

# A non-parameter global method of interpreting star count data

B. Chen

Departament d'Astronomia i Meteorologia, Universitat de Barcelona, Avda. Diagonal 647, E08028, Barcelona, Spain

Received June 26; accepted December 3, 1995

**Abstract.** — Galactic structure studies involve the manipulation of star catalogues and their global comparison with sophisticated synthesis models. In this paper, we present a detailed multivariate star count analysis algorithm. By combining the photometric and kinematic information, it performs a global analysis of the multivariate star counts including magnitude, colours, and proper motions. The method allows us to constrain galactic structure parameters and derive the relative densities of different components and populations in the Galaxy. Monte Carlo simulations have been carried out to demonstrate the viability of the method. A simple example of application to real star count data is given. We found the asymmetric drift of the thick disk to be  $80 \pm 10 \text{ km s}^{-1}$  at a 3 sigma level. Detailed investigations from a complete survey of several galactic directions are in preparation.

**Key words:** methods: data analysis — methods: statistical — Galaxy: structure — Galaxy: stellar content

## 1. Introduction

Star counts contain valuable information about the Galaxy and its stellar components, therefore, since the invention of the telescope, they have long been used to understand the fundamental features of the formation, evolution, and structure of the Galaxy in general (Bok 1937).

Over the last two decades, on the one hand, complete samples (Chiu 1980; Bienaymé et al.; 1992; Soubiran 1992; Majewski 1992; Ojha 1994a, b) of stars have been obtained with good accuracy in photometry, astrometry and radial velocities over sufficiently large areas of sky that random errors due to counting statistics are unimportant. On the other hand, the main features of Galaxy properties have been identified. It has therefore been possible to design synthetic models (Gilmore 1984; Bahcall & Soneira 1980; Buser & Kaeser 1985; Robin & Crézé 1986a) of galactic stellar populations with a certain degree of complexity. The model allows us to calculate the expected star counts and colour distributions explicitly.

Since the 1920's (Kapteyn & van Rhijn 1920), some astronomers have inferred the shapes and parameters of the different stellar populations by inverting the fundamental equation of stellar statistics. The classical procedures have been used recently by Becker (1980) and by Gilmore & Reid (1983). However, the inversion of the integral equation from the projected number of stars in the sky is an unstable mathematical procedure that can produce unreliable results when the number of stars is small.

The alternative analysis technique which has been used by many astronomers (Bahcall 1986; Pritchet 1983; Robin & Crézé 1986b; Reid & Majewski 1993) is to test the consistency of a given model with the observations in the observational space. The comparison between the model and the observations is made by binning the data to a histogram of counts as a function of the variables of interest.

It is evident that the limited star count observations which can be achieved on large collections of faint objects can hardly derive intrinsic stellar parameters such as distance, mass, age, space velocity, chemical composition, evolutionary stage or interstellar reddening for an individual star. However some information relevant to the distribution of these quantities is reflected in the  $n$ -dimensional distribution of observables: observed apparent magnitudes result from the combinations of absolute magnitudes and the distance; spectral types contain colours which may be altered by metallicity. Age also determines the velocity distribution, which, in combination with the distances, produces the proper motions. The connection of the observed distributions to the main processes they come from is a rather difficult problem in multivariate analysis (Pritchet 1983; Crézé et al. 1991). We need to perform a global analysis of the multivariate star count samples including all available observables simultaneously.

Buser & Kaeser (1985) and Buser & Rong (1995) have compared the predicted and observed distributions in the space of three-dimensional photometric survey data; they used a  $\chi^2$  statistic to investigate the large-scale structure of the Galaxy in a large number of galactic directions.

Crézé et al. (1991), Robin & Chen (1992), Chen (1993) and Chen (1996) have used classical multivariate data analysis (Cluster Analysis, Principal Components Analysis, Multiple Discriminant Analysis) to investigate the galactic structure parameters. They try to separate the thick disk from disk and halo and constrain the physical parameters of the thick disk.

In this paper, we present a multivariate star count analysis algorithm based on pattern recognition theory (Sect. 2). It uses a non-parameter approach to estimate the probability density function in the  $n$ -dimensional space of observables, and perform a global analysis of the multivariate star counts including magnitude, colours, and proper motions simultaneously.

The basic schematic structure in our algorithm is shown in Fig. 1. It can be described from top to bottom as follows:

Based on our knowledge of Galaxy properties, such as density law, luminosity function and colour-magnitude diagram for each stellar component, one can build a Galaxy model of population synthesis from the equation of stellar statistics. The model allows us to create a model simulated catalogue according to the same selection criteria as the real star counts. Physical as well as observed properties of each simulated star are known. The simulated catalogue is merged with the corresponding observed star counts. Then we calculate the probability density function (PDF, hereafter) at each point in this merged catalogue (Sect. 2.1). This probability density function is related to the physical quantities, which are used explicitly in the galactic model. By classifying the model catalogue data according to their physical properties, the contributions of observed stars can be fitted by a linear combination of the contributions of the model groups (Sect. 2.2). This method allows us to determine the physical parameters and derive the relative densities of different components and populations in the Galaxy.

In Sect. 3, a series of Monte Carlo simulations has been carried out in order to demonstrate the viability of the method. In Sect. 4, a simple example of application to real star count data is given. In Sect. 5, we summarize the main conclusions. This method has been used successfully to interpret three star count catalogues (Chen 1993). A detailed investigation from new star count surveys will be published in a forthcoming paper.

