Photo-D with LSST: stellar photometric distances out to the edge of the Galaxy

Lovro Palaversa [ID],[1] Željko Ivezić [ID],[2] Karlo Mrakovčić [ID],[3] Bob Abel [ID],[4] Oleksandra Razim [ID],[1]
Filip Matković [ID],[5] Connor Yablonski [ID],[6] Toni Šarić [ID],[7] Tomislav Jurkić [ID],[3] Neven Caplar [ID],[2] and
Mario Jurić [ID][2]

[1] Ruđer Bošković Institute, Bijenička cesta 54, 10000 Zagreb, Croatia
[2] Department of Astronomy and DiRAC Institute, University of Washington, Box 351580, Seattle, WA 98195, USA
[3] Faculty of Physics, University of Rijeka, Radmile Matejčić 2, 51000 Rijeka, Croatia
[4] Olympic College, 1600 Chester Avenue, Bremerton, WA 98337, USA, and DiRAC Institute, University of Washington, Box 351580, Seattle, WA 98195, USA
[5] Hvar Observatory, Faculty of Geodesy, University of Zagreb, Kačićeva 26, 10000 Zagreb, Croatia
[6] Department of Astronomy, University of Washington, Box 351580, Seattle, WA 98195, USA
[7] University of Split - FESB, R. Boškovića 32, 21000 Split, Croatia

Abstract

As demonstrated with SDSS, Pan-STARRS and most recently with Gaia data, broad-band near-UV to near-IR stellar photometry can be used to estimate distance, metallicity and interstellar dust extinction along the line of sight for stars in the Galaxy. Anticipating photometric catalogs with tens of billions of stars from Rubin's Legacy Survey of Space and Time (LSST), we present a Bayesian procedure and pipeline that build on previous work and can handle LSST-sized datasets. Likelihood computations utilize MIST/Dartmouth isochrones and priors are derived from TRILEGAL-based simulated LSST catalogs from Dal Tio et al. (2022). The computation speed is about 10 milliseconds per star on a single core for both optimized grid search and MCMC methods; we show in a companion paper by Mrakovčić et al. how to utilize neural networks to accelerate this performance by up to an order of magnitude. We validate our pipeline, named PhotoD (in analogy with photo-z, photometric redshifts of galaxies) using both simulated catalogs, and SDSS, DECam and Gaia photometry. We intend to make LSST-based value-added PhotoD catalogs publicly available via the Rubin Science Platform with every LSST Data Release.

1. INTRODUCTION

In order to map the Milky Way in three dimensions, distances to its stars must be accurately estimated. Kinematic studies based on proper motion data also require estimates of stellar distances. There are a variety of astronomical methods to estimate distances to stars, ranging from direct geometric (trigonometric) methods for nearby stars to indirect methods based on astrophysics for more distant stars.

As demonstrated with SDSS (Jurić et al. 2008; Ivezić et al. 2008), Pan-STARRS (Green et al. 2014; Green et al. 2019) and most recently with Gaia data (Bailer-Jones et al. 2021), broad-band near-UV to near-IR stellar photometry is sufficient to estimate distance, metallicity and interstellar dust extinction along the line of sight for stars in the Galaxy. In analogy with photo-z, photometric redshifts of galaxies, hereafter we refer to these methods as photo-D. The photo-D method is conceptually quite simple: multi-dimensional color tracks (either empirical or model-based), parametrized by luminosity, metallicity and extinction, are fit to observed colors and the best-fit produces estimates of the three model parameters. The method relies on strong correlations between stellar colors and stellar luminosity for dominant stellar populations such as main-sequence stars, red giants, white dwarfs, and even for the majority of

unresolved binary stars. These are the same correlations that are responsible for the abundant structure seen in the Hertzsprung-Russell diagram. In addition to colors and luminosity, these correlations also involve metallicity in the case of main-sequence and red giant stars, and surface gravity in the case of white dwarfs. For the youngest main-sequence stars, stellar age may play a role, too. Measured stellar colors are also affected by interstellar dust extinction along the line of sight towards the star. Consequently, sufficiently accurate measurements of apparent brightness and sufficient number of UV to IR colors, such as those that Rubin Observatory's Legacy Survey of Space and Time (LSST; Ivezić et al. 2019) will provide, can be used to accurately estimate these parameters, and ultimately stellar distances.

LSST-based stellar distance estimates will significantly improve available distance catalogs, such as those recently produced by Green et al. (2019) and Bailer-Jones et al. (2021). First, the sample size will be increased by more than an order of magnitude, and exceed 10 billion stars. Distance accuracy for stars with sufficiently small photometric errors will be within the 5-10% range, or about twice as accurate as for surveys lacking the UV u band (which provides metallicity constraints). LSST-based stellar distances will reach about 10 times further than Gaia's color-based distances and will be transformative for studies of the Milky Way in general, and of its stellar and dark matter halo in particular.

In this paper we present a Bayesian procedure and pipeline that build on previous work and can handle LSST-sized datasets. In §2, we describe our methodology and in §3 we test the pipeline using both simulated catalogs, and SDSS, DECam and Gaia photometry. We discuss possibilities for further improvements and catalog public release plans in §4.

2. METHODOLOGY

In this Section we discuss a Bayesian method for stellar photometric distance estimation and its implementation. We start with a brief overview of Bayesian methodology and then discuss in detail our choices of likelihoods and priors, and how they differ from previous work. A pipeline implementation of this method and a discussions of its performance are presented in the next Section.

2.1. Bayesian Approach to Stellar Photometric Distance Estimation

In the most general terms, our aim is to estimate for each star an array (a vector) of model parameters $\vec{\theta}$, for some model M (e.g., main-sequence stars), using data vector $\vec{D}$ and priors for M and $\vec{\theta}$. Data, or observations, include multi-band photometry that is used to construct colors, $\vec{c}$. In the case of SDSS, colors include $u-g$, $g-r$, $r-i$, $i-z$, and in the case of LSST also $z-y$. In addition to colors, $\vec{D}$ also includes an apparent magnitude and hereafter we choose the r-band magnitude. Therefore, $\vec{D} = (r, \vec{c})$. Observations also provide stellar sky coordinates and we address their role further below when discussing priors.

Models M (either empirical or computational, see below) need to provide stellar colors as functions of model parameters $\vec{\theta}$. The three principal parameters that control stellar colors at a fixed stellar age and at the accuracy level relevant here ($\sim$1 %) include absolute magnitude (here chosen in the r band, M_r), metallicity ($[Fe/H]$) and surface gravity. In the case of main-sequence stars and red giants, the color tracks can be expressed as functions of M_r and $[Fe/H]$ along an isochrone, without having to explicitly specify surface gravity. With other populations, such as white dwarfs, the roles of metallicity and surface gravity are different; we will assume here for notational simplicity that intrinsic stellar colors depend on M_r and $[Fe/H]$. We do not consider stellar age as a model parameter and use it as a model label for reasons discussed below.

The observed stellar colors also depend on the interstellar dust extinction along the line of sight. Hereafter we will assume (and justify further below) that dust extinction is fully specified by a single model parameter, A_r. A_r is extinction in the r band and extinction in other bands is proportional to A_r, with known constants of proportionality (this assumption can be relaxed, see below). Once M_r and A_r are constrained, that is, the posterior probability distribution $q(Q_r)$, where $Q_r = M_r + A_r$ is known, the distance modulus Δ can be computed using relationship

$$r = M_r + A_r + \Delta. \tag{1}$$

In practice the uncertainty of the observed magnitude, r^{obs}, is always much smaller than the width of $q(Q_r)$ distribution (at the bright end by an order of magnitude, $\sim$0.01 mag vs. $\sim$0.1 mag). Therefore,

$$p(\Delta) = q(r^{obs} - \Delta), \tag{2}$$

with the mean values related as $\overline{\Delta} = r^{obs} - \overline{Q_r}$ and the uncertainty of Δ approximately equal to the standard deviation of $q(Q_r)$. For related discussion, please see §2.4 in Bailer-Jones et al. (2021).

In summary, $\vec{D} = (r, \vec{c})$ and $\vec{\theta} = (M_r, [Fe/H], A_r)$. Data $\vec{D}$ and model parameters $\vec{\theta}$ are related via the Bayes theorem (see, e.g., Chapter 5 in Ivezić et al. 2020),

$$p(M, \vec{\theta}|D, I) = \frac{p(D|M, \vec{\theta}, I)\, p(M, \vec{\theta}|I)}{p(D|I)}. \tag{3}$$

where I is prior information. Strictly speaking, the vector $\vec{\theta}$ should be labeled by M since different models may be described by different parameters (e.g., main-sequence stars vs. white dwarfs).

