

Chapter 4

Colorful Stars and Black Boxes: Bayesian Analysis of Stellar Populations



Paul D. Baines, Xiao-Li Meng, Andreas Zezas, and Vinay Kashyap

4.1 Stellar Archeology

This article details a statistical analysis of a complex problem in astronomy and astrophysics, with a broader aim to suggest strategies and methodologies for similar “black-box” problems in physical sciences and beyond. For reasons that will become clear, the specific problem we address is sometimes known as the “stellar archeology” problem. The archeological analogy provides a nice overview of the scientific objective: we are interested in estimating the age of objects (stars) from measurements of their attributes (brightness).

P. D. Baines

Department of Statistics, University of California Davis, Davis, CA, USA
e-mail: pdbainesredsox@ucdavis.edu

X.-L. Meng (✉)

Harvard University, Department of Statistics, Cambridge, MA, USA
e-mail: meng@stat.harvard.edu

A. Zezas

University of Crete, Department of Physics, Institute for Astrophysics, Foundation for Research and Technology-Hellas, Heraklion, Greece

High Energy Astrophysics Division, Center for Astrophysics, Harvard & Smithsonian, Cambridge, MA, USA

e-mail: azezas@physics.uoc.gr

V. Kashyap

High Energy Astrophysics Division, Center for Astrophysics, Harvard & Smithsonian, Cambridge, MA, USA

e-mail: vkashyap@cfa.harvard.edu

Astronomy has a long history of using and developing statistical methodology to analyze experimentally collected data (e.g., see [8]). Despite the inability to directly manipulate the processes being studied, astronomers and astrophysicists have amassed a large body of knowledge by both indirect observation of the underlying processes and the construction of physics-based models. As the understanding of underlying physical systems develops, observed data typically can be characterized as noisy observations of a complex physical process involving the parameters of inferential and scientific importance. The link between parameters of interest and observed data is provided by problem-specific knowledge, often in the form of a system of partial differential equations (PDEs). This characterization is common in many problems in astronomy, as well as other scientific fields such as the environmental sciences. When the driving systems cannot be solved analytically, or are particularly computationally expensive, the relevant community often relies on lookup tables, describing the expected observation (i.e., mean) for a variety of input parameters. Given the huge amount of expertise devoted to developing these models, the analysis of observed data often lags behind. Statistics can play an important role in such settings, although the resulting computation can be challenging. There has been increased interest in this type of problem, where one or more components in the model are a “black-box,” lookup table, or computer-model output [3, 4, 17, 20].

In this article, we present an example of such a problem: a hierarchical Bayesian analysis of photometric data. The objective is to infer stellar properties such as the mass, age, and metallicity of individual and collections of stars. The mapping between the scientifically interesting properties (mass, age, and metallicity) and the observed data (photometric measurements) is governed by a series of isochrone tables: lookup tables derived under an assumed physics model. Isochrone tables are traditionally named after the location of the research groups that computed them: commonly used versions include the Geneva [13], Padova [14], and Dartmouth [6] isochrones. The highly structured mapping poses challenges for traditional computational methods, as discussed in Sect. 4.2. In Sect. 4.3 we present a generalizable and robust algorithm for posterior sampling that does not rely on any specific properties of the isochrone mapping. By avoiding the use of isochrone-specific fixes, we seek an algorithm that can successfully adapt to new lookup tables and could be applied to a wider class of problems. Combining the ideas of different augmentation schemes in [22] with an energy-based partition proposal distribution in the spirit of the Equi-Energy sampler [11], our “Equi-Expectation” MCMC scheme is both efficient and scalable to large datasets. The performance of competing sampling schemes is detailed in Sect. 4.4, together with an application to the 47 Tucanae dataset. In practice there are often uncertainties in the choice of deterministic physical model, and hence we investigate the issue of selecting between competing sets of isochrones in Sect. 4.5. The rest of Sect. 4.5 discusses some future directions and concludes.

4.2 Color-Magnitude Diagrams and Isochrone Tables

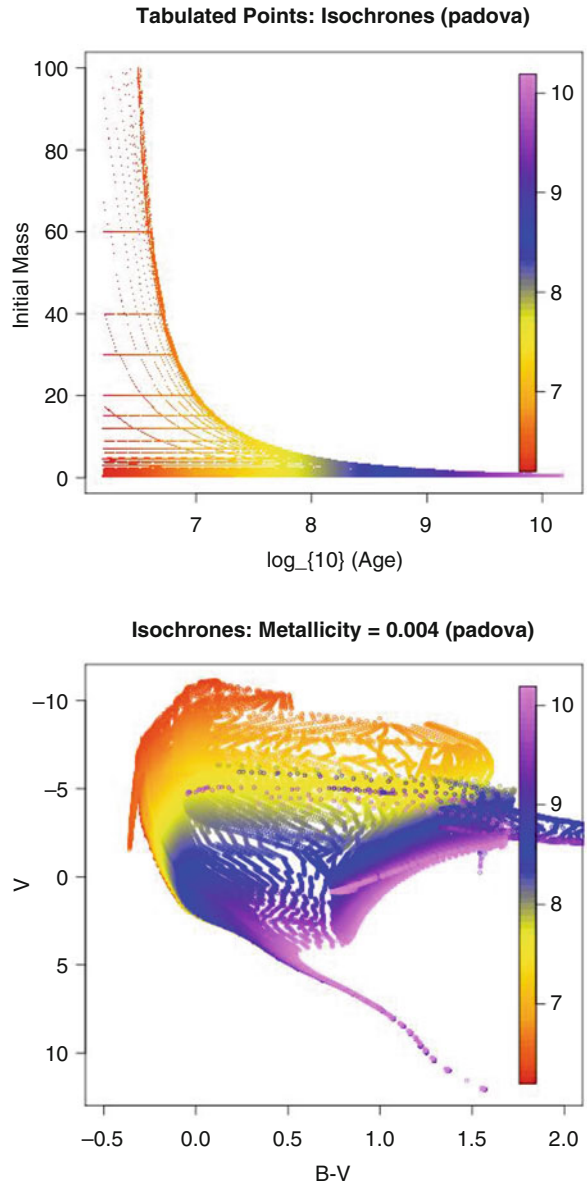
Photometric measurements are obtained by a detector, pointed at a particular region of the sky. Sources such as stars emit photons, which, together with additional background photons are counted whenever they pass through the detector. In crude terms, by counting the number and energy of photon arrivals at a particular detector location in a particular time interval, the photon counts can be calibrated to obtain the spectrum of a given source. The spectrum of a source represents the intensity across a continuous range of wavelengths and, as such, these observations can be expensive to obtain. An alternative is to use optical filters that allow only photons within a specified wavelength band/range to pass through. The measurements can then be thought of as estimating the integral of the spectrum over a small wavelength range. Depending on the number of bands, this approach yields a small number of measurements representing the brightness of the source that are both cheaper to obtain and simpler to analyze than their spectral counterparts. The brightness of a source in a photometric band such as B (blue) is also known as its B -band magnitude. Colors can be obtained as differences between magnitudes: for example, the color $B - V$ represents the difference in B - and V -band (visual) magnitudes. In light of this property we are able to freely switch between colors and magnitudes, and the analysis of Sect. 4.3 can be conducted across different combinations of colors and magnitudes.

To relate observed photometric data to the relevant physical quantities such as the age, mass, and metallicity of the stars, we use a theoretical collection of *isochrones*. The term “isochrone” is typically used to refer to the curve defined by tracing out the expected color and magnitude for stars of a fixed age and metallicity, for different initial masses. More generally, an isochrone can be viewed as a function that, given the physical properties of the star (mass, age, metallicity), returns the brightness of the star in a variety of different photometric bands. The metallicity of a star describes the relative abundance of elements such as oxygen and iron with respect to hydrogen.

The top panel of Fig. 4.1 displays all of the combinations of initial mass and age that appear in the (Padova) isochrone tables. The bottom panel of Fig. 4.1 displays the expected V -band magnitude and $B - V$ color of stars with a metallicity $Z = 0.004$, at each of the tabulated points of the isochrones. These plots correspond to the input and output spaces, with the isochrone mapping (i.e., the “black box”) between them. The color of each point in the plot indicates the age of the star, with younger stars typically being hotter and brighter than their older counterparts. The plot of color against magnitude is known as the Color-Magnitude Diagram (CMD), and forms the basis of the use of photometric data to infer stellar properties. Here, however, we use CMD to refer to the more general setting including higher-dimensional photometry and arbitrary (non-degenerate) color-magnitude combinations.

The initial mass of a star is a crucial factor in determining the evolution of the star. As stars age they burn off their component elements in order from the lightest

Fig. 4.1 Isochrone plot for (Top) the input/parameter space: initial mass and age, and, (Bottom) the output/observation-space: V -magnitude and $B - V$ photometry, for stars of metallicity $Z = 0.004$ from the Padova isochrones. The color of each point represents the age in \log_{10} -years (i.e., from $10^{6.0}$ to $10^{10.2}$ years), with the color-scale given on the right-hand side of each plot



to the heaviest, beginning with hydrogen and helium fusion. Since the chemical composition of the star is one of the determining factors in its photometry, and we have a physics model for the stellar evolution process, we can attempt to infer the age, initial mass, and metallicity of the star from observed data.

The bottom-left portion of the CMD (Fig. 4.1, bottom) is known as the *main sequence*. This is where stars spend most of their lives, usually before evolving into either a brown or white dwarf. On the main sequence, there are many different combinations of mass, age, and metallicity that produce the same expected photometry, leading to a degeneracy in measurements. Therefore, taken in isolation, the mass and age of a given star may or may not be identifiable. The applications we consider are those where we are interested in estimating the properties of a “cluster” of stars. Typically, by a cluster of stars we mean a collection of stars located in a similar physical location, and at a similar distance from the detector. Despite the individual-level potential non-identifiability, by combining observations, we can proceed to draw inference about both cluster-level and individual stellar properties. In addition to these challenges with identifiability, small changes in mass and age can potentially produce large changes in expected photometry, depending on the region of the CMD in which the star falls. These problems all add to the complexity of both the physical modeling and statistical analysis, but they are by no means unique to stellar archeology. We therefore believe that the strategies and methodologies in this article have general implications.

