Draft version December 3, 2021
Typeset using LaTeX twocolumn style in AASTeX63

Bayesian Inference of Globular Cluster Properties Using Distribution Functions

Gwendolyn M. Eadie,[1,2] Jeremy J. Webb,[1] and Jeffrey S. Rosenthal[2]

[1] *David A. Dunlap Department of Astronomy & Astrophysics, University of Toronto, 50 St George St, Toronto, ON M5S 3H4, Canada*
[2] *Department of Statistical Sciences, University of Toronto, 700 University Avenue, 9th Floor, Toronto, ON M5G 1Z5, Canada*

(Received April 29, 2021; Revised December 3, 2021)

Submitted to ApJ

ABSTRACT

We present a Bayesian inference approach to estimating the cumulative mass profile and mean squared velocity profile of a globular cluster given the spatial and kinematic information of its stars. Mock globular clusters with a range of sizes and concentrations are generated from lowered isothermal dynamical models, from which we test the reliability of the Bayesian method to estimate model parameters through repeated statistical simulation. We find that given unbiased star samples, we are able to reconstruct the cluster parameters used to generate the mock cluster and the cluster's cumulative mass and mean velocity squared profiles with good accuracy. We further explore how strongly biased sampling, which could be the result of observing constraints, may affect this approach. Our tests indicate that if we instead have biased samples, then our estimates can be off in certain ways that are dependent on cluster morphology. Overall, our findings motivate obtaining samples of stars that are as unbiased as possible. This may be achieved by combining information from multiple telescopes (e.g., *Hubble* and *Gaia*), but will require careful modeling of the measurement uncertainties through a hierarchical model, which we plan to pursue in future work.

Keywords: globular clusters: general — methods: data analysis — methods: statistical

1. INTRODUCTION

Globular clusters are nearly-spherical, massive collections of stars that are found in every type of galaxy. Upon formation, their early evolution is governed by stellar evolution in the sense that massive stars quickly lose mass, which causes the cluster's potential to weaken. However, over the majority of their lifetimes, two-body relaxation and the external tidal field of their host galaxy are the dominant mechanisms that govern a cluster's evolution (e.g. Heggie & Hut 2003). These two mechanisms lead to clusters becoming spherically symmetric, isotropic, and mass segregated over time as they evolve towards a state of partial energy equipartition while playing host to stellar collisions and mergers (Meylan & Heggie 1997; Spitzer 1987; Heggie & Hut 2003). Dynamically old clusters are even capable of having their core energetically decouple from the rest of the cluster, a process known as core collapse (Hénon 1961; Lynden-Bell & Wood 1968).

Given the bevy of dynamical processes that occur within globular clusters, the ability to accurately measure the current distribution of stars within a given cluster leads to a deeper understanding of how these processes work and shape cluster evolution. Reverse engineering the evolution of a system of clusters can then lead to constraining the conditions under which they form and therefore the formation and evolution of their host galaxy. A large number of distribution functions (DFs) have been proposed to represent the observed distribution of stellar positions and velocities in globular clusters (e.g., Woolley 1954; Michie 1963; King 1966; Wilson 1975; Gunn & Griffin 1979; Bertin & Varri 2008; Gieles & Zocchi 2015; Claydon et al. 2019). The general picture that emerges out of the models that best represent observations of Galactic globular clusters is that clusters are isotropic in their centre with density and velocity dispersion profiles that decrease to zero out to a truncation radius. The treatment of how the DF drops

Corresponding author: Gwendolyn M. Eadie
gwen.eadie@utoronto.ca

to zero out to the truncation radius varies from model to model, with additional treatments being necessary to address the presence of radial anisotropy (Michie 1963) and globular cluster rotation (Varri & Bertin 2012).

Complicating the situation slightly is that stars within globular clusters have a large range of masses, while most DFs assume all stars have the same mass. Hence mass segregation, which is a natural outcome of clusters evolving towards a state of partial energy equipartition, is not considered in the models. Failing to account for the presence of mass segregation has been shown to incur strong biases when fitting models to the surface brightness profile or number density profile of a cluster (Shanahan & Gieles 2015; Sollima et al. 2015). One solution is to treat a globular cluster system as the combination of several single mass models (Da Costa & Freeman 1976).

Historically, the application of the aforementioned models to observed globular clusters has been in the fitting of their observed number density or surface brightness profiles. From a given distribution function, it is possible to derive how the number of stars per unit area on the sky or volume decreases with clustercentric distance. Assuming a mass spectrum and mass-to-light ratio, a surface brightness profile can also be derived. Several different distribution function-based models have been successfully fit to Galactic (McLaughlin & van der Marel 2005; Miocchi et al. 2013; de Boer et al. 2019, e.g) and extragalactic (Woodley & Gómez 2010; Usher et al. 2013; Webb et al. 2013; Puzia et al. 2014, e.g) globular clusters.

Alternatives to fitting clusters with distribution function based models include comparing observations to large suites of N-body star cluster simulations (Heggie & Giersz 2014; Baumgardt & Hilker 2018) and Jeans Modelling (Cappellari 2008; Watkins et al. 2013). Direct N-body simulations can also be used to test and rule out different distribution function based models, as completeness, contamination and measurement errors will not contribute to the uncertainty in the fit. For example, Zocchi et al. (2016) successfully demonstrated that direct N-body simulations of star clusters could be well fit by the lowered isothermal models of Gieles & Zocchi (2015).

In addition to the issues associated with assuming what model best represents globular clusters in general, the process of finding the exact model parameters (or N-body simulation) that best represent a specific globular cluster is also challenging. Historically, globular clusters were fit with models by comparing observed and theoretical surface brightness profiles or density profiles (e.g, McLaughlin & van der Marel 2005). A typical approach to fitting observational data with models would be to radially bin the observed stars and then minimize the χ^2 between the observed surface brightness or density profile and the model profile. Such an approach will result in systematic error due to binning the data, with the completeness of the dataset, contamination from non-cluster stars, and measurement errors introducing additional uncertainty into the fit as well. Binning data is also undesirable as information is lost about each individual star. Furthermore, as previously mentioned, multi-mass models require either a mass-to-light ratio be added as a free parameter when fitting surface brightness profiles or a mass-to-light ratio be assumed for the observational data (Hénault-Brunet et al. 2019).

Gaia Data Release 2 (Gaia Collaboration et al. 2016, 2018) and the Hubble Space Telescope Proper Motion (HSTPROMO) Survey (Bellini et al. 2014) have helped usher in a new era of globular cluster studies, with spatial and kinematic information now available for a large number of cluster stars. Knowing the kinematic properties of individual stars can mitigate uncertainties related to contamination, as kinematics make it easier to determine what stars in the observed field of view are truly members of the cluster or are simply foreground or background stars. Combining membership constraints with spatial and photometric information of core stars in high-resolution images of cluster centres also allows for the radial coverage across a cluster to be improved (de Boer et al. 2019).

Kinematic information can also be taken into consideration when fitting clusters with models, as the cluster's density profile and velocity dispersion can be simultaneously fit by minimizing the combined χ^2 (Baumgardt & Hilker 2018). Extending the method even further, Zocchi et al. (2017) has fit lowered isothermal models to the Galactic globular cluster Omega Centauri by simultaneously fitting its surface brightness profile, line of sight velocity dispersion profile, radial proper motion dispersion profile, and tangential proper motion dispersion profile. Unfortunately, even with kinematic information, issues related to binning data, completeness, and measurement uncertainties remain when fitting data with models. Furthermore, when trying to simultaneously fit surface brightness profiles and kinematic profiles, one must assume how to weight the importance of each fit. For example, when fitting through the minimization of χ^2 between model and observed data, it must be decided whether the total χ^2 is simply the sum of the individual χ^2 values calculated for the density and kinematic profile fits or if they should be weighted differently. The advantages and disadvantages of fitting each of the models discussed above to observed cluster datasets are summarized by Hénault-Brunet et al. (2019).

The purpose of this study is to investigate and potentially improve the method in which distribution function-based models can be fit to observed star cluster datasets by avoiding systematic errors and loss of information associated with radially binning the data, contamination, and completeness. We instead estimate the model parameters, cumulative mass profile, and mean-square velocity profile of a globular cluster (GC) using the positions and velocities of individual stars and assuming a physical model for the GC through a DF and Bayesian method.

A Bayesian framework has at least four main advantages for this type of analysis. First, we wish to incorporate useful prior information about GCs to help constrain parameter estimates. Second, since kinematic data for GCs is often incomplete, using a Bayesian framework allows one to include both incomplete and complete data simultaneously. Third, astronomical data are also subject to measurement uncertainties that are well understood by astronomers, and that we can incorporate via a hierarchical Bayesian framework. Fourth, our ultimate goal is to infer the cumulative mass profile without having to make assumptions about the mass-to-light ratio of the GC, and this should be achievable given samples from the posterior distribution of model parameters.

For the current study, we work with simulated data generated using `limepy` (Gieles & Zocchi 2015) of lowered isothermal models for GCs and test the ability of a Bayesian framework to recover a cluster's true total mass, cumulative mass profile, mean-square velocity profile, and other parameters of interest. A related study was completed by Hénault-Brunet et al. (2019), where they used a single snapshot from a direct N-body simulation of the Galactic GC M4 (Heggie & Giersz 2014) to compare the ability of multiple *methods* to recover the simulated cluster's mass and mass profile. In the current paper, rather than comparing and contrasting the pros and cons of different methodological approaches on a single snapshot, we study the pros and cons of a single method to recover the mass profile of different types of of globular clusters (e.g., "average", "compact", "extended" GCs). This approach is especially important, as Hénault-Brunet et al. (2019) suggested that single-mass DF methods could lead to biases in the mass and mass profile. We would like to concretely quantify any possible biases, and identify whether they are dependent on certain types of GCs (e.g., average, compact, and extended).

The paper is structured as follows. In Section 2, we introduce the suite of simulated data used to test our approach, with the fitting routine and methods described in Section 3. In Section 4, we examine the estimated coverage probabilities of the Bayesian credible intervals for the model parameters, and discuss situations in which inference from the posterior distribution is (and is not) able to reproduce the true cumulative mass profile and mean-square velocity profile of the simulated GCs. Future applications of this work, including the use of observational data, are also discussed. Finally, we summarize our findings in Section 5.

2. SIMULATED DATA

We develop and test our method for GC parameter inference with simulated kinematic data $\boldsymbol{d} = (\mathbf{r}, \mathbf{v})$ of stars in a GC-centric reference frame, where $r_i = \sqrt{x_i^2 + y_i^2 + z_i^2}$ and $v_i = \sqrt{v_{x,i}^2 + v_{y,i}^2 + v_{z,i}^2}$ are the distance and speed of the i^{th} star. The data are generated using the python code `limepy` (Gieles & Zocchi 2015), which uses a four-parameter model for the phase-space distribution function $f(\boldsymbol{r}, \boldsymbol{v})$ of stars in the cluster (see Section 3). The `limepy` parameters are

$$\boldsymbol{\theta}_{\text{limepy}} = (g, \Phi_0, M_{total}, r_h) \tag{1}$$

where g (dimensionless) is a truncation parameter, Φ_0 (dimensionless) determines the central potential, M_{total} (in $M_\odot$) is the total mass, and r_h (in parsecs, pc) is the half-light radius. Overall, g and Φ_0 impact the shape of the GC profile, while M_{total} and r_h are scale parameters. In the case of isotropic GCs, a value of $g = 0$ in the `limepy` model is equivalent to the Woolley (1954) model, and a value of $g = 1$ is equivalent to the King models (Michie 1963; King 1966, see also Gieles & Zocchi 2015). The value of g is not only a truncation parameter but also plays a role in determining the spatial distribution of stars. The parameter Φ_0 — which determines the central gravitational potential — helps set the concentration of stars.

GCs with the same M_{total}, g, and Φ_0, but with different half-light radii r_h, have relatively different levels of *compactness*. That is, a GC with a small half-light radius is much more compact than a GC with a large half-light radius. At the same time, GCs with the same M_{total}, g, and r_h, but different Φ_0 values have relatively different *concentrations*. Lower (higher) Φ_0 values lead to a larger (smaller) concentrated region of stars at the GC center.

