LJMU Research Online

Everall, A, Wyn Evans, N, Belokurov, V, Boubert, D and Grand, RJJ

 The Photo-Astrometric vertical tracer density of the Milky Way-I. the method

http://researchonline.ljmu.ac.uk/id/eprint/19280/

Article

For more information please contact researchonline@ljmu.ac.uk

http://researchonline.ljmu.ac.uk/

The Photo-Astrometric vertical tracer density of the Milky Way – I. The method

Andrew Everall [ID],[1]★ N. Wyn Evans,[1] Vasily Belokurov [ID],[1] Douglas Boubert [ID][2,3]
and Robert J. J. Grand [ID][4,5,6]

[1]*Institute of Astronomy, University of Cambridge, Madingley Road, Cambridge CB3 0HA, UK*
[2]*Magdalen College, University of Oxford, High Street, Oxford OX1 4AU, UK*
[3]*Rudolf Peierls Centre for Theoretical Physics, Clarendon Laboratory, Parks Road, Oxford OX1 3PU, UK*
[4]*Max Planck Institute for Astrophysics, Karl-Schwarzschild-Str. 1, Postfach 1317, D-85741 Garching, Germany*
[5]*Instituto de Astrofísica de Canarias, Calle Vía Láctea s/n, E-38205 La Laguna, Tenerife, Spain*
[6]*Departamento de Astrofísica, Universidad de La Laguna, Av. del Astrofísico Francisco Sánchez s/n, E-38206 La Laguna, Tenerife, Spain*

ABSTRACT
We introduce a method to infer the vertical distribution of stars in the Milky Way using a Poisson likelihood function, with a view to applying our method to the *Gaia* catalogue. We show how to account for the sample selection function and for parallax measurement uncertainties. Our method is validated against a simulated sample drawn from a model with two exponential discs and a power-law halo profile. A mock *Gaia* sample is generated using the *Gaia* astrometry selection function, whilst realistic parallax uncertainties are drawn from the *Gaia* Astrometric Spread Function. The model is fit to the mock in order to rediscover the input parameters used to generate the sample. We recover posterior distributions that accurately fit the input parameters within statistical uncertainties, demonstrating the efficacy of our method. Using the GUMS synthetic Milky Way catalogue, we find that our halo parameter fits can be heavily biased by our overly simplistic model; however, the fits to the thin and thick discs are not significantly impacted. We apply this method to *Gaia* Early Data Release 3 in a companion paper where we also quantify the systematic uncertainties introduced by oversimplifications in our model.

Key words: methods: data analysis – methods: statistical – stars: statistics – Galaxy: kinematics and dynamics – Galaxy: stellar content.

1 INTRODUCTION

Models for the distribution of stars in the Milky Way are key to stellar physics, Galactic archaeology (study of the formation history of the Galaxy) and understanding observations of external galaxies.

A core scientific aim of the *Gaia* mission is to map the 3D distribution of sources throughout the Milky Way (Perryman et al. 2001). To achieve this, *Gaia* has measured parallaxes for 1467 744 818 sources (Gaia Collaboration et al. 2016, 2021a) providing geometric distance estimates with no assumptions about source intrinsic brightness (Lindegren et al. 2021). However, we cannot straightforwardly use these billions of distances to construct a map of stars throughout the Galaxy for two key reasons.

Until recently, the completeness limits of the *Gaia* catalogues were largely unknown. The observation strategy of the mission results in a completeness that varies significantly across the sky on sub-degree scales. Traditional methods of evaluating selection functions rely on the existence of a more complete source catalogue against which the sample can be compared; however, there is no such catalogue to compare against, due to the incredible depth and resolution of *Gaia* across the entire sky. Without a selection function, it is impossible to

generate an unbiased map of the Milky Way using the full power of the *Gaia* data. We refer the interested reader to Rix et al. (2021) for a detailed discussion on evaluating and using selection functions.

Furthermore, parallax-based distances are statistically awkward to work with. Much of our statistical methodology is constructed around the assumption of Gaussian measurement uncertainties, motivated by the central limit theorem. Parallax uncertainties are Gaussian distributed, which means that distances are reciprocal Gaussian distributed. This is a highly asymmetric distribution that, under an improper uniform prior, cannot be normalized. As such, the distribution does not have a finite mean. Detailed discussions on how to use *Gaia* parallaxes for distance inference on individual stars are given in Bailer-Jones (2015) and Luri et al. (2018).

In spite of these hurdles, the structure of the Milky Way has been studied in detail by many authors. A work-around to the challenges of parallax uncertainties is to focus on particular stellar populations for which the intrinsic brightness can be modelled. The distance can then be inferred from the measured apparent brightness. In some cases, simple stellar colour–absolute magnitude relations are used for either a large population of sources across the CMD (e.g. Bilir et al. 2006; Dobbie & Warren 2020) or a small subset (e.g. horizontal branch stars; Fukushima et al. 2019). This approach has been taken further by using full stellar evolution models to infer intrinsic source brightness (de Jong et al. 2010). Period–luminosity relations for certain variable

★ E-mail: aeverall2@gmail.com

sources are also incredibly valuable distance indicators. Ak et al. (2008) used cataclysmic variables to estimate the vertical profile of the Milky Way disc, whilst Mateu & Vivas (2018) used RR Lyrae to determine the structure of the old thick disc and radial profile of the halo.

Some of these approaches apply uncertain colour–magnitude relations to large populations across the CMD, leaving the results susceptible to systematic biases. Other approaches use more carefully chosen sub-samples of specific stellar types such that only a small fraction of the data are used.

In this paper, we develop a method to overcome the challenges of directly using *Gaia* parallaxes and applying the selection functions for the *Gaia* source catalogue and astrometry subset from Everall & Boubert (2021). We demonstrate its feasibility on a *Gaia*-like mock sample with a known ground-truth. We limit the scope of this work to a high latitude region of the sky for statistical and computational reasons, and due to the challenge of dust extinction that we do not attempt to solve here.

This paper is arranged as follows. In Section 2, we introduce the likelihood optimization method used for this work, followed by a full description of the model in Section 3. The *Gaia*-like mock sample is explained in Section 4 and we demonstrate the application of the method in Section 5. The method is tested on a more realistic mock catalogue in Section 6. In a companion paper (Everall et al. 2021a, henceforth Paper II), we apply this method to high latitude regions of the *Gaia* EDR3 catalogue to estimate the vertical stellar profile of the Milky Way at the Solar radius and quantify the systematic uncertainties introduced by the simplifications and assumptions used in our model.

2 METHOD

The probability of drawing a population of objects $\{x_i\}$ from a density profile $\lambda(x)$ is given by the Poisson likelihood function (for which derivations are given in Lombardi, Lada & Alves 2013; Everall & Das 2020),

$$\log \mathcal{L} = \sum_{i=1}^{N} \log (\lambda(x_i)) - \int dx \lambda(x). \tag{1}$$

The observed population of objects is drawn from the true underlying distribution of sources multiplied by a selection function that gives the probability of a source being included in the survey. Therefore, we can substitute $\lambda(x) = f(x, \psi)\mathcal{S}(x)$, where $\mathcal{S}$ is the selection function and f is the true underlying source density with model parameters ψ,

$$\log \mathcal{L} = \sum_{i=1}^{N} \log (f(x_i, \psi)\mathcal{S}(x_i)) - \int dx f(x, \psi)\mathcal{S}(x). \tag{2}$$

The aim of density estimation is to fit the parameters of the true underlying distribution, ψ. Since the selection function is independent of the model parameters, it can be dropped out of the first term in the likelihood function,

$$\log \mathcal{L} \sim \sum_{i=1}^{N} \log (f(x_i, \psi)) - \int dx f(x, \psi)\mathcal{S}(x). \tag{3}$$

The source properties, x, need to be chosen according to the dependencies of the model and selection function.

Mateu & Vivas (2018) use this method on a sample of RR Lyrae to constrain the structure of the thick disc and halo considering only spatial dimensions, while Bovy et al. (2012b) apply a more complex model to a population of G-dwarfs to fit the Milky Way disc

using measured apparent magnitude, colour, and metallicity. The aim of this work is to model the purely spatial distribution of sources, however the selection function, which will be introduced in more detail in Section 3.3, is a function of position on the sky and apparent magnitude. Therefore, we must also consider the intrinsic brightness of a source, so our source properties are $x = (l, b, s, M_G)$.

An additional complexity we introduce beyond previous works is accounting for parallax measurement uncertainties, which is vital when working with *Gaia* astrometry. Suppose instead that x are the *measured* source properties and $f(x, \psi)$ is the expected distribution of measured source properties given the model. Source measurements are drawn from an uncertainty distribution, $P(x \mid x_T)$, where x_T are the underlying *true* source properties. The measured model (f) is given by a convolution between the true underlying model (f_T) and the measurement error distribution,

$$f(x, \psi) = \int dx_T P(x \mid x_T) f_T(x_T, \psi). \tag{4}$$

Substituting this into the likelihood, we obtain

$$\log \mathcal{L} \sim \sum_{i=1}^{N} \log \left(\int dx_T P(x_i \mid x_T) f_T(x_T, \psi) \right)$$
$$- \int dx_T f_T(x_T, \psi) \int dx P(x \mid x_T) \mathcal{S}(x), \tag{5}$$

where we have reversed the order of integration in the second term and brought f_T outside the integral over measured parameters.

Our measured source properties are $x = (l, b, G, \varpi)$, or Galactic longitude and latitude, apparent magnitude, and parallax. In this work, we consider parallax error as the only significant measurement uncertainty. Positional uncertainties in (l, b) are extremely small and we will test the impact of neglecting error in G in Paper II. Therefore, the error term becomes

$$P(x \mid x_T) = \delta(l - l_T)\,\delta(b - b_T)\,\delta\,(G - G_T(s, M_G))\,P(\varpi \mid s). \tag{6}$$

We integrate over all delta functions in the first term of the likelihood function ,

$$\int dx_T P(x_i \mid x_T) f_T(x_T, \psi) = \int ds\, P(\varpi_i \mid s)\, f_T(l_i, b_i, G_i, s, \psi). \tag{7}$$

The selection function is a function of l, b, and G only; there is no dependence on measured parallax (Boubert & Everall 2020; Everall & Boubert 2021). This makes it easy to integrate over

$$\int dx\, P(x \mid x_T) \mathcal{S}(l, b, G) = \mathcal{S}\,(l_T, b_T, G_T(s, M_G)). \tag{8}$$

Finally, we can substitute this into the likelihood function,

$$\log \mathcal{L} \sim \sum_{i=1}^{N} \log \left(\int ds\, P(\varpi_i \mid s)\, f_T(l_i, b_i, G_i, s, \psi) \right)$$
$$- \int dx_T f_T(l_T, b_T, M_G, s, \psi)\, \mathcal{S}\,(l_T, b_T, G_T(s, M_G)). \tag{9}$$

This is the likelihood function that we use to fit the model parameters, ψ, to the observed data. For the remainder of this paper, we will drop the subscript T with f always referring to the underlying source distribution.

