PUB. IRMA, LILLE 2009
Vol. 69, Nº VIII

Order statistics and heavy-tail distributions for planetary perturbations on Oort cloud comets *

R. S. STOICA, S. LIU, Yu. DAVYDOV
Université Lille 1
Laboratoire Paul Painlevé, CNRS UMR 8524
59655 Villeneuve d'Ascq Cedex, France

M. FOUCHARD, A. VIENNE
Université Lille 1, LAL, 59000 Lille, France
Institut de Mécanique Céleste et Calcul d'Ephémérides,
77 av. Denfert-Rochereau, 75014 Paris, France

G . B. VALSECCHI
INAF-IASF, via Fosso del Cavaliere 100, 00133 Roma, Italy

Abstract

This paper tackles important aspects of comets dynamics from a statistical point of view. Existing methodology uses numerical integration for computing planetary perturbations for simulating such dynamics. This operation is highly computational. It is reasonable to wonder whenever statistical simulation of the perturbations can be much more easy to handle. The first step for answering such a question is to provide a statistical study of these perturbations in order to catch their main features. The statistical tools used are order statistics and heavy tail distributions. The study carried out indicated a general pattern exhibited by the perturbations around the orbits of the important planet. These characteristics were validated through statistical testing and a theoretical study based on Öpik theory.

Résumé

Cet article aborde des aspects de la dynamique des comètes du point de vue statistique. La méthodologie existante utilise l'intégration numérique pour calculer les perturbations planètaires dans une telle dynamique. Cette opération est très coûteuse en temps de calcul. Il est donc raisonable de se demander si la simulation statistique de ces perturbations ne constitue un cadre plus convenable pour ce problème. Ainsi, le premier pas, pour répondre à une telle question, est de fournir une étude statistique qui décrit ces perturbations. Les outils statistiques utilisés sont les statistiques d'ordre et les lois à queue lourde. Les resultats de cette étude

*Preprint

ont montré que ces perturbations induisent une structure spatiale autour de l'orbite des grandes planètes. Ces caractéristiques ont été validées en utilisant des tests statistiques et la théorie d'Öpik.

Keywords: order statistics, heavy-tail distributions, Celestial mechanics, Oort cloud

1 Introduction

Comet dynamics is one of the most difficult phenomena to model in celestial mechanics. Indeed their dynamics is strongly chaotic, thus individual motions of known comets are hardly reproducible for more than a few orbital periods. When the origin of comets is under investigation, one is thus constrained to make use of statistical tools in order to model the motion of a huge number of comets supposed to be representative of the actual population. Such statistical model should also be reliable on a time scale comparable to the age of the solar system.

Due to their very elongated shapes, comet trajectories are affected by planetary perturbations during close encounters with planets. Such perturbations turn out to be the main mechanisms able to affect comet trajectories. Consequently, it is of major importance to model these perturbations in a way which is statistically reliable and with the lowest cost in computing time.

A direct numerical integration of a 6 bodies restricted problem (Sun, Jupiter, Saturn, Uranus, Neptune, Comet) each time a comet enters the planetary region of the Solar System is not possible due to the cost in computer time.

Looking for an alternative approach, we can take advantage of the fact that planetary perturbation on Oort cloud comets are uncorrelated. In fact the orbital period of such comets are so much larger than those of the planets, that when the comet returns, the phases of the latter can be taken at random. Thus we can build a synthetic integrator à la Froeschlé and Rickman [15] to speed up the modeling. The criticism by [14] to such an approach does not apply in the present case because, as just said, successive planetary perturbations on an Oort cloud comets are uncorrelated.

The aim of this paper is to give a statistical description of a large set of planetary perturbations assumed to be representative of those acting on Oort cloud comets entering the planetary region. To this purpose we use order statistics and heavy tails distributions.

The rest of this paper is organised as follow. Section 2 is devoted to the presentation of the mechanism producing the data, *i.e.* the planetary perturbations and the statistical tools used to analyse the data. These tools are order statistics and heavy-tail distributions, that allow, respectively, the study and the modeling of the data distribution, with respect to its symmetry, skewness and tail fatness. The obtained results are shown and interpreted in the third section. The results are finally analysed from a more theoretical point of view using the Öpik theory in Section 4. The paper closes with conclusions and perspectives.

2 Statistical tools

2.1 Data compilation

By planetary perturbations, one intends the variations of the orbital parameters between their values before entering the planetary region of the Solar System, *i.e.* the barycentric orbital element of the osculating cometary orbit $(z_i, q_i, \cos i_i, \omega_i, \Omega_i)^T$ (where q, i, ω, Ω are the perihelion distance, the inclination, the argument of perihelion and the

longitude of the ascending node and $z = -1/a$ with a the semi-major axis), and their final values $(z_f, q_f, \cos i_f, \omega_f, \Omega_f)^T$, that is either when the comet is at its aphelion or when it is back on a keplerian barycentric orbit.

Between its initial and final values, the system Sun + Jupiter + Saturn + Uranus + Neptune + comet is integrated using the RADAU integrator at the 15th order [10] for a maximum of 2 000 yrs. Then the planetary perturbation obtained through this integration is $(\Delta z = z_f - z_i, \Delta q = q_f - q_i, \Delta \cos i = \cos i_f - \cos i_i, \Delta\omega = \omega_f - \omega_i, \Delta\Omega = \Omega_f - \Omega_i)^T$. The detail on the numerical experiment used to perform the integrations may be found in [26].

Repeating the above experiment with a huge number of comets (namely 9 600 000), one gets a set of planetary perturbations. The comets are chosen with uniform distribution of the perihelion distance between 0 and 32 AU, cosine of the ecliptic inclination between -1 and 1 and argument of perihelion, longitude of the ascending node between 0 and 360°. The initial mean anomaly is chosen such that the perihelion passage on its initial keplerian orbit occurs randomly with an uniform distribution between 500 and 1 500 years after the beginning of the integration.