Figure 1 shows examples of GCs with different levels of compactness and concentrations; the figure shows examples of the clustercentric x and y positions of GC stars (first and third rows) and the magnitude of the stars' velocities as a function of distance r from the center of the cluster (second and fourth rows). The three GCs shown in the top two panels of Figure 1 have the same

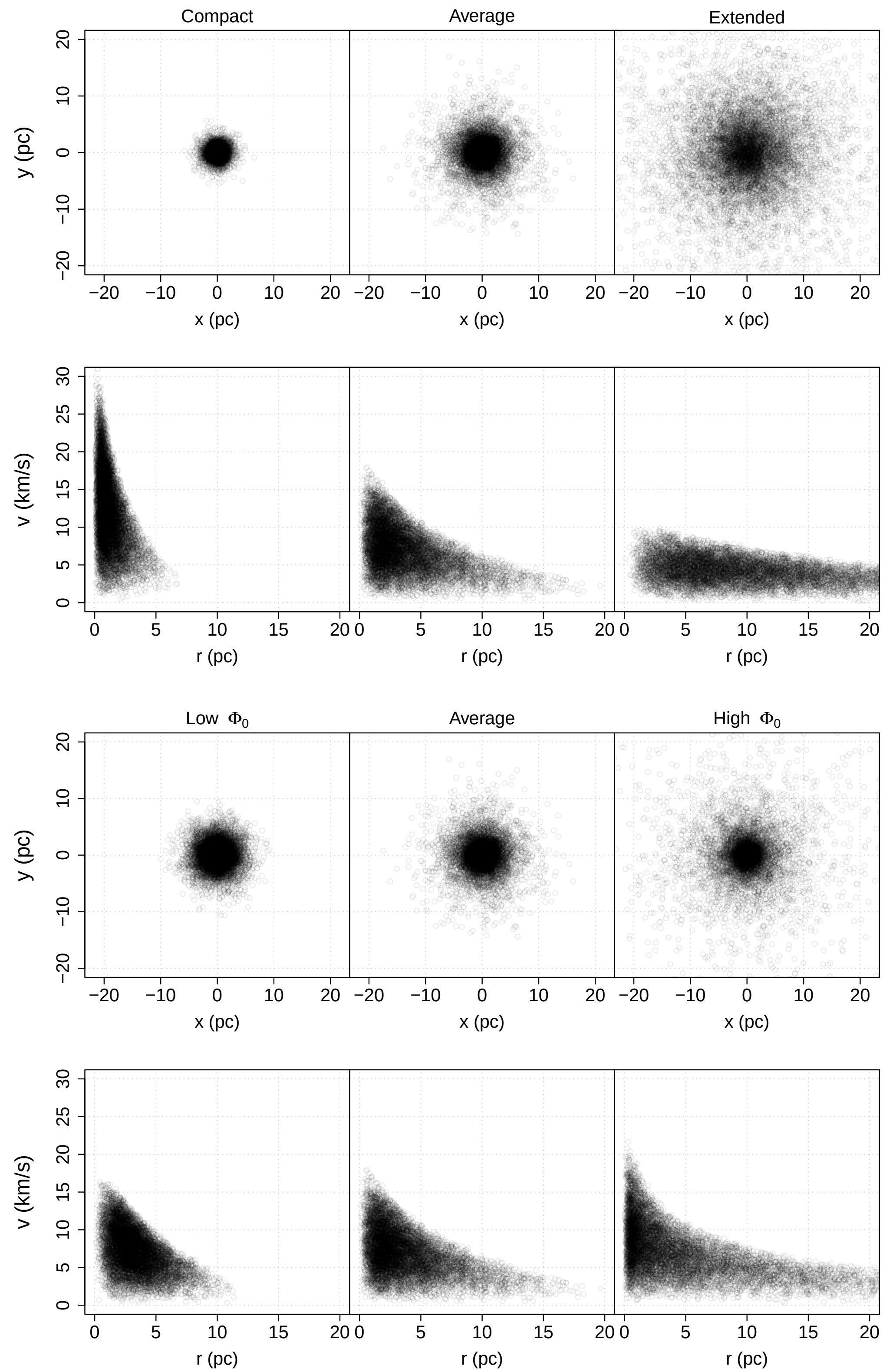

Figure 1. *First row:* The x and y coordinates of ten thousand randomly selected stars in three different simulated GCs: a compact (left), average (middle), and extended (right) GC with 10^5 stars and `limepy` parameters $g = 1.5$, $\Phi_0 = 5.0$, $M = 10^5$, and $r_h = 1.0, 3.0$ and 9.0 respectively. *Second row:* The magnitude of each star's velocity (semi-transparent circles) as a function of total distance r, using the same stars as in the top row. *Third and fourth rows:* The same as the top two rows, except for a GC with parameters $g = 1.5$, $M = 10^5$, and $r_h = 3.0$, and changing the Φ_0 parameter: $\Phi_0 = 2.0$ (left), average (middle, same as top two rows), and $\Phi_0 = 8.0$.

M_{total}, g, and Φ_0 values, but have increasing half-light radii r_h from left to right. In the bottom two panels of Figure 1, the GCs have the same g, M_{total} and r_h, but have increasing Φ_0 values from left to right. Note that the "average" GC is shown in the center of all rows of Figure 1, to show the transition from low r_h to high r_h and from low Φ_0 to high Φ_0. From Figure 1, we see that either a low (high) r_h or a low (high) Φ_0 leads to subtle differences in positional space but noticeably different distributions in velocity.

In this work, we explore different GC morphologies based on the parameter values listed in Table 1 (i.e., the types shown in Figure 1). Every simulated GC has the same total mass ($M_{total} = 10^5 M_\odot$) and truncation parameter $g = 1.5$, but has a different level of *compactness* (different r_h) or different *concentration* (different Φ_0). To simplify our terminology we refer to the five scenarios in Table 1 as: *compact* (small r_h), *average*, *extended* (large r_h), *Low* Φ_0, and *High* Φ_0. We create 50 GCs of each type in order to repeat our analysis many times.

GC Type	Parameter Values			
	g	Φ_0	M_{total} ($10^5 M_\odot$)	r_h (pc)
average	1.5	5.0	1.0	3.0
compact	1.5	5.0	1.0	1.0
extended	1.5	5.0	1.0	9.0
high Φ_0 GC	1.5	8.0	1.0	3.0
low Φ_0 GC	1.5	2.0	1.0	3.0

Table 1. Summary of `limepy` parameter values used to simulated different GC types analysed in this study.

Each simulated GC contains $N = 10^5$ stars. In real data sets, we do not have kinematic information for all n stars due to limited observations and observational selection effects. Thus, we study the effects of our mass profile estimates when selecting stars (a) randomly, (b) only in the outer regions (thereby mimicking *Gaia* data), and (c) only in the inner regions (thereby mimicking *HST* data). In each case, we use a subsample of 500 stars from each GC. Moreover, these three different tests, combined with the five different morphological GCs (Table 1), leads to fifteen different scenarios.

For this initial study and for the development and testing of our code, we use complete data in both position and velocity and assume there is no measurement uncertainty. We also work in the reference frame of the GC, where positions and velocities of individual stars are given with respect to the GC center. Of course, real data are collected in a Heliocentric reference frame, may be incomplete (e.g., only projected distances and line-of-sight velocities are known), and are subject to measurement uncertainty. However, it is worthwhile to investigate the ability of this method in an idealized case where we have complete data. Ultimately, our goal is to work in projected space on the plane of the sky (i.e., the reference frame in which actual data are measured), account for incomplete data (e.g., only one component of the velocity is known), and incorporate measurement uncertainty through a hierarchical model.

3. METHODS

Using the simulated spatial and kinematic data of stars from each GC mentioned in Section 2, we take a Bayesian approach to infer the model parameters of each GC. From Bayes' theorem (Bayes 1763), the posterior probability of a vector of model parameters $\boldsymbol{\theta} = (g, \Phi_0, M_{total}, r_h)$, given data $\boldsymbol{d}$, is

$$p(\boldsymbol{\theta}|\boldsymbol{d}) = \frac{p(\boldsymbol{d}|\boldsymbol{\theta})p(\boldsymbol{\theta})}{p(\boldsymbol{d})}, \tag{2}$$

where $p(\boldsymbol{d}|\boldsymbol{\theta})$ is the probability of the data conditional on the model parameters, $p(\boldsymbol{\theta})$ is the prior distribution on the model parameters, and $p(\boldsymbol{d})$ is the "evidence" or *prior predictive density*. The latter is a constant, leaving us with a target distribution proportional to the posterior distribution $p(\boldsymbol{\theta}|\boldsymbol{d})$, which we will estimate through sampling in order to perform parameter inference (Section 3.3). Our simulated data $\boldsymbol{d}$ described in Section 2 are the six Cartesian phase-space components $(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z}, \boldsymbol{v_x}, \boldsymbol{v_y}, \boldsymbol{v_z})$ of each star, which we treat as perfectly measured. An individual star's phase-space components $d_i = (x_i, y_i, z_i, v_{x,i}, v_{y,i}, v_{z,i})$ provide its cluster-centric distance r_i and speed v_i, which are needed for the calculation of the DF $f(\boldsymbol{\theta}; d_i)$.

In practice, $p(\boldsymbol{d}|\boldsymbol{\theta})$ is often taken to be the likelihood — a function of model parameters for fixed data $\mathcal{L}(\boldsymbol{\theta}; \boldsymbol{d})$ — which we define using the DF in Section 3.1. The prior distributions for the model parameters $\boldsymbol{\theta} = (g, \Phi_0, M_{total}, r_h)$ in the `limepy` model are described in Section 3.2.

3.1. *Likelihood*

In this study, we define the likelihood using a physical distribution function (DF), $f(\boldsymbol{\theta}; d_i)$ of the `limepy` lowered-isothermal model. Given a fixed set of data $\boldsymbol{d}$ of N stars, the likelihood is a function of the model parameters $\boldsymbol{\theta}$ and the total mass M_{total} of the GC:

$$\mathcal{L}(\boldsymbol{\theta}; \boldsymbol{d}) = \prod_{i=1}^{N} \frac{f(\boldsymbol{\theta}; d_i)}{M_{total}} \tag{3}$$

$$= \prod_{i=1}^{N} \frac{f(g, \Phi_0, M_{total}, r_h; r_i, v_i)}{M_{total}}, \tag{4}$$

where the stars are assumed to be independent.

For lowered-isothermal models, the DF f is calculated numerically via the `limepy` software (Gieles & Zocchi 2015), and thus the likelihood must be calculated numerically too.

As mentioned in Section 2, we simulate position and kinematic data of stars following a `limepy` model DF with parameters $\boldsymbol{\theta}$ shown in Table 1, assume the likelihood defined in equation 4, and define physically-motivated informative priors on the model parameters.

Given that the likelihood is defined by the DF that was used to generate the data, we expect to obtain reasonable parameter estimates through inference made from the posterior distribution using Markov Chain Monte Carlo (MCMC) sampling. However, we are also going to impose prior distributions that are at least weakly informative, and so it is good practice to test whether the posterior can still be used to reliably infer the model parameters. Moreover, in the cases where the sampling of stars from the cluster is biased to inside the core or outside the core, we aim to understand how this sampling bias affects parameter inference.

3.2. *Prior Distributions*

Two advantages of Bayesian inference are the necessity to incorporate meaningful prior information, and the requirement to state this explicitly. In order for the DF to correspond to a physically realistic collection of stars in a GC, all model parameters must be greater than zero. Negative parameter values are not allowed by the likelihood, but we also disallow negative parameter values via the priors (this increases efficiency and keeps the `limepy` model from returning errors).

One reason to use informative priors is that images and studies both within the Milky Way Galaxy and around other galaxies provide prior information on quantities like the mass and half-light radius of GCs. For example, GC masses span about an order of magnitude and most astronomers would be comfortable setting the prior $p(\log_{10} M_{total}) \sim N(\mu_M, \sigma_M)$, where the hyperparameters[1] μ_M and σ_M are defined in $\log_{10} M_{total}$. This is the prior we choose, and it is also supported by the near universal GC mass function (Brodie & Strader 2006; Harris 2010).

The `limepy` model works in M_{total} space, so we need to do a change of variables to obtain the prior $p(M_{total})$. Using a change of variables, the prior on M_{total} is

$$p(M_{total}) = \frac{N(\mu_M, \sigma_M)}{M_{total} \ln 10}. \qquad (5)$$

[1] the term hyperparameters is used to differentiate μ_M and σ_M from the model parameters of interest

The half-light radius is another quantity of GCs for which we have considerable prior information. Images of GCs give an independent estimate of r_h, with a conservative measurement uncertainty of roughly 0.4pc (e.g. de Boer et al. 2019). In this simulation study, we assume the observer has this prior information and set a truncated normal prior on r_h.

We have considerably less prior information on the values of g and Φ_0, aside from the physically allowable, positive values. For these parameters, we use truncated uniform distributions. In summary, we assume the parameters for the `limepy` model are distributed as

$$g \sim \text{unif}(0.001, 3.5), \qquad (6)$$

$$\Phi_0 \sim \text{unif}(1.5, 14), \qquad (7)$$

$$M_{total} \sim \frac{N(\mu_M, \sigma_M)}{M_{total} \ln(10)}, \qquad (8)$$

$$\text{and } r_h \sim N(a, b, \mu_{r_h}, \sigma_{r_h}), \qquad (9)$$

where $\mu_M = 5.85$ and $\sigma_M = 0.6$ (defined in $\log_{10} M_{total}$), and hyperparameters for the lower and upper bounds of r_h are $a = 0$ and $b = 30$ respectively. The mean and standard deviation for the r_h parameter (μ_{r_h} and σ_{r_h}) are chosen to reflect plausible information an observer would have for a given GC. Thus, for the average GCs in our analysis, we try different means, such as $\mu_{r_h} = 3.4$, $\mu_{r_h} = 3.1$, etc. with $\sigma_{r_h} = 0.4$pc. Our results are insensitive to the choice of the mean, as long as it is not too many standard deviations away from the true value.