The result $p(M, \vec{\theta}|D, I)$ is called the *posterior* pdf (probability density function) for model M *and* parameters $\vec{\theta}$, given data D and other prior information I. This term is a $(k+1)$-dimensional pdf in the space spanned by $k = 3$ model parameters and the model index M. The term $p(D|M, \vec{\theta}, I)$ is the *likelihood* of data *given* some model M and given some fixed values of parameters $\vec{\theta}$ describing it, and all other prior information I. The term $p(M, \vec{\theta}|I)$ is the a priori joint probability, or simply prior, for model M and its parameters $\vec{\theta}$ in the absence of any of the data used to compute likelihood. The prior can be expanded as

$$p(M, \vec{\theta}|I) = p(\vec{\theta}|M, I)\, p(M|I). \tag{4}$$

The term $p(D|I)$ is the *probability of data*, or the prior predictive probability for D. It provides proper normalization for the posterior pdf; for simplicity, it is usually not explicitly computed when estimating model parameters: rather,
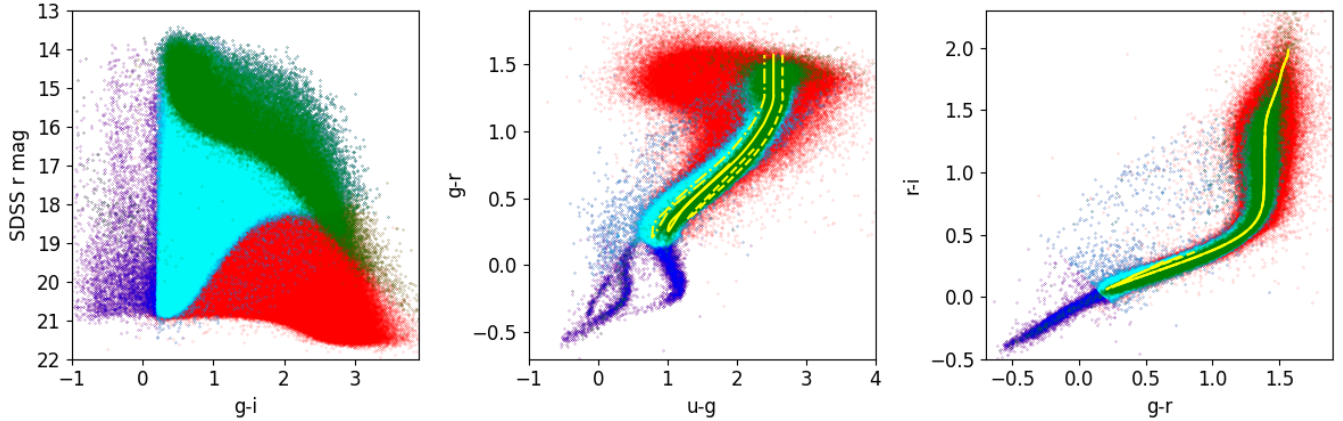
Figure 1. An illustration of multiple stellar populations. The left panel is a color-magnitude diagram for 841,000 stars from the SDSS Stripe 82 Standard Star Catalog (variable sources are excluded, Thanjavur et al. 2021) that have Gaia matches within 0.15 arcsec (after correcting for proper motion using Gaia measurements). A subset of 415,000 stars with $r < 22$ and $u < 22$ are overplotted as blue dots, and 409,000 of those that also have $0.2 < g - i < 3.5$ (dominated by main-sequence stars and red giants) are overplotted as cyan dots. Finally, 63,000 stars that have a signal-to-noise ratio for Gaia's parallax measurements of at least 20 are shown as green dots (these stars can be used for the calibration of luminosity-color relations). The same symbol color scheme is used in the other two panels. The three yellow lines in the middle panel show the stellar locus parametrization used by Green et al. (2014) for three values of metallicity (left to right): $[Fe/H] = -2, -1, 0$. In the right panel, the impact of metallicity on color-color tracks is negligible and all three are indistinguishable from each other. In the bottom of the middle panel, at $-0.5 < g - r < 0$, the three dark blue sequences correspond to (from left to right) He white dwarfs, H white dwarfs, and blue horizontal branch stars. The clouds of pale blue dots visible above the main stellar locus in the middle and right panels correspond to unresolved binary stars (Smolčić et al. 2004).

$p(M, \vec{\theta}|D, I)$ for a given M is renormalized so that its integral over all model parameters $\vec{\theta}$ is unity. The integral of the prior $p(\vec{\theta}|M, I)$ over all parameters should also be unity, but for the same reason, calculations of the posterior pdf are often done with an arbitrary normalization. An important exception is model selection discussed further below in §2.5, where the correct normalization of the product $p(D|M, \vec{\theta}, I)\, p(\vec{\theta}|M, I)$ is crucial.

Our approach adopted here is essentially the same as used in recent papers[1] by, e.g., Bailer-Jones (2011), Green et al. (2014), Lallement et al. (2014), Gordon et al. (2016), Queiroz et al. (2018), Green et al. (2019) and Bailer-Jones et al. (2021). The main differences compared to these works include:

1. The use of multiple stellar populations, in addition to main-sequence stars and red giants (white dwarfs, unresolved binary stars, blue horizontal branch stars; potentially Miras, quasars and RR Lyrae stars too, which can also be recognized and rejected using variability).

2. Improved color tracks for main-sequence stars and (especially) red giants, including the use of very young (<1 Gyr) populations and an extended $[Fe/H]$ range.

3. Priors based on sophisticated TRILEGAL simulations by Dal Tio et al. (2022) that include multiple stellar populations and also account for the Galaxy's bulge component (see §2.3 below for more details).

We discuss these improvements in detail in the next few sections.

2.2. *Likelihood Computation*

Given a chosen model (i.e., a stellar population) M, the likelihood $p(D|M, \vec{\theta}, I)$ can be explicitly written as

$$\mathcal{L} \equiv p(D|M, \vec{\theta}, I) = p(\vec{c}|M_r, [Fe/H], A_r). \tag{5}$$

---

[1] This approach greatly simplifies when studying faint and distant blue halo stars, as in e.g., Jurić et al. (2008). Such stars are beyond the dust layer which is confined close to the disk and thus A_r can be obtained from IR maps, they have halo metallicities ($[Fe/H] \sim -1.5$), and can be assumed dominated by main-sequence stars. As a result, a simple functional relationship, $M_r = f(g - i)$, or its generalized version that accounts for the shift of M_r as a function of metallicity (Ivezić et al. 2008), can be used to estimate distance in a straightforward manner.

Assuming Gaussian photometric errors that are parametrized by a vector of color uncertainties $\vec{\sigma}$, the log-likelihood is given by

$$\ln(\mathcal{L}) = -\frac{N}{2}\ln(2\pi) - \sum_{i=1}^{N}\ln(\sigma_i) - \frac{1}{2}\sum_{i=1}^{N}\left(\frac{c_i^{obs} - c_i^{mod}}{\sigma_i}\right)^2 \tag{6}$$

where the summation is over all colors (for example, $N = 4$ for SDSS, and $N = 5$ for LSST), c_i^{obs} are the observed colors and c_i^{mod} are the model colors (they are functions of M_r, $[Fe/H]$, and A_r but for notational simplicity we don't explicitly list model parameters). Note that only the last sum involves model predictions for colors (c_i^{mod}).

The model colors can be computed as

$$\vec{c}^{\,mod} = \vec{c}_0(M_r, [Fe/H]) + \vec{\delta c}(A_r), \tag{7}$$

where $\vec{c}_0(M_r, [Fe/H])$ are intrinsic stellar colors for a given stellar population and $\vec{\delta c}(A_r)$ are color corrections due to interstellar dust reddening. Eq. 7 can be thought of as a set of 3-dimensional data cubes, one for each color, that map the triplet $(M_r, [Fe/H], A_r)$ to that color. The likelihood function can be thought of as a 3-dimensional data cube that, for a given set of observed colors c_i^{obs}, maps the triplet $(M_r, [Fe/H], A_r)$ to a 1-dimensional scalar, that is, $ln(\mathcal{L})$ is a scalar function of M_r, $[Fe/H]$, and A_r.

The existence of multiple stellar populations in an SDSS photometric catalog is illustrated in Figure 1. With accurate multi-band photometry that includes a UV band (here SDSS u), main-sequence stars, red giants, white dwarfs, blue horizontal branch (BHB) stars, and unresolved binary stars can be reliably identified (UV photometry is also crucial for constraining metallicity). We discuss our choice of $\vec{c}_0(M_r, [Fe/H])$ for main-sequence stars and red giants next, and then for white dwarfs, unresolved binary stars and BHB stars.

2.2.1. *Empirical Luminosity-Color Tracks for Main-sequence Stars and Red Giants*

Both Berry et al. (2012) and Green et al. (2014) used empirical color tracks for main-sequence stars and red giants (see the left panel in Figure 2, modeled after Figure 1 in Green et al. 2014) derived from SDSS data for globular clusters
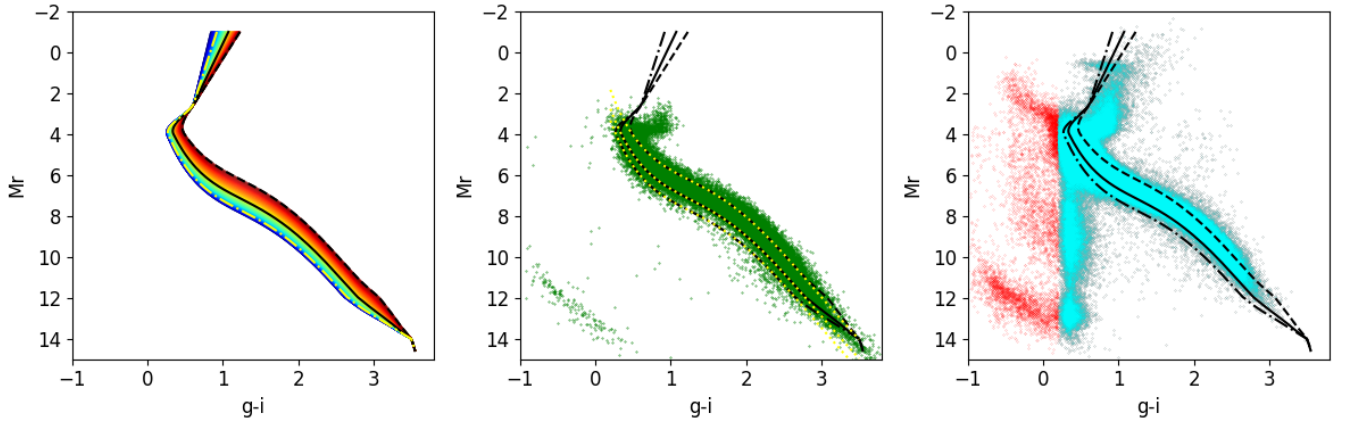