2.1 Parallax error integration

The biggest numerical challenge for our method is the parallax error convolution. We need to integrate over parallax for every source at

every proposed set of model parameters. In this section, we will use slightly different notation, where $\varpi = 1/s$ is the true parallax distance that we are marginalizing over, and ϖ_i is the measured parallax for source i. The integral we need to evaluate is

$$\int_0^\infty ds \, P(\varpi_i \,|\, s) \, f(l_i, b_i, G_i, s, \boldsymbol{\psi})$$
$$= \int_0^\infty d\varpi \, \varpi^{-2} \, \mathcal{N}(\varpi; \varpi_i, \sigma_{\varpi,i}) \, f(l_i, b_i, G_i, s, \boldsymbol{\psi})$$
$$\equiv \int_0^\infty d\varpi \, I(\varpi), \tag{10}$$

where $\mathcal{N}$ is a normal (Gaussian) distribution with standard deviation $\sigma_{\varpi, i}$, which is the parallax error of source i. In Section 3, we will introduce the absolute magnitude model that is broken into sections with an upper absolute magnitude limit (minimum brightness) for the model. We can then write the integral as a sum of definite integrals ,

$$\int_0^\infty d\varpi \, I(\varpi) = \sum_j \int_{\varpi_j}^{\varpi_{j+1}} d\varpi \, I(\varpi), \tag{11}$$

where

$$\varpi_j = 10^{(M_j + 10 - G_i)/5} \tag{12}$$

and M_j are the magnitude boundaries of the sections. For an unconstrained lower absolute magnitude limit, $\varpi_0 = 0$.

We numerically evaluate the integral of each section using the following five-step recipe:

(i) Transform into logit-parallax space using the substitution

$$x' = \log\left(\frac{\varpi - \varpi_j}{\varpi_{j+1} - \varpi}\right). \tag{13}$$

This gives

$$\int_{\varpi_j}^{\varpi_{j+1}} d\varpi \, I(\varpi) = \int_{-\infty}^{\infty} dx' \frac{I}{J}, \tag{14}$$

where the Jacobian is

$$J = \left|\frac{\partial x'}{\partial \varpi}\right| = \frac{\varpi_{j+1} - \varpi_j}{(\varpi - \varpi_j)(\varpi_{j+1} - \varpi)}. \tag{15}$$

(ii) Find the peak of the logit-transformed integrand by solving

$$\frac{\partial}{\partial x'}\left(\frac{I}{J}\right) = 0 \tag{16}$$

using the bisection algorithm with respect to ϖ initializing at the integration boundaries, ϖ_j, ϖ_{j+1}. Transform the parallax of the peak into logit space giving us the mode, x_0'.

(iii) Estimate the width of the peak from the curvature around x_0',

$$\sigma_{x'} = \left(\frac{\partial^2 I/J}{\partial x'^2}\right)^{-1/2}\bigg|_{x'=x_0'}. \tag{17}$$

(iv) Recentre and rescale via

$$x = \frac{x' - x_0'}{\sqrt{2}\sigma_{x'}}, \tag{18}$$

such that the integrand is approximately $I \sim \exp(-x^2)$ around the peak.

(v) Apply the Gauss–Hermite quadrature in x-space, which gives

$$\int_{\varpi_j}^{\varpi_{j+1}} d\varpi \, I = \sum_k w_k \frac{\sqrt{2}\sigma_{x'} \, I(\varpi(x_k))}{J(\varpi(x_k))} \exp\left(x_k^2\right). \tag{19}$$

In our application of the method, we use the Gauss–Hermite quadrature with 11 sample points. Increasing the number of sampling points has no appreciable effect on our inferred likelihood.

A major limitation of this method is that it cannot accurately integrate multimodal integrands. The integrand must be unimodal such that we can integrate around the single peak. We will discuss the implications of this in Section 3 when introducing our model. However, since this is purely a numerical rather than conceptual challenge, we hope future work can improve on our method to allow for more general models to be evaluated and with greater computational efficiency.

3 MODEL

For this work, we only consider high latitudes, $|b| > 80°$. There are several reasons for this:

(i) Dust extinction is negligible at high latitudes. Modelling the 3D distribution of dust throughout the Milky Way is a complicated problem on its own (Marshall et al. 2006; Green et al. 2014). We quantify the impact of dust extinction on our results in Paper II.

(ii) The in-plane structure of the Milky Way disc is complex with waves, spiral arms, and the bar that add vast numbers of free parameters to any spatial model.

(iii) Parallax integration is computationally expensive and scales linearly with the number of sources. By focusing on a subset of *Gaia* data, we are left with a computationally tractable problem.

The aim of this work is to demonstrate how *Gaia* parallax information can be used to obtain an unbiased model of the Milky Way's stellar content. The vertical distribution at the Solar neighbourhood is a tractable first step in this direction.

The vertical distribution of sources is assumed to be a mixture of three distinct components: thin disc, thick disc, and halo. This canonical model has been used for decades since the addition of the second disc component by Gilmore & Reid (1983). More recent work has shown that – rather than a dichotomy into thin and thick discs – there may be a continuous evolution of disc height with stellar metallicity (Bovy, Rix & Hogg 2012a; Bovy et al. 2016). However, since metallicity is not an observable in our sample, we keep to the canonical distinct thin and thick disc model.

Within each component, we assume the spatial and absolute magnitude distributions are separable such that

$$f(l, b, \varpi, M_G) = \sum_{c=\{\text{Tn,Tk,H}\}} w_c \, \nu_c(l, b, \varpi, \boldsymbol{\psi}_\nu) \, \phi_c(M_G, \boldsymbol{\psi}_\phi). \tag{20}$$

This is a significant assumption. The thin disc has undergone star formation over long periods and will have correlations between age and metallicity and the vertical and radial dispersion of orbits (e.g. Ivezić et al. 2008; Recio-Blanco et al. 2014; Martig et al. 2016; Snaith et al. 2015). Likewise, the halo is made of multiple stellar populations from *in situ* star formation and historical merger events (e.g. Helmi et al. 2018; Belokurov et al. 2018, 2020). None the less, we maintain this assumption here in the interests of keeping a simple and tractable model.

We have deliberately chosen to assume a separable thin disc – thick disc – halo Milky Way as this provides a simple and tractable application of our method that we introduced in Section 2 whilst still returning physically informative parameters. It will be worthwhile applying our method to *Gaia* data with more detailed model parametrizations. It is beyond the scope of this work because

the exact choice of model will depend on the scientific interests of the researcher so we wish to leave this open.

Note that w_c is a free parameter of the model for each component and gives the total number of stars for that component within the given region of the sky and absolute magnitude range.

3.1 Spatial distributions

We consider the thin and thick discs to have exponential profiles vertically $\nu_c \propto \exp(-|z|/h_c)$ similar to previous work (e.g. Jurić et al. 2008; Bovy et al. 2012b, 2016). Other possibilities include sec$h$ or sech^2 profiles, but there is a moderate preference in the data for an exponential profile (Dobbie & Warren 2020).

Since we are only considering high latitudes, we neglect any radial dependence of the vertical density profile. This makes the numerical integral described in Section 2.1 significantly more tractable. The complexity introduced by adding radial dependence is explained in more detail in Appendix A. The impact of this simplification on the results is tested and quantified in Paper II.

Transforming into heliocentric coordinates $z = s\sin(b)$ and normalizing, we obtain the density distribution ,

$$\nu_c(l, b, s)\mathrm{d}V = \frac{\tan^2(|b|_{\min})}{2\pi\, h_c^3} s^2 \exp\left(-\frac{|s\sin b|}{h_c}\right)\, \mathrm{d}l\, \mathrm{d}\sin(b)\, \mathrm{d}s, \tag{21}$$

where $|b|_{\min} = 80°$ is the on-sky latitude limit of our sample. In Paper II we consider the northern and southern high latitude samples independently; however, in this work, we assume that the Galaxy is symmetric above and below the Galactic mid-plane and that the Sun lies perfectly on the plane at $z = 0$ pc. This introduces a ~ 20.8pc systematic offset into our results (Bennett & Bovy 2019), whose effect on the posterior distributions is quantified in Paper II.

For the spatial distribution of the halo, we use a spherically symmetric single power-law profile centred on the Galactic Centre, $\nu_H(r)\mathrm{d}V \propto r^{-n_H}$. Many other works also include a free parameter for the halo axial ratio (Jurić et al. 2008; Mateu & Vivas 2018); however, as we are only using a narrow window on the sky, there will be limited information to independently constrain the profile and axial ratio of the halo. Furthermore, previous works have either implicitly or explicitly truncated the halo or included a broken power-law profile. The halo profile used in this work is assumed to extend infinitely and as such a normalization constraint is placed such that $n_H > 3$. This will be in tension with Deason et al. (2014) and Fukushima et al. (2019), who find a steeper halo profile beyond $r \sim 50$ and 160 kpc, respectively. This corresponds to a parallax $\varpi < 0.02$ mas, which is pushing the precision limit of *Gaia* parallaxes even for bright sources (see fig. 7 of Lindegren et al. 2021). Therefore, our model should not be significantly sensitive to this shift.

As we did for the disc profile, we neglect cylindrical radius dependence for the halo by placing all sources at the same projected distance from the Galactic Centre as the Sun, R_0, such that

$$r^2 = s^2 \sin^2(b) + R_0^2. \tag{22}$$

Again this is only valid at high latitudes. The dependence of the source distribution on Galactic longitude can then be neglected. This may lead to systematic biases that are tested in Paper II. The spatial model of the halo is given by

$$\nu_H(l, b, s)\mathrm{d}V = \mathcal{N}_{\nu_H} s^2 \left(s^2 \sin^2 b + R_0^2\right)^{-n_H/2}\, \mathrm{d}l\, \mathrm{d}\sin b\, \mathrm{d}s, \tag{23}$$
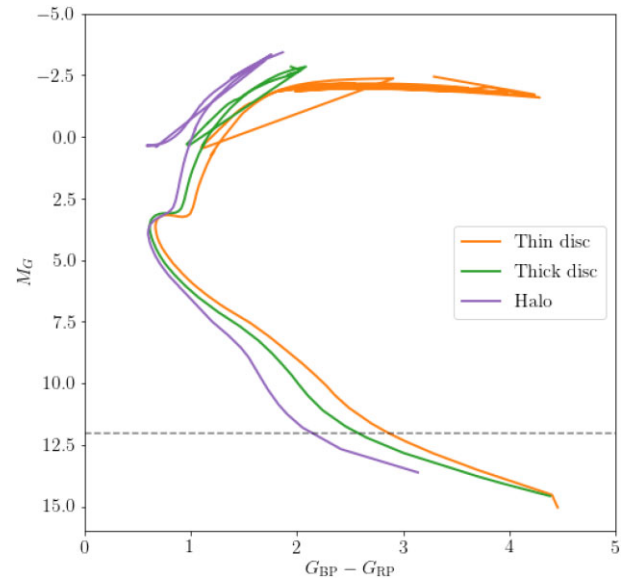

Figure 1. HR diagram showing the isochrones used for our mock model of Milky Way sources, with ages $\tau = 6.9$, 7.8, and 12.5 Gyr and metallicities [Fe/H] $= -0.3$, -0.7, and -1.5 for the thin disc, thick disc, and halo, respectively (orange, green, and purple). The grey dashed line shows the minimum absolute magnitude of our model – $M_G = 12$.

where

$$\mathcal{N}_{\nu_H} = \frac{1}{2\pi} \frac{8\tan^2(b_{\min})}{\sqrt{\pi}R_0^{3-n}} \frac{\Gamma(n/2)}{\Gamma(n/2 - 3/2)}. \tag{24}$$

This spatial distribution adds three parameters to the model: the exponential scaleheight of the thin and thick discs (h_{Tn} and h_{Tk}) and the power-law index of the halo n_H.

3.2 Luminosity functions

The luminosity distribution function of stars in the Milky Way is an intricate function of the star formation history, accretion history, and dynamical evolution of the Galaxy. The aim of this work is to derive the spatial distribution of sources in the Galaxy – independent of stellar populations – and so the magnitude distribution is only included in order to formally account for the survey selection function. In this section, we will explain how to derive an adequate parametrization for the luminosity function for each Milky Way component.