4.3 Hierarchical Modeling and Computation

Model Specifications

A photometric observation of source i , typically a star, is a vector of observed values in a combination of colors and magnitude bands, denoted by $Y_i \in \mathbb{R}^p$, where p is the number of bands. The (Gaussian) measurement errors from the detector are typically well understood, in the sense that variances are traditionally taken to be known for each band. Without loss of generality, we can assume unit variance (i.e., working with standardized Y_i). Here we allow for the measurement errors to be correlated across bands or colors: the correlation structure is assumed to be constant among all stars and is modeled with a weakly informative prior. Given the intrinsic properties of the stars, the measurement errors are assumed independent across different stars. The lower-level data generating process is thus given by

$$Y_i | M_i, A_i, Z \stackrel{ind}{\sim} N(f(M_i, A_i, Z), \mathbf{R}) \quad i = 1, \dots, n, \quad (4.1)$$

where M_i and A_i are the (initial) mass and age of the star, Z is the metallicity of the cluster and $f(M_i, A_i, Z)$ is a vector of the expected photometry of the star, found from the isochrone tables, and is standardized the same way as Y_i is. For

all applications here we consider Z to be known from external knowledge, as is standard in the astrophysics literature, although extending the model to include unknown metallicity is conceptually straightforward. The correlation matrix \mathbf{R} is assumed to be the same across all observations, following a common strategy for balancing between model adequacy and model complexity (e.g., [2, 12]).

The literature on CMDs has assumed almost exclusively that stars in the same subpopulation (e.g., cluster) have the same age, and sought the best-fit isochrone based on this single age (e.g., [21]). This comes despite knowledge in many contexts that the spread in stellar ages is sizable. Our approach remedies this problem but retains model simplicity by placing a common structure on the ensemble of star ages. Allowing flexibility of individual parameters yet utilizing the common structure motivates the following model. We assume the “population” of \log_{10} ages for a given cluster to be Gaussian (equivalently, age is \log_{10} -normal, not standard log-normal):

$$A_i | \mu_A, \sigma_A^2 \stackrel{iid}{\sim} N(\mu_A, \sigma_A^2), \quad (4.2)$$

where 10^{A_i} is the age of the star in years. The traditional single-age approach amounts to imposing $\sigma_A^2 = 0$ and finding a “best” choice of μ_A , the parameter of primary inferential importance. Here μ_A characterizes the theoretical mean age (on the \log_{10} -scale) of the collection of stars, while σ_A^2 specifies the intra-cluster variability of the individual ages. By modeling the distribution of individual star ages, we can potentially detect outlying stars or multi-cluster populations corresponding to multi-burst star formation processes. Although such discoveries are feasible when we move beyond the single-age paradigm, estimation in multi-clusters contexts should be redone with explicit multi-cluster models, as we discuss in Sect. 4.3.

To complete the model specification, we use the conjugate hyperprior:

$$\mu_A | \sigma_A^2 \sim N\left(\mu_0, \sigma_A^2 / \kappa_0\right), \quad \sigma_A^2 \sim \text{Inverse-}\chi^2\left(\nu_0, \sigma_0^2\right). \quad (4.3)$$

Typically we have prior knowledge that the stars in a given dataset are all of a similar, though not necessarily identical, age. The prior mean and variance for σ_A^2 are $m_0 \equiv \nu_0 \sigma_0^2 / (\nu_0 - 2)$ and $\tau_0^2 \equiv 2m_0^2 / (\nu_0 - 4)$, respectively. Therefore, in this setting σ_A^2 is given a prior where ν_0 is large, and σ_0^2 is set to the expected within-cluster variance of the individual stellar ages. The isochrone mapping is both highly nonlinear and degenerate in that many different parameter values lead to the same expected photometry. As a result, there is typically insufficient information in the data alone to give meaningful answers. The inclusion of external knowledge from previous literature or standard astrophysics theory is an important tool in breaking these degeneracies. Indeed, the entire statistical model represents a translation of scientific understanding into a collection of modeling assumptions, and the Bayesian framework makes this task relatively straightforward. But we are

mindful of the need to check prior sensitivity and more appropriately to quantify inferential uncertainty.

The initial mass of a star, together with its metallicity, is one of the primary factors that determine how that star will evolve. The initial masses of stars are known to have a distribution, or *initial mass function* (IMF) that, for stars above a threshold M_{brk} , typically around one solar mass, is described by a power-law with parameter $\alpha = 2.5$ [18]. For stars below the threshold, the distribution of masses is considered to be uniform. However, we are interested in placing a prior on a star in our dataset, not the population of all stars. For a star of a given age we know *a priori* that for a star of that age to potentially be observed, it must have a mass within a certain range of values. As can be seen in the top panel of Fig. 4.1, stars with a large initial mass have a shorter lifespan than those with smaller initial mass. This leads to constraints on the support of the joint distribution of mass and age, with the support defined by the tabulation in Fig. 4.1. In light of this, and to ensure our prior includes only feasible (M_i, A_i) pairs we assume a distribution of the form:

$$p(M_i|A_i) = \begin{cases} 0 & M_i < M_{\min} \\ k & M_{\min} < M_i \leq M_{\text{brk}} \\ \frac{\alpha-1}{M_{\min}} \left(\frac{M_i}{M_{\min}}\right)^{-\alpha} & M_{\text{brk}} < M_i \leq M_{\max}(A_i) \end{cases}, \quad (4.4)$$

where $M_{\max}(A_i)$ is the maximum possible mass for an “observable” star of age A_i , as determined by the theoretical isochrones. M_{\min} is selected to be the minimum mass that is scientifically reasonable for the dataset, or from the set of theoretical isochrones (usually 0.8 solar masses) and does not vary with age. The prior distribution on \mathbf{R} is taken to be uniform across all positive definite correlation matrices. Note that this is not uniform on the correlation parameters, but will typically be close to uniform since the number of observed bands, p , is relatively small (e.g., single digit); see [2].

Posterior Inference and Sampling

The model specified by (4.1)–(4.4) yields a joint posterior distribution of dimension $d = 2(n + 1) + 0.5p(p - 1)$. In practice, however, we are typically most interested in the marginal posterior distribution of μ_A and σ_A^2 . In real applications, n is usually on the order of tens of thousands, although the size of dataset can be anywhere from hundreds to millions of stars. The large amount of structure among the posterior distributions of the parameters poses a challenge to standard methods of approximating posterior quantities of interest. We now describe a Markov chain Monte Carlo (MCMC) scheme to sample from the posterior distribution. We utilize a Metropolis-within-Gibbs scheme, which sequentially draws from the $d - n$ full conditional distributions of each component of \mathbf{R} , of $\{(M_i, A_i), i = 1, \dots, n\}$, and of (μ_A, σ_A^2) ; here we have only $d - n$ full conditional distributions because we draw

(M_i, A_i) jointly. Performing the first and last of these updates is straightforward. Hence, the greater interest is in sampling the mass and age of individual stars, given the observed photometry and stellar cluster parameters.

First we describe updates for the hyperparameters (μ_A, σ_A^2) and the correlation matrix \mathbf{R} . By conjugacy, the full conditional posterior distributions of the cluster-level parameters reduce to

$$\sigma_A^2 | \mu_A, \mathbf{A} \sim \text{Inverse-}\chi^2 \left(\nu_0 + n, \frac{\nu_0 \sigma_0^2 + S_A^2(\mu_A) + \kappa_0 (\mu_A - \mu_0)^2}{\nu_0 + n} \right)$$

$$\mu_A | \sigma_A^2, \mathbf{A} \sim N \left(\frac{\kappa_0 \mu_0 + n \bar{A}}{\kappa_0 + n}, \frac{\sigma_A^2}{\kappa_0 + n} \right),$$

where $\mathbf{A} = \{A_i, i = 1, \dots, n\}$, $S_A^2(\mu_A) = \sum_{i=1}^n (A_i - \mu_A)^2$, and $\bar{A} = \frac{1}{n} \sum_{i=1}^n A_i$.

To draw \mathbf{R} , we use component-wise Metropolis–Hastings updates, with a uniform proposal over the range of values that result in a valid (positive definite) correlation matrix. It is shown in [2] that when we change one correlation at a time, the positive definiteness constraints reduce to solving a quadratic equation to find the conditional support of the correlation. While this method can be inefficient for large correlation matrices, typically the number of observed bands is small in our application, and hence the proposal is rapid to compute and performs well in most settings.

To sample from the conditional posterior distributions $p(M_i, A_i | \mu_A, \sigma_A^2, \mathbf{Y})$ for $i = 1, \dots, n$, we need to construct a proposal that is robust to both multi-modality and many different types of nonlinear dependencies that can be induced by different regions of the CMD. Figure 4.2 displays the “likelihood” of an old star as a function of initial mass and age, i.e., the contribution to the posterior from Eq. (4.1).

Ideally, to achieve this, we would utilize an energy-based sampler in the spirit of the Equi-Energy sampler of [11]. In its full incarnation the Equi-Energy sampler proceeds by constructing “energy bands” that attempt to empirically partition the full parameter space into posterior contours. Given the dimensionality, constructing such energy rings for the full $2(n + 1) + 0.5p(p - 1)$ dimensional posterior is infeasible in practice, as is constructing full energy bands for subsets of conditional distributions. Since the contours of the conditional posterior distributions depend on the conditioned values, it would be necessary to re-compute the partition for every star across every iteration. Nevertheless, we now describe how we can explicitly utilize the tabulated component of the posterior distribution to pre-compute a single partition that can be used across all conditional distributions (M_i, A_i) , independently of the conditioning variables. By constructing the partition in this way, we retain the fundamental location-independent nature of the Equi-Energy sampler.

