Mem. S.A.It. Vol. 78, 612
© SAIt 2007

Memorie della

Extrasolar planet taxonomy: a new statistical approach

S. Marchi and S. Ortolani

Dipartimento di Astronomia, Università di Padova, Italy
e-mail: `simone.marchi@unipd.it; sergio.ortolani@unipd.it`

Abstract. We present a new statistical analysis done via principal component analysis followed by hierarchical clustering analysis. We report the definition of five robust groups of planets. For instance, we were able to divide the hot Jupiters into two main groups, on the basis of their star masses and metallicities. Moreover, for some groups, we find strong correlations between metallicity and semi-major axis and eccentricity.

Key words. Stars

1. Introduction

The inputs to our model are the elements provided by the interactive extrasolar planets catalog mantained by J. Schneider[1]. These are: planetary projected mass (M_p), semi-major axis (a), eccentricity (e), stellar mass (M_s), stellar metallicity ([Fe/H]). Only objects having simultaneously estimates for $\{M_p, a, e, M_s,$[Fe/H]$\}$ have been used.

We consider 183 EPs (updated at 8 November 2006) and Jupiter. The first step is to perform a statistical analysis in order to find out if there are useless -or less significant- input variables. This is done using principal component analysis (PCA; for datails see Everitt & Dunn 2001). The basic idea of PCA is to combine the input variables in such a way as to show the most important ones. This is done by describing the data with a number of new variables (pc_i), ordered in terms of decreasing variance. On the basis of the variance attained by each pc_i we may reject some of them. According to general criteria, it seems reasonable to keep only the first three principal components which account for 73% of the total variance.

2. Cluster analysis

The choice of the clustering technique relies mostly on the kind of description of the data we are interested in. When the number of clusters are not known *a priori*, like in our case, hierarchical clustering is more suitable (Everitt et al. 2001). One of the advantages of this technique is that it provides a classification which consists of a series of nested partitions, which is well illustrated by a two-dimensional diagram known as *dendrogram*. First we decided to not standardize the clustering variables as this may reduce the difference among members, making the identification of clusters more difficult. Moreover we tested several metrics and inter-group algorithms. The full set of possibilities has been investigated by

---

Send offprint requests to: S. Marchi

[1] http://exoplanet.eu/

using the traditional means (e.g. cophenetic coefficient) and analyzing the corresponding dendrograms. We therefore identify a robust solution which is closely nested for small heights and stable against errors (e.g. observational errors) on the position of the EPs in the clustering space. The best solution was obtained with the Pearson correlation distance and weighted centroid merging. It has five robust clusters. We tested the solution against the absence of clusters with Monte Carlo simulations. We also tested the solution with respect to the presence of observational errors. We find that the solution is stable with respect to both tests. The process developed for EP taxonomy is sketched in fig. 1.

3. Analysing the clusters

Our best solution is composed of five robust clusters. In this section we present briefly the properties of each cluster. For more details see (Marchi 2007).

Cluster *C1*: Containing 11 EPs (see Table 1 for a detailed description) this is the least populated cluster. Significant intracluster correlations exist between $M_p - e$ and $M_p - M_s$. Notice that M_p is anticorrelated with M_s. This is somehow unexpected as for higher M_s we would expect higher dust surface density of the protoplanetary disks (Ida & Lin 2005) and hence more massive planets (consider also that here we have sub-solar stellar masses; on this point see also *C4*).

Cluster *C2*: This cluster contains 46 EPs (see Table 1). To this cluster belongs Gl 581 b, Gliese 876 b-c-d and GJ 436 b which orbit low mass stars (respectively with 0.31, 0.32 and 0.41 $M_\odot$ they are the lowest M_s in the sample) also with low metallicity (respectively -0.33, -0.12, -0.32 dex). *C2* contains 17 hot Jupiters (that is 37% of its members), and 4 EPs belong to MSSs. It contains also 5 transiting EPs (the total number of transiting EPs is 14 -at December 2006- but only 9 are involved in the present analysis; see Burrows et al. 2006). Finally, *C2* contains also 13 planets in multiple planet systems (MPS). The significant intra-cluster correlation are $a - e$ and $a-$[Fe/H]. The first one is very interesting because a is anti-correlated with e. Thus planets further away from the stars have higher eccentricities. In other words, either the excitation of e is more effective further away from the central star (assuming that planets form in circular orbits) and/or e dumping is more effective for lower a. This result is consistent with tidal circularization of close-in planets. Moreover, with a notably steep trend, a anticorrelates with [Fe/H]. Thus either the planetary migration is more pronounced for high [Fe/H] (which is in agreement to what the simulations predict, e.g. see Livio & Pringle 2003) or giant planets of this cluster may form close to stars ($\leq$ 1 AU) in high metallicity environment.

