Asteroids, Comets, Meteors 1991, pp. 61–64
Lunar and Planetary Institute, Houston, 1992

140867 P. 4

THE USE OF THE WAVELET CLUSTER ANALYSIS FOR ASTEROID FAMILY DETERMINATION

N93-19126

Bendjoya Ph., Slezak E., Froeschlé Cl.
O.C.A BP 139 06003 Nice cedex France

ABSTRACT

The asteroid family determination has been for a longtime analysis method dependent. A new cluster analysis based on the wavelet transform has allowed an automatic definition of families with a degree of significance versus randomness. Actually this method is rather general and can be applied to any kind of structural analysis. We will rather concentrate on the main features of the method. The analysis has been performed on the set of 4100 asteroid proper elements computed by Milani and Knézévic (see Milani and Knézévic 1990). Twenty one families have been found and influence of the chosen metric has been tested. The results have been compared to Zappala et al.'s ones (see Zappala et al 1990) obtained by the use of a completely different method applied to the same set of data. For the first time, a good overlapping has been found between the both method results, not only for the big well known families but also for the smallest ones.

INTRODUCTION

The purpose of this paper is to introduce a new method of cluster analysis based on the wavelet transform in order to determine asteroid families. The mathematical basis of the wavelet transform will be briefly presented and we will rather insist on the way it has been used. Then it will be shown how this method has been performed on a set of proper elements computed by Milani and Knezevic (version 4.2) (see Milani and Knezevic 1990) and how the families have been defined with a degree of significance versus randomness. Finally the results will be compared to those found by a complete different method used by Zappala et al. (see Zappala et al. 1990).

THE WAVELET TRANSFORM

Among many fields of applications the wavelet transform appears to be well suited for signal processing purpose since it enables one to get informations on a signal both in frequencies and space. The wavelet transform of a signal can be seen as a decomposition of this signal onto a base of functions defined from dilatations and translations of a unique function called the analyzing wavelet.

Let the one-dimensional signal $f(x)$, and the analyzing wavelet $\psi(x)$, belong to $L^2(\Re)$. The wavelet transform of $f(x)$ depends on the scale σ and the location b according to the expression:

$$C(\sigma,b)=\sigma^{-\frac{1}{2}}\int_{-\infty}^{+\infty} f(x)\psi^*(\frac{x-b}{\sigma})dx \tag{1}$$

where $\sigma^{-\frac{1}{2}}$ is a $L^2(\Re)$ normalization factor and * denotes the complex conjugate.

The wavelet transform can be seen as a mathematical zoom which is able to extract the σ-sized features from the signal around the location b. In fact when σ decreases thinner and thinner details can be extracted and finaly $f(x)$ can be seen as the sum of all its details at different scales.

In order to be an analyzing wavelet the function $\psi(x)$ must satisfy certain conditions. The admissibility condition (see Meyer 1990) can be written for a regular function as:

$$\int_{-\infty}^{+\infty} \psi(x)dx = 0 \tag{2}$$

Moreover $\psi(x)$ must nearly compact supported for numerical reasons and smooth both in the real and Fourier spaces in order to give a good localization of an event in both spaces.

These three conditions present the " mexican hat function" as a good candidate for being an analyzing wavelet (see fig.1). This function comes from the second derivative function of a Gaussian function and is defined, for isotropic analysis by:

$$\psi_\sigma(r) = (n - \frac{r^2}{\sigma^2})\exp(-\frac{r^2}{2\sigma^2}) \tag{3}$$

where r is the distance defined in the space in which the signal is defined and n its dimension.

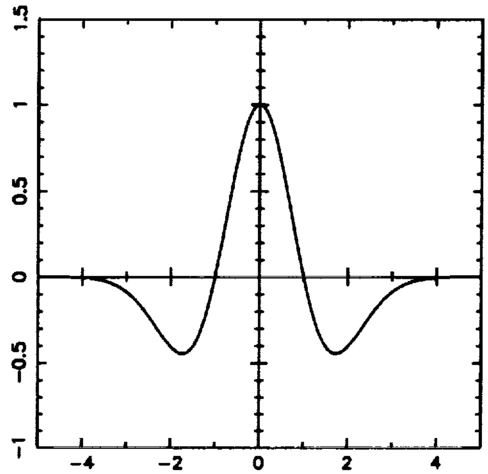

Fig 1. The $1D$ mexican hat function

THE WAVELET CLUSTER ANALYSIS METHOD

In the case of asteroid family determination, the signal under study is a discrete one:it is a set of point in the proper element space. As this signal can be modelled by a set of Dirac functions the integral in the wavelet transform definition (eq. 1) is replaced by a discrete sum. Because of the use of a computer the wavelet transform of the signal is computed for each node of a network and we will rather speak of a wavelet coefficient at a considered point $\vec{b}_i$ of the network. This coefficient is then discrete sum of the weights, w_j, of the N data points.

$$C(\sigma, \vec{b}_i) = \sum_{j=1}^{N} w_j \tag{4}$$

where:

$$w_j = (n - \frac{r_{ij}^2}{\sigma^2})\exp(-\frac{r_{ij}^2}{2\sigma^2}) \tag{5}$$

where r_{ij} is the distance between the node $\vec{b}_i$ and the j^{th} data point.

