STUDY OF THE SELF-ORGANIZING MAPS OF AN ASTRONOMICAL CATALOGUE USING THE STELLAR VELOCITY MOMENTS

M. Hernandez-Pajares, R.Cubarsi [1] *E. Monte* [2]

Abstract: We present the application of the learning vectorial quantification neural method (LVQ) to the study of stellar catalogue of the solar neighbourhood containing more than 12,000 stars. One result is the appearance of three groups of stars. Their kinematic properties are studied using the stellar velocity distribution moments up to order four.

Key words: *Neural networks, astronomical maps*

Received: June 23, 1992
Revised and accepted: August 18, 1992

1. Introduction

Observational astronomy presents some particular difficulties compared with other experimental sciences. First, the astronomers are observers, not experimenters. A second difference is that everything is variable in time such as positions, proper motions, radial velocities and colors. Finally, we can add that an additional complication appears when one speaks of bias. The intrinsic bias being that all astronomical observations are bound to one observational base - the earth.

Because of those particular features, observational astronomy has to look carefully at new trends and tools in the field of statistics and information theory. We could summarize the present aim of many astronomers with the phrase *New tools for old problems* [16].

[1]M. Hernandez-Pajares, R.Cubarsi
Department de Matemàtica Aplicada i Telemàtica, ETSETB, Universitat Politècnica de Catalunya, Apartat 30002, Barcelona, Spain
[2]E. Monte
Department de Teoria del Senyal i Comunicacions, ETSETB, Universitat Politècnica de Catalunya, Apartat 30002, Barcelona, Spain

One relevant aspect of studies that are carried out from stellar catalogues is the segregation of stars in populations in terms of spectral, photometric or kinematic criteria. For instance [19], [29], [32], [5] have also worked on this subject recently. The viewpoints adopted by most of these have been statistical, numerical or dynamical approaches.

This problem is important in itself, and for its consequences in particular, when we study the kinematics of the stellar fluid in the solar neighbourhood. Modern formulations and calculations are [4], [28], [22] in the first order, [11] in the second order, and [14], [15] in order n using orthogonal functions. Also [20], [5], [30], [17] have modeled and studied the moments of the residual stellar velocity field.

The LVQ is one recent statistical tool that can be used to study the possible stellar groups contained in a given stellar catalogue in an efficient way. It is also called the Kohonen maps or self-organizing feature maps algorithm. The basic aim of "self-organizing maps" is finding a smaller set W of "prototypes" w_r, so that the set $W = \{w_{r1}, w_{r2}, \ldots, w_{rn}\}$ of prototypes provides a good approximation of the original set V of patterns, encoded as "vectors" $v \in V$. Intuitively, this should mean that for each $v \in V$ the distance $\| v - w_{s(v)} \|$ between v and the closest prototype $w_{s(v)}$ in the set W shall be small. Here the "mapping function" $s(.)$ has been introduced to denote for each $v \in V$ the index of the closest prototype in W. However, the main advantage of the self-organizing feature maps algorithm is that it also arranges the w_r so that the associated mapping $s(.)$ from V to A maps the topology of the set V. This topology is defined by the metric relationships of its vectors $v \in V$, onto the topology of A in a least distorting way. Usually A is a bidimensional set of indexes (see Fig. 1). For further details see for example [18], [8].

In this context we consider the application of the LVQ to the problem of studying and distinguishing between different stellar groups of the solar neighbourhood. The final segregation is also analyzed taking into account the moments of the residual stellar velocities up to order four.

