Marginal likelihoods of distances and extinctions to stars: computation and compact representation

S. E. Sale, J. Magorrian

Monthly Notices
of the
ROYAL ASTRONOMICAL SOCIETY

MNRAS **448,** 1738–1750 (2015) doi:10.1093/mnras/stv068

Marginal likelihoods of distances and extinctions to stars: computation and compact representation

S. E. Sale[1]★ and J. Magorrian[1,2]

[1]*Rudolf Peierls Centre for Theoretical Physics, Keble Road, Oxford OX1 3NP, UK*
[2]*Institut d'Astrophysique de Paris, 98 bis Boulevard Arago, 75014 Paris, France*

Accepted 2015 January 10. Received 2015 January 9; in original form 2014 November 12

ABSTRACT

We present a method for obtaining the likelihood function of distance and extinction to a star given its photometry. The other properties of the star (its mass, age, metallicity and so on) are marginalized assuming a simple Galaxy model. We demonstrate that the resulting marginalized likelihood function can be described faithfully and compactly using a Gaussian mixture model. For dust mapping applications we strongly advocate using monochromatic over bandpass extinctions, and provide tables for converting from the former to the latter for different stellar types.

Key words: methods: statistical – stars: distances – dust, extinction.

1 INTRODUCTION

Our present lack of knowledge of the three-dimensional (3D) distribution of interstellar dust is a significant barrier to building a complete picture of our Galaxy. Like the Sun, most of the Galaxy's stars lie close to the plane, which means that their light is subject to significant extinction before it reaches us. Therefore any attempt to construct a complete model of the Galaxy's stellar density distribution must include a 3D model of the extincting dust distribution. Dust is also interesting in its own right as a tracer of the densest parts of the interstellar medium (ISM). Consequently there is now a growing industry devoted to understanding and mapping extinction, with a number of authors either presenting methods for mapping extinction (e.g. Majewski, Zasowski & Nidever 2011; Sale 2012; Green et al. 2014; Hanson & Bailer-Jones 2014) or constructing actual maps of extinction in two or three dimensions (e.g. Marshall et al. 2006; Lallement et al. 2014; Planck Collaboration XI 2014; Sale et al. 2014).

A superficially attractive and straightforward way of producing a 3D extinction map is by first using a method such as Berry et al. (2012) and Hanson & Bailer-Jones (2014) to calculate posterior expectations for the distances and extinctions to large numbers of stars individually, then binning the results spatially to produce a map. Unfortunately, this produces maps that are biased in a complicated manner. There are three principal sources of bias. First, almost any catalogue of stars will itself not be an unbiased sample of the stars in the Galaxy. Most catalogues are magnitude limited, which biases them towards less extinguished, and therefore brighter, stars. Dealing effectively with such selection effects is not trivial and is the subject of Sale (in preparation). Second, we expect that extinction along two nearby sightlines should be correlated: two-dimensional (2D) projected dust maps exhibit correlations on scales ranging from less than 1 pc (e.g. di Francesco et al. 2010) up to that of spiral arms. Third, the posteriors distributions of the distances and extinctions to individual stars are frequently extended and exhibit complicated forms. As a result the posterior expectations of distance and extinction will not transmit the full range of uncertainties nor the complex correlations that exist between distance and extinction.

In Sale & Magorrian (2014) we presented a new method for mapping extinction from star counts that avoids these problems. Building on earlier work in Vergely et al. (2001) and Sale (2012), we used a simple physical model of Kolmogorov turbulence to impose spatial correlations on the density map, which prevents the formation of non-physical 'fingers of God' as found in the maps in Marshall et al. (2006) and Sale et al. (2014). The method avoids the need for spatial binning, producing (realizations of) extinction maps whose resolution is set naturally by the available data.

Most of the 3D extinction mapping procedures mentioned above, including that of Sale & Magorrian (2014), share the requirement that one have some way of calculating the marginal likelihoods of distances and extinctions to individual stars. That is, having some observations $\boldsymbol{y}$ (photometry and/or spectroscopy and/or astrometry) of a single star at Galactic coordinates (l, b), they need the likelihood $\mathrm{p}(\boldsymbol{y}|s, l, b, A, \alpha, \beta)$ of the distance s and extinction A to the star, in which the details of the star's mass, age, metallicity and so on have been marginalized out assuming some Galaxy model β and set of extinction laws and isochrones α. The present paper provides one way of calculating such marginalized likelihoods. We begin though by considering the problem of how best to parametrize the extinction law included in α and how to calculate the effects of extinction in a range of popular photometric passbands. This is the subject of Section 2; tables giving the results of our calculation are available

★ E-mail: stuart.sale@physics.ox.ac.uk

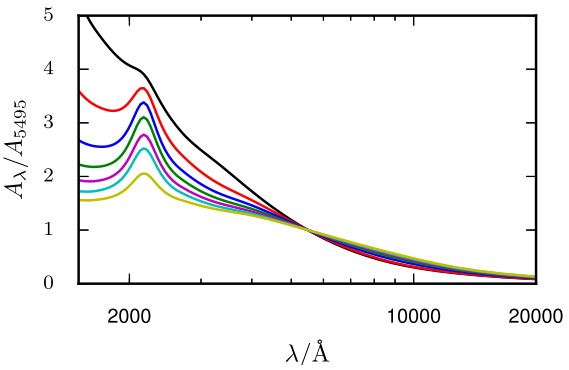

Figure 1. Fitzpatrick (2004) extinction laws for $R_V = 2.1$ (black), 2.6 (red), 3.1 (blue), 3.6 (green), 4.1 (magenta), 4.6 (cyan) and 5.1 (yellow). All have been normalized to 5495 Å.

online. Then in Section 3 we present a method for calculating the marginal likelihood[1] $\mathrm{p}(\boldsymbol{y}|s, l, b, A, R, \alpha, \beta)$ and constructing compact, accurate fits to its dependence on (s, A). Section 4 sums up.

2 PARAMETRIZING EXTINCTION

We start by defining extinction and its relationship to the column of dust between us and a star. Much of what we discuss in this section has previously appeared by various authors, including Golay (1974), McCall (2004), Sale et al. (2009), Stead & Hoare (2009), Bailer-Jones (2011) and Casagrande & VandenBerg (2014). None the less, we repeat it here for completeness and clarity.

Historically, extinction has usually been estimated by looking at the broad-band colours of stars. If one has a star of known spectral type, then by comparing, say, the measured $B - V$ colour of the star to its expected intrinsic colour, one obtains the colour excess

$$E(B-V) \equiv (B-V)_{\mathrm{measured}} - (B-V)_{\mathrm{intrinsic}}, \tag{1}$$

which is a direct estimate of the difference $A_B - A_V$ between the B- and V-band extinctions to the star. Typically (e.g. Cardelli, Clayton & Mathis 1989; Fitzpatrick 2004) the shape of the extinction law at optical and near-infrared wavelengths (see Fig. 1) is assumed to depend on a single parameter R_V, defined through

$$R_V \equiv \frac{A_V}{A_B - A_V} = \frac{A_V}{E(B-V)}. \tag{2}$$

Estimates of A_V, A_B then follow directly from $E(B - V)$ given an assumed R_V. The procedure for other bands (X, Y) is similar: measure a colour excess $E(X - Y)$, then use an assumed extinction law to obtain the broad-band extinctions A_X and A_Y.

Such broad-band extinctions are less than ideal for mapping dust, however. To see this, recall that the extinction in a band X to a distance s along a single line of sight[2] is given by

$$A_X(s) = -2.5 \log_{10} \left(\frac{\int_0^\infty \mathrm{d}\lambda F(\lambda) T_X(\lambda) \mathrm{e}^{-\int_0^s \mathrm{d}s' \kappa_\lambda(s')\rho(s')}}{\int_0^\infty \mathrm{d}\lambda F(\lambda) T_X(\lambda)} \right), \tag{3}$$

where $\rho(s)$ is the density of dust along the line of sight, $\kappa_\lambda(s)$ is its wavelength-dependent opacity, $F(\lambda)$ is the spectral energy distribution (SED) of the observed star and $T_X(\lambda)$ the combination of the transmission of the filter X, the transmission of the atmosphere, the transmission of the rest of the telescope, the detector efficiency and a function that characterizes how the detector responds to incident flux.[3] It is obvious from this equation that passband-based measurements of extinction and reddening, such as A_V and $E(B - V)$, depend not only on the dust column between us and a star (i.e. ρ and κ), but also on the star's SED (see also e.g. McCall 2004; Sale et al. 2009; Bailer-Jones 2011). Consequently it is possible to observe two stars of different spectral types behind the exact same dust column and obtain different measurements of e.g. A_V from each. It is perhaps less immediately obvious from equation (3) that the relationship between column density and the broad-band extinction A_X is not linear, even when the opacity κ_λ is independent of position. Therefore, although quantities such as $E(B - V)$ and A_V are conveniently close to observation, they mix together the effects of the dust column and stars' SEDs in a way that is non-trivial to disentangle: by considering passband-based measurements of extinction one obscures the true physics of the ISM behind a layer of obfuscating variables.

An alternative to passband-based measurements of extinction is to consider monochromatic measurements (e.g. McCall 2004; Sale et al. 2009; Bailer-Jones 2011). The monochromatic extinction at wavelength λ is given by

$$A_\lambda(s) = -2.5 \log_{10} \left(\mathrm{e}^{-\int_0^s \mathrm{d}s \kappa_\lambda(s')\rho(s')} \right) \tag{4}$$

$$= 1.086 \int_0^s \mathrm{d}s' \kappa(s', \lambda)\rho(s'), \tag{5}$$

which follows directly from equation (3) on adopting a Dirac delta function for the transmission filter $T_X(\lambda)$. It is immediately apparent that this $A_\lambda(s)$ does not depend on the SED of the observed star and that its derivative $\mathrm{d}A_\lambda/\mathrm{d}s$ is linear in $\kappa_\lambda \rho$. Therefore monochromatic extinction offers a much more direct view on the distribution of dust, mediated only by variations in dust opacity.

It might appear that using monochromatic extinctions would be significantly more complicated than employing band-based measurements. But, if working within a Bayesian framework, or indeed with any methodology that employs a forward model, building in a monochromatic measure of extinction is essentially trivial: one simply requires a model for how variations in monochromatic extinction will alter observed apparent magnitudes. We will develop this model in Section 2.3.

2.1 The wavelength dependence of extinction

The normalized form of the opacity (κ) dependence on wavelength is typically referred to as the reddening or extinction law, the shape of which depends on the dust grain size distribution, with a greater number of larger grains leading to a greyer extinction law

[1] We note that 'marginal likelihood' can refer to a likelihood with some *or* all of the parameters of the model employed marginalized out. However, this terminology is most frequently used in the case where all parameters have been marginalized, in which case the 'marginal likelihood' is sometimes also called the 'evidence' and is employed in model selection applications. Our use of the term is distinct to this case as we only marginalize some parameters.

