SIAM/ASA J. Uncertainty Quantification
Vol. 0, pp. 000–000

Minimum Distance Estimation of Milky Way Model Parameters and Related Inference[*]

Sourabh Banerjee[†], Ayanendranath Basu[‡], Sourabh Bhattacharya[‡], Smarajit Bose[‡], Dalia Chakrabarty[§], and Soumendu Sundar Mukherjee[¶]

Abstract. We propose a method to estimate the location of the Sun in the disk of the Milky Way using a method based on the Hellinger distance and construct confidence sets on our estimate of the unknown location using a bootstrap-based method. Assuming the Galactic disk to be two-dimensional, the sought solar location then reduces to the radial distance separating the Sun from the Galactic center and the angular separation of the Galactic center to Sun line, from a pre-fixed line on the disk. On astronomical scales, the unknown solar location is equivalent to the location of us earthlings who observe the velocities of a sample of stars in the neighborhood of the Sun. This unknown location is estimated by undertaking pairwise comparisons of the estimated density of the observed set of velocities of the sampled stars, with the density estimated using synthetic stellar velocity data sets generated at chosen locations in the Milky Way disk. The synthetic data sets are generated at a number of locations that we choose from within a constructed grid, at four different base astrophysical models of the Galaxy. Thus, we work with one observed stellar velocity data and four distinct sets of simulated data comprising a number of synthetic velocity data vectors, each generated at a chosen location. For a given base astrophysical model that gives rise to one such simulated data set, the chosen location within our constructed grid at which the estimated density of the generated synthetic data best matches the density of the observed data is used as an estimate for the location at which the observed data was realized. In other words, the chosen location corresponding to the highest match offers an estimate of the solar coordinates in the Milky Way disk. The "match" between the pair of estimated densities is parameterized by the affinity measure based on the familiar Hellinger distance. We perform a novel cross-validation procedure to establish a desirable "consistency" property of the proposed method.

Key words. Milky Way, Hellinger distance, density estimation, confidence sets, cross-validation

AMS subject classifications. Primary, 62P35, 85A35; Secondary, 65C60, 85A05, 85A15

DOI. 10.1137/13093552RR

1. Introduction and background.

The learning of structure in the space of parameters of a system, using available data, is an exercise that has gained increasing attention in the recent past. This includes attempts at finding intervariable relationships in large data using developed methods of scoring the association (Reshef et al., 2011), in graphical model contexts

[*] Received by the editors September 3, 2013; accepted for publication (in revised form) August 11, 2014; published electronically DATE.
http://www.siam.org/journals/juq/x-x/93552.html

[†] Department of Statistics, University of Illinois at Urbana-Champaign, Champaign, IL 61820 (sbanerj7@illinois.edu).

[‡] Bayesian and Interdisciplinary Research Unit, Indian Statistical Institute, 203, B. T. Road, Kolkata 700108, India (ayanbasu@isical.ac.in, sourabh@isical.ac.in, smarajit@isical.ac.in).

[§] Department of Statistics, University of Warwick, Coventry CV4 7AL, UK, and Department of Mathematics, University of Leicester, Leicester LE1 7RH, UK (d.chakrabarty@warwick.ac.uk, dc252@le.ac.uk).

[¶] Department of Statistics, University of California, Berkeley, Berkeley, CA 94720-3860 (soumendu@berkeley.edu).

(Heckerman, Geiger, and Chickering, 1995; Ghahramani, 2003), by searching for chosen features within the data (Lee, Pedersen, and Mumford, 2003; Zomorodian and Carlsson, 2005), by developing high-dimensional regression models in regression frameworks characterized by the number of covariates far exceeding the number of responses (Yuan and Lin, 2007; Simon, Friedman, and Hastie, 2012), and by developing density-based distances within the paradigm of semisupervised or unsupervised learning (Bijral, Ratliff, and Srebro, 2012; Orlitsky et al., 2005; Weinberger and Saul, 2006).

Indeed, unsupervised learning is often the relevant framework in real-world problems. However, within the framework of supervised learning, the aim is to predict values of the response variable $\mathbf{Y}$ corresponding to a given set of predictor variables $\mathbf{X}$, given the training sample $(\mathbf{x}_1, \mathbf{y}_1^{(\star)})$, $(\mathbf{x}_2, \mathbf{y}_2^{(\star)})$, $\ldots$, $(\mathbf{x}_n, \mathbf{y}_n^{(\star)})$. Here $\mathbf{y}_i^{(\star)}$ is a known value of $\mathbf{Y}$ at a chosen value $\mathbf{x}_i$ of $\mathbf{X}$. Here, the ith such chosen $\mathbf{x}_i$ is the ith design vector. The aim is to learn the model parameters that minimize the expected loss at each $\mathbf{x}$, where the loss function is appropriately chosen to embody the error in the estimation of the value of the response variable (Hastie, Tibshirani, and Friedman, 2001). The probability density of the response variable $\mathbf{Y}$, conditional on $\mathbf{X}$, is considered with the aim of learning the unknown model parameters. In contrast, in the framework of unsupervised learning, the joint probability density of the observations is examined, with the aim of making inference on the model parameters.

In this paper, we present a novel application in which the aim is to perform estimation of the unknown model parameters by comparing the conditional density of synthetic values of the response variable given a chosen set of predictor values with that of the measured values of the response variable given the same predictor set. Though this resonates with the supervised learning scheme, there are some features of this implementation that mark it as atypical in terms of a supervised learning scheme. First, here the response variable $\mathbf{Y}$ is a matrix; it is more the case in unsupervised learning that $\mathbf{Y}$ is a high-dimensional variable. Second, in this work, the loss function is itself defined in terms of the distance between the two aforementioned conditional probability density functions; the $\mathbf{x}$ for which this distance is minimized gives the unknown model parameter. Third, the density functions in question are not known to begin with but are estimated using kernel density estimation techniques. In fact, the estimated densities are found to be highly multimodal as well as sparse. The efficiency of this learning may, however, be compromised if the chosen minimum distance procedure is not robust against violations of the usual model assumptions (Basu, Shioya, and Park, 2011).

In particular, we invoke an affinity measure based on the Hellinger distance, between the densities that the observed data and the synthetic data are sampled from. The motivating idea in this work is that the synthetic data sets are realizations of simulations of the system under a variety of given values of the model parameter vector. Thus, the particular synthetic data set that maximizes the affinity between the said densities is the realization obtained from the model parameter value that corresponds best to the true value; the “true” value of the model parameter indicates the value which suitably describes the observations. Maximization of the affinity in this context is equivalent to the minimization of the Hellinger distance.

One fundamentally important aspect of statistical learning is to perform model selection (Kohavi et al., 1995; Kearns et al., 1997) and importantly to quantify accuracy of a given model, using available data (Last, 2006). It is in principle possible to extend parameter estimation using minimized Hellinger distance to higher dimensions (Tamura and Boos, 1986).

The accompanying parameter uncertainty estimation is possible by constructing a high dimensional confidence set within the region of interest, as distinguished from a product of confidence intervals of interest along each dimension. In our application we seek a similarly constructed confidence set on our estimate of the unknown parameters using a bootstrap-based method. It is also of vital importance to ensure generalization of the learned model to an independent data set and this is achieved using cross-validation techniques (Efron and Tibshirani, 1997). We include such validation of our learned model parameters by adopting a cross-validation technique where assuming a particular location as the true location we verify whether they are accurately estimated by the proposed method.

The paper is organized as follows. In section 2 we describe the experimental setup under which the data are generated. Some discussion of the existing literature related to this problem is presented in section 3. The method we advocate is described in section 4. Section 5 contains the results of our analysis. In particular, section 5.4 discusses the bootstrap-based method that we use to construct a confidence set on our estimation of the Milky Way parameters, and in section 5.5 we present our implementation of cross-validation. Finally, section 6 provides some concluding remarks.

2. The experimental setup. In this application, the system under consideration is the disk of the Milky Way that is assumed to be two-dimensional. The observed data comprise the $N \times 2$-dimensional matrix $\mathbf{Y} = (\mathbf{y}_1 : \mathbf{y}_2 : \ldots : \mathbf{y}_N)^T$, where $\mathbf{y}_j$ is a two-dimensional velocity vector, $j = 1, 2, \ldots, N$. Thus $\mathbf{Y}$ represents the two-component velocity vector measurement of N stars that were observed close to the Sun in our galaxy (Fux, 2001). For this astronomical observational data set, we have $N = 3500$.

Such a matrix of these velocity measurements is realized at location $\mathbf{X}$ of the observer who measures the velocities of these N stars. Nonlinear dynamical simulations of the Milky Way disk was performed by Chakrabarty (2007) by varying this physical location $\mathbf{X}$. We place the two-dimensional Milky Way disk on a two-dimensional polar coordinate system such that the spatial location vector $\mathbf{X}$ is given by the radial distance R from the defined center of this coordinate system (chosen to coincide with the center of the Milky Way disk) and the azimuthal or angular displacement θ (where $\theta = 0$ is chosen to be along the long axis of a feature in the Milky Way, namely, the central bar in the Galaxy). Thus, the value of $\mathbf{X}$ in a two-dimensional orthogonal basis is $\mathbf{x} = (r\cos\theta, r\sin\theta)^T$. In fact, in our work, it is this physical location $\mathbf{X}$ of the observer on the two-dimensional Milky Way disk that we want to learn. Thus, in this setup, what we referred to as our "unknown model parameters" in the introductory section concurs with the unknown physical location of the observer. We would like to emphasize that hereafter, "location" refers to the address of the observer on the Milky Way disk parameterized by $\mathbf{X}$. According to our model, the observed velocity matrix $\mathbf{Y}^{(obsvd)}$ corresponds to an unknown value of the location, i.e., at $\mathbf{X} = \mathbf{x}_\star = (r_\star \cos\theta_\star, r_\star \sin\theta_\star)^T$.