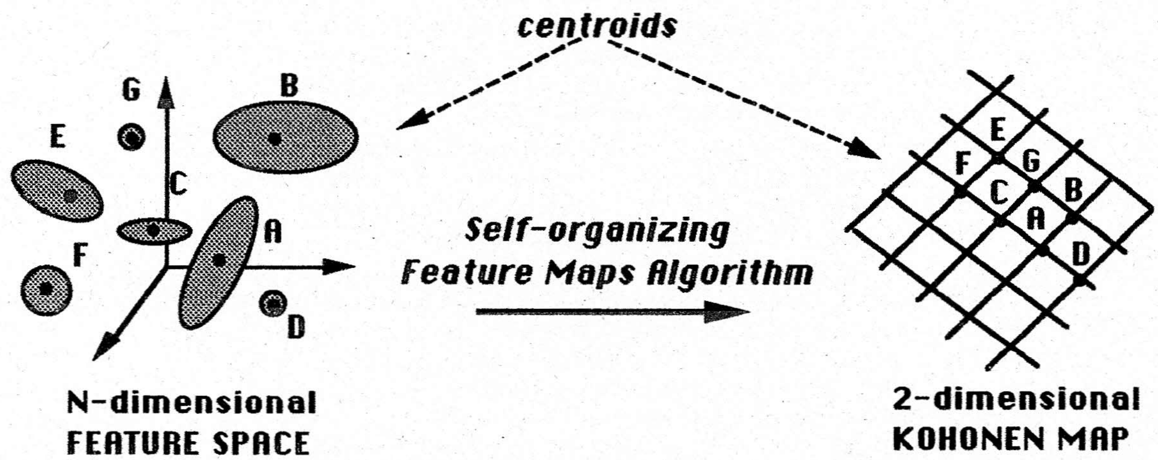

Fig.1. *The Kohonen map: the adjacent property in the characteristic space is preserved*

2. Neural Network Calculations and Results

The observational data considered is the [10] stellar catalogue (see [11]). It was made from the S.A.O. catalogue that contains all the kinematic and astrophysical information available about more than 250.000 stars [24], [25]. The final catalogue contains 12.824 stars, with enough information to estimate the spatial velocity. The most relevant data for our purposes are:

- The galactic longitude, latitude (degrees), and the heliocentric distance (parsecs), l, b, r respectively.
- The spectral type and luminosity class, sp and l, c respectively.
- The Johnson photometric magnitude and indexes m_v, $B-V, U-B$.
- The spatial residual velocities taking out the simple rotation model, U_1, V_1, W_1, in a galactic heliocentric reference frame (Km/s).
- The velocity components U, V, W in the same reference system, and with the same units as, the residual velocity components.

The LVQ method has been applied to working in a 14 dimensional characteristic space, i.e. the space formed by the 14 properties described above. Also, we have considered a symmetry assumption to improve the sensitivity of the process. We assume the symmetry referred to the galactic plane for the galactic latitude and for the residual velocity component perpendicular to galactic plane, this means considering its absolute values $|b|$ and $|W_1|$, respectively. In the calculations, we have taken $8 \times 8 = 64$ centroids to be determined after $4 \cdot 10^6$ training iterations of the neural network. So, the resultant Kohonen map consists of a two-dimensional grid of $8 \times 8 = 64$ points, with a 14 dimensional centroid vector for every one.

To evaluate the results, it is interesting to keep in mind what we can obtain and how the possible results in terms of the Kohonen map aspect are translated. Generally speaking, we can say that if the i-th characteristic is not relevant, the values of the centroid vector i-th components will be similar. Hence, a normalized image representation of the Kohonen map for the i-th characteristic will not present a systematic contrast between the cells representing every point. But when the i-th characteristic is significant in the segregation problem, a systematic trend in the darkness/brightness aspect of the Kohonen map cells must appear. Regarding the relevant characteristics i and j, with a systematic relation between them, the translation of the Kohonen map aspect for the i-th and j-th planes has a strong correlation between the systematic darkness/brightness trends.