3.3. *Sampling the Target Distribution*

Given the `limepy` model, we have a likelihood function $\mathcal{L}(g, \Phi_0, M, r_h\,; \boldsymbol{d})$ for the four unknown parameters, depending on the observed star data $\boldsymbol{d}$. Combining the above prior distributions with this `limepy` likelihood function leads to a posterior distribution or *target posterior density* via Bayes' theorem (eq. 2),

$$p(g, \Phi_0, M_{total}, r_h|\boldsymbol{d}) \propto p(\boldsymbol{d}|g, \Phi_0, M_{total}, r_h) \times p(g)p(\Phi_0)p(M_{total})p(r_h),$$

where we assume independent priors. Our goal is to sample from the target distribution $p(g, \Phi_0, M_{total}, r_h|\boldsymbol{d})$, and perform inference of the parameter values, the cumulative mass profile, and the mean-square velocity profile of the GC.

Ultimately, we explore and collect samples of this posterior density using a MCMC algorithm, specifically a version of the standard Metropolis algorithm (Metropolis et al., 1953) that includes automated, finite adaptive tuning (to be discussed later). First, however, we find optimal starting values; we use the differential evolution optimizer function `DEopt` from the `NMOF` package

(Schumann 2011–2021; Gilli et al. 2019) in **R** (R Core Team 2019) to find modal (i.e., argmax) values of the four parameters, and then use these values as the initial state of our MCMC algorithm. Differential evolution was first introduced by Storn & Price (1997), and we refer the reader to this paper for details on the algorithm. This initial step allows an automated selection of good starting values, which helps to overcome the complicated structure of the posterior distribution, thereby making sampling more efficient. Once the starting values are obtained, we run an automated, finite adaptive-tuning method during the burn-in of the Markov chain. To describe the finite adaptive-tuning method, we first provide a brief review of proposal distributions and sampling efficiency.

Sampling a target or posterior distribution using a standard Metropolis algorithm requires a choice of proposal or "jumping" distribution. The latter is used to randomly suggest a new place in parameter space, θ^*, based on the current location θ_i. Often, this suggestion is done using a normal distribution such that

$$\boldsymbol{\theta}^* = \boldsymbol{\theta}_i + \boldsymbol{Z}, \tag{10}$$

where $\boldsymbol{Z} \sim N(0, \boldsymbol{\Sigma})$. Here, $N(0, \boldsymbol{\Sigma})$ is the jumping distribution with a covariance matrix $\boldsymbol{\Sigma}$ set by the user. The value of Σ determines whether, on average, "big jumps" or "small jumps" are attempted from the current location of θ_i. These proposed jumps are either accepted or rejected according to the standard formula in the Metropolis algorithm. The efficiency of the sampling is dependent on the choice of this covariance matrix. For example, if the variance is too small then the algorithm make jumps that are too small. If the variance is too large, then the algorithm will make jumps that are too large.

Finding a $\boldsymbol{\Sigma}$ that enables the most efficient sampling is sometimes accomplished through manual *tuning*: adjusting $\boldsymbol{\Sigma}$ until the appropriate acceptance rate is achieved. Obviously, this can be a tedious and time-consuming process, especially in the case of multiple parameters. Thankfully, there are methods which automate this task and that are founded in statistical theory.

In this paper, we use an *automated, finite adaptive-tuning* method during the burn-in of the Markov chain. This adaptive-tuning method is one in which the proposal step sizes are adjusted automatically and iteratively. We obtain a good covariance matrix for the proposal distribution using an Adaptive Metropolis algorithm (Haario et al. 2001; Roberts & Rosenthal 2009) which repeatedly updates the Metropolis proposal distribution (i.e., the proposal covariance matrix) based on the empirical covariance of the run so far, in an effort to obtain a proposal covariance matrix equal to about $(2.38)^2$ times the target covariance matrix divided by the Markov chain's dimension, which has been shown to be optimal under appropriate assumptions (Roberts & Rosenthal 1997, 2001). Foundational works on the subject of adaptive Metropolis and convergence are found in the statistics literature (Haario et al. 2001; Roberts et al. 1997; Roberts & Rosenthal 2009).

The practice of using the Adaptive Metropolis algorithm for an *initial* run and then fixing the proposal variance for the final run corresponds to "finite adaptation" as in Proposition 3 of Roberts & Rosenthal (2007). We require a minimum of five initial runs to update the proposal variance, but also automatically allow for further iterations as needed to achieve efficient sampling. Almost all of the GCs we analyze take no more than five iterations of the finite adaptive tuning, which takes one to five minutes per cluster on a simple laptop computer.

Once the finite adaptive step is complete, we run a standard Metropolis algorithm using the final (hopefully approximately optimal) proposal distribution found by the Adaptive Metropolis step. The final sampling takes less than 15 minutes per cluster to complete. At the end, we discard an initial burn-in period, and take the remaining chain values as a sample from the posterior density.

The above procedure allows us to approximately sample from $p(g, \Phi_0, M_{total}, r_h|\boldsymbol{d})$, and hence (a) approximately compute the posterior means and other statistics of the four unknown parameters $(g, \Phi_0, M_{total}, r_h)$, including Bayesian credible intervals, and (b) calculate a cumulative mass profile of the GC for every sample from the target distribution.

3.4. *Different Cluster and Sampling Cases*

Very generally, GCs may be classified as having an average, compact, or extended morphology based on their radius r_h, or may be considered to have high or low concentration based on the value of Φ_0. Additionally, the spatial and kinematic data from stars may be a random sample from everywhere in the cluster, a random sample beyond some radius, or a random sample within some radius. We expect the ability of our method to recover the true mass, cumulative mass profile, and mean-square velocity profile to depend on both GC morphology and the type of sampling of its stars. Understanding the bias in parameter inference that can occur as a result of biased sampling is important, since in reality we sometimes lack position and kinematic data from the inner or outer regions of the cluster. Thus, we investigate multiple combinations of the aforementioned cases to understand any possible bias.

Table 1 summarize the types of GCs we investigate. In our simulated GCs, all stars have the same brightness and mass, and so the half-light radius corresponds to the half-mass radius. For each case, we simulate 50 GCs using the parameter values listed in Table 1, and subsample 500 stars either (1) randomly, (2) outside r_{cut}, or (3) inside r_{cut}. We choose an r_{cut} value of 1.5pc mostly for simplicity but also partly because recent work by the HSTPROMO Team indicates that proper motions are most often available for stars within the half-mass radius (Watkins et al. 2013) but not beyond. Our conservative choice for r_{cut} is therefore half of the average effective radius of Galactic clusters (excluding very extended clusters with effective radii greater than 10 pc) (Baumgardt & Hilker 2018). We use this same cut-off radius when sampling outer stars (i.e., situation (2) above) as well. In this way, we investigate what happens when data are only available for outer stars (e.g., when Gaia kinematic data are used).

By repeating the analysis on 50 randomly generated GCs, we estimate and examine the coverage probabilities for the Bayesian credible regions for all sampling scenarios, for all GCs listed in Table 1 (Section 4).

For example, for the average cluster, we generate 50 simulated GCs with parameter values $g = 1.5, \Phi_0 = 5, M_{total} = 10^5 M_{\odot}$, and $r_h = 3.0$pc, and randomly sample 500 stars from each GC. For each GC, we run the analysis on the subsample of stars, obtaining samples of the target distribution as described in the previous section. Next, we estimate the mean, interquartile range, and 95% credible interval of the posterior distribution using our MCMC samples from the target distribution. After doing this for all 50 average GCs, we count how many times the interquartile ranges and 95% credible intervals cover the true parameter value to estimate the coverage probability. If the Bayesian credible regions are reliable, then the interquartile ranges should cover the true parameter values 50% of the time, and the 95% credible intervals should cover the true parameter values 95% of the time.

The same procedure is repeated for all GC types listed in Table 1. For example, we look at GCs with different half-light radii, reflecting extended and compact clusters. For these clusters we use parameter values of $g = 1.5, \Phi_0 = 5, M_{total} = 10^5 M_{\odot}$, and $r_h = 9.0$pc and $g = 1.5, \Phi_0 = 5, M_{total} = 10^5 M_{\odot}$, and $r_h = 1.0$pc respectively. To further explore the parameter space believed to be covered by Galactic GCs, and specifically to explore GCs that are more (less) concentrated, we also look GCs with a high (low) Φ_0.

Using our estimate of the posterior distribution for a single GC, we can also estimate that GC's cumulative mass profile (CMP). The CMP is an estimate of the mass contained within some distance r of the GC. To estimate the CMP, we follow the same procedure as described in Eadie & Jurić (2019), who used this approach to estimate the Milky Way's CMP. For every set of model parameters $(g, \Phi_0, M_{total}, r_h)$ sampled by our algorithm (i.e., every row of parameter values in the Markov chain), we calculate the cumulative mass profile determined by the `limepy` model. Because we have 1000s of rows in our Markov chain, we obtain thousands of CMP estimates. These CMPs provide us with a visual and quantitative estimate that can be used to calculate Bayesian credible regions and that can be compared directly to the true CMP of the cluster.

Another quantity of interest that we can estimate using the posterior distribution for a single GC is the *mean-square velocity*; the mean-square velocity is equal to the sum of the velocity dispersion squared (or the variance) and the square of the mean velocity. The `limepy` code can calculate the mean-square velocity as a function of radius from the centre of the cluster, given a specific set of model parameters. Thus, the estimate of the GC's mean-square velocity profile can be calculated in much the same way as the CMP, using the parameter samples from the posterior distribution.

In all of the GC examples, we assume that we know the complete position and velocity components of the stars. However, in reality we often have incomplete data. For example, we may only have projected measurements on the plane of the sky (i.e., projected distances in the $x-y$ plane, and proper motions). This missing data may influence our mass and mass profile estimates in unexpected ways, and is important to study. In a Bayesian analysis one can treat the missing components as parameters in the model, but this also means that further prior distributions must be set. Given the complexity of the problem, we leave this to future work.

4. RESULTS & DISCUSSION

4.1. *Random Sampling*

For the cases in which we randomly sample stars from everywhere in the cluster, we find the Bayesian credible regions to be reliable for all five GC types.

As an example, Figure 2 shows the 95% credible intervals (error bars) for each model parameter, for 50 realizations of an average cluster. The true parameter values are shown as vertical blue lines, and the number of times out of 50 that the 95% credible interval of the target distribution overlaps the true value is shown at the top of each panel. We can see that the credible

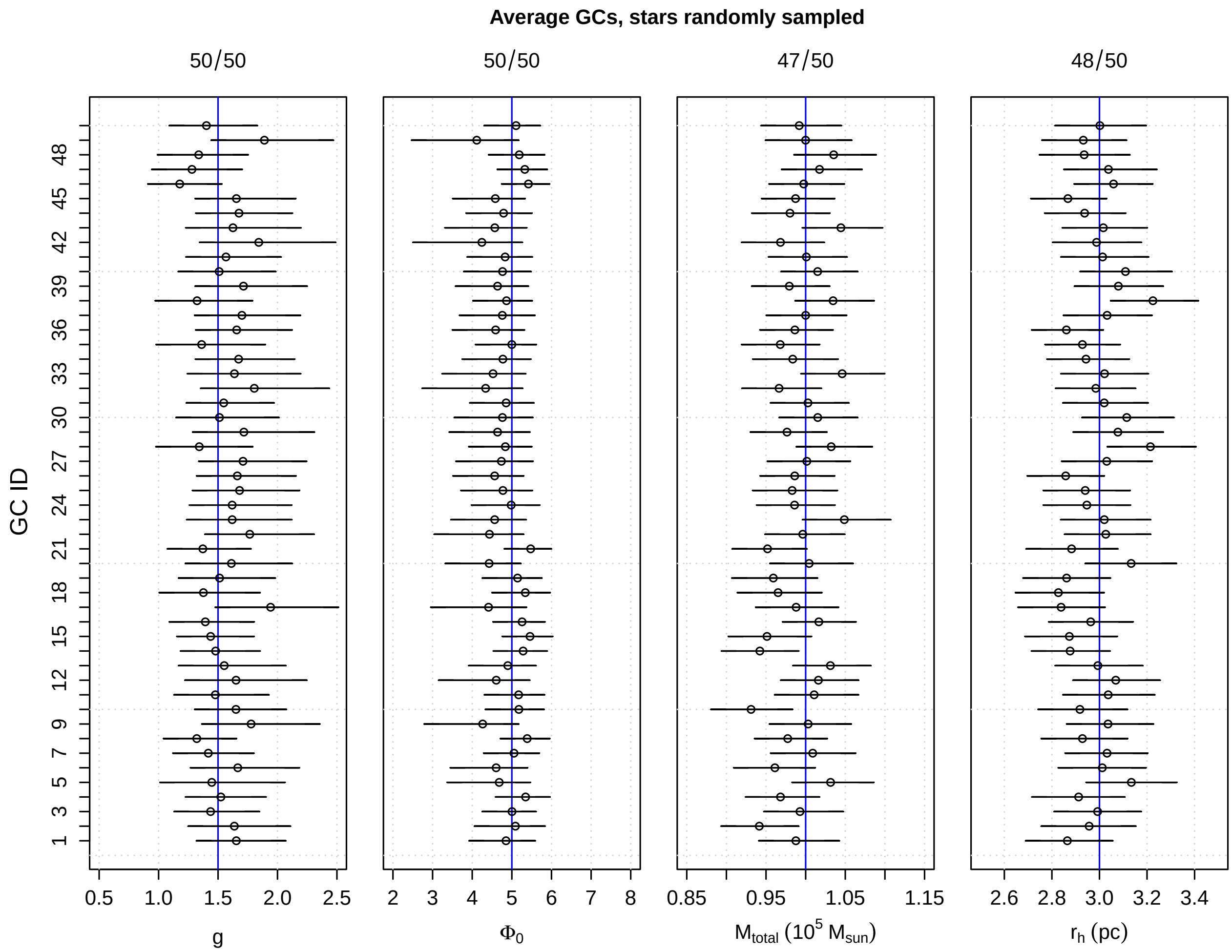

Figure 2. The parameter estimates and 95% credible intervals for fifty simulated "average" GCs. Each panel shows 50 credible intervals (error bars), the corresponding mean (points), and the true parameter value (vertical blue line). Each *row* of points across the four panels corresponds to the parameter estimates for the GC with ID given on the vertical axis. The fraction at the top of each panel indicates the number of times the 95% credible interval overlaps the true parameter value. The fractions are very large, as they should be for 95% intervals.

intervals for each parameter reliably contains the true parameter approximately 95% of the time (Figure 2).