Thus, the identification of $\mathbf{x}_\star$ is equivalent to identifying the radial location $r_\star$ of the observer from the center of the Galaxy and the angular location (separation) $\theta_\star$ of the observer from a chosen axis in the Galaxy, such that if from this location in the model Milky Way, the observer had tracked the stars in the neighborhood of the Sun for their velocity vectors, the collected data would have been "closest" to the observed data $\mathbf{Y}^{(obsvd)}$; here the aforementioned "closeness" is in the sense implied by our affinity measure (see section 4.3). Now, the

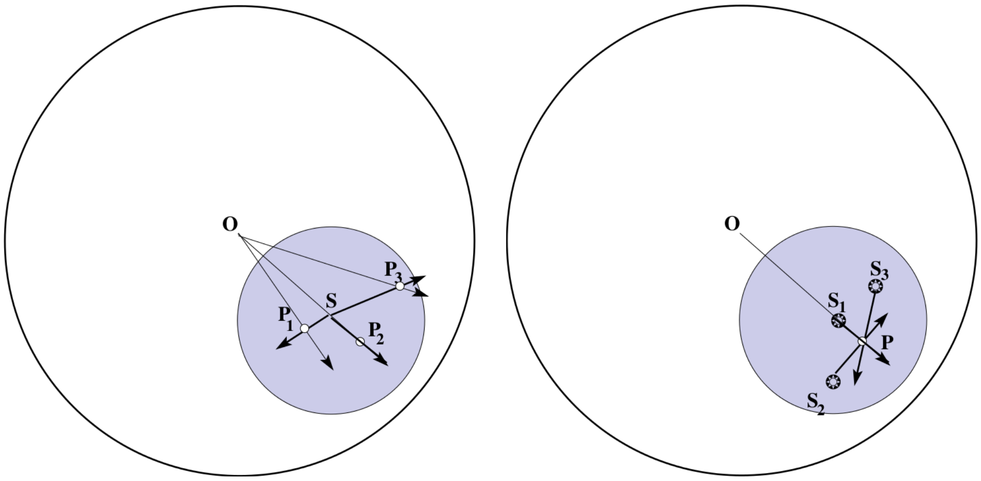


Figure 1. Left: *Figure showing locations of* 3 *of the sampled stars from center of circular patch (gray circle with center at location of the Sun—depicted at point* $\mathbf{S}$*) within which stars are sampled uniformly. The location and velocity vectors of these sampled stars are recorded by an observer at the Sun. However, the locations of these sampled stars with respect to the center of the Galaxy (at* $\mathbf{O}$*) are unknown. Thus, a measured heliocentric stellar location vector cannot constrain the unknown location vector of the Sun with respect to the center of the Galaxy (the vector* $\mathbf{OS}$*). Neither can the distribution of the measured heliocentric stellar locations constrain the unknown* $\mathbf{OS}$ *since the sample mean of the measured heliocentric stellar locations is zero.* Right: *Velocity* $\mathbf{V}$ *of an example star (at* $\mathbf{P}$*) along* $\mathbf{OP}$ *is viewed by observer at point* $\mathbf{S_1}$ *to lie entirely along the line that joins this observer to the star while the observer at point* $\mathbf{S_2}$ *views this stellar velocity to be entirely orthogonal to her line of sight to the star at* $\mathbf{P}$*. The velocity vector of a sampled star, as measured by an observer, is expressed to comprise a radial component that is along the line of sight of the observer to the star, and a transverse component that is orthogonal to this line of sight. Thus, the observer at* $\mathbf{S_1}$ *records the stellar velocity to be* $(V, 0)^T$*, while observer at* $\mathbf{S_2}m$ *records the stellar velocity to be* $(0, V)^T$*. The observer at* $\mathbf{S_3}$*, however, records the stellar velocity to have nonzero radial and transverse components. Thus, the set of velocities measured by an observer potentially bears information about the observer's location in the Galaxy.*

location of the observer is really *our* location as earthlings on the Galactic disk, i.e., seeking $(r_\star \cos\theta_\star, r_\star \sin\theta_\star)^T$ is the same as trying to estimate the location of the Sun in the Milky Way disk.[1]

It may be questioned why the velocity data—observed or synthetic—alone are invoked to help learn the unknown model parameter vector $\mathbf{X}$. Indeed, the data includes information on the spatial location of the stars as well as the velocity of the stars, but of these, only the velocity data can be implemented in the estimation of $\mathbf{X}$. This is understood by consulting Figure 1. The stars that are tracked for their locations and velocities live in a circular patch in the neighborhood of the Sun in the two-dimensional Milky Way disk; thus, the center of this circular patch is at the location of the Sun and the radius of this patch is small (ϵ) compared to $\|\mathbf{X}\|$, where $\|\cdot\|$ is the Euclidean norm of a vector. It is noted that the spatial location vector $\mathbf{p}_k$ of the kth star and its velocity vector $\mathbf{y}_k$, are as recorded by the observer seated at the Sun; $k = 1, 2, \ldots, N$, where N stars constitute a data set. Here $\mathbf{p}_k = (s_k \cos\alpha_k, s_k \sin\alpha_k)^T$, where s_k is the radial location of the kth sampled star, as recorded by the heliocentric observer and

[1] On Galactic length scales, the location of us, i.e., the Earth in the Galaxy, is very well approximated by the location of the Sun in the Galaxy.

α_k is the angular displacement from a chosen line. The sampling of the spatial locations of the stars is such that s_k is uniform in the interval $[0, \epsilon]$ and α_k is uniform in $[0, 2\pi]$. Then the mean of $s_k \cos \alpha_k$ over all k is zero, as is the mean of $s_k \sin \alpha_k$, i.e., the sample mean of the measured $\mathbf{p}_k$ is zero.

The left panel of Figure 1 shows the location vectors of three example stars at points $\mathbf{P_1}, \mathbf{P_2}$, and $\mathbf{P_3}$ inside this circular patch (marked in gray) where the location of the observer (Sun) is at point $\mathbf{S}$ and that of the center of the Galaxy is at point $\mathbf{O}$ in the Milky Way disk. Then, in reference to this figure, the heliocentric location to the jth of these example stars is the vector $\mathbf{SP_j} = \mathbf{OP_j} - \mathbf{OS}$, $j = 1, 2, 3$, in the figure. But the galactocentric location $\mathbf{OP_j}$ to the star is unknown, implying that the measured heliocentric location $\mathbf{SP_j}$ cannot be used in this equation to constrain the location of the observer with respect to the center of the Galaxy, i.e., the unknown model parameter vector $\mathbf{X}$ that we are after (or $\mathbf{OS}$ in reference to this figure). Again, the uniform distribution of s and α suggests that the mean of the recorded stellar location vectors is zero so that the unknown $\mathbf{X}$ is the mean of the galactocentric locations of the sampled stars, which, however, is unknown. In other words, no matter what the galactocentric location of the Sun is, the average of the measured heliocentric locations of the sampled stars is identically zero; these heliocentric location measurements do not offer any information about $\mathbf{X}$.

On the other hand, as is depicted in the right panel of Figure 1, the velocity (vector) of the sampled star at point $\mathbf{P}$ as measured by an observer at point $\mathbf{S_1}$ is distinct from that measured by the observer at points $\mathbf{S_2}$ and $\mathbf{S_3}$ on the Milky Way disk. Here, the velocity of the star measured by the observer at point $\mathbf{S_j}$ is considered to have a radial component along the line $\mathbf{S_jP}$ that joins the observer to the star, and the transverse component is orthogonal to this line; $j = 1, 2, 3$ in this figure. Thus, we see in this panel that the velocity vector $\mathbf{V}$—that is, along $\mathbf{OP}$—of an example sampled star at $\mathbf{P}$, will appear to be entirely along the line $\mathbf{S_1P}$ and entirely orthogonal to line $\mathbf{S_2P}$, so that its velocity as measured by the observer at point $\mathbf{S_1}$ will be recorded as $(\|\mathbf{V}\|, 0)^T$ while the observer at point $\mathbf{S_2}$ will record its velocity as $(0, \|\mathbf{V}\|)^T$. The observer at point $\mathbf{S_3}$ will record the velocity of the star to have nonzero radial and transverse components. Then, a data set that comprises stellar velocities as recorded by an observer in the Milky Way disk bears information about the location of this observer, i.e., about $\mathbf{X}$. Thus, such a velocity data set can be inverted to help estimate the location of the observer, i.e., the Sun.

The radial units used by Chakrabarty (2007) are motivated by the physics of interaction of the stars in the model Milky Way disk and one of the most conspicuous features in the Galaxy, namely, the elongated stellar bar that rotates with its own rotational frequency Ω_b, pivoted at the center of the Galaxy. The radius at which the (radius-dependent) rotational frequency of the stars in the Milky Way disk equals Ω_b is called the co-rotation radius or R_{CR} of that model of the Galaxy. The radial unit used in our work is equivalent to $1R_{CR}$ for the choices of the Milky Way astrophysical model and Ω_b used by Chakrabarty (2007).

Chakrabarty (2007) motivates the observer radial location variable to lie in the interval $[1.7, 2.3]$ radial units and the observer angular location variable θ to lie in $[0°, 90°]$ on the basis of the relevant physics. These intervals are discretized with $N_R = 24$ different values of the radial location and $N_\theta = 9$ values of the angular location. The left edge of the radial bin is $r_0 = 1.7$ radial units, that of the angular bin is $\theta_0 = 0°$, the radial bin width is $\delta_r = 0.025$

radial units, and the angular bin width is $\delta_\theta = 10°$. Thus, the kth radial bin is said to be centered around $1.7 + (k-1)\delta_r + \delta_r/2$, $k = 1, 2, \ldots, N_R$. Similarly, the jth angular bin is centered around $(j-1)\delta_\theta + \delta_\theta/2$, $j = 1, 2, \ldots, N_\theta$. As described above, all radial distances expressed here are in units of R_{CR} and all angles in units of degrees.

The simulations carried out by Chakrabarty (2007) correspond to variation over the values of the location vector $\mathbf{X}$, the components of which are the two components of the spatial location vector of the observer on the Milky Way disk. In these simulations, $\mathbf{X}$ is chosen to take values $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_d$ such that at $\mathbf{X} = \mathbf{x}_i$, the simulated synthetic velocity matrix is $\mathbf{Y}_i^{(sim)}$; here $i = 1, 2, \ldots, d$. In these simulations, d was chosen to be 216. Now, $\mathbf{x}_i = (r_i \cos\theta_i, r_i \sin\theta_i)^T$, where the ith value of the observer radial location is r_i and that of the observer angular location is θ_i.

We consider the stellar location and velocity coordinates simulated from a model of the Milky Way and for each $i \in \{1, 2, \ldots, d\}$ identify the N_i stars that have location vectors such that these stars lie within a neighborhood of the ith proposal for the solar location, i.e., in a neighborhood of bx_i. Here, the size of the neighborhood is chosen to mimic the extent of the circular patch of radius ϵ centered at the Sun, from within which the real stars are sampled, to generate the data set $\mathbf{Y}^{(obsvd)}$. The ith such neighborhood then defines the intersection of the kth radial bin and the jth angular bin; $i = N_\theta(k-1) + j$ and in the simulations performed by Chakrabarty (2007), ϵ motivated δ_θ and δ_r via the suggestion that $\pi\epsilon^2$ is roughly approximated by the area of the intersection of a radial and an angular bin. The velocity vectors of these N_i stars are then implemented to estimate the density function from which the discrete $\mathbf{Y}_i^{(sim)}$ are sampled. This is repeated for each $i \in \{1, 2, \ldots, d\}$. A density function is also estimated from the real velocity data $\mathbf{Y}^{(obsvd)}$. Pairwise comparison of this density is undertaken with the density estimated using $\mathbf{Y}^{(sim)}$. The comparison is parametrized by an affinity parameter (see section 4).

It merits mention that a set of the synthetic velocity data matrices $\mathbf{Y}_i^{(sim)}$, $i = 1, \ldots, d$, is obtained with nonlinear dynamical simulations of one of four different base astrophysical models of the Milky Way. However, we do not include any reference in the notation to the base astrophysical model that the corresponding synthetic data set is generated from, as we perform the analysis with each such set of synthetic data, one at a time. Along with the estimation of the observer, i.e., the solar location in the Milky Way, our investigation aims to determine which of the four astrophysical models best explains the observed data.

So to summarize, if $\mathbf{X} = \mathbf{x}_\star$ represents the location where the estimated density of the simulated synthetic data has the maximum affinity with the estimated density of the observed data, our inference chooses the estimate of the unknown model parameter vector to be $\mathbf{x}_\star$. We now begin discussion of the details of this inference that is based on distances between the estimated density of the observed velocity data at the unknown location and the estimated density of the synthetic velocity data generated at a chosen value of $\mathbf{X}$. Along the way, we will also develop first the motivation and then the methodology used to implement validation.