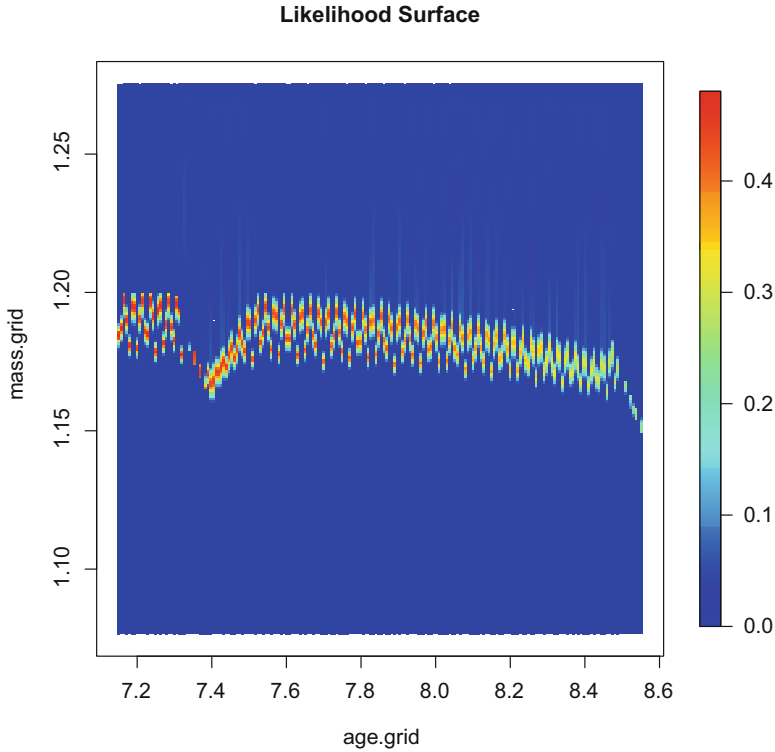


Fig. 4.2 The “likelihood” surface as a function of initial mass and age for a typical observation of an old star (10^{10} years)

A Partition Strategy Inspired by Equi-Energy Sampler

To avoid the additional complications of the CMD application, we first describe the construction of the proposal distribution for a simplified example. Consider two input (physical) parameters x and y that are related to two output quantities u and v on which measurements (with error) can be made. In the context of the CMD example x and y might correspond to the mass and age of an individual star, and u and v might correspond to two photometric bands. Suppose that the expected output for each of 2601 different combinations of input parameters (a regular grid of 51 unique values for each parameter) are given in a lookup table. The grid of input values is shown in the left-hand panel of Fig. 4.3. Our proposal distribution will be constructed from a partition of the parameter space: typically formed by polygons with corners at tabulated input points. The right-hand panel of Fig. 4.3 shows a possible partition of the input grid, obtained by Delaunay triangulation [16]. For each distinct polygon (triangle) we take the centroid as a “representative” of that region. Next, we compute the output value corresponding to the centroid. Since each

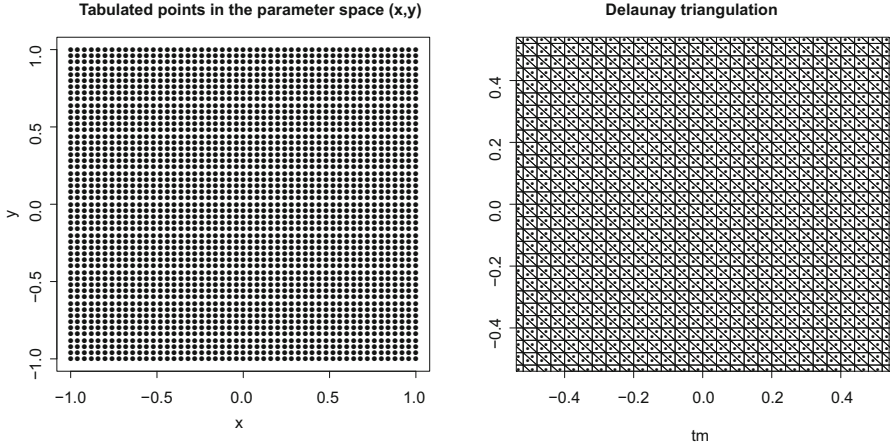


Fig. 4.3 (L) Tabulated combinations of the input parameters x and y . (R) Partition of the parameter space (x, y) into non-overlapping triangles, and centroids of those triangles. The right plot depicts only a subset of the parameter space, as the triangulation is regular. Vertices of the partitioning triangles are tabulated (x, y) points

vertex is a tabulated input point, the interpolated output value corresponding to the centroid is a distance-weighted average of the output values at the vertices.

For this example, we consider the following (isochrone) mapping from parameters to data space:

$$u = a_u(y - c_u)^2 + \sin(y) - |y + x| \quad (4.5)$$

$$v = a_v(x - c_v)^2 + \sin(x) + |y - x|, \quad (4.6)$$

where we select $a_u = 0.8$, $a_v = 1.2$, $c_u = -0.55$, and $c_v = 0.05$. Here we take the dimension of the data space to match the dimension of the parameter space, although this is not required. The mapping for an arbitrary point in the input, (x, y) -space, to the output, (u, v) -space, is done by interpolating the points in (u, v) -space corresponding to nearby points in (x, y) -space. For the interpolation to make sense in practice we require local continuity of the mapping between neighboring points, i.e., the tabulation must be sufficiently high resolution to enable safe interpolation of nearby values. This is not a restrictive requirement; all methods of analysis for CMDs rely upon sufficiently high-resolution tables. For the toy example, we now proceed as if the functional form of the mapping were not known: only tabulated values and interpolation are used.

Figure 4.4 illustrates some of the properties of the functional mapping from the parameters to the data. Firstly, as with the isochrone tables, the mapping is non-invertible: multiple (x, y) -values can lead to the same (u, v) -value, as manifested by intersecting cross-sections in the bottom panel of the figure. As with the isochrone tables, portions of the mapping are potentially invertible, but we want to derive a general method that does not rely on this fact. Secondly, as a result of this non-

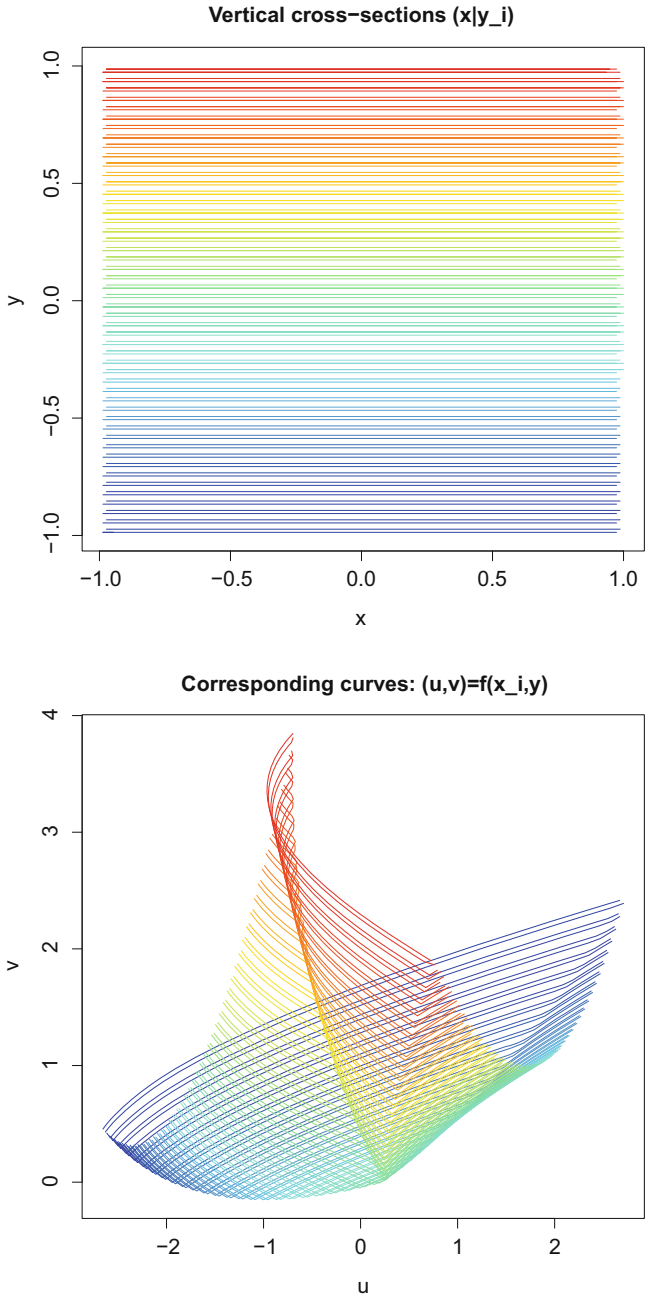


Fig. 4.4 (Top) Cross-sections of the parameter space, colored according to the fixed value of x . (Bottom) Each of the cross-sections maps to a curve in $(u, v) = f(x, y)$ as defined in Eqs. 4.5 and 4.6, where each curve is plotted in the same color as its corresponding cross-section

invertibility, similar observed data can arise from disjoint regions of the parameter space, hence the need to construct an efficient proposal distribution. Also, the mapping is not differentiable along the planes $x - y = 0$ and $x + y = 0$, also mimicking similar behavior to the isochrone tables.

The centroids of the input partition have their corresponding counterparts in the output space, shown in Fig. 4.5. The important observation is that, in a likelihood setting, similar expected values in the output space correspond to similar values of the target distribution. Hence, regions in the input space that correspond to nearby centroids in the output space will have similar likelihood values. In mapping back to the input space, we have essentially constructed a crude approximation to the inverse of the (many-to-one) mapping from input to output. The primary advantage of these “Equi-Expectation” contours is that they are expressed in a functional form. That is, given an arbitrary input point (x_0, y_0) , we have instant access to a set of points with “similar” expected values, without knowledge of the observed data or conditioning parameters.

Exact contours of the likelihood surface depend on the observed data, and hence require fresh computation for each observation. However, we can form a random-walk style proposal in the output space that produces approximately location-independent moves in the input space. For sufficiently high-resolution tables, the regions of the input space that are nearby in terms of their expected output value will have similar values of the likelihood. Larger distances between points in the output space correspond to larger differences in likelihood, with the Euclidean distance providing a natural metric when observations are made with Gaussian measurement errors. In practice, computing distances between all of the (u, v) -centroids is computationally expensive if the tables are very high resolution, as the isochrones tables are. So, to reduce the computational burden, we define “similar” in this context by running a clustering algorithm on the centroids in (u, v) -space, and tracking the accompanying (x, y) -clusters. In the CMD example, these clusters correspond to values of mass and age that have similar expected photometry: essentially the banded inversion of the isochrone mapping, f , as in (4.1). In general, the dimensions of the input and the output spaces do not need to match, and we can have input parameters defined on \mathbb{R}^k mapping to an output space on \mathbb{R}^p or subsets thereof.