In the present study, because the perturbations are mainly depending on q_i and $\cos i_i$ [13], each perturbation is associated to the couple $(\cos i_i, q_i)$. Similarly, since the orbital energy is the main quantity which is affected by the planetary perturbations, we will consider only these perturbations here.

Consequently, our data are composed by a set of triplets $(\cos i_i, q_i, Z)$ where $Z = z_f - z_i$ denotes the perturbations of the cometary orbital energy by the planets, and $(\cos i_i, q_i)$ a point in a space denoted by K. In the following, we call Z the perturbation mark.

2.2 Exploratory analysis based on order statistics

Let $Z_1, \ldots, Z_n$ be a sequence of independent identically distributed random variables and let $F(z) = P(Z \leq z), z \in \mathbb{R}$ be the corresponding cumulative distribution function. Let us consider also Σ_n, the set of permutations on $\{1, \ldots, n\}$.

The order statistics of the sample $(Z_1, \ldots, Z_n)$ is the rearrangement of the sample in increasing order and it is denoted by $(Z_{(1,n)}, \ldots, Z_{(n,n)})$. Hence $Z_{(1,n)} \leq \ldots, \leq Z_{(n,n)}$ and there exists a random permutation $\sigma_n \in \Sigma_n$ such that

$$(Z_{(1,n)}, \ldots, Z_{(n,n)}) = (Z_{\sigma_n(1)}, \ldots, Z_{\sigma_n(n)}). \tag{1}$$

In the following, some classical results from the literature are presented [4, 8]. If F is continuous, then almost surely $Z_{(1,n)} < \ldots, < Z_{(n,n)}$ and the permutation σ_n in definition (1) is unique. If Z_1 has a probability density f, then the probability density of the order statistics is given by

$$n!\mathbf{1}\{z_1 < \ldots z_n\} f(z_1) \ldots f(z_n).$$

A major characteristic of order statistics is that they allow quantiles approximations. The quantiles are one of the most easy to use tool for characterising a probability distribution. In practice, the data distribution can be described by such empirical quantiles.

Two important results are now presented. The first result shows how to compute empirical quantiles using order statistics. Let us assume that F is continuous and there exists an unique solution z_q to the equation $F(z) = q$ with $q \in (0,1)$. Clearly, z_q is the $q-$quantile of F. Let $(k(n), n \geq 1)$ be an integers sequence such that $1 \geq k(n) \geq n$ and $\lim_{n\to\infty} \frac{k(n)}{n} = q$. Then the sequence of the empirical quantiles $(Z_{(k(n),n)}, n \geq 1)$ converges almost surely towards z_q.

The second result allows the computation of confidence intervals and hypothesis testing. If Z_1 has a continuous probability density f such that $f(z_q) > 0$ for $q \in (0,1)$ and if it is supposed that $k(n) = nq + o(\sqrt{n})$, then $Z_{(k(n),n)}$ converges in distribution towards z_q as it follows

$$\sqrt{n}(Z_{(k(n),n)} - z_q) \xrightarrow{\mathcal{L}} \mathcal{N}\left(0, \frac{q(1-q)}{f(z_q)^2}\right).$$

The exploratory analysis we propose for the perturbation data sets is based on the computation of empirical quantiles. There are several reasons motivating such a choice. First, there is not too much a priori knowledge concerning the perturbations marks, except that they are distributed around zero and that they are uniformly located in K. This implies that very few hypothesis with respect to the data can be done. Clearly, in order to apply such an analysis the only assumptions needed are the conditions of validity for the central limit theorem. From a practical point of view, an empirical quantiles based analysis allows for checking the tails, the symmetry and the general spatial pattern of the data distribution. From a theoretical point of view, the mathematics behind this tool allow a rather rigorous analysis.

2.3 Stable distributions models

Stable laws are a rich class of probability distributions that allow heavy tails, skewness and have many nice mathematical properties. They are also known in the literature under the name of α-stable, stable Paretian or Lévy stable distributions. These models were introduced by [18]. In the following some basic notions and results on stable distributions are given [1, 12, 27].

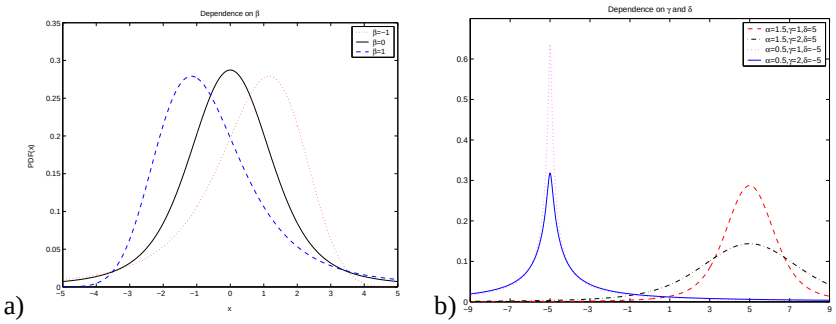
A random variable Z has a *stable distribution* if for any $A, B > 0$, there is a $C > 0$ and $D \in \mathbb{R}^1$ such that

$$AZ_1 + BZ_2 \stackrel{\mathcal{L}}{=} CZ + D,$$

where Z_1 and Z_2 are independent copies of Z, and "$\stackrel{\mathcal{L}}{=}$" denotes equality in distribution.

A stable distribution is characterised by four parameters $\alpha \in (0,2]$, $\beta \in [-1,1]$, $\gamma \geqslant 0$ and $\delta \in \mathbb{R}^1$ and it is denoted by $\mathcal{S}_\alpha(\beta, \gamma, \delta)$. The role of each parameter is as it follows : α determines the rate at which the distribution tail converges to zero, β controls the skewness of the distribution, whereas γ and δ are the scale and shift parameters, respectively. Figure 1 shows the influence of these parameters on the distribution shape.