3. Literature review.

The squared Hellinger distance is one of the most popular measures used in robust minimum distance inference and has a one-to-one relationship with the Bhattacharyya distance (Bhattacharyya, 1943); the Hellinger affinity is also referred to as the Bhattacharyya coefficient. The technical definitions of the distance measures, affinities, and coefficients are given in the subsequent sections. Although it does not satisfy the triangle inequality, the Bhattacharyya distance is nonnegative and equals zero if and only if the component densities are identically equal. See Kailath (1967), Djouadi, Snorrason, and Garber (1990), and Aherne, Thacker, and Rockett (1998), among others, for some useful applications of the Bhattacharyya distance in real-life problems. The Hellinger distance is also referred to as the Matusita distance (Matusita, 1953; Kirmani, 1971) or the Jeffreys–Matusita distance in the literature. Both the Bhattacharyya distance and the Matusita (Hellinger) distance (or the corresponding affinities) are extensively used as measures of separation between probability densities in many practical problems such as remote sensing (Landgrebe, 2003; Canty, 2007).

A method for estimating the solar location in the Milky Way disk, as proposed by Chakrabarty (2007) involved performing a d number of tests to test for the null that the observed data is sampled from the density estimated using the ith synthetic data set $\mathbf{Y}_i^{(sim)}$ which is generated at the ith value of the chosen location, i.e., at $\mathbf{x}_i$, where $i = 1, 2, \ldots, d$. The chosen location at which the p-value of the test statistic (employed by Chakrabarty (2007)) is maximized is considered an estimate of the unknown location at which the observed data is realized, i.e., the solar location. In this work, however, our approach is different as we attempt a direct comparison of the density estimated using the synthetic data $\mathbf{Y}_i^{(sim)}$ and that estimated using the observed data $\mathbf{Y}^{(obsvd)}$, $i = 1, 2, \ldots, d$. Then $\mathbf{x}_i$ is our estimated solar location where the closeness of the comparison is quantified by the Hellinger distance. Thus, our work represents an application of the Hellinger distance measure.

4. Proposed method.

4.1. Motivation. Since simulated velocity data at each of the d different chosen locations are available, the velocity densities at each such point can be estimated. This we have achieved by fitting a standard bivariate Gaussian kernel. Subsequently, we have calculated affinity measures of each of these densities with the estimated density of the observed velocity data. The affinity measures so obtained have then been maximized over the (r_i, θ_i) grid points to derive the estimate of the true location from which the observed data may have been generated.

Many choices of density-based distances (more generally divergences) are available in the literature. Depending on their choice, the distances can exhibit very different characteristics. See Basu, Shioya, and Park (2011) for a comprehensive description of the topic of density-based distances and their use in statistical theory. In this particular work we have chosen to use the affinity measure based on the Hellinger distance. This affinity measure, also linked to the Bhattacharya distance, takes the value 1 when the densities coincide and takes the value 0 when they are singular (i.e., their supports are nonoverlapping). We will give a very brief introduction to density-based distances in section 4.3.

4.2. Novelty of the density-based method. The approach that we adopt in this paper has, in our opinion, the following advantages to distinguish itself. First of all, we feel that this is a more natural approach to identifying the unknown location compared to the p-value approach (Chakrabarty, 2007). The p-value approach for finding the location depends on repeated generation of data from the physical system or the estimated velocity distribution for the particular location to create estimates of the Kullback–Leibler divergence (KLD) necessary for the generation of the estimated quantiles for the construction of the p-value. We take the view that since the estimated density for the location is already available, this is unnecessary. This also greatly reduces the computational burden of the procedure.

Another issue, which has not been sufficiently addressed in the previous approaches dealing with this problem, is the issue of revalidation of the procedure. Would this method of finding the maximum of the affinity over the different grid points be "consistent" in the sense that should the optimal data generating location be removed from the data, would the maximum of the affinity be obtained at one of its immediate neighbors? Essentially we are demanding a continuity property for the affinity surface over the grids in question. We will observe later in the article that in most situations this is indeed the case for the affinity measure based on the Hellinger distance, giving us the confidence that the determination of the location based on the affinity measure is doing the appropriate thing.

There is another point in favor of the particular approach chosen here. The different models used for the description of the velocity data set are, after all, only abstractions of reality. While we expect that these models will satisfactorily explain the pattern of the majority of the data, there is always the chance (in fact it is practically expected) that there could be small subsets of the data which would not follow the pattern dictated by the bigger majority. In such situations, the Hellinger distance is a more dependable measure for identifying the model which fits the large majority of the data, sacrificing a small group of outlying observations (see, e.g., Basu, Shioya, and Park (2011)). The same is not true for the version of the KL divergence which generates the p-values for the likelihood-based method.

Another approach that has been advanced to learn $\mathbf{x}_\star$ is independent of density estimation (Chakrabarty, Biswas, and Bhattacharya, 2013). In this method, the velocity data is expressed as a function of the solar location vector $\mathbf{X}$ and this unknown function is modeled with a Gaussian process. The posterior probability density of $\mathbf{x}_\star$ given the simulated and observed data is computed in this Bayesian approach. We compare our results to those obtained by Chakrabarty, Biswas, and Bhattacharya (2013).

4.3. Distance methods.

4.3.1. Affinity measure based on the Hellinger distance. Let f and g be two probability density functions with respect to the Lebesgue measure (or any other appropriate measure). Then the squared Hellinger distance $\mathrm{HD}(g, f)$ between the densities g and f is defined as

$$\mathrm{HD}(g,f) = \int \left(g^{\frac{1}{2}}(x) - f^{\frac{1}{2}}(x)\right)^2 dx. \tag{4.1}$$

The Hellinger distance is one of the few genuine metrics in the large class of density-based divergences widely used in statistics. The measure HD is bounded from above by 2, a value which is attained when the densities are singular. Similarly, the lower bound of the measure is 0, obtained when the densities are identically equal. Notice that the measure in (4.1) may be represented as

$$\begin{aligned}\mathrm{HD}(g,f) &= \int g(x)dx + \int f(x)dx - 2\int g^{\frac{1}{2}}(x)f^{\frac{1}{2}}(x)dx \\ &= 2\left(1 - \int g^{\frac{1}{2}}(x)f^{\frac{1}{2}}(x)dx\right). \qquad (4.2)\end{aligned}$$

Thus the minimization of the Hellinger distance is equivalent to the maximization of the affinity measure

$$\rho(g, f) = \int g^{\frac{1}{2}}(x) f^{\frac{1}{2}}(x) dx \tag{4.3}$$

which varies between 0 and 1; the end points are obtained when the densities are singular and identical, respectively. The quantity in (4.3) is linked to the Bhattacharyya distance (Bhattacharyya, 1943)

$$B(g, f) = -\log\left(\int g^{\frac{1}{2}}(x) f^{\frac{1}{2}}(x) dx\right) \tag{4.4}$$

and is widely used as a measure of closeness between two probability densities.

4.3.2. The KLD. The KLD (also known as information divergence, information gain, or relative entropy) is a nonsymmetric measure of the difference between two probability distributions G and F (Kullback and Leibler, 1951). The distribution G typically represents the "true" distribution of the data, while the distribution F represents a theory, model, description, or approximation of G.

Although it is often intuited as a metric or distance, the KLD is not a true metric. In particular it is not a symmetric measure; the KLD between G and F is generally not the same as that between F and G. The divergence is computed between the corresponding densities g and f and is defined as

$$\delta(g, f) = \int g(x) \log\left(\frac{g(x)}{f(x)}\right) dx. \tag{4.5}$$

This divergence measure is not bounded above; however, a zero value of this measure indicates zero distance between f and g, i.e., the densities are identically equal. In spirit, the divergence measure can be considered to be similar to the inverse of the affinity measure. Both the KL and HD measures are special cases of the Cressie–Read family of power-divergences (Cressie and Read, 1984).

4.3.3. Relative Pearson divergence. The Pearson (PE) divergence is a squared-loss variant of the KLD. It is basically an extension of the Pearson's χ^2 divergence and is defined as

$$PE(g, f) = \int g(x) \left(\frac{f(x)}{g(x)} - 1\right)^2 dx. \tag{4.6}$$

It also belongs to the family of f-divergences and shares many theoretical properties of the KLD. This divergence measure is also not bounded above; a zero value of this measure indicates zero distance between f and g, i.e., the densities are identically equal.

The relative Pearson (rPE) divergence is a variant of the PE divergence (see, e.g., Sugiyama et al. (2013)). It is defined as

$$rPE(g, f) = PE(h_\alpha, f) = \int h_\alpha(x) \left(\frac{f(x)}{h_\alpha(x)} - 1\right)^2 dx, \tag{4.7}$$

where

$$h_\alpha(x) = \alpha f(x) + (1-\alpha) g(x) \quad \text{for} \quad 0 \leqslant \alpha < 1. \tag{4.8}$$

For $\alpha = 0$, the rPE divergence reduces to the normal PE divergence. However, the relative density ratio in this case, i.e., f/h_α, is bounded above by $1/\alpha$ for $\alpha > 0$:

$$\frac{f(x)}{h_\alpha(x)} = \frac{1}{\alpha + (1-\alpha)\frac{g(x)}{f(x)}} < \frac{1}{\alpha}.$$

Thus it overcomes the problem of the unboundedness of the density ratio f/g in the PE divergence. The tuning parameter α is chosen by cross-validation.

5. Results. The four astrophysical models will henceforth be referred to as 18sp3bar3, 25sp3bar3, sp3bar3, and bar6. This nomenclature involves the values of the bar and the spiral parameters which specify the models.

5.1. Density estimation. We use the bivariate kernel density estimation with the kernel $K(\cdot,\cdot)$ being the standard two-dimensional Gaussian kernel with covariance matrix $\mathbf{I}_2$, the two-dimensional identity matrix, i.e., the kernel function is given by

$$K(x,y) = \frac{1}{2\pi} \exp\left(-\frac{x^2+y^2}{2}\right). \tag{5.1}$$

Based on a set of n independent and identically distributed observations $(X_1, Y_1), \ldots,$ (X_n, Y_n) from the data generating density, our density estimate is given by

$$\hat{f}(x,y) = \frac{1}{nh^2} \sum_{i=1}^{n} K\left(\frac{x - X_i}{h}, \frac{y - Y_i}{h}\right), \tag{5.2}$$

where h is the smoothing parameter. We have chosen the smoothing parameter h as

$$h = \sigma n^{-\frac{1}{6}}, \tag{5.3}$$

where

$$\sigma^2 = \frac{s_X^2 + s_Y^2}{2}.$$

Here s_X^2 and s_Y^2 are the sample variances of the X and the Y observations, respectively. See, e.g., Silverman (1986) for a discussion on the choice of the smoothing parameter.

For a fixed model, let us denote the true density at the location (r, θ) under this model by $g_{(r,\theta)}(x,y)$ and its kernel density estimate by $\hat{g}_{(r,\theta)}(x,y)$. Also we shall denote the true density of the observed velocity data by $f(x,y)$ and its kernel density estimate by $\hat{f}(x,y)$. As the analysis for the simulated data generated at the $d = 216$ chosen locations each for each of the base astrophysical models is done separately, we do not attach another index for this base model to the density $g(\cdot,\cdot)$.

