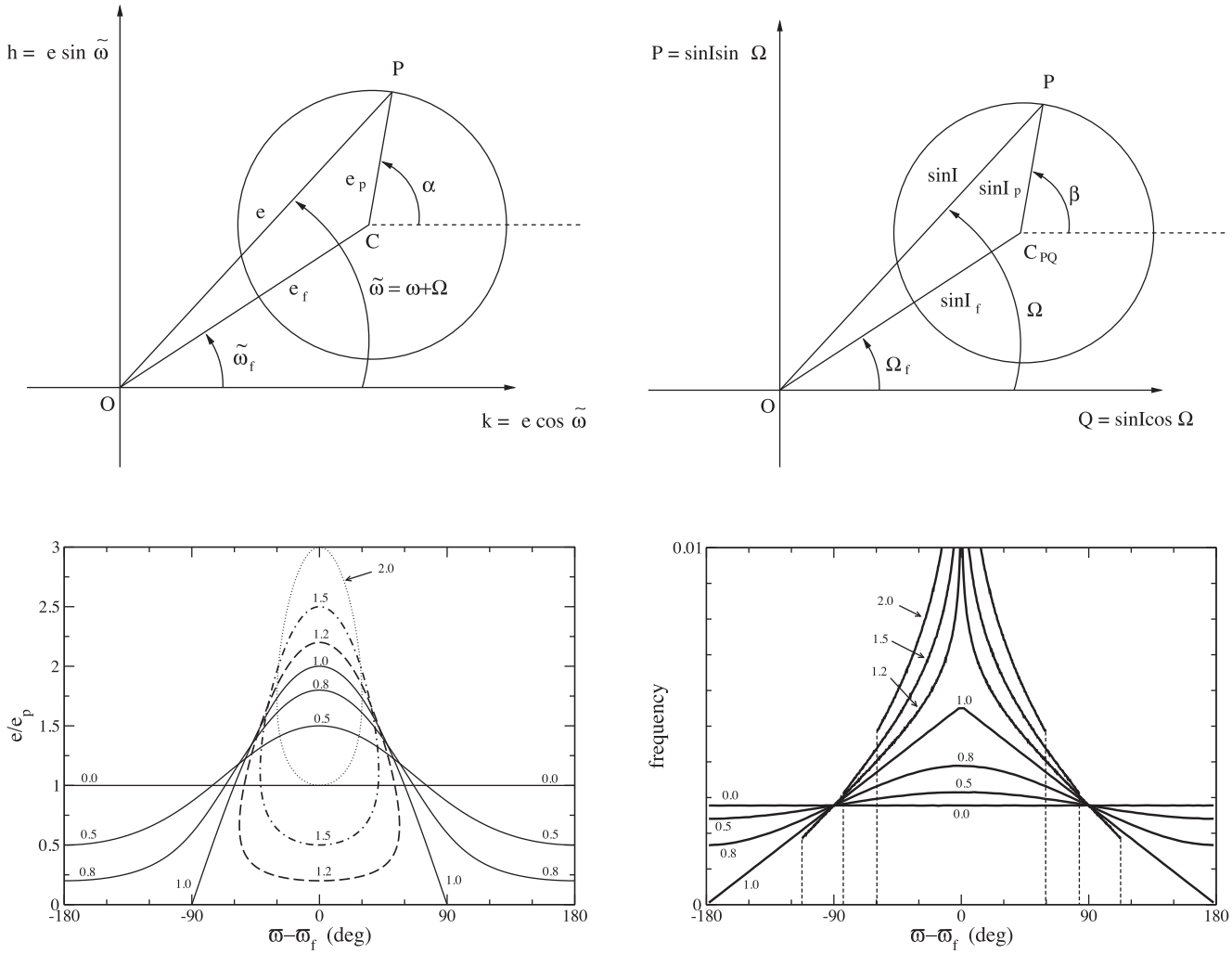

Figure 6. Top left: Plane of the non-singular elements h and k. Top right: Plane of the non-singular elements P and Q. Bottom left: Ratio e/e_p versus the difference $\tilde{\omega} - \tilde{\omega}_f$ (the numbers are the values of the ratio e_f/e_p). Bottom right: Distribution of the angle $\tilde{\omega} - \tilde{\omega}_f$ (the numbers are the values of the ratio e_f/e_p).

of collision depends on the values of e_f and I_f with respect to e_p and I_p (Fig. 6, bottom left). If the forced elements are small enough compared to the corresponding proper elements, the correlation tends to become weaker and the method of Dell'Oro & Paolicchi (1998) can again be used, such as for the large majority of the Main Belt asteroids.

The method described in this paper does not require any assumption about the instantaneous orbital elements of the bodies. So, we can use it to study the statistics of collisions among particles the orbits of which are affected by secular perturbations. The system is composed, besides the central mass m_0, of the perturber B and of a group of test bodies of type A, which are perturbed by B but do not perturb B. In other words, we are treating groups of particles the orbits of which are affected by the same perturber. We are interested in the statistics of close encounters among the test bodies. Each one of the test bodies is characterized by fixed values of the parameters a, e_p, e_f, I_p and I_f, while the fixed values of the forced angles $\tilde{\omega}_f$ and Ω_f are the same for all the test bodies, being specific to the orbit of the perturber only. On the other hand, the particular values of $\tilde{\omega}_f$ and Ω_f are not important because, since they are common to all the particles, a change of those parameters produces only rigid rotations of the entire system, without affecting the relative geometric configurations of the particle trajectories.

The sampling model required to randomly generate the samples (r_k, v_k) of the relative position and the relative velocity between two test bodies, which is necessary for equations (1)–(3), consists of the following procedure. Once two test particles are chosen, for each of them:

(S1) Random and independent values of the angles α and β are generated in the interval $[0, 2\pi]$.
(S2) The values of e and $\tilde{\omega}$ are computed from the fixed parameters e_f, e_p and $\tilde{\omega}_f$, and the current value of α.
(S3) The values of I and Ω are computed from the fixed parameters I_f, I_p and Ω_f, and the current value of β.
(S4) The value of the argument of the pericentre is simply derived as $\omega = \tilde{\omega} - \Omega$.
(S5) A random value of the mean anomaly M is generated in the interval $[0, 2\pi]$.
(S6) Finally, starting from a, e, I, ω, Ω and M, the vector of the position (x, y, z) and the vector of the velocity $(\dot{x}, \dot{y}, \dot{z})$ are computed.

This procedure is followed for each particle independently. As done in the previous section, the pair of particles can be fixed once for all or randomly chosen from a list of sets of elements (a, e_p, e_f, I_p, I_f). In the second case, the final results obtained by equations (1)–(3) correspond to the averages of the quantities $\phi(R)$, $\Psi(R, \mathcal{Q})$ and

$\langle q(R)\rangle$ with respect to the distribution of the secular elements of the particles a, e_p, e_f, I_p and I_f. Once we have obtained the components of the positions and the velocities of the two particles, we can determine the relative distance r and the relative velocity v, closing the sample model loop.

We examine an example of such a dynamical system consisting of a couple of particles sharing the same perturber. The semi-major axes are set very similar but not identical, precisely as 3.0 and 3.05 au. Note that if the semi-major axes were exactly the same, the sampling model (S1)–(S6) would not reproduce the dynamical properties of the system. In fact, in this case, the rate of variation of the mean anomalies would be the same, and the difference between the two mean anomalies would be constant in time. Consequently, the two mean anomalies cannot be generated independently, as in step (S5). The forced inclinations are set null, for simplicity. This last choice entails fixed instantaneous inclinations, which in this case are set to 2.5 and 5 deg. We assume that both particles have the same proper eccentricity and forced eccentricity, and that $e_p = e_f$ for both. We consider two cases, one with low proper/forced eccentricity 0.1 and the other with high proper/forced eccentricity 0.4.

Since our main goal here is to focus on the effects produced by the correlation between the eccentricity and the longitude of the pericentre, we compare this dynamical system with another pseudo-secular system sharing with the original one all the properties except for the correlation between e and $\tilde{\omega}$. The sampling model of this pseudo-secular system is obtained from that of the secular one previously described, except between steps (S5) and (S6) a randomization step (SP5bis) is inserted, which consists of forgetting the computed values of ω and Ω in steps (S3) and (S4) and substituting them with new values randomly and independently generated in the interval $[0, 2\pi]$. This modification is equivalent to replacing each particle with an ensemble of pseudo-orbits with the same semi-major axes but with eccentricities and inclinations obtained by randomly generating the angles α and β. These are circulating uniformly, i.e. they are characterized by a uniform distribution of the angles ω and Ω. The possible values of the eccentricities of these pseudo-orbits belong to the interval [0, 0.2] for $e_p = e_f = 0.1$, or [0, 0.8] if $e_p = e_f = 0.4$. In a few words, the pseudo-secular system is, in reality, a perfect canonical system, sharing with the real secular system the same distributions of the osculating eccentricities and inclinations. So, it is correct to use the methods of Bottke et al. (1994) or Dell'Oro & Paolicchi (1998) to compute the probability of collisions among the particles of the pseudo-secular system.

Fig. 7 shows the results of this exercise. The values of $\phi(R)/R^2$ and $\langle v(R)\rangle$ versus R for the secular system are represented with black dots, and with white dots for the pseudo-secular one. The horizontal long-dashed lines represent the mean intrinsic probability of collision P_i and the mean impact velocity U_m for the pseudo-secular systems computed using the methods of Bottke et al. (1994) and Dell'Oro & Paolicchi (1998). Again, for the pseudo-secular systems, the asymptotic values of $\phi(R)/R^2$ and $\langle v(R)\rangle$ are fully compatible with the corresponding values of P_i and U_m. The horizontal dotted lines represent the estimates of the asymptotic values of $\phi(R)/R^2$ and $\langle v(R)\rangle$ for $R \to 0$ for secular systems (black dots). The scattering of the points for $R \lesssim 10^6$ km is due to numerical random fluctuations of the sampling series of the phase space.

The impact velocities in the pseudo-secular systems are larger than in the corresponding secular systems because in the former case, all possible relative orientations among the orbits are allowed, in particular with large and small difference of the angles $\tilde{\omega}$. In the second case, the distribution of the difference of the angles $\tilde{\omega}$ has the triangular shape shown in Fig. 6, bottom right, for $e_f/e_p = 1$, with the consequence that the lines of the apses of the orbits tend to be oriented preferably towards the direction $\tilde{\omega} \sim \tilde{\omega}_f$, reducing the relative velocities at the close encounter.

For the frequency of the close encounters, the intrinsic rate $\phi(R)/R^2$ for secular systems (black dots) is systematically larger than the corresponding pseudo-secular case (white dots). This is due again to the non-random orientation of the lines of apses. The orbits getting closer reduces the available volume for the particles. Indeed, as pointed out by Farinella & Davis (1992), the order of magnitude of the intrinsic probability of collisions could be roughly estimated using a simple particle-in-a-box model as $P_i \approx \pi U/W$, where U is the average relative velocity and W the interaction volume of the particles. Labelling with the subscript s the secular case and with ps the pseudo-secular case, we can write $P_\mathrm{s}/P_\mathrm{ps} = (U_\mathrm{s}/U_\mathrm{ps})/(W_\mathrm{s}/W_\mathrm{ps})$. The reduction $W_\mathrm{s}/W_\mathrm{ps}$ of the available volume from the pseudo-secular to the secular case, i.e. from a completely random orientation of the apses to more or less a high degree of orbit polarization, is not easily computable a priori.