The linear transformation of stable random variable is also a stable variable. If $\alpha \in (0,2)$, then $\mathbb{E}|Z|^p < \infty$ for any $0 < p < \alpha$ and $\mathbb{E}|Z|^p = \infty$ for any $p \geqslant \alpha$. The distribution is Gaussian if $\alpha = 2$. The stable variable with $\alpha < 2$ has an infinite variance and



a) b)

Figure 1: Influence of the parameters on the shape of a stable distribution : a) β parameter, b) α, γ and δ parameters.

the corresponding distribution tails are asymptotically equivalent to a Pareto law [28]. More precisely

$$\begin{cases} \lim_{z\to\infty} z^{\alpha}\mathbb{P}\{Z > z\} &= \frac{(1+\beta)}{2}\sigma, \\ \lim_{z\to\infty} z^{\alpha}\mathbb{P}\{Z < -z\} &= \frac{(1-\beta)}{2}\sigma. \end{cases} \tag{2}$$

where $\sigma = C_\alpha \gamma^\alpha$, $C_\alpha = \frac{1-\alpha}{2\Gamma(2-\alpha)\cos(\pi\alpha/2)}$ if $\alpha \neq 1$ and $C_\alpha = \frac{2}{\pi}$ elsewhere. The distribution is symmetric whenever $\beta = 0$, or skewed otherwise. In the case $\alpha < 1$, the support of the distribution $\mathcal{S}_\alpha(\beta, \gamma, 0)$ is the positive half-line when $\beta = 1$ and the negative half-line when $\beta = -1$. If $\alpha > 1$, then the first order moment exists and equals the shift parameter δ.

One of the technical difficulty in the study of stable distribution is that except for a few cases (Gaussian, Cauchy and Lévy), there is no explicit form for the densities. The characteristic function can be used instead, in order to describe the distribution. There exist numerical methods able to approximate the probability density and the cumulative distribution functions [21]. Simulation algorithms for sampling stable distribution can be found in [1, 3].

Due to the previous considerations, parameter estimation is still an open and challenging problem. Several methods are available in the literature [11, 19, 20, 22, 25]. Nevertheless, these methods have all the same drawback, in the sense that the data is supposed to be a sample of a stable law. It is a well known fact, that if the data comes from a different distribution, the inference of the tail index may be strongly misleading. A solution to this problem is to estimate the tail exponent [17] and then estimate distribution parameters if $\alpha \in (0, 2]$.

Still, it remains to solve the problem of parameter estimation whenever the tail exponent is greater than 2. Under these circumstances, distributions with regularly varying tails can be considered. A random variable has a distribution with regularly varying tails of index $\alpha \geq 0$ if there exist $p, q \geq 0, p + q = 1$ and a slowly varying function L, *i.e* $\lim_{z\to\infty} \frac{L(\lambda z)}{L(z)} = 1$ for any $\lambda > 0$, such that

$$\begin{cases} \lim_{z\to\infty} z^{\alpha}L(z)\mathbb{P}\{Z > z\} &= p, \\ \lim_{z\to\infty} z^{\alpha}L(z)\mathbb{P}\{Z < -z\} &= q. \end{cases} \tag{3}$$

It is important to notice that the conditions (2) can be obtained from (3) whenever $L(z) = 1/\sigma$ and $p = (1+\beta)/2$.

The parameter estimation algorithm proposed by [6, 7] is constructed under the assumption that the sample distribution has the asymptotic property (2). The algorithm gives three estimated values $\widehat{\alpha}, \widehat{\beta}, \widehat{\sigma}$. The $\widehat{\delta}$ can be computed easily whenever $\alpha > 1$, by approximating it using the empirical mean of the samples. This parameter estimation method can be used for stable distribution and in this case, $\widehat{\alpha}$ should indicate positive values lower than 2. In the same time, the strong point of the method is that it can be used for data not following stable distributions. In this case the data distribution is assumed to have regularly varying tails. The weak point of this algorithm is that in this case, it does not give indications concerning the body of the distribution. Nevertheless, in both cases, this method allows a rather complete characterisation of a wide panel of probability distributions. The code implementing the algorithm is available just by simple demand to the authors.

3 Results

3.1 Empirical quantiles

The lack of stationarity of the perturbations marks imposes the partitioning of the location space in a finite number cells. Let us consider such a partition $K = \cup_{i=1}^{n} K_i$. The cells K_i are disjoint and they all have the same volume. The size of volume has to be big enough in order to contain a sufficient number of perturbations. In the same time, the volume has to be small enough to allow stationarity assumptions for the perturbations marks inside a cell. After several trial and errors, we have opted for a partition made of square cells K_i, having all the same volume 0.1×0.1 AU, so that each cell contains about 1 500 perturbations.

We were interested in three questions concerning the perturbations marks distributions. The first two questions are related to the tails and the symmetry of the data distribution. The third question is related to a more delicate problem. It is a well known fact that the perturbations locations follow an uniform distribution in K. Nevertheless, much few is known about the spatial distribution of the perturbations marks, except that they are highly dependent on their corresponding locations. So, the third question to be formulated is the following : do the distributions of the perturbations marks exhibit any pattern depending on the perturbation location ?

For this purpose, empirical $q-$quantiles were computed in each cell. The most part of these values were indicating that the perturbations marks are distributed around the origin, while no particular spatial pattern is exhibited in the perturbation location space.

On the other hand, the situation is completely different for extremal $q-$values such as : $0.01, 0.05, 0.95, 0.99$. These quantiles were indicating rather important values around the semi-major axis of each planet. In order to check if these values may reveal heavy tail distributions, the difference based indicator $z_q - \widehat{n}_q$ was built. The first term of this indicator represents an empirical $q-$quantile. The second term is the theoretical $q-$quantile of the normal law with mean and standard deviation given by $z_{0.50}$ and $0.5(z_{0.84} - z_{0.16})$. Hence, for values of q approaching 1, positive values of the indicator

may suggest heavy-tail behaviour for the data. Clearly, this indicator may be used also for quantiles approaching 0. In this case, it is the negative sign that reveals the fatness of the distribution tail.