Each of the three Milky Way components is assumed to be a single mono-age, mono-abundance stellar population. Using the results of Kilic et al. (2017) from white dwarf populations, the ages used for the thin disc, thick disc, and halo are 6.9, 7.8, and 12.5 Gyr, respectively. Using SDSS spectroscopy, Ivezić et al. (2008) derived halo and thick disc metallicities of [Fe/H] $= -1.5$ and -0.7, respectively, whilst Recio-Blanco et al. (2014) used the *Gaia*-ESO survey (Gilmore et al. 2012) to find the thin disc metallicity fell in the range $[-0.8, 0.2]$ and the thick disc in the range $[-1.0, -0.25]$. Combining these results, we assume the thin disc, thick disc, and halo have metallicities of -0.3, -0.7, and -1.5. The HR diagram in Fig. 1 shows the three isochrones that are taken from PARSEC v1.2s (Bressan et al. 2012; Tang et al. 2014; Chen et al. 2014, 2015).

We then draw a random sample from the broken power law initial mass function (IMF) of Kroupa (2001) for initial masses greater than $0.09\,\mathrm{M}_\odot$ with $\mathcal{M}_{\mathrm{ini}} \sim \mathcal{M}_{\mathrm{ini}}^{-1.3}$ for $\mathcal{M}_{\mathrm{ini}} < 0.5\,\mathrm{M}_\odot$ and $\mathcal{M}_{\mathrm{ini}} \sim \mathcal{M}_{\mathrm{ini}}^{-2.3}$
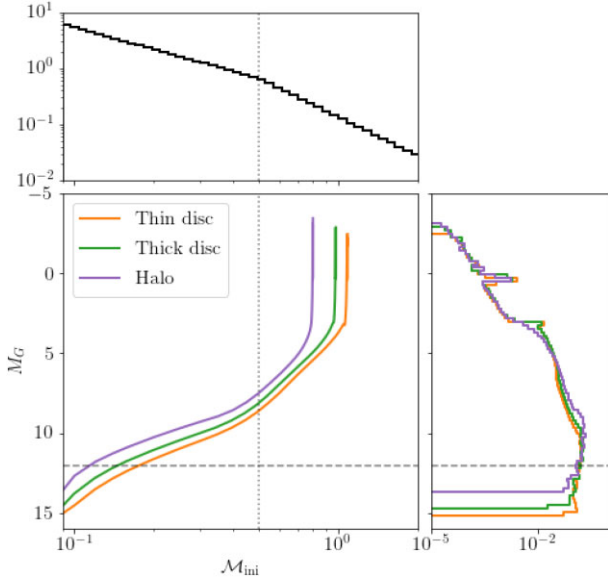
Figure 2. The thin disc, thick disc, and halo isochrones (orange, green, and purple) are used to transform a mock sample of stars from initial mass ($\mathcal{M}_{\mathrm{ini}}$) to absolute magnitude (M_{G}). The initial mass (top panel) is drawn from a Kroupa IMF (Kroupa 2001) with $\mathcal{M}_{\mathrm{ini}} > 0.09 \mathrm{M}_\odot$ where the vertical grey dotted line is the break mass $0.5\,\mathrm{M}_\odot$. This produces the absolute magnitude distribution shown in the right-hand panel. The horizontal grey-dashed line shows the maximum absolute magnitude as our model only includes sources with $M_{\mathrm{G}} < 12$.

otherwise. This is shown in the top panel of Fig. 2. The individual component isochrones, shown in the middle panel, are then used to transform the IMF into an absolute magnitude distribution that is shown in the right hand panel. This sample is not used as our mock catalogue, it is only for deriving our model absolute magnitude distribution.

The absolute magnitude distributions of the three components from the right-hand panel of Fig. 2 are shown as shaded histograms in Fig. 3. They are made up of four regimes. At the bright end ($M_{\mathrm{G}} \lesssim 3$), sources evolve much faster along the giant branch than the main sequence (MS), generating a sharp drop at the turn-off above which the number density of sources falls quickly aside from a spike at the red clump ($M_{\mathrm{G}} \sim 0$). The MS has three components, a relatively shallow upper sequence for $M_{\mathrm{G}} \sim [3, 7]$, a steeper section for $M_{\mathrm{G}} \sim [7, 9]$ where the slope of the MS in Fig. 1 shifts, which is also around the power-law break of the IMF (we will refer to this section as the 'gap'), and a very flat lower MS for $M_{\mathrm{G}} \gtrsim 9$. Sources continue fainter to the brown dwarf regime; however, stellar models in these regions of parameter space are poorly constrained by observations as there are few stars this dim yet bright enough for current observatories. For this reason, we only consider sources with $M_{\mathrm{G}} > 12$ in this work. This will be especially beneficial when we model the *Gaia* data in Paper II as the majority of sources with spurious astrometric solutions as classified by Rybizki et al. (2021) and Gaia Collaboration et al. (2021b) have absolute magnitudes fainter than $M_{\mathrm{G}} = 12$.

Each component of the absolute magnitude distribution is modelled by an exponential distribution. Here we state the parametrization; however, a full derivation of the absolute magnitude profile is given in Appendix B. The absolute magnitude is drawn from a broken exponential distribution,

$$M_{\mathrm{G}} \sim \exp(-\alpha M_{\mathrm{G}}), \qquad (25)$$
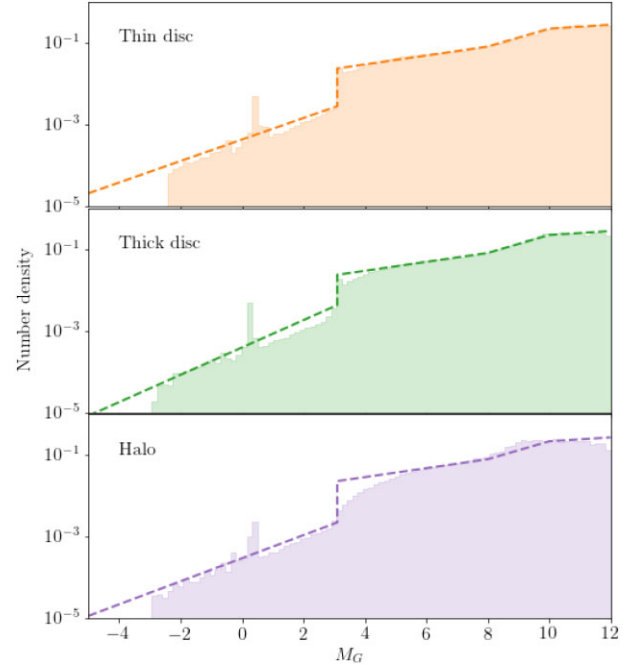

Figure 3. The mock distributions produced by transforming the Kroupa IMF through isochrones from Fig. 1 (shaded histograms) are fit with the approximate absolute magnitude distribution used for the model (dashed lines).

with four components

$$\alpha = \begin{cases} \alpha_1 & 9 < M_{\mathrm{G}} < 12 & (\text{Lower MS}) \\ \alpha_{\mathrm{g}} & 7 < M_{\mathrm{G}} < 9 & (\text{MS, 'gap'}) \\ \alpha_2 & M_{\mathrm{TO}} < M_{\mathrm{G}} < 7 & (\text{Upper MS}) \\ \alpha_{\mathrm{G}} & M_{\mathrm{G}} < M_{\mathrm{TO}} & (\text{Giants}) \end{cases} \qquad (26)$$

where M_{TO} is the turn-off magnitude.

The distribution is continuous everywhere except from at the turnoff where the discontinuous change in the gradient of the magnitude–initial mass relation leads to a discontinuity in the magnitude distribution. Continuity conditions at $M_{\mathrm{G}} = 7$ and 9 constrain the exponential profile α_{g} and the normalization A_{g} of the gap profile.

The full magnitude distribution is given by

$$f(M)\mathrm{d}M = \begin{cases} (1 - f_{\mathrm{G}})\mathcal{N}_{\mathrm{D}}\frac{1}{a_1} \\ \quad \times \exp(-\alpha_1(M - M_{\mathrm{MS}}))\,\mathrm{d}M & 9 < M < 12 \\ (1 - f_{\mathrm{G}})\mathcal{N}_{\mathrm{D}}A_{\mathrm{g}} \\ \quad \times \exp\left(-\alpha_{\mathrm{g}}(M - M_{\mathrm{MS}})\right)\mathrm{d}M & 7 < M < 9 \\ (1 - f_{\mathrm{G}})\mathcal{N}_{\mathrm{D}}\frac{1}{a_2} \\ \quad \times \exp(-\alpha_2(M - M_{\mathrm{MS}}))\,\mathrm{d}M & M_{\mathrm{TO}} < M < 7 \\ f_{\mathrm{G}}\mathcal{N}_{\mathrm{G}} \\ \quad \times \exp(-\alpha_{\mathrm{G}}(M - M_{\mathrm{TO}}))\mathrm{d}M & M < M_{\mathrm{TO}} \end{cases}, \qquad (27)$$

where $\mathcal{N}_{\mathrm{D}}$ and $\mathcal{N}_{\mathrm{G}}$ are the normalizations of the dwarf and giant magnitude distributions, respectively. $M_{\mathrm{MS}} = 8$ is the magnitude of the transition from the lower to upper MS.

The magnitude distribution introduces five parameters: α_1, α_2, M_{TO}, α_{G}, and f_{G}, the fraction of the population that are giants, which constrains the size of the discontinuity at the turn-off. We could fix all parameters using the IMF-isochrone sample just constructed. However, this is only an approximate representation of the magnitude distribution that may introduce large systematics. To avoid this problem, we free up α_1, α_2, and f_{G} to be constrained by the real

data. α_1 and α_2 are assumed to be the same for all populations as the MS is dominated by older stars that show a similar distribution independent of population parameters.

The position of the turn-off, M_{TO}, defines a discontinuity for the model. Depending on the location of individual sources in relation to the turnoff, this can generate sample-dependent local optima in the likelihood space that is challenging for optimization. For this reason, we fix $M_{TO} = 3.1$ for all models and address the implications of this in Paper II. α_G has a strong degeneracy with f_G as both control the number of sources at bright magnitudes. We also fix α_G in all optimizations to avoid this degeneracy to values that are discussed in Section 4. All free parameters are listed in Table 1 with their respective components.

This fully defines the model that we fit to the *Gaia* data. In total, there are 11 free parameters of the model.

3.3 Selection function

A major obstacle to using a catalogue of sources to fit a distribution is the selection function. Many surveys have complex and unknown observation limitations that are a strong function of observatory properties and observing conditions. *Gaia* is no exception due in part to the complexity of the scanning law (Boubert, Everall & Holl 2020; Boubert et al. 2021).

In most previous works, the sample is either assumed to be magnitude complete to some limit (e.g. Jurić et al. 2008; Bilir et al. 2006; Ak et al. 2008), or the sample is bright and nearby for which there are larger, complete catalogues against which the selection function has been estimated (e.g. Bovy 2017; Mateu & Vivas 2018; Bennett & Bovy 2019). *Gaia* is neither complete in position on the sky or apparent magnitude, nor is there a larger, more complete sample against which to compare the *Gaia* catalogue.