## 2. The method

### 2.1. Multivariate Kernel estimation

Galactic structure studies involve the manipulation of multivariate catalogues and their comparisons with sophisticated synthesis models. The observational star counts are derived from a complete sample of stars in a certain region and brighter than a given apparent magnitude.

Model simulated catalogues have been created according to the same selection criteria as the real star counts.

The real stars are merged with model predicted ones in the  $d$ -dimensional space of observables, including magnitude, colours, and proper motions. Suppose  $p(\mathbf{x})$  is the true PDF, then a multivariate kernel estimator  $\hat{p}(\mathbf{x})$  is given by:

$$\hat{p}(\mathbf{x}) = \frac{1}{nh^d} \sum_{i=1}^n \frac{1}{|\Sigma|^{1/2} (2\pi)^{d/2}} e^{-\frac{1}{2h^2}(\mathbf{x}-\mathbf{x}_i)' \Sigma^{-1}(\mathbf{x}-\mathbf{x}_i)} \quad (1)$$

where  $n$  is the number of stars in the merged sample,  $h$  is the window width,  $d$  is the dimension of the variables,  $x_i$ , ( $i = 1, \dots, n$ ) is the sample set defined in  $d$ -dimensional space, and  $\Sigma$  is the variance-covariance matrix of the sample.

Choice of window width plays a very important role in the kernel estimator. An optimal window width has been derived by Silverman (1986) from minimizing the approximate mean integrated square error. This optimal window can be written as:

$$h = \left(\frac{4}{d+2}\right)^{1/(d+4)} \sigma n^{-1/(d+4)} \quad (2)$$

where  $\sigma$  is the average marginal variance,  $\sigma^2 = d^{-1} \sum_{i=1}^d \sigma_i^2$

This optimal  $h$  is adequate for a wide range of densities (Hand 1981) and allows a reliable, automatic and therefore objective choice of the smoothing parameter. The detailed discussions of the kernel estimation can be found in Silverman (1986) and Cabrera-Cano & Alfaro (1990).

### 2.2. Fitting procedure

We classify simulated stars into several homogeneous groups according to their physical properties, for example, groups can be formed with different populations or age groups (disk, thick disk, halo, ...) or/and luminosity class (dwarfs, giants).

One can determine the contributions of each physical group  $\omega_m$  in the simulated catalogue by  $\hat{p}(\mathbf{x}|\omega_m)$  and the total observed points by  $\hat{p}(\mathbf{x})$  at each point  $\mathbf{x}$  by Eq. (1). To fit the model parameters to the data we suppose that the contributions of observed stars in each point of the  $d$ -dimensional space can be fitted by a linear combination of the contributions of the model groups. Let  $\alpha_{\omega_m}$  be the coefficient to apply to the group  $\omega_m$  in order to fit the data:

$$\Sigma \alpha_{\omega_m} \hat{p}(\mathbf{x}|\omega_m) = \hat{p}(\mathbf{x}) \quad (3)$$

This equation must be valid at each point  $\mathbf{x}$  leading to a system of  $n$  equations if  $n$  is the total number of stars in the merged catalogue.

Suppose  $\hat{p}(\omega_m)$  is the a priori probability density of  $\omega_m$ , by Bayes rules, the joint probability density can be expressed as:

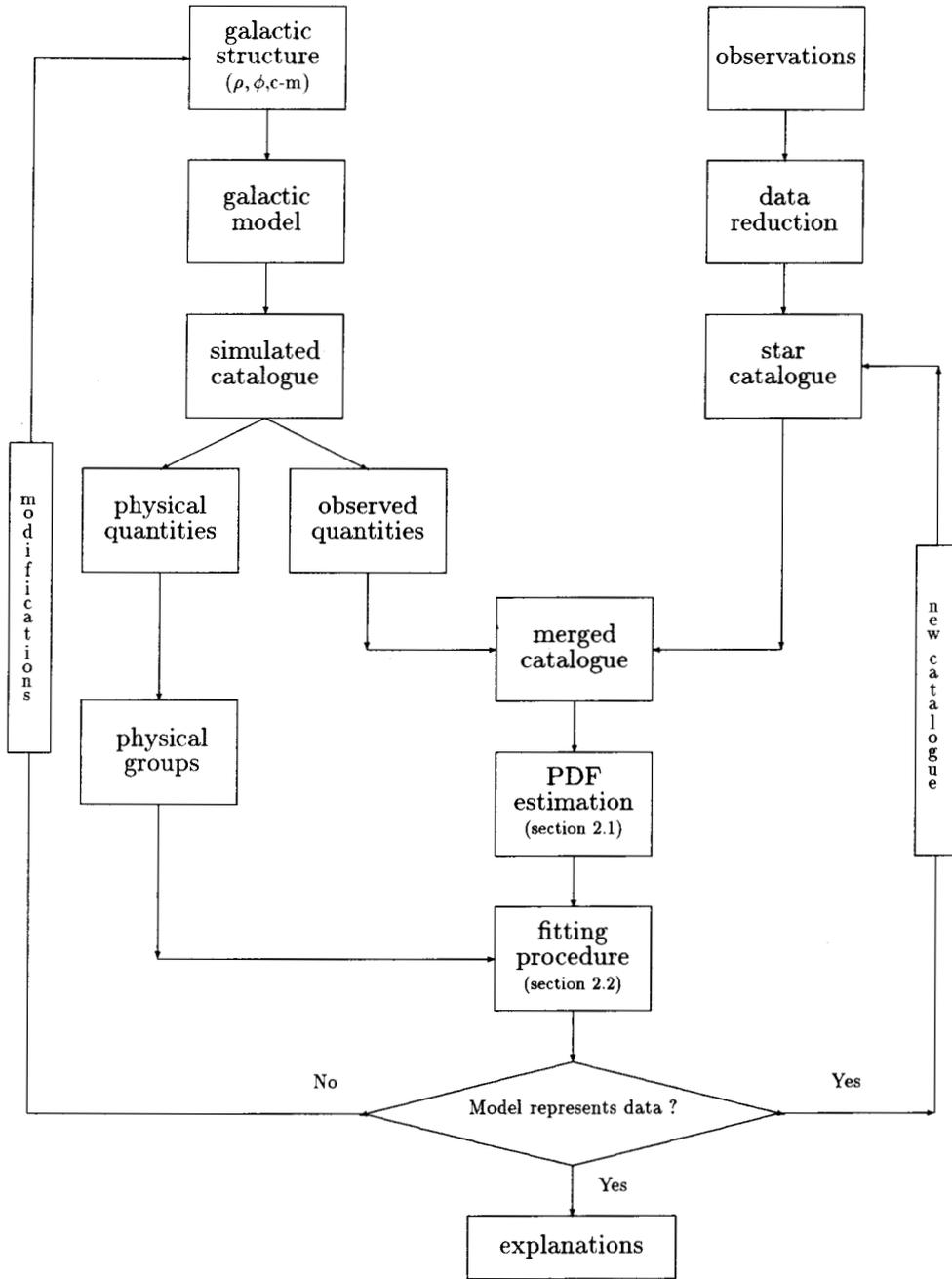


Fig. 1. Algorithm flow chart of the method, showing the principal procedural steps

$$\hat{p}(\omega_m | \mathbf{x}) = C \times \hat{p}(\mathbf{x} | \omega_m) \hat{p}(\omega_m) \quad (4)$$

where  $C$  is a normalization. We derive weighted means and the covariance matrix for each group from the observed data as follows:

$$\overline{X_l^{\omega_m}} = \frac{\sum_{i=1}^{n_1} \hat{p}(\omega_m | \mathbf{x}_i) x_{il}}{\sum_{i=1}^{n_1} \hat{p}(\omega_m | \mathbf{x}_i)} \quad (5)$$

$$\sigma_{lj}^{\omega_m} = \frac{\sum_{i=1}^{n_1} \hat{p}(\omega_m | \mathbf{x}_i) (x_{il} - \overline{x}_l)(x_{ij} - \overline{x}_j)}{\sum_{i=1}^{n_1} \hat{p}(\omega_m | \mathbf{x}_i)} \quad (6)$$

where the weight is the posterior probability ( $\hat{p}(\omega_m | \mathbf{x})$ ),  $l, j$  are two axes of the multi-dimensional space,  $x_{il}$  indicates the  $i$ th stars with its  $l$ th variable in the observed catalogue, and  $n_1$  is the number of stars in the observed catalogue. Equations (5) and (6) describe the statistical properties of observed stars in each group. We compare them with

that of the model catalogue, and disagreement between the model and the data being interpreted as erroneous physical properties of the model groups.

### 3. Monte Carlo simulations

In order to test the reliability of the method it has been applied to simulated 2-dimensional data  $(X, Y)$  containing three physical groups, in which the observables have multivariate normal distributions. We also produce a set of simulated data that will be used as “observed” data for comparison (Hereafter the term “observed” refer to this second set of simulations). They differ from the model data in the number of stars, the mean and standard deviation of each physical group. Our goal is to detect these differences by our multivariate kernel estimation method.

A series of Monte Carlo simulations has been carried out to mimic all probable cases happening between the model and the data. We demonstrate them as follows:

#### 3.1. Well separated groups and no shifts

As the first test, we suppose that the model exactly represents the data, it puts all the real groups at the right place with the right covariance matrix and number of stars. In Table 1, the input parameters and the derived parameters are shown on the left hand side and the right hand side, respectively.  $C$  is the mean,  $\sigma$  the standard deviation, and  $\alpha_{\omega_m}$  the ratio of number of stars in the model to the “observed” data in each group. 500 stars have been generated in each group. For the model catalogue, each star is defined in a given group, the derived parameters ( $C$ ,  $\sigma$ ) are computed directly from each group, while for the “observed” data they are computed from the multivariate kernel fitting method (Eqs. (5) and (6)). The coefficient  $\alpha_{\omega_m}$ , its error ( $\epsilon_{\omega_m}$ ) and the correlation coefficients ( $r_{lj}$ ) are estimated from the least-square solutions (Eq. 3). From table 1, by comparing the derived parameters ( $C$ ,  $\sigma$ ) in each group between the model and the “observed” data and  $\alpha_{\omega_m}$ , we can see that the model can well represent the “observed” data and the method is able to derive the values of the parameters ( $\alpha_{\omega_m}$ ) correctly within an error of 1%.

#### 3.2. Poorly separated groups and no shifts

The situation is the same as in the first case, except that the groups are not well separated, their centres are closer, and the correlation coefficients among groups ( $r_{lj}$ ) are larger. The results are shown in Table 2. From the derived parameters ( $C$ ,  $\sigma$ ), we can see that the mean and the dispersion of each group in the model are in good agreement with that in the ‘observed’ data, but the relative densities of the group 1 (0.89 in place of 1.0) and

group 3 (1.11 in place 1.0) are not well represented because of the large correlation coefficients between group 1 and group 2 ( $r_{12}=-0.72$ ), and between group 2 and group 3 ( $r_{23}=-0.70$ ).