As a second example, we show a similar plot for the case of the extended GCs (Figure 3). Here too, we find the 95% credible intervals to be reliable for the most part. The credible intervals for g and Φ_0 are slightly overconfident, since the true parameter value lies within the 95% credible intervals only 90% and 92% of the time respectively.

As a final and third example, Figure 4 shows the same type of plot for a more concentrated cluster with $\Phi_0 = 8$. Again, the credible intervals are reliable, showing good coverage probabilities.

Table 2 shows the estimated coverage probabilities for the M_{total} parameter in the case of random sampling, for all five types of clusters, found by calculating the fraction of times that the true M_{total} is contained within

GC Type	C.I.	Coverage Prob. for M_{total}
average		0.50
compact		0.42
extended	50%	0.52
high Φ_0		0.48
low Φ_0		0.38
average		0.94
compact		0.90
extended	95%	1.00
high Φ_0		0.94
low Φ_0		0.92

Table 2. *Estimated coverage probabilities under the random sampling case, for different GC morphologies. In the table heading, C.I. stands for credible interval.*

.

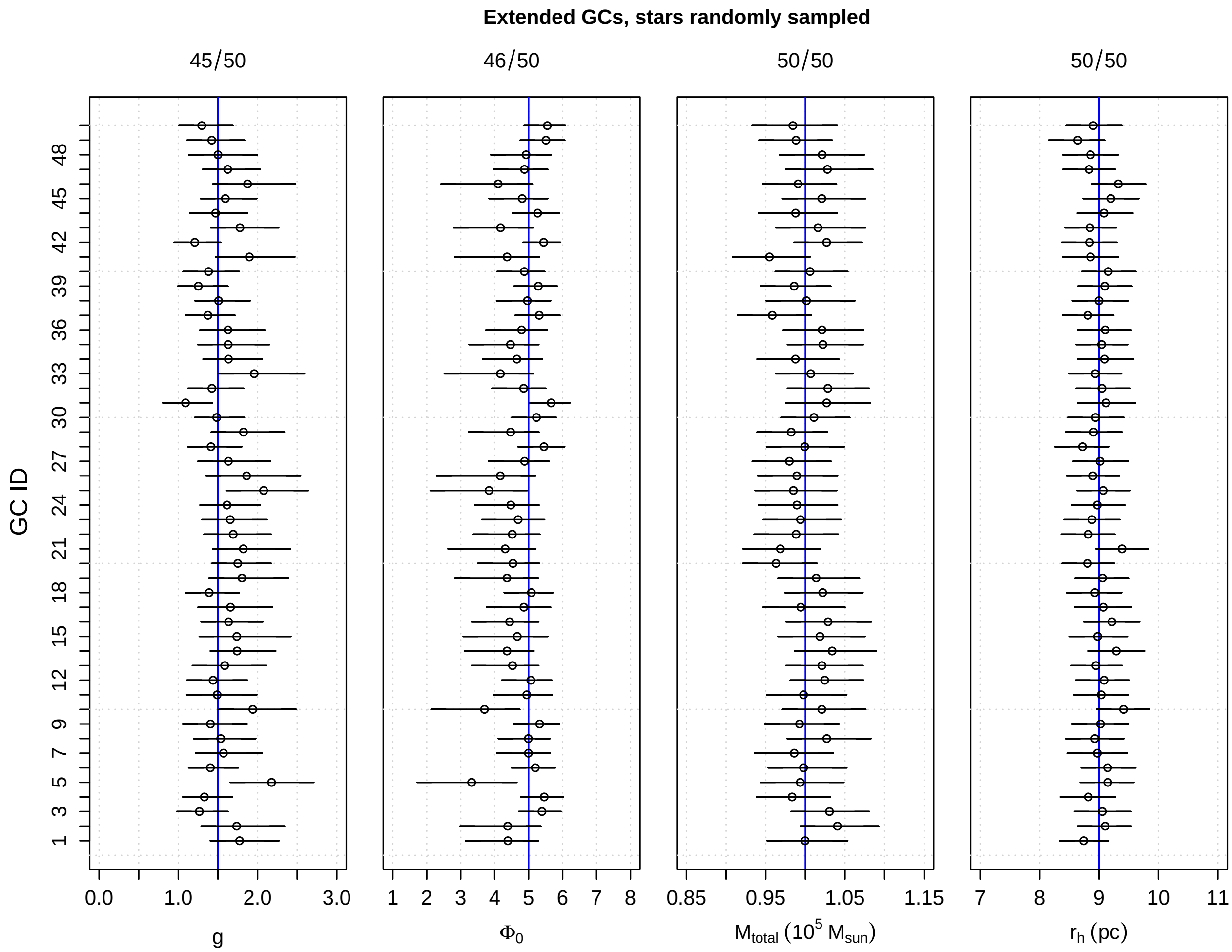

Figure 3. The parameter estimates and 95% credible intervals for fifty simulated "extended" GCs, when stars are randomly sampled at all radii. The fractions are again very large, as they should be.

the Bayesian credible interval. We can see that both the 50% and 95% credible intervals for M_{total} are reliable when the stars are randomly sampled throughout the cluster, despite cluster type.

The MCMC samples can also be used to infer the cumulative mass profile (CMP) of the cluster under the `limepy` model. Figure 5 shows the CMP inferred for one example of an average, compact, extended, low Φ_0, and high Φ_0 cluster in the random sampling case. The posterior distribution samples of g, Φ_0, M_{total} and r_h from the Markov chains are used to calculate the posterior estimate of the CMP, shown as transparent black curves. The red curve shows the true CMP given by the `limepy` model with the correct parameters.

The CMPs provide not only a visual inspection of our method, but also a quantitative one. The posterior curves for a given GC (e.g., the collection of black curves for the average GC in Figure 5) can be used to construct Bayesian credible intervals at all radii (e.g., the teal regions for the average GC in Figure 6). After constructing these credible regions for each GC, we can ask: *"how often does the true CMP lie within these credible regions, at different radii?"*.

As an example of this quantitative comparison, we use the results of all 50 realizations of average GCs to calculate the reliability of the CMP 95% credible regions. Table 3 shows how often the true $M(r < R)$ fell within the 95% credible region at 10 logarithmically-spaced distances r, for the average GCs. The results show that the credible regions are reliable, with the true $M(r < R)$ being recovered approximately 95% of the time at all radii.

In general, we find that the credible regions and CMPs are reliable for all types of GCs when the stars are sampled randomly throughout the cluster. It is reassuring that we can recover the true parameter values and the CMPs reliably from a random sample of only 500 stars.

In the case of real data, we will not know the true CMP of a GC. Thus, one might like to check whether the

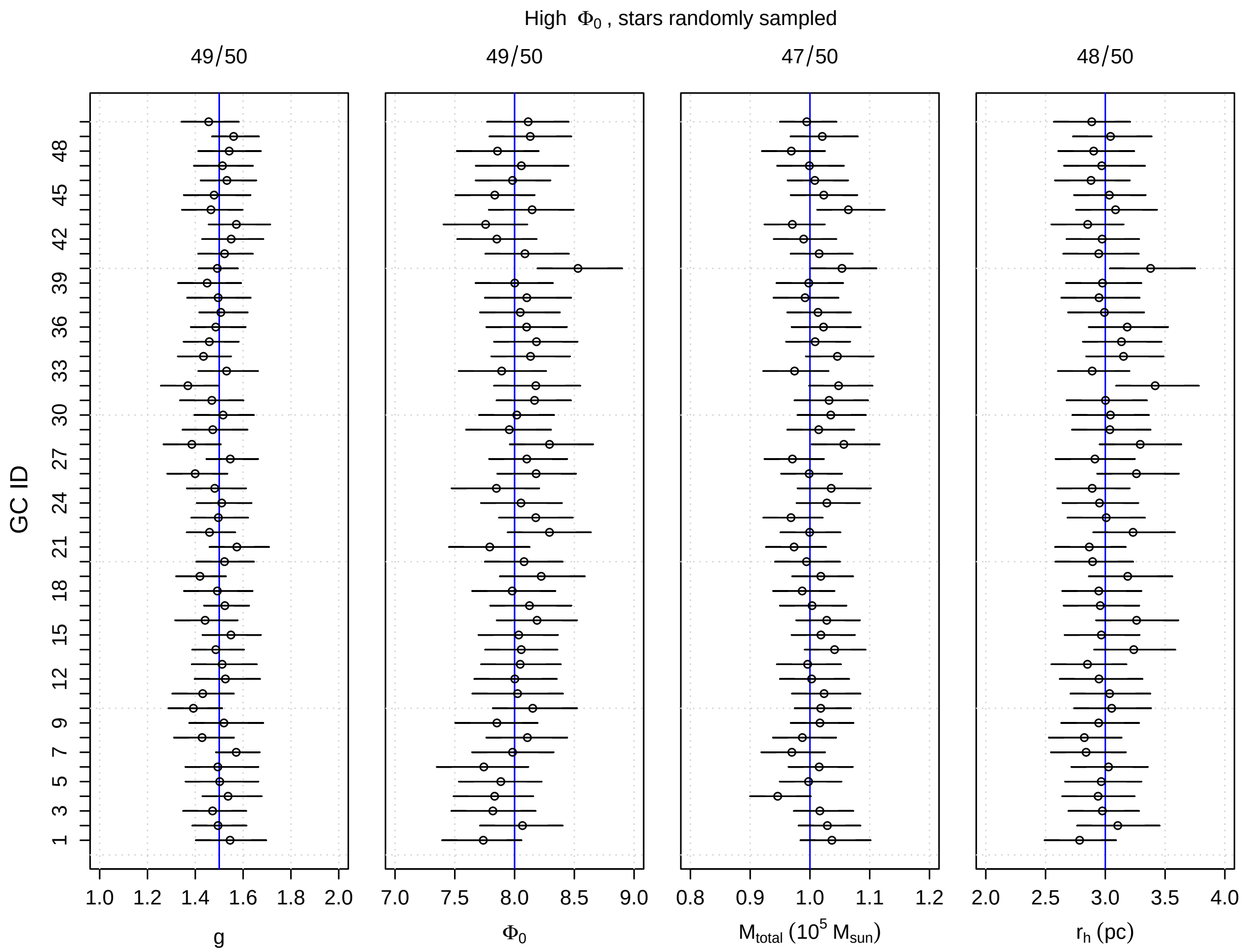

Figure 4. The parameter estimates and 95% credible intervals for fifty simulated high Φ_0 GCs, when stars are randomly sampled at all radii. The fractions are again very large, as they should be.

Average GCs, stars randomly sampled	
r (pc)	within 95% c.r.
1.00	49/50
1.39	49/50
1.95	50/50
2.71	50/50
3.79	50/50
5.28	46/50
7.37	46/50
10.28	46/50
14.34	47/50
20.00	47/50

Table 3. Reliability of CMP credible regions (c.r.) for the average GCs, under random sampling of stars.

CMP inference is reasonable given the observed data. One way to compare the Bayesian inferred CMP to the observed data is shown in Figure 6. Here, the 50, 75, and 95% credible regions are compared to the empirical cumulative distribution function of the 500 stars' positions, r. Another way to do this kind of comparison or check, which is not done here, would be to perform *posterior predictive checks* — simulate data from the posterior distribution, and compare these simulated data to the real data (e.g., see Shen et al. 2021, where Bayesian posterior predictive checks are used to check inferences about the CMP of the Milky Way).