In Figure 2 the values obtained for the difference indicator $z_{0.99} - \widehat{n}_{0.99}$ are shown. It can be observed that its rather important values appeared whenever the perturbations are located in the vicinity of a planet orbit. All these values tend to form a spatial pattern similar to an arrow-like shape. As it can be observed, this shape is situated around the planet orbit and it is pointing from the right to left. It tends to vanish, while the cosine of the inclination angle approaches -1. The prominence of this arrow shape clearly depends on the closest planet : bigger the planet is, sharper is the arrow-like shape. This can be observed by looking at the change of values for the difference indicator with respect the size of the planet. These observations fulfil some good sense expectations : the comets perturbations tend to be more important whenever a comet cross the orbit of a giant planet.

Since these phenomena are observed for extremal $q-$quantiles, they indicate that the distribution tails may be an important feature for the data. Hence, a statistical model for the data should be able to catch these characteristics of the perturbations marks.

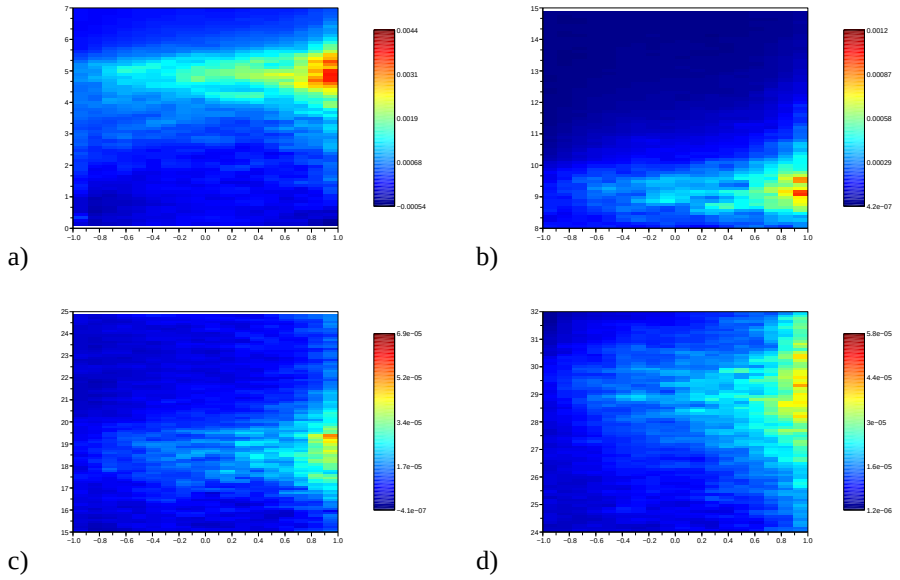
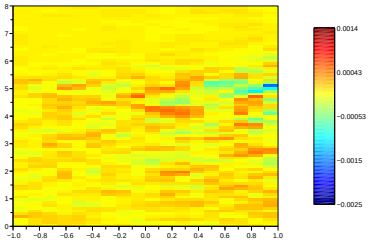


Figure 2: Empirical quantiles based difference indicator $z_{0.99} - \widehat{n}_{0.99}$ for the perturbation marks around the big planets : a) Jupiter, b) Saturn, c) Uranus d) Neptune. For each diagram the y-axis correspond to initial perihelion distance in AU, and the x-axis to cosine of the inclination. We recall that the respective semi-major axis of the four giant planets are: $a_J = 5.2$ AU, $a_S = 9.6$ AU, $a_U = 19.2$ AU, $a_N = 30.1$ AU.

Empirical quantiles can be also used in straightforward way as symmetry indicators of the data distribution. Clearly, by just checking whenever the difference $z_q - |z_{1-q}|$ tends to 0, this may suggest a rather symmetric data distribution. Figure 3 shows the computation of such differences for each data cell. The values obtained are rather small all over the studied region. Nevertheless, there are some regions and especially around the Jupiter's orbit we may suspect the data distributions are a little bit skewed. Still, since the perturbations have rather small numerical values, assessing symmetry using the proposed indicator has to be done cautiously.

It is reasonable to expect a more reliable answer concerning this question by using a statistical model. Clearly, such a model should be able to catch the symmetry of the data distribution as well.



a)

Figure 3: Exploring symmetry using empirical quantiles difference $z_{0.99} - |z_{0.01}|$ for the perturbations marks around Jupiter. Axis are as for Fig. 2

The central limit theorem available for the order statistics allows the construction of an hypothesis test. Since our analysis leads us towards heavy-tailed distributions models, as a precaution, a statistical test was performed to verify if a rather simpler model can be fitted to the data. The normality assumption was considered as null hypothesis for the test. The test was performed for the data in each cell, by considering that the normal distribution parameters are given by the empirical quantiles as expained previously. The $p-$values were computed using a χ^2 distribution. In this context, the local normality assumption for the perturbation marks is globally rejected. Figure 4 shows the result of testing the normality of the $z_{0.95}$ empirical quantile computed around the Jupiter's orbit.

Indeed, there exist regions where the normality assumptions cannot be rejected for the considered quantile. Still, the regions where this hypothesis is rejected clearly indicate that normality cannot be assumed entirely. Therefore, a parametric statistical model has to be able to reflect this situation : indicate whenever is the case how "heavy" or how stable are the distributions tails.

The only parameter used during this exploratory analysis was the partitioning of the location domain K. There is one more question to answer : do the obtained results depend on the patterns exhibited by the data, or they are just a consequence of the partitioning in cells of the data locations ? To answer this question, a bootstrap procedure and a permutation test were implemented [5].

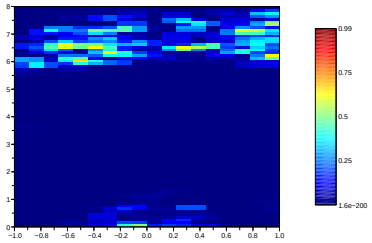


Figure 4: $p-$values computed for testing the normality of the empirical quantiles $z_{0.95}$ around Jupiter.