In Fig. 8, we show the distribution of the positions of the two particles, or more exactly of their coordinates (x, y) on the ecliptic plane generated by the sampling model. In the pseudo-secular systems, the particles move inside a torus-shaped region, centred at the origin of the coordinates (Sun), with internal radius $r_\mathrm{min} = a(1 - 2e_f)$ and external radius $r_\mathrm{min} = a(1 + 2e_f)$. For the secular systems, the interaction volume is again a torus, not centred around the origin but completely internal to the region filled by the pseudo-secular system. Note that, in the sampling model, we have set $\omega_f = 0$, so the mean line of the apses is along the x-axis and the mean pericentre has a positive abscissa. The distribution of the coordinate z is the same both in the pseudo-secular and in the corresponding secular system, because we have set $I_f = 0$ in both systems. So, the ratio $W_\mathrm{s}/W_\mathrm{ps}$ between the volumes of the secular and pseudo-secular tori are equal to the ratio between the areas of the regions filled with points in Fig. 8 at right and the corresponding areas at left. From the figure, it is possible to estimate that $W_\mathrm{s}/W_\mathrm{ps} \simeq 0.5$ in both cases $e_p = e_f = 0.1$ and $e_p = e_f = 0.4$. On the other hand, from Fig. 7, then $U_\mathrm{s}/U_\mathrm{ps} \simeq 0.76$ for $e_p = e_f = 0.1$ and $U_\mathrm{s}/U_\mathrm{ps} \simeq 0.58$ for $e_p = e_f = 0.4$. So, in conclusion $P_\mathrm{s}/P_\mathrm{ps}$ should be around 1.5 for $e_p = e_f = 0.1$, and $P_\mathrm{s}/P_\mathrm{ps} \simeq 1.16$ if $e_p = e_f = 0.4$. Such an evaluation, based on a simplistic particle-in-a-box model, overestimates the results of our simulations, which are respectively $\sim$1.16 and $\sim$1.05, but it helps us to understand the general trend shown in Fig. 7. In particular, in this case, the intrinsic probability of collision increases a little on taking into account the correlation between the eccentricities and longitude of the pericentre because the reduction of the interaction volume is more than balanced by the reduction of the average relative velocity.

We want to stress that the results shown in this section are not representative of the general features of the statistics of close encounters in systems affected by secular perturbations, and they are only an instructive example of the methodology presented in this paper. We refer to future and more comprehensive works on the effect of the secular perturbations on the statistics of impacts.

5 CONCLUSIONS

In this paper, we have presented a general method of studying the impact statistics among asteroids. The method belongs to the family of methods based on a Monte Carlo strategy that consists of deriving the collision rate and the related statistical distributions starting from a random sampling of the phase space. It differs from

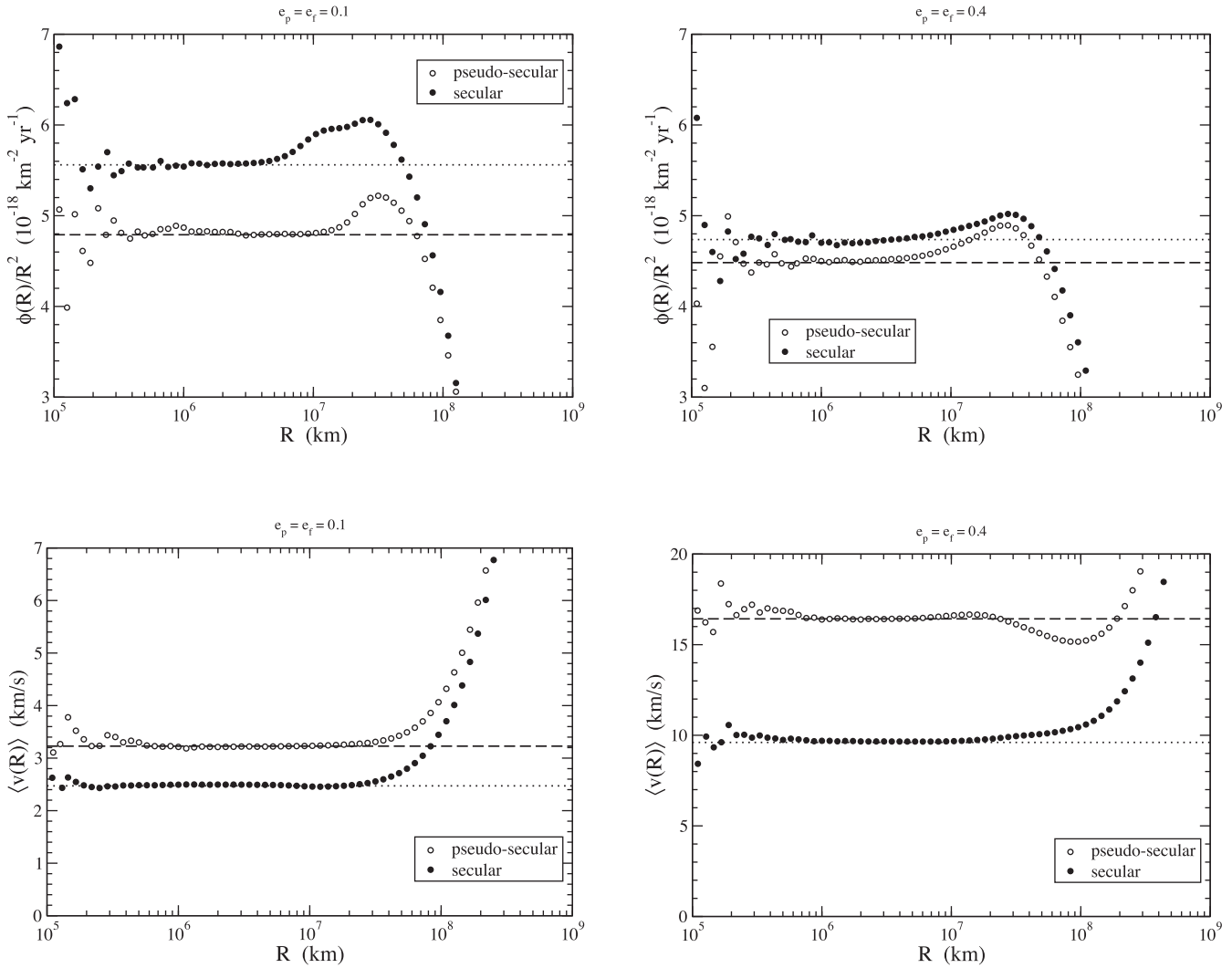

Figure 7. Statistics of collisions between two orbits evolving according to the linear perturbation theory (see text for details). The horizontal long-dashed lines represent P_i and U_m for the pseudo-secular systems computed using the method of Bottke et al. (1994). The dotted lines highlight the asymptotic values of $\phi(R)/R^2$ and $\langle v(R)\rangle$ for $R \to 0$. On the left, $e_p = e_f = 0.1$, while on the right, $e_p = e_f = 0.4$.

other Monte Carlo methods already described in the literature in the way it derives the rate of close encounters and the impact probability of collision from a series of the phase space samplings. The mathematical formalism to do this is very simple and consists only of the computation of sums of sample values of the relative velocity.

The method requires an algorithm to generate randomly the relative distance r and relative velocity v of the particles under investigation. This sampling algorithm must correctly reproduce the dynamical behaviour of the system, in the sense that in repeating a large number of times the random generation of the couples of values (r, v), their joint distribution must match the probability distribution for finding their values in the intervals $[r, r + \mathrm{d}r]$ and $[v, v + \mathrm{d}v]$ on observing the real system at an instant of time chosen at random. Indeed, the sampling model is not really part of our method, but is rather a prerequisite for using it. The way to sample the phase space depends on the particular case under investigation, and it characterizes the type of collision statistics. For instance, a flat random sampling of the pericentre arguments and node longitudes corresponds to the canonical statistics introduced by Öpik–Wetherill, while the procedure described in Section 4 reproduces the dynamical properties of systems affected by secular perturbations. In both cases, and in any other case, we are able to generate correctly the sampling values (r, v). Equations (1)–(4), which are the basic tools of the method, are always valid.

Equations (1)–(4) provide the mean temporal rate $\phi(R)$ of close encounters within R and the distribution Ψ of any parameters related to the kinematic circumstances of the close encounters. The validity of equations (1)–(4) relies on some simplified assumptions that are not valid in general, but which are more and more correct for smaller and smaller values of R, like the linearity of the relative motion of particles during the close encounters. On the other hand, the main goal of this method is not to produce reliable frequencies of close approaches for any value of R. This problem is well beyond the scope of this work. Instead, the present method is specifically designed to study the statistics of impacts among orbiting bodies, the sizes of which are very small compared to the dimensions of their orbits. For this reason, we are interested only in the asymptotic values of $\phi(R)$ and Ψ for $R \to 0$, for which the underlying assumptions are valid. On the other hand, computing $\phi(R)$ and Ψ for smaller and smaller values of R becomes more and more numerically