The centroids obtained for the stellar catalogue present, as the main significant characteristics, the distance r and the absolute value of the residual velocity component $|W_1|$. The distance is directly correlated with other significant characteristics, such as the spectral type and the luminosity class. Also a less direct correlation is referring to the galactic longitude. Other significant characteristics strongly correlated with the distance distribution are $|b|$, the m_v magnitude, $U-B$ color index, and especially $|W_1|$ (isoline plot of Fig. 2). However, the characteristics U_1 and V_1 present a systematic contrast and correlation with the distance for values of this variable greater than 1000-1200pc. For the rest of the domain in the Kohonen map, a systematic contrast does not appear. The $B-V$ color index and the velocity components U, V present the same features over the entire map. But the W is strongly correlated with the galactic longitude.

Using the Kohonen map we can segregate the catalogue from an astronomical point of view. Indeed, the characteristic $|W_1|$ (spatial velocity perpendicular to the galactic plane), gives us the maximum perpendicular distance to which the star can climb away from the plane. This parameter is related directly with its metalicity, and with the population to which the star can belong: Disk and Halo populations with low and high values, and the recent proposal of a third population, the *Thick Disk*, with intermediate values of $|W_1|$ [12].

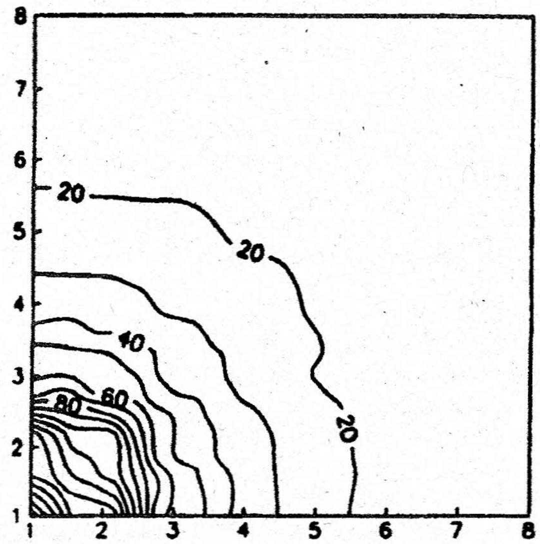

Fig. 2. *Isoline Kohonen map for the characteristic absolute value of the velocity perpendicular to the galactic plane, $|W_1|$ in Km/s.*

In Fig. 3, the $|W_1|$ in function of the distance perpendicular to the galactic plane $|Z|$ appears, calculated as $r \sin |b|$, for the 64 centroids obtained. We can observe that in general $\frac{d|W_1|}{d|Z|} > 0$, a result that is concordant with the intrinsic information contained in $|W_1|$ mentioned above. Also, and more importantly, we can distinguish between three groups of neighbouring centroids in the Kohonen map: (A) with $|W_1| \leq 24 Km/s$ (distances between 1 and 550 pc), (B) with $24 < |W_1| \leq 60 Km/s$ (distances between 380 and 1400 pc), and (C) with $|W_1| > 60 Km/s$ and distances greater than 1400 pc (Fig. 4). These intervals, distinguished without any extra-statistical consideration, agree very well with the kinematic bins considered by [1]; and related to the metalicity and galaxy populations: region (A) with a predominant thin disk component, (B) with the thick disk and (C) with the Halo population.