[2] To keep this and subsequent expressions readable, we suppress the dependence on the line of sight (l, b) in this and subsequent expressions.

[3] For most modern detectors this is a factor proportional to λ since the detector counts incident photons (Bessell 2005).

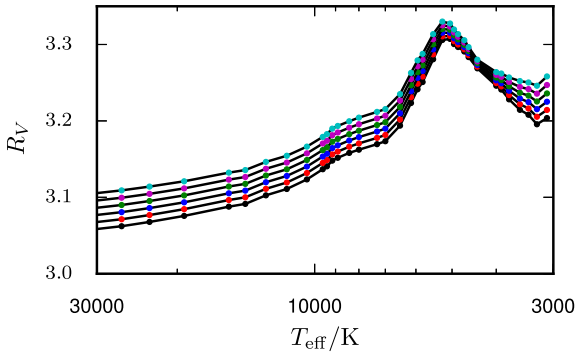

Figure 2. The R_Vs implied by column densities of $R_{5495} = 3.056$ dust as a function of effective temperature along the main sequence defined by Straizys & Kuriliene (1981). Different colours correspond to different extinctions: $A_{4000} = 0$ (black), 2 (red), 4 (blue), 6 (green), 8 (magenta) and 10 (cyan).

(Weingartner & Draine 2001). There exist a number of parametrizations of extinction laws, inferred from a range of sightlines within the Galaxy (e.g. Cardelli et al. 1989; O'Donnell 1994; Fitzpatrick 2004). In Fig. 1 we show a number of those given by Fitzpatrick (2004).

Typically (e.g. Cardelli et al. 1989; Fitzpatrick 2004) the shape of extinction law at optical and near-infrared wavelength is assumed to be a function of the single parameter R_V defined in equation (2) above. This is not ideal, as the broad-band extinctions A_B and A_V that define R_V depend on the SED of the star being observed. Consequently, as shown in Fig. 2 and by McCall (2004), one can place different sources behind the same dust column and still obtain significantly different measurements of R_V. Moreover, as the broad-band A_V and A_B do not depend linearly on dust column density, the inferred R_V depends also on the depth of the dust column in front of the star.

Ideally we would like to have a dust extinction law that depends only on the dust's intrinsic opacity κ_λ. One possibility would be to use the value of R_V that one would measure if a vanishingly small amount of the dust were placed in front of a standard star (McCall 2004), but this is unnecessarily complicated. Instead we follow Maiz Apellániz (2013) and adopt the more straightforward quantity:

$$R_{5495} \equiv \frac{A_{5495}}{A_{4405} - A_{5495}}, \tag{6}$$

where A_{5495} and A_{4405} are the monochromatic extinctions at 5495 and 4405 Å, respectively.[4] This is designed to be similar to R_V, but, as R_{5495} is defined using monochromatic extinctions, it does not depend on the SED of the star observed and will vary along a line of sight only if the grain size distribution and therefore κ varies.

With this choice of monochromatic wavelengths, the values of R_{5495} for the Fitzpatrick (2004) selection of extinction laws are similar to the R_Vs they quote. For example, their '$R_V = 2.1, 3.1, 4.1$' curves give $R_{5495} = 2.097, 3.056, 4.034$, respectively. In Fig. 2 we plot the R_V implied by the $R_{5495} = 3.056$ ('$R_V = 3.1$') extinction law of Fitzpatrick (2004) for a range of SEDs along the main sequence and for various quantities of extinction. As in McCall (2004), it is apparent that there are significant variations in R_V along the main sequence, in addition to smaller variations in response to increasing extinction. We note that this procedure typically gives $R_V = 3.1$ for late B-type stars, a not unexpected result given that Fitzpatrick (2004) used a sample of O, B and A stars to determine their extinction laws.

[4] A_{5495} is sometimes denoted as A_0 (e.g. Bailer-Jones 2011; Sale et al. 2014), following its use in Cardelli et al. (1989).

2.2 Selecting a wavelength for monochromatic extinctions

Now that we have defined our extinction law, we can easily transform monochromatic extinction given at one wavelength to any other wavelength. Therefore, we are free to choose the reference wavelength at which monochromatic extinctions are defined. When monochromatic extinctions have been used in earlier work, the choice of wavelength has generally been made for reasons of convenience. For example, Hanson & Bailer-Jones (2014) and Sale (2012) followed Bailer-Jones (2011) in adopting 'A_0', the monochromatic extinction at 5495 Å, chosen to enable easy use of the Cardelli et al. (1989) extinction laws, which are anchored at this wavelength. In contrast, Sale et al. (2009) used A_{6250}, the monochromatic extinction at 6250 Å. As this wavelength lies near the centre of the INT/WFC Photometric Hα Survey of the Northern Galactic Plane (IPHAS) r band used in that study, the resulting measurements were less affected by variations in R_{5495}.

From equation (5) we have that

$$\frac{{\rm d}A_\lambda}{{\rm d}s}(s) = 1.086\kappa_\lambda(s)\rho(s). \tag{7}$$

If κ did not change along a sightline, it would be trivial to obtain the dust column density from A_λ if R_{5495}, and therefore κ_λ, were known. In reality, however, we expect that the grain size distribution, and consequently R_{5495} and κ, will vary along lines of sight as well as between them. Instead we can look for a wavelength where κ is approximately independent of R_{5495}. Examination of the Draine (2003) models indicates that κ_λ varies only weakly with changing dust grain distribution at around $\lambda = 4000$ Å. Therefore we have that

$$A_{4000}(s) \simeq 1.086\kappa_{4000}\int_0^s \rho(s')\,{\rm d}s', \tag{8}$$

where κ_{4000} is the opacity at 4000 Å that Draine (2003) quotes as $3.8 \times 10^{-3}\,{\rm m^2\,kg^{-1}}$ for his $R_V = 3.1$ grain distribution. So, adopting this $\lambda = 4000$ Å anchor point, we now have a measure of extinction that – to a reasonable approximation – depends on the column density of dust and is independent of variations in opacity. In order to facilitate comparisons to existing results, we note that, if $R_{5495} = 3.056$, then

$$A_V \simeq 0.6929A_{4000} + 0.0018A_{4000}^2, \tag{9}$$

$$R_V \simeq 3.1 \tag{10}$$

for an A0V star.

Before proceeding further we pause to comment that, in the earlier models of Weingartner & Draine (2001), the opacity per unit dust mass is a stronger function of R_{5495} at 4000 Å, but is nearly independent of R_{5495} at approximately 8000 Å. Therefore we advise that it may prove necessary to renormalize to monochromatic extinction to a different wavelength in the future, should improved dust models that contradict those of Draine (2003) become available. Changing this reference wavelength would have no effect on the methods we propose below; the key point is that we choose a wavelength that minimizes opacity variations, the actual wavelength itself is not directly important.

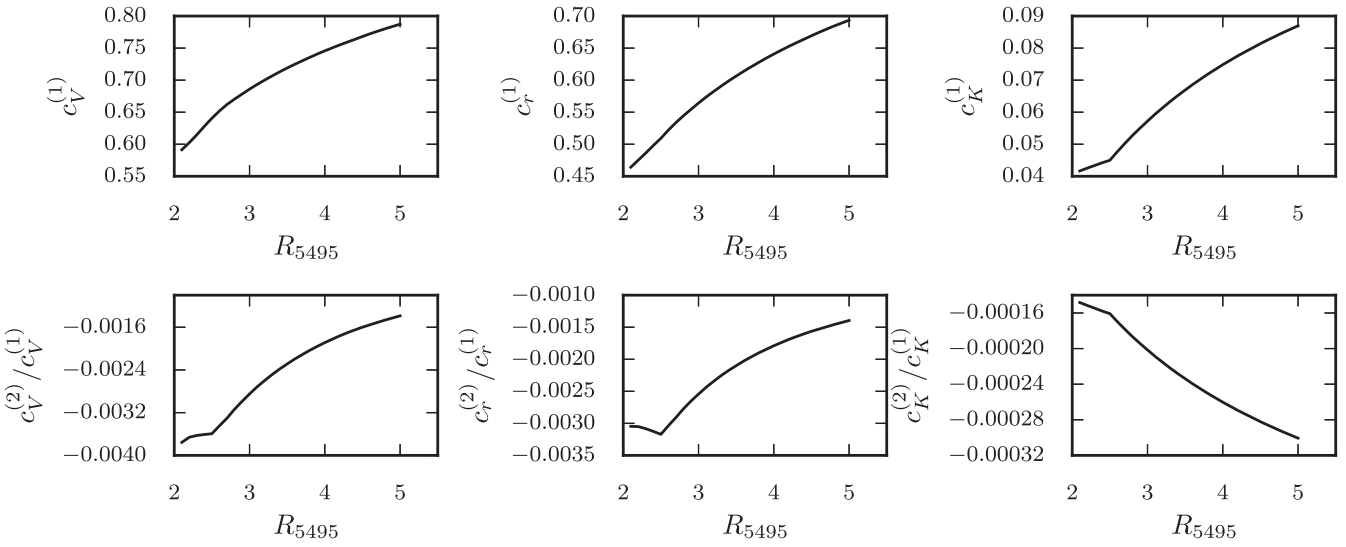

Figure 3. Examples of the dependence of $c_X^{(1)}$ and $c_X^{(2)}$ (equation 11) on R_{5495} for an approximate A0V star with $T_{\rm eff} = 9600$ K and $\log g = 4.07$ and solar metallicity. The values in the left-hand column are for the Johnson–Cousins V band, as defined by Bessell (1990). The centre column is for the IPHAS r band and the right-hand column for the UKIDSS K band as defined by Hewett et al. (2006).

2.3 Variation of broad-band extinction with dust column and stellar type

We now examine how passband extinctions A_X depend on the extinction law (i.e. R_{5495}) and stellar type. We use a method similar to that employed in Sale et al. (2009). For stars cooler than 12 000 K we adopt the closest match from the Phoenix library of synthetic spectra (Husser et al. 2013) to the star's SED. The Phoenix library does not cover hotter stars, so, for $T_{\rm eff} > 12\,000$ K we instead draw from the Munari et al. (2005) library. As the Munari et al. (2005) library does not cover wavelengths redder than 10 500 Å, we do not provide results for the combination of hot stars and passbands with measured transmission beyond this wavelength. Then we use equation (3) to calculate A_X, with κ_λ given by the appropriate Fitzpatrick (2004) extinction law for the assumed R_{5495}.

Clearly one does not want to repeat this procedure every time one seeks to determine the model colours and apparent magnitude of some star. Casagrande & VandenBerg (2014) have suggested precomputing the implied passband extinctions for all combinations of model stars and extinction laws for a range of A_{4000}. However, storing all these extinctions for a reasonably dense isochrone library and for a decent range of R_{5495} and A_{4000} would be impractical. Our approach is instead to use quadratic relations,

$$A_X = c_X^{(1)} A_{4000} + c_X^{(2)} A_{4000}^2, \tag{11}$$

to fit the dependence of A_X on A_{4000} in the range $0 \leq A_{4000} < 10$ for a grid of stellar parameters ($T_{\rm eff}$, $\log g$, [Fe/H]) and extinction laws R_{5495}.