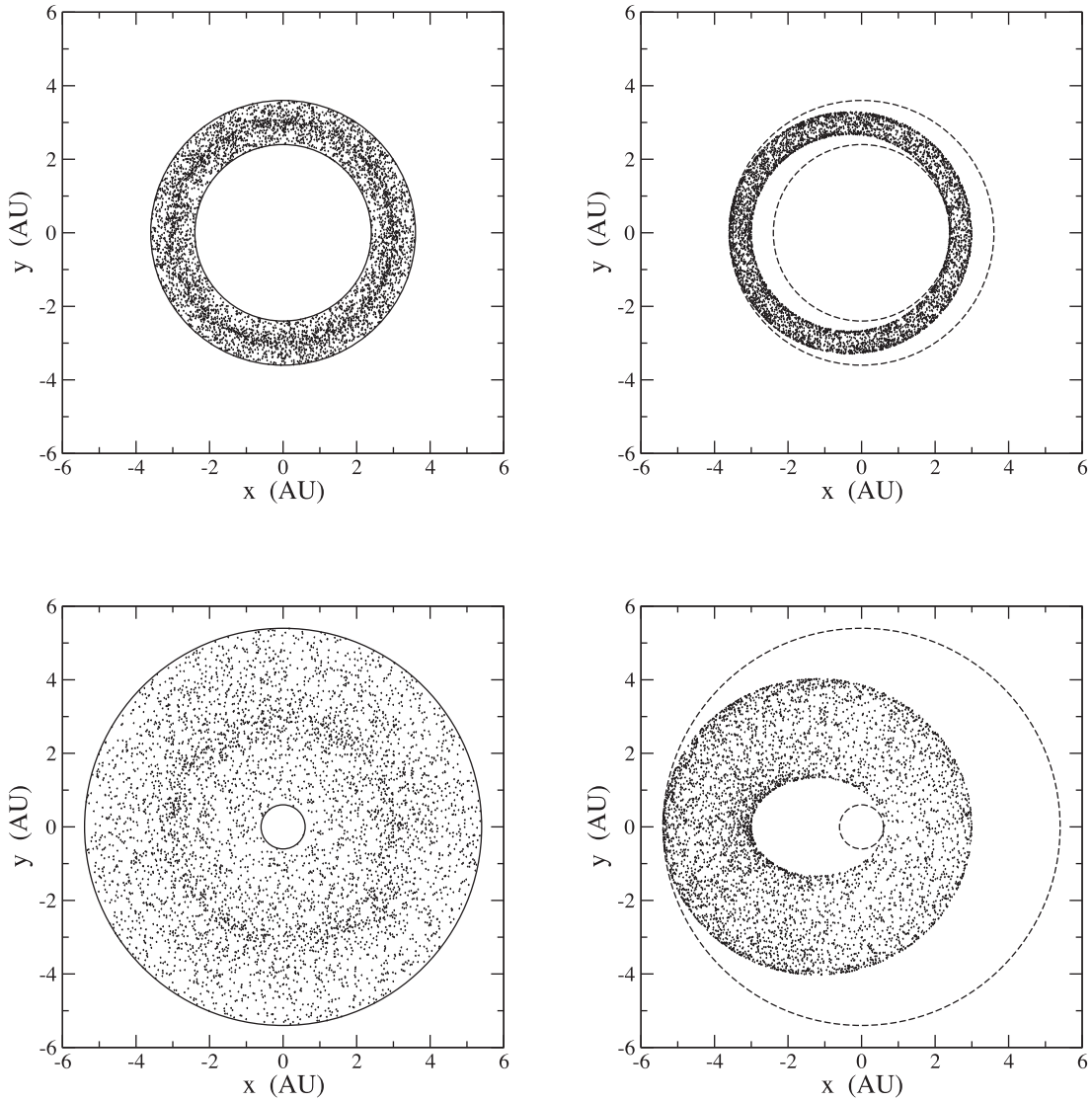

Figure 8. Distribution of the positions of the particles of the pseudo-secular (left) and secular (right) systems. Each dot represents the position of a particle projected on the ecliptic plane at a given epoch. At the top, $e_p = e_f = 0.1$, and at the bottom $e_p = e_f = 0.4$. In the plots on the right, the long-dashed circles redraw the regions filled by the points in the pseudo-secular system (at left).

prohibitive. So, the usual approach is to extrapolate $\phi(R)$ down to values of R comparable with the physical sizes of the bodies under investigation.

This extrapolation must be done carefully for two reasons. Two values $\phi(R_1)$ and $\phi(R_2)$ of the mean rate of close encounter computed by our method for two different values of R_1 and R_2 of the close encounter distance, with $R_2 > R_1$, will be affected by numerical random errors due to the fluctuations of the phase space sampling. Such errors are not independent because the samplings with $r < R_1$ contribute to the computation of both $\phi(R_1)$ and $\phi(R_2)$ (see equation 1). The correlation between the errors is larger when R_1 and R_2 are closer. For this reason, in some cases, the random sampling of the phase space produces false trends in the function $\phi(R)$ within intervals of under-sampled values of R, corresponding to large errors of $\phi(R)$, like that shown in Fig. 2. The second reason for caution in extrapolating is that, in general, there is no guarantee that the trend $\phi(R) \propto R^2$ really holds for $R \to 0$. This can happen for two nearly tangential but not crossing orbits, for which $\phi(R) = 0$ for values of R smaller than the minimum possible distance between the orbits, and that can be hardly explored by a random sampling. For these reasons, the extrapolation should be accompanied by a careful check of the possibility of intersection of the trajectories to identify when an orbital crossing is not possible for geometrical reasons or because of the effect of a dynamical protection mechanism (like that of the Neptune–Pluto system). This topic will be studied in detail in a forthcoming paper.

Concerning the hypothesis about the asymptotic trend of ϕ for $R \to 0$, note that some cases exist for which $\lim_{R \to 0} \phi(R) \propto R^{\gamma}$ with $\gamma \neq 2$ (Milani, Carpino & Marzari 1990; Vedder 1996). An example is two coplanar orbits with zero inclination, where for small values of R, the rate ϕ is proportional to R instead of R^2, simply because the third dimension is missing and we have a perfectly bi-dimensional problem. In this case, the relevant parameter is no longer $\phi(R)/R^2$ but rather the ratio $\phi(R)/R$, the limit of which for $R \to 0$ is a kind of intrinsic probability of collision measured in yr^{-1} km^{-1} Even if in this paper we focused on the standard case $\gamma = 2$, the method can be easily adapted to other situations, as briefly mentioned in Appendix A.

As in any Monte Carlo approach, even this method has the major drawback of being computationally slow. In general, the method requires a very large number of samples. However, for asteroids, it is prohibitive to obtain a reliable estimation of $\phi(R)$ for values of the order of 1–100 km. In the examples shown in this work, we explored the function $\phi(R)$ for values of R larger than a few 10^5 km, which was enough to obtain the intrinsic probability of collision, the mean impact velocity and its frequency distribution with acceptable accuracy. The computational effort depends on the number of samples necessary to sample the phase space densely enough, which in turn depends on the dynamical and geometrical properties of the system under investigation. The computational time depends on the kind of hardware available and the way the sampling model simulating the system is implemented. For instance, our software implementation of the sampling model for canonical statistics (Section 3) is able to generate 10^9 samplings in about 0.56 h using one single Intel i3-2100 CPU. The values of $\phi(R)/R^2$ and $\langle v(R)\rangle$ reported in Table F3 of Appendix F2 were obtained with simulations of 10^9 samplings for each of the targets in the list. The distributions shown in Fig. 3 at the top and at bottom left were obtained with a simulation of 10^{10} samplings (taking about 5 h), while the distribution at the bottom left required a much longer computational time (a 20-day-long simulation consisting of 10^{12} samplings).

On the other hand, as we have seen, the computational effort (and time) does not depend on the number of particles forming the system under investigation if we are interested in the mean value of the intrinsic probability of collision and the average impact velocity distribution. The plots in Fig. 5 showing $\phi(R)$ and $\langle v(R)\rangle$ versus R for a continuous distribution of orbits required the same computational time as for the single pair of orbits of Fig. 1. For this reason, the Monte Carlo approach is particularly suitable for studying the statistics of impacts in systems consisting of large numbers of particles, or even described by continuous distributions of mass.

The major advantage of a Monte Carlo approach is its flexibility together with robustness and ease of implementation. Methods based on an analytical description of the geometry of orbital crossing require the development of formulas (mainly integrals) specifically valid for a particular type of evolution of the orbits. Also, in the general method of Dell'Oro & Paolicchi (1998), the analytical form of the density distribution of the angular orbital elements has to be provided explicitly. This, in some cases, can be very difficult. Instead, in many cases, it is much easier to develop an algorithm to generate numerically the same orbital elements. Indeed, the general formulas of Section 2 are valid regardless of the way the samples (r, v) are generated, provided that such samplings faithfully reproduce the dynamics of the system.

We have discussed the formulation of the sampling model in the simple but important case of orbits with evolution governed by the canonical Öpik–Wetherill hypotheses (a problem already solved analytically), and then we extended the sampling model to include the secular perturbations produced by one perturbing body. Other kinds of dynamical systems could be treated if the corresponding sampling models are correctly developed. The extension of the sampling model described in Section 4 to secular perturbations due to two or more perturbing bodies is straightforward, because it simply consists of the superposition of further epicycles in the planes of the variables (h, k) and (P, Q). The dynamical behaviour of asteroids locked in some kinds of resonance, like mean motion or secular and Kozai resonances, can be reproduced if the orbital elements are randomly generated but taking into account the constraints due to the critical arguments characterizing the specific resonances. In this case, the constraints manifest as statistical correlations among the elements involved in the definition of the critical arguments. However, we do not intend discussing those cases here. The main aim of this paper is to provide only the basic principles of the method, with forthcoming papers focussing on the study and detailed discussion of the statistics of impacts in specific dynamical systems.

ACKNOWLEDGEMENTS

This research has made use of data and/or services provided by the International Astronomical Union's Minor Planet Center.

REFERENCES

Bottke W. F., Greenberg R., 1993, Geophys. Res. Lett., 20, 879
Bottke W. F., Nolan M. C., Greenberg R., Kolvoord R. A., 1994, Icarus, 107, 255
Dahlgren M., 1998, A&A, 336, 1056
Dell'Oro A., Paolicchi P., 1997, Planet. Space Sci., 45, 779
Dell'Oro A., Paolicchi P., 1998, Icarus, 136, 328
Dell'Oro A., Paolicchi P., Marzari F., Dotto E., Vanzani V., 1998, A&A, 339, 272
Dell'Oro A., Marzari F., Paolicchi P., Vanzani V., 2001, A&A, 366, 1053
Dell'Oro A., Paolicchi P., Cellino A., Zappalà V., 2002, Icarus, 156, 191
Dell'Oro A., Campo Bagatin A., Benavidez P. G., Alemañ R. A., 2013, A&A, 558, A95
Farinella P., Davis D. R., 1992, Icarus, 97, 111
Greenberg R., 1982, AJ, 87, 185
Horner J., Jones B. W., 2008, Int. J. Astrobiology, 7, 251
JeongAhn Y., Malhotra R., 2015, Icarus, 262, 140
Kessler D. J., 1981, Icarus, 48, 39
Knežević Z., Milani A., 2003, A&A, 403, 1165
Marsden B. G., 1980, Celest. Mech., 22, 63
Marzari F., Scholl H., Farinella P., 1996, Icarus, 119, 192
Milani A., Knežević Z., 1990, Celest. Mech. Dynamical Astron., 49, 247
Milani A., Carpino M., Marzari F., 1990, Icarus, 88, 292
Murray C. D., Dermott S. F., 1999, Solar System Dynamics. Cambridge Univ. Press, Cambridge
Öpik E. J., 1951, Proc. R. Irish Acad. Section A, 54, 165
Pokorny P., Vokrouhlicky D., 2013, Icarus, 226, 682
Rickman H., Wiśniowski T., Wajer P., Gabryszewski R., Valsecchi G. B., 2014, A&A, 569, A47
Vedder J. D., 1996, Icarus, 123, 436
Vedder J. D., 1998, Icarus, 131, 283
Vokrouhlický D., Pokorny P., Nesvorny D., 2012, Icarus, 219, 150
Wetherill G. W., 1967, J. Geophys. Res., 72, 2429
Wiźniowski T., Rickman H., 2013, Acta Astronomica, 63, 293

APPENDIX A: MEAN TRANSIT TIME ACROSS A SPHERE

One of the basic assumptions of the approach followed in this paper is that during the interval of time in which the relative distance between the two particles is smaller than R, the relative motion is rectilinear with constant velocity. In a reference frame where one particle (target) is at rest and placed at the origin of a Cartesian system of coordinates x, y and z, the position $\mathbf{r} = (x, y, z)$ of the other particle (projectile) at the time t is $\mathbf{r} = \mathbf{v}t + \mathbf{r}_0$, where $\mathbf{r}_0$ is the position at $t = 0$. The time t_m corresponding to the minimum value of $\mathbf{r}$ is $t_\mathrm{m} = -(\mathbf{v} \cdot \mathbf{r}_0)/v^2$, where v is the modulus of $\mathbf{v}$. For $t = t_\mathrm{m}$, the position of the projectile is $\mathbf{b} = \mathbf{r}_0 - (\hat{\mathbf{n}} \cdot \mathbf{r}_0)\hat{\mathbf{n}}$, where $\hat{\mathbf{n}}$ is the unit vector such that $\mathbf{v} = v\hat{\mathbf{n}}$, that is $\hat{\mathbf{n}}$ is the direction of the velocity. It is simple to see that $\mathbf{b}$ is always perpendicular to $\hat{\mathbf{n}}$, and it expresses the impact parameter of the close encounter. In other words, $\mathbf{b}$ is a vector in the plane $\mathcal{P}$ passing through the origin and

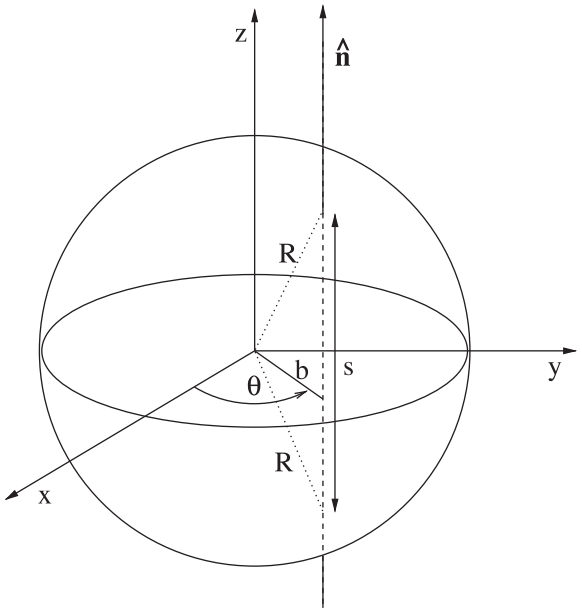

Figure A1. Geometric parameters of the intersection between a sphere and a line (see text for details).

perpendicular to $\hat{\mathbf{n}}$. From now on, we choose as $t = 0$ the epoch $t_{\rm m}$, and the motion of the projectile is given as $\mathbf{r} = v\hat{\mathbf{n}}t + \mathbf{b}$.