Figure 2. The left panel shows SDSS-based empirical absolute magnitude vs. color parametrization for main-sequence stars and red giants. The data are color-coded by metallicity, ranging from $[Fe/H] = -2.5$ to 0 (blue to red). The three lines correspond to three values of metallicity: $[Fe/H] = -2, -1, 0$ (dot-dashed, solid and dashed, respectively). The middle panel shows a sample of 63,000 stars that have signal-to-noise ratio for Gaia's parallax measurements of at least 20 (white dwarfs can be seen in the lower left corner. Their absolute magnitudes are derived from parallax measurements. The dot-dashed, solid and dashed black lines are the same as in the left panel. For comparison, the essentially identical yellow dotted lines were computed using eqs. A2 and A7 from Ivezić et al. (2008). Note the discrepancy between these parametrizations and data for sub-giant stars ($M_r \sim 3 - 4$ and $g - i \sim 0.8 - 1.1$). The right panel shows a sample of 415,000 stars with $r < 22$ and $u < 22$ as red dots (shown by blue dots in Figure 1), and 409,000 of those that also have $0.2 < g - i < 3.5$ as cyan dots. Their absolute magnitudes were computed using the so-called "photogeometric" distances from Bailer-Jones et al. (2021). The dot-dashed, solid and dashed black lines are the same as in the left and middle panels. About 10,000 stars below the main sequence (about 2.5% of the full sample) seen at $g - i = 0.4$ and $Mr > 7$ are predominantly found at the faint end ($r > 20$) and may be outliers due to the low photometric signal-to-noise ratio.

(for technical details, see Appendix A in Ivezić et al. 2008). These color tracks suffer from three problems. First, as can be seen in Figure 11 from Green et al. (2014), their predicted colors for sub-giant stars between main-sequence turn-off and red giant branch are too blue by about 0.1–0.2 mag. Second, their metallicity grid does not extend to the $[Fe/H] > 0$ range relevant for some disk stars. Finally, they correspond to very old populations (older than a few Gyr) and cannot be used for stars younger than about 1-2 Gyr.

We use a combination of SDSS and Gaia data to demonstrate the first problem. The middle panel in Figure 2 shows a clear discrepancy between SDSS-based empirical tracks and data for sub-giant stars, using parallax-based absolute magnitudes. Nevertheless, it is noteworthy that for main-sequence stars the agreement is excellent. Furthermore, the right panel in Figure 2 demonstrates that SDSS-based photometric distances and Gaia-based "photogeometric" distances from Bailer-Jones et al. (2021) for main-sequence stars are on the "same scale". We discuss these distance scales in more detail in §3.2.2.

We address all three problems by augmenting the SDSS-based empirical isochrones that correspond to old populations with model-based isochrones that span a range of ages and a wider range of metallicities, which results in better agreement with the observations.

2.2.2. *Model-based Isochrones for Main-sequence Stars and Red Giants*

We considered two sets of isochrones: the Dartmouth Stellar Evolution Database[2], hereafter DSED (Dotter et al. 2008), and PARSEC[3] (Bressan et al. 2012) isochrones. Both isochrone sets span adequate ranges of age and metallicity. Unfortunately, the computed color sequences show discrepancies with the observed SDSS stellar locus at the level of 0.1-0.2 mag and cannot be used without further adjustments.

We "augment" the empirical SDSS-based isochrones in two steps:

1. We extend its metallicity range from $[Fe/H] = 0$ to $[Fe/H] = +0.5$ by linear extrapolation of the color vs. metallicity dependence around $[Fe/H] = 0$. We adjust the gradient by a multiplicative factor (0.7) to ensure that the bright end of the stellar locus at $[Fe/H] = +0.5$ agrees with SDSS-Gaia data (shown in the middle panel in Figure 2).

2. For $M_r < M_{TO}$, where turn-off absolute magnitude M_{TO} depends on age and ranges from $M_{TO}=4$ for age = 1 Gyr to $M_{TO} = 5$ for age = 10 Gyr, we "attach" model-based isochrones (we use DSED isochrones hereafter) to the empirical SDSS-based isochrones.

Examples of the resulting color tracks for two representative ages are shown in Figure 3. Sub-giant stars that could not be fit with empirical SDSS-based tracks can now be explained with tracks for old stars and intermediate-range metallicity. We have computed such tracks for a grid of ages but believe that the two choices shown in Figure 3 (1 Gyr and 10 Gyr) should suffice for "non-specialized" bulk processing. The reason is that the locii for ages above 1-2 Gyr look very similar to each other, while the fraction of stars younger than 1 Gyr is very small at the faint apparent magnitude levels probed by SDSS and LSST. Of course, in sky regions with intensive star formation, a "specialized" approach with a fine age grid can be easily executed.

We note strong degeneracies in color space between giants and main sequence stars (see Figure 4). They are especially strong for sub-giant stars with $2 < M_r < 4$, for which it is always possible to find a matching main sequence star closer in 4-dimensional color distance than 0.02 mag (and for $3 < M_r < 4$ even closer than 0.01 mag). Due to photometric scatter, such subgiant stars can be misidentified as a main sequence star and absolute magnitude errors can range up to several magnitudes.

These color tracks for main-sequence stars, blue horizontal branch stars, and red giants account for an overwhelming majority of stars expected in SDSS and LSST catalogs (approximately >95% but the fraction varies with apparent magnitude and sky position). Nevertheless, we also explicitly account for a few additional populations: white dwarfs and unresolved binary stars.

2.2.3. *Luminosity-Color Tracks for White Dwarfs*

High-precision SDSS photometry clearly shows two white dwarf (WD) sequences in the $g - r$ vs. $u - g$ color-color diagram (see, e.g., Figures 23 and 24 in Ivezić et al. 2007, as well as the middle panel in Figure 1). A comparison with

---

[2] See http://stellar.dartmouth.edu/models

[3] http://stev.oapd.inaf.it/cgi-bin/cmd_3.7

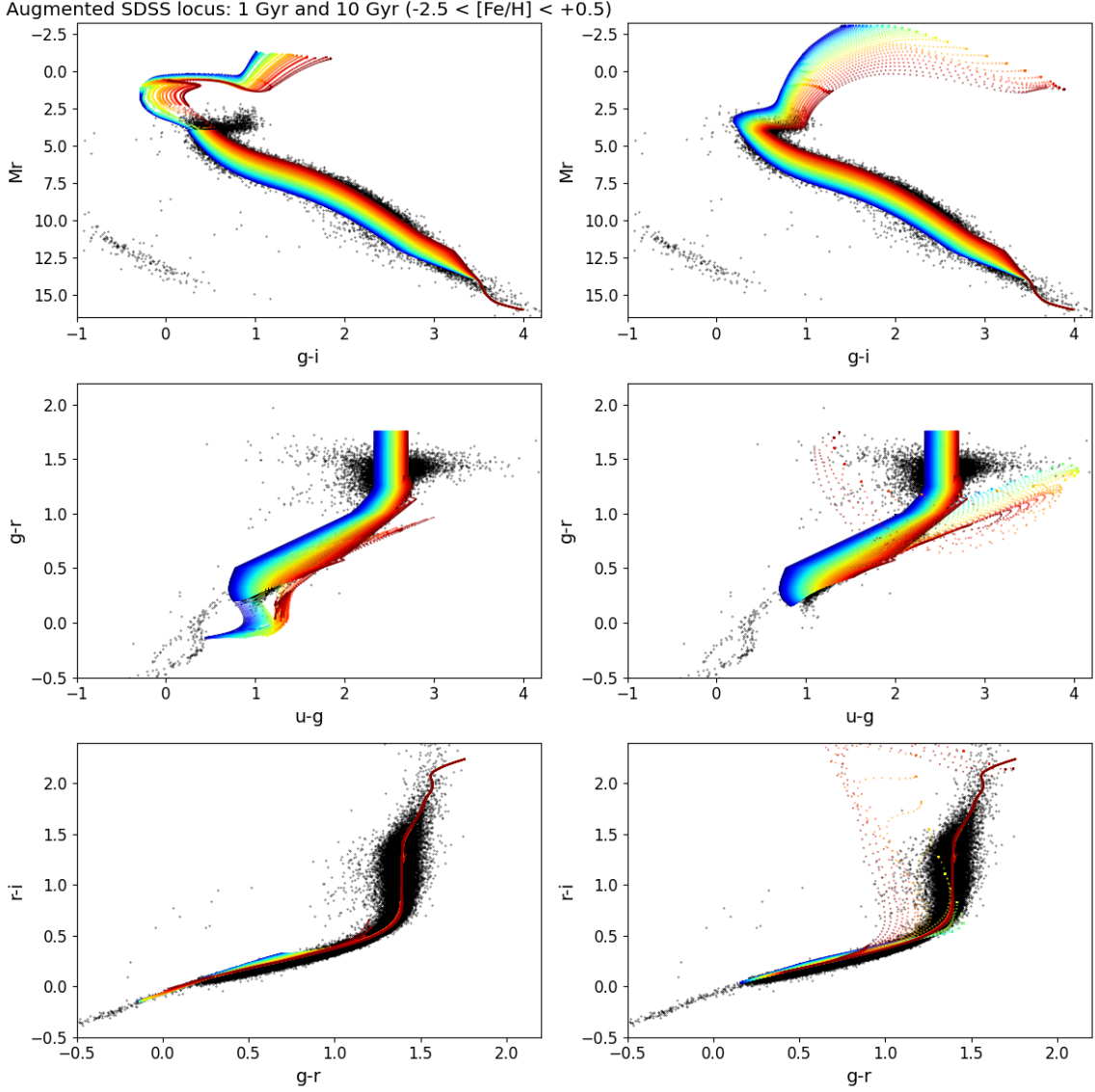Augmented SDSS locus: 1 Gyr and 10 Gyr (-2.5 < [Fe/H] < +0.5)