Fortunately, a solution for the *Gaia* source catalogue selection function has been developed and applied to *Gaia* DR2 (Boubert & Everall 2020). Appendix A of Everall & Boubert (2021) provides a simple extension to model the selection function of the *Gaia* EDR3 source catalogue using the nominal EDR3 scanning law. This may have some limitations in crowded regions due to changes in *Gaia*'s data processing pipeline. However, since we are only considering high latitude fields, it should be sufficient for our purposes. The selection probability as a function of apparent magnitude for $b = 90°$ is given by the green dashed line in Fig. 4, showing that the source catalogue is nearly complete for $3 < G < 21$.

Given the source catalogue selection function, the selection functions of subsets can be estimated by comparison (Boubert & Everall 2021; Everall & Boubert 2021). In Paper II, we will use the *Gaia* astrometry catalogue with RUWE < 1.4 where apparent G-band magnitude is available. The selection function for this data set is given by the product of the source catalogue and subset selection functions :

$$\mathcal{S}_{subset}(l, b, G) = P(\mathcal{S}_{subset} \,|\, \mathcal{S}_{source}, l, b, G) \, P(\mathcal{S}_{source} \,|\, l, b, G), \quad (28)$$

where $P(\mathcal{S}_{source} \,|\, l, b, G)$ is the probability of selection in the *Gaia* source catalogue with published G and $P(\mathcal{S}_{subset} \,|\, \mathcal{S}_{Gaia}, l, b, G)$ is the probability of an object in the source catalogue having published parallax with RUWE < 1.4, modelled in Everall & Boubert (2021), as a function of G and position on the sky only. When fitting the model parameters to data, equation (28) is substituted into equation (9).

The results are applied in 0.2-mag bins in G in NSIDE $= 64$ HEALPIX pixels (Górski et al. 2005) across the sky. The selection probability for $b = 90°$ is given by the red line in Fig. 4. Due to the

challenges of modelling sources that saturate the *Gaia* CCDs at the bright end of the magnitude distribution, we use a selection function that truncates at $G = 5$. Our sample will also only include those sources with $G > 5$.

4 MOCK

To test and demonstrate the efficacy of the method, we generate a mock catalogue from our model with realistic parameters. Information on the true parameters is then removed, the *Gaia* selection function and *Gaia*-like parallax uncertainties are applied, and we attempt to infer the input parameters from the mock sample. We note that this only tests the method. Because the data are drawn from the same model that is being refit, any inconsistencies between the model and true Milky Way distribution of stars do not show up here. These inconsistencies are discussed, tested and quantified in Paper II.

4.1 Input parameters

Parameters for the scaleheights and power-law indices of the discs and halo respectively are taken from the literature. For the thin disc, $h_{Tn} = 300$pc, and for the thick disc, $h_{Tk} = 900$pc (Jurić et al. 2008). The power-law index used is $n_H = 3.74$ from Fukushima et al. (2019).

The relative stellar mass density of the discs is $\rho_{Tk}/\rho_{Tn} = 0.12$ and $\rho_H/\rho_{Tn} = 0.005$ (Jurić et al. 2008). Instead of local mass density, our model fits the total number of sources in each component with $|b| > 80°$. To convert mass density into number density in the Solar neighbourhood, we divide by the mean mass of a star. The mean mass is estimated using the IMF-isochrone sample in Section 3.2 as $\mathcal{M} \sim 0.413, 0.369, 0.308$ M$_\odot$ for the thin disc, thick disc, and halo, respectively. We then divide the number density by the value of the normalized component at $s = 0$ to get the total number of sources in each component. The result is that $w_{Tn}/w_{Tk} = 0.275$ and $w_{Tn}/w_H = 0.0127$. The halo dominates the total counts because our observing volume is a cone with $|b| > 80°$. This significantly reduces the relative contribution from the disc to the sample.

The absolute magnitude distributions for each Milky Way component are shown by the shaded histograms in Fig 3. To estimate magnitude parameters for the luminosity function described in Section 3.2, we directly fit the parameters to the magnitude distributions. For each component, the turn-off magnitude is at $M_G \sim 3.1$. f_G is approximated from the ratio of sources with $G < 3.1$ to those with $G > 3.1$. For $G < 3.1$, we fit a power-law profile to each component independently using the Poisson likelihood function from equation (1). This gives $\alpha_3 = -0.60, -0.77$, and -0.64 and $f_G = 0.0045, 0.0054$, and 0.0035 for the thin disc, thick disc, and halo, respectively.

The lower MS is dominated by old, long-lived stars that evolve slowly on the HR diagram. Therefore, we assume that the MS profiles are similar between different Milky Way components such that the values of α_1, α_2 are shared between profiles. We draw a sample of sources from each of the components according to the component's respective weight and fit the MS profiles to the sources with $G > 3.1$, which gives $\alpha_1 = -0.12$, $\alpha_2 = -0.26$. The dashed lines in Fig. 3 give the absolute magnitude distributions implied by the parameters we have just derived.

All selected and evaluated parameter values are listed as 'Input' in Table 2.

Table 1. The 11 free parameters used to model the spatial and absolute magnitude distributions of sources along with their priors.

Component	Parameter	Prior	Transformation	Bounds
Thin disc	w	Dirichlet($a = 2$)	$\log(w)$	$[-10, 50]$
	h_{Tn}	$U[0.1, 0.6]$		$[-10, 10]$
			$\text{logit}((h - 0.1)/(0.6 - 0.1))$	
	f_{D}	$U[0, 1]$	$\text{logit}(f_{\text{D}})$	$[-10, 10]$
Thick disc	w	Dirichlet($a = 2$)	$\log(w)$	$[-10, 50]$
	h_{Tk}	$U[0.6, 3.0]$		$[-10, 10]$
			$\text{logit}((h - 0.6)/(3.0 - 0.6))$	
	f_{D}	$U[0, 1]$	$\text{logit}(f_{\text{D}})$	$[-10, 10]$
Halo	w	Dirichlet($a = 2$)	$\log(w)$	$[-10, 50]$
	n_{H}	$U[3, 7.3]$		$[-10, 10]$
			$\text{logit}((h - 3)/(7.3 - 3))$	
	f_{D}	$U[0, 1]$	$\text{logit}(f_{\text{D}})$	$[-10, 10]$
Shared	α_1	$-\alpha_1 \sim \log U[e^{-5}, e^3]$	$\log(-\alpha_1)$	$[-5, 3]$
	α_2	$-\alpha_2 \sim \log U[e^{-5}, e^3]$	$\log(-\alpha_2)$	$[-5, 3]$

Notes. The method fits directly to the parameters under the given transformations where logistic priors are also included to correct for the logit transform. The bounds are applied to the transformed parameters for numerical stability of the optimization.
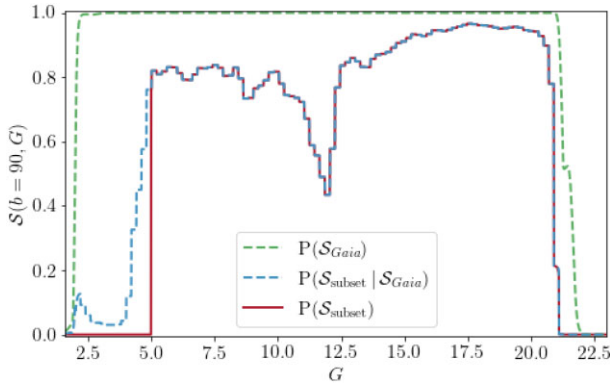

Figure 4. The selection function probability at $b = 90°$ for the *Gaia* EDR3 source catalogue (green dashed) drops off at bright magnitudes ($G < 2$, due to CCD oversaturation) and faint magnitudes ($G \gtrsim 21$); however, it remains high across the rest of apparent magnitude space. The *Gaia* EDR3 astrometry with RUWE < 1.4 relative selection function (blue dashed) is more restrictive over the entire magnitude range and dominates the total selection function (red solid). The cut-off at $G < 5$ is deliberately imposed to remove regions of apparent magnitude with poor astrometry calibration.

4.2 Parallax error

To generate a realistic mock, we also need to sample measurement uncertainties. Since the *Gaia* astrometry was fit using an iterative linear regression process, the covariance may be estimated from information theory (neglecting excess noise) using only the scanning law and individual observation centroid uncertainties. This process is performed in Everall et al. (2021b) for *Gaia* DR2 and we use the *Gaia* EDR3 nominal scanning law to extend this to the EDR3 baseline.

The covariance estimates break down for sources with significant excess noise, such as in heavily crowded regions and for sources with intrinsic astrometric variability like binaries. Since we will only consider sources with $|b| > 80°$, crowding is negligible. By focusing on the sample with RUWE < 1.4, we expect to have removed sources with observable binary motion.

4.3 Mock samples

A sample of one million sources with distance, latitude and absolute magnitude is drawn from the model using MCMC sampling (Foreman-Mackey et al. 2013). Since all sources are assumed to be at the projected distance from the Galactic Centre of the Sun, the full model is Galactic longitude-independent so the longitude is drawn from a uniform distribution $l \sim U[0, 2\pi]$. The distribution of drawn sources as a function of distance from the Galactic disc and absolute magnitude is given by the blue histograms in the top panels of Fig. 5.

The selection function probability is evaluated for all sources, given their position on the sky and apparent magnitude as described in Section 3.3. To generate the mock *Gaia* astrometry with RUWE < 1.4 sample, the event of a source being included is drawn from a Bernoulli distribution with the given selection probability $S_i \sim \text{Bernoulli}(\mathcal{S}(l_i, b_i, G_i))$, where $S_i = 0, 1$. Of the 1 000 000 source in the full sample, 73 132 survive the selection cuts, shown by the red histograms in the middle and bottom panels of Fig. 5.

Parallax error is evaluated from the Astrometric Spread Function described in Section 4.2. The observed parallax is drawn from a Gaussian distribution with the given error for each source $\varpi \sim \mathcal{N}(1/s, \sigma_\varpi)$. The red histograms in the bottom panels of Fig. 5 show the distribution of measured $z = \sin(b)/\varpi$, $M_{\text{G}} = G - 10 + 5\log_{10}(\varpi/\text{mas})$ after sampling ϖ from the parallax error. This significantly affects the distributions, demonstrating the importance of properly accounting for parallax uncertainty when modelling the structure of the Milky Way from *Gaia* data.

This produces three samples that can each be used to independently fit the model parameters demonstrating each stage of the method:

(i) Full sample fit with equation (1): $l^i, b^i, si, G^i \, \forall \, i$,
(ii) SF sample fit with equation (3): $l^i, b^i, s^i, G^i \, \forall \, i$, where $S_i = 1$,
(iii) SF and σ_ϖ fit with equation (9): $l^i, b^i, \varpi^i, G^i \, \forall \, i$, where $S_i = 1$.

To be clear, in sample (iii), the selection function is not dependent on measured parallax or parallax error as discussed in Section 2. We simply mean that the selection function is applied and parallax error on sources is also included. Samples (ii) and (iii) contain the exact

Table 2. The input parameters for the mock sample catalogue generation and the results of the fit to the data are shown when using the full sample with no observational errors ('Full'), the selection function with no observational errors ('SF'), and the sample with both the selection function and the added parallax errors ('SF and σ_ϖ').