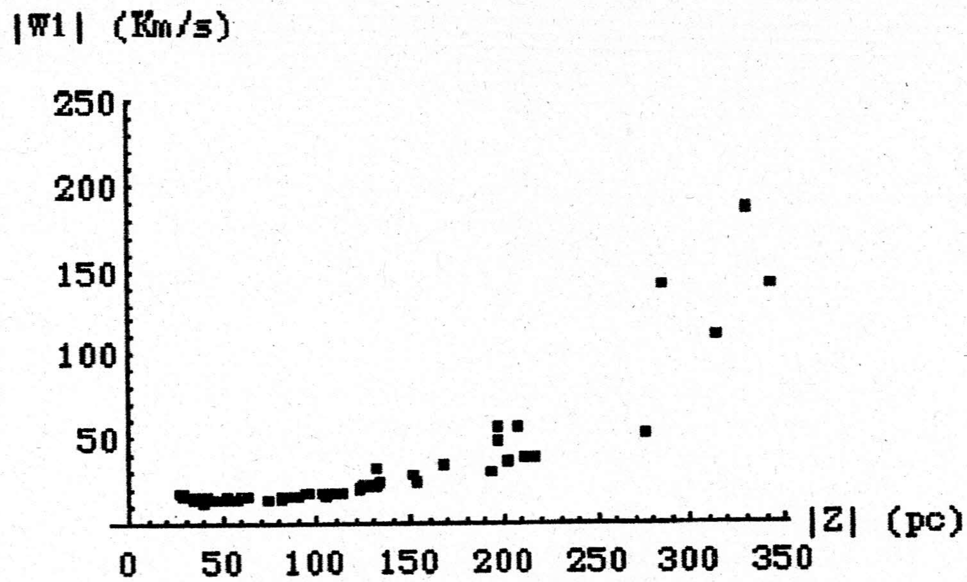

Fig. 3. *$|W_1|$ plotted against the height above the galactic plane, $|Z|$, for the 64 centroids calculated.*

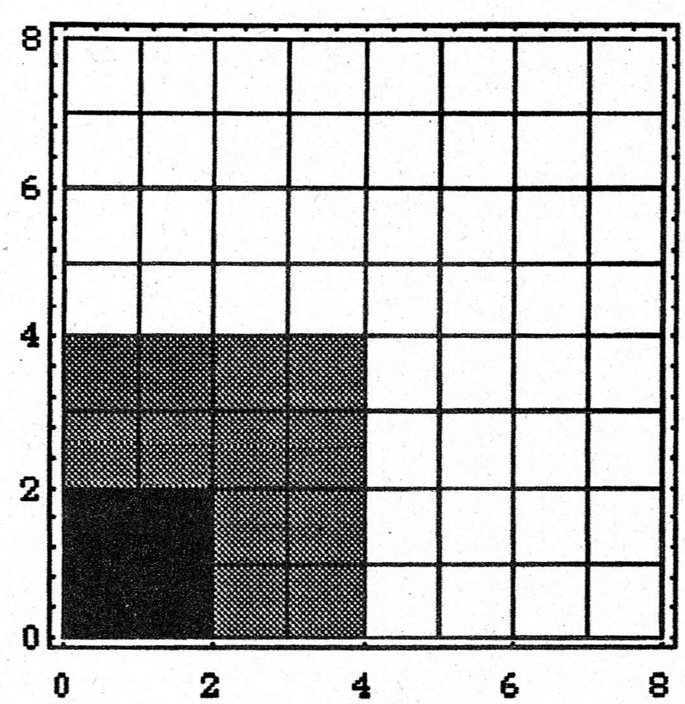

Fig. 4. *Distribution of the three families of centroids in the Kohonen map, regarding the $|W_1|$ characteristic: (A) with $|W_1| \leq 24Km/s$ (white squares), (B) with $24 < |W_1| \leq 60Km/s$ (gray squares) and (C) with $|W_1| > 60Km/s$ (black squares).*

3. Kinematical Interpretation of the Stellar Groups

Starting from the total stellar catalogue, three groups have been obtained with some common features. The interpretation of the resulting subsamples has been done according to the average properties of the stars belonging to each group. In particular, the spatial velocity perpendicular to the galactic plane and the heliocentric distance. However, since the central velocity moments of each stellar group have been computed up to fourth-order, it is possible to analyze their kinematical features by other well known statistical criterions and, therefore compare and interpret the resulting LQV method segregation in terms of kinematical behavior of stellar systems.