As one might expect, the correction coefficients $c_X^{(1)}$ and $c_X^{(2)}$ depend sensitively on the parameter R_{5495} that sets the shape of the extinction law (see e.g. Fig. 3). Varying the effective temperature of the star is also significant, a fact that is often overlooked (e.g. Fig. 4). The extent and form of the temperature dependence varies from band to band, and is governed by the position of the band relative to the peak of the star's spectrum.

Varying $\log g$, [Fe/H] or [α/Fe] affects A_X much less than variations in $T_{\rm eff}$. For example, Fig. 5 plots the dependence of $c_X^{(1)}$ and $c_X^{(2)}$ on $\log g$, showing that variations in the latter produce changes that are more than an order of magnitude smaller than those caused by varying $T_{\rm eff}$. The response to changes in [Fe/H] or [α/Fe] is significantly smaller still; only in u or similar bands, where variations in [Fe/H] most strongly affect spectra, can the chemical composition of the star measurably affect the effect of reddening and then typically only at a level comparable to $\log g$.

Therefore, in general we approximate $c_X^{(1)}$ and $c_X^{(2)}$ as functions of R_{5495} and $T_{\rm eff}$ only, neglecting the dependence on $\log g$, [Fe/H] and [α/Fe]. In Appendix A we provide tables of p and q for a variety of different popular photometric systems for a range of $T_{\rm eff}$ and R_{5495}. We encourage readers to adopt this calibration when considering extinction for stars within our Galaxy, since it accounts for SED and extinction law variation as well as the non-linear response of extinction in any given photometric band due to increasing dust column.

In the absence of a proper calibration, such as that in Appendix A, many have turned to table 6 of Schlegel et al. (1998) to convert between extinctions in different bands. We warn that this table of relative extinctions was not intended to be used for stars within the Galaxy, but rather to deredden photometry of galaxies behind the relatively sparse dust columns that characterize high Galactic latitudes. As such their table was calibrated using an elliptical galaxy as a source in the limit of small extinctions. Therefore, the calculated ratios are not suitable for detailed use when considering stars subject to significant extinction.

In Fig. 6 we show the fractional errors that arise due to assuming parametrizations of extinction given here and those that result from assuming the $A_X/E(B-V)$ ratios given by Schlegel et al. (1998) and Schlafly & Finkbeiner (2011). The errors due to the calibration we propose here are extremely small, generally less than 0.1 per cent. In contrast, the error that arises by following the Schlegel et al. (1998) or Schlafly & Finkbeiner (2011) ratios is frequently on the order of 5 per cent.

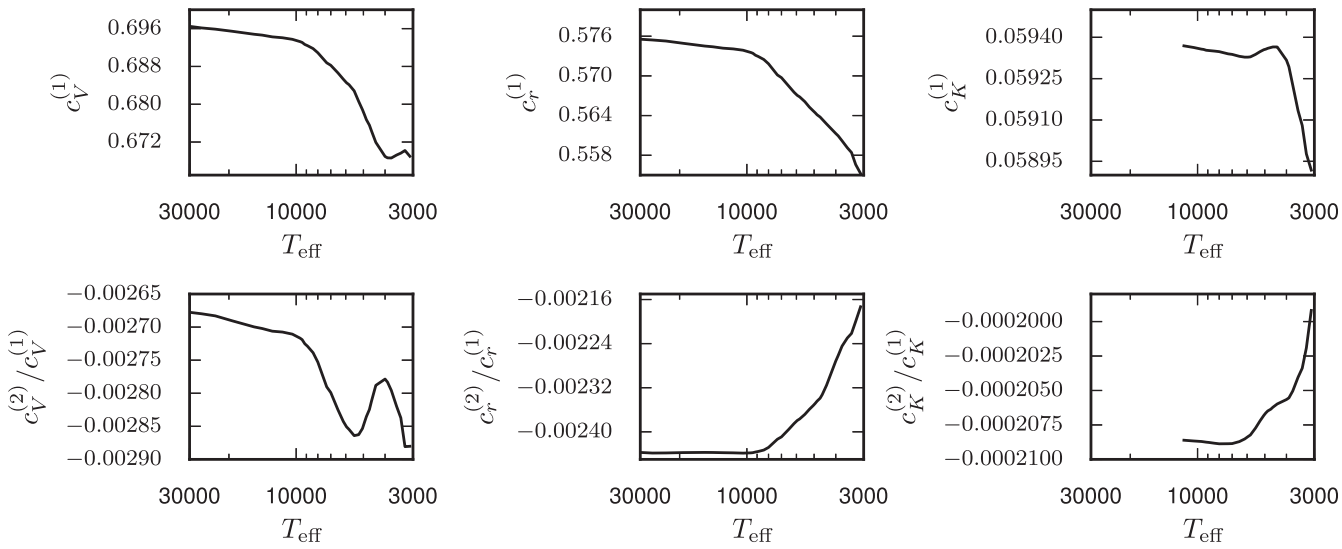

Figure 4. Examples of the dependence of $c_X^{(1)}$ and $c_X^{(2)}$ on $T_{\rm eff}$ along the main sequence, as defined in $(T_{\rm eff}, \log g)$ by Straizys & Kuriliene (1981) for stars having solar metallicity. The columns are the same as Fig. 3.

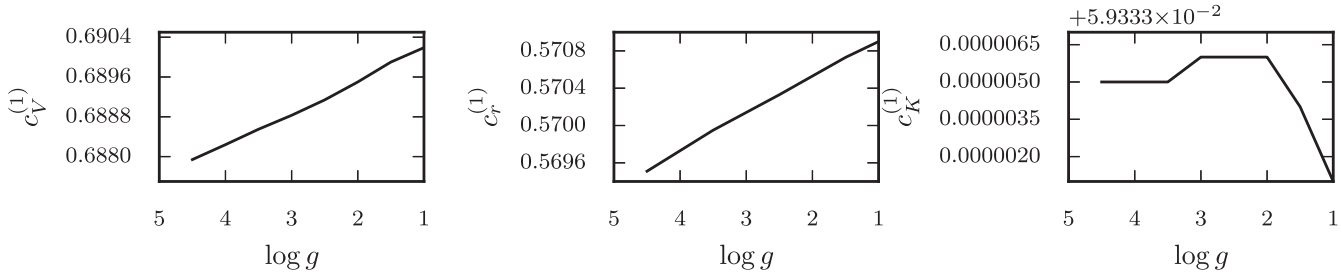

Figure 5. Examples of the dependence of $c_X^{(1)}$ on $\log g$. We fix $T_{\rm eff} = 7000$ K and metallicity to solar. The columns are the same as Fig. 3.

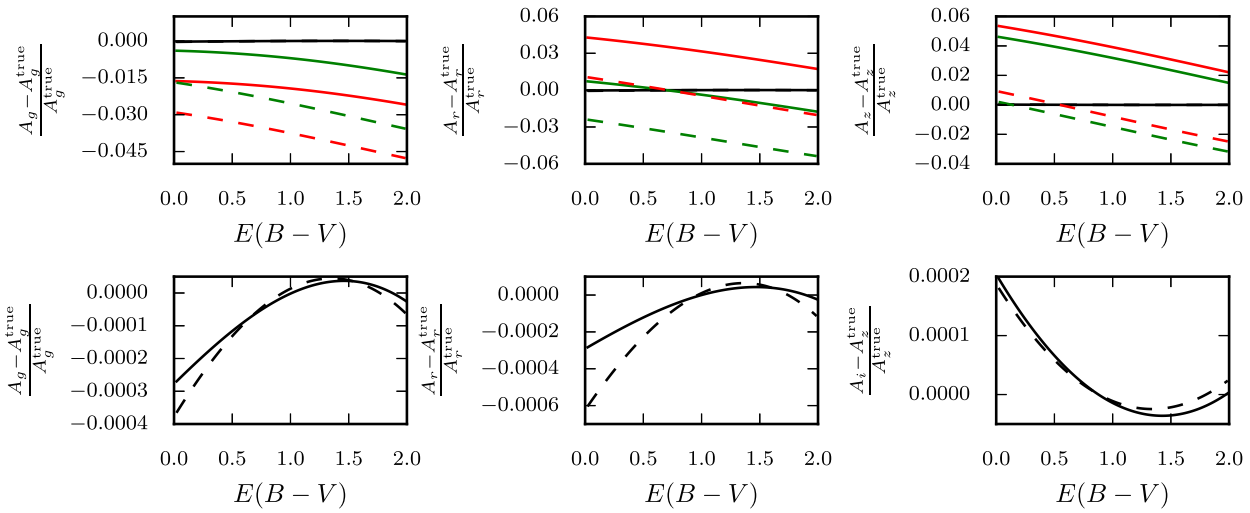

Figure 6. Comparison of responses to extinction in different SDSS bands using stars of type A0V (solid curves) and M3V (dashed) as examples. Upper row: fractional errors in A_g (left), A_r (middle) and A_z (right) obtained assuming the $A_X/E(B-V)$ ratios given by Schlegel, Finkbeiner & Davis (1998) (red), using $A_X/E(B-V)$ ratios from Schlafly & Finkbeiner (2011) (green) and by using our fit (equation 11) (black). The solid (dashed) curves plot results for A0V (M3V) stars. Lower row: zoomed-in views of the fractional errors produced by our fits (equation 11).

3 MARGINAL LIKELIHOOD OF DISTANCE AND EXTINCTION

Having discussed how best to model the effects of extinction, we now turn to the problem of estimating distances and extinctions to individual stars in situations where we are not interested in the details of each star's spectral type. This problem occurs when constructing 3D extinction maps from stellar catalogues (e.g. Sale & Magorrian 2014).

It is convenient to replace the distance s by the distance modulus,

$$\mu \equiv 5 \log_{10}(s/10\,\mathrm{pc}), \tag{12}$$

and the extinction A_{4000} by its logarithm,

$$a_{4000} \equiv \ln A_{4000}. \tag{13}$$

As $s > 0$ and we assume that $A_{4000} > 0$, it is sensible to consider the logarithms of both values, since both μ and a_{4000} span the entire real line. In addition, the use of the distance modulus is sensible, as uncertainties on it are often approximately Gaussian. Meanwhile, the use of a_{4000} is largely motivated by the fact that in Sale & Magorrian (2014) we place a Gaussian random field prior on it to create an extinction map.