Additionally, we can inspect other physical quantities provided by the `limepy` model fit. For example, Figure 7 shows the mean-square velocity $\overline{v^2}$ profiles as a function of radius for one GC in each of the five morphologies. Under random sampling of the stars, we observe that the true mean-square velocity profile is well-recovered by the MCMC samples. Similarly to the CMPs discussed above, Bayesian credible regions for velocity profiles could also be calculated, and posterior predictive

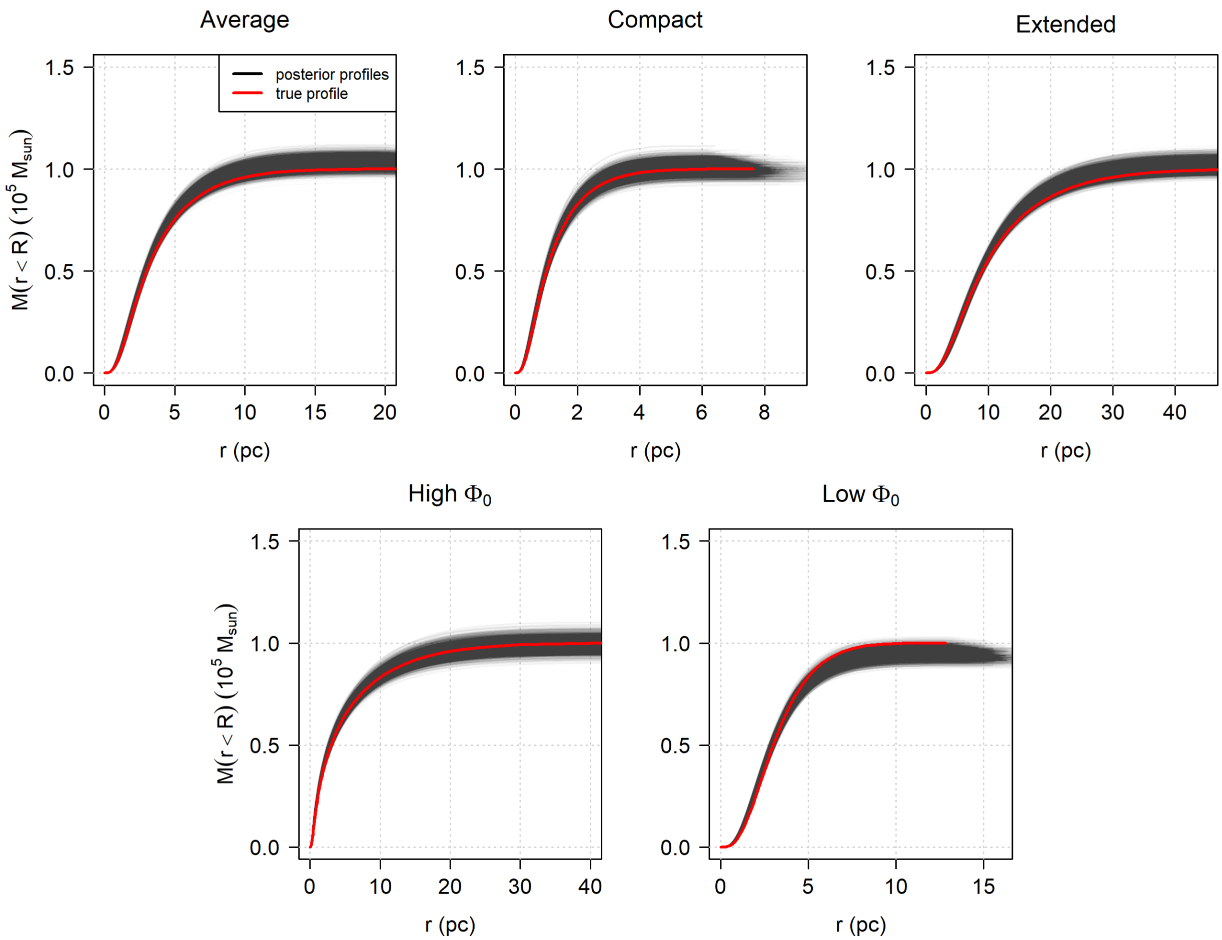

Figure 5. Example cumulative mass profiles (CMPs) calculated from the posterior samples (black curves) for the five GC types (Table 1), in the case of random sampling. Each plot shows the posterior samples for a *single* GC, with the type of GC (average, compact, extended, high Φ_0, or low Φ_0) indicated above each figure. The red solid curves show the true mass profile from the `limepy` model, showing excellent agreement.

checks could be performed to compare simulated data from the posterior to the observed data.

4.2. *Biased Sampling*

In general, we find that biased sampling of stars from only inside or outside the cluster core results in model parameter estimates that are biased and in Bayesian credible intervals that are unreliable. While obtaining biased estimates from a biased data sample is not surprising, the reality is that this type of sampling mimics the data from some telescopes. Investigating these cases can illuminate the kind of biases we should expect and possibly correct for. Indeed, through our investigations of biased sampling, we find the success of the parameter inference and CMP inference is a combination of *both* the cluster's morphology *and* the type of biased sampling.

As an example, Figure 8 shows the 95% credible intervals for an average GC when only the outer stars' data are sampled. We can see that the credible intervals are unreliable, and that parameter estimates are biased. In particular, M_{total}, g, and r_h are consistently overestimated, while Φ_0 is underestimated.

In contrast, biased sampling of outer stars of an extended cluster result in parameter estimates that are much more reliable (Figure 9). In this case, the extended GC's mass M_{total} and half-light radius r_h can actually be estimated reliably.

In Table 4, we summarize how reliably we can recover M_{total} in the biased sampling cases. Only the extended cluster with sampling in the outer regions is reliable. Also note the $*$'s in the table, which indicate when the MCMC algorithm had trouble finding a stationary distribution with good mixing, leading to biased estimates of the total mass (Table 2). In these particular cases, the behaviour of the Markov chain would be a clue to the observer that the model is having trouble describing the data.

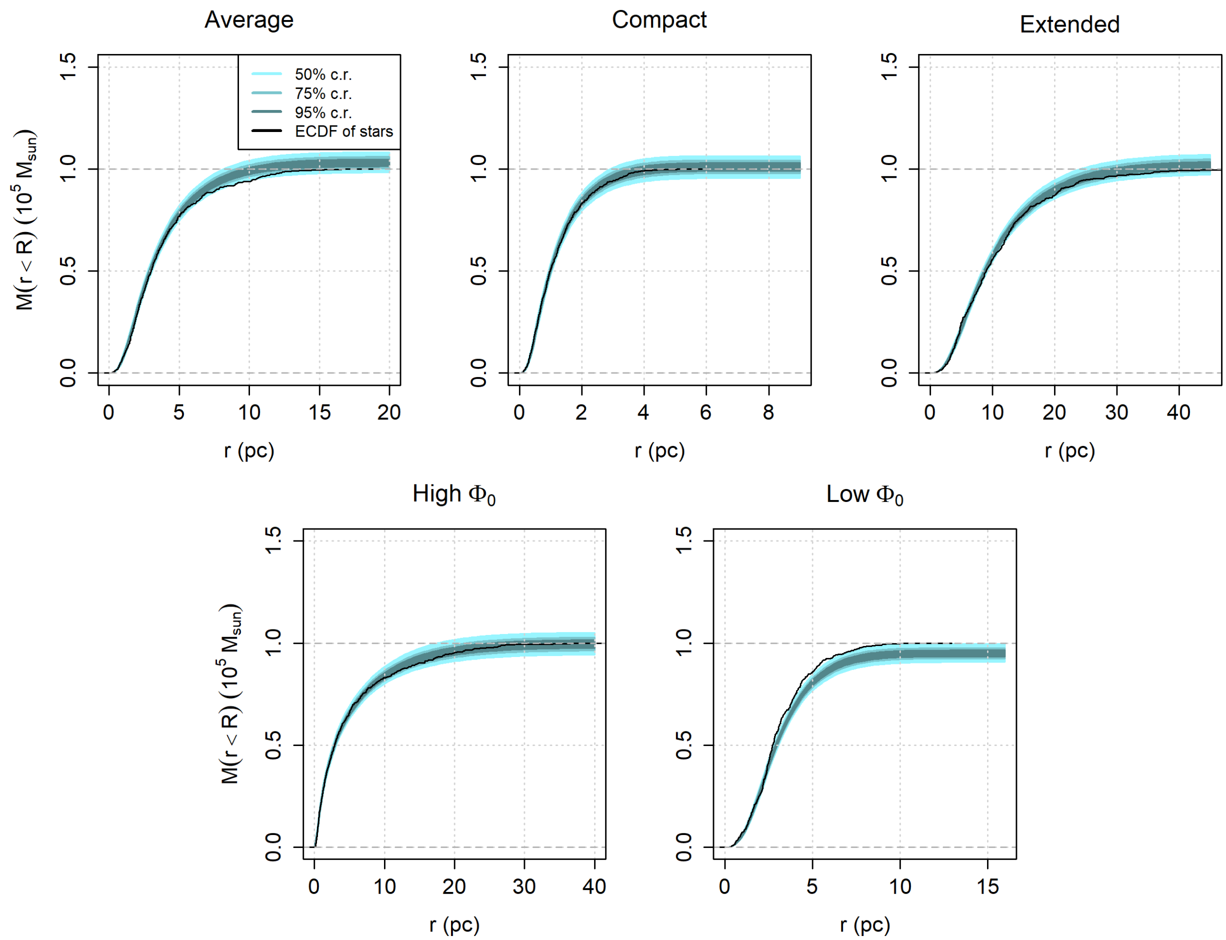

Figure 6. The cumulative mass profile 50, 75, and 95% Bayesian credible regions (dark to light teal-shaded regions respectively) for the examples shown in Figure 5. Comparing these inferred CMPs to the empirical cumulative distribution function (ECDF) of the 500 stars' distances r (black curves) could act as a check that the Bayesian inference is reasonable given the data, when the true CMP is unknown.

GC TYPE	C.I.	COVERAGE PROB. FOR M_{total}	
		outside core	*inside core*
average	50%	0.02*, +	0.00*, −
compact		0.00*, +	0.14*, −
extended		0.60, −	0.00*, −
high Φ_0		0.00, +	0.00, −
low Φ_0		0.12, +	0.00*, −
average	95%	0.08*, +	0.00*, −
compact		0.00*, +	0.62*, −
extended		0.96, −	0.00*, −
high Φ_0		0.00, +	0.00*, −
low Φ_0		0.48, +	0.00*, −

Table 4. *Estimated coverage probabilities and bias in mass estimates.* Also shown is whether the mass parameters are on average overestimated (+) or underestimated (−), or unbiased (no symbol). A $*$ indicates the chains had trouble converging and/or the estimates are at the lower or upper end of the prior distribution(s).

For GCs with high Φ_0, biased sampling of stars in the inner regions also leads to poor parameter estimates and unreliable credible regions (Figure 10). For some of the GCs in the scenario, the Markov chains become stuck in one location. The estimates of the mean from these bad chains are shown as the open circles with a small dot in the middle (i.e., the "estimates" have a variance of zero because the chains became stuck at a single place in parameter space). The exact estimated parameter values in these bad cases are rather meaningless and random. Moreover, if a scientist were to see this behaviour in a Markov chain from a real data analysis, then they would know not to trust the solution. However, in many cases of randomly generated GCs with high Φ_0 and biased sampling in the inner regions, the Markov chains do look reasonable even when their estimates are not. Thus, a scientist could mistakenly assume the convergence is giving reliable parameter estimates. We will return to this scenario shortly.

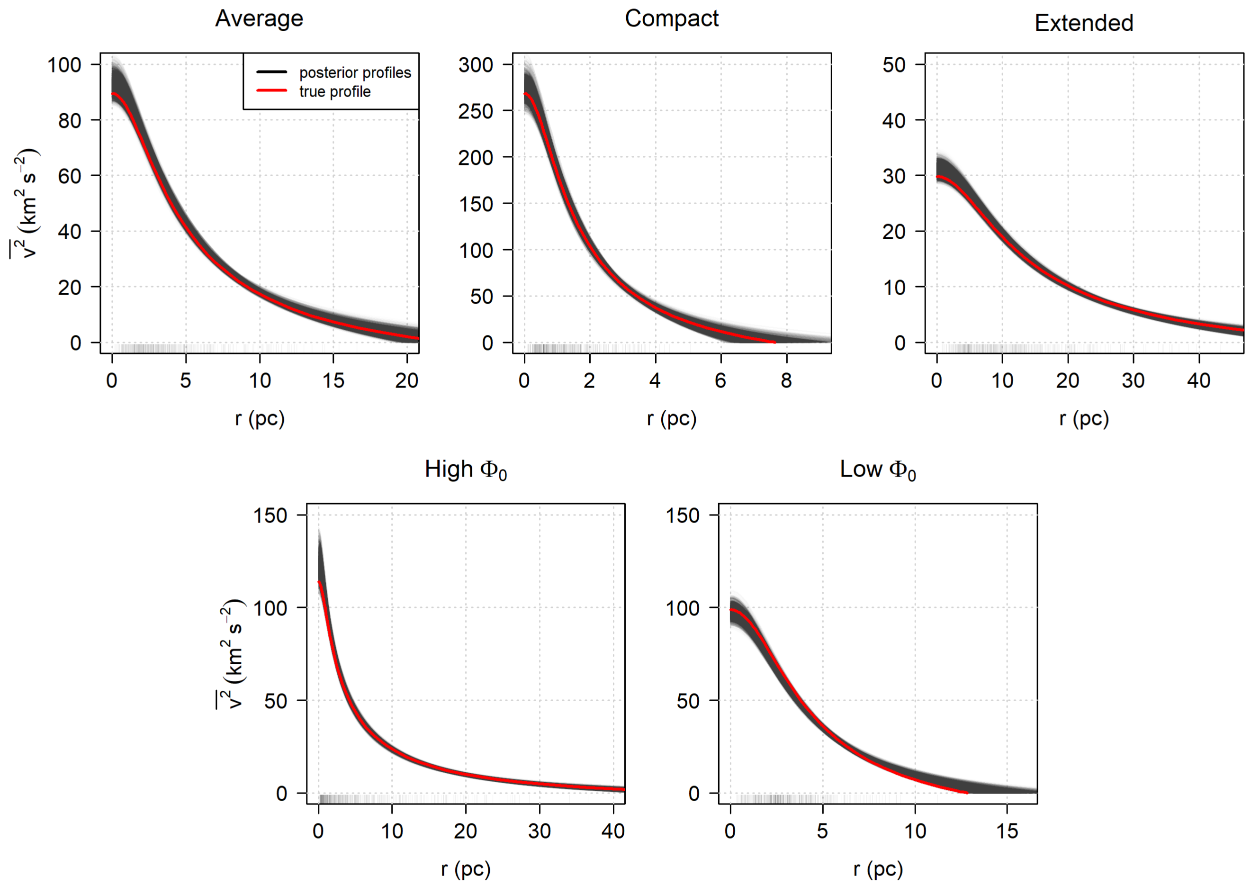

Figure 7. Example mean-square velocity profiles calculated from the posterior samples (black curves) for the case of random sampling. Each plot title indicates the type of GC (average, compact, extended, high Φ_0, or low Φ_0). The red solid curves show the true mean-square velocity profile from the `limepy` model, again showing good agreement. The semi-transparent vertical dashes along the bottom of each plot show the exact location r of the randomly sampled stars in the GC.