A Proposal Distribution for (M_i, A_i) via Ancillary Augmentation

Now we return to the CMD example and address additional implementation challenges. Unlike the toy example, the isochrones are given on an irregular (m, a) grid, so the choice of partition is not immediate. The partition can either be constructed manually or using a standard technique such as the Delaunay triangulation [16] of the input vertices; we use the latter method for all applications presented here. Figure 4.6 shows an example of the partition formed using Delaunay

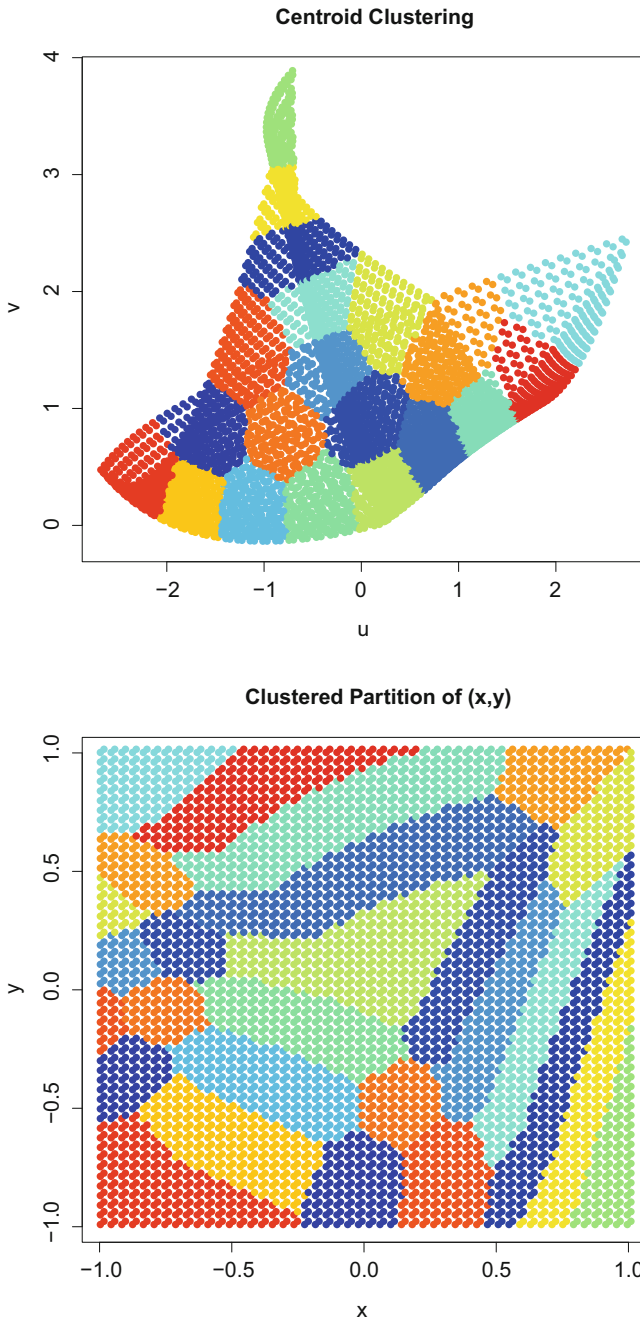


Fig. 4.5 (Top) Interpolated values $(u, v) = f(x, y)$ for each of the centroids defined by the (x, y) -partition in Fig. 4.3. The values are clustered into 50 groups containing “similar” output values, each represented with a different color. (Bottom) The corresponding clustered partition of the input. The color of each triangle reflects the cluster to which its centroid belongs

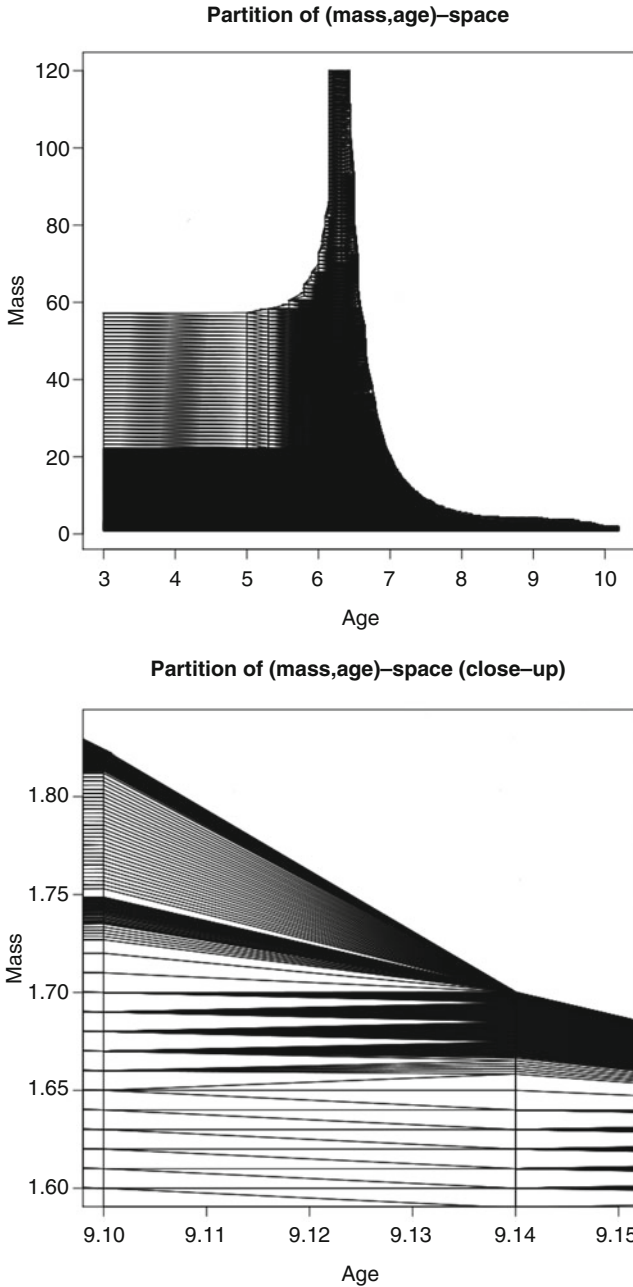


Fig. 4.6 (Top) Partition of the parameter space using the Delaunay triangulation, and (Bottom) a close-up of the partition. Note that blank regions in the upper right of each plot correspond to infeasible mass-age combinations

triangulation for the Geneva isochrones. Note that the tabulation is very irregular, with much higher resolution tabulation for masses close to the maximum allowable mass for each given age. Each corner of the polygon corresponds to a tabulated value that has a corresponding vector of expected photometric values. As discussed, assuming the isochrone tabulation is of sufficiently high resolution, the implied isochrone values within a given polygon can be approximated by interpolation of the (vector) values at the corner points. Next, taking the centroid of each polygon as a representative of that particular region of the parameter space, we proceed to construct approximate “contours” of the conditional posterior distributions that correspond to these centroid values. Each centroid is comprised of two components: (i) a pair of mass and age values (m_j, a_j) , and (ii) an (interpolated) isochrone value describing the expected photometry with the given mass and age $f(m_j, a_j) \in \mathbb{R}^p$.

After running the clustering algorithm on the photometry vectors at the centroids, each cluster is simply a list of polygons defining a collection of possibly disconnected regions of the parameter space. For computational simplicity, we use k -means clustering to form C clusters. Given a set of C clusters of polygons, we can quantify approximate measures of the “distance” between points in each pair of distinct clusters. Finally, after reparameterization, we ensure that nearby clusters, as quantified by their distance in the observed photometric bands, will provide similar values of the conditional posterior—yielding a proposal that enables both location-independent movement throughout the mass-age parameterization, and approximate contour-based sampling for *all* of the n independent conditional distributions $p(M_i, A_i | \mu_A, \sigma_A^2, \mathbf{Y})$. As presented however, the motivation for the partitions was that they allow location-independent exploration of multiple modes and diverse regions of the parameter space. However, in our hierarchical model we must deal with the additional contributions from the informative prior distributions in Eqs. (4.2) and (4.4). To do this, we perform the proposal using the *ancillary parameterization* [22]. For applications where the lowest mass stars are above the IMF break-point $M_{brk} = M_{min}$, this becomes

$$\tilde{A}_i = \Phi\left(\frac{A_i - \mu_A}{\sigma_A}\right), \quad \tilde{M}_i = \frac{M_{min}^{-(\alpha-1)} - M_i^{-(\alpha-1)}}{M_{min}^{-(\alpha-1)} - M_{max}(A_i)^{-(\alpha-1)}}, \quad (4.7)$$

where $\Phi(x)$ is the CDF for the standard normal variable. Under this augmentation scheme the model becomes

$$Y_i | \tilde{\mathbf{M}}, \tilde{\mathbf{A}}, \mathbf{R}, \mu_A, \sigma_A^2 \sim N\left(f(\tilde{M}_i, \tilde{A}_i, \mu_A, \sigma_A), \mathbf{R}\right), \quad (4.8)$$

$$\tilde{A}_i | \mu_A, \sigma_A^2 \sim \text{Unif}[0, 1], \quad \tilde{M}_i | \tilde{A}_i, \mu_A, \sigma_A^2 \sim \text{Unif}[0, 1], \quad (4.9)$$

$$\mu_A | \sigma_A^2 \sim N\left(\mu_0, \sigma_A^2 / \kappa_0\right), \quad \sigma_A^2 \sim \text{Inverse-}\chi^2\left(\nu_0, \sigma_0^2\right). \quad (4.10)$$

Hence, the conditional distribution of any given individual mass-age pair reduces to

$$p(\tilde{m}_i, \tilde{a}_i | \mu_A, \sigma_A^2, \mathbf{Y}) \propto \exp \left\{ -\frac{1}{2} \text{tr} \left(\mathbf{R}^{-1} \tilde{\mathbf{F}}_i \right) \right\}, \quad (\tilde{M}_i, \tilde{A}_i) \in [0, 1]^2, \quad (4.11)$$

where

$$\tilde{\mathbf{F}}_i = \left(\mathbf{y}_i - f(\tilde{m}_i, \tilde{a}_i, \mu_A, \sigma_A^2) \right) \left(\mathbf{y}_i - f(\tilde{m}_i, \tilde{a}_i, \mu_A, \sigma_A^2) \right)^\top. \quad (4.12)$$

By essentially placing all of the additional non-likelihood terms inside the mapping between sufficient and ancillary augmentation, we can help facilitate the improved performance of our likelihood-based proposal distribution. The impact of the transformation can be seen by the relative differences in areas between corresponding regions of the parameter space, i.e., the Jacobian. If the current state of the MCMC chain for star i is (m_i, a_i) , which is contained in polygon k , in cluster l , then we implement the partition-based proposal as follows:

Algorithm 1 [For Computing the Proposal Distribution]

1. Select a cluster l^* with probability $p_{ll^*}^C$.
2. Select a polygon k^* from within cluster l^* with probability $p_{l^*k^*}^W$.
3. Propose a point (m_i^*, a_i^*) uniformly within polygon k^* , and map to $(\tilde{m}_i^*, \tilde{a}_i^*)$.