The method of analysis used in this section is based on the fact that, according to gas dynamics, a galactic component can be locally described by a multivariate Gaussian function in the particular velocities. In astronomical literature, this velocity distribution of the Schwarzschild type. It has two important properties [26]: (1) The odd-order central moments are zero. In particular, this is significant for the third moments. (2) The second and fourth central moments satisfy, by components, the following relationship:

$$\mu_{ijkl} - \mu_{ij}\mu_{kl} - \mu_{ik}\mu_{jl} - \mu_{il}\mu_{jk} = 0$$

Thus, it is easy to identify this type of distribution. Of course, the foregoing assumption can be adopted in the solar neighborhood to study some particular stellar components. Also, the local stellar kinematics and dynamics can be described by a superposition of two Schwarzschild velocity distributions [5], associated with two stellar components: thin disk and thick disk stars. In fact, in the solar neighborhood, there is also a few halo stars. On the other hand, the characteristic scale heights over the galactic plane for the thin disk and thick disk populations are 0.3 kpc and 1.3 kpc [13] respectively. For these reasons, we shall focus on the interpretation of the kinematic features of the stellar groups A and B, which can be clearly considered as local samples.

It is worth noting that neither the A group velocity moments nor the B ones (Tab. 1.), have the values usually associated with thin disk or thick disk components, according to astronomical literature. However, the B group velocity moments are very near to thick disk values, and A group moments represents the local velocity distribution, according to those of [9], and with those of Figueras' sample [6] selected from several astronomical criterions. Moreover, the central moments of each stellar group cannot be explained with a Schwarzschild velocity distribution alone, since the third moments are not null, and the second and fourth moments do not satisfy the corresponding relationships.

Thus we may suspect that the A group corresponds to a typical local stellar sample containing a main component of thin disk population, being contaminated by a significant number of stars belonging to the thick disk component [3], [33], and, perhaps, also a few halo components. Similarly for the B group central velocity moments, taking into account the characteristic scale heights for the stellar components, we may think of a main thick disk component with some contamination of halo population.

The central velocity moments resulting from a superposition of two Schwarzschild distributions are not arbitrary. The main properties of such distributions [6] can be summarized as follows:

1. Concerning the odd-order moments of the total distribution (in this case they are not null), there exist four vanishing independent linear combinations of the third moments.

2. The total second, third, and fourth moments are constrained by a set of twelve relationships, providing the fundamental superposition parameters, which are associated with the skewness and the peakedness of the distribution.

These properties lead us to compute the partial second moments of the components, the mean velocity difference between components, and the percentage of mixture.

If we apply this method, by assuming that the A group is a superposition of two-Schwarzschild components, a mixture with a percentage of 90% (+/− 6%) for the first component is obtained. This considers around 9.535 stars for the first component and 1.060 stars for the second one. The partial second central moments and the components of the mean velocity difference **w** are listed in Tab. 2a. Let us note that the non-diagonal partial moment corresponding to the indices 12 are nearly zero or very small, and the other non-diagonal partial moments are null. This fact must be interpreted as an almost insignificant vertex deviation of the velocity ellipsoids associated with the stellar components. Furthermore, besides the expected differential movement in the rotational direction between the stellar components, a significant differential radial movement has also been obtained. The velocity dispersions (square root of the diagonal moments) of the first component (27:15:11) are in total accordance with the thin disk population [31], and those corresponding to the second component (72:55:56) agree with thick disk population, perhaps biased towards older disk stars. This second component is likely contaminated by halo stars.

Similarly, if we assume that the B group is a superposition of two-Schwarzschild components, a mixture with a percentage of 95% (+/− 4%) for the first component is obtained. The partial second central moments and the vector components of the mean velocity difference are listed in Tab. 2b. In this case as well, the non-diagonal partial moments are zero. The velocity dispersions of the first component are 46:40:42, in total accordance with the typical thick disk population [13], and, for the second component, the dispersions are 130:110:110, corresponding to the values of the halo population [2], [21]. The result implies that the stellar group B consists in a nearly pure thick disk population with 1.590 stars, contaminated with approximately 85 halo stars. It is worth mentioning that for the second component the z-velocity dispersion is nearly the same as that for stellar group C, a typical halo component value. However, for this stellar group, the central moments associated with the other components of the velocity show an important presence of non-halo components.