The particular problem we address is the following. Given a set of observations $\boldsymbol{y}$ of some star, we would like to compute the marginal likelihood,

$$\mathrm{p}(\boldsymbol{y}|\mu, a_{4000}, R_{5495}, \alpha, \beta) = \int [\mathrm{p}(\boldsymbol{y}|\mu, a_{4000}, R_{5495}, \boldsymbol{x}, \alpha) \times \mathrm{p}(\boldsymbol{x}|\mu, \beta)\,\mathrm{d}\boldsymbol{x}], \tag{14}$$

of the distance modulus μ and log extinction a_{4000} to the star by marginalizing the star's intrinsic parameters $\boldsymbol{x}$, which include its mass, age, metallicity and so on. We assume a mix of stellar populations β, which specifies the (possibly position dependent) prior $\mathrm{p}(\boldsymbol{x}|\mu, \beta)$ on $\boldsymbol{x}$.[5] When the observations $\boldsymbol{y} = \boldsymbol{y}_{\mathrm{phot}}$ are limited to the star's photometric apparent magnitudes the likelihood $\mathrm{p}(\boldsymbol{y}_{\mathrm{phot}}|\mu, a_{4000}, R_{5495}, \boldsymbol{x}, \alpha)$ is straightforward to calculate using the extinction model α described in Section 2. If one has independent additional data $\boldsymbol{y}_{\mathrm{other}}$, such as a spectroscopic metallicity or a trigonometric parallax, then the likelihood becomes

$$\mathrm{p}(\boldsymbol{y}|\mu, a_{4000}, R_{5495}, \alpha, \beta) = \mathrm{p}(\boldsymbol{y}_{\mathrm{phot}}|\mu, a_{4000}, R_{5495}, \alpha, \beta)\mathrm{p}(\boldsymbol{y}_{\mathrm{other}}|\mu, \beta), \tag{15}$$

in which $\mathrm{p}(\boldsymbol{y}_{\mathrm{other}}|\mu, \beta)$ is (usually) independent of extinction and is relatively easy to treat. In the following we ignore $\mathrm{p}(\boldsymbol{y}_{\mathrm{other}}|\mu, \beta)$ and assume that $\boldsymbol{y} = \boldsymbol{y}_{\mathrm{phot}}$ only.

The dependence of the marginalized likelihood $\mathrm{p}(\boldsymbol{y}_{\mathrm{phot}}|\mu, a_{4000}, R_{5495}, \alpha, \beta)$ on $(\mu, a_{4000}, R_{5495})$ is usually difficult to predict directly from the observed $\boldsymbol{y}_{\mathrm{phot}}$. We might reasonably expect that it will typically have two maxima – one that corresponds to the star being on the main sequence, the other to the giant branch – but the locations $(\mu, a_{4000}, R_{5495})$ and extent of these maxima cannot be found without some exploration of the $(\mu, a_{4000}, R_{5495})$ space. We use a Markov chain Monte Carlo (MCMC) algorithm to carry out this exploration, and then fit a simple mixture model to the $(\mu, a_{4000}, R_{5495})$ dependence of the likelihood function.

[5] Recall that $\mathrm{p}(\boldsymbol{x}|\mu, \beta)$ is actually $\mathrm{p}(\boldsymbol{x}|\mu, l, b, \beta)$, as we have suppressed the dependence on the line of sight (l, b).

The marginal likelihood depends on a calibration α that includes the extinction calibration discussed in Section 2 in addition to a set of isochrones rendered in the appropriate filter system. In this paper we employ Padova isochrones (Bressan et al. 2012) that use bolometric corrections calculated from ATLAS9 (Castelli & Kurucz 2003) model spectra. To investigate the extent of the systematic error stemming from the use of *ATLAS9* derived bolometric corrections, we have repeated the tests of Section 3.2 but with bolometric corrections derived from Phoenix model spectra. We found that the posterior expectations of distance modulus and extinction typically change by $\lesssim$0.02. We caution that, although this uncertainty may appear small, as it is systematic it will potentially have a measurable impact on extinction maps.

3.1 Sampling the likelihood function

Although the marginalized likelihood (equation 14) is a function of $(\mu, a_{4000}, R_{5495})$, it is not a probability density in these parameters and therefore cannot directly be sampled using MCMC methods. So, we begin by using Bayes' theorem to express it as

$$\mathrm{p}(\boldsymbol{y}|\mu, a_{4000}, R_{5495}, \alpha, \beta) = \mathrm{p}(\boldsymbol{y}|\alpha, \beta)\frac{\int \mathrm{p}(\mu, a_{4000}, R_{5495}, \boldsymbol{x}|\boldsymbol{y}, \alpha, \beta)\,\mathrm{d}\boldsymbol{x}}{\mathrm{p}(\mu, a_{4000}, R_{5495}|\beta)}, \tag{16}$$

in which the prior on $(\mu, a_{4000}, R_{5495})$ that appears in the denominator is given by

$$\mathrm{p}(\mu, a_{4000}, R_{5495}|\beta) = \int \mathrm{p}(\mu, a_{4000}, R_{5495}, \boldsymbol{x}|\beta)\,\mathrm{d}\boldsymbol{x}, \tag{17}$$

which is completely determined by our choice of Galaxy model β. Now the posterior

$$\mathrm{p}(\mu, a_{4000}, R_{5495}, \boldsymbol{x}|\boldsymbol{y}, \alpha, \beta) = \frac{\mathrm{p}(\boldsymbol{y}|\mu, a_{4000}, R_{5495}, \boldsymbol{x}, \alpha, \beta)\mathrm{p}(\mu, a_{4000}, R_{5495}, \boldsymbol{x}|\beta)}{\mathrm{p}(\boldsymbol{y}|\alpha, \beta)} \tag{18}$$

that appears within the integral in the numerator of equation (16) can be sampled using any convenient MCMC method. The likelihood $\mathrm{p}(\boldsymbol{y}|\mu, a_{4000}, R_{5495}, \boldsymbol{x}, \alpha, \beta)$ is easy to calculate and the normalizing factor $\mathrm{p}(\boldsymbol{y}|\alpha, \beta)$ is important only if we want to compare models with different population mixes β or dust properties α.

3.1.1 The prior p(s, A4000, R5495, x|β)

In the examples that follow we adopt a prior $\mathrm{p}(\boldsymbol{x}|\mu, \beta)$ on the stellar parameters that comes from an intentionally simple Galactic model β. We include in β a Salpeter initial mass function (IMF) and a constant star formation history. We model metallicity variations as being Gaussian with a standard deviation of 0.2 dex and a mean that declines by 0.06 dex kpc^{-1} with Galactocentric radius, following Luck & Lambert (2011), normalized to solar metallicity at the solar circle. We could also impose a vertical metallicity gradient, but opt not to do so here since the stars we use as examples later in this section lie very close to the Galactic mid-plane.

In the light of equation (16), we are free to use any convenient prior on the parameters μ, a_{4000} and R_{5495} provided only that it is sufficiently broad to cover plausible regions of parameter space. We adopt a flat prior in all three parameters, but with R_{5495} limited to the range $2.1 \lesssim R_{5495} \lesssim 5.5$ over which the Fitzpatrick (2004)

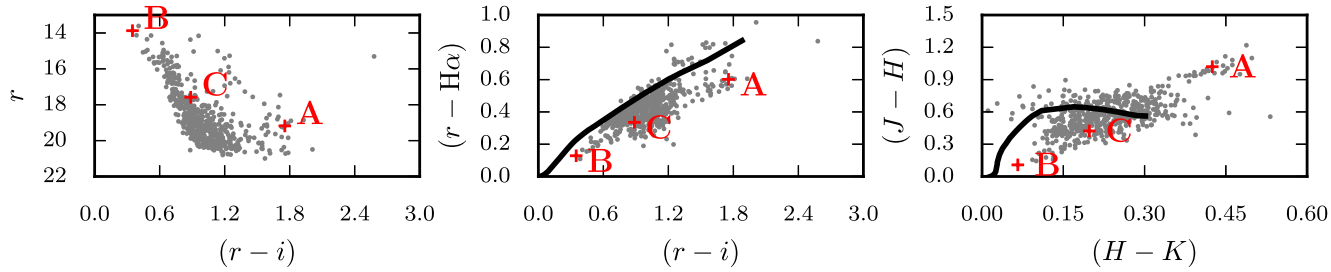

Figure 7. Colour–colour diagrams in the IPHAS and UKIDSS-GPS systems of our sample catalogue. The three stars studied in detail are marked with red crosses, labelled with their corresponding letter, where A is IPHAS2 J211210.70+482106.8, B is IPHAS2 J211225.40+481927.6 and C is IPHAS2 J211223.10+481656.4. Solid lines show unreddened main sequences. Only stars that appear in all six bands and are flagged as stellar in both surveys are shown.

reddening laws are available. When written out, the prior in equation (18) is

$$\begin{aligned} \mathrm{p}(\mu, a_{4000}, R_{5495}, \boldsymbol{x}|\beta) &\propto \mathrm{p}(\boldsymbol{x}|\mu, \beta)\mathrm{p}(\mu, a_{4000}, R_{5495}), \\ \mathrm{p}(\boldsymbol{x}|\mu, \beta) &\propto \mathcal{M}^{-2.35}\mathcal{N}([\mathrm{M/H}]|0.06(\mathcal{R}_\odot - \mathcal{R}), 0.2), \\ \mathrm{p}(\mu, a_{4000}, R_{5495}) &\propto \begin{cases} 1, & \text{if } 2.097 \leq R_{5495} < 5.402, \\ 0, & \text{otherwise}, \end{cases} \end{aligned} \tag{19}$$

where $\mathcal{M}$ is the initial mass of the star, [M/H] its metallicity, $\mathcal{R}$ the Galactocentric radius of the star implied by the distance modulus μ and Galactic coordinates (l, b), $\mathcal{R}_\odot$ the Galactocentric radius of the Sun and $\mathcal{N}(\cdot|\cdot, \cdot)$ denotes a normal distribution.

3.1.2 *The choice of MCMC scheme*

There are many MCMC algorithms that could be used to sample the probability density function (pdf; equation 18). We use the affine invariant ensemble sampler of Goodman & Weare (2010) as implemented in EMCEE (Foreman-Mackey et al. 2013). This algorithm employs a collection of 'walkers' that explore the parameter space. At each iteration each walker attempts to move some distance along the vector towards another randomly chosen walker. We initialize the sampler with an array of 100 walkers positioned along the main sequence and red giant branch in $(T_{\rm eff}, \log g)$ space, with [M/H] drawn from the prior distribution, R_{5495} drawn from a uniform distribution on the full range of Fitzpatrick (2004) reddening laws and μ and a_{4000} chosen to be the maximum likelihood values given all the other parameters. However, when running this algorithm it became clear that groups of walkers would occasionally become stuck in islands of low probability, with the relatively high dimensionality making it difficult for them to transition out to higher probability regions. We therefore adopt a similar approach to Hou et al. (2012) and 'prune' the set of walkers at the end of burn-in, moving some walkers when a disproportionally large number are stuck in islands of low probability.