Here the observed velocity vectors are assumed to be independent and identically distributed. Such assumptions are generally reasonable and frequently employed in astronomical studies. See, e.g., Feigelson and Babu (2012) and Way et al. (2012).

5.2. Maximum affinity estimation of the location parameter. Here we present the results of the proposed method for each of the four simulation models. For a given model, let us define

$$\rho_{(r,\theta)} := \rho(g_{(r,\theta)}, f), \tag{5.4}$$

where $\rho(\cdot,\cdot)$ is the Hellinger affinity defined in (4.3). We compute the density estimate $\hat{f}$ for the observed data $\mathbf{Y}^{(obsvd)}$. We also compute the density $\hat{g}_{(r_i,\theta_i)}$ for the synthetic data $\mathbf{Y}_i^{(sim)}$ that is simulated at the ith chosen location $(r_i \cos\theta_i, r_i \sin\theta_i)^T$ of the $d = 216$ such chosen locations. We use the former and latter density estimates as surrogates for f and $g_{(r,\theta)}$, respectively. Thus for each base astrophysical model, we have 216 affinity values corresponding to each i; for brevity's sake, we use the notation

$$\hat{\rho}_{(r_i,\theta_i)} = \rho(\hat{g}_{(r_i,\theta_i)}, \hat{f}). \tag{5.5}$$

In Figure 2, we show the affinity surfaces generated over the chosen locations at which the simulated velocity data are generated in each of the base astrophysical models, i.e., the surface plot of $\hat{\rho}_{(r_i,\theta_i)}$ against (r_i, θ_i) for $i = 1, 2, \ldots, d$, $d = 216$, for each base model. The plot provides visualization of where the surfaces attain their maxima.

To avoid the dependency on perspective while viewing a surface plot, it sometimes helps to look at a more mundane contour plot. Figure 3 shows a contour plot of the affinity surface in grayscale. It is clear from these contour plots that in the bar6 model there is a single mode, while in the sp3bar3 model there are at least two pronounced modes. The other two models fall somewhere in between. This is in concert with the results obtained earlier by Chakrabarty and Sideris (2008) and by Chakrabarty, Biswas, and Bhattacharya (2013).

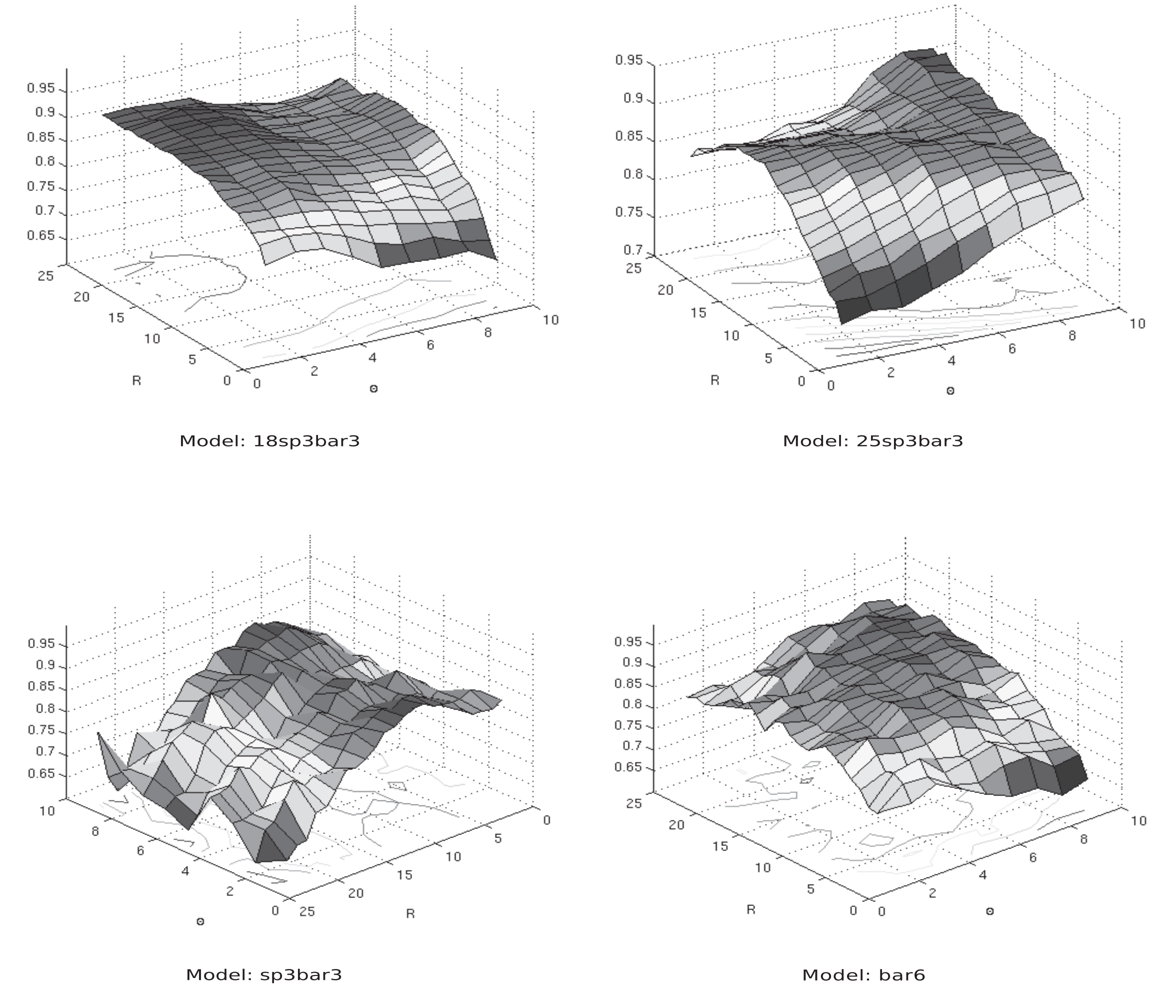
The chosen locations at which the synthetic data are simulated from the base astrophysical model in question are in fact arranged over a uniform rectangular grid. Thus, the ith point in this grid would represent the ith such chosen location, $i = 1, 2, \ldots, d$, $d = 216$, as per the nonlinear dynamical simulations of the Milky Way disk reported in Chakrabarty (2007). Taking advantage of the uniform nature of this grid, any grid point could have an alternative, two-dimensional representation, (k, j), $k = 1, 2, \ldots, 24$, $j = 1, \ldots, 9$, so that there are $24 \times 9 = 216$ such chosen locations. In this treatment, let the (k, j)th grid point be the physical location (r_k, θ_j).

For the base astrophysical model in question, we define $\max(k, j)$ as the indices for the particular chosen location where the affinity measure is maximized. Let the corresponding radial and angular coordinates of this location be

$$(r_{\max}, \theta_{\max}) := \arg\max_{(k,j)} \rho(g_{(r_k,\theta_j)}, f). \tag{5.6}$$

That is, $(r_{\max}, \theta_{\max})$ is the actual physical location where the true distribution of the synthetic data is closest to the true distribution of the observed data in the sense of having highest affinity, while $\max(k, j)$ represents the indices for this location. We have estimated $(r_{\max}, \theta_{\max})$ by

$$(\hat{r}_{\max}, \hat{\theta}_{\max}) = \arg\max_{(k,j)} \hat{\rho}_{(r_k,\theta_j)}, \tag{5.7}$$



Model: 18sp3bar3

Model: 25sp3bar3

Model: sp3bar3

Model: bar6

Figure 2. *Hellinger affinity surfaces under the four different base astrophysical models.*

and the corresponding indices provide an estimate of $\max(k, j)$. We refer to this estimate of $\max(k, j)$ as $\widehat{\max(k, j)}$.

Of the chosen 216 grid points, the location $(\hat{r}_{\max}, \hat{\theta}_{\max})$ corresponding to the $\widehat{\max(k, j)}$th grid point is the one that maximizes the affinity of the density of the observed data to that of the simulated data. In other words, of the 216 chosen locations in our work, this location best represents the value of the unknown model parameter vector $\mathbf{X}$ at which the observed data $\mathbf{Y}^{(obsvd)}$ are realized. Since $\mathbf{X}$ is the unknown location of the observer who observes data $\mathbf{Y}^{(obsvd)}$, $\hat{r}_{\max}$ and $\hat{\theta}_{\max}$ best represent the values of the unknown radial and angular location of the observer, respectively, of the set of chosen locations that we use in our work, following the astrophysically motivated choice of such parameters by Chakrabarty (2007).

In Table 1 we present the estimated locations (and their indices) where the affinities are maximized for the four base astrophysical models.

Thus the location of the maximum affinities is quite different for the four models. Note that these point estimates are not going to be very precise owing to the multimodal and flat

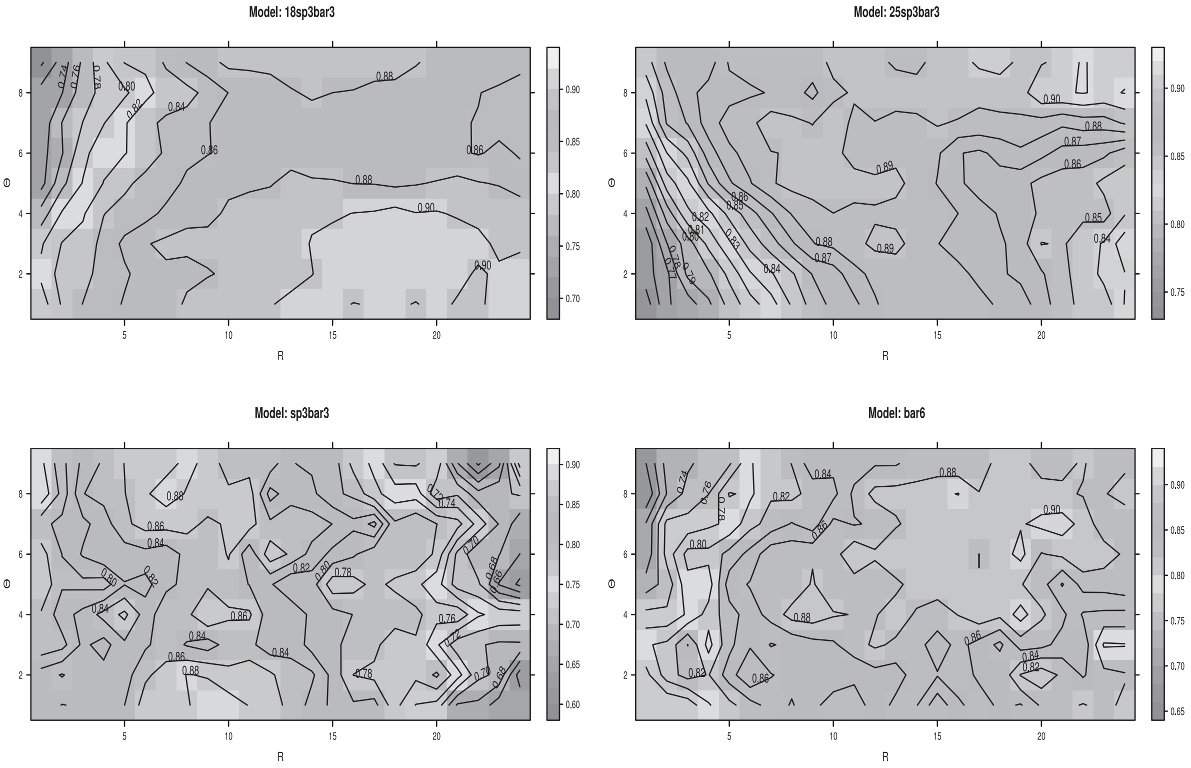


Figure 3. *Contour plots of the Hellinger affinity surfaces under the four different base astrophysical models.*

Table 1

Location of maximum affinities for the four base astrophysical models.