It is obvious that for a node located in uniformely populated area the size of which is large compared to the studied scale, and because of eq.2, the wavelet coefficient will be near 0 since the positive weights will be compensated by the negative ones. On the other hand, for a node centered on an area which is a cluster of points the size of which is about the studied scale, the wavelet coefficient will be positive and negative if this area is a hole of the same size. The discret wavelet transform of a set of points yield to a map of coefficients, for a given scale, where high positive values point out structures of about the same size. The advantage of the wavelet transform over other cluster analysis method is to enable one to get informations both on the location and the size of the structures within the signal.

The major problem however is to distinguish the physical structures from structures due to chance. A threshold has to be introduced in the wavelet coefficients in order to quantify the chance. The idea is to build a pseudo random distribution in the sense that: being given a certain number of boxes in the original distribution of points, the pseudo random distribution is built in putting randomly the same number of points in each equivalent box in order to keep the same local density of points. The wavelet map of coefficients is then computed on each distribution of points. An histogram of the wavelet coefficients coming from a pseudo random distribution is made. The value of the coefficient for which 99.5% of the other coefficients have a lower value than this one, defines the threshold. There are indeed 5 chances over 1000 to find a higher valued coefficient. By reporting the threshold, averaged over many pseudo random distribution realisations, in the wavelet coefficient map of the original distribution and by keeping only the coefficients higher than this threshold, it is obvious that the structures associated to these wavelet coefficients have less than 5 over 1000 risks to be due to chance. It is then possible to extract the structures at the studied scale by superimposing the grid with the conserved coefficients over the original set of data.

The next step is to choose a metric well fitted to the studied problem in order to compute the wavelet coefficients. Because of the break-up origin of the families, a metric with the dimension of a velocity has been chosen. From the Gauss equations and by substituting the proper elements to the osculating ones and after having averaged these equations over the fast variables,we obtain a relative velocity δv between the points of the set of proper elements:

$$\delta v = na'\sqrt{k_1(\frac{\delta a'}{a'})^2 + k_2(\delta e')^2 + k_3(\delta i')^2} \tag{6}$$

n is the mean motion, a', e', i' are respectively the proper semi-major axis, the proper eccentricity and the proper inclination.
k_1, k_2, k_3 is a set of arbitrary coefficients that has been chosen equal to $\{\frac{5}{4}, 2, 2\}$.

The wavelet analysis is so performed on the same set of data for several scales with the fitted networks and structures are pointed out for each scale with the same degree of significance versus chance. The ultimate problem is to cut the hierarchy in order to define the families. A criterion based on the philosophy of the wavelet transform has been chosen. This criterion is based on the additional information got from one scale to the successive one. The largest scale to consider in order to define families is the one for which the number of added asteroids is either 0 or for which the next scale brings obviously too much new asteroids. The first case indicates that the families are well isolated, the second one shows that structures are merging and bridges are made between structures.

THE RESULTS

Twenty one families have been defined from this analysis. All the well-known big families have been found again (Eunomia,Koronis, Eos and Themis). For the other smaller and much more debated families, the results have been compared with those obtained by Zappala et al. (see Zappala et al. 1990) with a complete different method and from the same set of data. It has been pleasant to constat that the overlapping of the results of both methods was good even for small families. Table I. presents the names of the twenty one families, the number of members of each families, the percentage of asteroids found in the same family by both methods and the number of asteroid found by the wavelet transform clustering method and not itemized by the hierarchical single linkage clustering method. In 'Flora' region the comparison is less good than for the other zones (defined by the mean motion resonances), but this can be explained by two reasons at least. First it seems that Zappala et al. have been a little severe in the way of cutting their hierarchy in this region and this explained by the very dense background of this zone. Secondly the poor quality code of the proper elements given by Milani and Knezevic (see Milani and Knezevic 1990) in this region has not been taken into account.
On the other hand the percentage of same results for Themis can grows up to 87% by cutting the hierarchy of Zappala et al. at a lower level.

Nevertheless the good agreement of the whole between the two different methods permits to be more confident in the existence of the families.

zones	names of families	number of members	scale (m/s)	comparison idem (%)	number of added asteroids
zone 2	Lucretia	6	110		
	Berolina	114	450		
	Auravictrix	26	230		
	Iduberga	5	110		
zone 3	Vesta	5	110	71.4	0
	Amalasuntha	6	80	75	0
	Leonce	5	80	100	0
	Polona	20	230	0	20
	1969UN	19	320	63	7
	Tinchen	7	110	50	1
zone 4	Eunomia	114	450	95.6	31
	Adeona	17	160	100	2
	Leto	37	450	100	15
	Lydia	22	320	100	16
	Maria	34	320	94	4
	Dora	18	320	100	2
	Agnia	12	320	86	6
zone 5	Koronis	125	320	91	1
zone 6	Eos	172	230	87	0
zone 7	Themis	117	160	51	0
	Veritas	6	110	86	0

CONCLUSION

The wavelet transform is a powerfull tool in a lot of domains. It has succesfully been applied in a cluster analysis method in order to point out dynamical families in the asteroid proper element space. This first incursion in the celestial mechanic field has permit to get a certain coherence about the family definition by the comparison with results obtained by a complete different method.

References :

Bendjoya Ph., Slezak E., Froeschlé (1990) The wavelet transform a new tool for asteroid family determination.
Astronomy and Astrophysic J. (in press)

Meyer,Y.: 1989, Orthonormal wavelets in Wavelets, time-frequency methods and phase space,
(J.M. Combes, A. Grossmann, Ph. Tchamitchian Eds.), Springer-Verlag

Milani,A. and Knežević,Z. : 1990, Celestial Mechanic 4, 347-412

Zappalà,V.,Cellino,A.,Farinella,P.,Knežević,Z.: 1990, Astron. J. 100, 2030