Figure 3. Augmented SDSS color tracks for two isochrone ages (left: 1 Gyr; right: 10 Gyr) and the full metallicity range $(-2.5 < [Fe/H] < +0.5$, color-coded linearly from blue to red). The black dots show the same sample of 63,000 stars from the middle panel in Figure 2. As can be seen in the top right panel, sub-giant stars that could not be fit with empirical SDSS-based tracks (see the middle panel in Figure 2) can be explained with tracks for old stars and intermediate-range metallicity. The sharp feature protruding from the main locus in the middle two panels corresponds to most luminous and evolved high-metallicity stars. Note that diagrams in the bottom row, which do not include the u band, show very little dependence on metallicity.

models from Bergeron et al. (1995), as well as with SDSS spectra, reveals that the two sequences correspond to H and He white dwarfs, with the mean $\log(g) = 8.0$ for the H sequence and $\log(g) = 8.5$ for the He sequence. Upper limits for the scatter of $\log(g)$ around these mean sequences appear as no more than 0.5. We use three modern white dwarf catalogs: the Montreal White Dwarf Database[4] (Dufour et al. 2017), the Gaia EDR3 White Dwarf Catalog[5] (Gentile Fusillo et al. 2021), and the Gaia-based White Dwarf Database[6] (García-Zamora et al. 2023) to validate the Bergeron et al. (1995) models and derive small color offsets that bring models in perfect agreement with these three datasets.

[4] https://www.montrealwhitedwarfdatabase.org

[5] https://warwick.ac.uk/fac/sci/physics/research/astro/research/catalogues/

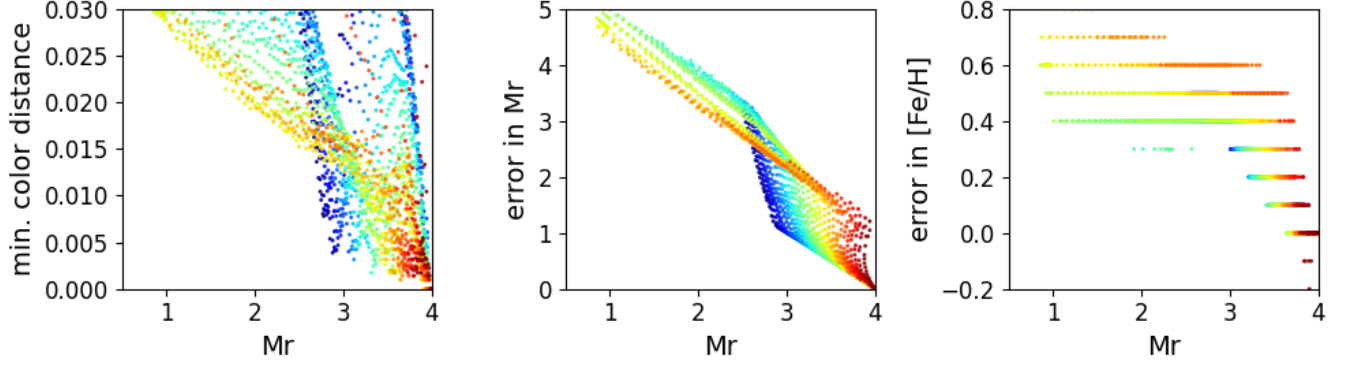[6] https://cdsarc.cds.unistra.fr/viz-bin/cat/J/A+A/679/A127

Figure 4. Analysis of color degeneracies between giants and main sequence stars. The left panel shows minimum color distance in 4-dimensional color space ($u - g$, $g - r$, $r - i$, $i - z$) between a position on the locus with $M_r < 4$ (giants) and any position on the main-sequence $M_r > 4$ locus (in magnitudes). Symbols are color-coded by metallicity, linearly from blue to red for the range -2.5 to 0.5). The middle and right panels show errors in absolute magnitude and metallicity when a giant is misidentified as a main sequence star closest to it in color space.

We consider only white dwarfs with cataloged DA, DB or DC spectral class, at galactic latitudes further than 10 degrees from the plane, with SDSS photometry, apparent magnitudes $5 \leq m \leq 22$ in any band m, and $15 \leq r \leq 19$, where r is the SDSS r-band magnitude. Observed magnitudes are corrected for interstellar dust extinction using maps from Schlegel et al. (1998) and per-band extinction coefficients discussed in §2.2.6 below. We group the DB and DC spectral classes as He dwarfs (1,939 objects), while the DA spectral class corresponds to H white dwarfs (9,307 objects). Their absolute magnitudes were calculated using "photogeometric" distances from Bailer-Jones et al. (2021).

For each WD type (H and He) and SDSS color, we bin the data into 20 M_r bins and compute the median value of a given color in each bin. The color vs. M_r sequences from Bergeron et al. (1995) are then slightly (up to 0.1 mag) shifted in color so that the mean offset for all bins vanishes. The only case where a small (up to 0.1 mag) linear adjustment of the model track was needed is the $u-g$ color for H models at $g-r < -0.3$. After these color adjustments are applied, all color vs. M_r model sequences were linearly interpolated to a common M_r grid ($8.5 < M_r < 14.5$, with a step of 0.02 mag). The resulting two model tracks for H and He white dwarfs are shown in Figure 5.
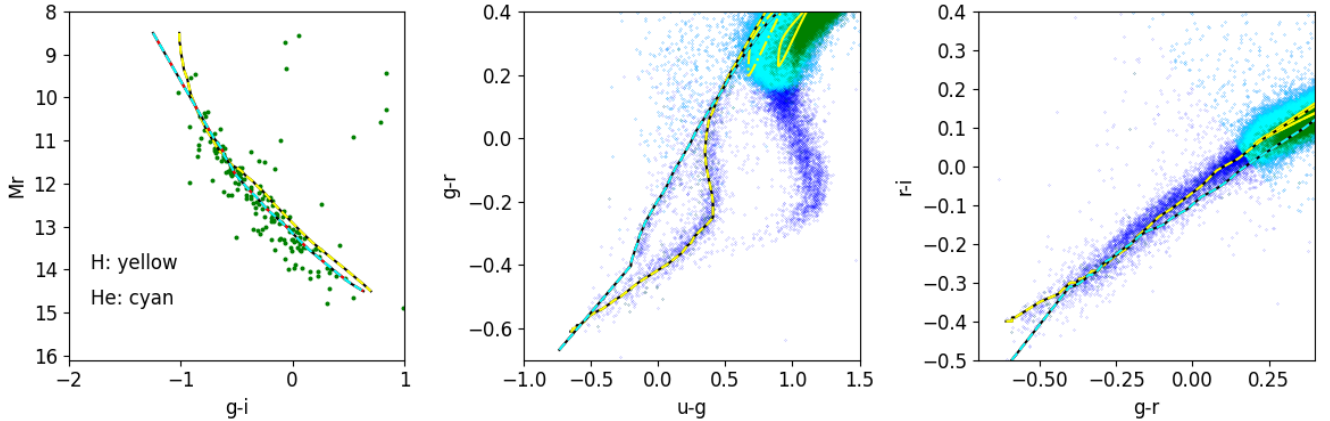

Figure 5. Luminosity-color tracks for H and He white dwarfs based on models from Bergeron et al. (1995), with colors slightly shifted to bring them in agreement with SDSS photometry. Data, shown as symbols, are the same as in Figure 1 and in the middle panel in Figure 2. Note the crucial role of the $u - g$ color for distinguishing H and He tracks.

2.2.4. Luminosity-Color Tracks for Unresolved Binaries

Following Smolčić et al. (2004), who discovered the so-called "second stellar locus" of unresolved binary stars in SDSS dataset, we generate luminosity-color tracks for unresolved binaries consisting of an M dwarf and a white dwarf. We limit models to M dwarfs because white dwarfs would have negligible impact on colors of more luminous main-sequence stars. M dwarfs are parametrized with M_r and $[Fe/H]$, and H/He white dwarfs with M_r. Therefore, model tracks for unresolved binary stars would have three model parameters. After numerical experimentation that revealed small model sensitivity to metallicity, we decided to consider only M dwarfs with two metallicity values corresponding to mean metallicity for disk and halo populations: $[Fe/H] = 0$ and $[Fe/H] = -1.5$, respectively. For the remaining two model parameters, we selected M_r corresponding to the total system luminosity and the component luminosity ratio in the r band.