As mentioned in the previous section, the CMPs provide more insight than simply looking at the parameter estimates and their credible intervals, both visually and quantitatively. Figure 11 shows example CMPs for each GC morphology when the stars in these GCs are sampled only in their outer or inner regions (first and second column respectively). Looking at the first column in Figure 11, we see that when stars are sampled *outside* the core, the *inner region* of the cluster's profile tends to be underestimated — regardless of the GC morphology. The opposite is true for sampling *inside* the core (the second column). At the same time, sampling outside the core tends to lead to an overestimate of the total mass, while sampling inside the core leads to a (sometimes severe) underestimate.

There are two exceptions to the observation that biased samples lead to biased CMPs, namely (1) when extended and low Φ_0 clusters are sampled in the outer regions, and (2) when compact clusters are sampled in the inner regions. For the extended and low Φ_0 GC, our method is able to recover the true CMP reasonably well when stars outside the core are sampled, whereas this is certainly not the case when stars inside the core are sampled. For the compact GC, we see the opposite case — the CMP is reasonably-well estimated when the sample contains stars inside the core versus outside the core.

These cases where biased samples still lead to unbiased estimates are not surprising — sampling stars in the outer region of an *extended* or *less concentrated* cluster will provide a better representation of the true stellar distribution than sampling stars in its core, because these types of GCs are less dense in their inner regions (Figure 1). Likewise, sampling stars in the inner region of a *compact* cluster will be a better representation of the true stellar distribution than a sample from the outer region because compact GCs are more dense towards their centers.

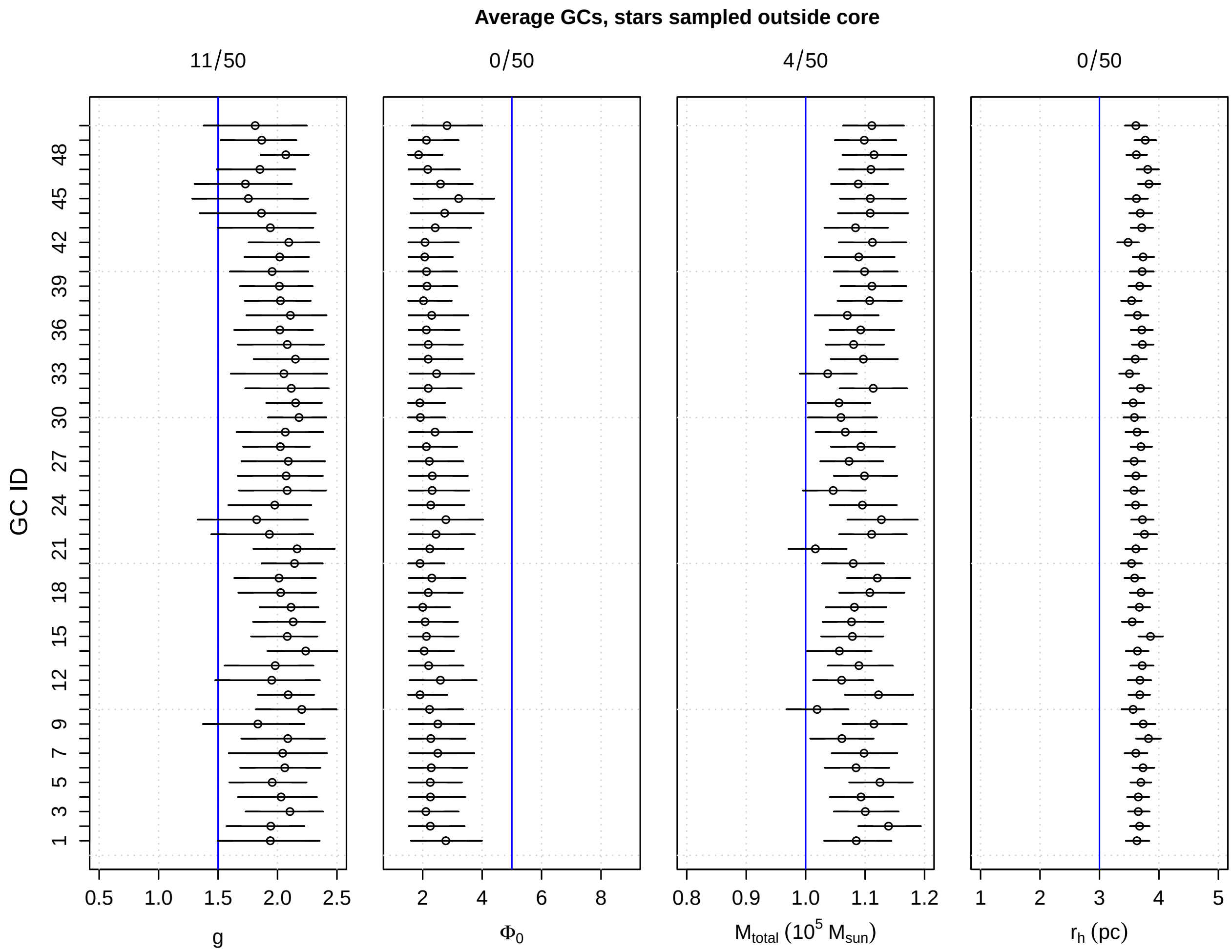

Figure 8. The mean estimate and 95% credible intervals for average GCs whose stars were sampled outside the core. Due to the biased sampling, most of the intervals miss the true value.

Next, we test the reliability of the CMP Bayesian credible regions. As an example, we show the results for low-Φ_0 GCs when stars are sampled outside the core. (Table 5). It is clear that the 95% c.r. at all radii are unreliable, with the inner regions being the most unreliable.

Next, we use the MCMC samples to estimate the mean-square velocity $\overline{v^2}$ profile as a function of radius. In Figure 12, each row corresponds to a specific GC type, and the columns indicate whether stars were sampled outside (left) or inside (right) the core of the GC. The light blue, dashed line shows the r_{cut} value, and along the bottom are semi-transparent marks showing the exact positions of the stars in the sample.

In the left-hand column of Figure 12, the estimated $\overline{v^2}$ profiles are reasonably-well matched to the true profiles for three morphologies (average, extended, and low-Φ_0 GCs). Notably, the corresponding mass profile CMPs in Figure 11 are also some of better estimates of the entire set. For the other two types of GCs, it is the inner part of the profiles that do not match; the true mean-square velocity profile (red curve) in the center of

Low Φ_0 GCs, stars sampled outside core	
r (pc)	within 95% c.r.
1.00	0/50
1.36	0/50
1.85	0/50
2.52	0/50
3.43	6/50
4.67	39/50
6.35	39/50
8.64	28/50
11.76	23/50
16.00	23/50

Table 5. Reliability of CMPs for Low Φ_0 GCs, in the case of biased sampling.

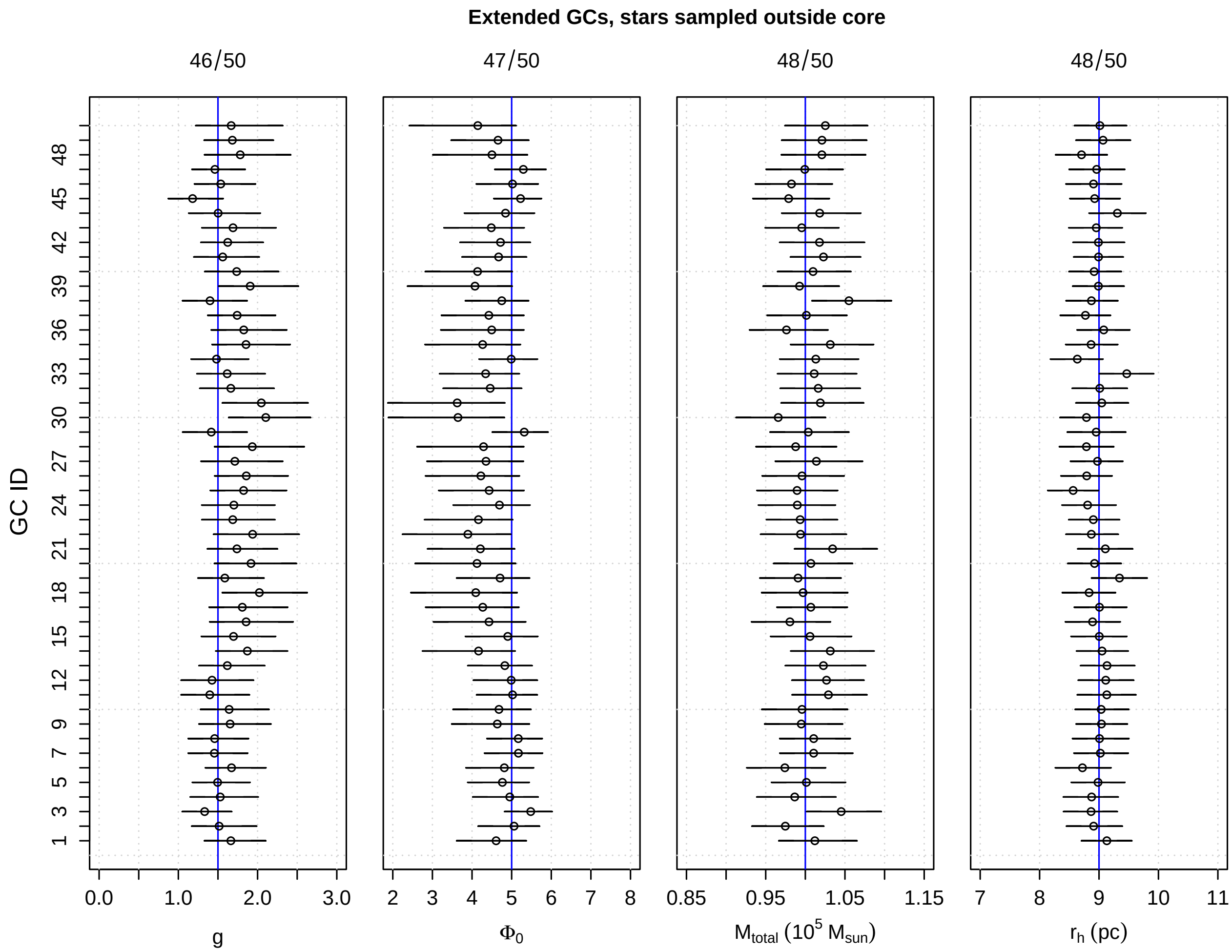

Figure 9. Same as Figure 2, but for extended GCs whose stars are sampled only from the outer regions.

the GC is much higher than the predicted profiles (black curves). Our findings suggest that a reasonable estimate of the $\overline{v^2}$ profile *might* be possible for outer regions of the GC when stars are sampled outside the core, but that it would be ill-advised to extrapolate the model fit to the inner regions when only stars outside of the core are available.

In the right-hand column of Figure 12, we see that for every type of GC the true mean-square velocity profile is poorly matched by the predictions at all radii r. Within the r_{cut} value, the true profile is generally lower than the black curves, whereas it is much higher than the black curves outside r_{cut}. Thus, the kinematic information from inner GC stars alone is not enough to constrain the model at any radii.

One aspect that we have not explored in the biased sampling cases is whether the r_{cut} value plays a significant role in determining parameter estimates — especially if that r_{cut} value was more directly linked to GC morphology. Here, we have used a fixed r_{cut} value mostly for simplicity — but in future work it would be worth exploring the impact of r_{cut} more fully. For example, the r_{cut} value for an extended or less-concentrated GC might be relatively smaller than that for the r_{cut} value for a compact or highly-concentrated GC.

It is also worth mentioning that for the fits in the right-hand column of Figure 12, the Markov chains had trouble converging and/or the estimates of the parameter were at the lower or upper end of the prior distributions (see Table 4). Both of these issues are red flags; the model has not been fit well to the data and any inference would be imprudent.