Let us imagine that we know only the value v, but not the other details about the motion of the projectile, that is neither the direction $\hat{\mathbf{n}}$ of the velocity vector nor the initial position $\mathbf{b}$. So, we assume that (1) the direction of the velocity is isotropically random and (2) the points described by the vector $\mathbf{b}$ are randomly but uniformly distributed on the plane $\mathcal{P}$. The hypothesis (2) is simply the expectation that the transit probability is proportional to the cross-section area. In the hypothesis that the modulus of $\mathbf{b}$ is equal or smaller than R, i.e. that the trajectory of the projectile really intersects the sphere of radius R centred on the origin, we want to compute the average crossing time. Since the modulus of the velocity is fixed, this problem is equivalent to computing the average length of the part of the trajectory inside the sphere, i.e. the length of the chord.

The length s of the chord depends only on the modulus of $\mathbf{b}$, regardless of the direction $\hat{\mathbf{n}}$, thanks to the spherical symmetry of the system. So, putting $\hat{\mathbf{n}} = (0, 0, 1)$, the plane $\mathcal{P}$ is the xy plane and $\mathbf{b} = (x, y, 0)$, where x and y are any values such that $x^2 + y^2 < R^2$ (Fig. A1). The length of the chord is $s = 2(R^2 - b^2)^{1/2}$, where $b^2 = x^2 + y^2$. The coordinates x and y are independent random deviates with uniform density distribution per unit area $\delta(x, y) = 1/(\pi R^2)$. Therefore, the mean value of s is expressed by the integral:

$$\overline{s} = \frac{2}{\pi R^2} \int\int_{x^2+y^2<R^2} (R^2 - b^2)^{1/2} {\rm d}x{\rm d}y, \tag{A1}$$

where the region of integration is the circle of radius R around the origin. This integral can be easily solved in polar coordinates θ and b, where $x = b\cos\theta$ and $y = b\sin\theta$:

$$\begin{aligned}\overline{s} &= \frac{2}{\pi R^2} \int_0^R \int_0^{2\pi} b(R^2 - b^2)^{1/2} {\rm d}\theta {\rm d}r \\ &= \frac{4}{\pi R^2} \int_0^R b(R^2 - b^2)^{1/2} {\rm d}r = \frac{4}{3} R,\end{aligned} \tag{A2}$$

where we have taken into account that the Jacobian of the transformation of the variables is $\partial(x, y)/\partial(\theta, b) = b$.

In conclusion, the mean duration of the transit inside the sphere of radius R is

$$\overline{\Delta t} = \frac{4}{3}\frac{R}{v}. \tag{A3}$$

Consider the special case of two-dimensional systems, where the sphere is substituted with a circle of radius R around the origin in the xy plane. In a similar way as before, we can consider the family of lines parallel to the y-axis, with x the abscissa of the point of their intersection with the x-axis. In this case, the length s of the chord inside the circle is $s = 2(R^2 - x^2)^{1/2}$. Assuming that the quantity x is a random variate with uniform density distribution $\delta(x) = 1/(2R)$ in the interval $[-R, +R]$, the mean value of s is

$$\overline{s} = \int_{-R}^{+R} \delta(x){\rm d}x = \frac{1}{R}\int_{-R}^{+R} (R^2 - x^2)^{1/2}{\rm d}x = \frac{\pi}{2}R, \tag{A4}$$

entailing a mean transit time inside the circle $\overline{\Delta t} = (\pi/2)(R/v)$.

APPENDIX B: RATE OF CLOSE ENCOUNTERS

Let us consider two particles, both moving inside a limited region of space. This means that the particles cannot get away to infinity. Let $r(t)$ be the relative distance between the particles at the epoch t. We are interested in the close encounters within R, i.e. in events during which the two particles pass close to each other within a distance R. In fact, since r is a function of time, a close encounter occurs at the epoch t' if $r(t) < R$ for $t \in [t', t' + \Delta t]$, and $r(t) > R$ for t just before t' and just after $t' + \Delta t$ (Fig. B1). The quantity Δt is the duration of the close encounter. Note that the ensemble of the close encounters within R includes also the ensemble of close encounters within a distance $R' < R$.

During an interval of time T, a given number $N(R)$ of close encounters within a mutual distance $r < R$ occur. By definition, the mean temporal rate of the close encounters is

$$\phi(R) = \frac{N(R)}{T}. \tag{B1}$$

We can rewrite this quantity in the following way:

$$\phi(R) = \frac{N(R)}{S(R)}\frac{S(R)}{T} = \frac{p(R)}{\tau(R)}, \tag{B2}$$

where $S(R)$ is the sum of the durations of the close encounters that occurred within R during the interval of time T. Or in other words,

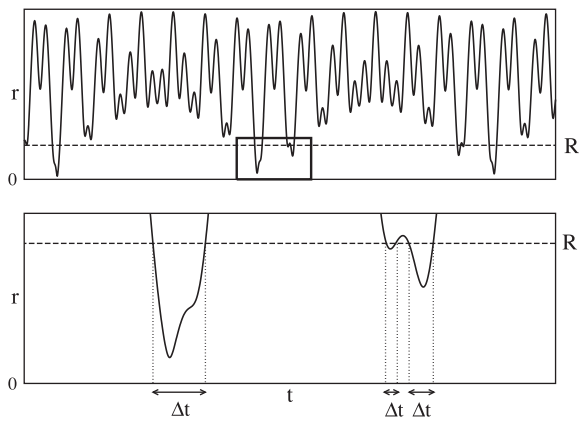

Figure B1. Example of function $r(t)$ for two points moving along Keplerian orbits. At the bottom, some intervals of time during which $r < R$ are highlighted.

$S(R)$ is the total amount of time during which the mutual distance between the two particles is less than R. Therefore, the ratio

$$p(R) = \frac{S(R)}{T} \tag{B3}$$

is the probability of finding, *at any randomly chosen instant of time*, the two particles at a distance smaller than R. The ratio

$$\tau(R) = \frac{S(R)}{N(R)} \tag{B4}$$

is the average duration of close encounters within R.

Moreover, we assume that during the interval of time when the mutual distance r is less than R, the relative motion is rectilinear with constant velocity v. However, the relative velocity v has a different value in each close encounter. Let $\mathrm{d}N(R, v)$ be the number of close encounters within R *and* with relative velocity within the infinitesimal interval $[v, v + \mathrm{d}v]$. So, we can define the mean temporal rate of close encounters with velocity v as:

$$\mathrm{d}\phi(R, v) = \frac{\mathrm{d}N(R, v)}{T}. \tag{B5}$$

This quantity can be rewritten as:

$$\mathrm{d}\phi(R, v) = \frac{\mathrm{d}N(R, v)}{\mathrm{d}S(R, v)} \frac{\mathrm{d}S(R, v)}{T}, \tag{B6}$$

where $\mathrm{d}S(R, v)$ is the sum of the durations of the close encounters with relative velocity v and $r < R$. As before, the ratio

$$\tau(R, v) = \frac{\mathrm{d}S(R, v)}{\mathrm{d}N(R, v)} \tag{B7}$$

is the mean value of the duration of the close encounters with velocity v and within $r < R$. From equation (A3), we can put

$$\tau(R, v) = \frac{4}{3} \frac{R}{v}. \tag{B8}$$

Moreover,

$$\mathrm{d}p(R, v) = \frac{\mathrm{d}S(R, v)}{T} \tag{B9}$$

is the probability of finding, *at any randomly chosen instant of time*, the two particles at a distance r smaller than R and with relative velocity in the infinitesimal interval $[v, v + \mathrm{d}v]$.

Taking into account that

$$\phi(R) = \int_0^\infty \frac{\mathrm{d}\phi(R, v)}{\mathrm{d}v} \mathrm{d}v,$$

from equations (B6), (B8) and (B9), it follows that

$$\phi(R) = \frac{3}{4} \frac{1}{R} \int_0^\infty v \frac{\mathrm{d}p(R, v)}{\mathrm{d}v} \mathrm{d}v. \tag{B10}$$

Note that $\phi(R)$ is not proportional to $p(R)$. While $\phi(R)$ is the *temporal* rate of close encounters, or, in other words, $\phi(R)$ is the mean number of times per unit of time that $r(t)$ becomes less than R, $p(R)$ is simply the probability that $r(t) < R$ if the epoch t is chosen at random. Let us imagine a simplified situation in which N close encounters occur during the interval of time T, and the duration Δt is the same for all events. In this case, $\phi = N/T$ and $p = N\Delta t/T$. If the duration of close encounters was shorter (for instance if the particles were faster), the probability p would be smaller, but the frequency ϕ would be the same.

In conclusion, from (B10), we can compute correctly the temporal frequency of the close encounters within a distance R if we are able to evaluate the probability $\mathrm{d}p(R, v)$ to find the two particles at a mutual distance smaller than R and with a relative velocity in the interval $[v, v + \mathrm{d}v]$. This goal can be easily reached if we have at our disposal a suitable sampling model correctly describing the dynamics of the system during the interval of time T. A sampling model is an algorithm able to generate the mutual distance r and the relative velocity v at an instant of time chosen at random. Given that, let us imagine we can generate a series of n independent pairs (r_k, v_k) using our sampling model. Let $\mathrm{d}n(R, v)$ be the number of cases for which $r_k < R$ and $v_k \in [v, v + \mathrm{d}v]$. Then, by definition of the probability as the ratio between the number of favourable outcomes and the total number of possible outcomes:

$$\frac{\mathrm{d}n(R, v)}{n} = \mathrm{d}p(R, v). \tag{B11}$$

Substituting this expression for $\mathrm{d}p(R, v)$ in equation (B10), we can write:

$$\phi(R) = \frac{3}{4} \frac{1}{R} \frac{1}{n} \int_0^\infty v \frac{\mathrm{d}n(R, v)}{\mathrm{d}v} \mathrm{d}v, \tag{B12}$$

but, since $v\mathrm{d}n(R, v)$ is nothing else than the sum of the sample values v_k for which $r_k < R$ and $v_k \in [v, v + \mathrm{d}v]$, it is easy to realize that the integral is simply the sum of all generated values v_k for which $r_k < R$. Therefore,

$$\phi(R) = \frac{3}{4} \frac{1}{R} \frac{1}{n} \sum_{k:r_k<R} v_k = \frac{3}{4} \frac{1}{R} \frac{w(R)}{n}, \tag{B13}$$

where

$$w(R) = \sum_{k:r_k<R} v_k. \tag{B14}$$

In conclusion the mean temporal rate of close approaches at a distance r smaller than R is simply related to the ratio between the sum of sample values v_k for which $r_k < R$ and the total number of samples. The ratio $w(R)/n$ is not the average of the values v_k because the number of terms of the sum $w(R)$ is lower than n. Alternatively, it is possible to rewrite the last formula in this way:

$$\phi(R) = \frac{3}{4} \frac{1}{R} \left(\frac{w(R)}{n(R)} \right) \left(\frac{n(R)}{n} \right), \tag{B15}$$

where $n(R)$ is the number of samples for which $r_k < R$, and $w(R)/n(R)$ is the mean value of the sample velocities v_k.