Component	Parameter	Input	Full	SF	SF and σ_ϖ
Thin disc	$\log_{10}(w)$	4.0792	$4.0700^{+0.0194}_{-0.0191}$	$4.0637^{+0.0331}_{-0.0359}$	$3.9816^{+0.0586}_{-0.0657}$
	h_{Tn}	0.300	$0.301^{+0.007}_{-0.006}$	$0.301^{+0.010}_{-0.010}$	$0.281^{+0.015}_{-0.015}$
	f_G	4.50×10^{-3}	$3.73^{+1.00}_{-0.99} \times 10^{-3}$	$3.91^{+1.23}_{-1.22} \times 10^{-3}$	$3.76^{+1.43}_{-1.30} \times 10^{-3}$
	M_{TO}	3.1			
	α_3	-0.6			
Thick disc	$\log_{10}(w)$	4.6335	$4.6249^{+0.0051}_{-0.0050}$	$4.6253^{+0.0092}_{-0.0093}$	$4.6221^{+0.0200}_{-0.0198}$
	h_{Tk}	0.900	$0.891^{+0.012}_{-0.011}$	$0.884^{+0.029}_{-0.028}$	$0.812^{+0.052}_{-0.045}$
	f_G	5.40×10^{-3}	$5.76^{+0.54}_{-0.52} \times 10^{-3}$	$5.80^{+0.64}_{-0.60} \times 10^{-3}$	$5.83^{+0.69}_{-0.66} \times 10^{-3}$
	M_{TO}	3.1			
	α_3	-0.77			
Halo	$\log_{10}(w)$	5.9754	$5.9759^{+0.0005}_{-0.0005}$	$5.9662^{+0.0106}_{-0.0105}$	$5.9450^{+0.0247}_{-0.0229}$
	n_H	3.740	$3.745^{+0.001}_{-0.001}$	$3.753^{+0.020}_{-0.020}$	$3.812^{+0.068}_{-0.066}$
	f_G	3.50×10^{-3}	$3.47^{+0.06}_{-0.06} \times 10^{-3}$	$3.49^{+0.10}_{-0.09} \times 10^{-3}$	$3.48^{+0.15}_{-0.15} \times 10^{-3}$
	M_{TO}	3.1			
	α_3	-0.64			
Shared	α_1	-0.1100	$-0.1109^{+0.0003}_{-0.0004}$	$-0.1094^{+0.0014}_{-0.0015}$	$-0.1098^{+0.0020}_{-0.0020}$
	α_2	-0.2500	$-0.2524^{+0.0020}_{-0.0019}$	$-0.2534^{+0.0045}_{-0.0046}$	$-0.2521^{+0.0089}_{-0.0084}$

Note. For all parameters, we provide the median and 16th and 84th percentile uncertainties.
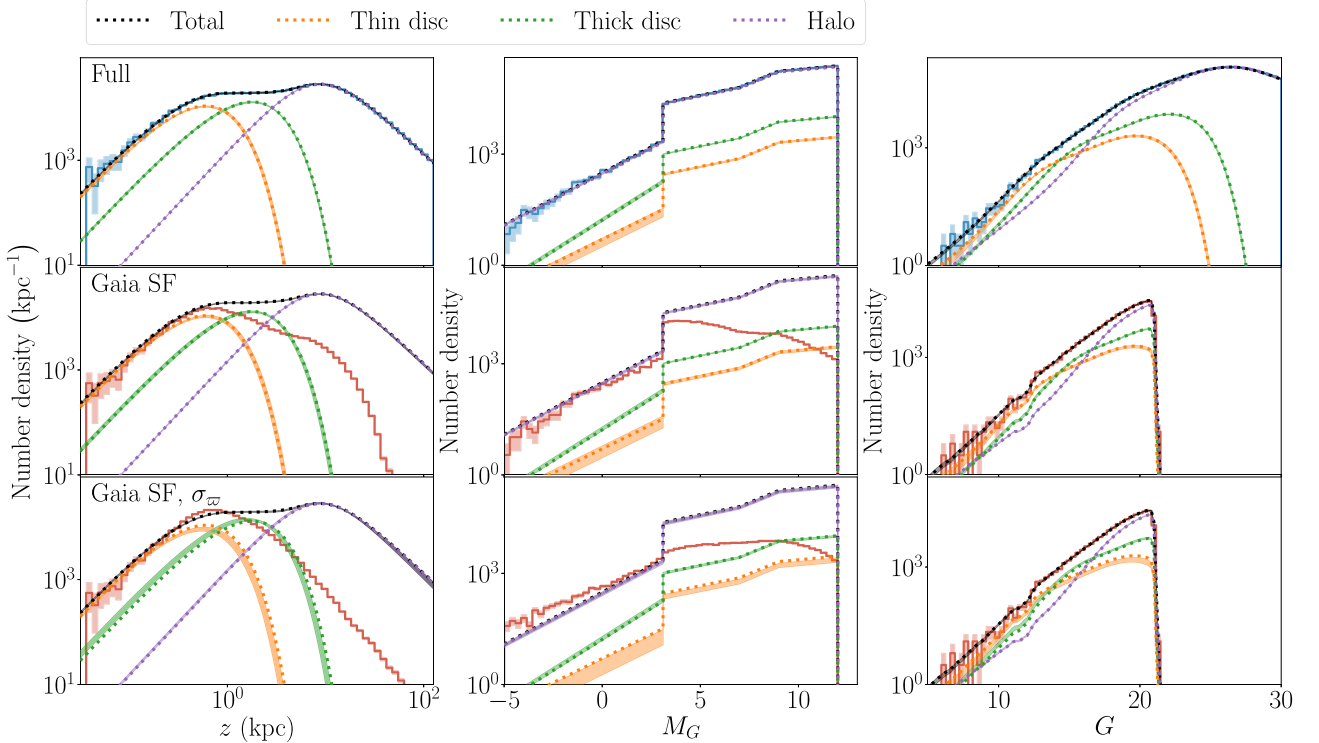

Figure 5. The posterior distribution of fits to the mock sample are shown by the shaded regions for the thin disc (orange), thick disc (green), halo (purple), and the sum total (black) as a function of vertical height (z, left-hand panel), absolute magnitude (M_G, middle panel), and apparent magnitude (G, right-hand panel). Blue histograms in the top row show the full sample that the model fit perfectly cut through. The red histograms in the middle and bottom rows show the distribution of SF selected samples with the bottom row showing the distribution of $z = \sin(b)/\varpi$ and $M_G = G + 5\log_{10}(\varpi) - 10$, demonstrating the impact of parallax uncertainties on measured quantities. The posteriors agree extremely well with the ground truth shown by the dotted lines in all panels. This is true when fitting to the full sample (top panel), the SF-limited sample (middle panel), and the SF-limited sample with measured parallaxes sampled from their error distributions (bottom panel). The posterior distributions are evaluated by randomly selecting 100 samples from the MCMC posteriors and taking the 16th–84th percentile range. In several cases, particularly for the 'Full' fits in the top row, the posterior is so tight that the distribrution appears as a line in the figure.

same subset of sources from the mock catalogue. Sample (ii) has no parallax error, whilst measured parallaxes in (iii) have been drawn from the parallax uncertainties.

5 PARAMETER INFERENCE

In this section, we will use the method introduced in Section 2 to fit the model parameters to the three mock samples described in Section 4.

5.1 Priors

Priors for all free parameters of the fits are given in Table 1. As is common with mixture model fits to density distributions, the likelihood space is strongly multimodal. For the thin and thick discs, there is of course a complete degeneracy where the components can be switched, but there are also problematic modes where, for example, a single component is expanded to fit the full data set, whilst remaining components are suppressed.

Priors are chosen specifically to avoid local optima in the model. All weights are assumed to be drawn from a Dirichlet distribution with $a = 2$ to remove modes where any component is completely suppressed relative to the others. To avoid the disc degeneracy, the possible disc scaleheights are limited to non-overlapping ranges with $h_{Tn} \sim U[0.1\text{kpc}, 0.6\text{kpc}]$ and $h_{Tk} \sim U[0.6\text{kpc}, 3.0\text{kpc}]$. The power-law index of the halo is also limited to $n_H \sim U[3.0, 7.3]$ as $n_H < 3.0$ would produce an unnormalized halo and $n_H > 7.3$ produces an incredibly steep halo profile that can mimic the exponential discs (for $n_H = 7.3$, the mean halo source distance is the same as an exponential profile with $h = 3.0$ kpc).

For numerical stability, the fits are made on the transformed parameters where transformations are given in Table 1. The transformations scale parameters to the range $[-\infty, \infty]$ in all cases. For logit transformed parameters, we include a logistic prior in logit space that is equivalent to a uniform prior in untransformed space. Therefore the logit transformation has no effect on the prior.

The L-BFGS-B algorithm requires boundaries on all parameters that are given in the final column of Table 1. The boundaries are chosen to avoid regions of parameter space that suffer from numerical precision issues. None of the parameter posterior distributions push up against the boundaries.

5.2 Optimization

The likelihood optimization is performed in three stages. All MCMC processes used EMCEE (Foreman-Mackey et al. 2013). First, a set of samples is drawn from the parameter priors using MCMC with 44 walkers (this is four times the number of free parameters in our model), with 100 step burn-in and 100 steps of sampling. Secondly, 10 samples are randomly selected from the prior samples as initialization for gradient descent using L-BFGS-B (Zhu et al. 1997) as implemented in SCIPY. Finally, the maximum likelihood estimate with the highest likelihood is taken as the best fit solution. A secondary MCMC process is initialized with 44 walkers drawn from a Gaussian ball around the maximum likelihood estimate with variance of 10^{-10} times the boundary width. These walkers were run with the likelihood $\times$ prior for 5000 steps. The latter 2500 steps are used at five-step intervals as the posterior samples. This process is used for fitting all mock samples and the real *Gaia* data in Paper II.

5.3 Results

The 'Full' sample posteriors, given by the blue contours in Fig. 6, provide tight solutions around the input parameter values that are shown by the black dot. A more quantitative comparison can be made from Table 2 that shows that the majority of input parameters fall within the 16th–84th percentile range of the posterior distribution. The top panels of Fig. 5 compare the ground truth input model, shown with dotted lines, to the refit model, shown by the narrow shaded regions. To produce the shaded posteriors in Fig. 5, we draw 100 samples from the posterior parameter distributions and plot the 16th–84th percentile range as a function of z, M_G, and G. The posteriors are so tight in most cases that the shaded regions appear as lines perfectly tracking the input model and the total of the components in black sits exactly on top of the blue histograms that show the distribution of the data in the sample.

The 'SF' sample, fit to only 73 132 of the initial one million mock sources, has a significantly less tight constraint around the true parameters, shown by the red contours in Fig 6, but the parameters show no significant bias. The fits to the halo parameters are slightly shifted from the true values but all parameters are well within 2σ of the input so this can be well explained by correlated noise, particularly considering the negative correlation between the halo weight and power-law index, n_H. The red histograms in the middle panels of Fig. 5 show the selection-limited sample that drops significantly at large vertical heights and faint apparent magnitudes demonstrating how much the model has to extrapolate using the selection function. Again, the model posteriors sit perfectly on the input model shown by the dotted lines.

For the apparent magnitude distribution in the middle right-hand panel of Fig. 5, we show the model multiplied by the selection function probability. The total model (black) sits perfectly on top of the red sample histograms demonstrating how successfully the model is fit to the data. This distribution will be especially important when analysing fits to the real *Gaia* data when we cannot directly infer the distance of stars from the Galactic plane or their absolute magnitudes due to significant parallax uncertainties.

The 'SF and σ_ϖ' posterior, given by the purple contours in Fig. 6, has significantly enhanced uncertainty compared with the solely SF limited data. This demonstrates how much information is held in the parallax and how information is lost when realistic *Gaia* parallax uncertainties are included. In spite of this, the input parameters are still recovered with reasonable precision and good accuracy. In the bottom panels of Fig. 5, we can see the posterior samples produce a clearer spread around the input distribution. This time the thin and thick discs have not been perfectly fit within the posteriors however the difference is still small enough to be well explained by statistical noise.