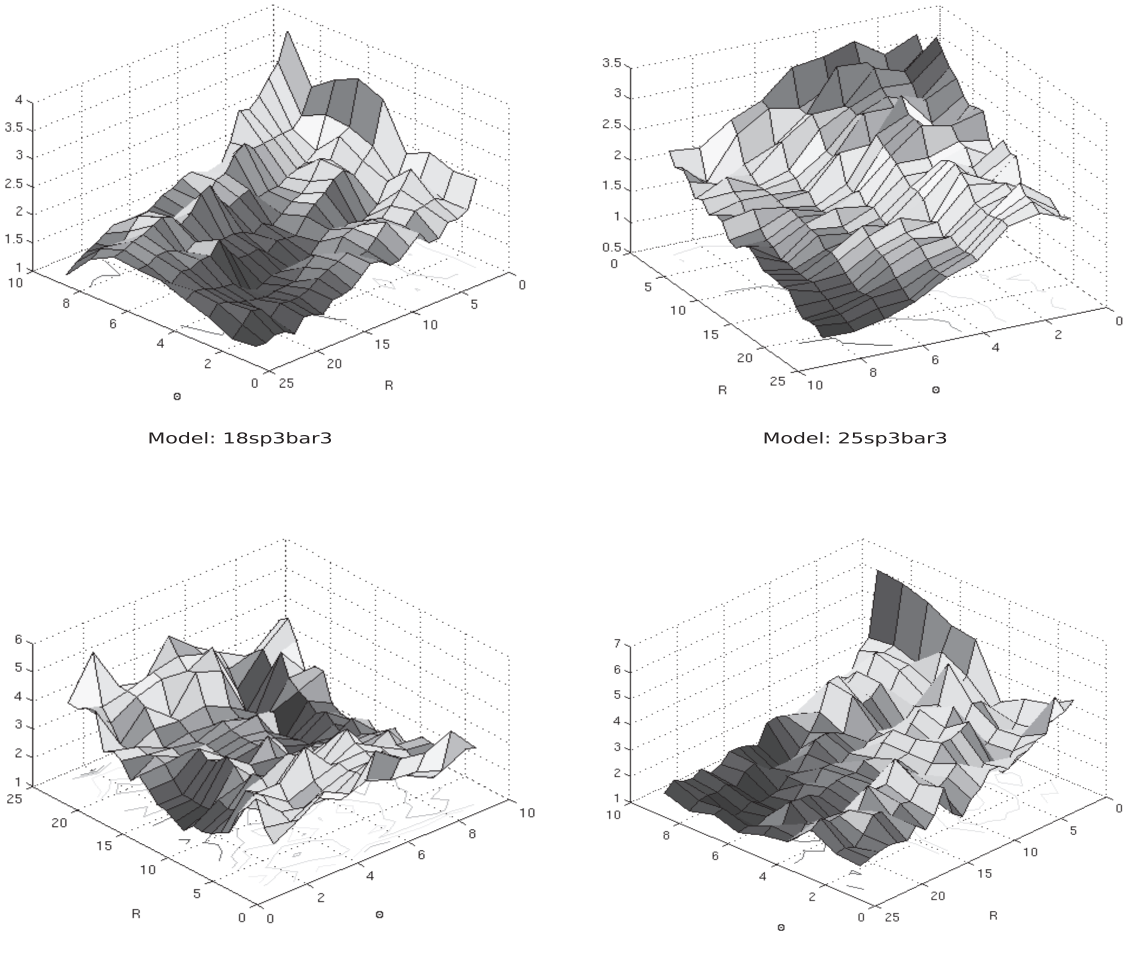
Model	$\widehat{\max(i,j)}$	$(\hat{r}_{\max}, \hat{\theta}_{\max})$
18sp3bar3	(20, 2)	(2.1875, 15°)
25sp3bar3	(22, 9)	(2.2325, 85°)
sp3bar3	(10, 1)	(1.9375, 5°)
bar6	(21, 7)	(2.2125, 65°)

character of the affinity surfaces. The contour plots in Figure 3 provide more meaningful information. In section 5.4 we shall provide confidence sets around these point estimates and, in light of those results, will carry out a comparison of our results to those reported by Chakrabarty (2007) and Chakrabarty, Biswas, and Bhattacharya (2013).

5.3. Maximum entropy estimation of the location parameter. For a given base model, we define

$$\delta_{(r,\theta)} := \delta(g_{(r,\theta)}, f), \tag{5.8}$$

where $\delta(\cdot,\cdot)$ is the KLD defined in (4.5). Let the density estimate for the observed data $\mathbf{Y}^{(obsvd)}$ be abbreviated as $\hat{f}$ and the estimated density of the data $\mathbf{Y}_i^{(sim)}$ simulated at the chosen location $(r_i \cos\theta_i, r_i \sin\theta_i)^T$ be $\hat{g}_{(r_i,\theta_i)}$, $i = 1, 2, \ldots, d$, $d = 216$. Thus for each base



Model: 18sp3bar3

Model: 25sp3bar3

Model: sp3bar3

Model: bar6

Figure 4. *KLD surfaces under the four different base astrophysical models.*

astrophysical model we have 216 KLD values at each of the 216 chosen locations at which the synthetic data sets are generated. For simplicity of notation we again denote

$$\hat{\delta}_{(r_i,\theta_i)} = \delta(\hat{g}_{(r_i,\theta_i)}, \hat{f}). \tag{5.9}$$

In Figure 4, we show the KLD surface generated at the 216 chosen locations for each of the base astrophysical models, i.e., the surface plot of $\hat{\delta}_{(r_i,\theta_i)}$ against (r_i, θ_i). The plot helps to visually detect where the surfaces attain their minima.

As with the affinity plots, here too we display KLD contours in grayscale; see Figure 5. Note that the overall appearance of these contour plots is in agreement with Figure 3. Again, as in the discussion of section 5.2, here too we invoke the construct that the d (=216) chosen locations are placed on a uniform two-dimensional rectangular grid. Then each grid point can be represented by a pair of indices such as (k, j), $k = 1, 2, \ldots, 24$, $j = 1, 2, \ldots, 9$. The location of the (k, j)th grid point is (r_k, θ_j). Let $\min(k, j)$ represent the indices for the particular

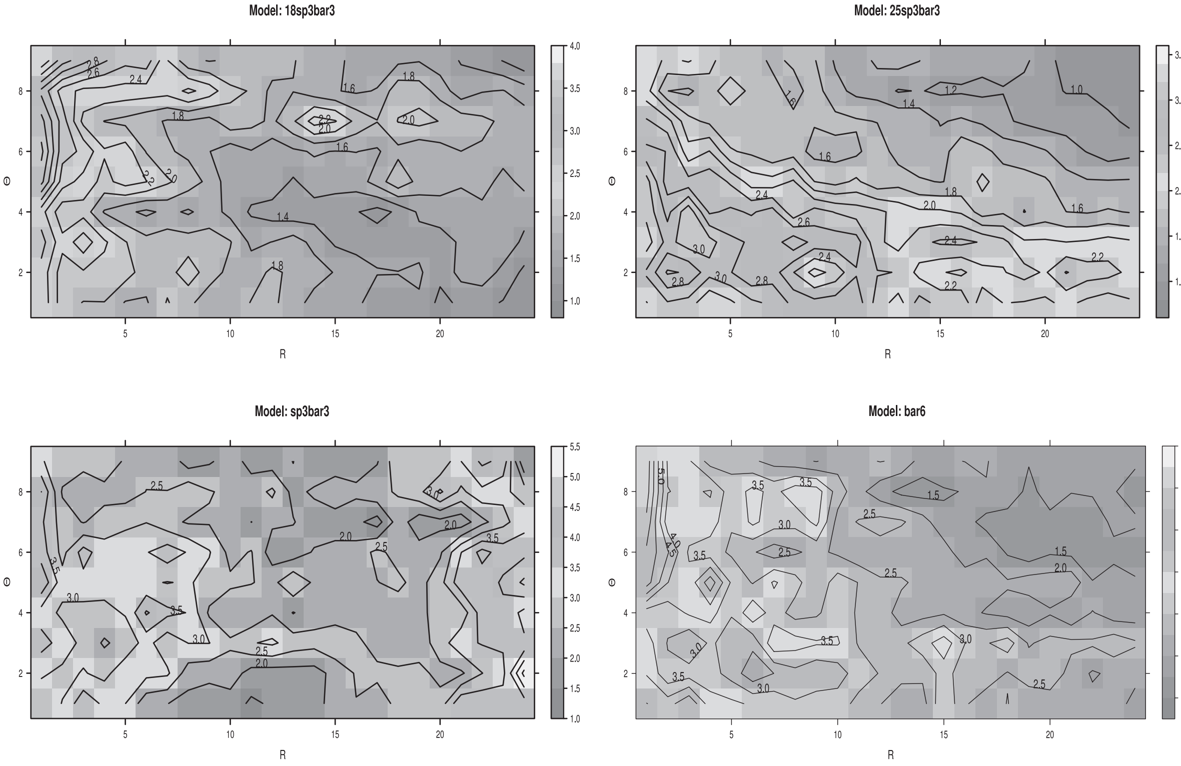


Figure 5. *Contour plots of KLD surfaces under the four different astrophysical models.*

grid point where the KLD values are minimized with the physical location of this grid point represented by

$$(r_{\min}, \theta_{\min}) = \arg\min_{(k,j)} \delta(g_{(r_k,\theta_j)}, f). \tag{5.10}$$

Thus, $(r_{\min}, \theta_{\min})$ is the actual physical location where the true distribution of the simulated data is closest to the true distribution of the observed data in the sense of having lowest KLD, while $\min(k, j)$ represents the indices for this location. We estimate this location (r_{min}, θ_{min}) by

$$(\hat{r}_{\min}, \hat{\theta}_{\min}) = \arg\min_{(k,j)} \hat{\delta}_{(r_k,\theta_j)}, \tag{5.11}$$

and the corresponding indices provide an estimate of $\min(k, j)$.

In Table 2 we present the coordinates of the location where the KLD values are minimized for the four base astrophysical models. Figure 6 shows level-plots of the affinity surface along with the KLD estimates. A corresponding KLD version is shown in Figure 7. Note that the estimates provided by these two approaches are quite close.

It is interesting to note that the surfaces are quite flat (particularly in the case of the base models 18sp3bar3 and bar6) near the peaks. Therefore, estimation of the location for which the affinity attains the maximum becomes difficult. One needs to investigate further to see whether this method will produce the right location consistently.

Table 2
Location of minimum KLD for the four models.