□

By encouraging moves between nearby clusters we can effectively explore different regions of the parameter space with similar photometry, and hence, similar likelihood. Note that although the partition is constructed in terms of the stellar mass and age, the transformation defined by Algorithm 1 is one-to-one and monotonic, and hence it forms a valid partition in the ancillary parameterization for any values of μ_A and σ_A^2 . However, the transformation is not affine and the partition no longer consists of polygons. The transition probability corresponding to Algorithm 1 is given by $q((\tilde{m}_i, \tilde{a}_i), (\tilde{m}_i^*, \tilde{a}_i^*)) = p_{ll^*}^C p_{l^*k^*}^W |J(\tilde{m}_i, \tilde{a}_i)| / |\mathcal{U}_{k^*}|$, where $J(\tilde{m}_i, \tilde{a}_i)$ is the Jacobian of the transformation from the sufficient to ancillary augmentation evaluated at the proposed state, and $|\mathcal{U}_{k^*}|$ is the area (in the sufficient augmentation) of the k^* -th unique polygon within cluster l^* . There is some freedom in choosing both the cluster-to-cluster and within-cluster proposal probabilities. For the cluster-to-cluster probabilities we compute the centroid of all centroids of polygons within the cluster, providing an approximate “center” of the cluster (in \mathbb{R}^p) and then compute Euclidean distances between all cluster centers. The cluster-to-cluster proposal probabilities are then selected to be $p_{ll^*} = \exp \{-d^2(x_l, x_{l^*}) / \beta\}$, where $d(\cdot, \cdot)$ is the Euclidean distance, x_l and x_{l^*} are the cluster centroids in photometric-space, and β is a tuning parameter controlling how freely we propose to move to nearby clusters.

All distances here are computed with respect to the Euclidean norm in photometric-space independently of the mass-age location, thus imitating the

posterior distribution and allowing free movement across modes in terms of the stellar mass and age. For example, two distant regions of the parameter space that produce the same expected photometry would be placed in the same cluster, and the proposal distribution as constructed provides a high probability of proposing to move between these disconnected regions. We note that the acceptance probability is strongly influenced by the area of the polygon in the ancillary scheme; an artifact of the ancillary transformation (4.7). This choice of cluster-to-cluster probabilities mimics a random-walk Metropolis proposal, “centering” the proposal around the current cluster, and proposing to move to regions of the parameter space with a probability that reflects the similarity of the photometry to the photometry at the current state.

As discussed, the size of the polygon in the ancillary parameterization is a function of the hierarchical structure and could also be accounted for in selecting the within-cluster probabilities if desired. For example, if the within-cluster proposal probability is chosen to be proportional to the area of the region in the ancillary parameterization, then it yields a uniform proposal over the area defined by the cluster. In practice this is implemented by computing the polygon areas in the sufficient parametrization and incorporating the Jacobian term. If this approach is taken then, since the mapping between parametrizations depends on the hyperparameters μ_A and σ_A^2 , the Jacobian terms within the cluster would need to be recomputed at each iteration. Uniform proposals across the cluster do not require this extra computation but can be less efficient. In practice we combine this proposal distribution with a random-walk proposal of the form: $(m_i^*, a_i^*)^\top \sim N((m_i, a_i)^\top, \text{Diag}(\lambda_1, \lambda_2))$, where λ_j are proposal variances that can be tuned to achieve desired acceptance rates. This combination of proposal helps to facilitate both rapid local and global exploration of the posterior distribution. Since the correlation between mass and age depends on the region of the CMD, we do not attempt to approximate the correlation between the variables. In our experience, there is little performance change when using either the cluster-based or random-walk proposal distributions between 20–80% of the time.

Checking the Effectiveness of Our Proposal

To understand the impact of the transformation, and the resulting proposal distribution, we begin by examining the components of the posterior distribution within the original (m, a) -parametrization. Figure 4.7 shows the prior (Top) and posterior (Bottom) surfaces for an individual star. The posterior surface is obtained by combining the likelihood surface in Fig. 4.2 with the prior as shown. As we can see from the bottom panel of Fig. 4.7, the posterior surface is challenging to sample efficiently from, particularly given that the presence and scale of any large-scale ellipsoidal trends can vary dramatically across stars. In light of this, to retain computational robustness to the form of isochrone being used, and

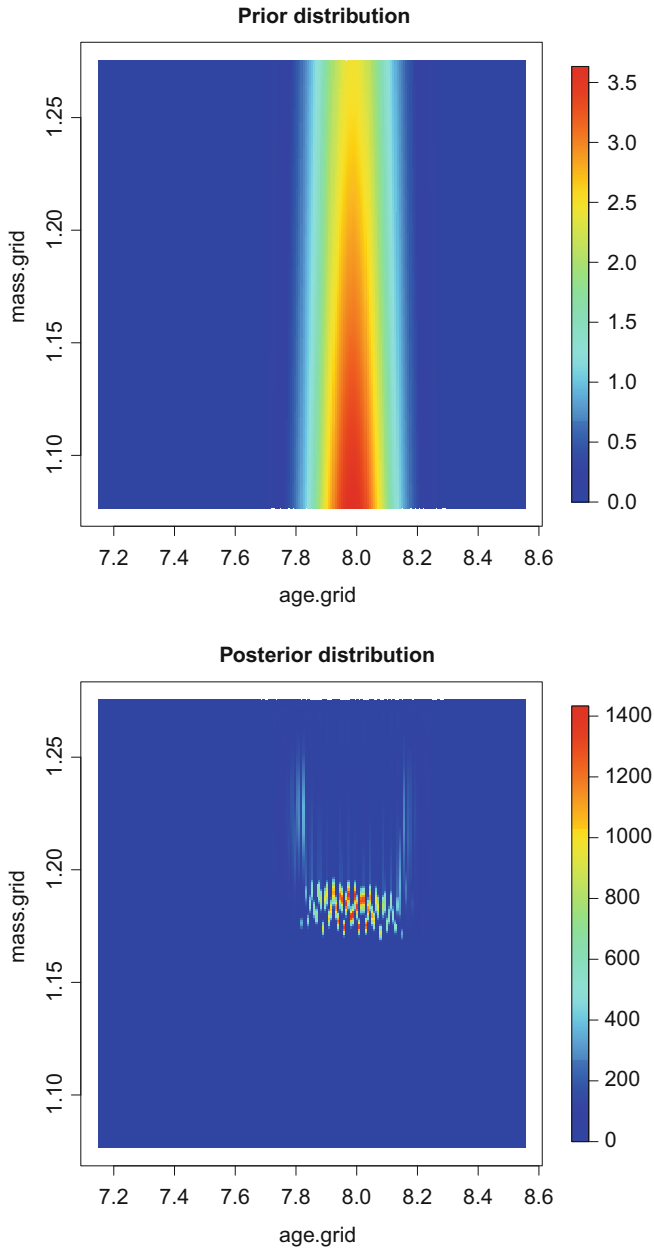


Fig. 4.7 (Top) Conditional prior surface for $(M_0, A_0)|\mathbf{Y}, \mu_a, \sigma_a^2$, and, (Bottom) conditional posterior surface for $(M_0, A_0)|\mathbf{Y}, \mu_a, \sigma_a^2$. The corresponding likelihood surface is given in Fig. 4.2

to maintain generality for non-isochrone settings, we avoid making observation-specific approximations to these conditional posterior distributions.

Our alternative approach, using the ancillary transformation, is depicted in Fig. 4.8. The top plot displays the transformation of the posterior distribution in Fig. 4.7 to the ancillary parametrization. Since the posterior distribution in the ancillary parametrization is simply a rescaling of the likelihood surface, we can observe the similarity in structure to Fig. 4.2. The bottom plot of Fig. 4.8 displays a proposal distribution obtained using our algorithm. The current state of the chain is highlighted by the black dot, and the proposal the proposal distribution mimics the contours of the ancillary posterior, albeit wrongly centered around the current state of the MCMC algorithm due to the random-walk style as implemented here. However, for this particular example the variance of our proposal is considerably greater than is desirable. This is the result of the observation falling in a region of insufficiently high resolution relative to the observational errors. This lack of resolution also illustrates the limitation of the “equi-expectation” approximation for low-resolution tables. Given higher resolution tables (or observations with higher measurement variance), and thus a higher resolution polygon-cluster proposal distribution, we will steadily obtain more appropriate contours and variance in the proposal.

As an analogous alternative to the random-walk style proposal, an independence style proposal could also be used where the cluster weights are proposed based on the distance between the observed photometry and the cluster centroids. This strategy would likely be more effective than the cluster-based alternative, but the large amount of computation required for each star and at each iteration renders it considerably more computationally expensive. In seeking the optimal trade-off between improved mixing and implementation speed, we elect not to pursue this further.

Although the clustering of polygons is not, in principle, necessary, the large number of polygons ($>200,000$) makes the construction of a polygon-to-polygon proposal more challenging, more memory-intensive, and more time-consuming. By adding the clustering of polygons, we need only store the much smaller $B \times B$ cluster-to-cluster proposal probability matrix, and possibly the within-cluster proposal probabilities (although this is not required for uniform proposals within the cluster).