5. CONCLUSION

This paper has investigated the estimation of globular cluster properties based upon a sample of their constituent stars. We have developed a Markov chain Monte Carlo (MCMC) algorithm to compute the four parameters of a lowered isothermal model that is used to represent a GC system. Our algorithm uses a version

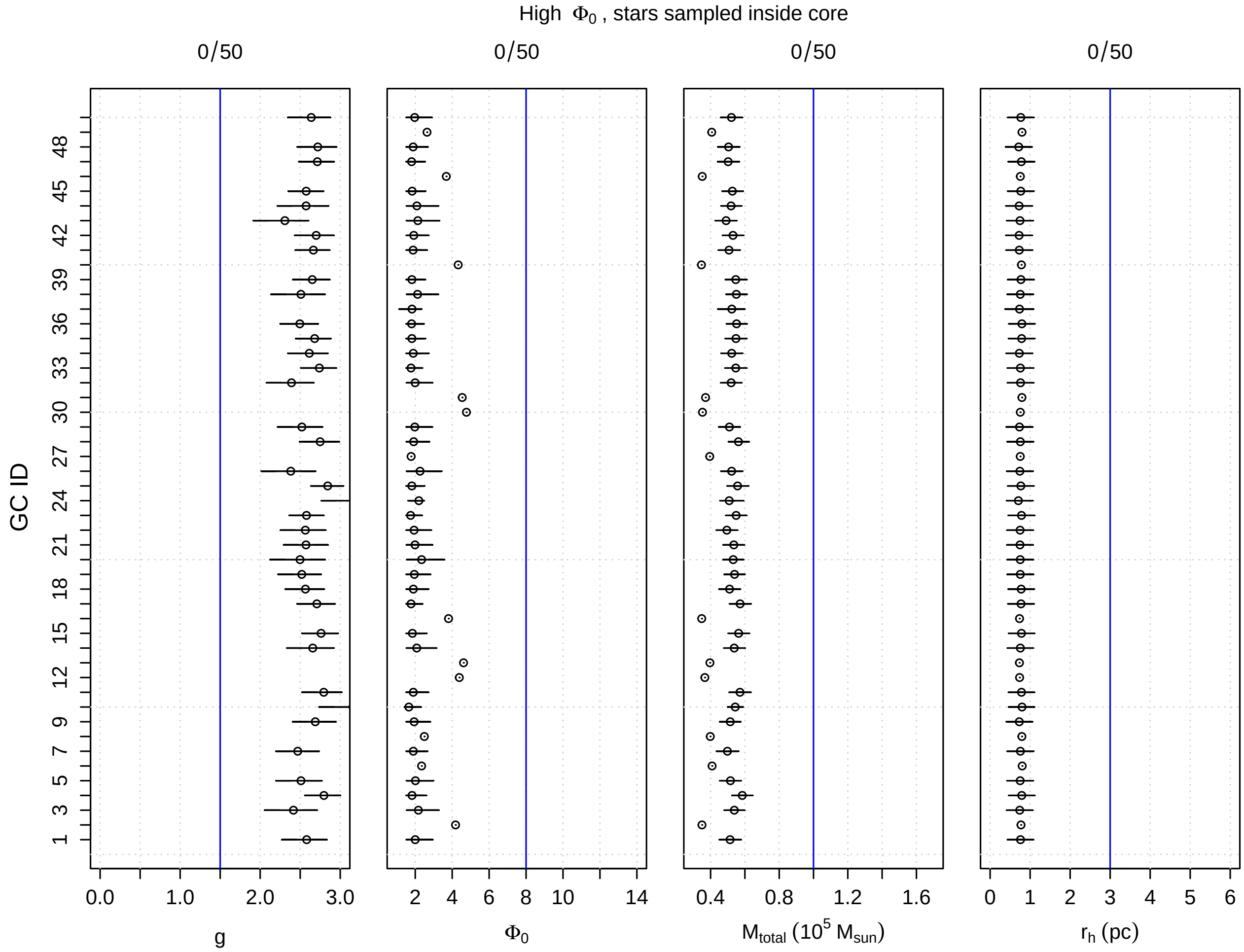

Figure 10. Same as Figure 2, but for high Φ_0 GCs sampled inside the core.

of the Metropolis algorithm, together with a numerical optimisation to find good starting values, and a finite adaptation tuning phase to find a good proposal covariance matrix. We then applied our algorithm to simulated data generated using the `limepy` package (Gieles & Zocchi 2015), and examined the extent to which the parameters, mass profile, and mean-square velocity profile of the original cluster are recovered by our algorithm.

A major goal for this study was to investigate what types of bias can occur when the GC's stars are sampled (a) randomly, (b) from the outer regions of the cluster, and (c) from the inner regions of the cluster. In summary, are findings are:

- Using all spatial and kinematic information and sampling stars randomly from throughout the cluster, our method gives reliable credible intervals for the parameter values, as well as reliable cumulative mass profiles (CMPs), and mean-squared velocity profiles.

- Using a biased sample of stars (i.e., within/outside r_h) gives unreliable credible intervals, leads to biased parameter estimates, and provides poor inference of the CMP and mean-square velocity profile.

- There are two possible exceptions where even biased samples still tend to be reliable: (1) extended and low Φ_0 clusters that are sampled in the outer regions, and (2) compact clusters that are sampled in the inner regions. In these cases, we believe the credible intervals for the parameters and CMPs are more reliable because the distribution of the sampled data is more similar the true distribution of stars in the cluster.

These results are quite promising. If the stellar data is sampled randomly in an unbiased fashion, then our algorithm's estimates are quite accurate. The mass profiles correspond closely to the theoretical curves, and the parameter estimates are close to the true parameters. We are also able to accurately estimate our error

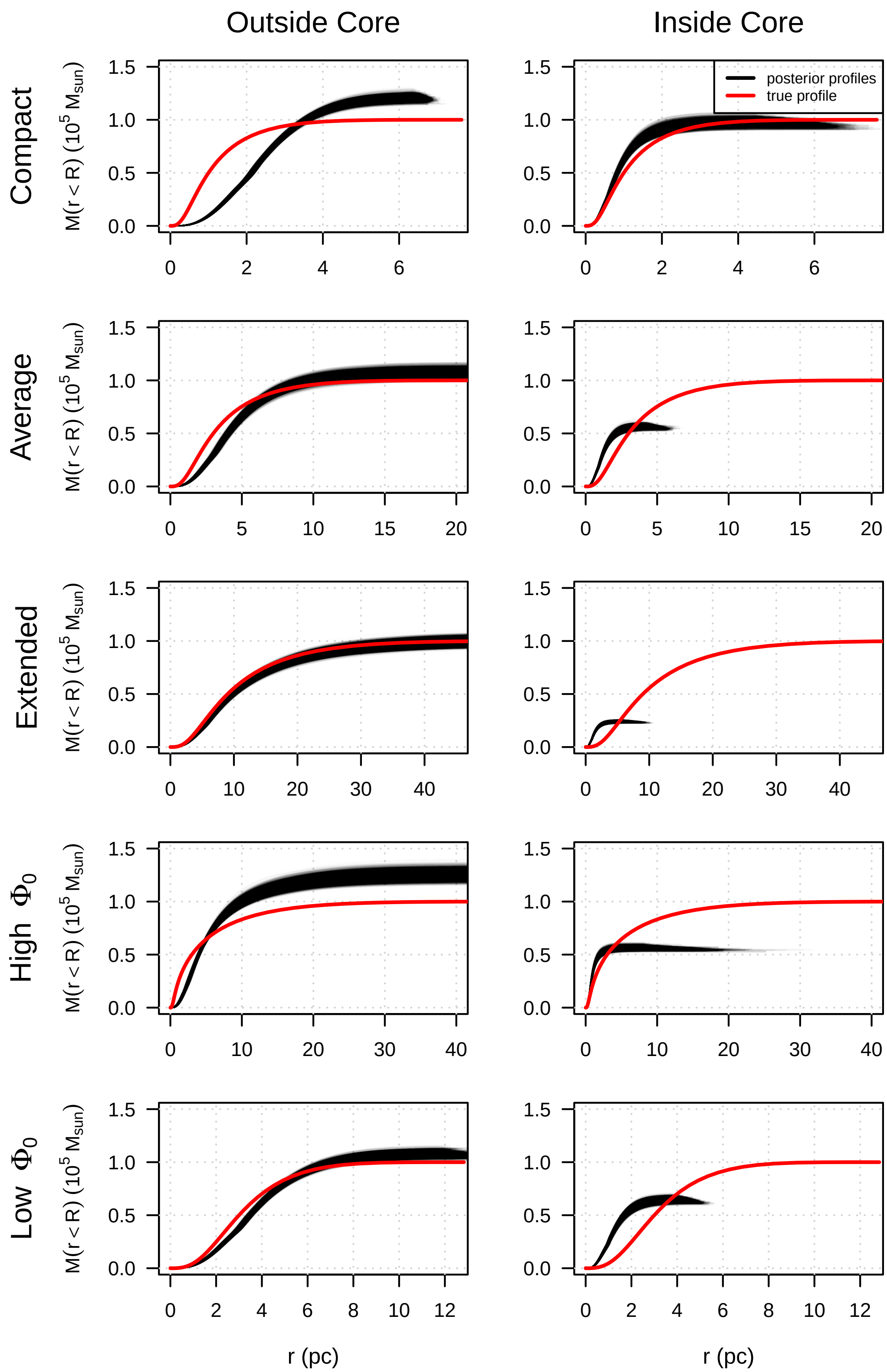

Figure 11. Example cumulative mass profile estimates (black curve) when stars are subject to selection bias either outside or inside the core of the GC. The black semi-transparent curves show the mass profiles predicted by the MCMC samples, and the solid red curves show the true mass profiles. Each row corresponds to the type of GC, and each column corresponds to the type of biased sampling — stars sampled outside or inside the core of the GC. The biased samples lead to very poor estimates in most cases, with the exception of the morphology-sampling combinations of extended cluster-outside core, the compact cluster-inside core, and low-Φ_0-inside core.

range, so that our 50% and 95% credible regions for the parameters have very close to the correct coverage probabilities.

If the stars are instead sampled in a biased fashion, then the results are more mixed. Biased sampling of outer stars only for an extended and low Φ_0 cluster still works well, since the essential information is preserved. However, in other cases, biased samples lead to biased estimates with poor coverage probabilities. This is not surprising, since our model assumes that the star sample is truly random (i.e., unbiased).

As we have seen, the bias in parameter estimates and profiles can be quite pronounced and consistent among the simulations when the data sample is biased. We could propose a "calibration" to correct for these parameter and profile biases, and such a calibration would allow us to re-scale the parameters and profiles to better match the truth. However, this calibration would only be valid for the specific analysis of full 6-D phase-space information that we have presented here. Ultimately, we plan to expand our method in future work to deal with projected position data and missing velocity components (i.e., a more realistic data scenario). At that stage, the biases in the mass and velocity profile estimates could change substantially. Thus, we leave any calibration to future work, when its application will be most useful.

There are many avenues to pursue for future work. We are currently investigating how to modify the model to give more accurate estimates in the face of biased samples, and similarly when only projected values of the star positions and velocities are known.

Both biased samples and missing data are an astronomer's reality. For example, kinematic data of stars measured by HST typically sample only a portion of the cluster, whereas the Gaia satellite mostly provides kinematic data from stars in a GC's outer regions with the inner regions being incomplete. Without accounting for a biased sample, parameter inference is less reliable.

Real kinematic data from HST and Gaia also have well-understood measurement uncertainties. We have not included measurement uncertainties in our simulation study, but a valuable next step would be to generate noisy measurements and then include a measurement model for each star that takes into account the sampling distribution of the measured kinematic components. This step could be accomplished through a hierarchical model. Additionally, one could use this framework as a way to *combine* data from different telescopes that have different measurement properties (e.g. HST and Gaia), and thus obtain a less-biased sample of the stars in the cluster. As we have shown in this work, an unbiased sample of stars is key to reliable parameter inference and recovering a good estimate of the CMP.

Ultimately, astronomers are not only interested in the intrinsic properties of GCs, but are also interested in comparison and selection of GC models. The latter will help our understanding of internal GC dynamics and the larger story of GC evolution as GCs traverse the Galactic potential. For example, the recently developed `SPES` model (Claydon et al. 2019) allows some of the stars in a GC to be "potential escapers". The existence of energetically unbound stars within clusters is, again, an astronomer's reality and could strongly affect how well a given distribution function is fit to observations. In fact, de Boer et al. (2019) found that the `SPES` models were a better representation of Galactic GCs than `limepy` models when fitting to GC density profiles. We are currently investigating some preliminary model comparison tests with simulated data from the `limepy` and `spes` models (Lou et al, in prep).

It is also important to compare the method presented here to traditional methods in the literature that use the projected distances of stars to estimate density and mass profiles, and that combine data sets from different telescopes to use stars at all radii (e.g., de Boer et al. 2019). However, at this stage of our research we have assumed an "ideal" scenario in which we have the full 6-dimensional phase-space information of stars — a comparison of our results to other methods which use only projected distances of the stars will unfairly favour our method simply because we have more positional information. In a follow-up study, we plan to improve our Bayesian approach so that it can be applied to the measurements of projected distances, and at this stage a more fair comparison of methods could be made.

The ability to attribute a given dynamical model to an observed GC is a key step towards unravelling a GC's current properties as well as its evolutionary history. Understanding the underlying distribution function of stars within a cluster allows for more complex GC features, like its dark remnant population, binary population, degree of mass segregation, and its tidal history to be more thoroughly explored. Using a model that incorporates all these components — while also improving the statistical framework to account for sampling bias in observations — will allow us to better understand the dynamical state of globular clusters. Knowing a cluster's dynamical state also places constraints on the cluster's properties at birth and how it has evolved over time. Hence, being able to fit a dynamical model to an observed GC strengthens the cluster's utility as a tool to study the Universe around it.