Our general schema then consists of using 100 walkers, with a burn-in of 1000 iterations, of which the last 100, thinned by a factor of 10, are used to facilitate the pruning. After the pruning, we then iterate for a further 9000 iterations to obtain our final MCMC chain, thinning the chain by a factor of 10. The thinned chain typically has an autocorrelation length of around 1, implying an autocorrelation length of roughly 10 for the unthinned chain, with a total sample size of 90 000 and an effective sample size of roughly 45 000. Although the moments of the distribution could be found to sufficient precision with a much smaller effective sample size, capturing some of the detail in the posterior requires such a large sample. We can then perform the integration in equation (16) by simply ignoring all parameters other than μ, a_{4000} and R_{5495} in the MCMC chain. Dividing the result by our (trivial) prior gives the desired marginal likelihood $\mathrm{p}(\boldsymbol{y}|\mu, a_{4000}, R_{5495}, \alpha, \beta)$.

We demonstrate how we obtain the marginal likelihoods using photometry from IPHAS (Drew et al. 2005; Barentsen et al. 2014) and UKIRT Infrared Deep Sky Survey-Galactic Plane Survey (UKIDSS-GPS; Lucas et al. 2008). In particular we use a cross-matched catalogue that covers $5 \times 5\,\mathrm{arcmin}^2$ centred on $(l, b) = (90.04, -0.04)$. We use a 1 arcsec matching radius and only stars flagged as stellar in both surveys are included. We show in Fig. 7 colour–colour plots of this catalogue. From this catalogue we select three stars to concentrate on IPHAS2 J211210.70+482106.8, IPHAS2 J211225.40+481927.6 and IPHAS2 J211223.10+481656.4, which we label as stars A, B and C, respectively. These span a range of colours and apparent magnitudes. We add, in quadrature, to the stated photometric uncertainties an additional factor of 2 per cent to account for systematic uncertainties, such as those on the photometric zero-points. This additional factor dominates the uncertainty budget for stars B and C and makes an important contribution for star A.

2D histograms of the marginal likelihoods obtained for the three stars are displayed in Fig. 8. It is apparent that, as in Green et al. (2014), some exhibit complicated shapes, largely due to the irregular shape of the stellar locus in colour–magnitude space. In particular, star A could be either a main-sequence star or on the red giant branch: from its photometry alone we are unable to make a distinction. On the other hand, qualitative examination of the colour–magnitude diagram in Fig. 7 indicates that the star should be on the giant branch, due to its position in a redder sequence (Sale et al. 2009). However, this qualitative analysis has been implicitly conditioned upon the photometry of all the stars in the catalogue – we would not have been able to identify a red sequence if we only had the photometry of star A. In contrast, the likelihood in Fig. 8 is conditioned upon only the photometry of star A. In order to condition it on the entire photometric catalogue we require a method such as that of Sale & Magorrian (2014) (see in particular their equation 19), in which case the construction of the extinction map would break the degeneracy between the main sequence and red giant branch. Both stars B and C appear too hot to be on the red giant branch.

Although we have assumed a flat prior on R_{5495}, the combination of optical and near-infrared photometry has enabled us to narrow the range of possible extinction laws (see also Berry et al. 2012). If our data had not constrained R_{5495} our uncertainties on both μ and a_{4000}

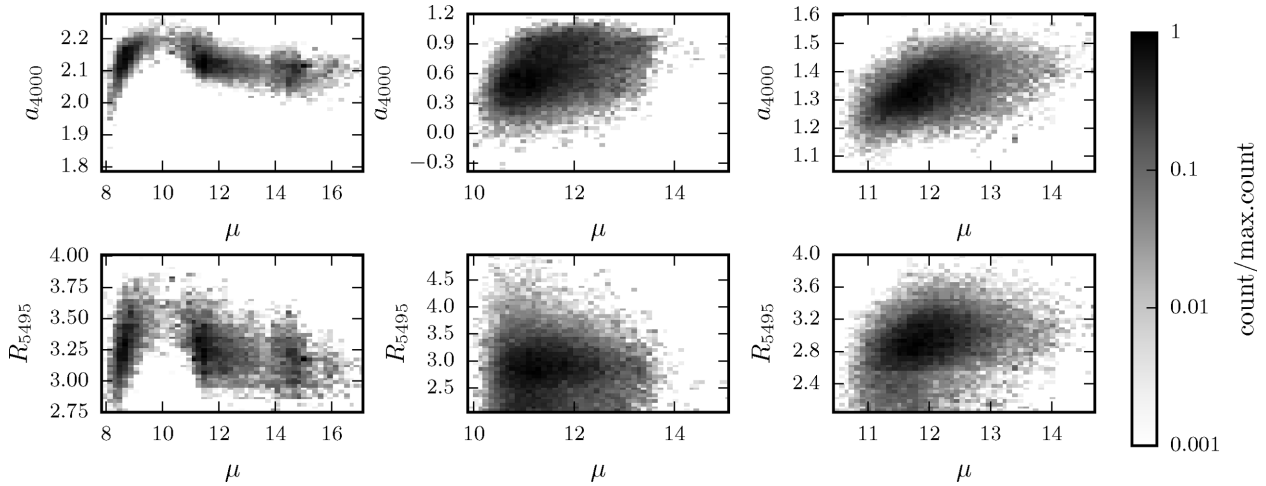

Figure 8. Example marginal likelihoods shown for stars A (left), B (centre) and C (right). We have binned the MCMC samples only for the purposes of producing histograms.

would have been increased, since R_{5495} is covariant with both. We note that the uncertainties on R_{5495} depend, to a large degree, on the number and wavelength range of the photometric bands employed: if, as in Berry et al. (2012), we had used Sloan Digital Sky Survey (SDSS) data in place of IPHAS we would have eight bands instead of six and the bluer coverage of the *u* and *g* bands and so should be able to achieve more precise estimates of R_{5495}.

3.2 Fitting a mixture model to the marginalized likelihood function

Having explored the $(\mu, a_{4000}, R_{5495}, \boldsymbol{x})$ posterior, we carry out the marginalization of $\boldsymbol{x}$ in the numerator of equation (16) by simply ignoring the $\boldsymbol{x}$ values returned by the sampler and focusing only on the distribution of the $(\mu, a_{4000}, R_{5495})$ samples. These samples are drawn from the marginal likelihood function (14) weighted by the prior (17) of our assumed Galaxy model β.

Although it would be possible to use the full set of samples from the MCMC chain (reweighted to account for the prior) as our description of the marginalized likelihood, this is far from ideal. The chains are long. Therefore the cost of storing them is high, all the more so when one considers that, when constructing coherent maps of extinction or stellar density, one will typically want to use marginalized likelihoods for many stars simultaneously. The data volumes can be reduced by thinning the MCMC chain (i.e. by removing all but every *n*th entry). Drastic thinning would enable the data volumes to be manageable, even with a large catalogue of stars. The cost associated with thinning, however, is that it reduces the ability of the chain to represent the true underlying marginalized likelihood function, particularly in the relatively low likelihood regions. This can be a key problem if the likelihood is then fed into a hierarchical model, as in Sale & Magorrian (2014), where data from other stars suggest that the distance or extinction to this star may lie in such a low probability region. For example, if the range of possible extinctions to a particular star were constrained by other nearby stars to a region that is only sampled by a single point in the thinned MCMC chain, the resultant marginal posterior distribution of extinction to this star will take the form of a delta function and the uncertainty would therefore be drastically underestimated.

At the other extreme, the most common solution to this problem is to simply report the mean and covariance matrix of the likelihood function. But doing this does not pass on any detailed information about the shape of the likelihood function, which, as demonstrated by Fig. 8, may well be somewhat irregular. In particular it will not reveal multimodality, as might be the case if there are two peaks in the likelihood corresponding to the observed star being on the main sequence or on the giant branch.

An alternative is to describe the likelihood function using some mixture of simple distributions. For example, Carrasco Kind & Brunner (2014) depict the posterior distributions of photometric redshifts to galaxies using a mixture of Gaussians and Voigt profiles. We instead fit a mixture of trivariate Gaussians to the marginalized posterior $\mathrm{p}(\mu, a_{4000}, R_{5495}|\boldsymbol{y}, \alpha, \beta)$. As we assume a flat prior on $(\mu, a_{4000}, R_{5495})$, this is equivalent to fitting Gaussians to the marginalized likelihood function $\mathrm{p}(\boldsymbol{y}|\mu, a_{4000}, R_{5495}, \alpha, \beta)$. So, writing $\boldsymbol{\theta} \equiv (\mu, a_{4000}, R_{5495})$, our goal is to fit a function,

$$\mathrm{p}(\boldsymbol{\theta}|\boldsymbol{y}, \beta, \alpha) \approx \sum_{k=1}^{K} w_k \mathcal{N}(\boldsymbol{\theta}|\boldsymbol{m}_k, \mathbf{C}_k), \tag{20}$$

to our MCMC sample $(\boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_N)$ by adjusting the weights w_k, means $\boldsymbol{m}_k$ and covariances $\mathbf{C}_k$ of the Gaussians on the right-hand side, along with their number *K*.

Before explaining our procedure for carrying out the fitting, we note that using Gaussians here has the key advantage that one can often carry out further marginalization analytically. An example of this is given in section 4.2 of Sale & Magorrian (2014), in which the distances and extinctions to individual stars were marginalized in order to obtain the pdf of the parameters describing the large-scale extinction distribution. Similarly, having fit the trivariate Gaussian mixture model above, one could later decide to take the prior $\mathrm{p}(R_{5495}|\beta)$ to be Gaussian with a mean and standard deviation from e.g. Fitzpatrick & Massa (2007) and then marginalize R_{5495} analytically to obtain the marginal likelihood $\mathrm{p}(\boldsymbol{y}|\mu, a_{4000}, \alpha, \beta)$. This new, 2D marginal likelihood would still be expressed as a sum of Gaussians.

3.2.1 *Fitting a Gaussian mixture model with K components*

One way of addressing the problem of fitting the Gaussian mixture model (20) to the MCMC sample would be by modelling the latter as a Dirichlet process mixture of Gaussians. Our goal here though is not to consider all possible Gaussian mixture descriptions of the MCMC chain, but instead to obtain a *single*, compact, 'best' description of

the marginalized likelihood. Generally a single Gaussian will not describe the marginalized posterior well, but a mixture of two or more Gaussians will do better.