#### 3.3. Wrong model density and no shifts

Here, we suppose the model gives all the real groups at the right place with the right covariance matrix, but the global densities deviate by unknown factors  $\alpha_{\omega_m}$ . The test has been made with the ratio of observed groups to model groups of 1, 0.5, and 1.5 respectively (see left hand side of Table 3). From the right hand side of the Table 3, we can see the derived coefficients in group 1 and group 3 are basically recovered.

#### 3.4. Wrong model density and systematic shifts

We now discuss the case where the model differs from the “observed” data by the number of stars, and the mean of each physical group. The test has been made with the relative density of observed groups to model groups of 1, 0.5 and 2.0, respectively, and a systematic shift exists between the model and the ‘observed data’ in group 3 (see left hand side of Table 4).

From the derived parameters in Table 4, we notice that the systematic errors (italic) introduced in the simulation have been mostly recovered while the proportions  $\alpha_{\omega_m}$  are not well recovered in group 3. This is mainly due to the fact that the groups are not well centered because of the shift introduced.

We give to the model the shifted mean derived from the data in the first step (i.e. 7.8 and 12.8 in  $X$  and  $Y$  in group 3) and we recompute the derived parameters (Table 5) as before. This step is a simulation of a correction applied to a model in order to fit the data.

From Table 5, we see that, once the shift (hence the systematic errors) in group 3 has been corrected, the correct values of the mean in group 2 and group 3 are well recovered (0.5 in the place of 0.5 in group 2 and 1.91 in place of 2.0 in group 3)

#### 3.5. Incorrect model and wrong variance

The dispersion of a variable such as velocity, metallicity for a stellar population is a very important physical parameter in the field of the galactic structure.

In this last test, we suppose that the data (“observed”) have a larger dispersion in group 3, and twice the number of stars in the group 2. The input parameter can be seen in detail in Table 6.

The fitting process has been applied in two steps:

From the parameters derived in Table 6, we notice that the systematic errors (italics) introduced in the simulation have mostly been recovered while the proportions  $\alpha_{\omega_m}$  are not well recovered in group 3. This is mainly due to the

**Table 1.** Well separated groups and no shifts: the derived parameters( $C, \sigma$ ) and the coefficients ( $\alpha_{\omega_m}$ ) show that the model can well represent the “observed” data

		input parameters			derived parameters		
		group 1	group 2	group 3	group 1	group 2	group 3
		(X, Y)	(X, Y)	(X, Y)	(X, Y)	(X, Y)	(X, Y)
Model	$C$	(1,6)	(3,8)	(5,10)	(1.0,6.1)	(3.0,8.1)	(4.9,10.0)
	$\sigma$	(1,1)	(1,1)	(1,1)	(1.0,1.0)	(1.0,1.0)	(1.0,1.0)
Data	$C$	(1,6)	(3,8)	(5,10)	(1.1,6.1)	(3.0,8.0)	(4.8,10.0)
	$\sigma$	(1,1)	(1,1)	(1,1)	(1.1,1.0)	(1.1,1.1)	(1.1,1.1)
$\alpha_{\omega_m}$		1	1	1	0.97	0.97	0.96
$\epsilon_{\omega_m}$					0.01	0.01	0.01
$r_{lj}$					$r_{12}=-0.23$	$r_{13}=0.07$	$r_{23}=-0.31$

**Table 2.** Poorly separated groups and no shifts: the derived parameters ( $C, \sigma$ ) are correctly recovered, but the relative densities of the group 1 (0.89 in place of 1.0) and group 3 (1.11 in place 1.0) are not well represented because of the large correlation coefficients between group 1 and group 2 ( $r_{12} = -0.72$ ), and between group 2 and group 3 ( $r_{23} = -0.70$ )

		input parameters			derived parameters		
		group 1	group 2	group 3	group 1	group 2	group 3
		(X, Y)	(X, Y)	(X, Y)	(X, Y)	(X, Y)	(X, Y)
Model	$C$	(1,6)	(2,5)	(3,4)	(1.0,6.1)	(1.9,5.0)	(3.0,4.0)
	$\sigma$	(1,1)	(1,1)	(1,1)	(1.0,1.0)	(1.0,1.0)	(1.0,1.0)
Data	$C$	(1,6)	(2,5)	(3,4)	(1.0,6.0)	(2.0,5.0)	(3.0,4.1)
	$\sigma$	(1,1)	(1,1)	(1,1)	(1.0,1.0)	(1.0,1.0)	(1.0,1.0)
$\alpha_{\omega_m}$		1	1	1	0.89	1.02	1.11
$\epsilon_{\omega_m}$					0.01	0.01	0.01
$r_{lj}$					$r_{12}=-0.72$	$r_{13}=0.40$	$r_{23}=-0.70$

**Table 3.** Wrong model density and no shifts: the derived coefficients are basically recovered

		input parameters			derived parameters		
		group 1	group 2	group 3	group 1	group 2	group 3
		(X, Y)	(X, Y)	(X, Y)	(X, Y)	(X, Y)	(X, Y)
Model	$C$	(1,6)	(3,8)	(5,10)	(1.0,6.1)	(3.0,8.1)	(4.9,10.0)
	$\sigma$	(1,1)	(1,1)	(1,1)	(1.0,1.0)	(1.0,1.0)	(1.0,1.0)
Data	$C$	(1,6)	(3,8)	(5,10)	(1.0,6.0)	(3.1,8.2)	(5.0,10.1)
	$\sigma$	(1,1)	(1,1)	(1,1)	(1.0,1.0)	(1.3,1.2)	(1.0,1.0)
$\alpha_{\omega_m}$		1	0.5	1.5	0.98	0.42	1.52
$\epsilon_{\omega_m}$					0.01	0.01	0.01
$r_{lj}$					$r_{12}=-0.24$	$r_{13}=0.07$	$r_{23}=-0.31$

fact that the data have a larger dispersion in group 3. In the second step, we increase the dispersion in group 3 in the model (i.e. 1.53, instead of 1 in group 3). In Table 7, we see that, once the dispersion in group 3 has been corrected, the correct values of the  $\alpha_{\omega_m}$  for each group are almost recovered.