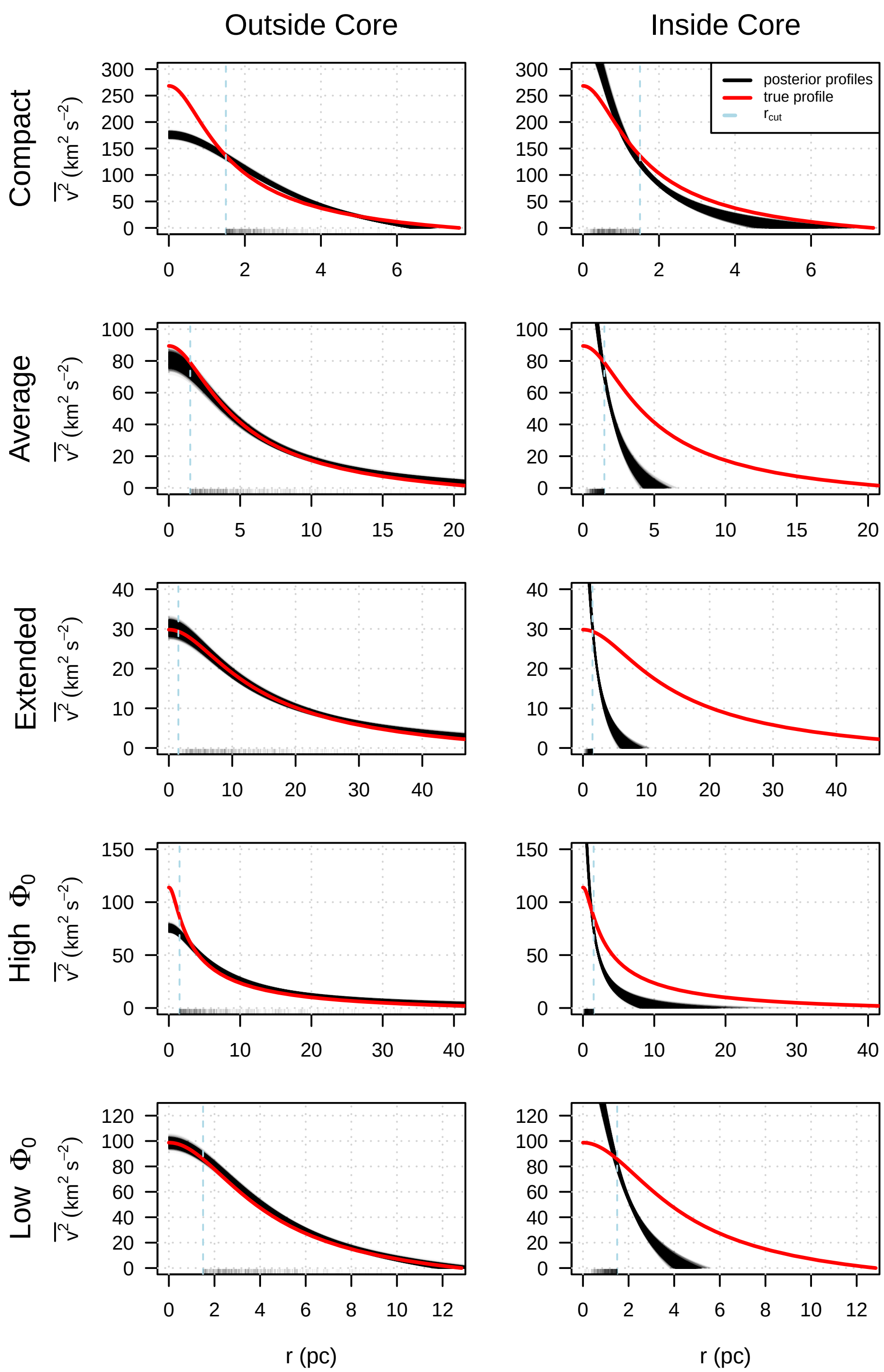

Figure 12. Example mean-square velocity profile estimates (black curves) from the MCMC samples when stars are subject to selection bias. The solid red curves show the true $\overline{v^2}$ profiles. Each row corresponds to the type of GC, and each column corresponds to the type of biased sampling — stars sampled outside or inside the core of the GC. The vertical, light-blue dashed line indicates the $r_{cut} = 1.5$pc and the semi-transparent vertical dashes along the bottom of each plot shows the individual positions of each star in the (biased) sample. The biased samples lead to very poor estimates in most cases, with the exception of the morphology-sampling combinations of average-outside core, extended cluster-outside core, and low-Φ_0-outside core.

ACKNOWLEDGEMENTS

GME acknowledges the support of a Discovery Grant from the Natural Sciences and Engineering Research Council of Canada (NSERC, RGPIN-2020-04554), and a Connaught New Reseacher grant from the University of Toronto. JSR was supported by NSERC grant RGPIN-2019-04142. JW would like to thank Mark Gieles for helpful discussions regarding the `limepy` software package. GME would like to thank Joshua Speagle for helpful discussions regarding the differential optimization algorithm. The authors would also like to thank the referee for their very helpful report that helped improve this paper.

Software: The code for this research can be found at https://github.com/gweneadie/GCs. Our code makes use of the following software and software packages: `astropy` (Astropy Collaboration et al. 2013), `Cairo` (Urbanek & Horner 2020), `coda` (Plummer et al. 2006), `dplyr` (Wickham et al. 2020), `limepy` (Gieles & Zocchi 2015), `MASS` (Venables & Ripley 2002), `NMOF` (Schumann 2011–2021; Gilli et al. 2019), **R** (R Core Team 2019), `reticulate` (Ushey et al. 2020), `tibble` (Müller & Wickham 2020), and `tidyverse` (Wickham et al. 2019).

REFERENCES

Astropy Collaboration, Robitaille, T. P., Tollerud, E. J., et al. 2013, A&A, 558, A33, doi: 10.1051/0004-6361/201322068

Baumgardt, H., & Hilker, M. 2018, MNRAS, 478, 1520, doi: 10.1093/mnras/sty1057

Bayes, T. 1763, Philosophical transactions of the Royal Society of London, 370

Bellini, A., Anderson, J., van der Marel, R. P., et al. 2014, ApJ, 797, 115, doi: 10.1088/0004-637X/797/2/115

Bertin, G., & Varri, A. L. 2008, ApJ, 689, 1005, doi: 10.1086/592684

Brodie, J. P., & Strader, J. 2006, ARA&A, 44, 193, doi: 10.1146/annurev.astro.44.051905.092441

Cappellari, M. 2008, MNRAS, 390, 71, doi: 10.1111/j.1365-2966.2008.13754.x

Claydon, I., Gieles, M., Varri, A. L., Heggie, D. C., & Zocchi, A. 2019, MNRAS, 487, 147, doi: 10.1093/mnras/stz1109

Da Costa, G. S., & Freeman, K. C. 1976, ApJ, 206, 128, doi: 10.1086/154363

de Boer, T. J. L., Gieles, M., Balbinot, E., et al. 2019, MNRAS, 485, 4906, doi: 10.1093/mnras/stz651

Eadie, G., & Jurić, M. 2019, The Astrophysical Journal, 875, 159

Gaia Collaboration, Prusti, T., de Bruijne, J. H. J., et al. 2016, A&A, 595, A1, doi: 10.1051/0004-6361/201629272

Gaia Collaboration, Brown, A. G. A., Vallenari, A., et al. 2018, A&A, 616, A1, doi: 10.1051/0004-6361/201833051

Gieles, M., & Zocchi, A. 2015, MNRAS, 454, 576, doi: 10.1093/mnras/stv1848

Gilli, M., Maringer, D., & Schumann, E. 2019, Numerical Methods and Optimization in Finance, 2nd edn. (Waltham, MA, USA: Elsevier/Academic Press). http://www.enricoschumann.net/NMOF/

Gunn, J. E., & Griffin, R. F. 1979, AJ, 84, 752, doi: 10.1086/112477

Haario, H., Saksman, E., Tamminen, J., et al. 2001, Bernoulli, 7, 223

Harris, W. E. 2010, arXiv e-prints, arXiv:1012.3224. https://arxiv.org/abs/1012.3224

Heggie, D., & Hut, P. 2003, The Gravitational Million-Body Problem: A Multidisciplinary Approach to Star Cluster Dynamics (Cambridge University Press)

Heggie, D. C., & Giersz, M. 2014, MNRAS, 439, 2459, doi: 10.1093/mnras/stu102

Hénault-Brunet, V., Gieles, M., Sollima, A., et al. 2019, MNRAS, 483, 1400, doi: 10.1093/mnras/sty3187

Hénon, M. 1961, Annales d'Astrophysique, 24, 369

King, I. R. 1966, AJ, 71, 64, doi: 10.1086/109857

Lynden-Bell, D., & Wood, R. 1968, MNRAS, 138, 495, doi: 10.1093/mnras/138.4.495

McLaughlin, D. E., & van der Marel, R. P. 2005, ApJS, 161, 304, doi: 10.1086/497429

Meylan, G., & Heggie, D. C. 1997, A&A Rv, 8, 1, doi: 10.1007/s001590050008

Michie, R. W. 1963, MNRAS, 125, 127, doi: 10.1093/mnras/125.2.127

Miocchi, P., Lanzoni, B., Ferraro, F. R., et al. 2013, ApJ, 774, 151, doi: 10.1088/0004-637X/774/2/151

Müller, K., & Wickham, H. 2020, tibble: Simple Data Frames. https://CRAN.R-project.org/package=tibble

Plummer, M., Best, N., Cowles, K., & Vines, K. 2006, R News, 6, 7. https://journal.r-project.org/archive/

Puzia, T. H., Paolillo, M., Goudfrooij, P., et al. 2014, ApJ, 786, 78, doi: 10.1088/0004-637X/786/2/78

R Core Team. 2019, R: A Language and Environment for Statistical Computing, R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/

Roberts, G. O., Gelman, A., & Gilks, W. R. 1997, The annals of applied probability, 7, 110

Roberts, G. O., & Rosenthal, J. S. 1997, Electronic Communications in Probability, 2, 13

—. 2001, Statistical science, 16, 351

—. 2007, Journal of applied probability, 44, 458

—. 2009, Journal of Computational and Graphical Statistics, 18, 349, doi: 10.1198/jcgs.2009.06134

Schumann, E. 2011–2021, Numerical Methods and Optimization in Finance (NMOF) Manual. Package version 2.4-1). http://enricoschumann.net/NMOF/

Shanahan, R. L., & Gieles, M. 2015, MNRAS, 448, L94, doi: 10.1093/mnrasl/slu205

Shen, J., Eadie, G. M., Murray, N., et al. 2021, arXiv e-prints, arXiv:2111.09327. https://arxiv.org/abs/2111.09327

Sollima, A., Baumgardt, H., Zocchi, A., et al. 2015, MNRAS, 451, 2185, doi: 10.1093/mnras/stv1079

Spitzer, L. 1987, Dynamical evolution of globular clusters (Princeton University Press)

Storn, R., & Price, K. 1997, Journal of global optimization, 11, 341

Urbanek, S., & Horner, J. 2020, Cairo: R Graphics Device using Cairo Graphics Library for Creating High-Quality Bitmap (PNG, JPEG, TIFF), Vector (PDF, SVG, PostScript) and Display (X11 and Win32) Output. https://CRAN.R-project.org/package=Cairo

Usher, C., Forbes, D. A., Spitler, L. R., et al. 2013, MNRAS, 436, 1172, doi: 10.1093/mnras/stt1637

Ushey, K., Allaire, J., & Tang, Y. 2020, reticulate: Interface to 'Python'. https://CRAN.R-project.org/package=reticulate

Varri, A. L., & Bertin, G. 2012, A&A, 540, A94, doi: 10.1051/0004-6361/201118300

Venables, W. N., & Ripley, B. D. 2002, Modern Applied Statistics with S, 4th edn. (New York: Springer). http://www.stats.ox.ac.uk/pub/MASS4/

Watkins, L. L., van de Ven, G., den Brok, M., & van den Bosch, R. C. E. 2013, MNRAS, 436, 2598, doi: 10.1093/mnras/stt1756

Webb, J. J., Sills, A., & Harris, W. E. 2013, ApJ, 779, 94, doi: 10.1088/0004-637X/779/2/94

Wickham, H., François, R., Henry, L., & Müller, K. 2020, dplyr: A Grammar of Data Manipulation. https://CRAN.R-project.org/package=dplyr

Wickham, H., Averick, M., Bryan, J., et al. 2019, Journal of Open Source Software, 4, 1686, doi: 10.21105/joss.01686

Wilson, C. P. 1975, AJ, 80, 175, doi: 10.1086/111729

Woodley, K. A., & Gómez, M. 2010, PASA, 27, 379, doi: 10.1071/AS09059

Woolley, R. V. D. R. 1954, MNRAS, 114, 191, doi: 10.1093/mnras/114.2.191

Zocchi, A., Gieles, M., & Hénault-Brunet, V. 2017, MNRAS, 468, 4429, doi: 10.1093/mnras/stx316

Zocchi, A., Gieles, M., Hénault-Brunet, V., & Varri, A. L. 2016, MNRAS, 462, 696, doi: 10.1093/mnras/stw1104