Model	$\widehat{\min(k,j)}$	$(\hat{r}_{\min}, \hat{\theta}_{\min})$
18sp3bar3	(24, 9)	(2.2875, 85°)
25sp3bar3	(24, 8)	(2.2875, 75°)
sp3bar3	(17, 7)	(2.1125, 65°)
bar6	(22, 7)	(2.2375, 65°)

In particular, we are concerned that the method of estimation used in our work abides by the undertaken assumptions. To this effect, we seek validation of our results.

At the same time, we are interested in quantifying uncertainties in the estimated locations of the chosen d locations at which the density of the synthetic data approaches the density of the observed data most closely, in the sense that the affinity measure between this pair of densities is the highest. In order to perform parameter uncertainty estimation, we undertake the construction of confidence sets using a bootstrap-based method.

5.4. Confidence sets. We are interested in quantifying uncertainties in the estimation of the locations (inside the grid of our choice) at which the affinity measure in maximized. We recall that $(r_{\max}, \theta_{\max})$ is the location at which the true distribution under the model is closest to the true distribution of the observed data in the sense of having the highest affinity among densities. We generated 300 bootstrap samples from the density of the synthetic data generated at $(r_{\max}, \theta_{\max})$. We then computed the affinity measures between the true density of the observed data and the bootstrap samples. This gave rise to a sampling distribution of the affinity measures between the density of the observed data and the bootstrap samples from the $(r_{\max}, \theta_{\max})$ location. Locations at which the values of the affinity measures (i.e., $\hat{\rho}(r_i, \theta_i)$) are above the cutoff point were included in the confidence set. For a 95% confidence set, we chose the lower fifth percentile of the empirical affinity distribution obtained through the above described bootstrap exercise as the cutoff point. However, we acknowledge that the suggested confidence sets will be valid under the assumption that the contours of constant affinity are shift invariant as $(r_{\max}, \theta_{\max})$ is varied.

In Figure 8 we show the confidence sets. The actual point where the affinity is maximized is indicated in gray, while the other points in the confidence set are indicated in light gray. It is interesting that the confidence sets for all the models are fairly small, and for the last two models the sets have just two members each. This shows that the estimation procedure is quite precise.

These estimates overlap moderately well with those reported by Chakrabarty (2007) as well as those by Chakrabarty, Biswas, and Bhattacharya (2013). For the base astrophysical model bar6, Chakrabarty (2007) reports that the angular location of the Sun lies between 0° and 49° with a median at 22°, while the radial location $\in [1.9625, 2.1975]$. For this model, Chakrabarty, Biswas, and Bhattacharya (2013) suggests that the mode of the marginal posterior probability density of $r_\star$ occurs at 2.2 and of $\theta_\star$ at 23.5°. In this Bayesian estimate of Chakrabarty, Biswas, and Bhattacharya (2013), the estimates lie in 95% highest probability density (HPD) credible regions that are, respectively $[2.04, 2.3]$ and about $[21°, 26°]$. As evident in Table 1, our point estimate for this base model is too high to fit into this interval. However, the confidence

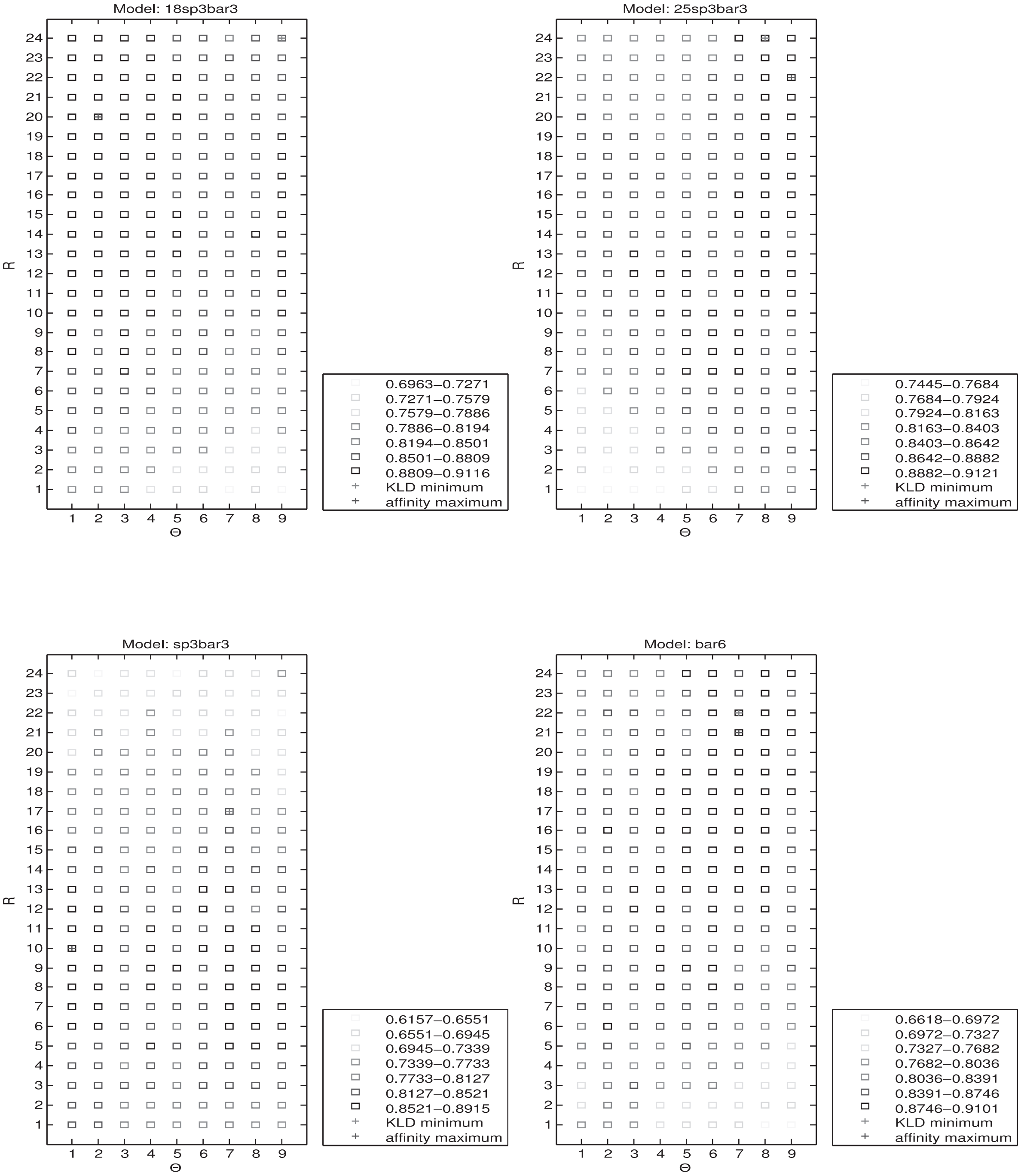


Figure 6. *A discrete representation of the level-plots of the affinity measure recovered in the two-dimensional grid of our chosen locations. Locations at which values of the recovered affinity measure lie in the same band are displayed in the same grayscale level. The grayscale coding of the affinity measure values is presented in the key adjoining each panel.*

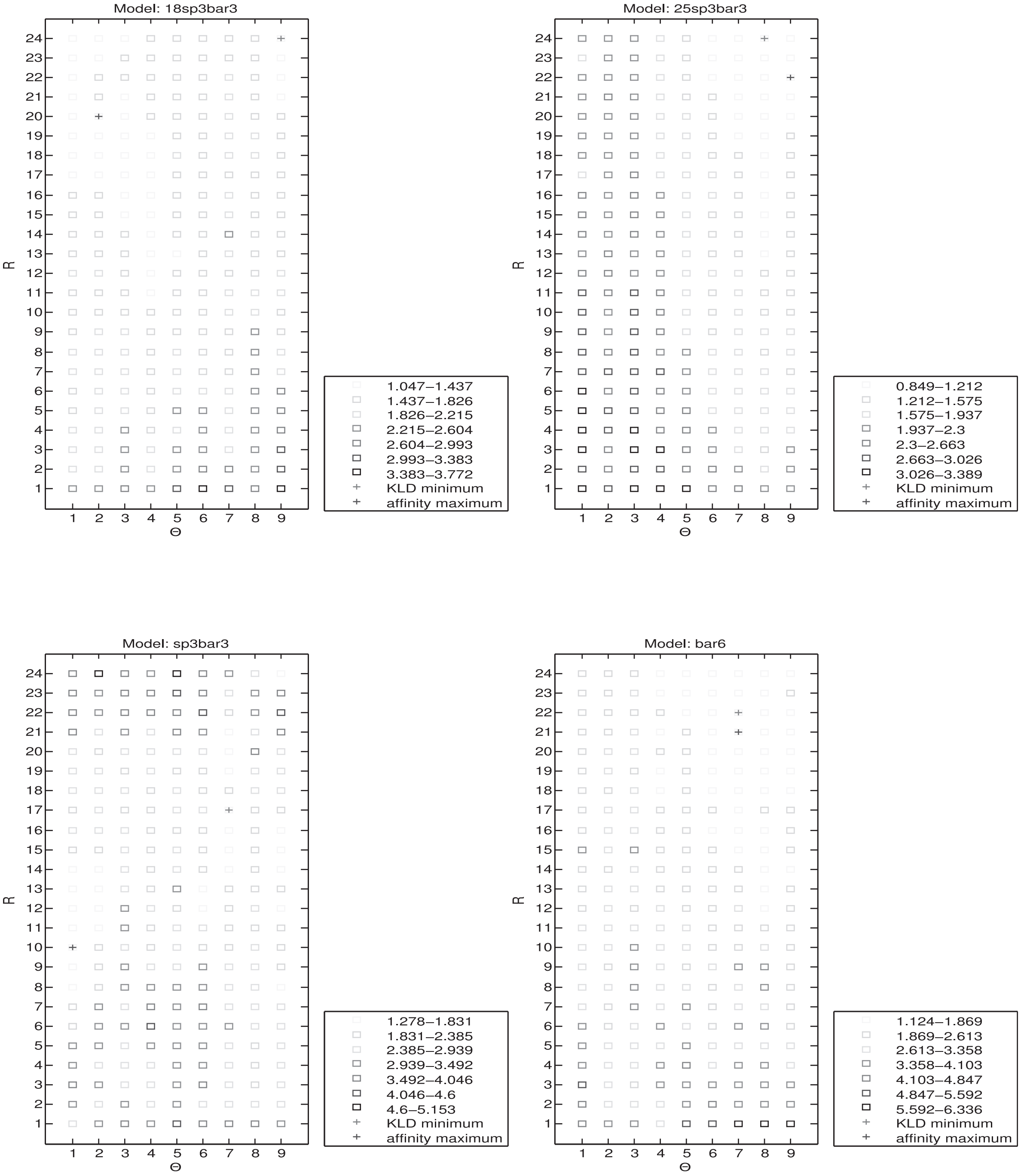


Figure 7. *Level-plots of the KLD surface, the analogous plot to Figure 6.*

set estimated for this base model includes locations at lower values of the radial location as well as lower angular location values (shown in light gray in Figure 8), such that these values are in conformity with the findings of Chakrabarty (2007) and Chakrabarty, Biswas, and Bhattacharya (2013).