Bootstrap samples were randomly selected by uniformly choosing 20% from the entire perturbations data set. Difference indicators were computed for this special data set. This operation was repeated 100 times. At the end of the procedure, the empirical means of the difference indicators were computed. In Figure 5a the bootstrap mean of the indicator $z_{0.99} - \widehat{n}_{0.99}$ around Jupiter's orbit is showed. As expected, the same pattern is obtained as in Figure 2a : important values are grouped around the planet's orbit while exhibiting an arrow-like shape pointing from right to left.

The permutation test follows the same steps as the bootstrap procedure except that the perturbations are previously permuted. This means that all the perturbations are modified as it follows : for a given perturbation, its mark is kept while its location is exchanged with the location of another randomly chosen perturbation. This procedure should destroy any pre-existing structure in the data. In this case, we expect that applying a bootstrap procedure on this new data set will indicate no relevant patterns. In Figure 5b the result of such permutation test is showed. The experiment was carried out in the vicinity of Jupiter's orbit. After permuting the perturbations as indicated, the previously described bootstrap procedure was applied in order to estimate bootstrap means of the difference indicator $z_{0.99} - \widehat{n}_{0.99}$. The result confirmed our expectations, in the sense that no particular structure or pattern is observed. This clearly indicates, that the analysis results were due mainly to the original data structure and not to the partitioning of the perturbations location domain in cells.

In the same time, the permutation test is also a verification of the proposed exploratory methodology. This methodology depends on a precision parameter for characterising the hidden structure or pattern exhibited by the data. Still, whenever such a structure does not exist at all, the present method detects nothing.

3.2 Inference using heavy-tail distributions

The empirical observations of the perturbations marks distributions indicated fat tails and skewness behaviour. This leptokurtic character of the perturbation distributions was observed especially in the vicinity of the planets orbits. In response to this empirical evidence heavy-tail distribution modelling was chosen.

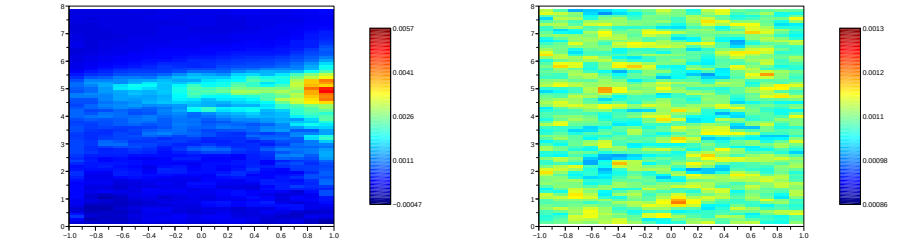


Figure 5: Validation of the analysis based on the computation of the difference indicator $z_{0.99} - \widehat{n}_{0.99}$ around Jupiter : a) bootstrap procedure ; b) permutation test.

The same cell partitioning as for the exploratory analysis is maintained. The previously mentioned algorithm for estimating stable laws parameters was run for the data in each cell.

In Figure 6 the estimation result of the tail exponent is shown. Clearly, it can be observed a region formed by the cells corresponding to estimated α values lower than 2. This kind of region may be located around each orbit corresponding to a big planet. The shape of this region is less picked than the region obtained using empirical quantiles. Still, the two results are coherent. Both results indicate that the heavy-tailed character of the perturbations distributions exhibits a spatial pattern. This spatial pattern is located around the orbits of the big planets.

The skewness of the data distribution can be analysed by looking at the results shown in Figure 7. Indeed, it can be observed that there are cells containing perturbations following a skewed distribution. The obtained results indicate neither the presence of a pattern by such distributions, nor the presence of such a pattern around the orbits of the big planets.

The estimation results for the σ and δ parameters are presented in Figure 8. The scale parameter indicates how heavy are the distribution tails. In Figure 8a, it may be observed that the most important values of σ tend to form a spatial pattern similar with the patterns formed by the difference indicator based on order statistics and the tail exponent, respectively. The results obtained for the δ parameter indicate that a shift of the perturbation may exist around the orbit of the corresponding big planets.

In order to check these results a statistical test using the central limit theorem for order statistics was built. Clearly, this result can be used in order to verify if the empirical quantiles from a cell are coming rather from the distribution characterised by the parameters previously estimated. Figure 9 shows the result of a test verifying that the $z_{0.99}$ quantiles around the Jupiter's orbit are originated from a heavy-tail distribution, while the quantiles outside this region are coming rather from Pareto distribution. It can be observed that high values for the $p-$values are spread around the entire region : for 81.5% of the cells we cannot reject the null hypothesis. Clearly, this result shows a far better characterisation of the distribution tails of the perturbations than the test for the normality assumption performed in the preceding section.

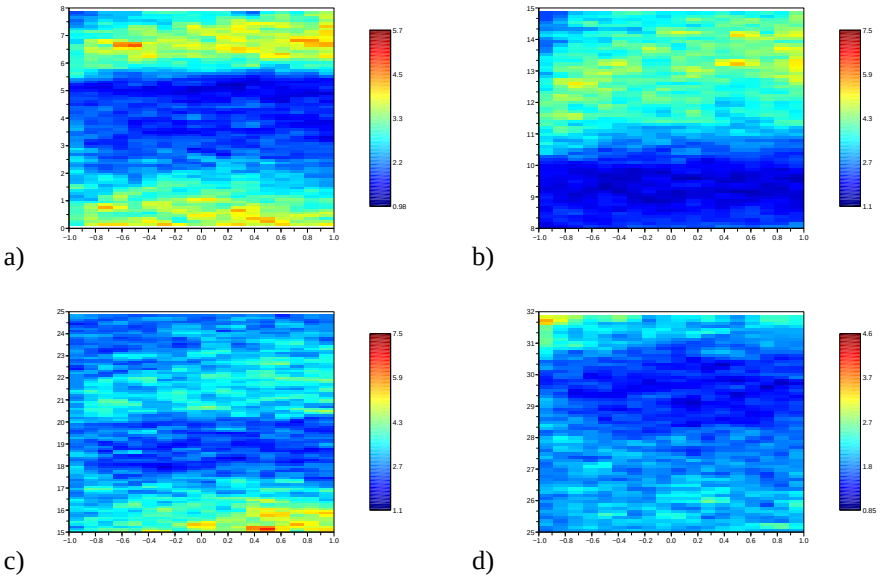


Figure 6: Estimation result of the tail exponent α for the perturbation marks around the big planets : a) Jupiter, b) Saturn, c) Uranus d) Neptune.