A simple and robust approach is to iterate of a range of possible K. For each K we use the expectation–maximization (EM) algorithm, as implemented in SCIKIT-LEARN (Pedregosa et al. 2011), to find the parameters $(w_k, \boldsymbol{m}_k, \mathbf{C}_k)$ of each of the K Gaussians that maximize the likelihood

$$\mathcal{L}_K \equiv \prod_{n=1}^{N} \sum_{k=1}^{K} w_k \mathcal{N}(\boldsymbol{\theta}_n | \boldsymbol{m}_k, \mathbf{C}_k), \tag{21}$$

subject to the constraint that $\sum_k w_k = 1$. The EM algorithm functions by introducing $N \times K$ new latent variables $\{z_{nk}\}$ that allow the awkward product of sums in this likelihood to be rewritten as the easier-to-handle sum of products:

$$\mathcal{L}_K \equiv \sum_{\{z_{nk}\}} \prod_{n=1}^{N} \prod_{k=1}^{K} [w_k \mathcal{N}(\boldsymbol{\theta}_n | \boldsymbol{m}_k, \mathbf{C}_k)]^{z_{nk}}. \tag{22}$$

The new variables z_{nk} indicate the probability that MCMC sample n was drawn from Gaussian k. The algorithm proceeds by alternately updating the latent membership probabilities $\{z_{nk}\}$ holding $\{w_k, \boldsymbol{m}_k, \mathbf{C}_k\}$ fixed, then, for this choice of $\{z_{nk}\}$, finding the $\{w_k, \boldsymbol{m}_k, \mathbf{C}_k\}$ that maximize the likelihood. We initialize the EM run with the means of the components given by the mean of the MCMC sample and with diagonal covariance matrices with the variance for each parameter being the corresponding variance from the MCMC sample. The EM algorithm is then run for 100 iterations to find optimal values of $\{w_k, \boldsymbol{m}_k, \mathbf{C}_k\}$.

3.2.2 *How many components K?*

Having obtained maximum likelihoods for $K = 1, 2, 3, \ldots$ the question then becomes one of deciding how many Gaussians are actually justified. For example, if we chose $K \geq N$ (i.e. there are as many Gaussians as there are MCMC samples), then the likelihood would be unbounded: simply centre one Gaussian on each point from the MCMC sample and let its covariance shrink to zero. We would like to avoid fitting the shot noise in our MCMC samples like this, or, more practically, requiring such a large number of Gaussians that they cause data volume problems.

A natural way of comparing models with different components would be to adopt uninformative priors on $\{w_k, \boldsymbol{m}_k, \mathbf{C}_k\}$ and to marginalize the likelihoods $\mathcal{L}_k$ to obtain the marginal likelihoods $\mathrm{p}(\{\boldsymbol{\theta}\}|K)$ for each K. These $\mathrm{p}(\{\boldsymbol{\theta}\}|K)$ could be estimated by a variational Bayes method (see e.g. appendix C of Magorrian 2014), but doing so would be overkill for our present purposes. As a straightforward alternative, we instead employ the Bayesian information criterion (BIC; Schwarz 1978):

$$\mathrm{BIC} = -2 \ln \hat{\mathcal{L}}_K + (10\,K - 1) \ln N, \tag{23}$$

where $\hat{\mathcal{L}}_K$ is the maximum likelihood of the K-component Gaussian mixture model, as found by the EM algorithm. The second term in this expression acts as a penalty on the number of components, with the $(10K - 1)$ factor accounting for the number of free parameters in a K-component trivariate Gaussian mixture model: $3K$ numbers are needed to specify the means $\boldsymbol{m}_k$, $6K$ for the symmetric covariance matrices $\mathbf{C}_k$ and $K - 1$ for the weights w_k. Our favoured model is simply the one that minimizes the value of BIC. We find that this minimum is typically achieved for mixtures having $K \sim 5$ Gaussians.

Fig. 9 shows our Gaussian mixture approximations to the MCMC-sampled marginal likelihoods of Fig. 8. As it is difficult to compare these 2D projections by eye, in Fig. 10 we also show one-dimensional (1D) projections of both the MCMC chain and our Gaussian mixture fits. The Gaussian mixture model provides a good, compact descriptions of the MCMC samples.

The one area in which we find that the mixture model fails to perform well is when R_{5495} takes on values close to the cut-offs imposed by the range covered by the Fitzpatrick (2004) extinction curves; our marginalized likelihoods fall sharply to zero at these extreme values, a behaviour which the Gaussian mixture has difficulty reproducing. However, we note that the impact of such issues will be dramatically reduced by the imposition of any sensible prior on R_{5495}. For example, one could place a simple Gaussian prior on R_{5495} with a mean and variance taken from e.g. Fitzpatrick & Massa (2007). Under such a prior the probability of the problematic

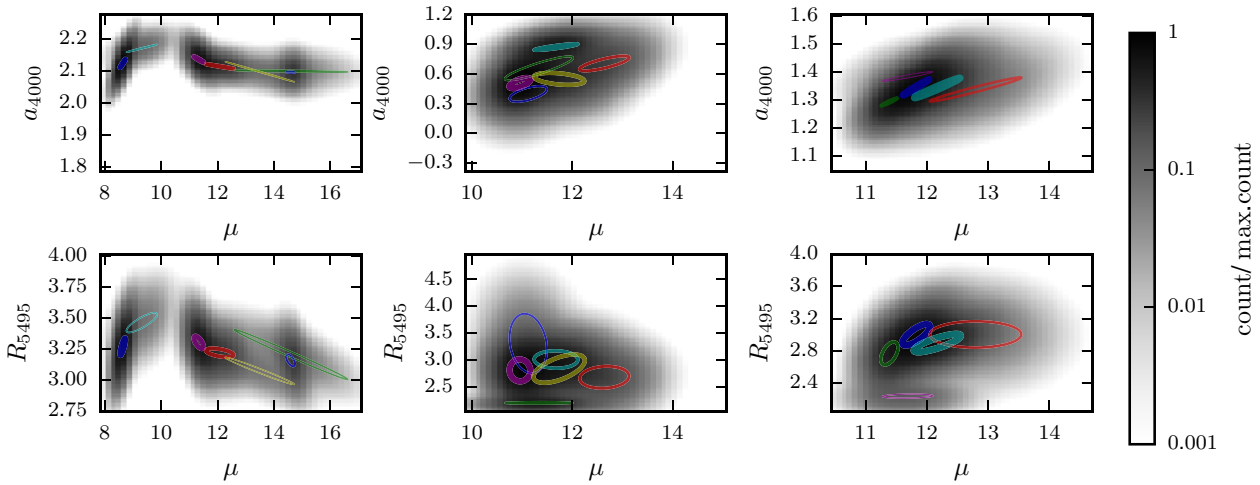

Figure 9. Examples of the Gaussian mixture approximations, plotted in histograms to match Fig. 8. The coloured ellipses show the 2σ contours of each of the Gaussian mixture model components, with the width of the ellipses' curves linearly increasing with the weight of the corresponding component in the Gaussian mixture model. As in Fig. 8 star A is in the left-hand column, B in the centre and C on the left.

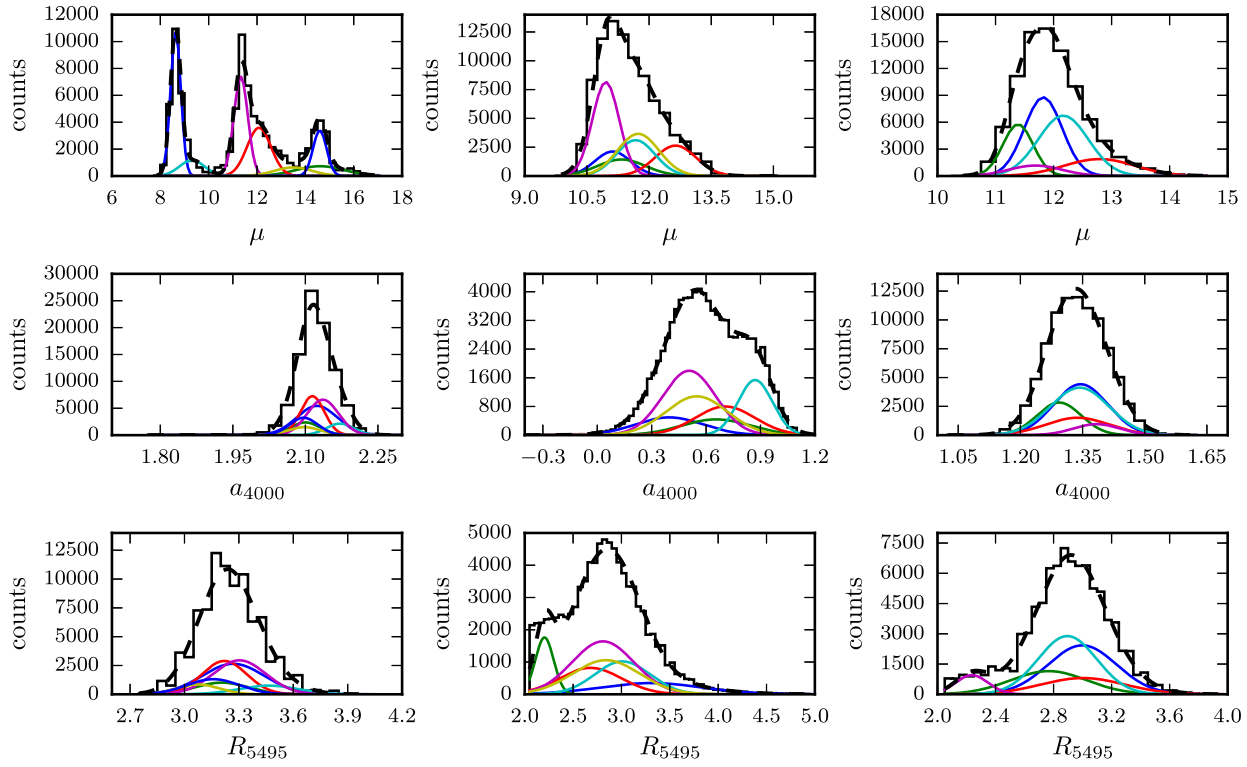

Figure 10. Examples of 1D marginal likelihoods for each star shown fitted with Gaussian mixtures. The black solid line shows a histogram of the MCMC samples. As in Fig. 8 we have only performed the binning of the MCMC samples to produce the plotted histograms. The coloured lines show the contribution of each of the components in the Gaussian mixture model, whilst the dashed black line shows the total 1D marginal likelihood implied by the Gaussian mixture model. Once again star A is in the left-hand column, B in the centre and C on the left.

extreme values of R_{5495} would be very low and so the issues related to the fit would become essentially irrelevant.

3.3 The quality of the Gaussian mixture model approximation

One of our primary goals in this paper is to find a compact description of the marginal likelihood $\mathrm{p}(\boldsymbol{y}|\mu, a_{4000}, R_{5495}, \alpha, \beta)$. It is natural then to ask whether our Gaussian mixture model provides a more compact summary of this function than, say, a thinned sample of points from an MCMC chain. In the following we consider two different measures of how well such fits reproduce the true marginal likelihood.