Finally, we note that for a bi-dimensional system, the factor $3/(4R)$ in equations (B10) and (B13) is replaced by $2/(\pi R)$.

APPENDIX C: DISTRIBUTIONS OF CLOSE ENCOUNTERS

Usually we are not interested only in the temporal rate of close encounters, since we need to know the statistical distribution of the values of any parameter q related to the close encounters. As for the relative velocity, we assume that the value of the parameter q does not change during each close approach. So, among all close encounters occurring during the interval of time T, we select those with $r < R$ *and* q belonging to an ensemble of values $\mathcal{Q}$, and their number is $N(R, \mathcal{Q})$. The corresponding temporal frequency is

$$\phi(R, \mathcal{Q}) = \frac{N(R, \mathcal{Q})}{T}. \tag{C1}$$

For the subset of the close encounters with $q \in \mathcal{Q}$, equation (B10) can be generalized as

$$\phi(R, \mathcal{Q}) = \frac{3}{4} \frac{1}{R} \int_0^\infty v \frac{\mathrm{d}p(R, v, \mathcal{Q})}{\mathrm{d}v} \mathrm{d}v, \tag{C2}$$

where $\mathrm{d}p(R, v, \mathcal{Q})$ is the probability of finding, at any instant of time, the system with $r < R$, $v \in [v, v + \mathrm{d}v]$ and $q \in \mathcal{Q}$.

The fraction of events per unit of time with $r < R$ and $q \in \mathcal{Q}$, is

$$\Psi(R, \mathcal{Q}) = \frac{\phi(R, \mathcal{Q})}{\phi(R)}, \tag{C3}$$

or, in other words, $\Psi(R, \mathcal{Q})$ is the statistical distribution of the variable q. The corresponding continuous distribution is obtained be letting the ensembles $\mathcal{Q}$ become infinitesimal.

To evaluate the expected value of the parameter, let us consider the total ensemble of all values of q. If $\mathrm{d}\Psi(R, q) = \mathrm{d}\phi(R, q)/\phi(R)$ is the frequency of the close approaches with the parameter between q and $q + \mathrm{d}q$, the expected value of q is, by definition,

$$\langle q(R) \rangle = \int q \frac{\mathrm{d}\Psi(R, q)}{\mathrm{d}q} \mathrm{d}q, \tag{C4}$$

where $\langle q(R) \rangle$ highlights that we are talking about the mean value of the parameter q for the close encounter within a distance R. So, from equations (C3) and (C2), we have:

$$\langle q(R) \rangle = \frac{\int \int q v \frac{\mathrm{d}p(R, v, q)}{\mathrm{d}q \mathrm{d}v} \mathrm{d}q \mathrm{d}v}{\int v \frac{\mathrm{d}p(R, v)}{\mathrm{d}v} \mathrm{d}v}. \tag{C5}$$

Note that the true statistical distribution $\Psi(R, \mathcal{Q})$ is proportional to $\phi(R, \mathcal{Q})$ and not to the probability $p(R, \mathcal{Q})$. The quantity $p(R, \mathcal{Q})$ is not the temporal frequency or the mean number of times per unit of time that $r(t)$ becomes smaller than R while $q \in \mathcal{Q}$; $p(R, \mathcal{Q})$ is only the probability that, choosing at random the epoch t, $r < R$ and $q \in \mathcal{Q}$. Indeed, for a given ϕ, i.e. for a given number of favourable events during the period T, p increases if the corresponding durations Δt increase (for example, if the relative velocity during the close encounters is smaller), but ϕ does not change. For instance, in a simplified case, let us imagine we have only two different kinds of close approaches during the interval T: $N/2$ events of type A for which $q = q_A$ and duration $\Delta t = 1$, and $N/2$ events of type B with duration $\Delta t = 2$ and $q = q_B$. If we choose at random an instant of time, the probability of it falling in an event A is half of the probability of it falling in an event B. However, the temporal frequency of the two kinds of events is the same, that is $N/(2T)$. In addition, the expected value of q is $(q_A + q_B)/2$ and not $(q_A + 2q_B)/3$.

We can evaluate numerically the integrals (C2) and (C5) using a sampling model, following the same approach described in Appendix B to evaluate the temporal frequency $\phi(R)$. This time our sampling model has to provide for each sample pair (r_k, v_k) the corresponding sample value q_k of the parameter. If, using the sample model, we generate a random sequence of n triplets (r_k, v_k, q_k), with n large enough, the probability of finding $r_k < R$, $v_k \in [v, v + \mathrm{d}v]$ and $q_k \in \mathcal{Q}$ is given by:

$$\mathrm{d}p(R, v, \mathcal{Q}) = \frac{\mathrm{d}n(R, v, \mathcal{Q})}{n}, \tag{C6}$$

where $\mathrm{d}n(R, v, \mathcal{Q})$ is the number of the random samples with $r_k < R$, $v_k \in [v, v + \mathrm{d}v]$ and $q_k \in \mathcal{Q}$. So, substituting in (C2):

$$\phi(R, \mathcal{Q}) = \frac{3}{4} \frac{1}{R} \frac{1}{n} \int_0^\infty v \frac{\mathrm{d}n(R, v, \mathcal{Q})}{\mathrm{d}v} \mathrm{d}v, \tag{C7}$$

but again the integral is simply the sum of sample values of the velocity v_k for which $r_k < R$ and $q_k \in \mathcal{Q}$. Therefore,

$$\phi(R, \mathcal{Q}) = \frac{3}{4} \frac{1}{R} \frac{w(R, \mathcal{Q})}{n}, \tag{C8}$$

where

$$w(R, \mathcal{Q}) = \sum_{k: r_k < R, q_k \in \mathcal{Q}} v_k. \tag{C9}$$

In conclusion, substituting equations (C8) and (B13) into equation (C3), we have:

$$\Psi(R, \mathcal{Q}) = \frac{\sum_{k: r_k < R, q_k \in \mathcal{Q}} v_k}{\sum_{k: r_k < R} v_k}. \tag{C10}$$

For instance, the frequency of close approaches with relative velocity belonging to an interval of values Δv is obtained as:

$$\phi(R, \Delta v) = \frac{3}{4} \frac{1}{R} \frac{\sum_{k: r_k < R, v_k \in \Delta v} v_k}{n}, \tag{C11}$$

and consequently the normalized frequency distribution of the relative velocity built with respect to a series of bins Δv_j is $\Psi(R, \Delta v_j) = \phi(R, \Delta v_j)/\phi(R)$.

Finally, for the expected value of q, from equation (C5), it is easy to see that

$$\langle q(R) \rangle = \frac{\sum_{k: r_k < R} q_k v_k}{\sum_{k: r_k < R} v_k}, \tag{C12}$$

where the final sums include all the generated pairs (r_k, v_k) with $r_k < R$, and q_k is the corresponding value of the parameter for each pair. Summarizing, the expected value of a parameter q for the close approaches within R is simply the weighted average of the sample values q_k by the corresponding relative velocity.

Important cases of expected values are the mean value of the relative velocity and its square:

$$\langle v(R) \rangle = \frac{\sum_{k: r_k < R} v_k^2}{\sum_{k: r_k < R} v_k} \tag{C13}$$

and

$$\langle v^2(R) \rangle = \frac{\sum_{k: r_k < R} v_k^3}{\sum_{k: r_k < R} v_k}. \tag{C14}$$

APPENDIX D: MULTIPLE PAIRS OF PARTICLES

In Appendices B and C, we focused on close encounters between two particles only. If we are interested in the statistics of close approaches among m different particles, the important quantity to be determined is the mean rate of close approaches, defined as the average of the rates for each possible couple of particles. It is given by

$$\overline{\phi(R)} = \frac{1}{m(m-1)} \sum_{i \neq j} \phi_{ij}(R), \tag{D1}$$

where $\phi_{ij}(R)$ is the rate of close encounters between the ith and jth particles. It is important to note that, by definition, also the pairs

for which no close encounter within R is possible contribute to the mean rate of close encounters with null terms in the sum (D1).

At first sight, it seems to be necessary to evaluate one by one the rates $\phi_{ij}(R)$ for each couple (i, j) of particles before we can compute the overall average $\overline{\phi(R)}$. This is not necessarily true.

Let us consider a list of n_{ij} sample pairs of relative distances and velocities (r_{ijk}, v_{ijk}) generated by a suitable sampling model for the couple of orbits of the ith and the jth particles. From equation (B13), we know that

$$\phi_{ij}(R) = \frac{3}{4}\frac{1}{R}\frac{w_{ij}(R)}{n_{ij}}, \tag{D2}$$

where

$$w_{ij} = \sum_{k:r_{ijk}<R} v_{ijk}. \tag{D3}$$

Substituting into (D1), we obtain:

$$\overline{\phi(R)} = \frac{3}{4}\frac{1}{R}\frac{1}{m(m-1)}\sum_{i\neq j}\frac{1}{n_{ij}}\sum_{k:r_{ijk}<R} v_{ijk}. \tag{D4}$$

If, for each pair of particles (i, j), the corresponding list of random samplings (r_{ijk}, v_{ijk}) contains the same number of samples $n_{ij} = \overline{n}$, the last equation becomes

$$\overline{\phi(R)} = \frac{3}{4}\frac{1}{R}\frac{1}{n}\sum_{k:r_{ijk}<R} v_{ijk}, \tag{D5}$$

where n is

$$n = m(m-1)\overline{n}, \tag{D6}$$

the total number of samples for all the particle pairs.

Equation (D5) is formally like (B13), but it is valid only if the random samples are evenly allocated among all the $m(m-1)$ possible pairs of particles. From an operative point of view, we can generate all the samples (r_{ijk}, v_{ijk}) for all the couples of particles (i, j) in a unique sequence, provided that every time we generate a different sample we randomly chose the two particles i and j (with $i \neq j$). In this way, each pair (i, j) has the same probability of being selected, and the number of samples n_{ij} for each pair is, on average, the same for all pairs.

Another important case is the statistics of close encounters between a given test particle (target) and an ensemble of other particles (impactors). If the number of impactors is m, the mean rate of close encounters is

$$\overline{\phi(R)} = \frac{1}{m}\sum_i \phi_i(R), \tag{D7}$$

where $\phi_i(R)$ is the rate of close encounters between the target and the ith impactor. In this case, the formula analogous to (D5) is

$$\overline{\phi(R)} = \frac{3}{4}\frac{1}{R}\frac{1}{n}\sum_{k:r_{ik}<R} v_{ik}, \tag{D8}$$

provided that any time we generate the sample (r_{ik}, v_{ik}), the index i of the of the impactor is randomly chosen. Again, n is the total number of samples.