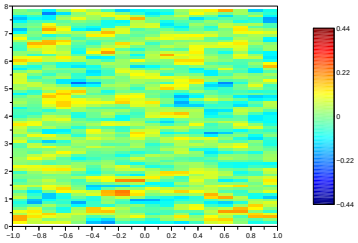


Figure 7: Estimation result of the skewness parameter β for the perturbations marks around Jupiter.

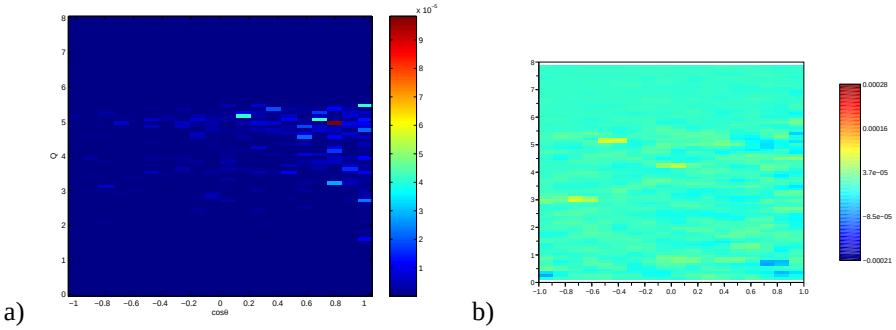


Figure 8: Estimation result of the scale parameter σ and shift parameter δ for the perturbation marks around Jupiter.

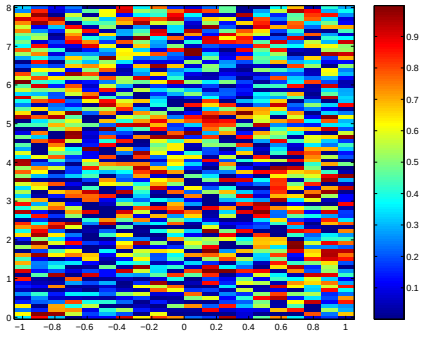


Figure 9: $p-$values computed for testing if the empirical quantiles $z_{0.99}$ around the Jupiter's orbit are originated from a heavy-tail distribution.

The previous test certifies the perturbations distributions tails exhibit a stable or regular variation behaviour. If the perturbations are close to the orbit of a big planet then they have rather a stable behaviour. Figure 10 shows the $p-$values of a χ^2-test implemented for the perturbations with estimated tail exponent $\alpha < 2$. This test allows to check the perturbations also for their distribution body. It can be observed that almost in all these regions the assumption of stable distributions is not rejected.

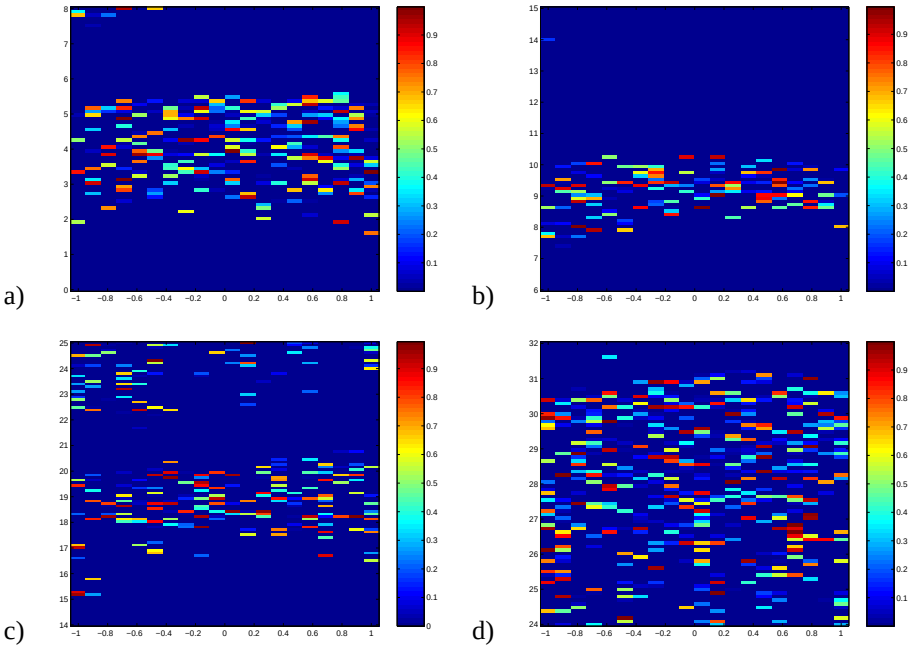


Figure 10: $p-$values of a χ^2 statistical test for the perturbations with $\alpha < 2$ around the big planets : a) Jupiter, b) Saturn, c) Uranus, d) Neptune.

For the perturbations with a tail exponent greater than 2, an alternative family of distributions with regularly varying tails was considered for modelling. Its expressions is given below :

$$f(z) = \frac{C_{\kappa,\alpha}}{1+ \mid \kappa z - \omega \mid^{\alpha+1}}, \tag{4}$$

with $C_{\kappa,\alpha}$ the normalising constant, κ the scale parameter, ω the location parameter and α the tail exponent.

The parameter estimation for such distributions was done in several steps. First, the tail exponent α was considered obtained from the previous algorithm. Second, the location parameter ω was estimated by the empirical mean of the data samples. Finally, the normalising constant $C_{\kappa,\alpha}$ and the scale parameter κ were estimated using the method of moments.