These results have demonstrated that the Poisson-likelihood method accounting for the *Gaia* selection function and parallax error is a powerful tool for recovering the spatial distribution of sources in the Milky Way. However, this only tests the self-consistency of the method; the results may still be susceptible to systematic uncertainties if the model does not represent the real Milky Way.

6 GUMS

So far, we have only tested the method on data drawn from the fitted model. But what happens when we attempt to fit our model to a more general and realistic catalogue? To test this we use the *Gaia* Universe Model Snapshot (GUMS; Robin et al. 2012), a synthetic Milky Way based on the Besançon Galaxy Model (Robin et al. 2003) that was
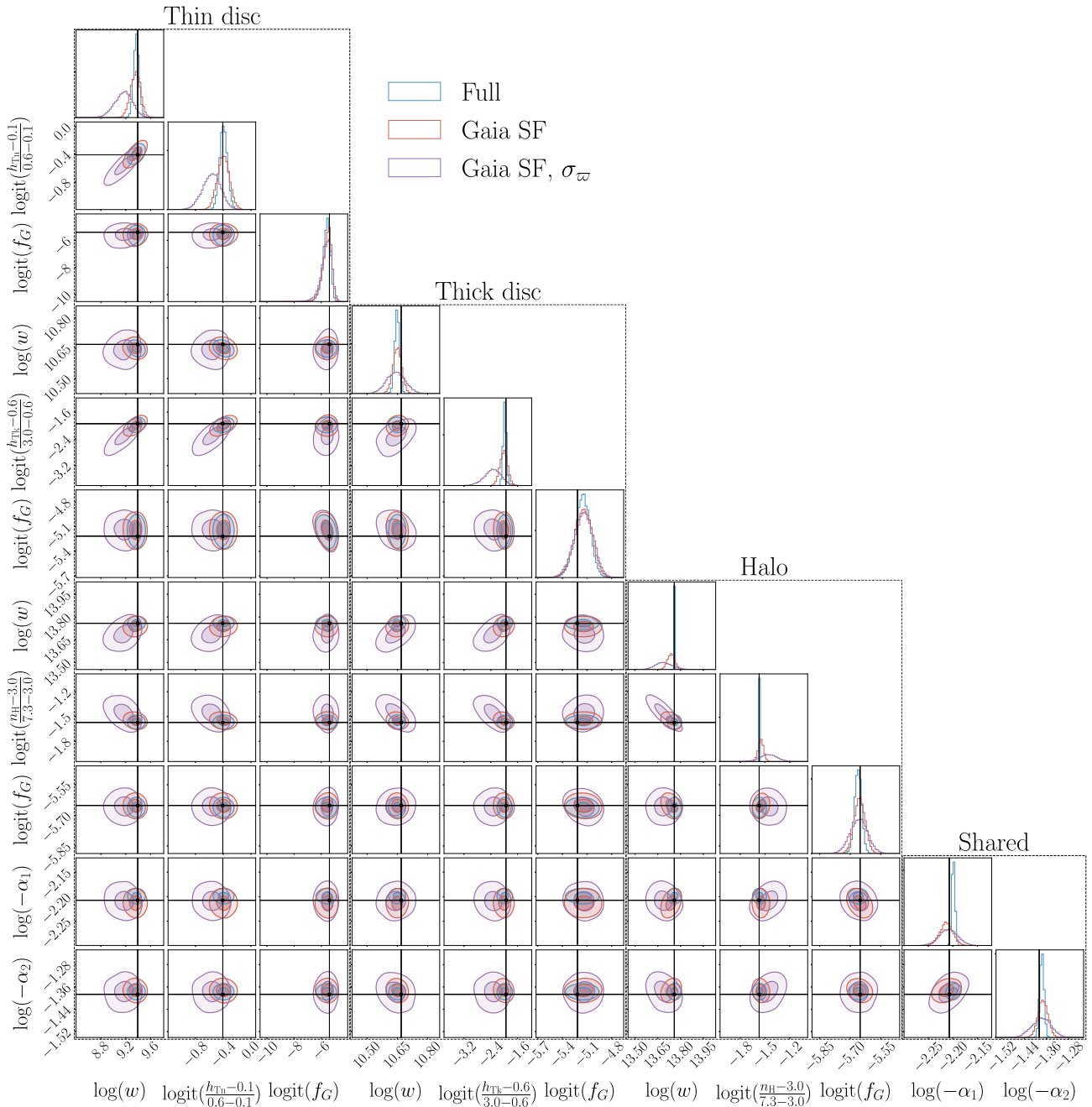
Figure 6. The posterior distributions for all mock samples are shown as a function of transformed parameters that are fit to the data. The Full sample fits (blue), SF sample (red), and SF with parallax error (purple) all show strong agreement with one another and the input parameters (black lines). The enhancement of the statistical uncertainty by introducing parallax error can clearly be seen by the increased spread of the posterior for the purple contours.

developed to test the *Gaia* data processing pipeline with a realistic population of sources.

The latest GUMS sample is provided with *Gaia* EDR3 and described in the *Gaia* documentation.[1] We will provide a very brief overview of the key components.

The thin disc is composed of seven mono-age populations each contributing a sum of square-exponential radial and vertical profiles

to the Milky Way disc. The scaleheights of the profiles increase with age, with ages ranging from 0 to 10 Gyr. The thick disc is based on the results of Robin et al. (2014) and consists of a sum of two components that are exponentially distributed in Galactocentric radius and reciprocal-cosh -square distributed in vertical height above the mid-plane with ages 10 and 12 Gyr, and scaleheights 400 and 795 pc, respectively. Flaring is also applied to the disc profiles; however, this only takes effect for $R > 10$ kpc so should not affect our analysis. The spheroidal halo is power law distributed with $n_{\rm H} \sim 3.77$ for Galactocentric distances of $r > > 2.2$ kpc and is slightly oblate with $q = 0.77$.

[1] https://gea.esac.esa.int/archive/documentation/GEDR3/Data_processing/chap_simulated/sec_cu2UM/ssec_cu2starsgal.html.
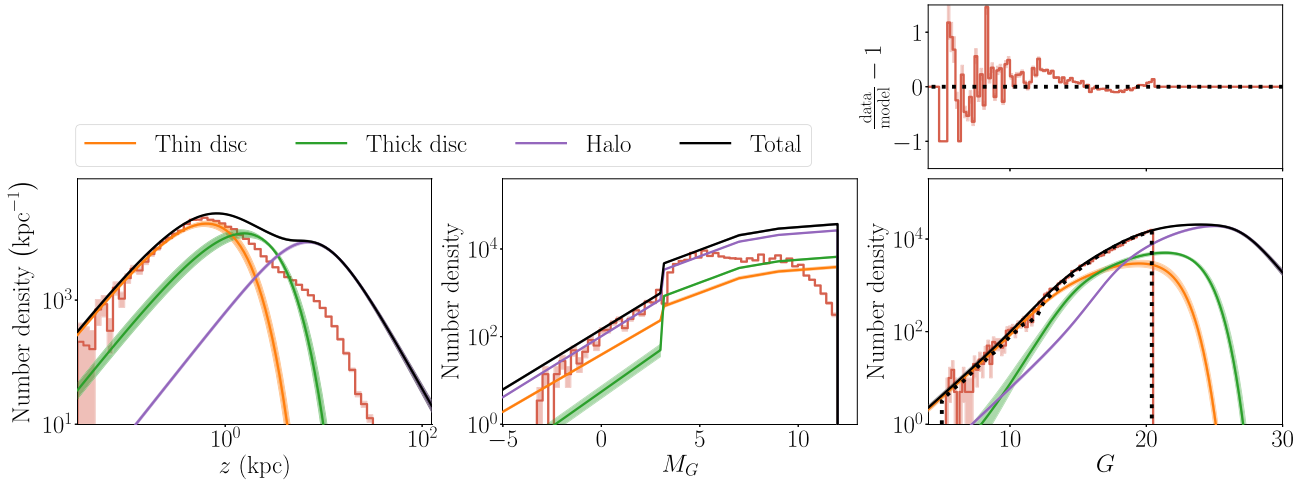
Figure 7. Bottom row: number density of sources in sample (red histograms) and predicted by the model in the thin disc (orange), thick disc (green), halo (purple), and sum total (black). Model shaded regions show the 1st–99th percentiles of posterior parameter fits to the data. Top right-hand panel: relative residual of the data from the median model fit showing that the model produces a small but significant underestimate of the data at bright magnitudes and overestimate at fainter magnitudes. The shaded red regions show the 1 standard deviation Poisson uncertainties of the bin counts.

The absolute magnitudes of the populations are determined by sampling from an IMF and star formation history and using stellar evolution tracks. The thin discs use a constant star formation history whilst instantaneous bursts of star formation at 10 and 12 Gyr are used for the thick disc and 14 Gyr for the halo.

The sample provided through the *Gaia* archive includes binary and higher order systems for a significant fraction of stars. For the purposes of this study, we treat all systems as unresolvable with *Gaia*, i.e. we only include them as a single point source with flux given by the sum of all stars in the system.

We select sources with $b > 80°$ in the north and $b < -80°$ in the south from the GUMS catalogue. Due to computational limitations, we work only with a randomly-drawn 10 per cent subsample when testing our method on GUMS. The GUMS sample was cut internally to only include sources with $G < 21$, where G was estimated from simple colour relations. The published apparent G-band magnitude is computed from the GUMS synthetic spectra such that the originally sharp cut becomes a smooth drop off at $G \sim 21$ (Robin, private communication). To avoid this, we cut the sample at $G = 20.5$ and set the selection function to $\mathcal{S} = 0$ for $G > 20.5$. We then produce a *Gaia*-like mock catalogue by resampling the data from the selection function introduced in Section 3.3. This produces 47 027 sources in the north and 49 508 in the south. The distributions of the north sample as a function of height above the mid-plane, absolute magnitude, and apparent magnitude are shown by the red histograms in the lower panels of Fig. 7.

As we did for the mock in Section 4.2, we also resample a realistic observed parallax measurement for each source from the Astrometric Spread Function (Everall et al. 2021b). We then run two fits for each of the north and south samples, one fit to the sample with no parallax error applied and the other with parallax error applied. These are equivalent to fits (ii) and (iii) in Section 4.

The results of the fits to the north sample with parallax error are shown in Fig. 7. The model is only slightly above the data at small z and bright absolute magnitudes, which is unsurprising as few sources will have been removed by the selection function in these regions of parameter space. For the apparent magnitude distribution, we also apply the selection function to the total model that produces the black

dotted line. The top right-hand panel of the figure shows the relative residuals of the data from the model. At bright magnitudes, the residuals are very large but decline to approximately a few per cent for $G \gtrsim 15$. This demonstrates that our model is not flexible enough to accurately reproduce the data; however, at the fainter magnitudes, this inaccuracy is small relative to the scale of the model.

The parameter posteriors for all four fits are shown in Fig. 8. The first thing to note is that the north and south posteriors, shown by blue and red contours respectively, are consistent with one another for all parameters whether parallax error is applied or not. This demonstrates that we are correctly finding no asymmetry north or south of the Milky Way disc. The glaring problem with our results is that the halo fits are significantly different when exact distances are used (solid contours) and when parallaxes are drawn from uncertainties (dashed contours). The halo distribution peaks at $z \sim 10$ kpc corresponding to a parallax of ~ 0.1 mas, which is smaller than the parallax uncertainty for sources fainter than $G \sim 18$. Many faint sources will have low parallax signal-to-noise ratio and so will only have weakly constrained distances. At these large distances, we expect that the method is using the absolute magnitude distribution combined with the measured apparent magnitude of sources to estimate the distance distribution. Oversimplifications in our absolute magnitude model therefore significantly bias the inferred halo profile. In spite of this, the thin and thick disc profiles are remarkably resilient to the halo systematics with no parameters producing significant offsets between the two fits.