	Group A (a)		Group B (b)		Group C (c)	
	Moment	Error	Moment	Error	Moment	Error
U_o	10.87	0.34	4.5	1.3	27.1	4.7
V_o	18.09	0.26	13.0	1.1	3.6	3.7
W_o	7.65	0.20	5.0	1.2	21.6	5.7
μ_{11}	1257	31	2820	150	8000	700
μ_{22}	735	24	2110	130	5040	470
μ_{33}	435	16	2440	130	11900	890
μ_{12}	115	18	-	-	1200	400
μ_{112}	-23700	2300	-	-	-	-
μ_{222}	-36300	3900	-	-	-	-
μ_{233}	-8700	1400	-	-	-	-
μ_{1111}	11870000	950000	48200000	5800000	246000000	43000000
μ_{1122}	3280000	300000	15700000	2200000	61800000	8800000
μ_{2222}	6740000	820000	33100000	5800000	106000000	20000000
μ_{1133}	1810000	180000	13100000	1900000	83000000	11000000
μ_{2233}	1580000	180000	10600000	1200000	52300000	6400000
μ_{3333}	3050000	600000	34600000	4200000	434000000	57000000

Tab. 1. *The mean residual velocities,* U_0, V_0, W_0 *and the non-vanishing central moments (3-sigma level) of order two, three and four* $\mu_{ij}, \mu_{ijk}, \mu_{ijkl}$ *respectively, with the associated errors are listed for the groups A, B and C of stars (those with residual velocity greater than 300 Km/s have not been taken into account). The units of the* n-th *order moments are* $(Km/s)^n$.

At this point it seems interesting to know what would be the percentage of mixture in the second component of the A group in terms of both components obtained in the B group. By applying the superposition criterion, we find that this second component is composed of a 82% (+/− 5%) of pure thick disk stars with dispersions 46:40:42. The remaining halo stars have dispersions of 130:110:110. Therefore, the stellar group A would be composed of three stellar components with 9.535 thin disk stars, 870 thick disk stars, and 190 halo stars.

	Group A (a)		Group B (b)	
	Thin	Thick	Thick	Halo
% of population	90	10	95	5
μ_{11}	710	5200	2100	17000
μ_{22}	220	3000	1600	12000
μ_{33}	130	3100	1800	12000
μ_{12}	45	130	-	-
μ_{13}	-	-	-	-
μ_{23}	-	-	-	-
w_1	12		-	
w_2	50		55	
w_3	-		-	

Tab. 2. *The non-vanishing partial second central moments and the non-vanishing vector components of the mean velocity difference are listed for groups A and B of stars taken as in Tab. 1.*

4. Conclusions

In this paper we have applied the learning vectorial quantification method to the study of a stellar catalogue that contains 3D positions jointly with spectral, photometric and kinematic data for a total of more than 12.000 stars in the solar neighbourhood. We have found the existence of three regions of neighbouring centroids in the resulting Kohonen map from the $|W_1|$ characteristic. Another important feature in the classification has been the distance. Hence, the resulting groups present properties related with the *localicity*. These three regions seem to correspond to the Thin Disk (A), Thick Disk (B), and Halo populations (C). These results have been contrasted (for samples A and B) assuming that every sample is a superposition of two Schwarzschild components for the velocity distribution. Under these assumptions, the analysis gives a *purity* of 90 and 95 percent for the samples A and B as thin and thick disk stars.

Acknowledgements

This work has been completed with the support from Spanish goverment funds PB90-0478. We would like to acknowledge to M.Juan-Zornoza for his interested discussions of this paper.

References

[1] Carney B.W., Latham D.W., Laird J.B.: Astron. J. **97**, 1989, 423.