A χ^2 statistical test was done for the perturbations with $\alpha \geq 2$. The null hypothesis considered was that the considered perturbations follow a regularly varying tails distribution (4) with parameters given by the previously described procedure. The obtained $p-$values are shown in Figure 11. It can be noticed that in the majority of considered cells the null hypothesis is not rejected.

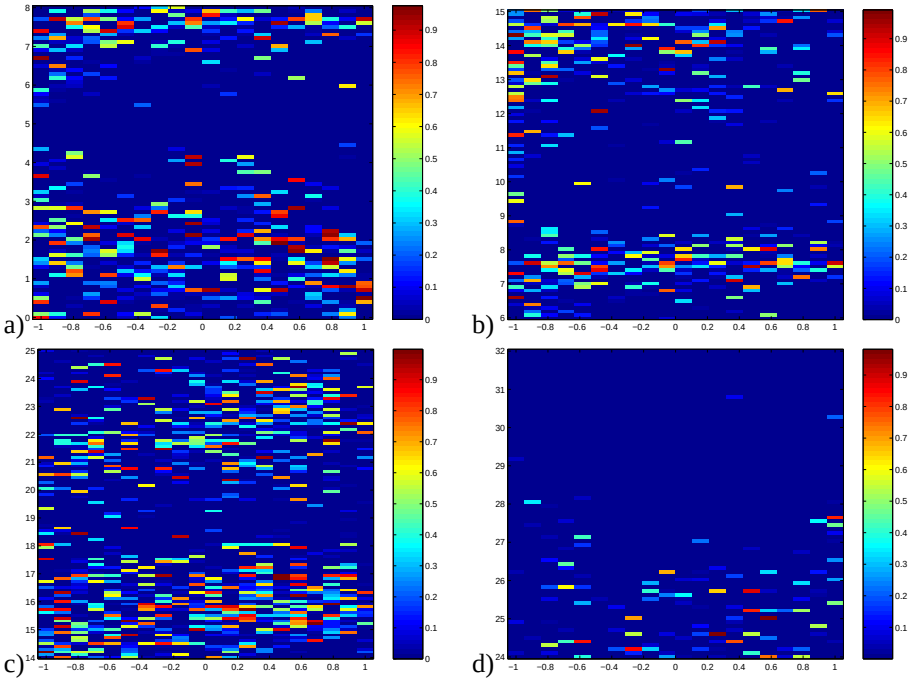


Figure 11: $p-$values of a χ^2 statistical test for the perturbations with $\alpha \geq 2$ around the big planets : a) Jupiter, b) Saturn, c) Uranus, d) Neptune.

4 Discussion and interpretation

Some of the features present in the Figures can be explained in the framework of the analytical theory of close encounters [24, 16, 2, 29].

Let us consider the magnitude of the perturbations in the vicinity of $a = a_J = 5.2$ AU (Jupiter). The colour coding of the Figure 2 represents the magnitude P of the perturbation, corresponding to

$$Z = -\frac{1}{a_f} + \frac{1}{a_i} \propto h_f - h_i \tag{5}$$

where a and h are respectively the orbital semi-major axis and the orbital energy of the heliocentric keplerian motion of the comet. The subscripts $_i$ and $_f$ stand, respectively,

for *initial* and *final*, *i.e.*, before and after the interaction with Jupiter).

Perturbations at planetary encounters are characterised by large and in general asymmetric tails, as was shown by various authors [9, 23, 15]; an analytical explanation of these features was given by [2] and by [30], and the consequences on the orbital evolution of comets was discussed by [29].

Let us consider the case of parabolic initial orbits (our orbits are in fact very close to parabolic). In the q-$\cos i$ plane, the condition for the tails of the energy perturbation distribution to be symmetric is:

$$0 = \frac{1 - 3 + 2\sqrt{2q/a_p}\cos i}{2\sqrt{3 - 2\sqrt{2q/a_p}\cos i}},$$

where a_p is the orbital semi major axis of the planet encountered.

However, the finite size of the available perturbation sample must be taken into account, as the tails would become sufficiently populated to show any asymmetry only for very large samples.

A way to take this effect into account is to consider that in different regions of the q-$\cos i$ plane the probability p for the comet on a parabolic orbit to pass within a given unperturbed distance b from the planet would be, according to [24]:

$$p = \frac{b^2}{a_p^2}\frac{\sqrt{3 - 2\sqrt{2q/a_p}\cos i}}{\pi \sin i \sqrt{2 - 2q/a_p}}.$$

To take into account the size of the perturbation, we consider that the angle γ by which the planetocentric velocity of the comet is rotated is given by:

$$\tan\frac{\gamma}{2} = \frac{a_p m_p}{b m_\odot \left(3 - 2\sqrt{2q/a_p}\cos i\right)};$$

we then define a function f as:

$$\begin{aligned} f &= p\tan\frac{\gamma}{2} \\ &= \frac{b m_p}{a_p m_\odot \pi \sin i \sqrt{2 - 2q/a_p}\sqrt{3 - 2\sqrt{2q/a_p}\cos i}}. \end{aligned} \tag{6}$$

Figure 12 shows the level curves of f; as can be seen, in it are reproduced the main features of Figure 2. The arrow-like shape observed during the statistical study can be now observed on the definition domain imposed by (6). This strenghten our interpretation of the features of Fig. 2 as due to the geometry of close approaches described by Öpik theory.

5 Conclusion and perspectives

In this paper a statistical study of the planetary perturbations on Oort cloud comets was carried out. The exploratory analysis of the perturbations distributions based on

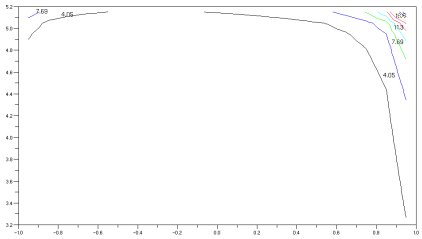


Figure 12: Level curves for the function f around the semi-major axis of Jupiter.

order statistics indicated the tail behaviour as determinant feature. Following this idea, parametric inference for heavy-tail distributions was implement. The obtained results indicated that the perturbations following heavy-tail stable distributions that are not always symmetric while tending to form a spatial pattern. This pattern is rather arrow-like shaped and is situated around the orbits of the big planets. A theoretical study was carried out, and it was observed that this pattern is similar with the theoretical curves derived from the Öpik theory. The perturbations outside this arrow shaped region were not exhibiting a stable character and they were modelled by a family of distributions with regularly varying tails. In both cases, stable and non-stable distributions, the modelling choices were confirmed by a statistical test.