We sample the M dwarf luminosity in the range $8.5 \leq M_r \leq 14.5$, with a step of 0.1 mag, and for each value generate a track by adding the luminosity of a white dwarf, where the track is sampled on the same grid of M_r but using white dwarf models (separately for H and He models). There are four model families (for two families of white dwarf models and two values of $[Fe/H]$ for M dwarfs), each with 3,721 M_r vs. color entries. The color tracks for H white dwarfs and $[Fe/H] = -1.5$ M dwarfs are shown in Figure 6. Analogous tracks for binaries composed of an M dwarf with $[Fe/H] = 0$ and either an H or He white dwarf look very similar though not identical (differences are at most a few tenths of a magnitude). In practice, it will be nearly impossible to constrain M dwarf metallicity, and very hard to distinguish H and He white dwarfs. It may turn out that it will be sufficient to use a single model family to fit LSST data (e.g., a combination of an H white dwarf and a $[Fe/H] = -1.5$ M dwarf).
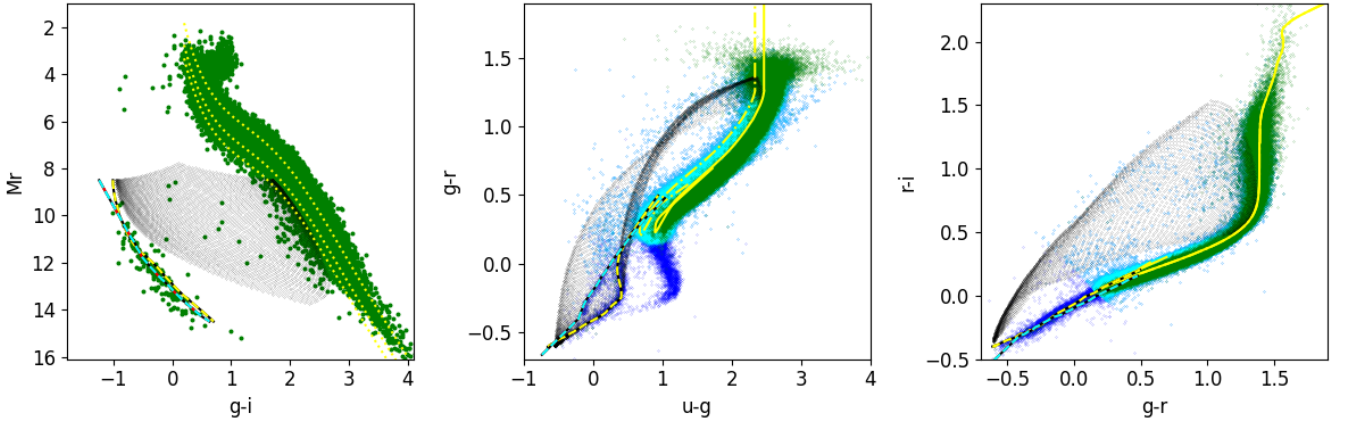

Figure 6. Luminosity-color tracks, shown with gray symbols, for unresolved binaries composed of an M dwarf with $[Fe/H] = -1.5$ and an H white dwarf. Analogous tracks for binaries composed of an M dwarf with $[Fe/H] = 0$ and either an H or He white dwarf look very similar though not identical. Data, shown as colored symbols, are the same as in Figure 1 and in the middle panel in Figure 2.

2.2.5. Luminosity-Color Tracks for Unaccounted Populations

There will always be sources that cannot be fully explained with any of the available families of luminosity-color tracks. For example, quasars and RR Lyrae are not considered here but can be easily recognized and removed from the sample in a straightforward manner since they are variable (e.g., see Sesar et al. 2007). The remaining non-variable sources that are not well fit with available models can be recognized as "bad fits" (see the last subsection on model selection below).

There will be impostors, too, with measured colors consistent with available model(s), but with a very different nature (e.g., ultra-cold white dwarfs "hiding" in the main stellar locus). As the time-domain information is built up with the progress of LSST, it might be possible to recognize them, for example, as high proper motion sources or perhaps using a broader wavelength range through cross-correlation with other surveys (e.g., WISE or Roman surveys). While they will require specialized studies, based on TRILEGAL simulations we expect that their fraction in the sample should be very small (most likely $< 1\%$).

2.2.6. Accounting for Interstellar Dust Extinction

Given a stellar population and resulting color tracks for intrinsic un-extincted colors, colors need to be corrected for interstellar dust extinction using an adequate dust extinction model. Additive color corrections $\vec{\delta c}(A_r)$ (see eq. 7) can be computed as

$$\delta c = (C_{m2} - C_{m1})\, A_r, \tag{8}$$

where $m1$ and $m2$ stand for two bandpasses that define the color (e.g., u and g). Dust extinction models, such as Cardelli et al. (1989) and Fitzpatrick (1999), parametrize extinction coefficients C_m as functions of R_V, where $R_V = A_V/E(B-V)$ and $E(B-V)$ is the stellar "color excess" (Cardelli et al. 1989).

Studies, such as Berry et al. (2012), based on SDSS data (including the Galactic plane), find that the scatter of R_V around its mean value $R_V = 3.1$ is very small. For this reason, we adopt empirical results from their Table 1: $C_m = (1.810, 1.400, 0.759, 0.561)$ in (u, g, i, z), respectively ($C_r = 1$ by definition). This choice does not preclude the use of our framework to fit for R_V as the fourth free model parameter in Galactic plane regions with very large A_r (R_V is poorly constrained when A_r is small); we can simply add multiple models with extincted colors generated using different values of C_m.

In our implementation, constraints on A_r for two nearby stars, even if they have similar distances, are independent. We note that one could use the so-called hierarchical Bayesian modeling and specify the prior for the line-of-sight extinction profile: nearby stars would then jointly constrain it. For more details, see Green et al. (2014).

Finally, it is noteworthy to point out that the dust extinction vector is nearly parallel to the main stellar locus, and this fact may cause model parameter degeneracies for certain choices of colors (e.g., see Figure 9 in Berry et al. 2012). Such degeneracies are broken when multiple colors that span a wide wavelength range and extend into near-IR (such as the $i - z$ color) are available. For more details, see Section 2.8 in Berry et al. (2012).

2.3. TRILEGAL-based Priors

In addition to specifying likelihood $p(\vec{c}|M_r, [Fe/H], A_r)$ for a given model M, we need to specify the prior probability distribution for model parameters, $p(M_r, [Fe/H], A_r)$ (see eq. 4; we address the model probability $p(M|I)$ further below). First, we assume that the stellar model parameters are unrelated to the distribution of interstellar dust along the line of sight,

$$p(M_r, [Fe/H], A_r) = p(M_r, [Fe/H])\, p(A_r). \tag{9}$$

We adopt a uniform prior for $p(A_r)$ using the values A_r^{SFD} taken from the Schlegel et al. (1998) dust extinction maps. To account for potential map errors, we set the maximum allowed value of A_r as

$$A_r^{max} = a\, A_r^{SFD} + b, \tag{10}$$

and $a = 1.3$, $b = 0.1$, where the choice of these parameters allows for plausible upper limits for additive and multiplicative errors in the dust extinction maps. Therefore, $p(A_r) = 1/A_r^{max}$ for $0 \leq A_r \leq A_r^{max}$ and $p(A_r) = 0$ for $A_r > A_r^{max}$.

The prior distribution of stellar model parameters, $p(M_r, [Fe/H])$, depends on sky position and apparent magnitude (here r) due to the complex structure of the Milky Way. For example, Green et al. (2014) generated priors using SDSS-based analytic descriptions of the 3-dimensional stellar distribution in the Milky Way (Jurić et al. 2008) and metallicity distribution for disk and halo components (Ivezić et al. 2008). Since LSST aims to provide good coverage of the Galactic plane and the Bulge, where SDSS data did not provide strong constraints, we opted to utilize recent simulations that implement constraints from a variety of modern surveys (some include the Galactic plane and Bulge regions). An analogous approach was taken by Bailer-Jones et al. (2021).