[2] Carney B.W., Latham D.W.: Astron. J. **92**, 1986, 60.

[3] Chiu L.T.G.: Astron. J. **85**, 1980, 812.

[4] Clube S.V.M.: Mont. Not. R. Astron. Soc. **159**, 1972, 289.

[5] Cubarsí R.: Astron. J. **99**, 1990, 1558.

[6] Cubarsí R.: Astron. J. (To appear on May). 1992.

[7] Cubarsí R., Juan M., Sanz J.: Astrophys. Space Sci. **170**, 1990, 197

[8] Domany E., van Hemmen J.L., Schulten K.: Models of Neural Networks. Springer-Verlag, Berlin, 1991, 281-305.

[9] Erickson R.R.: Atrophys. J. **195**, 1975, 343.

[10] Figueras F.: Doctoral thesis. University of Barcelona, Barcelona, 1986.

[11] Figueras F., Núñez J.: Astrophys. Space Sci. **177**, 1991, 483.

[12] Gilmore G., Reid N.: Mont. Not. R. Astron. Soc. **202**, 1983, 1025.

[13] Gilmore G., Wise R.F.G.: The Galaxy. G.Gilmore and R.F.Carswell (eds), Reidel, Dordrecht, 1987, 247.

[14] Hernández-Pajares M., Núñez J.: Astrophys. Space Sci. **170**, 1990a, 187.

[15] Hernández-Pajares M., Núñez J.: Error, bias and uncertainties in Astronomy, C. Jaschek and F.Murtagh, (eds), Cambridge Univ. Press, 1990b, 339.

[16] Jaschek C.: Proceedings of the Colloquium HIP-PARCOS. Scientific Aspects of the Input Catalogue Preparation II, held at Sitges, Catalonia, Spain from 25 to 29 January 1988, J.Torra, C.Turon, (eds).

[17] Juan-Zornoza J.M., Sanz i Subirana J.: Astrophys. Space Sci. **185**, 1991, 95.

[18] Kohonen T.: Self-Organization and Associative Memory. Springer Series in Information Sciences, **8**, Springer, Berlin-heidelberg, 1984.

[19] Kondratev B.P., Ozernoy L.M.: Astrophys. Space Sci. **84**, 1982, 431.

[20] Moreno M., Torra J.: Error, bias and uncertainties in Astronomy. C.Jaschek and F.Murtagh, (eds), Cambridge Univ. Press, 1990, 377.

[21] Norris J.E.: The Galaxy. G.Gilmore and R.F.Carswell (eds), Reidel, Dordrecht, 1987, 297.

[22] Núñez J.: Rev. Mex. Astron. Astrof. 8, 1983, 91.

[23] Oblak E.: Astron. and Astrophys. **123**, 1983, 238.

[24] Ochsenbein F.: CDS Inf. Bull. **19**, 1980, 74.

[25] Ochsenbein F., Bischoff M., Egret D.: Astron. Astrophys. Suppl. Ser. **43**, 1981, 259.

[26] Orús J.J.: Apuntes de Dinámica Galáctica. University of Barcelona, 1977.

[27] Palous J., Piskunov A.E.: Astron. and Astrophys. **143**, 1985, 102.

[28] Rius A.: Doctotral thesis. University of Barcelona, Barcelona, 1974.

[29] Ros R.M.: Rev. Mex. Astron. Astrof. **11**, 1985, 23.

[30] Sanz Subirana J., Juan Zornoza J.M., Catalá M.A.: Boletin Astronómico del Observatorio de Madrid, XII, n.4, 1990, 136.

[31] Strömgren B.: The Galaxy. G.Gilmore and R.F.Carswell (eds), Reidel, Dordrecht, 1987, 229.

[32] Wyse R.F.G., Gilmore G.: Astron. Astrophys. **60**, 1986, 263.

[33] Wyse R.F.G., Gilmore G.: Astron. J. **91**, 1986, 855.