3.3.1 Kullback–Leibler divergences

One way of quantifying the fidelity of different descriptions of the marginal likelihood is by using the Kullback–Leibler (KL) divergence. Let P be the true marginalized likelihood and Q a fit from either the Gaussian mixture or the thinned MCMC sample. The divergence of the fit Q from the true function P is given by

$$D_{\mathrm{KL}}(P \parallel Q) = \int_{-\infty}^{\infty} \mathrm{d}\boldsymbol{\theta}\; P(\boldsymbol{\theta}) \log\left(\frac{P(\boldsymbol{\theta})}{Q(\boldsymbol{\theta})}\right). \quad (24)$$

This can be recognized as the entropy of P relative to Q, a measure of how much more information there is in P than in the fitted Q.

One problem with applying this is that we do not know the true marginal likelihood function $P(\boldsymbol{\theta})$: we only have discrete samples $(\boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_N)$ of it from the MCMC chain. So, to construct our reference P we take a very long chain of $N \sim 10^5$ samples and then, as a simple kernel density estimator, we replace each sample point $\boldsymbol{\theta}_n$ (which has density $\delta(\boldsymbol{\theta} - \boldsymbol{\theta}_n)$) by a narrow Gaussian kernel centred on $\boldsymbol{\theta}_n$. Then the value of the function P at any point $\boldsymbol{\theta}$ is given by the sum of the contributions from all $N \sim 10^5$ kernels at that point.

We set the kernel width using 10-fold cross-validation. That is, each point from the MCMC chain is assigned at random to one of 10 subsamples. Then, for a given trial kernel width, we construct a kernel density estimate using nine of the 10 subsamples. We use this kernel density estimate to calculate the log-likelihood of the points in the remaining subsample. This is then repeated for all 10 subsamples and the average log-likelihood found. By considering a range of kernel widths we can choose an optimum value by maximizing the mean log-likelihood. Typically the kernel widths found by this procedure are small – on the order of 0.01 in μ, for example – and much smaller than the bin sizes adopted in Fig. 8. Consequently, if one applies a binning to the kernel density estimate of the marginal likelihood to match that employed in Fig. 8, one would obtain a distribution that will closely resemble the histograms in Fig. 8.

We use a similar procedure to reconstruct $Q(\boldsymbol{\theta})$ for the thinned MCMC chains. We do so by thinning the main MCMC chain and reapplying the cross-validation procedure to each thinned chain. Finally, we estimate the integral in equation (24) using Monte Carlo integration with 10 000 samples drawn from $P(\boldsymbol{\theta})$.

We are interested in how the KL divergences from $P(\boldsymbol{\theta})$ of Gaussian mixture fits and of the thinned MCMC chains scale with the number of parameters needed to describe each fit. As discussed in Section 3.2 we require $(10K - 1)$ parameters to describe a K component Gaussian mixture model, whilst the number of parameters

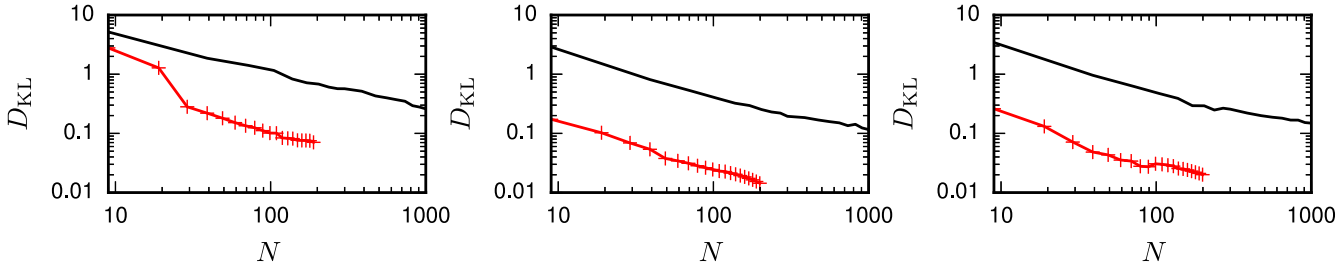

Figure 11. Kullback–Leibler divergences as a function of the number of parameters required. Smaller divergences indicate a greater degree of similarity between the two distributions and so a more successful approximation. In black we plot $D_{\rm KL}$ for thinned MCMC chains relative to the long unthinned chain, using kernel density estimates of both and in red we plot values the $D_{\rm KL}$ between the Gaussian mixture approximation and the kernel density estimate of the unthinned chain, with crosses indicating different values of K running from $K = 1$ to 20. In the left-hand plot we show values for star A, in the middle star B and star C on the right.

needed to describe a thinned MCMC chain is the dimensionality (i.e. three) multiplied by the number of samples in the chain.

In Fig. 11 we compare the $D_{\rm KL}$ found using the Gaussian mixture model approximation to those obtained using thinned MCMC chains as a function of the number of parameters required. To achieve a given $D_{\rm KL}$, the Gaussian mixture model requires an order of magnitude fewer points than the thinned MCMC chain.

We note that $D_{\rm KL}$ for the Gaussian mixture does not pass below $\sim$0.02 for any of the three stars shown. There are a number of sources of error and noise that will prevent a perfect agreement between Q and P, and so $D_{\rm KL} = 0$ being achieved. Most fundamentally, the exact marginal likelihood will not, in general, take a form that can be fit with $K \leq 20$ Gaussian components. In particular the marginal likelihood for star A takes a more complicated form than that of B or C (Fig. 8), which is reflected in correspondingly large values of $D_{\rm KL}$. In addition, we do not actually know the exact marginal likelihood. Instead we have only a noisy kernel density estimate of it, which limits our ability to fit smooth functions, such as Gaussian mixtures to it. Also, we limit the EM algorithm that fits the Gaussian mixture to a maximum number of iterations. Consequently it will generally not achieve the absolute best fit. The resulting error will be manifested in a small contribution to the measured $D_{\rm KL}$. Despite these shortcomings of our $D_{\rm KL}$ tests, we nevertheless believe that it is evident that our Gaussian mixture fits produce very good descriptions of the marginal likelihoods. The $D_{\rm KL} \simeq 0.02$ achieved for stars B and C indicate that our Gaussian mixture fit differs from the marginal likelihood by, at most, $\sim$2 per cent on average.

3.3.2 *Kolmogorov–Smirnov tests*

With 1D data it is common to compare samples and/or distributions using the Kolmogorov–Smirnov distance,

$$D_{\rm KS}(P, Q) \equiv \sup_x |P(x) - Q(x)|, \tag{25}$$

where P and Q are cumulative distributions, either derived directly from a probability distribution, or found empirically from a sample of points. The Kolmogorov–Smirnov distance is not immediately applicable to our situation, as it is defined only for 1D distributions $P(x)$ and $Q(x)$. In this 1D case there are only two possible cumulative distribution functions, either $\mathrm{p}(x < X)$ or $\mathrm{p}(x > X)$, each of which is uniquely defined by the other, because

$$\mathrm{p}(x \leq X) = 1 - \mathrm{p}(x > X). \tag{26}$$

In $p \geq 2$ dimensions the notion of a cumulative distribution function breaks down. One way of proceeding (Peacock 1983) is by constructing cumulative distribution functions (CDFs) with respect to the coordinate axes, such as $\mathrm{p}(x_1 < X_1, x_2 < X_2, \ldots, x_p < X_p)$ or $\mathrm{p}(x_1 > X_1, x_2 < X_2, \ldots, x_p > X_p)$ and so on. For each of our $p = 3$ variables we are free to choose either sign of the inequality when constructing the CDF, giving $2^p = 8$ different possibilities. We follow Peacock (1983) in calculating the 1D KS distance for all eight possibilities of CDFs for P and Q, then taking the maximum such distance as our measure of the 'similarity' of the two functions.[6]

An advantage of this scheme over the KL divergence is that it can be applied directly to the samples from MCMC chains: it avoids the need for kernel density estimates of either P or, in the case of thinned MCMC chains, Q. Fig. 12 shows the results. As with the KL divergences, for a given number of parameters the Gaussian mixture model provides a far better approximation to the marginal likelihood than a thinned MCMC chain can.

4 SUMMARY

We have considered how one should measure the distance and extinction to individual stars for use in constructing extinction maps of the whole Galaxy. We advocate the use of monochromatic extinctions, since, unlike bandpass measures such as A_V and $E(B-V)$, monochromatic extinctions are linear functions of the dust column density and are independent of the source SED. In particular we suggest the use of A_{4000}, the monochromatic extinction at 4000 Å because of its insensitivity to the dust grain size distribution.

We have developed one way of calculating the marginal likelihood $\mathrm{p}(\tilde{\boldsymbol{y}}|\mu, a_{4000}, R_{5495}, \alpha, \beta)$ by marginalizing the (unknown and, for our purposes, uninteresting) fundamental parameters of the star in order to estimate the marginal likelihood. As this integration is not possible analytically, we suggest a scheme for doing so using MCMC methods, specifically the affine invariant ensemble sampler of Goodman & Weare (2010).

We find that the resulting marginal likelihood function can be described very well using a Gaussian mixture model composed of only $K \simeq 5$ Gaussians. Using thinned MCMC chains would require vastly more parameters to achieve the same level of fidelity. Having such a compact description of $\mathrm{p}(\tilde{\boldsymbol{y}}|\mu, a_{4000}, R_{5495}\alpha, \beta)$ is

[6] The multidimensional analogue of equation (26) means that any one of these CDFs is completely determined by the other $2^p - 1$. So, there are $2^p - 1 = 7$ independent CDFs, but this redundancy does not affect Peacock's argument.

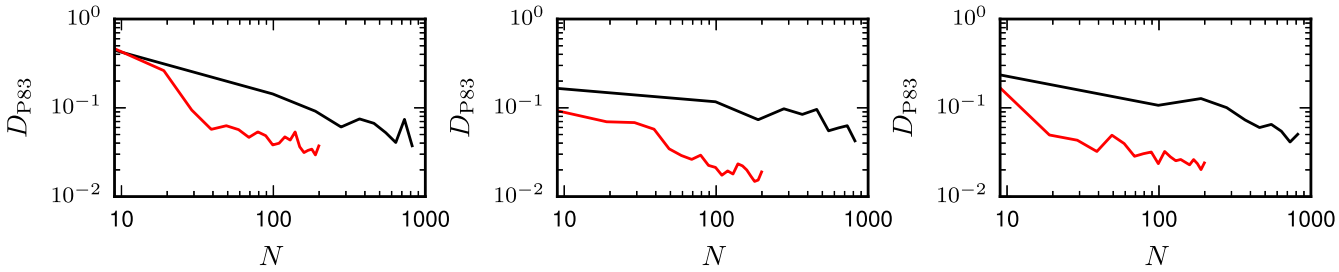

Figure 12. Peacock (1983) distances as a function of the number of parameters required. Smaller distances indicate a greater degree of similarity between the two distributions and so a more successful approximation. In black we plot distances for thinned MCMC chains relative to the unthinned chain and in red we plot values the distances between the Gaussian mixture approximation and the unthinned chain. In the left-hand plot we show values for star A, in the middle star B and star C on the right.

vital when one is constructing maps from large catalogues of stars. Another advantage of expressing the marginal likelihood as a sum of Gaussians is that it makes further marginalization of any or all of the parameters (μ, a_{4000}, R_{5495}) straightforward. This is particularly important if one models the dust density distribution as a Gaussian random field (Sale & Magorrian 2014).