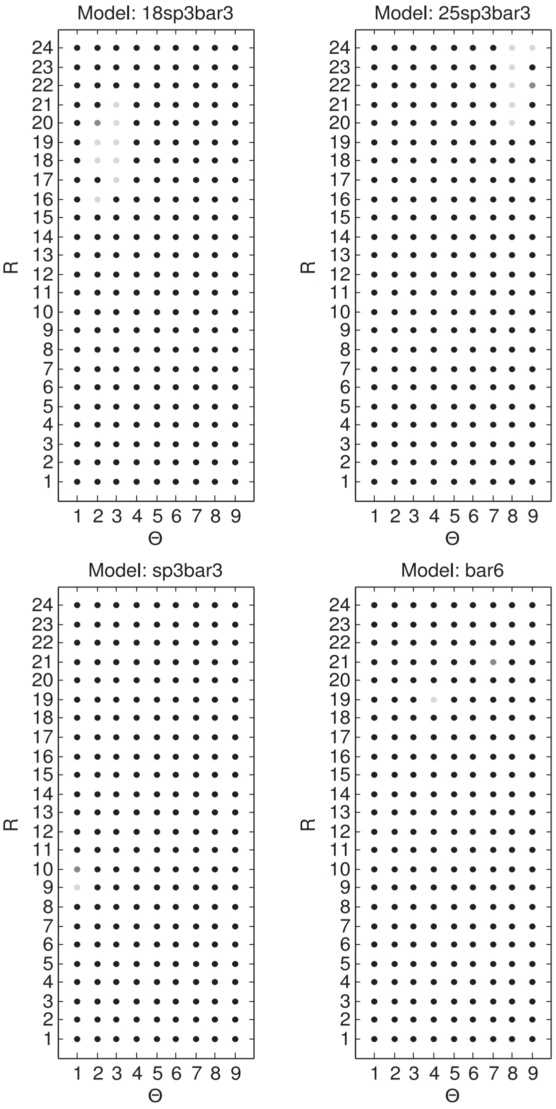


Figure 8. 95% *confidence set for* $(r_{\max}, \theta_{\max})$ *under each model. The elements of the confidence sets are depicted as light gray dots, the gray dots representing the point-estimates obtained earlier.*

For the base astrophysical model 18sp3bar3, the radial and angular location estimates of Chakrabarty (2007) are [1.95, 2.21] and [0°, 30°], respectively. The estimates of Chakrabarty, Biswas, and Bhattacharya (2013) are similar, with the 95% HPD credible region given by [1.7, 2.29] and about [10°, 62°] for the solar radial and angular coordinates, respectively. Our point estimate of (2.1875, 15°) for this base model then lies comfortably within these intervals; the confidence set recovered for this base model suggests that the observed data are consistent with radial location values lower than 2.1875 at the angular location value of 15° as well as at a higher angular value of 25°. In fact, a slightly higher radial location value of 2.2125 at an angular value of 25° is also included in our constructed confidence set for this base model.

This example helps to bring to the fore a salient advantage of the uncertainty estimation in our work, compared to that in Chakrabarty (2007). Given that we are performing a joint (radial and angular) parameter uncertainty estimation, we present our results as confidence sets on the tow-dimensional grid of our chosen locations. This allows for identification of the interval estimate of the solar location more clearly than in Chakrabarty (2007), in which the intervals represent uncertainties on the radial or angular values obtained using the marginal distribution of the radial or angular location values. Thus, it needs to be emphasized that the interval estimation of Chakrabarty (2007) are not to be directly compared to our estimated uncertainties. Additionally, the 95% HPD credible regions that Chakrabarty, Biswas, and Bhattacharya (2013) report are fundamentally different from our uncertainty estimates. We merely explore the possibility of an overall overlap between the results obtained using our methodology here with what exists in the literature.

In the context of our uncertainty estimation, we would also like to emphasis that it has been discussed in the literature that the underlying chaos in the base astrophysical model drives the estimated locations to be scattered over the constructed grid of the chosen locations (Chakrabarty, 2007; Chakrabarty and Sideris, 2008). In fact, a necessary condition for chaos to occur is the increasing noninjectivity of stellar velocity as a function of the unknown solar location $\mathbf{X}$ (Sengupta, 2003). In the results presented by Chakrabarty, Biswas, and Bhattacharya (2013), the models that manifest such chaos are those for which the posterior probability density of the location parameters are rendered multimodal. In other words, the distribution of the locations that are compatible with the observed data (i.e., the locations at which the affinity measure is high in our work) may be multimodal. This further suggests that a visual representation of the confidence sets (as in Figure 8) allows for easy reading of the interval estimation of the unknown solar location.

Of all the base models, Chakrabarty (2007) had found the distribution of locations compatible with the observed data to be most scattered over the grid of chosen locations for sp3bar3. This scatter disallowed the interval estimation of the unknown solar location in this earlier work. Chakrabarty, Biswas, and Bhattacharya (2013) agree with this trend in that the posterior densities of the location parameters are most multimodal for this model. We confirm a similar trend in our recovery of the affinity surfaces (Figure 6). However, our method of estimating uncertainties works for this base model and we recover a very small confidence set adjoining the point estimate at (1.9375, 5°); see Figure 8.

For the base model 25sp3bar3, our point estimate equals (2.2325, 85°) (see Table 1), while the recovered confidence set suggests that at a slightly lower angular location value of 75°, radial location values in [2.1875, 2.2875] are also compatible with the observed data as they are within the 95% confidence interval; the situation is the same for the location 2.2875 at the higher angle of 85°. While these radial location values overlap with the estimate from Chakrabarty (2007), our estimate of the angular locations are slightly in excess of the earlier estimate of angular location value.

5.5. Cross-validation.

In this context it is recalled that the estimation procedure is based on the assumption that the velocities observed from nearby locations and hence their corresponding densities will be more similar to each other than those observed from distant locations. It is important to verify that the affinity values obtained by this method show such desirable property. For this purpose, we used a cross-validation approach where one of the grid

Table 3
Results of cross-validation.

Model	1st nbhood	2nd nbhood
18sp3bar3	24	0
25sp3bar3	24	0
sp3bar3	24	0
bar6	22	1

points was chosen as the "true" location, and the corresponding kernel density estimate was chosen as the "true" density. The affinity values between this density and the density estimates at all the other grid points are obtained. Under the aforementioned assumption it is expected that the maximum of these affinity values should occur at one of the nearest neighbors of the "true" locations. For this analysis, the 24 × 9-sized grid $(r_k, \theta_j), k = 1, \ldots, N_R, j = 1, \ldots, N_\theta$, was broken into 24 blocks of size 3×3 each. The midpoint of each block was chosen as the representative for that block for the purpose of cross-validation; thus for each base astrophysical model, we had 24 points implemented in cross-validation.

When the midpoint (r_{k_m}, θ_{j_n}) is chosen as the true location, we define its first neighborhood points as the set of points (r_k, θ_j) such that $\max(|k - k_m|, |j - j_n|) = 1$, its second neighborhood points as the set of points (r_k, θ_j) such that $\max(|k - k_m|, |j - j_n|) = 2$, and so on. In Table 3, the first column gives the number of times the maximum occurred within the first neighborhood, and the second column gives the number of times the maximum occurred outside the first but within the second neighborhood. It is quite clear from the table that the maximum did occur closest to the true locations in a overwhelmingly large majority of cases, underscoring the effectiveness of the proposed method. These results give us the required confidence in our estimation. See Figure 9 for a visual idea about the locations of the maxima during cross-validation. The points which are chosen for the implementation of the cross-validation algorithm are indicated in gray; light gray lines join them to the point where the corresponding maximum of the affinity was observed.

In addition to performing cross-validation to check against internal inconsistencies, we have successfully compared our results with those reported by Chakrabarty, Biswas, and Bhattacharya (2013) on the basis of their Bayesian method that is independent of density estimation (see section 5.4).

5.6. Direct divergence estimation. On the suggestion of one of the reviewers, we explored some methods of construction of divergences avoiding density estimation. In particular we considered the construction of the rPE divergence (introduced in section 4.3.3) using the direct method of estimating the density ratio. The surfaces of the new divergence (Figure 10) have reasonable similarity with the KLD surfaces (Figure 4), and general conclusions based on the new surfaces are largely compatible with our previous findings. Thus it appears that the methods that bypass the issue of density estimation can have some real utility in practice. We note, however, that at present theoretical consistency results about the direct density ratio method are limited in number as well as scope.

6. Concluding remarks. In this paper we have developed a new method for estimating the location of the Sun with respect to the center of the Milky Way. Observed two-dimensional