#### 4. Example of application

Bienaymé et al. (1992) have carried out a magnitude-colour-proper motion survey (M5 sample hereafter) at in-

termediate galactic latitude towards the galactic center based on *UBV* Schmidt plates on a time baseline of 30 years. It covers 1.78 square degrees towards  $l = 2.7^\circ$  and  $b = 47.3^\circ$ . The precision of proper motions ( $\mu_l, \mu_b$ ) is  $0.2''/\text{century}$ . The photometric accuracy is about 0.08 to 0.10 in *B* and *V* to magnitude 17.5 and 0.12 in *U* at the plate limit.

The existence of an intermediate stellar population, call it a thick disk, is now well established (Gilmore & Reid 1983; Robin et al. 1989; Chen 1995). However, the structural parameters of the thick disk remain

**Table 4.** Wrong model density and systematic shifts: from the derived parameters on the right hand side, we can see the systematic errors introduced on the left hand side have mostly been recovered, while the proportions  $\alpha_{\omega_m}$  are not well recovered in group 3. This is mainly due to the fact that the groups are not well centered because of the shift introduced

		input parameters			derived parameters		
		group 1	group 2	group 3	group 1	group 2	group 3
		(X, Y)	(X, Y)	(X, Y)	(X, Y)	(X, Y)	(X, Y)
Model	<i>C</i>	(1,6)	(3,8)	(5,10)	(1.0,6.1)	(3.0,8.1)	(4.9,10.0)
	$\sigma$	(1,1)	(1,1)	(1,1)	(1.0,1.0)	(1.0,1.0)	(1.0,1.0)
Data	<i>C</i>	(1,6)	(3,8)	(8,13)	(1.0,6.0)	(2.4,7.6)	(7.8,12.8)
	$\sigma$	(1,1)	(1,1)	(1,1)	(1.1,1.1)	(1.1,1.3)	(1.4,1.4)
$\alpha_{\omega_m}$		1	0.5	2	1.0	0.42	0.23
$\epsilon_{\omega_m}$					0.03	0.04	0.04
$r_{lj}$					$r_{12}=-0.43$	$r_{13}=0.20$	$r_{23}=-0.49$

**Table 5.** Wrong model density and systematic shifts: we give the model the shifted mean derived from the data in the first step (see Table 4) and recompute the results. This step is a simulation of a correction applied to a model in order to fit the data

		input parameters			derived parameters		
		group 1	group 2	group 3	group 1	group 2	group 3
		(X, Y)	(X, Y)	(X, Y)	(X, Y)	(X, Y)	(X, Y)
Model	<i>C</i>	(1,6)	(3,8)	(7.8,12.8)	(1.0,6.1)	(3.0,8.1)	(7.8,12.9)
	$\sigma$	(1,1)	(1,1)	(1,1)	(1.0,1.0)	(1.0,1.0)	(1.0,1.0)
Data	<i>C</i>	(1,6)	(3,8)	(8,13)	(1.1,6.1)	(2.7,7.7)	(8.0,13.0)
	$\sigma$	(1,1)	(1,1)	(1,1)	(1.1,1.1)	(1.2,1.3)	(1.0,1.0)
$\alpha_{\omega_m}$		1	0.5	2.0	0.96	0.50	1.91
$\epsilon_{\omega_m}$					0.01	0.01	0.01
$r_{lj}$					$r_{12}=-0.46$	$r_{13}=0.01$	$r_{23}=0.0$

**Table 6.** Incorrect model and wrong variance: from the derived parameters, we notice that the systematic errors (italic) introduced in the simulation have mostly been recovered while the proportions  $\alpha_{\omega_m}$  are not well recovered in group 3. This is mainly due to the fact that the data have a larger variance in group 3

		input parameters			derived parameters		
		group 1	group 2	group 3	group 1	group 2	group 3
		(X, Y)	(X, Y)	(X, Y)	(X, Y)	(X, Y)	(X, Y)
Model	<i>C</i>	(1,6)	(3,8)	(8,13)	(1.0,6.1)	(3.0,8.1)	(8.0,13.1)
	$\sigma$	(1,1)	(1,1)	(1,1)	(1.0,1.0)	(1.0,1.0)	(1.0,1.0)
Data	<i>C</i>	(1,6)	(3,8)	(8,13)	(1.4,6.3)	(3.0,8.1)	(8.2,13.1)
	$\sigma$	(1,1)	(1,1)	(1.5,1.5)	(1.2,1.2)	(1.1,1.1)	(1.53,1.53)
$\alpha_{\omega_m}$		1	2	1	0.93	1.95	0.57
$\epsilon_{\omega_m}$					0.01	0.01	0.01
$r_{lj}$					$r_{12}=-0.40$	$r_{13}=0.00$	$r_{23}=-0.00$

controversial; the difficulty is that, until now, we cannot separate the thick disk stars from the disk or the halo in a histogram of colour or even in a colour-colour diagram. An illustrative example is that both 2-component and 3-component galactic models can fit the star count data.