In common with Green et al. (2014), the approach adopted in Sale & Magorrian (2014) is to split the production of 3D dust maps into two distinct steps. First we estimate the marginal likelihood $\mathrm{p}(\tilde{\boldsymbol{y}}|\mu, a_{4000}, R_{5495}, \alpha, \beta)$ of distance modulus μ and (log) extinction a_{4000} to each star in the catalogue. Then we construct maps from these distances and extinctions. The method we present in this paper for carrying out the first of these two steps is very similar to the method Green et al. (2014) use for calculating their posterior pdf $\mathrm{p}(\mu, A|\tilde{\boldsymbol{y}})$. The most important differences are that we use monochromatic extinctions and we return the result in a compact multi-Gaussian form. Our sample of Galactic plane stars meant that we could reasonably use a simple prior β on stellar distances and intrinsic parameters $\boldsymbol{x}$: this is easy to change for more extended samples.

The alternative to these two-step approaches would be to infer simultaneously the distance–extinction relationship and the properties of all the stars that trace it (Sale 2012). The benefit of this is that MCMC schemes operating in the extended space of the stars' intrinsic parameters and their (μ, a_{4000}, R_{5495}) would tend to avoid regions of (μ, a_{4000}, R_{5495}) that are a posteriori unlikely, reducing the computing load. The downside is that parallelization becomes very difficult, making it infeasible to scale up to large data sets. In contrast, in the two-step procedure one has no way of knowing what portions of (μ, a_{4000}, R_{5495}) parameter space are going to be important, and so it has to be explored thoroughly. But this is a small price to pay for the trivial parallelization opportunities.

The software libraries used to obtain the results in this paper are available online, including a library for manipulating isochrones[7] and the code used to sample the marginal likelihood and fit it with a Gaussian mixture model.[8]

ACKNOWLEDGEMENTS

The research leading to the results presented here was supported by the United Kingdom Science Technology and Facilities Council (STFC, ST/K00106X/1), the European Research Council under the European Union's Seventh Framework Programme (FP7/2007-2013)/ERC grant agreement no. 321067. JM thanks the Institut d'Astrophysique de Paris for their hospitality and Ville de Paris for support through their 'Research in Paris' programme.

[7] https://github.com/stuartsale/iso_lib
[8] https://github.com/stuartsale/marg_iso

REFERENCES

Bailer-Jones C. A. L., 2011, MNRAS, 411, 435
Barentsen G. et al., 2014, MNRAS, 444, 3230
Berry M. et al., 2012, ApJ, 757, 166
Bessell M. S., 1990, PASP, 102, 1181
Bessell M. S., 2005, ARA&A, 43, 293
Bessell M., Bloxham G., Schmidt B., Keller S., Tisserand P., Francis P., 2011, PASP, 123, 789
Bressan A., Marigo P., Girardi L., Salasnich B., Dal Cero C., Rubele S., Nanni A., 2012, MNRAS, 427, 127
Cardelli J. A., Clayton G. C., Mathis J. S., 1989, ApJ, 345, 245
Carrasco Kind M., Brunner R. J., 2014, MNRAS, 441, 3550
Casagrande L., VandenBerg D. A., 2014, MNRAS, 444, 392
Castelli F., Kurucz R. L., 2003, in Piskunov N., Weiss W. W., Gray D. F., eds, Proc. IAU Symp. 210, Modelling of Stellar Atmospheres. Kluwer, Dordrecht, p. 20P
Cohen M., Wheaton W. A., Megeath S. T., 2003, AJ, 126, 1090
Crawford D. L., Barnes J. V., 1970, AJ, 75, 978
Crawford D. L., Mander J., 1966, AJ, 71, 114
di Francesco J. et al., 2010, A&A, 518, L91
Draine B. T., 2003, ARA&A, 41, 241
Drew J. E. et al., 2005, MNRAS, 362, 753
Drew J. E. et al., 2014, MNRAS, 440, 2036
Fitzpatrick E. L., 2004, in Witt A. N., Clayton G. C., Draine B. T., eds, ASP Conf. Ser. Vol. 309, Astrophysics of Dust. Astron. Soc. Pac., San Francisco, p. 33
Fitzpatrick E. L., Massa D., 2007, ApJ, 663, 320
Foreman-Mackey D., Hogg D. W., Lang D., Goodman J., 2013, PASP, 125, 306
Golay M., ed., 1974, Astrophysics and Space Science Library, Vol. 41, Introduction to Astronomical Photometry. Reidel, Dordrecht
Goodman J., Weare J., 2010, Commun. Appl. Math. Comput. Sci., 5, 65
Green G. M. et al., 2014, ApJ, 783, 114
Hanson R. J., Bailer-Jones C. A. L., 2014, MNRAS, 438, 2938
Hewett P. C., Warren S. J., Leggett S. K., Hodgkin S. T., 2006, MNRAS, 367, 454
Hou F., Goodman J., Hogg D. W., Weare J., Schwab C., 2012, ApJ, 745, 198
Husser T.-O., Wende-von Berg S., Dreizler S., Homeier D., Reiners A., Barman T., Hauschildt P. H., 2013, A&A, 553, A6
Lallement R., Vergely J.-L., Valette B., Puspitarini L., Eyer L., Casagrande L., 2014, A&A, 561, A91
Lucas P. W. et al., 2008, MNRAS, 391, 136
Luck R. E., Lambert D. L., 2011, AJ, 142, 136

McCall M. L., 2004, AJ, 128, 2144
Magorrian J., 2014, MNRAS, 437, 2230
Maiz Apellániz J., 2013, in Guirado J. C., Lara L. M., Quilis V., Gorgas J., eds, Highlights of Spanish Astrophysics VII. Spanish Astronomical Society (SEA), Valencia, p. 583
Majewski S. R., Zasowski G., Nidever D. L., 2011, ApJ, 739, 25
Marshall D. J., Robin A. C., Reylé C., Schultheis M., Picaud S., 2006, A&A, 453, 635
Munari U., Sordo R., Castelli F., Zwitter T., 2005, A&A, 442, 1127
O'Donnell J. E., 1994, ApJ, 422, 158
Patat F. et al., 2011, A&A, 527, A91
Peacock J. A., 1983, MNRAS, 202, 615
Pedregosa F. et al., 2011, J. Machine Learning Res., 12, 2825
Planck Collaboration XI, 2014, A&A, 571, A11
Sale S. E., 2012, MNRAS, 427, 2119
Sale S. E., Magorrian J., 2014, MNRAS, 445, 256
Sale S. E. et al., 2009, MNRAS, 392, 497
Sale S. E. et al., 2014, MNRAS, 443, 2907
Schlafly E. F., Finkbeiner D. P., 2011, ApJ, 737, 103
Schlegel D. J., Finkbeiner D. P., Davis M., 1998, ApJ, 500, 525
Schwarz G., 1978, Ann. Stat., 6, 461
Stead J. J., Hoare M. G., 2009, MNRAS, 400, 731
Straizys V., Kuriliene G., 1981, Ap&SS, 80, 353
Stubbs C. W., Doherty P., Cramer C., Narayan G., Brown Y. J., Lykke K. R., Woodward J. T., Tonry J. L., 2010, ApJS, 191, 376
Vergely J.-L., Freire Ferrero R., Siebert A., Valette B., 2001, A&A, 366, 1016
Weingartner J. C., Draine B. T., 2001, ApJ, 548, 296

APPENDIX A: TABULATED RESPONSE TO A_{4000} FOR A VARIETY OF PHOTOMETRIC BANDS

We make available with this paper[9] a tabulation of the coefficients $c_X^{(1)}$ and $c_X^{(2)}$ of equation (11) for a variety of filters and the full range of Fitzpatrick (2004) reddening laws. We do this for SEDs along a solar metallicity main sequence, defined in ($T_{\rm eff}$, $\log g$) by Straizys & Kuriliene (1981). We also include a Rayleigh–Jeans spectrum to demonstrate the limiting behaviour for extremely hot stars.

In Table A1 we list the photometric systems and their constituent filters that we employ. For all the survey filter sets we also employ the detector quantum efficiency curve and atmospheric transmission for the instrument and site used.

We include two 'standard' filter sets: the Bessell (1990) *UBVRI* set and a Strömgren filter set with the *uvby* transmissions taken from Crawford & Barnes (1970) and H$\beta_{\rm wide}$ and H$\beta_{\rm narrow}$ from Crawford & Mander (1966). We present results for these filters using the INT/WFC CCD quantum efficiency and the Patat et al. (2011) model for atmospheric absorption at Cerro Paranal. In addition, for reference purposes, we also provide results for the Bessell (1990) filter set with no atmospheric absorption and a 100 per cent efficient detector.

A sample of the table of values of $c_X^{(1)}$ and $c_X^{(2)}$ is given in Table A2.

Table A1. A list of the systems and filters for which we tabulate the response to extinction.

System	Filters	Source
Bessell	*UBVRI*	Bessell (1990)
Strömgren	*ubvy*H$\beta_{\rm narrow}$H$\beta_{\rm wide}$	Crawford & Barnes (1970), Crawford & Mander (1966)
2MASS	$JHK_{\rm s}$	Cohen, Wheaton & Megeath (2003)
Gaia	*G*	http://www.cosmos.esa.int/web/gaia/transmissionwithoriginal
INT (IPHAS/UVEX)	*Ugri*Hα	http://www.ing.iac.es/astronomy/instruments/wfc/
PAN-STARRS	*grizy*	Stubbs et al. (2010)
SDSS	*ugriz*	https://www.sdss3.org/instruments/camera.php#Filters
Skymapper	*uvgriz*	Bessell et al. (2011)
UKIDSS	*ZYJHK*	Hewett et al. (2006)
VISTA	$ZYJHK_{\rm s}$	http://www.eso.org/sci/facilities/paranal/instruments/vircam/inst.html
VST	*ugri*Hα	http://www.eso.org/sci/facilities/paranal/instruments/omegacam/tools.html, Drew et al. (2014)

SUPPORTING INFORMATION

Additional Supporting Information may be found in the online version of this article:

Table A2. An extract from the compilation of the coefficients $c_X^{(1)}$ and $c_X^{(2)}$ of equation (11) for the filters listed in Table A1 and the Fitzpatrick (2004) reddening laws (http://mnras.oxfordjournals.org/lookup/suppl/doi:10.1093/mnras/stv068/-/DC1).

Please note: Oxford University Press is not responsible for the content or functionality of any supporting materials supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.

This paper has been typeset from a TEX/LATEX file prepared by the author.

[9] https://github.com/stuartsale/A4000_coeffs