As discussed in detail in Appendix E, the value of $\phi(R)$ provided by equation (B13) is not deterministic but it is rather a random deviate fluctuating around a mean value. Both equations (D1) and (D7) are special cases of the general expression:

$$\overline{\phi(R)} = \frac{1}{M}\sum_{i=1}^{M}\phi_i(R), \tag{D9}$$

where M is the number of terms of the sum, which in the two previous examples can be equal to $m(m-1)$ or m. So, $\overline{\phi(R)}$ is a random deviate too. From equation (E12), the variance of $\phi(R)$ is inversely proportional to the number of samples used to evaluate it. Therefore, the variance σ_i^2 of each term $\phi_i(R)$ can be written as c_i^2/n_i, where c_i^2 is a constant and n_i the number of samples used to compute $\phi_i(R)$. Since the fluctuations of one term are not correlated with the fluctuations of another, the variance of $\overline{\phi(R)}$ is

$$\sigma^2 = \frac{1}{M^2}\sum_{i=1}^{M}\sigma_i^2 = \frac{1}{M^2}\sum_{i=1}^{M}\frac{c_i^2}{n_i}. \tag{D10}$$

If for all terms $n_i = \overline{n}$, the last equation can be rewritten as:

$$\sigma^2 = \frac{1}{n}\left(\frac{1}{M}\sum_{i=1}^{M}c_i^2\right) = \frac{\overline{c}^2}{n} = \frac{\overline{c}^2}{\overline{n}M}, \tag{D11}$$

where $n = M\overline{n}$ is the total number of samples and $\overline{c}^2$ is the average of the constants c_i^2. In conclusion, σ^2 is M times smaller than the typical variance $\overline{c}^2/\overline{n}$ of the single terms. In other words, given the desired level of accuracy in computing $\overline{\phi(R)}$, i.e. the desired value of σ^2, requiring a total number of samples n, it is not necessary to evaluate each of the M terms $\phi_i(R)$ with the same accuracy, but it is sufficient for each one to have a number of samples equal to $\overline{n} = n/M$. This means that the computational time necessary to evaluate $\overline{\phi(R)}$ does not depend on the number m of particles. The same argument is valid also for the average in equation (D7), computed using equation (D8).

APPENDIX E: ERROR EVALUATION

According to equation (B13), the computation of the rate $\phi(R)$ of close approaches within a given distance R, using a sampling model, requires the generation of a finite number n of sample values (r_k, v_k) of the relative position and velocity, and the computation of the sum $w(R) = \sum_{k:r_k<R} v_k$. The sum contains a number $n(R)$ of terms corresponding to the number of samples for which the corresponding relative distance r_k is less than R. Each v_k and the number of generated samples $n(R)$ are random variables. Consequently $w(R)$ and $\phi(R)$ computed by equation (B13) are random variables too. Therefore, depending on the exact sequence of the random sampling, at each repeated computation, the value of $w(R)$ (and $\phi(R)$) fluctuates according to a particular parent distribution. Here we want to provide an estimation of the entity of such random fluctuations. To do this we have to study the statistical properties of the random variate $w(R)$.

Let $\psi(r, v)$ be the parent distribution of the pair of random variables (r, v), corresponding to a dynamical status of the system and generated using a suitable sampling model. Let $\psi(v|R)$ be the marginal distribution of the relative velocity for all the pairs with $r < R$:

$$\psi(v|R) = \int_0^R \psi(v, r)\mathrm{d}r. \tag{E1}$$

Let $v(R)$ be the generic value of the relative velocity for dynamical states with $r < R$. The expected value of the sum $w(R)$ is the product of the expected number of terms of the sum, i.e. the expected value of $n(R)$, and the expected value of $v(R)$:

$$\mathrm{E}\{w(R)\} = \mathrm{E}\{n(R)\}\mathrm{E}\{v(R)\}, \tag{E2}$$

where

$$\mathrm{E}\{v(R)\} = \int_0^\infty v\psi(v|R)\mathrm{d}v. \tag{E3}$$

Computing the square of $w(R)$,

$$w^2(R) = \sum_i v_i^2 + \sum_{i \neq j} v_i v_j, \tag{E4}$$

we can write

$$\begin{aligned} \mathrm{E}\{w^2(R)\} &= \mathrm{E}\{n(R)\}\mathrm{E}\{v^2(R)\} \\ &\quad + \mathrm{E}\{n^2(R) - n(R)\}\mathrm{E}\{v_i v_j\}_{i \neq j}, \end{aligned} \tag{E5}$$

since $n^2(R) - n(R)$ is the number of terms of the second sum, and where

$$\mathrm{E}\{v^2(R)\} = \int_0^\infty v^2 \psi(v|R) \mathrm{d}v. \tag{E6}$$

On the other hand, each random sample is generated independently and no correlation exists between v_i and v_j if $i \neq j$. Therefore,

$$\mathrm{E}\{v_i v_j\}_{i \neq j} = \mathrm{E}\{v_i\}\mathrm{E}\{v_j\} = \mathrm{E}^2\{v(R)\}, \qquad i \neq j. \tag{E7}$$

Substituting (E7) into (E5), we obtain:

$$\mathrm{E}\{w^2(R)\} = \mathrm{E}\{n(R)\}\mathrm{Var}\{v(R)\} + \mathrm{E}\{n^2(R)\}E^2\{v(R)\}, \tag{E8}$$

where $\mathrm{Var}\{v(R)\}$ is the variance of $v(R)$:

$$\mathrm{Var}\{v(R)\} = \mathrm{E}\{v^2(R)\} - \mathrm{E}^2\{v(R)\}. \tag{E9}$$

Finally, from (E9) and (E2), we can write the variance of $w(R)$:

$$\begin{aligned} \mathrm{Var}\{w(R)\} &= \mathrm{E}\{w^2(R)\} - \mathrm{E}^2\{w(R)\} \\ &= \mathrm{E}\{n(R)\}\mathrm{Var}\{v(R)\} + \mathrm{Var}\{n(R)\}\mathrm{E}^2\{v(R)\}. \end{aligned} \tag{E10}$$

We can simplify this last equation taking into account that the number $n(R)$ of samples (r, v) with $r < R$ is a binomial random variate with probability of success much lower than 1, or, in other words, it is a Poisson random variable. In this case, we can put $\mathrm{E}\{n(R)\} = \mathrm{Var}\{n(R)\}$ and the last equation becomes:

$$\mathrm{Var}\{w(R)\} = \mathrm{E}\{n(R)\}\mathrm{E}\{v^2(R)\}. \tag{E11}$$

Once we have obtained an evaluation of $\mathrm{Var}\{w(R)\}$, we can use it to compute the variance of $\phi(R)$, or of any other parameter the definition of which contains only $w(R)$ as random variate (for example the ratio $\phi(R)/R^2$ for the computation of the intrinsic probability of collision). In particular,

$$\mathrm{Var}\{\phi(R)\} = \frac{9}{16}\frac{p(R)}{R^2}\mathrm{E}\{v^2(R)\}\frac{1}{n}, \tag{E12}$$

where $p(R)$ is the probability that r_k is smaller than R and then $\mathrm{E}\{n(R)\} = np(R)$.

In practical terms, in a random generation of n samples (r_k, v_k), of which $n(R)$ samples have $r_k < R$, our best estimations of $\mathrm{E}\{n(R)\}$ and $\mathrm{E}\{v^2(R)\}$ are

$$\mathrm{E}\{n(R)\} \sim n(R)$$

and

$$\mathrm{E}\{v^2(R)\} \sim \frac{1}{n(R)} \sum_{k:r_k<R} v_k^2.$$

Then our best estimation of $\mathrm{Var}\{w(R)\}$ is simply:

$$\mathrm{Var}\{w(R)\} \sim \sum_{k:r_k<R} v_k^2. \tag{E13}$$

The evaluation of the variance of the sample mean of a parameter q as computed from equation (C12) with a finite number n of samples is more difficult, requiring the study of the ratio of two random variates $\sum_{k:r_k<R} q_k v_k$ and $\sum_{k:r_k<R} v_k$. We limit ourselves only to a rough estimation. If $n(R)$ is the number of samples with $r_k < R$ and σ is the sample standard deviation of the parameter q, evaluated using the $n(R)$ values q_k, we simply put $\sigma/\sqrt{n(R)}$ as the uncertainty of the computation of the expected value of q.

The previous evaluation of the uncertainties has been numerically verified in some test cases, which consists of repeating the computations of $\phi(R)$ from equation (B13) and $\mathrm{Var}\{w(R)\}$ from equation (E13) with independent sampling series. The random fluctuations of those values were fully compatible with the expected analytical estimation.

APPENDIX F: VALIDATION TESTS

In this appendix, we report the results of some tests done to validate the Monte Carlo algorithm described in this paper, and in particular to check the correctness of equations (B13), (C11) and (C13) using independent methods.

As test cases, we consider a set of systems with canonical behaviour, i.e. groups of intersecting orbits with fixed semi-major axes a, eccentricities e and inclinations I, and with longitudes of the nodes Ω and arguments of the pericentres ω circulating uniformly with time. The motion along each orbit is obviously fully described by Kepler's second law, that is in terms of uniform variation of the mean anomaly M with time.

For each case, we computed $\phi(R)/R^2$ versus R using equation (B13), the distribution $\Psi(R, \Delta v)$ of the relative velocity using equation (C11) for different values of R, and the value of $\langle v(R) \rangle$ versus R using equation (C13). The sampling algorithm used to generate the series of samples (r_k, v_k) consists of the steps (C1)–(C3) described in Section 3.

Moreover, for each test case, we computed the mean intrinsic probability of collision P_i, the frequency distribution of the impact velocity $\Psi(U)$ and its mean value U_m, using one or more classical

Table F1. First 10 numbered Main Belt asteroids with the orbital elements used in this paper for the validation tests.

Name	a (au)	e	I (deg)
(1) Ceres	2.76797	0.075783	10.592
(2) Pallas	2.772	0.231158	34.84
(3) Juno	2.67107	0.25583	12.987
(4) Vesta	2.36191	0.088834	7.14
(5) Astraea	2.57348	0.191188	5.369
(6) Hebe	2.42604	0.201507	14.748
(7) Iris	2.3858	0.231116	5.522
(8) Flora	2.20158	0.156688	5.887
(9) Metis	2.38657	0.122178	5.574
(10) Hygiea	3.14212	0.114692	3.838

Table F2. Collisions among the first 10 numbered Main Belt asteroids (see text for details): mean intrinsic probability of collision P_i expressed in 10^{-18} km^{-2} yr^{-1}, and mean impact velocity U_m in km s^{-1}, computed using the methods of Dell'Oro & Paolicchi (1998) and Bottke et al. (1994). See text for the meaning of the two cases A and B.