Addressing Block Correlations

The proposal distribution for the individual stellar masses and ages is useful only in helping to sample efficiently from the series of conditional posteriors. As should be anticipated from the hierarchical specification of the model, there remains a large posterior correlation between (μ_A, σ_A^2) and (A_1, \dots, A_n) .

To help address these problems, we embed our sampler within a parallel tempering (PT) framework [9] to facilitate easier movement around the posterior space.

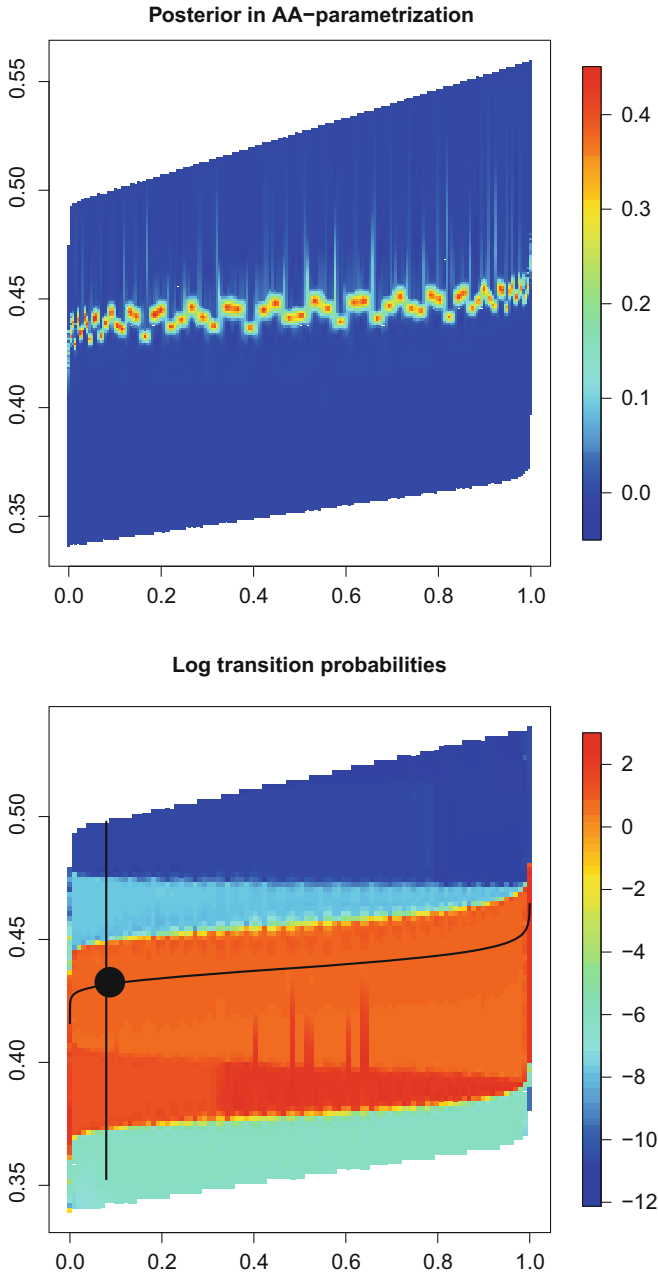


Fig. 4.8 (Top) Conditional posterior surface for $(\tilde{M}_0, \tilde{A}_0)|\mathbf{Y}, \mu_A, \sigma_A^2$, and, (Bottom) log-transition probabilities for (M_0, A_0) obtained by applying algorithm 1. The current state is shown by the black dot. The proposal mimics a random-walk across contours of the posterior surface

To sample $p(\theta)$ with energy $H(\theta) = -\log p(\theta)$, PT proceeds by constructing a sequence of tempered distributions, $\{p_1(\theta), \dots, p_N(\theta)\}$ of the form $p_j(\theta) \propto \exp\{-H(\theta)/T_j\}$, with $T_N > \dots > T_1 = 1$. By applying a series of monotonic transformations to the full posterior density, the full conditional densities in the Gibbs sampler are also transformed in an identical manner. An attractive feature of parallel tempering is that the modified conditional distributions require only trivial modifications. The tempered conditional posterior of (μ_A, σ_A^2) is given by

$$p_j(\mu_A, \sigma_A^2 | \mathbf{M}, \mathbf{A}, \mathbf{R}, \mathbf{Y}) \propto \exp \left\{ -\frac{1}{2T_j\sigma_A^2} \left[v_0\sigma_0^2 + \kappa_0(\mu_A - \mu_0)^2 + \sum_{i=1}^n (A_i - \mu_A)^2 \right] \right\} \cdot (\sigma_A^2)^{-\left(1 + \frac{(v_0+n+3-2T_j)/T_j}{2}\right)}$$

The conjugate marginal/conditional formulation can be shown to yield:

$$\sigma_A^2 | \mathbf{M}, \mathbf{A}, \mathbf{R}, \mathbf{Y} \sim \text{Inverse-}\chi^2 \left(\frac{v_n}{T_j}, \frac{v_0\sigma_0^2 + (n-1)s_A^2 + \frac{\kappa_0 n}{\kappa_0+n}(\bar{A} - \mu_0)^2}{v_n} \right), \quad (4.13)$$

$$\mu_A | \sigma_A^2, \mathbf{M}, \mathbf{A}, \mathbf{R}, \mathbf{Y} \sim N \left(\frac{\kappa_0\mu_0 + n\bar{A}}{\kappa_0 + n}, \frac{T_j\sigma_A^2}{\kappa_0 + n} \right), \quad (4.14)$$

where $v_n = v_0 + n + 3(1 - T_j)$ and $s_A^2 = \frac{1}{n-1} \sum_{i=1}^n (A_i - \bar{A})^2$. Since v_n cannot be negative, we must impose $\max_j T_j < (v_0 + n + 3)/3$. Typically either n , v_0 or both are large, and hence this condition is not generally restrictive.

4.4 Empirical Investigations

Simulation Studies

Given the complex properties of the isochrone tables, it is important to validate that the sampling algorithm can reliably converge to the correct posterior distribution. See [21] for an illustration of typical complications when using MCMC with isochrone tables. We approach this with an aggregate check of coverage properties. That is, we simulate many datasets from the model, and we then fit the model to obtain posterior intervals and check nominal and actual coverage are consistent. This is a special case of the more general framework in [5]. For this aggregate check we simulate 1000 datasets from the model for each of the parameter configurations detailed in Table 4.1.

Table 4.1 Details of coverage simulations used to validate the algorithm described in Sect. 4.3

Sim.	n	p	Bands	μ_0	σ_0^2	ν_0	κ_0	Number of datasets
1	100	3	(M(U),B,V)	8.1	0.01	1,000,000.0	1.0	1000
2	1000	3	(M(U),B,V)	9.2	0.05	1000.0	100.0	1000
3	100,000	3	(M(U),B,V)	6.6	0.16	1000.0	10.0	1000

Table 4.2 Coverage properties of the different sampling algorithms for simulation configuration 1 ($\mu_0 = 8.1$)

	1%	2.5%	5%	25%	50%	75%	95%	97.5%	99%
<i>M_i</i>									
MH	0.8	2.1	4.3	24.4	49.0	73.9	94.0	96.9	99.0
MH + PT	0.7	2.0	4.1	23.4	49.0	74.2	94.9	97.6	99.1
PC	0.7	2.1	4.3	24.3	48.8	74.2	94.4	97.1	99.0
<i>A_i</i>									
MH	1.7	3.4	5.5	25.2	48.0	72.3	92.5	95.2	96.9
MH + PT	1.2	2.9	5.0	26.2	49.4	75.2	94.7	97.6	98.9
PC	1.2	2.8	4.9	25.4	49.3	75.4	95.0	97.6	99.0
μ_a									
MH	1.7	2.9	5.0	24.8	46.9	72.0	91.6	94.7	96.3
MH + PT	0.7	2.2	4.3	25.8	49.3	76.5	95.3	97.7	99.2
PC	0.7	1.9	4.2	25.1	48.1	75.0	94.6	97.3	99.0
σ_a^2									
MH	0.7	1.8	3.4	24.2	48.0	75.1	95.2	97.1	99.1
MH + PT	0.7	1.5	2.8	21.5	45.9	72.4	95.0	97.0	99.5
PC	0.7	1.7	3.4	23.8	48.2	75.2	95.1	97.3	99.2

For each of the three settings, there are four different MCMC schemes:

1. **(MH)**: Vanilla scheme using only random-walk proposals for the individual masses and ages, without tempering,
2. **(MH + PT)**: As in (MH), with additional parallel tempering,
3. **(PC)**: The Polygon-Cluster scheme of Algorithm 1, without tempering,
4. **(PC + PT)**: As in (PC), with additional parallel tempering.

Due to limited computational resources, we did not implement the 4th scheme. In all cases we combine the results from four chains, and each algorithm is run for approximately the same total CPU time across the four chains. Remaining tuning parameters such as the variance of the random-walk proposal, the number of clusters and the cluster-to-cluster ‘‘variance’’ parameter β were chosen after pilot runs on a subset of the datasets. Table 4.2 shows the coverage properties for a subset of the parameters for simulation number 1.

The first, simplest method struggles to effectively sample the tails of the posterior distributions, particularly for the main parameter of interest, and most computationally challenging parameter, μ_a . Adding in tempering we do better

in many cases, but some potentially worrisome discrepancies between actual and nominal coverage still remain, even with tempering. The third method, using the cluster-based partition proposal without tempering appears to do slightly better than the standard approach both with and without parallel tempering. Although we do not directly employ the combination of the cluster-based proposal with parallel tempering across all 1000 datasets, we do recommend such an approach for the analysis of a single dataset. While on aggregate the differences between the approaches do not appear to be drastic, the results for any given dataset can differ by a non-negligible amount. Brute force numerical integration for a subset of the datasets suggests that the cluster-based proposal and the cluster-based proposal with tempering better capture the tails of the distribution, although we defer a fuller analysis for future investigation.

Results from configurations 2 and 3 are very similar to those presented above and omitted for brevity. One important difference that we note here though is the size of the dataset; configuration 3 analyzes 1000 datasets of 100,000 observations each: an important test of the scalability of our approach. We run the analysis for each dataset for a maximum of 24 h: a reasonable computational cost for such large-scale analysis.