Clearly, these choices and the estimation parameter estimation procedures can be further improved. Nevertheless, the obtained results give good indications and also good reasons for developing a probabilistic methodology able to simulate such planetary perturbations.

Acknowledgements

The authors are grateful to Alain Noullez for very useful comments and remarks. Part of this research was supported by the University of Lille 1 through the financial BQR program.

References

[1] S. Borak, W. Härdle, and R. Weron. Stable distributions. In P. Cizek, W. Härdle, and R. Weron, editors, *Statistical tools for finance and insurance*, pages 21–44. Springer, 2005.

[2] A. Carusi, G. B. Valsechi, and R. Greenberg. Planetary close encounters - Geometry of approach and post-encounter orbital parameters. *Celestial Mechanics and Dynamical Astronomy*, 49:111–131, 1990.

[3] J. Chambers, C. Mallows, and B. Stuck. A method for simulating stable random variables. *Journal of the American Statistical Association*, 70:340–344, 1976.

[4] H. A. David. *Order statistics*. John Wiley and Sons, 1981.

[5] A. C. Davison and D. V. Hinkley. *Bootstrap methods and their application*. Cambrdige University Press, 1997.

[6] Y. Davydov and V. Paulauskas. On the estimation of the parameters of multivariate stable distributions. *Acta Applicandae Mathematicae*, 58:107–124, 1999.

[7] Y. Davydov and V. Paulauskas. Recent results on p-stable convex compact sets with applications. *Fields Institute Communications*, 44:127–139, 2004.

[8] J.-F. Delmas and B. Jourdain. *Modèles aléatoires. Application aux science de l'ingénieur et du vivant*, volume 57 of *Mathématiques et Applications (SMAI)*. Springer, 2006.

[9] E. Everhart. Close Encounters of Comets and Planets. *Astronomical Journal*, 74:735–+, June 1969.

[10] E. Everhart. An efficient integrator that uses Gauss-Radau spacings. In A. Carusi and G. B. Valsecchi, editors, *ASSL Vol. 115: IAU Colloq. 83: Dynamics of Comets: Their Origin and Evolution*, page 185, 1985.

[11] E. Fama and R. Roll. Parameter estimates for symmetric stable distributions. *Journal of the American Statistical Association*, 66:331–338, 1971.

[12] W. Feller. *An introduction to probability theory and its applications - Vol. II*. Wiley, 1971.

[13] J.A. Fernández. New and evolved comets in the solar system. *å*, 96:26–35, 1981.

[14] M. Fouchard, Ch. Froeschlé, and G. B. Valsecchi. Is the dynamics of Jupiter family comets amenable to Monte Carlo modelling? *Monthly Notices of the Royal Astronomical Society*, 344:1283–1295, 2003.

[15] Cl. Froeschlé and H. Rickman. A monte carlo investigation of jovian perturbations on short-period comet orbits. *Icarus*, 46:400–414, 1981.

[16] R. Greenberg, A. Carusi, and G. B. Valsecchi. Outcomes of planetary close encounters - A systematic comparison of methodologies. *Icarus*, 75:1–29, July 1988.

[17] B. M. Hill. A simple general approach to inference about the tail of a distribution. *Annals of Statistics*, 3:1163–1174, 1975.

[18] P Levy. *Calcul des probabilités*. Gauthier Villars, 1925.

[19] J. H. McCulloch. Simple consistent estimators of stable distribution parameters. *Communications in Statistics - Simulation and Computation*, 15:1109–1136, 1986.

[20] S. Mittnik, T. Doganoglu, and D. Chenyao. Computing the probability density function of the stable paretian distribution. *Mathematical and Computer Modelling*, 29:235–240, 1999.

[21] J. P. Nolan. Numerical calculation of stable densities and distribution functions. *Communications in Statistics - Stochastic Models*, 13:759–774, 1997.

[22] J. P. Nolan. Maximum likelihood estimation and diagnostics for stable distributions. In *Lévy Processes*, 2001.

[23] S. Oikawa and E. Everhart. Past and future orbit of 1977 UB, object Chiron. *Astronomical Journal*, 84:134–139, January 1979.

[24] E. J. Opik. *Interplanetary encounters : close-range gravitational interactions*. 1976.

[25] S. J. Press. Estimation in univariate and multivariate stable distributions. *Journal of the American Statistical Association*, 67:842–846, 1972.

[26] H. Rickman, G. B. Valsecchi, and C. Froeschlé. From the Oort cloud to observable short-period comets - I. The initial stage of cometary capture. *Monthly Notices of the Royal Astronomical Society*, 325:1303–1311, August 2001.

[27] G. Samorodnitsky and M. S. Taqqu. *Stable non-Gaussian random processes : stochastic models with infinite variance*. Chapman & Hall/CRC, 1994.

[28] A.V. Skorokhod. Asymptotic formulas for stable distribution laws. In *Selected Translations in Mathematical Statistics and Probability 1*, pages 157–161, 1961.

[29] G. B. Valsecchi and A Manara. Dynamics of comets in the outer planetary region. II. Enhanced planetary masses and orbital evolutionary paths. *Astronomy and Astrophysics*, 323:986–998, July 1997.

[30] G. B. Valsecchi, A. Milani, G. F. Gronchi, and S. R. Chesley. The Distribution of Energy Perturbations at Planetary Close Encounters. *Celestial Mechanics and Dynamical Astronomy*, 78:83–91, 2000.