Cluster *C3*: Containing 48 EPs (see Table 1) this cluster, along with *C4*, is the most populated one. This cluster contains Jupiter. The significant intra-cluster correlations are: $a-$[Fe/H], $e-$[Fe/H], and M_s-[Fe/H]. First of all, a is anti-correlated with [Fe/H]. Thus either the planetary migration is more pronounced for high [Fe/H] (see also *C2*) or planets with lower [Fe/H] form at larger distances. A striking result is that Jupiter fits very well in this cluster: its large a would be the result of its formation in a solar-like metallicity disk. Moreover, e is anti-correlated with [Fe/H]. In order to understand the meaning of these correlations, two further points have to be considered. First, the existence of $a-$[Fe/H] and $e-$[Fe/H] correlation does not imply that a correlation between a and e exists. Indeed they are not correlated to the level of confidence adopted here. Moreover, the members of this cluster have remarkable super-solar metallicities. Here we suggest the idea that the metallicity acts in some way in determining e. For instance, since high [Fe/H] produces a faster migration, the low values of e observed for high [Fe/H] may be the result of the migration process, e.g. tidal circularization (see also Halbwachs et al. 2005). Alternatively the condition for the pumping up of e during planet-disk interactions (Sari & Goldreich 2004; Matsumura & Pudritz 2006) is not achieved in high [Fe/H] environments: indeed higher [Fe/H] imply higher disk viscos-

Table 1. Relevant data for extrasolar planet clusters. Corr: All significant intracluster correlations, i.e. having 2-tailed probability less then 5%. Those of *C2*, *C3* and *C4* are the most important ones. HT: Hot Jupiters. T: Transiting planets. MSS: Planets in multiple star systems (data on multiple star systems have been taken from Desidera & Barbieri 2006). MPS: Multiple planetary systems.

Cluster #	Prototype	Members	Corr	HJ	T	MSS	MPS
C1	HD 41004 A b	11	$M_p - e$, $M_p - M_s$	2	-	6	-
C2	HD 69830 c	46	$a - e$, $a-$[Fe/H]	17	5	4	13
C3	HD 11964 b	48	$a-$[Fe/H], $e-$[Fe/H], M_s-[Fe/H]	23	4	11	8
C4	HD 142022 A b	48	$M_p - e$, $M_p - M_s$	-	-	12	12
C5	HD 117207 b	31	-	1	-	7	13

ity and hence lower probability of e excitation (see equ. 8 and 9 of Sari & Goldreich 2004). Finally, M_s is anti-correlated with [Fe/H]. It is not easy to understand this correlation, and in particular if it has something to do with the planetary formation process. If we assume that higher M_s implies higher masses of proto-planetary disks from which EPs formed, this implies that to form planets in lower metallicity environments a more massive disk is required in agreements with the core accretion theory. Eleven EPs are in MSSs. It contains also 23 hot Jupiters (48% of the members) and 8 MPS planets. They are well spread and seem not to affect the $a-$[Fe/H], $e-$[Fe/H] correlations. Notice also the MPS planets of this cluster have $e < 0.2$ and [Fe/H]> 0.2 and for this reason they differ from those of *C2* (which have $e < 0.3$, [Fe/H]< 0.2) and those of from those of *C4* (which have $e > 0.3$).

Cluster *C4*: This cluster contains 48 EPs (see Table 1). Two significant correlations exist for this cluster: $M_p - e$ and $M_p - M_s$. The first implies that lower massive EPs have higher e, thus the mechanisms for the pumping-up of the eccentricity are more active in low mass planets, at least for the high semi-major axes and moderate positive metallicities of this cluster. Moreover, EPs masses are correlated with stellar masses. This may confirm the fact that higher M_s implies larger proto-planetary disk surface density and hence larger M_p (Ida & Lin 2005). *C4* contains 12 EPs in MSSs and 12 in MPS (see Table 1). Notice that the MSS planets may be responsible of the M_p-e correlation as many of them have low M_p and high e.