Dal Tio et al. (2022) have generated a mock catalog of the Milky Way stars to LSST depth (r=27.5) and over the entire LSST survey area. The simulation is based on the TRILEGAL code, incorporates all principal stellar populations, includes about 10 billion stars, and the catalog is publicly accessible through the NOIRLab Astro Data Lab[7]. We have developed code to query this mock LSST catalog using Astro Data Lab portal. For a given position on the sky, we extract all catalog entries from an area of $\sim$10 deg^2 (the size of the LSST Camera's field of view) around it, then bin the sample by apparent r-band magnitude (27 bin centers from $r = 14$ to $r = 27$, with a bin width of 1 mag). Given an r-band selected subsample, we separate all modeled populations and then bin them using adequate model
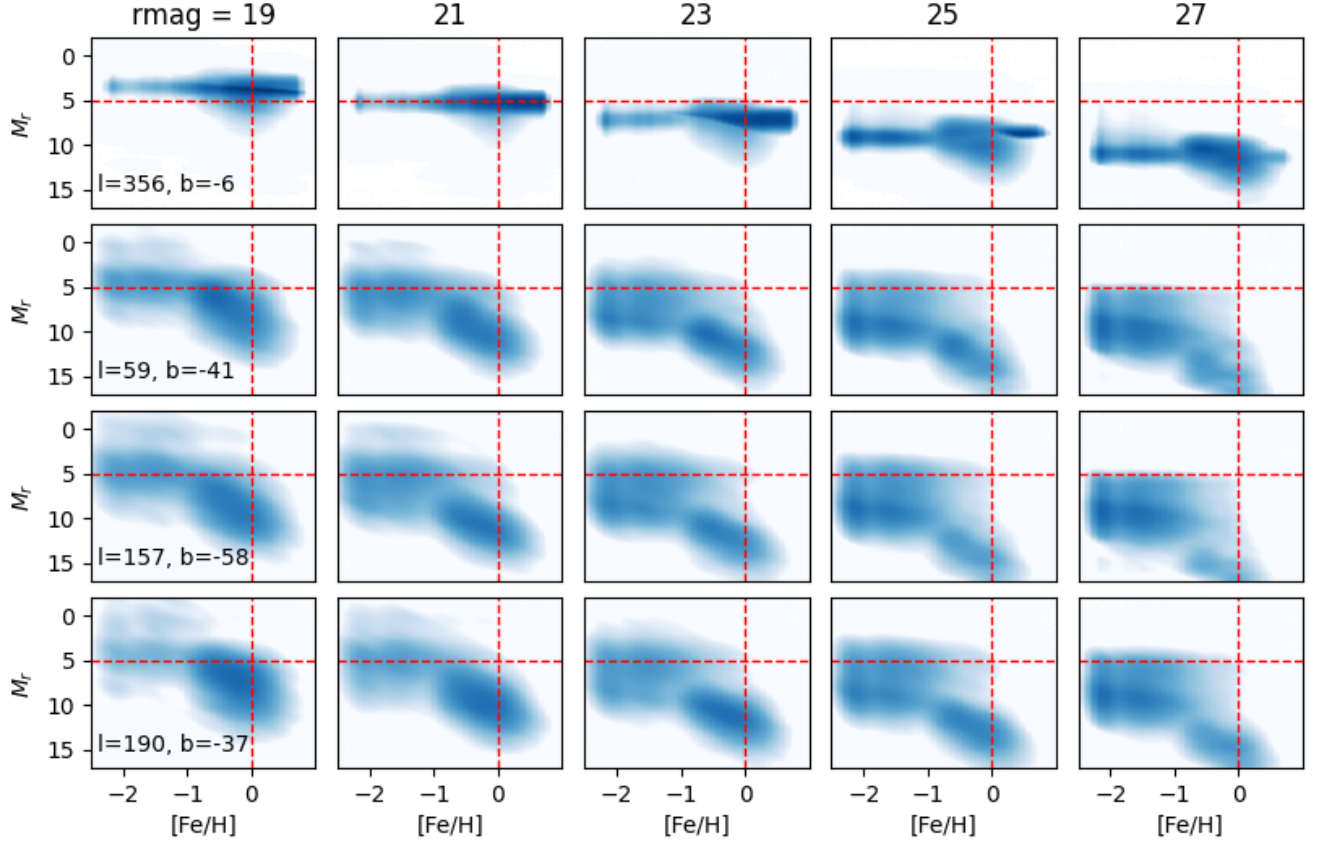
---

[7] Available at https://datalab.noirlab.edu/

Figure 7. Examples of priors for model parameters, $p(M_r, [Fe/H])$, for main-sequence stars and red giants, for four sky positions (rows, for locations see the first column) and for several brightness levels in the LSST magnitude range (left to right: bright to faint). The integral of each map over M_r and $[Fe/H]$ is unity. The dashed lines are the same in each panel and are added to guide the eye.

parameters (e.g., M_r and $[Fe/H]$ in the case of main-sequence stars and red giants). The resulting 2-dimensional map is renormalized so that its integral over all model parameters is unity.

Figure 7 shows examples of such $p(M_r, [Fe/H])$ maps for a grid of apparent magnitudes and for several characteristic sky positions. Several features are noticeable: i) there are two "clouds", at low and high metallicity, that correspond to halo and disk stars; ii) the distribution towards the Galactic center (top row) is much more compact in the M_r direction (because these subsamples are dominated by the Bulge stars at similar distances); iii) the whole distribution shifts to fainter levels as the bin center (r) becomes fainter. We have experimented with different region sizes[8] and found that $\sim$10 deg^2 (HEALPix N_{side}=16 and 32, with 3,072 and 12,288 pixels over the full sky) is sufficiently small to capture the variation of $p(M_r, [Fe/H])$ maps across the sky.

2.4. An Example of Bayesian Model Parameter Estimation

We use isochrones shown in the right panels of Figure 3 and priors shown in the second row of Figure 7 to illustrate the Bayesian model parameter estimation. Figure 8 shows prior, likelihood and posterior for a main-sequence star close to the turn-off point. Note how the posterior, although much wider than the likelihood map, helps break the degeneracy[9] in the likelihood map (the two "islands").

Figure 9 shows two-parameter covariances and marginal distributions for this case. Note that true values are recovered within expected uncertainties, as well as non-vanishing covariances between the parameters. The marginal

---

[8] Following Bailer-Jones et al. (2021), we use HEALPix geometry, see https://healpix.jpl.nasa.gov.

[9] Priors can break degeneracies between the giant and dwarf stars because luminous stars become strongly disfavored at faint magnitudes (because an apparently faint giant star would imply a very large distance, beyond the presumed edge of the Galaxy at $\sim$100 kpc; for example a giant star with M_r=0 and r=22 would imply a distance of $\sim$250 kpc).
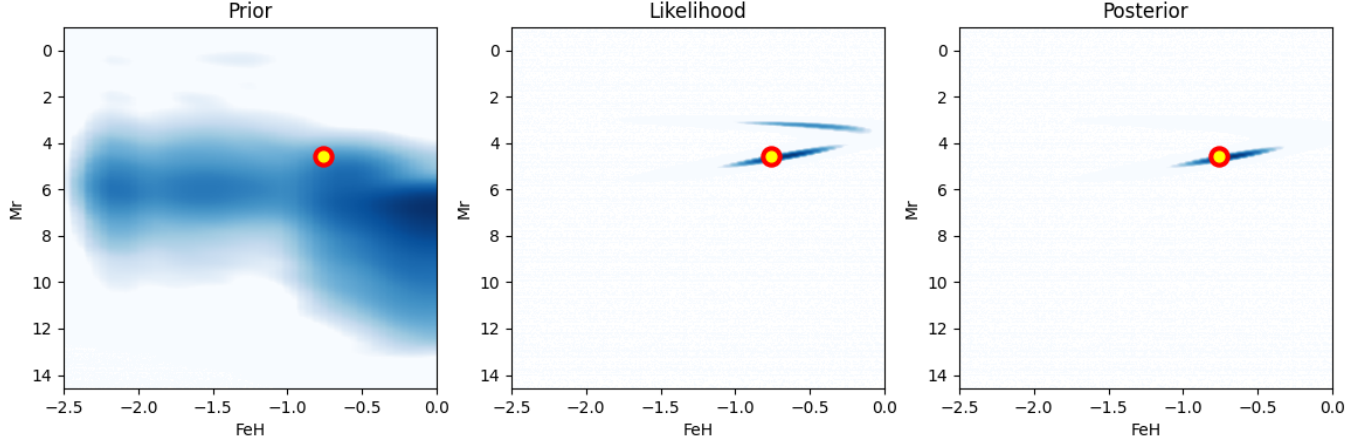
Figure 8. Maps of the prior (left), likelihood (middle; computed using eq. 6) and posterior (right; computed using eq. 3) for a simulated main-sequence star with $r = 21.2$, $(g - r) = 0.48$, $M_r = 4.58$, $[Fe/H] = -0.76$ and $A_r = 0.37$. The prior in the A_r direction is uniform. The circles mark the values of input M_r and $[Fe/H]$. The posterior marginal distributions for all three model parameters are shown in Figure 9.

distributions produced with the prior, likelihood and posterior maps, shown in Figure 9, illustrate the improvement in the "knowledge" of model parameters between prior and posterior, brought by color measurements via the likelihood map.

In the next Section, we discuss a fast numerical pipeline implementation that can perform this computation for LSST-sized catalogs and provide a more quantitative analysis of the method's performance. We conclude this Section with a brief discussion of the Bayesian model selection method for assigning posterior probabilities to each stellar population used to interpret observations.

2.5. *Bayesian Model Selection*

For each star and for each stellar population (that is, a model for color-magnitude tracks), a posterior 3-dimensional data cube is produced. These posteriors are used when choosing the best model, as follows.