As the model parametrization used for GUMS is significantly different to our own, we cannot make direct comparisons between our parameter values and a 'ground truth' input. The thick disc scaleheight of 670–780 pc is broadly consistent with the scaleheights used to generate the population however the GUMS sample has two profiles for different age populations that bracket our inferred value.

We conclude from this that the inferred halo profile parameters are susceptible to significant systematic uncertainties when applying our method to realistic *Gaia* samples likely due to an oversimplified absolute magnitude model. However, the disc parameters are more reliable and we can draw information about the structure of the Milky
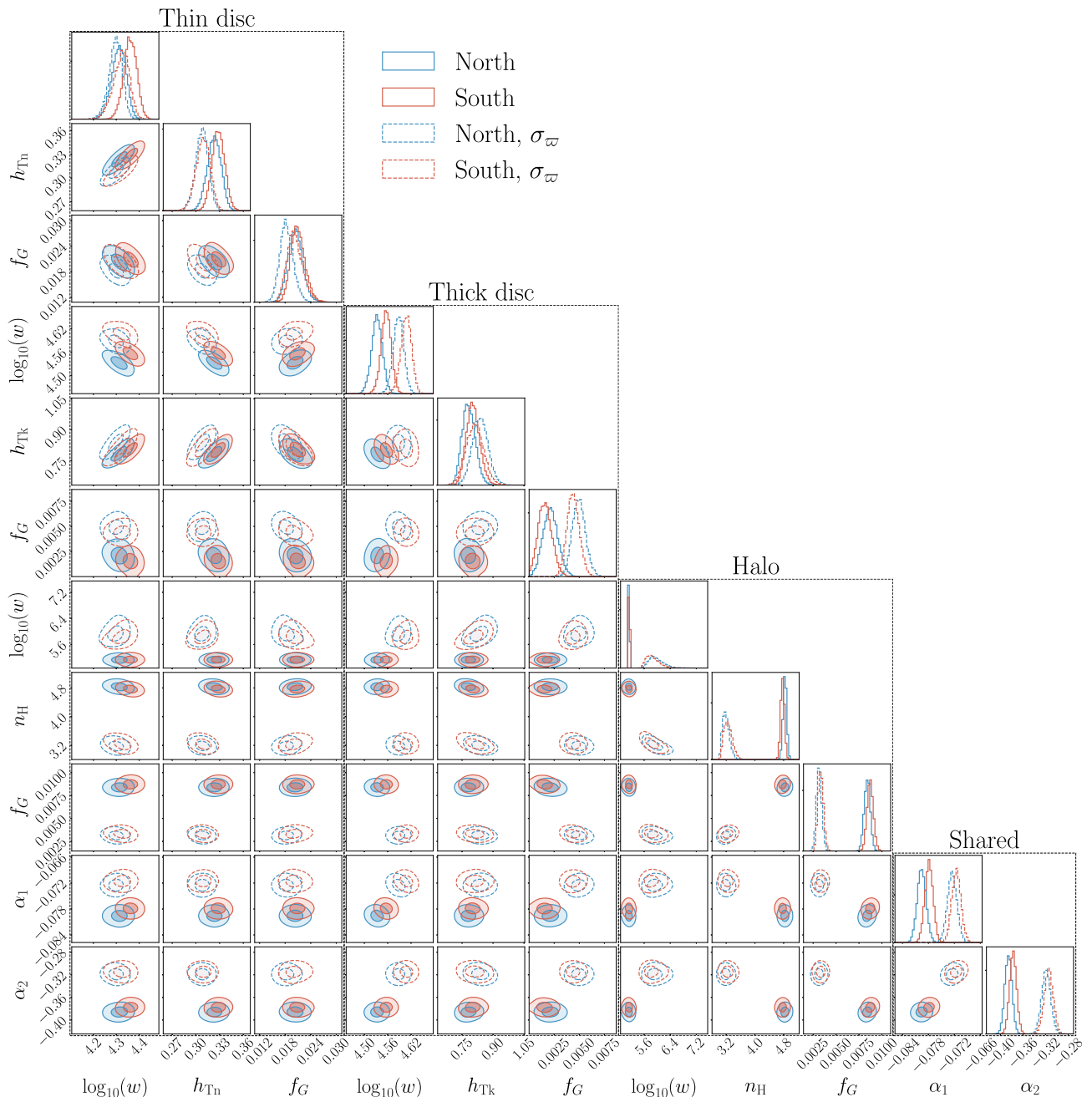
Figure 8. Posterior distribution of fits to GUMS data for the north (blue) and south (red) samples show reasonable agreement across all parameters. The fits with precise distances (solid) and parallaxes drawn from uncertainty distributions (dashed) are consistent for disc parameters but significantly disagree for the halo, likely related to a overly simplistic magnitude model.

Way disc profiles from our results when applying to the real *Gaia* data in Paper II.

Even given the consistency of our disc parameter results, we cannot guarantee that we have managed to separate out the two components. Since our model is different to the one used to generate GUMS, it is likely that some thin disc stars will be contributing to the thick disc and *vice versa*. This will happen to some extent with the real *Gaia* data too assuming that the Galactic disc can even be decomposed into discrete components, which is contested (e.g. Bovy et al. 2012a). In the scenario that the Milky Way disc does not decompose well into exponential thin and thick discs, the sum total of our disc models

that describes the total tracer density may be of greater interest to the community than the individual components that contribute to it.

In Paper II, we quantify the systematic uncertainty introduced by some of the assumptions we have applied. We test the impact of Solar position offset from the Galactic mid-plane, dust extinction, magnitude uncertainty, parallax zero-point offset, shifting the turn-off absolute magnitude, Galactocentric-radius-dependent disc and halo distributions, and an oblate stellar halo. This provides a comprehensive overview of the systematic uncertainties introduced to parameter estimates by oversimplifications in our model of stars in the Galaxy.

7 CONCLUSIONS

We have developed a method to fit the distribution of stars in the Milky Way using the Poisson likelihood function. Our method correctly accounts for the sample selection function and parallax measurement uncertainty.

The method is used to fit the vertical distribution of stars with $|b| > 80°$. For the model, we use two exponential disc components and a power-law halo. The data are also simultaneously fit with a four-piece exponential absolute-magnitude distribution.

The efficacy of our method is demonstrated against a mock sample. By refitting the model parameters, we demonstrate that the method produces results that are accurate to within the statistical uncertainties of the parameter posteriors.

We apply our method to the GUMS mock sample to infer the parameters of a population that are drawn from a far more complex Milky Way model. We obtain consistent fits when applying our model with and without parallax error for disc parameters but not for halo parameters. This suggests that our results for disc parameters are reliable when fit to data that does not exactly represent our model but that our disc parameters should be viewed with caution.

In Paper II, we set the machinery working on *Gaia* EDR3. We undertake a set of strenuous tests to quantify the systematic uncertainties in our parameter estimates due to oversimplifications in the model.

DATA AVAILABILITY

The data underlying this paper are publicly available from the European Space Agency's *Gaia* archive (https://gea.esac.esa.int/archive/). The EDR3 nominal scanning law is also available from the *Gaia* archive (http://cdn.gea.esac.esa.int/Gaia/gedr3/auxiliary/commanded_scan_law/).

The selection function implementations used in this work are described in Boubert & Everall (2020) and Everall & Boubert (2021), and made publicly accessible through the PYTHON package SELECTIONFUNCTIONS (https://github.com/gaiaverse/selectionfunctions). The Astrometric Spread Function used to generate realistic uncertainties can be accessed through the PYTHON package SCANNINGLAW (https://github.com/gaiaverse/scanninglaw).

The code used to fit the model and produce all figures is made publicly available as a GitHub repository (https://github.com/averall/mwtrace.git).

REFERENCES

Ak T., Bilir S., Ak S., Eker Z., 2008, New Astron., 13, 133
Bailer-Jones C. A. L., 2015, PASP, 127, 994
Belokurov V., Erkal D., Evans N. W., Koposov S. E., Deason A. J., 2018, MNRAS, 478, 611
Belokurov V., Sanders J. L., Fattahi A., Smith M. C., Deason A. J., Evans N. W., Grand R. J. J., 2020, MNRAS, 494, 3880
Bennett M., Bovy J., 2019, MNRAS, 482, 1417
Bilir S., Karaali S., Ak S., Yaz E., Hamzaoğlu E., 2006, New Astron., 12, 234
Boubert D., Everall A., 2020, MNRAS, 497, 4246
Boubert D., Everall A., 2022, MNRAS, 510, 4626
Boubert D., Everall A., Fraser J., Gration A., Holl B., 2021, MNRAS, 501, 2954
Boubert D., Everall A., Holl B., 2020, MNRAS, 497, 1826
Bovy J., 2017, MNRAS, 470, 1360
Bovy J., Rix H.-W., Hogg D. W., 2012a, ApJ, 751, 131
Bovy J., Rix H.-W., Liu C., Hogg D. W., Beers T. C., Lee Y. S., 2012b, ApJ, 753, 148
Bovy J., Rix H.-W., Schlafly E. F., Nidever D. L., Holtzman J. A., Shetrone M., Beers T. C., 2016, ApJ, 823, 30
Bressan A., Marigo P., Girardi L., Salasnich B., Dal Cero C., Rubele S., Nanni A., 2012, MNRAS, 427, 127
Chen Y., Bressan A., Girardi L., Marigo P., Kong X., Lanza A., 2015, MNRAS, 452, 1068
Chen Y., Girardi L., Bressan A., Marigo P., Barbieri M., Kong X., 2014, MNRAS, 444, 2525
de Jong J. T. A., Yanny B., Rix H.-W., Dolphin A. E., Martin N. F., Beers T. C., 2010, ApJ, 714, 663
Deason A. J., Belokurov V., Koposov S. E., Rockosi C. M., 2014, ApJ, 787, 30
Dobbie P. S., Warren S. J., 2020, Open J. Astrophys., 3, 5
Everall A., Belokurov V., Evans N. W., Vasiliev E., Boubert D., Grand R., 2021a, MNRAS(arXiv:2111.08043)
Everall A., Boubert D., 2022, MNRAS, 509, 6205
Everall A., Boubert D., Koposov S. E., Smith L., Holl B., 2021b, MNRAS, 502, 1908
Everall A., Das P., 2020, MNRAS, 493, 2042
Foreman-Mackey D., Hogg D. W., Lang D., Goodman J., 2013, PASP, 125, 306
Fukushima T. et al., 2019, PASJ, 71, 72
Gaia Collaboration et al., 2016, A&A, 595, A1
Gaia Collaboration et al., 2021a, A&A, 649, A1
Gaia Collaboration et al., 2021b, A&A, 649, A6
Gilmore G. et al., 2012, The Messenger, 147, 25
Gilmore G., Reid N., 1983, MNRAS, 202, 1025
Górski K. M., Hivon E., Banday A. J., Wandelt B. D., Hansen F. K., Reinecke M., Bartelmann M., 2005, ApJ, 622, 759
Grand R. J. J. et al., 2018, MNRAS, 481, 1726
Green G. M. et al., 2014, ApJ, 783, 114
Helmi A., Babusiaux C., Koppelman H. H., Massari D., Veljanoski J., Brown A. G. A., 2018, Nature, 563, 85