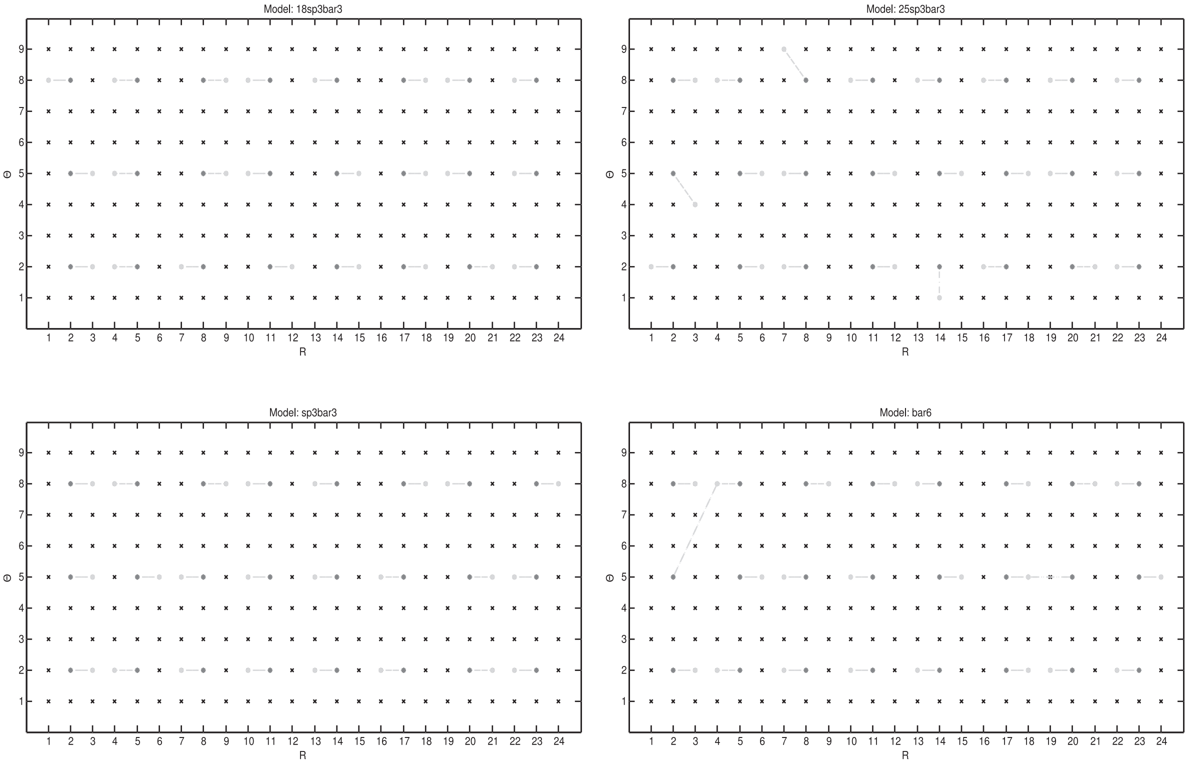
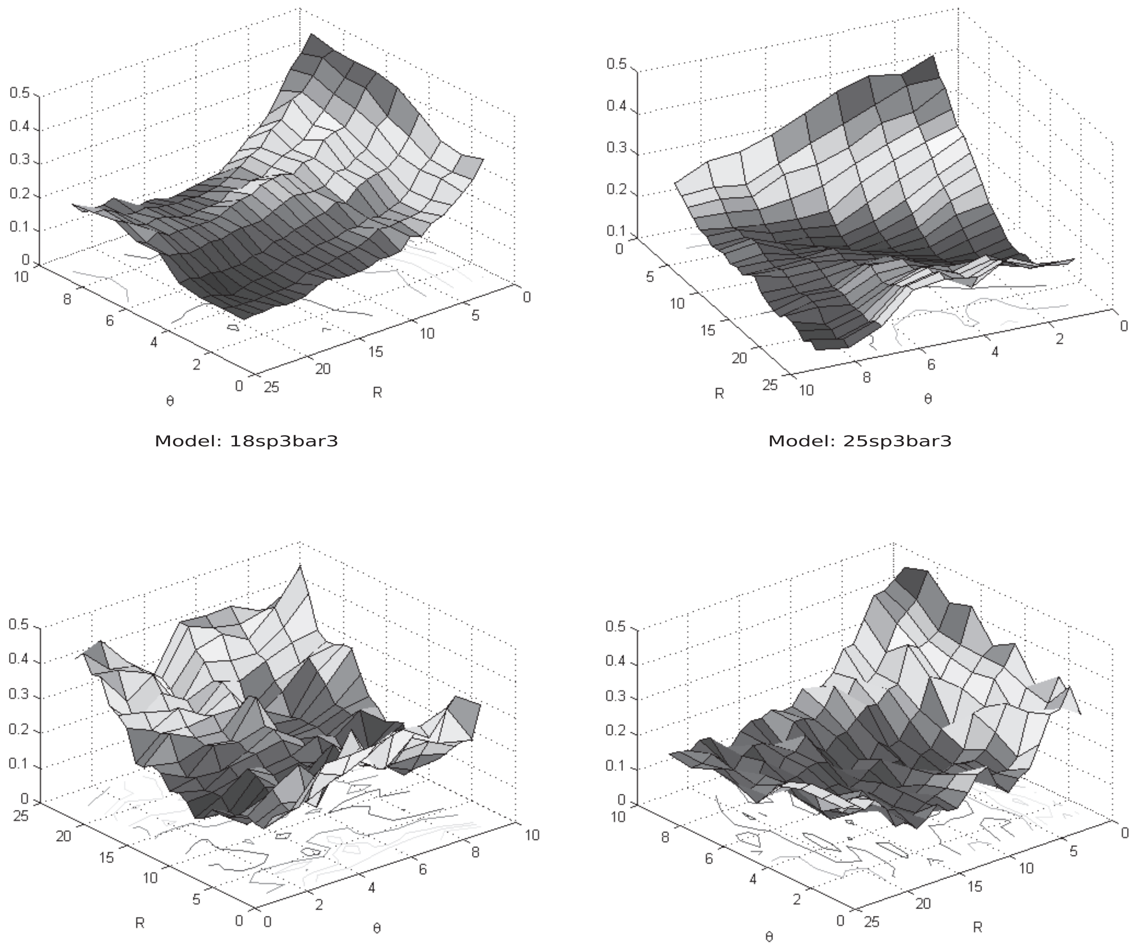


Figure 9. *Graph depicting the locations where the maxima occurred during cross-validation. The gray stars represent the coordinates chosen for cross-validation and light gray lines connect these coordinates to their corresponding maxima (represented by light gray stars).*

velocity vectors of stars were used to estimate the distribution of the observed stellar motion where the location of the observer, i.e., the location of the Sun, with respect to the center of the Galaxy is unknown. This distribution was compared to distributions estimated using synthetic stellar velocities generated at known locations in the Milky Way disk, where such synthetic data were taken from the astronomical literature. The comparison was performed by considering affinity measures based on the Hellinger distance. In doing so we have made a direct determination of the compatibility of the location from which the observed stellar velocities were recorded with these synthetic data sets. Our procedure allows us to estimate the observer location directly as a point on the (radial, angular) plane, rather than estimating the components of the location vector individually. Indeed, the confidence set of the estimated positions that we develop, based on the bootstrap technique, is a set of locations on the two-dimensional plane rather than a product of intervals. As a final test we run a consistency check on the estimates through a cross-validation experiment which indicates that the estimation procedure has some desirable continuity properties. The method provides a new perspective on the problem under consideration without contradicting the general belief about the behavior of the astronomical models under study.

Acknowledgments. The authors gratefully acknowledge the detailed comments of two anonymous referees, which led to a significantly improved version of the paper. The major



Model: 18sp3bar3

Model: 25sp3bar3

Model: sp3bar3

Model: bar6

Figure 10. *Direct rPE divergence surfaces.*

part of this work was done when the first and last authors were graduate students at the Indian Statistical Institute, Kolkata.

REFERENCES

F. J. Aherne, N. A. Thacker, and P. I. Rockett (1998), *The Bhattacharyya metric as an absolute similarity measure for frequency coded data*, Kybernetika (Prague), 34, pp. 363–368.

A. Basu, H. Shioya, and C. Park (2011), *Statistical Inference: The Minimum Distance Approach*, Monogr. Statist. Appl. Probab. 120, CRC Press, Boca Raton, FL.

A. Bhattacharyya (1943), *On a measure of divergence between two statistical populations defined by their probability distributions*, Bull. Calcutta Math. Soc., 35, pp. 99–109.

A. S. Bijral, N. Ratliff, and N. Srebro (2012), *Semi-Supervised Learning with Density Based Distances*, preprint, arXiv:1202.3702.

M. J. Canty (2007), *Image Analysis, Classification and Change Detection in Remote Sensing: With Algorithms for ENVI/IDL*, CRC Press, Boca Raton, FL.

D. Chakrabarty, M. Biswas, and S. Bhattacharya (2013), *Bayesian Nonparametric Estimation of Milky Way Model Parameters Using a New Matrix-Variate Gaussian Process Based Method*, preprint, arXiv:1304.5967.

D. Chakrabarty and I. Sideris (2008), *Chaos in models of the solar neighbourhood*, Astronom. Astrophys., 488, pp. 161–165.

D. Chakrabarty (2007), *Phase space around the solar neighbourhood*, Astronom. Astrophys., 467, p. 145.

N. Cressie and T. R. Read (1984), *Multinomial goodness-of-fit tests*, J. Roy. Stat. Soc. Ser. B Methodol., pp. 440–464.

A. Djouadi, O. Snorrason, and F. Garber (1990), *The quality of training sample estimates of the Bhattacharyya coefficient*, IEEE Trans. Pattern Anal. Mach. Intell., 12, pp. 92–97.

B. Efron and R. Tibshirani (1997), *Improvements on cross-validation: The* 632+ *bootstrap method*, J. Amer. Statist. Assoc., 92, pp. 548–560.

E. D. Feigelson and G. J. Babu (2012), *Modern Statistical Methods for Astronomy: With R Applications*, Cambridge University Press, Cambridge, UK.

R. Fux (2001), *Order and chaos in the local disc stellar kinematics induced by the galactic bar*, Astronom. Astrophys., 373, pp. 511–535.

Z. Ghahramani (2003), *Graphical Models: Parameter Learning*, MIT Press, Cambridge, MA.

T. Hastie, R. Tibshirani, and J. J. H. Friedman (2001), *The Elements of Statistical Learning*, Vol. 1, Springer, New York.

D. Heckerman, D. Geiger, and D. M. Chickering (1995), *Learning Bayesian networks: The combination of knowledge and statistical data*, Machine Learning, 20, pp. 197–243.

T. Kailath (1967), *The divergence and Bhattacharyya distance measures in signal selection*, IEEE Trans. Comm. Tech., 15, pp. 52–60.

M. Kearns, Y. Mansour, A. Y. Ng, and D. Ron (1997), *An experimental and theoretical comparison of model selection methods*, Machine Learning, 27, pp. 7–50.

S. N. Kirmani (1971), *Some limiting properties of Matusita's measure of distance*, Ann. Instit. Statist. Math., 23, pp. 157–162.

R. Kohavi et al. (1995), *A study of cross-validation and bootstrap for accuracy estimation and model selection*, in Proceedings of IJCAI 14, pp. 1137–1145.

S. Kullback and R. A. Leibler (1951), *On information and sufficiency*, Ann. Math. Statist., 22, pp. 79–86.

D. Landgrebe (2003), *Signal Theory Methods in Multispectral Remote Sensing*, Wiley, Hoboken, NJ.

M. Last (2006), *The uncertainty principle of cross-validation*, in Proceedings of IEEE International Conference on Granular Computing, pp. 275–280.

A. B. Lee, K. S. Pedersen, and D. Mumford (2003), *The nonlinear statistics of high-contrast patches in natural images*, Int. J. Comput. Vision, 54, pp. 83–103.

K. Matusita (1953), *On the estimation by the minimum distance method*, Ann. Inst. Statist. Math., 5, pp. 59–65.

A. Orlitsky et al. (2005), *Estimating and computing density based distance metrics*, in Proceedings of the 22nd International Conference on Machine Learning, ACM, pp. 760–767.

D. N. Reshef, Y. A. Reshef, H. K. Finucane, S. R. Grossman, G. McVean, P. J. Turnbaugh, E. S. Lander, M. Mitzenmacher, and P. C. Sabeti (2011), *Detecting novel associations in large data sets*, Science, 334, pp. 1518–1524.

A. Sengupta (2003), *Toward a theory of chaos*, Internat. J. Bifur. Chaos, 13, pp. 3147–3233.

B. W. Silverman (1986), *Density Estimation for Statistics and Data Analysis*, Monogr. Statist. Appl. Probab. 26, CRC Press, Boca Raton, FL.

N. Simon, J. Friedman, and T. Hastie (2012), *A Blockwise Descent Algorithm for Group-Penalized Multiresponse and Multinomial Regression*; http://www-stat.stanford.edu/~jhf/ftp/noah.pdf.

M. Sugiyama, S. Liu, M. C. Du Plessis, M. Yamanaka, M. Yamada, T. Suzuki, and T. Kanamori (2013), *Direct divergence approximation between probability distributions and its applications in machine learning*, J. Comput. Sci. Engrg., 7, pp. 99–111.

R. Tamura and D. D. Boos (1986), *Minimum Hellinger distance estimation for multivariate location and covariance*, J. Amer. Statist. Assoc., 81, pp. 223–229.

M. J. Way, J. D. Scargle, K. M. Ali, and A. N. Srivastava (2012), *Advances in Machine Learning and Data Mining for Astronomy*, CRC Press, Boca Raton, FL.

K. Q. Weinberger and L. K. Saul (2006), *Unsupervised learning of image manifolds by semidefinite programming*, Int. J. Comput. Vis., 70, pp. 77–90.

M. Yuan and Y. Lin (2007), *Model selection and estimation in the Gaussian graphical model*, Biometrika, 94, pp. 19–35.

A. Zomorodian and G. Carlsson (2005), *Computing persistent homology*, Discrete Comput. Geom., 33, pp. 249–274.