M5 sample contains five observational parameters ( $V$ ,  $B - V$ ,  $U - B$ ,  $\mu_l$ ,  $\mu_b$ ) providing an opportunity to separate the thick disk and determine its physical parameters. Because different stellar populations with the same  $B - V$

colour index can be separated from their space velocities, and  $U - B$  colour gives a good discrimination between the three populations because of its sensitivity to the metallicity.

Trying to fit the observed distribution with a suitable model for the thick disk we found that these data are not at all sensitive to the scale height and the velocity dispersion used for the thick disk because they are correlated by the potential (through the Boltzmann equation).

**Table 7.** Incorrect model and wrong variance: in the second step, we increase the dispersion in group 3 in the model (i.e. 1.53, instead of 1 in group 3), we can see that, once the variance in group 3 has been corrected, the correct values of the  $\alpha_{\omega_m}$  for each group are almost recovered

		input parameters			derived parameters		
		group 1	group 2	group 3	group 1	group 2	group 3
		(X, Y)	(X, Y)	(X, Y)	(X, Y)	(X, Y)	(X, Y)
Model	C	(1,6)	(3,8)	(8,13)	(1.0,6.1)	(3.0,8.1)	(8.0,13.0)
	$\sigma$	(1,1)	(1,1)	(1.53,1.53)	(1.0,1.0)	(1.0,1.0)	(1.52,1.64)
Data	C	(1,6)	(3,8)	(8,13)	(1.3,6.4)	(3.0,8.1)	(8.1,13.0)
	$\sigma$	(1,1)	(1,1)	(1.5,1.5)	(1.2,1.1)	(1.1,1.1)	(1.53,1.55)
$\alpha_{\omega_m}$		1	2	1	1.02	1.97	0.99
$\epsilon_{\omega_m}$					0.01	0.01	0.01
$r_{lj}$					$r_{12}=-0.40$	$r_{13}=0.0$	$r_{23}=0.0$

However, these data are very sensitive to the circular velocity of the thick disk. We here derive the circular velocity of the thick disk using the M5 sample in the 5-dimensional space of observables simultaneously.

A grid of model simulated catalogues with different asymmetric drifts of the thick disk has been generated from Bienaymé et al. (1987) model of population synthesis. In the model, the galactic thin disk is represented by a sum of 7 components with different scale heights, and thin disk metallicities are a function of age, ranging from +0.03 to -0.37 dex. The metallicities for thick disk and halo are  $-0.6 \pm 0.25$  dex and  $-1.7 \pm 0.25$  dex, respectively, and the thick disk and halo luminosity functions are those of 47 Tuc and M3, respectively. Details of the model can be found in Robin & Crézé (1986a) and Bienaymé et al. (1987).

We classify the model simulated sample into three groups, - disk, thick disk, and halo stars. In Fig. 2a, the stars are plotted on  $(B - V, \mu_l)$  plane from the simulated catalogue with the asymmetric drift of the thick disk;  $-80 \text{ km s}^{-1}$ . In Table 8, we give the statistical properties of each observable  $(B - V, U - B, V, \mu_l, \mu_b)$  in each group.

In Fig. 2b, the stars are plotted on  $(B - V, \mu_l)$  plane from the observed data.

The ability of different models to represent the data is tested in the 5-dimensional space  $(B - V, U - B, V, \mu_l, \mu_b)$  of observables through the coincidence between real data and the simulated data. In Table 9, in Col. 1, we give several values of the asymmetric drift of the thick disk. The baricenter of  $\mu_l$  of the thick disk for the model ( $\mu_l^{\text{model}}$ ) is computed directly from the model simulated catalogue, while for the real data,  $\mu_l^{\text{data}}$  is computed from our fitting method (Eq. 5). In Col. 4, we give the statistical error and Col. 5 shows the difference between the model simulated baricenter and the observed baricenter. From this table, we can see that, when the asymmetric drift of the thick disk is chosen as  $-80 \text{ km s}^{-1}$ , the mean  $\mu_l^{\text{model}}$  for the thick disk group is in very good agreement with the observed data.