Case	P_i	U_m
A	3.169 ± 0.002	5.217 ± 0.001
B	5.035 ± 0.002	5.910 ± 0.002

methods based on the Öpik–Wetherill dynamical hypotheses. As a first check, we used the method of Dell'Oro & Paolicchi (1998), in the version corresponding to the particular case of canonical statistics [see Dell'Oro & Paolicchi (1998) for details]. The results are then cross-checked using the independent method of Bottke et al. (1994). In this appendix, we refer to the two methods of Bottke et al. (1994) and Dell'Oro & Paolicchi (1998) as integral methods, because they consist of computing an integral that gives the value of the intrinsic probability of collision or the distribution of the impact velocity, which differs from our method based on a Monte Carlo simulation. Recall that our method does not directly provide P_i and $U_{\rm m}$, but its basic outputs are the values of $\phi(R)$ and $\langle v(R)\rangle$ as functions of R, from which P_i and $U_{\rm m}$ are then estimated.

The validations consist of verifying:

(i) If a threshold R_ϕ can be identified such that $\phi(R)$ tends to be proportional to R^2 for $R < R_\phi$, and if the asymptotic value of the ratio $\phi(R)/R^2$ for $R \to 0$ is compatible with the value of P_i (apart from random fluctuations).

(ii) If a threshold R_v can be identified such that $\langle v(R)\rangle$ tends to be constant for $R < R_v$, and if the asymptotic value of $\langle v(R)\rangle$ for $R \to 0$ is compatible with the value of $U_{\rm m}$ (apart from random fluctuations).

(iii) If the limit distribution $\Psi(R, \Delta v)$ of the relative velocity computed for small values of R is compatible with the distribution $\Psi(U)$ of the impact velocity (apart from random fluctuations). In practical terms, this test is done comparing $\Psi(U)$ with $\Psi(R, \Delta v)$ computed for $R < R_v$.

F1 Multiple pairs of orbits

In Appendix D, we have described how to use our Monte Carlo method for systems composed of more than two particles, in order to compute the average value $\overline{\phi(R)}$ of the close encounter rate over different pairs of particles. We have shown how to avoid the two-step procedure of calculating at first the values of $\phi(R)$ for each pair of particles and then computing their mean value. In fact with our technique, it is possible to directly compute the mean value with only one sampling of the phase space of the ensemble of particles as a whole. A test concerning the close encounters among a group of different orbits has been done to verify the correctness of the equations (D5) and (D8). The orbits used in this test are those of the

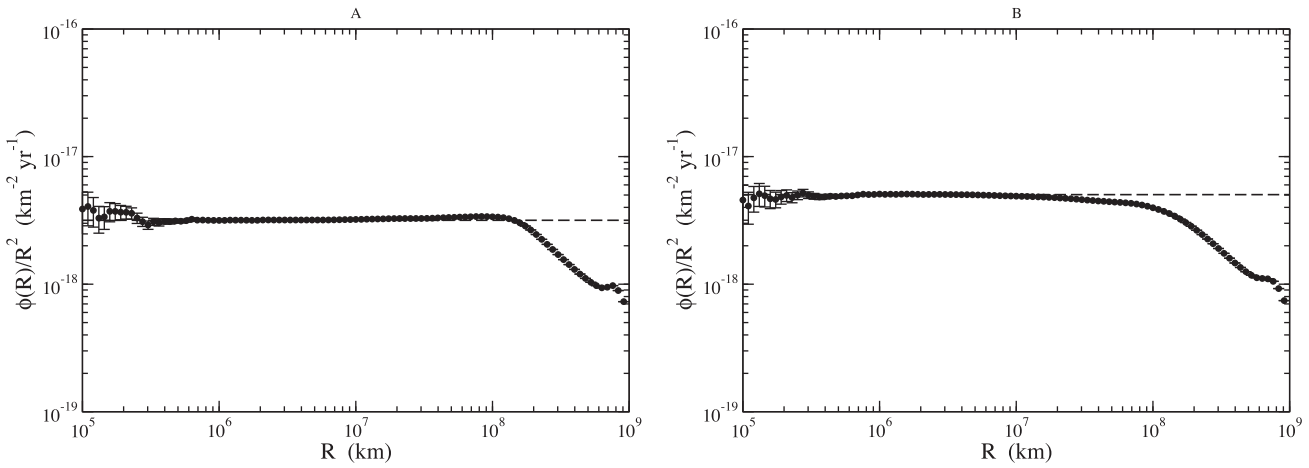

Figure F1. Close approaches among the first 10 numbered Main Belt asteroids. Value of $\phi(R)/R^2$ versus R. The horizontal long-dashed line represents the value of P_i. Left: Collision between (1) Ceres and one of the other nine asteroids (A in the text). Right: Collisions between all possible pair of orbits (B).

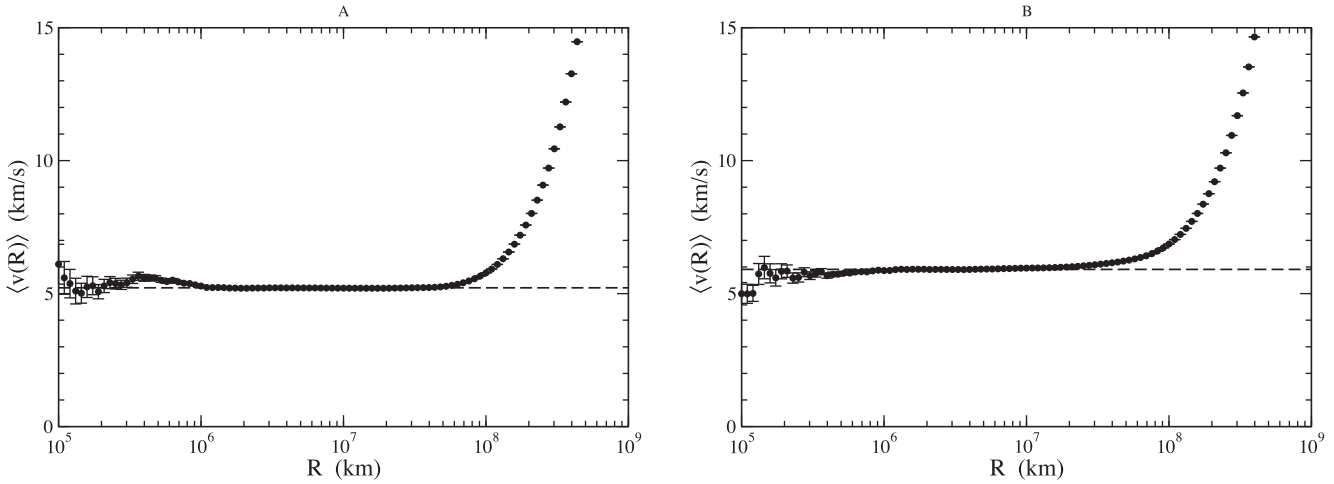

Figure F2. The mean relative velocity during close approach versus R, for the same cases of Fig. F1. The horizontal long-dashed line represents the value of $U_{\rm m}$.

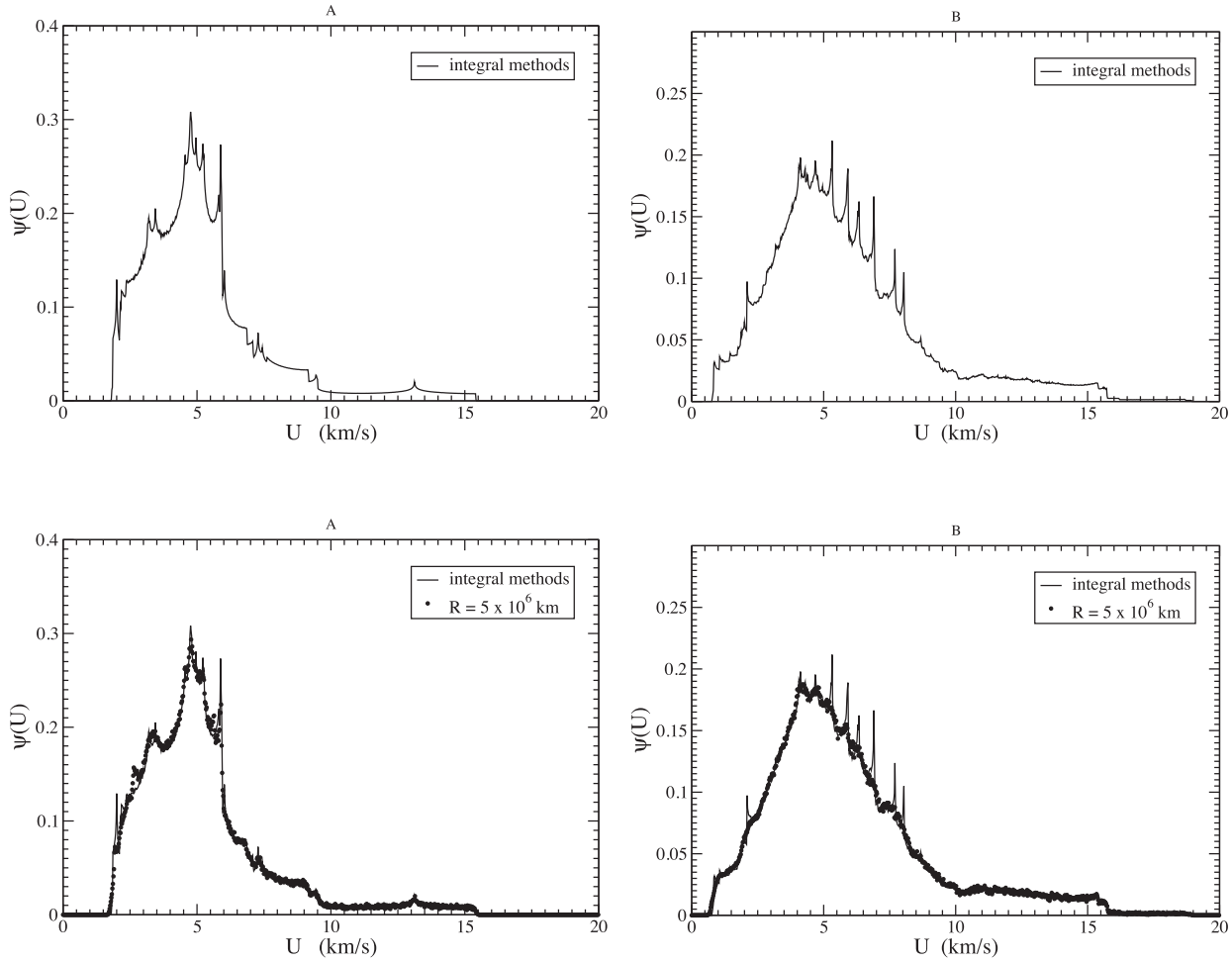

Figure F3. Distributions of the impact velocities among the first 10 number Main Belt asteroids. Solid lines: Distributions of the impact velocity computed using the integral methods. Black dots: Encounter velocity distribution produced by the present method for the close encounters within a distance $R = 5 \times 10^6$ km. Left: Collision between (1) Ceres and the other nine asteroids. Right: Collisions between all possible pair of orbits.

first 10 numbered Main Belt asteroids. The values of their orbital elements used in our computations are listed in Table F1.

We have considered two different cases:

A: The rate, and corresponding distribution of the relative velocity, of close approaches between (1) Ceres and the other asteroids of the group.

B: The rate, and corresponding distribution of the relative velocity, of close approaches between any member of the group and a different member of the group.

For A, nine different pairs of orbits exists, while for B, 45 combinations are possible.