NGC 104: 47 Tucanae

“47 Tuc” is a globular cluster estimated to be 13,000–17,000 light years away, originally discovered by Abbe Nicolas Louis de Lacaille in 1751 [7]. Being the second largest and second brightest globular cluster, it has been extensively studied in recent years. Examples include [10] and [19]. Here we reanalyze photometric data to investigate possible age differences within the cluster and to assess the sensitivity of estimates to the choice of hyperprior. The 47 Tuc data consists of 1,697 observed stars in V and B bands ($p = 2$), with no missing data. For the analysis here we consider the distance modulus to be fixed at 13.33, although the extension to estimating the distance modulus is, in principle, straightforward. Figure 4.9 shows the data that we analyze: each dot corresponds to a star, with accompanying measurement error. The colored dots in the figure represent the theoretical isochrones: our model essentially seeks a distribution over these curves that best represents the 47 Tucanae cluster. CMD-based estimates of the (single \log_{10}) age of 47 Tuc typically range from 9.95 to 10.10 (9.0–12.5 billion years). In light of this, we select the hyperparameters for the analysis to reflect the estimates and uncertainty ranges in the literature:

$$\mu_0 = 10.025, \quad \kappa_0 = \frac{9}{64}, \quad \nu_0 = 1000 \text{ and } \sigma_0^2 = 0.03^2. \quad (4.15)$$

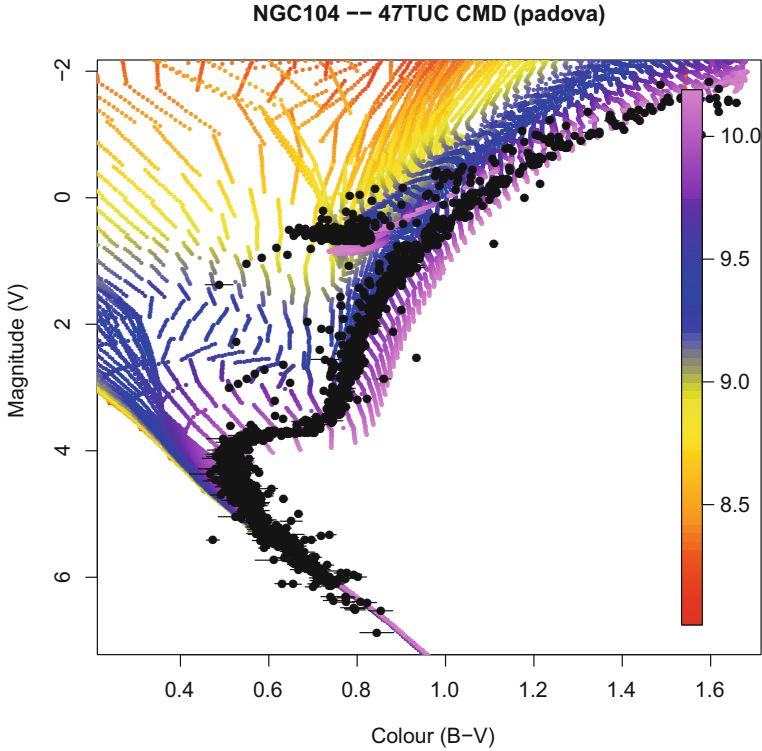


Fig. 4.9 Color-Magnitude Plot of the 47 Tucanae dataset. Each black dot represents a star in the dataset, colored points represent tabulated theoretical isochrone values. The color of each point represents its corresponding age, the mass, and metallicity of each point is not shown

These correspond to approximately

$$\mu_A \sim N(10.025, 0.08^2), \quad \sigma_A^2 \sim N(0.03, 8.1 \times 10^{-10}).$$

The analysis was performed using the Polygon-Cluster proposal distribution for the individual mass-age distributions, and a ladder of 8 logarithmically spaced tempering distributions. We run a total of 10 chains, each for approximately 24 h, and combine the results for estimation. Relevant convergence diagnostics were checked, but we omit the details for brevity. Figure 4.10 shows the posterior median and 95% intervals for each of the stars in the dataset, sorted by increasing posterior median. We clearly see a heavy left-tail: a collection of approximately 100 stars that appear to have a lower age than the rest of the cluster. An alternative cruder but simpler visualization is given by simply plotting a histogram of the posterior medians of the individual stars, as shown in Fig. 4.11.

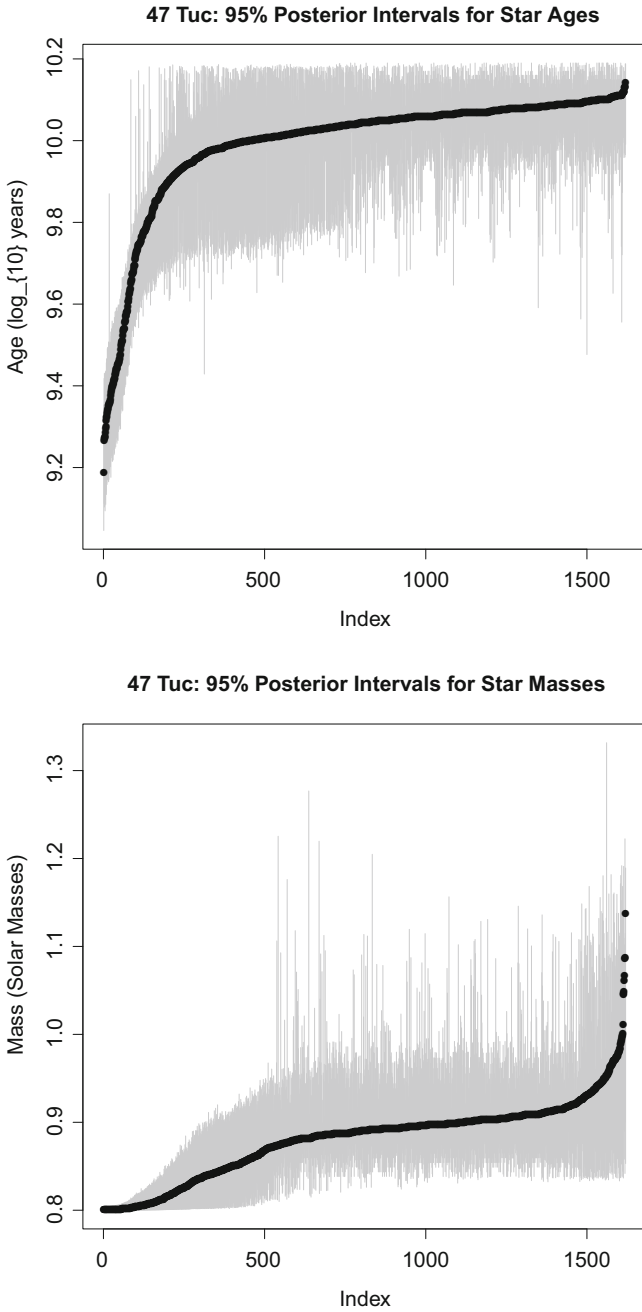


Fig. 4.10 Posterior intervals for the individual stellar ages a_i (Top) and masses m_i (Bottom). The stars are sorted in order of the posterior median, shown as a black dot: the accompanying 95% intervals are shown as gray bars

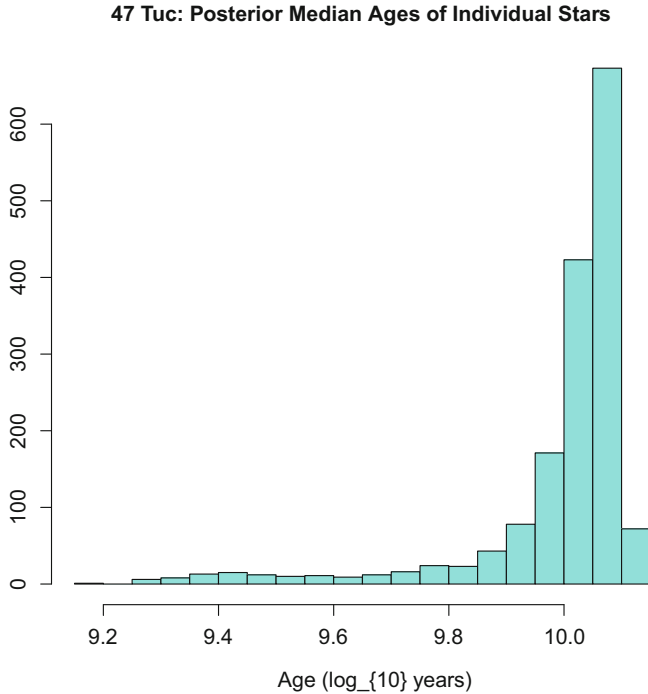


Fig. 4.11 Histogram of the posterior medians of all stars in the dataset. While providing a simpler visual, the distribution of posterior medians is a less complete representation than the posterior intervals of Fig. 4.10

The long-held belief has been that globular clusters are formed in a single burst from a single cloud of material. Based purely on the data and analysis here, however, there is a suggestion that 47 Tuc may contain multiple star formation bursts. Alternative explanations for the phenomena in Fig. 4.10 include contamination by foreground stars, misspecification or uncertainty in the distance modulus, or bias induced by extinction: see Sect. 4.5 for more details. Examining Fig. 4.10, there appear to be two bursts of star formation approximately 3 Gyr (billion) years apart at 7.9 Gyr and 11.5 Gyr ago, respectively. Recent independent work [1] using different techniques also suggests multi-burst SF in 47 Tuc, although analysis with higher quality multi-band photometric data would be required before drawing such scientific conclusions. Importantly, however, the flexibility in our model provides sufficient richness to be able to investigate previously untestable assumptions. Indeed, it is this additional flexibility and the appropriate modeling of uncertainty that is the primary contribution of statistical research in astrophysics.

4.5 Extensions and Future Work

Uncertainty in the “Black Box”

All of the previous analysis was predicated on the assumption that the “black box” describing the relationship between the physical parameters and the observed data (i.e., the isochrone mapping) is correct. In practice there is also uncertainty in these mappings, and we now consider some approaches to investigate this. Ideally, uncertainty in the mappings would be proliferated down through the mapping in the form of uncertainties in previously fixed quantities, i.e., essentially creating an expanded black-box/lookup table incorporating both different inputs and different physical assumptions. In practice, however, this is rarely feasible without access to the models that generate the lookup tables. In light of this, we consider a simpler problem: comparing competing sets of isochrones. For simplicity we consider comparison of two competing black boxes, although the extension to the comparison of more than two is straightforward.