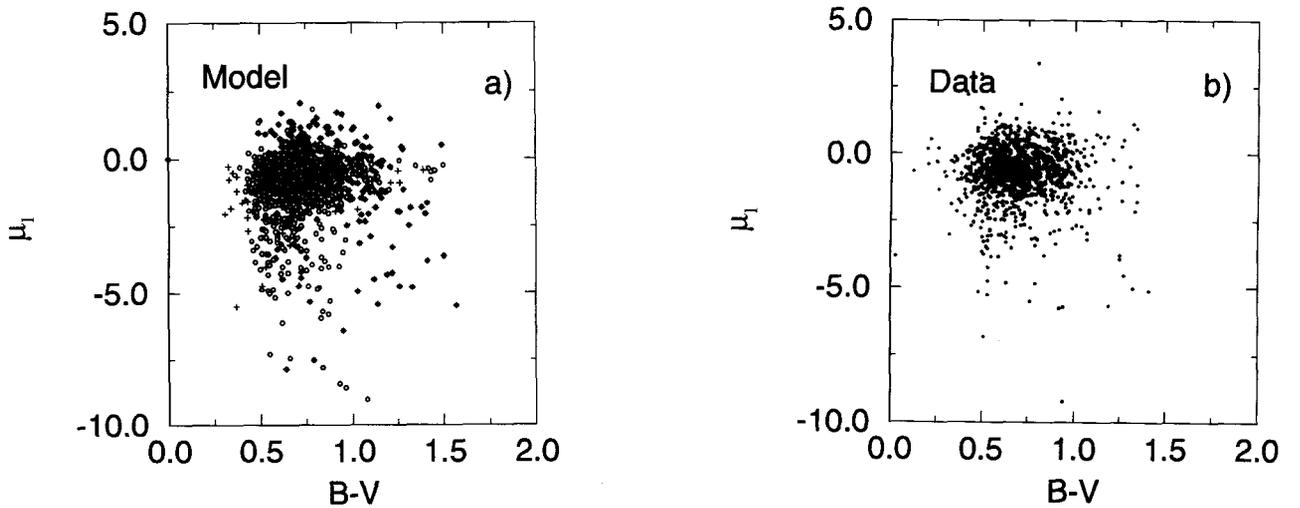
Controversial results for the thick disk asymmetric drift have been published in the literature. Norris (1987) found a value of  $-20 \text{ km s}^{-1}$ , Ratnatunga & Freeman (1989) and Carney et al. (1989)  $-30 \text{ km s}^{-1}$ , Sandage & Fouts (1987)  $-50 \text{ km s}^{-1}$ , Spaenhauer (1989)  $-80 \text{ km s}^{-1}$  at a distance of 2 kpc, Wyse & Gilmore (1986)  $-100 \text{ km s}^{-1}$ . Recently, Robin & Chen (1992) have used the M5 sample to investigate the circular velocity of the thick disk in the 5-dimensional space of observables by multiple discriminant analysis and found it to be  $-80 \pm 20 \text{ km s}^{-1}$  at a 2 sigmas level. Majewski (1992) concluded that the thick disk asymmetric drift varies linearly with distance from the Galactic plane, and has its most extreme value,  $\sim -120 \text{ km s}^{-1}$  at 5.5 kpc.

Since in the M5 sample, the mean distance of the thick disk is about 1.5 to 2 kpc above the plane, we find that our result is in very good agreement with Robin & Chen's result ( $-80 \pm 20 \text{ km s}^{-1}$  at 2 sigma level) and Spaenhauer's result, who found a drift of  $-80 \text{ km s}^{-1}$  at a distance of  $z = 2$  kpc.

The method described in this paper can not only determine the physical parameters but also derive the relative densities of different components and populations in the Galaxy. In Table 10, we show the results from the least squares solution (Eq. 3). The estimated unknown parameters ( $\alpha_{\omega_m}$ ), - the ratio of number of stars in the model to the observed data in each group-, are shown at the bottom line of the matrix. The values on-diagonal are the error on the estimated coefficients ( $\epsilon_{\omega_m}$ ), the values off-diagonal give the correlation coefficient among the groups. From Table 10, we find that the model can predict correct disk stars, and it cannot predict enough thick disk stars (by 35%). However, because the thick disk is not well separated from the halo star (the correlation coefficient is 0.84) in this sample, the observational constraints are limited.

**Table 8.** The means and standard deviations (between brackets) of the observational parameters- the colours ( $B - V$ ,  $U - B$ ), magnitude  $V$  and proper motions ( $\mu_l$ ,  $\mu_b$ ) for disk, thick disk and halo from the simulated M5 catalogue.  $N$  is the number of stars in the group

Group	$N$	Name	$B - V$	$U - B$	$V$	$\mu_l$	$\mu_b$
Group 1	719	disk	0.73 (0.19)	0.52 (0.35)	14.97 (1.30)	-0.56 (1.53)	0.13 (1.49)
Group 2	377	thick disk	0.73 (0.20)	0.47 (0.39)	15.75 (1.15)	-0.98 (1.41)	0.05 (0.98)
Group 3	115	halo	(0.67) (0.21)	(0.30) (0.43)	(15.95) (1.13)	(-1.22) (1.37)	(0.04) (0.78)



**Fig. 2.** The stars are plotted on  $(B - V, \mu_l)$  plane. a) Model predictions, star: disk; open circle: thick disk; plus: halo. b) The observed data

## 5. Conclusions

Star count survey provides a promising opportunity to understand the fundamental features of the large-scale structure of the Galaxy. The extraction of the galactic structure parameters requires a global analysis including magnitudes, colours and proper motions in an  $n$ -dimensional space of observables (Chen et al. 1992). In this paper, we have developed a multivariate star count algorithm including non-parameter kernel estimation in pattern recognition. The method allows us to determine the physical parameters and to estimate the density of each stellar population accurately.

We have tested the viability of the method by a series of Monte Carlo simulations. Results show that the methodology is powerful and significant.

When the model is basically correct (Sects. 3.1-3.3), the fitting procedure is oversimplified. However, in most cases, there are some systematic shifts between the model catalogue and the observed data. From the Monte Carlo

simulations in this investigation, we stress that any systematic error can lead to serious errors in the determination of the relative density of each physical group. Therefore, we have applied a stepwise approach. A first comparison of the multivariate densities of the model (defined by a sum of populations or physical groups) and the data allows us to detect systematic shifts between the group characteristics and the observed distributions. Then, applying these shifts to the model, the same computation allows us to measure the relative density of each physical group. Because the problem addressed is universal, this method may be applied to other astronomical fields, involving multivariate data (Chen et al. 1994).