Cluster *C5*: This cluster contains 31 EPs (see Table 1). The formally significant correlation are: $M_p - M_s$, M_p-[Fe/H] and $a - e$. However they are all due to outlier planets and thus cannot be considered as real (notice that if we do not consider the outliers, the correlation $a-e$ is very close to the 5% significancy level) It contains 7 EPs in MSSs and only one hot Jupiter (namely, HD 118203 b). It is interesting to understand why this hot Jupiter is in this cluster. The peculiarity of HD 118203 b is that it has the highest eccentricity (0.309) among hot Jupiters. Notice that also HD 185269 b (in *C3*) has a very high eccentricity of 0.3. All the other input variables are the same, except for the planetary mass. HD 118203 b with a mass of 2.13 M_p is one of the most massive hot Jupiters. This explain why HD 118203 b has been put in this cluster. This cluster contains 13 EPs in MPS.

4. Discussion and conclusion

In this paper we develop a multivariate statistical analysis (PCA) to find the most important variables, and then hierarchical clustering analysis. The best result is achieved with non traditional metric and merging algorithms, namely the Pearson correlation metric and weighted centroid cluster merging. We reject the absence of clustering structure with Monte Carlo simulations, and also tested the stability of the solution against observational errors of the input variables.

Our best solution consists of five clusters. We show the importance of including the en-

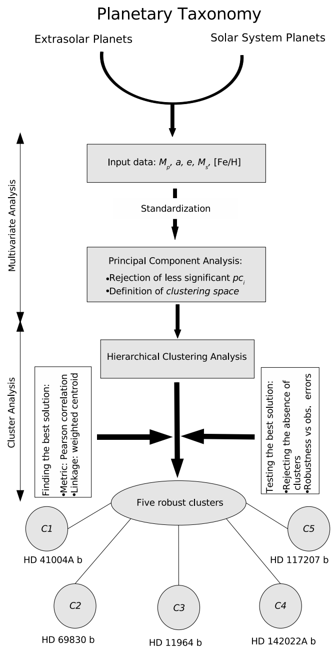

Fig. 1. Guidelines for the planetary taxonomy developed in this paper. As for the Solar System, only Jupiter has been included in this analysis so far (see text for further detail). Below each cluster the corresponding prototype is reported.

vironmental variables (M_s and [Fe/H]) to discriminate between otherwise similar planets; and also to merge together different bodies (like EP in MSSs and orbiting single stars). For instance, we were able to divide the hot Jupiters into -at least- 2 main groups (see Table 1). This division is mainly due to the stellar mass and metallicity. Those belonging to *C3* basically orbit around stars with super-solar masses and high metallicities; those of *C2* orbit mostly sub-solar mass stars with moderate (both positive and negative) metallicities. This may reflect differences of the formation processes of these EPs.

Jupiter belongs to the cluster *C3*. Much has been speculated about the similarity of our Solar System and extrasolar systems, in particular concerning the formation histories. With the help of cluster analysis we may identify those EPs which are more similar -in the 3-fold clustering space- to Jupiter. We suggest that the actual large semi-major axis of Jupiter is the result of its formation in a solar-like metallicity disk.

We also analyzed the intra-cluster correlations. Remarkably, for *C2* and *C3* we find important trends between metallicity and orbital parameters (see Table 1). It results that [Fe/H] has very important effects on the semi-major axis (and thus on the migration processes) and the eccentricity. It may also happen that the same variables correlate in an opposite way between two different clusters (see the $M_p - M_s$ correlations for *C1* and *C4*). Moreover, we also studied the distribution of planets in multiple star systems in each cluster. They do not seem to play a particular role in the corresponding cluster correlations. Similar considerations apply also for multiple planet systems.

References

Burrows, A., Hubeny, I., Budaj, J., & Hubbard, W. B. 2006, ArXiv Astrophysics e-prints, (`arXiv:astro-ph/0612703`)

Desidera, S., & Barbieri, M. 2006, ArXiv Astrophysics e-prints, (`arXiv:astro-ph/0610623`)

Everitt, B.S., & Dunn, G. 2001, Applied multivariate data analysis, second edition, Arnold Ed., London

Everitt, B.S., Landau, S., & Leese, Morven. 2001, Cluster analysis, fourth edition, Arnold Ed., London

Goldreich, P., & Sari, R. 2003, ApJ, 585, 1024

Halbwachs, J. L., Mayor, M., & Udry, S. 2005, A&A, 431, 1129

Ida, S., & Lin, D. N. C. 2005, ApJ, 626, 1045

Livio, M., & Pringle, J. E. 2003, MNRAS, 346, L42

Marchi, S. 2007, ArXiv e-prints, 705, (`arXiv:0705.0910`)

Matsumura, S., & Pudritz, R. E. 2006, MNRAS, 365, 572

Sari, R., & Goldreich, P. 2004, ApJ, 606, L77