Given two competing models, \mathcal{M}_1 and \mathcal{M}_2 , differing only by the specific choice of isochrone table, i.e., f in Eq. (4.1), we specify prior probabilities for each model: $p(\mathcal{M}_1)$ and $p(\mathcal{M}_2)$. In all cases here we begin with a neutral prior, selecting $p(\mathcal{M}_1) = p(\mathcal{M}_2) = 0.5$. The posterior model probabilities are then given by

$$p(\mathcal{M}_1|\mathbf{Y}) = \frac{1}{1 + \frac{p(\mathcal{M}_2) \int p(\theta) p(\mathcal{Y}|\theta, \mathcal{M}_2) d\theta}{p(\mathcal{M}_1) \int p(\theta) p(\mathcal{Y}|\theta, \mathcal{M}_1) d\theta}}, \quad (4.16)$$

thus requiring only additional computation of the ratio of normalizing constants for the two competing posterior distributions. More general model comparisons allow for different priors $p(\theta|\mathcal{M}_j)$, although in our application the prior $p(\theta)$ is the same for both models. Meng and Wong [15] show how one can estimate ratios of normalizing constants using the bridge sampling identity:

$$\frac{c_2}{c_1} = \frac{\mathbb{E}_1 [q_2(\theta)\alpha(\theta)]}{\mathbb{E}_2 [q_1(\theta)\alpha(\theta)]}, \quad (4.17)$$

where α is an arbitrary function providing a “bridge” between the two densities. They also provide a fast-converging iterative scheme to approximate the estimator under the optimal α . Note that the expectations of each unnormalized posterior are taken with respect to the other model. Therefore, if posterior samples are available for the two competing models, then implementing this model comparison boils down to the evaluation of the unnormalized posterior density for each draw from its rival model.

Multi-Cluster Models

As discussed in the context of the 47 Tucanae analysis, there are potentially applications where we want to allow for the possibility of multiple stellar clusters. The model defined by (4.1)–(4.4) is explicitly designed for single-cluster populations, although one possible generalization is conceptually straightforward. We could consider replacing (4.2) by an alternative mixture distribution:

$$A_i | H_i, \mu_{A, H_i}, \sigma_{A, H_i}^2 \stackrel{ind}{\sim} N\left(\mu_{A, H_i}, \sigma_{A, H_i}^2\right), \quad H_i \in \{1, 2, \dots, K\}, \quad (4.18)$$

where H_i is the cluster membership of star i . In most applications H_i would be given a uniform prior. When combined with identical priors on the cluster-specific hyperparameters, the posterior is defined only up to label switching. While, in principle, the number of clusters K could also be estimated, this would likely be fixed as part of the analysis. The additional computational burden induced by (4.18) rests primarily in the additional block correlations between the cluster- and individual-level variables.

Extinction and Non-ignorable Missingness

In many examples it is possible that observations for some stars are either partially (i.e., one or more bands) or fully (i.e., all bands) missing. The missing data mechanism for this missingness can potentially depend on the intrinsic brightness of the stars. That is, brighter stars are more likely to be observed than dimmer ones. Thus, the missing data mechanism can potentially provide information about the model parameters. For a given detector the detection/missingness probabilities are often well understood by careful calibration and testing. In such cases we often have access to a series of functions that express the probability of missingness as a function of the brightness of the star, a functional form that can then be coherently built into our hierarchical model. Again, within the Bayesian framework the extra layer can be added in a relatively straightforward way, although this will entail an additional computational burden. The importance of this missingness mechanism varies depending upon the type of stellar cluster being analyzed, and thus we currently restrict to those datasets where it is unlikely to affect the resulting inference.

Going Beyond Stellar Populations

Computer models and “black-box” likelihoods are increasingly common in many scientific disciplines, and can pose some interesting challenges to traditional com-

putational methods. In the case of analyzing stellar populations, that the likelihood is tabulated is both a blessing and a curse. We benefit in that much of the structure in the model is known *a priori*, and we show how an effective proposal distribution can be pre-computed independent of data. However, the black box proves to be a curse in that understanding and intuition are harder to come by, as are analytic simplifications and approximations. Despite this, one can construct an efficient and effective sampling scheme even for highly nonlinear and degenerate likelihoods that are more robust to the properties of the black box than naive methods.

The frequency of statistical applications involving components of the model that cannot be written down analytically is likely to increase in the coming years. There is much work to be done to better understand the computational and inferential implications of such models, and we hope the strategies and methods explored in this article can contribute to further research in this area.

Acknowledgments We dedicate this article to the memory of Stephen Fienberg, a stellar statistician and academic leader, who spent many decades unveiling the true colors of many black boxes in and outside of statistics.

We thank David van Dyk, Aneta Siemiginowska, and other members of CHASC for many helpful conversations and suggestions. PB and XLM acknowledge partial support from multiple NSF grants, and VLK from the NASA Contract to the Chandra X-ray Center NAS8-03060. AZ acknowledges support from the European Research Council under the European Union's Seventh Framework Programme (FP/2007–2013)/ERC Grant Agreement n. 617001 (A-BINGOS). This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie RISE action, grant agreement No. 691164 (ASTROSTAT).

References

1. J. Anderson, G. Piotto, I. R. King, L. R. Bedin, and P. Guhathakurta. Mixed populations in globular clusters: Et Tu, 47 Tuc? *The Astrophysical Journal*, 697:58–62, 2009.
2. J. Barnard, R. McCulloch, and X.-L. Meng. Modeling covariance matrices in terms of standard deviations and correlations, with application to shrinkage. *Statistica Sinica*, 10:1281–1311, 2000.
3. M. J. Bayarri, J. O. Berger, J. A. Cafeo, G. Garcia-Donato, F. Liu, J. Palomo, R. J. Parthasarathy, R. Paulo, J. Sacks, and D. Walsh. Computer model validation with functional output. *Annals of Statistics*, 35:1874–1906, 2007.
4. M. J. Bayarri, J. O. Berger, R. Paulo, J. Tu, J. Sacks, J. A. Cafeo, J. Cavendish, and C. H. Lin. A framework for validation of computer models. *Technometrics*, 49:138–154, 2007.
5. S. R. Cook, A. Gelman, and D. B. Rubin. Validation of software for Bayesian models using posterior quantiles. *Journal of Computational and Graphical Statistics*, 15(3):675–692, 2006.
6. A. Dotter, B. Chaboyer, D. Jevremovic, V. Kostov, E. Baron, and J. W. Ferguson. The Dartmouth stellar evolution database. *The Astrophysical Journal Supplement Series*, 178:89–101, 2008.
7. D. S. Evans. *Lacaille: astronomer, traveller; with a new translation of his journal*. Pachart, 1992.
8. E. D. Feigelson. Statistics in astronomy. *arXiv*, [astro-ph.IM](0903.0416v1):1–12, 2008.
9. C. J. Geyer. Markov chain Monte Carlo maximum likelihood. In *Computing Science and Statistics: Proceedings of the 23rd Symposium on the Interface*, pages 156–163, 1991.

10. C. O. Heinke, J. E. Grindlay, P. D. Edmonds, H. N. Cohn, P. M. Lugger, F. Camilo, S. Bogdanov, and P. C. Freire. A deep CHANDRA survey of the globular cluster 47 Tucanae: Catalog of point sources. *The Astrophysical Journal*, 625:796–824, 2005.
11. S. C. Kou, Q. Zhou, and W. H. Wong. Equi-energy sampler with applications in statistical inference and statistical mechanics. *Annals of Statistics*, 34-4:1581–1619, 2006.
12. E. Lawrence, D. Bingham, C. Liu, and V. N. N. Nair. Bayesian inference for ordinal data using multivariate probit models. Technical report, Misc, 2006.
13. Thibault Lejeune and Daniel Schaerer. Database of Geneva stellar evolution tracks and isochrones for (UBV)_J(RI)_CJHKLL'M, HST-WFPC2, Geneva and Washington photometric systems. *Astronomy & Astrophysics*, 366(2):538–546, 2001.
14. P. Marigo, L. Girardi, A. Bressan, M. A. T. Groenewegen, L. Silva, and G. L. Granato. Evolution of asymptotic giant branch stars. II. Optical to far-infrared isochrones with improved TP-AGB models. *Astronomy and Astrophysics*, 482:883–905, 2008.
15. X.-L. Meng and W. H. Wong. Simulating ratios of normalizing constants via a simple identity: a theoretical explanation. *Statistica Sinica*, 6:831–860, 1996.
16. R. J. Renka. Algorithm 751: TRIPACK: a constrained two-dimensional Delaunay triangulation package. *ACM Transactions on Mathematical Software*, 22(1):1–8, 1996.
17. J. Rougier, S. Guillas, and A. Maute. Expert knowledge and multivariate emulation: The thermosphere-ionosphere electrodynamics general circulation model (TIE-GCM). *Technometrics*, 51(4):414–424, Nov 2009.
18. E. E. Salpeter. The luminosity function and stellar evolution. *The Astrophysical Journal*, 121:161–167, 1955.
19. L. Sbordone, M. Limongi, A. Chieffi, E. Caffau, H.-G. Ludwig, and P. Bonafacio. Sulfur in the globular clusters 47 Tucanae and NGC 6752. *Astronomy and Astrophysics*, 503:121–127, 2009.
20. C. M. Schafer and P. B. Stark. Constructing confidence regions of optimal expected size. *Journal of the American Statistical Association*, 104:1080–1089, 2009.
21. D. A. van Dyk, S. DeGennaro, N. Stein, W. H. Jefferys, and T. von Hippel. Statistical analysis of stellar evolution. *Annals of Applied Statistics*, 3:117–143, 2009.
22. Y. Yu and X.-L. Meng. To center or not to center: That is not the question – an ancillarity-sufficiency interweaving strategy (ASIS) for boosting MCMC efficiency (with discussion). *Journal of Computational and Graphical Statistics*, 20(3):531–570, 2010.