Bayes' theorem as introduced by eq. 3 quantifies the posterior pdf of parameters describing a single model, with that model *assumed to be true*. To find out which of two models, say M_1 and M_2, is better supported by data, we compare their posterior probabilities via the *odds ratio* in favor of model M_2 over model M_1 as

$$O_{21} \equiv \frac{p(M_2|D, I)}{p(M_1|D, I)}. \tag{11}$$

The posterior probability for model M (M_1 or M_2) given data D, $p(M|D, I)$ in this expression, can be obtained from the posterior pdf $p(M, \vec{\theta}|D, I)$ in eq. 3 using marginalization (integration) over the model parameter space spanned by $\vec{\theta}$. The posterior probability that the model M is correct given data D (a number between 0 and 1) can be derived using eqs. 3 and 4 as

$$p(M|D, I) = \frac{p(D|M, I) \, p(M|I)}{p(D|I)}, \tag{12}$$

where

$$E(M) \equiv p(D|M, I) = \int p(D|M, \vec{\theta}, I) \, p(\vec{\theta}|M, I) \, d\vec{\theta} \tag{13}$$

is called the *marginal likelihood* (or *the evidence*) for model M and it quantifies the probability that the data D would be observed *if* the model M were the correct model. Since the marginal likelihood $E(M)$ involves integration of the data likelihood $p(D|M, \vec{\theta}, I)$, it is also called the *global likelihood* for model M.

The global likelihood is a *weighted average* of the likelihood function, with the prior for model parameters acting as the weighting function. Alternatively, $E(M)$ is simply the integral over allowed parameter space of the posterior pdf before its renormalization to set this integral to unity (e.g., the integral of the posterior shown in the right panel in Figure 8). If the chosen model color tracks cannot explain the observed colors, the likelihood (eq. 6) will never be very
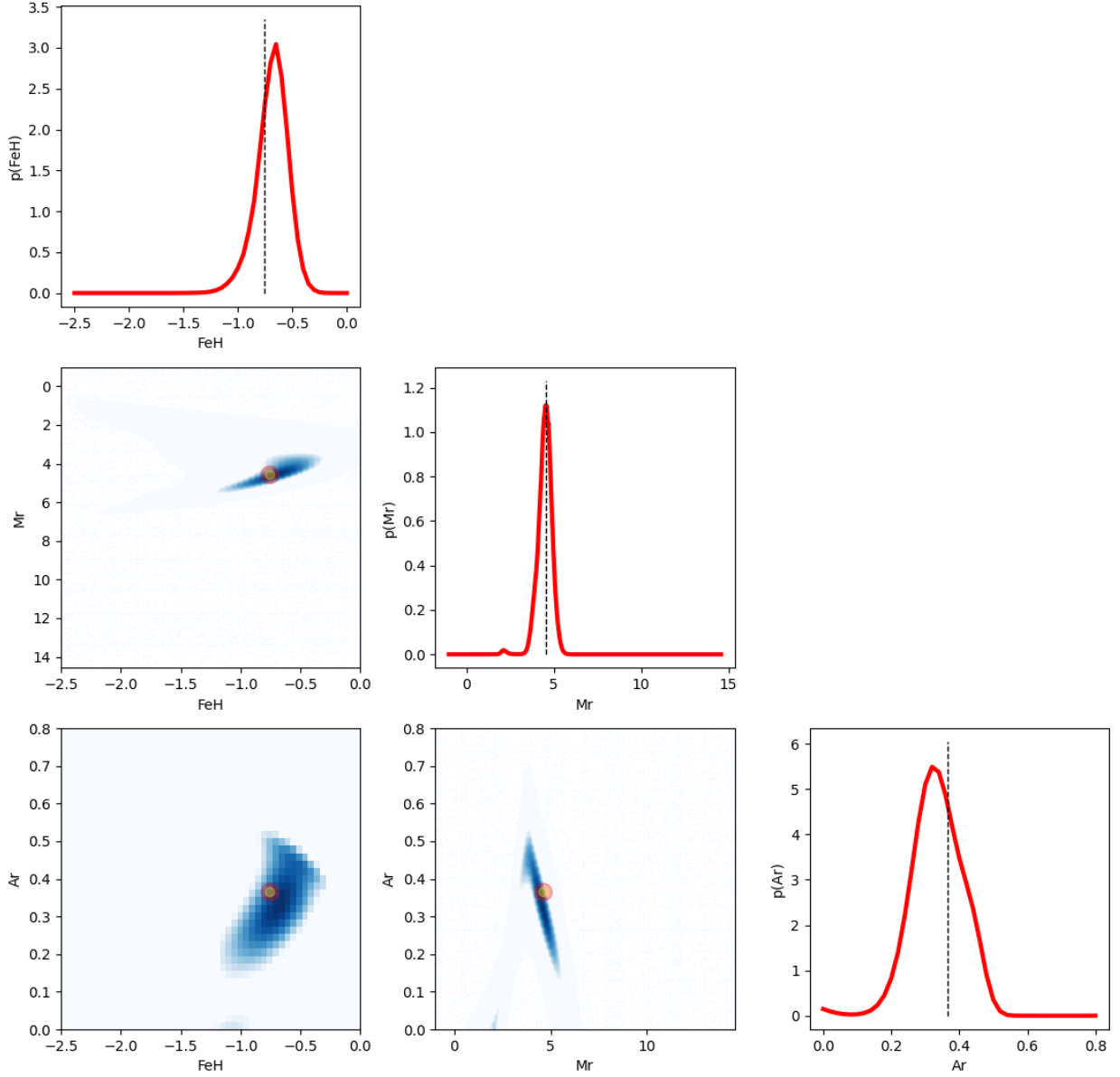
Figure 9. Two-parameter covariances and marginal distributions for the posterior map from Figure 8. The symbols and dashed lines show true input values.

high and the resulting $E(M)$ will be low. We note that in the limit of Gaussian posterior and flat priors, Bayesian evidence-based model selection becomes equivalent to χ^2 selection from the frequentist statistical framework.

The probability of data, $p(D|I)$, cancels out when the odds ratio is considered:

$$O_{21} = \frac{E(M_2)\, p(M_2|I)}{E(M_1)\, p(M_1|I)} = B_{21}\, \frac{p(M_2|I)}{p(M_1|I)}. \tag{14}$$

In practice, the values of the odds ratio are interpreted using Jeffreys' scale (see, e.g., Chapter 5 in Ivezić et al. 2020); in particular, $O_{21} > 10$ represents "strong" evidence in favor of M_2 (M_2 is ten times more probable than M_1).

The ratio of global likelihoods, $B_{21} \equiv E(M_2)/E(M_1)$, is called the *Bayes factor*, and is equal to

$$B_{21} = \frac{\int p(D|M_2, \vec{\theta}_2, I)\, p(\vec{\theta}_2|M_2, I)\, d\vec{\theta}_2}{\int p(D|M_1, \vec{\theta}_1, I)\, p(\vec{\theta}_1|M_1, I)\, d\vec{\theta}_1}. \tag{15}$$
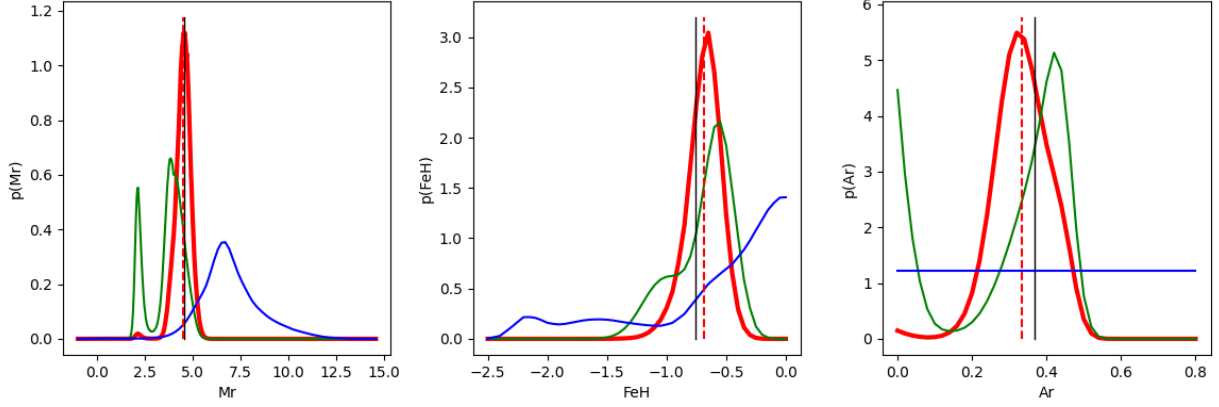
Figure 10. An illustration of the improvement in the "knowledge" of model parameters between prior and posterior, brought by measurements via the likelihood map. The lines show marginal distributions produced with the prior (blue), likelihood (green) and posterior (red) maps from Figure 8. The dashed line shows the expectation value for the marginalized posterior and the vertical solid line marks the true input value.

The vectors of parameters, $\vec{\theta_1}$ and $\vec{\theta_2}$, are explicitly indexed to emphasize that the two models may span vastly different parameter spaces (including the number of parameters per model).

The prior model probabilities, $p(M_1|I)$ and $p(M_2|I)$, are determined using TRILEGAL simulated catalog. For example, if model 1 is main-sequence stars and model 2 is white dwarfs, we estimate the $p(M_2|I)/p(M_1|I)$ ratio by simply counting main-sequence stars and white dwarfs in the corresponding r magnitude bin.

In case of N_M models, we assume that they represent an exhaustive model set and estimate the posterior probability of model k, $k = 1...N_M$, as

$$p(k) = \frac{O_{k1}}{\sum_{j=1}^{j=N_M} O_{kj}}, \tag{16}$$

where the model $k = 1$ is chosen arbitrarily without a loss of generality ($O_{11} = 1$).