Ivezić Ž. et al., 2008, ApJ, 684, 287

Jurić M. et al., 2008, ApJ, 673, 864

Kilic M., Munn J. A., Harris H. C., von Hippel T., Liebert J. W., Williams K. A., Jeffery E., DeGennaro S., 2017, ApJ, 837, 162

Koposov S., Bartunov O. 2006, Astronomical Society of the Pacific Conference Series, Astronomical Data Analysis Software and Systems XV , 351, 735

Kroupa P., 2001, MNRAS, 322, 231

Lindegren L. et al., 2021, A&A, 649, A2

Lombardi M., Lada C. J., Alves J., 2013, A&A, 559, A90

Luri X. et al., 2018, A&A, 616, A9

Marshall D. J., Robin A. C., Reylé C., Schultheis M., Picaud S., 2006, A&A, 453, 635

Martig M., Minchev I., Ness M., Fouesneau M., Rix H.-W., 2016, ApJ, 831, 139

Mateu C., Vivas A. K., 2018, MNRAS, 479, 211

Perryman M. A. C. et al., 2001, A&A, 369, 339

Recio-Blanco A. et al., 2014, A&A, 567, A5

Rix H.-W. et al., 2021, AJ, 162, 142

Robin A. C. et al., 2012, A&A, 543, A100

Robin A. C., Reylé C., Derrière S., Picaud S., 2003, A&A, 409, 523

Robin A. C., Reylé C., Fliri J., Czekaj M., Robert C. P., Martins A. M. M., 2014, A&A, 569, A13

Rybizki J., Green G., Rix H.-W., Demleitner M., Zari E., Udalski A., Smart R. L., Gould A., 2021, MNRAS, 510, 2597

Snaith O., Haywood M., Di Matteo P., Lehnert M. D., Combes F., Katz D., Gómez A., 2015, A&A, 578, A87

Tang J., Bressan A., Rosenfield P., Slemer A., Marigo P., Girardi L., Bianchi L., 2014, MNRAS, 445, 4287

Zhu C., Byrd R. H., Lu P., Nocedal J., 1997, ACM Trans. Math. Softw., 23, 550

APPENDIX A: INTEGRAND LIMITATIONS

In Section 2.1, we stated that the integral over parallax uncertainty becomes intractable for more complex models, we will briefly justify that statement here where we will use the example of the exponential disc model to demonstrate.

The integrand including R-dependence is

$$I \, \mathrm{d}\varpi \propto \varpi^{-4} \exp\left(-\frac{z}{h} - \frac{R}{L}\right) \exp\left(\frac{(\varpi - \varpi_i)}{2\sigma_{\varpi i}^2}\right) \mathrm{d}\varpi \qquad (A1)$$

where h and L are the scaleheight and scalelength of the disc being considered. The Jacobian for the logit transformation we applied is

$$J \propto \frac{1}{(\varpi_{j+1} - \varpi)(\varpi - \varpi_j)}. \qquad (A2)$$

Taking the gradient of I/J, setting to zero (as in equation 16) and simplifying down, we are left with

$$-\frac{4}{\varpi} - \frac{1}{h}\frac{\partial z}{\partial \varpi} - \frac{1}{L}\frac{\partial R}{\partial \varpi} - \frac{(\varpi - \varpi_i)}{\sigma_{\varpi i}^2} + \frac{1}{(\varpi_{j+1} - \varpi)} - \frac{1}{(\varpi - \varpi_j)} = 0, \qquad (A3)$$

where

$$z = \frac{\sin b}{\varpi} \quad \text{and} \quad R^2 = R_0^2 + \left(\frac{\cos b}{\varpi}\right)^2 - \frac{2R_0 \cos b \, \cos l}{\varpi}. \qquad (A4)$$

In our application, we have assumed no R-dependence, i.e. setting $L = \infty$. We have

$$\frac{\partial z}{\partial \varpi} = -\frac{\sin b}{\varpi^2}, \qquad (A5)$$

and equation (A3) simplifies to a quintic polynomial in terms of ϖ. We know that at least two solutions of the quintic are outside $[\varpi_j,$

$\varpi_{j+1}]$, since

$$\frac{I}{J} \begin{cases} = 0 & \text{for } \varpi = \varpi_j, \varpi_{j+1} \\ < 0 & \text{for } \varpi \lesssim \varpi_j, \, \varpi \gtrsim \varpi_{j+1}, \\ = 0 & \text{for } \varpi = 0, \infty \end{cases} \qquad (A6)$$

so there must be a stationary point above and below the boundaries. This leaves three stationary points in the integration range corresponding to two peaks or modes. Our model is equivalent to the exponentially-decreasing square distance prior used by section 7 of Bailer-Jones (2015) and they also find the same two modes. However, two of the roots are often either complex, or, for $\varpi_i < 0$, there will be a mode with negative parallax that is outside the integration limits. Whilst we cannot guarantee that the integrand is always unimodal, Section 5 demonstrates that this does not have a measurable affect on our results.

If, however, we include R-dependence and have L of order unity (kpc), then the integrand significantly changes. Equation (A3) now includes

$$\frac{\partial R}{\partial \varpi} = \frac{1}{R}\left(-\frac{\cos^2 b}{\varpi^3} + \frac{R_0 \cos b \cos l}{\varpi^2}\right), \qquad (A7)$$

where R is given in equation (A4). Expanding this out, equation (A3) is now an 11th-order polynomial in ϖ. Again, two of the stationary points are outside the integration bounds due to the logit transformation but that leaves nine stationary points meaning up to five modes in the integrand.

One simplification we could take that would avoid adding any more modes to the integrand is

$$R \approx R_0 - X = R_0 - s \cos l \cos b. \qquad (A8)$$

This would provide a slight improvement on our previous models; however, it also makes the model normalization non-analytic that adds another layer of complexity. This may be an avenue worth pursuing however we consider it beyond the scope of this work.

APPENDIX B: MAGNITUDE DISTRIBUTION

To derive the absolute magnitude distribution, we start from a power-law IMF with a break at $\mathcal{M}_b = 0.5 M_\odot$:

$$f(\mathcal{M})\mathrm{d}\mathcal{M} = \begin{cases} \mathcal{N}_{\mathcal{M}_1} \exp(-\epsilon_1) & \mathcal{M} < \mathcal{M}_b \\ \mathcal{N}_{\mathcal{M}_2} \exp(-\epsilon_2) & \mathcal{M} > \mathcal{M}_b \end{cases}. \qquad (B1)$$

The continuity boundary condition at $\mathcal{M} = \mathcal{M}_b$ constrains $\mathcal{N}_{\mathcal{M}_1}\mathcal{M}_b^{-\epsilon_1} = \mathcal{N}_{\mathcal{M}_2}\mathcal{M}_b^{-\epsilon_2}$.

The initial–mass luminosity relation is assumed to approximately follow a set of power-law slopes:

$$L \propto \mathcal{M}^a, \qquad (B2)$$

such that the absolute magnitude distribution is given by

$$M = \frac{-2.5a}{\log(10)} \log(\mathcal{M}) + C, \qquad (B3)$$

where C is an unknown normalization constant.

Fig. 2 shows that the magnitude–luminosity relation changes approximately around the mass break in the IMF. We introduce the magnitude boundary M_{MS} that is the absolute magnitude approximately corresponding to the mass $\mathcal{M}_b$:

$$M = \begin{cases} \frac{-2.5a_1}{\log(10)} \log(\mathcal{M}) + C_1 & M > M_{MS} \\ \frac{-2.5a_2}{\log(10)} \log(\mathcal{M}) + C_2 & M < M_{MS} \end{cases}. \qquad (B4)$$

Applying a continuity condition for the magnitude–initial mass relation at M_{MS} constrains :

$$C_2 = C_1 + \frac{\log(10)M_{MS}}{2.5a_2}\left(\frac{1}{a_2} - \frac{1}{a_1}\right).$$

(B5)

We can now construct the absolute magnitude distribution:

$$f(M)dM = f(\mathcal{M})d\mathcal{M}$$

(B6)

$$= f(M(\mathcal{M}))d\left|\frac{\partial \mathcal{M}}{\partial M}\right|dM$$

(B7)

$$= \mathcal{N}\begin{cases}\frac{1}{a_1}\exp\left(-\alpha_1(M - M_{MS})\right)dM & M > M_{MS} \\ \frac{1}{a_2}\exp\left(-\alpha_2(M - M_{MS})\right)dM & M < M_{MS}\end{cases}.$$

(B8)

The model derived so far assumes a discontinuous change in the gradient of the magnitude–initial mass relation; however, Fig. 2 clearly shows that there is a continuous change between modes. To reflect this, the model shifts between regimes across a range of apparent magnitudes. We refer to the intermediate magnitude range as the 'gap' and use an extra exponential profile that connects smoothly into the lower and upper MSs:

$$f(M)dM = \mathcal{N}\begin{cases}\frac{1}{a_1}\exp\left(-\alpha_1(M - M_{MS})\right)dM & M_{MS1} < M \\ A_g\exp\left(-\alpha_g(M - M_{MS})\right)dM & M_{MS2} < M < M_{MS1} \\ \frac{1}{a_2}\exp\left(-\alpha_2(M - M_{MS})\right)dM & M < M_{MS2}\end{cases}.$$

(B9)

This introduces two boundary conditions that are continuity conditions at M_{MS1} and M_{MS2}. Applying the boundary conditions fully constrains both A_g and α_g:

$$\alpha_g = \frac{\log\left(\frac{a_1}{a_2}\right) - \alpha_1(M_{MS} - M_{MS1}) + \alpha_2(M_{MS} - M_{MS2})}{M_{MS1} - M_{MS2}}$$

(B10)

$$A_g = \frac{1}{a_1}\exp\left((\alpha_g - \alpha_1)(M_{MS} - M_{MS1})\right).$$

(B11)

This gives us our MS distribution. However, the giants follow a steeper track with a sharp drop at the turn-off magnitude, M_{TO}. For this, a final exponential component is included with an independent normalization to the MS:

$$f_G(M)dM = \mathcal{N}_G\exp(-\alpha_G(M - M_{TO}))\quad M < M_{TO},$$

(B12)

where $\mathcal{N}_G = -1/\alpha_G$, to normalize the giant distribution.

Finally, putting this all together, the fraction of all sources that are dwarfs (i.e. have $M > M_{TO}$) is parametrized by f_D. The full magnitude distribution is given by

$$f(M)dM = \begin{cases}f_D\mathcal{N}_D\frac{1}{a_1}\exp\left(-\alpha_1(M - M_{MS})\right)dM & M_{MS1} < M < M_X \\ f_D\mathcal{N}_D A_g\exp\left(-\alpha_g(M - M_{MS})\right)dM & M_{MS2} < M < M_{MS1} \\ f_D\mathcal{N}_D\frac{1}{a_2}\exp\left(-\alpha_2(M - M_{MS})\right)dM & M_{TO} < M < M_{MS2} \\ (1 - f_D)\mathcal{N}_G\exp(-\alpha_G(M - M_{TO}))dM & M < M_{TO}\end{cases},$$

(B13)

where $\mathcal{N}_D$ is the normalization of the full MS. In order to make this well normalized, an upper absolute magnitude limit, M_X, has been placed on the lower MS. This also cuts the distribution off before it reaches the end of the information from isochrones. At these magnitudes, there are very few visible stars and those that are in the *Gaia* data set will be nearby with well constrained parallaxes enabling them to be easily removed from the sample as contamination.

This paper has been typeset from a TeX/LaTeX file prepared by the author.