An example of application to a real star count sample (M5 sample) has been presented; it shows how the method can be used to constrain the galactic structure parameters and derive the relative density of each stellar population. A grid of model simulated catalogues with different asymmetric drifts of the thick disk has been compared with the

**Table 9.** The determination of the asymmetric drift of the thick disk: Models differ by their asymmetric drift for the thick disk (Col. 1). The Col. 2, the baricenter of  $\mu_l$  for the thick disk from the model simulated catalogue. The Col. 3, the baricenter of  $\mu_l$  for the thick disk from the real M5 sample. The Col. 4, the statistical uncertainty, and the Col. 5 shows the difference between the model simulated baricenter and the observed baricenter

$V_{asy}$ km s <sup>-1</sup>	$\mu_l^{model}$	$\mu_l^{data}$	error	note
-150	-1.624	-1.145	0.048	$\sim 10 \sigma$
-90	-1.076	-0.937	0.046	$> 3 \sigma$
-80	-0.849	-0.885	0.048	$< 0.8 \sigma$
-70	-0.739	-0.891	0.049	$> 3 \sigma$

**Table 10.** The results from the least square solution. The estimated unknown parameters are shown at the bottom line of the matrix. The values on-diagonal are the error on the estimated coefficients ( $\epsilon_{\omega_m}$ ), the values off-diagonal give the correlation coefficient among the groups

Disk	Thick Disk	Halo
0.03		
-0.35	0.06	
-0.04	-0.84	0.38
0.98	1.35	0.73

observation. We found that the asymmetric drift of the thick disk to be  $80 \pm 10$  km s<sup>-1</sup> at a 3 sigma level, and the model cannot predict enough thick disk stars. However, because we only use the star count in one field, the observational constraints are limited.

When applying this method, both multivariate star count survey and high-quality data over large areas of sky are desired in order to give better separation between groups and reduce the statistical error. This can be achieved by comparing model predictions and data in a complete survey of several galactic directions simultaneously (Chen 1993). A detailed investigation from new observations in several different galactic directions will be published in a forthcoming paper.

*Acknowledgements.* I thank Drs. Michel Crézé and Annie Robin, who motivated me to use multivariate data analysis to interpret star count survey and gave me many useful suggestions; without their help, this work would not have been possible. I wish to acknowledge the warm hospitality of the CDS, Observatoires de Strasbourg and Besançon, where this research was started. I thank Dr. Olivier Bienaymé, who provided me with his M5 sample. Many thanks to Drs. J. Torra and F. Figueras for some stimulating discussions. The remarks

of the referee, R. Buser, which helped improve the presentation, are gratefully acknowledged. This work is supported by the Generalitat de Catalunya (PIEC), CICYT under contract ESP94-1133-E, and DGICYT under contract PB91-0857.

## References

- Bahcall J.N., Soneira R.M., 1980, ApJS 47, 357  
 Bahcall J.N., 1986, ARA&A 24, 577  
 Becker W., 1980, A&A 87, 80  
 Bienaymé O., Mohan V., Crézé M., Considère S., Robin A.C., 1992, A&A 253, 389  
 Bienaymé O., Robin A.C., Crézé M., 1987, A&A 180, 94  
 Bok B.J., 1937, The distribution of stars in space. University of Chicago Press  
 Buser R., Kaeser U., 1985, A&A 145, 1  
 Buser R., Rong J., 1995, Baltic Astron. 4, 1-24  
 Cabrera-Cano J., Alfaro, E.J., 1990, A&A 235, 94,  
 Carney B.W., Latham D.W., Laird J.B., 1989, AJ 97, 423  
 Chen B., 1993, Ph.D thesis, CDS, Université de Strasbourg  
 Chen B., 1996, A&A 306, 733  
 Chen B., Crézé M., Robin A.C., Bienaymé O., 1992, 2<sup>nd</sup> Annual Conference on astronomical Data Analysis Software and systems. Boston, U.S.A., 489  
 Chen B., Figueras F., Torra J., Asiain R., Jordi C., IAU Symp. 169, 1994  
 Chiu L.T.G., 1980, ApJS 44, 31  
 Crézé M., Chen B., Robin A.C., Bienaymé O., 1991, The Stellar Populations of Galaxies, IAU Symp. 149, 410  
 Gilmore G., 1984, MNRAS 207, 223  
 Gilmore G., Reid N., 1983, MNRAS 202, 1025  
 Hand D.J., Discrimination and classification. John Wiley & Sons Ltd. 1981  
 Kapteyn J.C., van Rhijn P.J., 1920, ApJ 52, 23  
 Majewski S.R., 1992, ApJS 78, 87  
 Norris J., 1987, AJ 93, 616  
 Ojha D.K., Bienaymé O., Robin A.C., Mohan V., 1994a, A&A 284, 810  
 Ojha D.K., Bienaymé O., Robin A.C., Mohan V., 1994b, A&A 290, 771  
 Pritchett C., 1983, AJ 88, 1476  
 Ratnatunga K.U., Freeman K.G., 1989, ApJ 339, 126  
 Reid N.I., Majewski S.R., 1993, ApJ 409, 635  
 Robin A.C., Chen B., 1992, Proc. of the third annual October Astrophys. Conf. in Maryland College Park, Maryland  
 Robin A.C., Crézé M., 1986a, A&A 157,71  
 Robin A.C., Crézé M., 1986b, A&AS 64, 53  
 Robin A.C., Crézé M., Bienaymé O., Oblak E., 1989, Proc. of a Conference held at Danbury. In: Philip A.G.D. and Lu P.K. (eds.). Davis press, 33  
 Sandage A., Fouts G., 1987, AJ 92, 74  
 Silverman B.W., 1986, Density estimation for statistics and data analysis, J.W. Arrowsmith Ltd, Bristol  
 Soubiran C., 1992, A&A 259, 394  
 Spaenhauer A., 1989, The Gravitational Force Perpendicular to the Galactic Plane. In: Philip A.G.D. and Lu P.K. (eds.), p. 45  
 Wyse R.F.G., Gilmore G., 1986, AJ 91, 855