Table F2 reports the values of the mean intrinsic probability of collision P_i and mean impact velocity $U_{\rm m}$, computed using the integral methods. For the two cases A and B, we computed $\phi(R)/R^2$ and $\langle v(R)\rangle$ for different values of R using our method. The results are plotted in Figs F1 and F2, respectively. Long-dashed horizontal lines in both figures represent the values of P_i and $U_{\rm m}$ in Table F2. In both cases, the asymptotic values of $\phi(R)/R^2$ and $\langle v(R)\rangle$ for $R < 10^7$ km are compatible with the values of P_i and $U_{\rm m}$, respectively, apart from the numerical fluctuation for values close to 10^5 km. The evidence that the mean relative velocity $\langle v(R)\rangle$ becomes constant for values of R of the order of 10^7 km suggests that the the distribution of the relative velocity does not change any more below

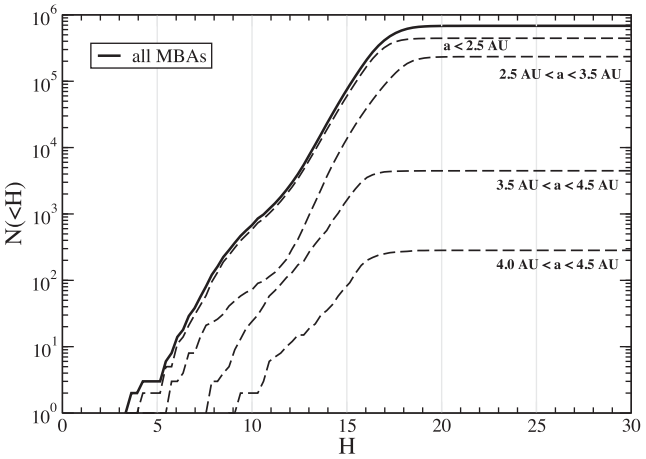

Figure F4. Cumulative distribution of the absolute magnitudes of the Main Belt asteroids (data from Minor Planet Center, March 2016). The bold line represents the distribution of all Main Belt asteroids, while long-dashed lines are the distributions for some sub-groups, selected according to different intervals of the semi-major axis.

Table F3. Statistics of collisions between selected targets and all Main Belt asteroids with absolute magnitude less than 15. Shown are the mean intrinsic probability of collision P_i and the mean impact velocity $U_{\rm m}$ computed by the integral methods, compared, respectively, to the estimated asymptotic values of $\phi(R)/R^2$ and $\langle v(R)\rangle$ for $R \to 0$, obtained using the method described in this work. For the parameter Δ/σ, see text.

Target	P_i (10^{-18} km^{-2} yr^{-1})	$\lim_{R\to 0} \phi(R)/R^2$ (10^{-18} km^{-2} yr^{-1})	Δ/σ	$U_{\rm m}$ (km s^{-1})	$\lim_{R\to 0} \langle v(R)\rangle$ (km s^{-1})	Δ/σ
(1) Ceres	3.661 ± 0.002	3.64 ± 0.02	−0.88	4.7773 ± 0.0009	4.80 ± 0.01	2.4
(2) Pallas	1.974 ± 0.003	1.95 ± 0.02	−0.88	11.719 ± 0.002	11.73 ± 0.03	0.23
(4) Vesta	3.809 ± 0.002	3.79 ± 0.02	−0.85	4.549 ± 0.001	4.551 ± 0.009	0.17
(5) Astraea	4.711 ± 0.003	4.71 ± 0.02	−0.23	4.8701 ± 0.0008	4.860 ± 0.009	−1.2
(10) Hygiea	2.698 ± 0.002	2.72 ± 0.02	1.3	4.217 ± 0.001	4.23 ± 0.01	1.3
(15) Eunomia	3.518 ± 0.003	3.54 ± 0.02	0.97	5.670 ± 0.001	5.66 ± 0.01	−0.45
(20) Massalia	5.117 ± 0.003	5.10 ± 0.02	−0.59	4.4114 ± 0.0009	4.429 ± 0.008	2.2
(21) Lutetia	4.846 ± 0.003	4.83 ± 0.02	−0.81	4.208 ± 0.001	4.224 ± 0.008	2.1
(24) Themis	3.364 ± 0.002	3.36 ± 0.02	−0.35	4.020 ± 0.001	4.056 ± 0.008	4.4
(93) Minerva	3.708 ± 0.003	3.69 ± 0.02	−0.68	4.839 ± 0.001	4.85 ± 0.01	1.5
(135) Hertha	5.112 ± 0.004	5.10 ± 0.02	−0.45	4.544 ± 0.001	4.571 ± 0.008	3.3
(145) Adeona	3.575 ± 0.003	3.58 ± 0.02	0.30	5.498 ± 0.001	5.51 ± 0.01	1.1
(158) Koronis	4.382 ± 0.003	4.35 ± 0.02	−1.8	3.465 ± 0.001	3.482 ± 0.006	2.7
(170) Maria	3.353 ± 0.003	3.36 ± 0.02	0.19	6.061 ± 0.001	6.07 ± 0.01	0.70
(221) Eos	3.038 ± 0.002	3.03 ± 0.02	−0.31	4.4392 ± 0.0008	4.45 ± 0.01	1.4
(243) Ida	4.423 ± 0.002	4.38 ± 0.02	−2.1	3.4624 ± 0.0008	3.477 ± 0.006	2.3
(253) Mathilde	4.157 ± 0.003	4.17 ± 0.02	0.73	5.2546 ± 0.0008	5.26 ± 0.01	0.82
(434) Hungaria	1.346 ± 0.004	1.37 ± 0.02	1.7	8.961 ± 0.007	9.00 ± 0.03	1.2
(490) Veritas	2.30 ± 0.04	2.27 ± 0.02	−0.80	4.255 ± 0.008	4.26 ± 0.01	0.031
(668) Dora	3.601 ± 0.002	3.60 ± 0.02	−0.21	5.1605 ± 0.0007	5.16 ± 0.01	−0.38
(847) Agnia	4.296 ± 0.002	4.29 ± 0.02	−0.26	3.7702 ± 0.0008	3.773 ± 0.007	0.42
(951) Gaspra	3.720 ± 0.004	3.70 ± 0.02	−0.73	4.848 ± 0.002	4.85 ± 0.01	0.048
(1040) Klumpkea	2.040 ± 0.002	2.03 ± 0.02	−0.46	6.791 ± 0.001	6.82 ± 0.02	1.5
(1726) Hoffmeister	4.188 ± 0.002	4.19 ± 0.02	0.24	3.7540 ± 0.0009	3.768 ± 0.007	1.9
(2076) Levin	3.860 ± 0.003	3.86 ± 0.02	0.055	4.722 ± 0.001	4.74 ± 0.01	2.2
(2867) Steins	3.386 ± 0.002	3.38 ± 0.02	−0.38	5.118 ± 0.001	5.10 ± 0.01	−1.4

$R \sim 10^7$ km. In Fig. F3, we have plotted the distribution of the impact velocity provided by the integral methods, drawn as a solid line, superimposed on the distribution of the encounter velocity for close approaches within a distance less than 5×10^7 km obtained with the present method and represented by black dots. The distributions of the impact velocity show the characteristic peak-like features due to the contribution of the different pairs of orbits of the group, as in Fig. 3. It is difficult to reproduce with our method the details of each peak of the impact velocity distribution, as this requires an excessive oversampling of the phase space and consequently a very long computational time. On the other hand, the distributions obtained with the present method reproduce well the fundamental features of the impact velocity distribution.

F2 Collisions in the Main Belt

Finally, we performed a series of tests to study the statistics of collisions between a list of targets and the whole population of the Main Belt asteroids. The targets were selected from objects having some particular scientific interest, such as the largest remnant of an asteroid family (and, therefore, with an orbit probably close to that of the family's parent body) or an asteroid visited by a probe during a space mission. For the list of impactors, to mitigate the effect of the observational bias on the distribution of orbital elements, we selected only the Main Belt asteroids with absolute magnitude H less than 15. As shown in Fig. F4, the list of known Main Belt asteroids with $H < 15$ is probably complete, regardless of the interval of the semi-major axes. The list of selected projectiles includes 71 066 objects. The data were downloaded from the Minor Planet Center website (Marsden 1980)[1] on 2016 March 17.

For the targets, we used the proper elements (Milani & Knežević 1990) from the AstDys-2 website (Knežević & Milani 2003).[2] The reason is that osculating elements of Main Belt asteroids have short- and long-term variations due to planetary perturbations. The probability of collision depends on the orbital elements, but we cannot know which osculating orbit the target had at the moment of collision. Proper elements, obtained by filtering out secular variations, represent in some sense a sort of average orbit of the target. For projectiles, we used the osculating elements instead. The reason is that the distribution of osculating elements represents a sort of snapshot of the typical dynamical configuration of the collisional environment, including the extension of its variations.

As always done in this paper, the values of P_i and $U_{\rm m}$ for each target include the contributions of all selected projectiles, regardless of whether the orbit of a particular projectile can cross the orbit of the target. In the second case, the projectile contributes with a null term to the average value P_i. The advantage of this approach is that the mean number N of impacts undergone by a target of diameter D_0 with Main Belt asteroids during a time span T can be written as

$$N = \frac{1}{4} P_i T \int (D_0 + D)^2 F(D) \mathrm{d}D, \tag{F1}$$

where $F(D)$ is the size distribution of the overall population, without worrying about which asteroids can actually collide with the target.

[1] http://www.minorplanetcenter.net/
[2] http://hamilton.dm.unipi.it/astdys/

The results of this test are shown in Table F3. For each selected target, the mean intrinsic probability of collision P_i and the mean impact speed $U_{\rm m}$ obtained using the integral methods are reported, together with the corresponding estimated limit values of $\phi(R)/R^2$ and $\langle v(R)\rangle$ for $R \rightarrow 0$, obtained using our method. The parameter Δ/σ is defined as follows: if x' is the value provided by the present method and σ' its error, and x_0 and σ_0 are, respectively, the same quantity and its error computed using the integral methods, then

$$\frac{\Delta}{\sigma} = \frac{x' - x_0}{\sqrt{\sigma'^2 + \sigma_0^2}}. \tag{F2}$$

The value of the ratio Δ/σ tells us how much the two values x' and x_0 are statistically different. If the two methods are equivalent, we expect that Δ/σ would be of the order of a few units. Table F3 confirms the mutual consistency of the two methods. In some cases, for $U_{\rm m}$ and the corresponding $\lim_{R\rightarrow 0}\langle v(R)\rangle$, like (24) Themis, the value of Δ/σ is rather large. This is because our rough determination of the error of $\langle v(R)\rangle$ (see Appendix E) can in some cases underestimate the real size of numerical fluctuations. This has been confirmed by repeating the computation of the values in Table F3. Moreover, the values of $\lim_{R\rightarrow 0}\langle v(R)\rangle$ computed by our method seem systematically slightly larger than the corresponding $U_{\rm m}$. This depends on the values of R we use to estimate $\lim_{R\rightarrow 0}\langle v(R)\rangle$ and whether the function $\langle v(R)\rangle$ versus R is perfectly constant in this range of values of R. In general, $\langle v(R)\rangle$ decreases with decreasing R, and the limit value of $\langle v(R)\rangle$ for $R \rightarrow 0$ is, thus, always smaller than all other values. A similar situation, although less evident in Table F3, occurs for P_i: in general, but not always, $\phi(R)/R^2$ increases with decreasing R. Therefore, its limit value for $R \rightarrow 0$ is, in this case, an upper limit, resulting in it being slightly underestimated.

This paper has been typeset from a TEX/LATEX file prepared by the author.