3. NUMERICAL IMPLEMENTATION AND PERFORMANCE TESTS

We first discuss our numerical implementation choices and then test the method's performance using simulated TRILEGAL catalogs. We validate adopted luminosity-color models and implied distance scale using SDSS and Gaia observations. An all-sky pipeline implementation and its performance are discussed in the next Section.

3.1. *Accelerated Exhaustive Grid Search Method*

Luminosity-color models introduced in §2 are expressed as lookup tables, rather than as analytic functions. They are initially defined on a two-dimensional grid, and after accounting for dust extinction, on a three-dimensional grid. We use equidistant grids for all three model parameters: 0.01 mag for M_r, 0.05 dex for $[Fe/H]$ and a range of steps from 0.005 mag to 0.02 mag for A_r, depending on its prior range. These grids have sufficient resolution for numerical analysis of the posterior probability distributions and typically result (depending on the maximum allowed A_r value) in several million grid points. For a given star, the exhaustive grid search method simply iterates over all these points and for each computes likelihood and ultimately the posterior pdf.

Significant acceleration can be achieved by using a two-step iteration, with degraded grids in the first step, but still fine enough to sufficiently resolve the posterior pdf to approximately estimate its behavior around its maximum. We find that degrading the M_r step to 0.05 mag and the $[Fe/H]$ step to 0.1 dex works well in practice and results in a about ten times shorter run time. After the first iteration, the range $\pm 3\sigma$ around the posterior maximum, where σ is standard deviation, is estimated using the marginal distribution for each model parameter. The calculation of the posterior is then repeated using the finest grid but only in the $\pm 3\sigma$ neighborhood (cube) around the posterior maximum, which includes 100 to 1000 times fewer grid locations.

With this acceleration trick, runtime per star is typically about 10 millisec per star (it varies nearly proportionally to the maximum allowed A_r value). While exceedingly simple, this method is sufficiently fast and robust to offer a convenient practical solution for processing samples of billions of stars.

3.1.1. *Markov Chain Monte Carlo Method*

Markov Chain Monte Carlo methods have been successfully used in this context (e.g., Green et al. 2014). Given that we aim to process LSST photometry for about 10^{10} stars, it is worthwhile to compare the MCMC runtime to the runtime for our accelerated exhaustive grid search method. We considered two implementations, based on *PyMC3* (Oriol et al. 2023) and *emcee* (Foreman-Mackey et al. 2013). The latter was about 10-20 times faster than the former; for example, with 4 "walkers" and 500 iterations, it took on average about 0.1 sec per star. Therefore, in this low-dimensional case (there are only three model parameters) the grid search method described above is about an order of magnitude faster than the *emcee* implementation.

3.1.2. *Neural Network Method*

Further acceleration can be achieved using the Neural Network method. In a companion paper by Mrakovčić et al., we describe how to use a small subsample of stars to train a neural network model and then estimate model parameters and their uncertainty for the remaining stars using the trained network. This approach yields runtimes of about 1 milisec per star, that is, shorter by an order of magnitude than when the grid search method is used for the full sample. With this additional speed up, a sample of 10 billion stars can be processed in about 10 hours using a 300-core machine and the pipeline described in the next Section.

3.2. *Performance Testing*

We first validate numerical implementation of our method using a simulated dateset, where the true model parameters are known. The luminosity-color model tracks are then validated using SDSS photometry from the so-called Stripe 82 region and trigonometric distances obtained using Gaia data. Finally, we demonstrate performance in the large A_r regime and three-dimensional mapping of interstellar dust distribution using SDSS scans that cross high-extinction regions in the Galactic plane.

3.2.1. *Simulated Data Set*

Our simulated catalogs are based on TRILEGAL simulations of LSST stellar content by Dal Tio et al. (2022), already mentioned in the context of priors in §2.3. We use these simulations to generate distributions in the M_r–$[Fe/H]$–A_r–magnitude–sky position space. Given these quantities for a sample of simulated TRILEGAL stars, we generate simulated LSST photometry for each star as follows:

1. Given relevant model parameters (e.g., M_r and $[Fe/H]$ for main sequence and red giant stars, M_r and $\log(g)$ for white dwarfs, the total system luminosity and the component luminosity ratio in the r band for unresolved binaries), we use luminosity-color tracks (see Figure 3) to assign photometric noise-free colors. We do not use the original TRILEGAL colors because for proper statistical tests colors must be consistent with luminosity-color tracks used for likelihood computations.

2. Using the apparent r band magnitude from the TRILEGAL simulation, corresponding to simulated distance and M_r, and colors from the previous step, we generate magnitudes in all the remaining bands.

3. We generate per-band photometric uncertainties, σ_b, using eqs. 4 and 5 from Ivezić et al. (2019) and per-band 5-σ depths (the so-called m_5) expected for co-added LSST data from Bianco et al. (2022).

4. We draw Gaussian noise from N(0, σ_b) and add it to each magnitude to obtain "observed" magnitudes.

5. Finally, we recompute photometric uncertainties to be treated as "observational" uncertainties using "observed" magnitudes (because the first set of photometric uncertainties are derived using true noise-free magnitudes).

For numerical tests, we select a region from the so-called SDSS Stripe 82, defined by $340° <$ R.A. $< 350°$ and $-1.3° < \delta < 1.3°$, which contains 280,000 simulated main sequence and red giants stars that satisfy $r < 26$ and $u < 27$. Their distributions of model parameters M_r, $[Fe/H]$, and interstellar dust extinction along the line of sight in the r band, A_r, in the u vs. $g - i$ color-magnitude diagram are shown in Figure 11.
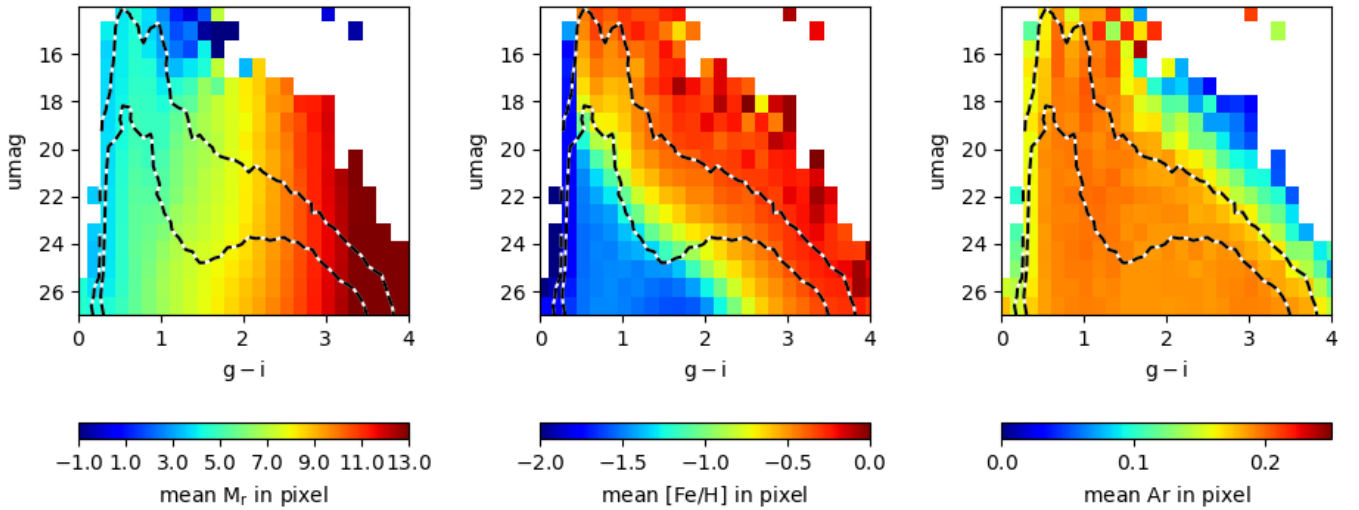
Figure 11. The mean per-pixel values of absolute magnitude M_r, metallicity $[Fe/H]$, and interstellar dust extinction along the line of sight in the r band, A_r, for a TRILEGAL-simulated sample of 280,000 main sequence and red giants stars with $r < 26$, $u < 27$, $340° <$ R.A. $< 350°$ and $-1.3° < \delta < 1.3°$ (a small patch from the SDSS Stripe 82 region) in the u vs. $g - i$ color-magnitude diagram. The mean values are color-coded according to the legend below each panel. Contours visualize the sample distribution in each diagram. The strong variation of M_r with the $g - i$ color is seen in the left panel, except for the small patch with $u < 16.5$ and $1.0 < g - i < 1.5$ where red giant stars dominate. The diagonal iso-metallicity boundaries in the middle panel closely correspond to distance (in the range from about 1 kpc to about 100 kpc in the lower left corner). The values of dust extinction, shown in the right panel, are not large (< 0.2) because the selected field is at high Galactic latitudes (centered on $b = -52°$).

Figures 12 and 13 show the summarized statistical performance of the Bayesian method (bias, scatter, χ^2) in the u vs. $g - i$ color-magnitude diagram. In this test based on a high-galactic latitude field, we assumed that all stars are beyond the dust layer and fixed A_r to its known true value (in practice provided by dust maps such as those from Schlegel et al. 1998). When A_r is considered as a free parameter in sky regions with small A_r, the best-fit A_r values are often underestimated for stars in the blue part of the stellar locus (because the reddening vector and locus are nearly parallel; for more details, see §2.7.1 in Berry et al. 2012). We note that at faint magnitudes probed by LSST, this assumption will only break down very close to the Galactic plane.
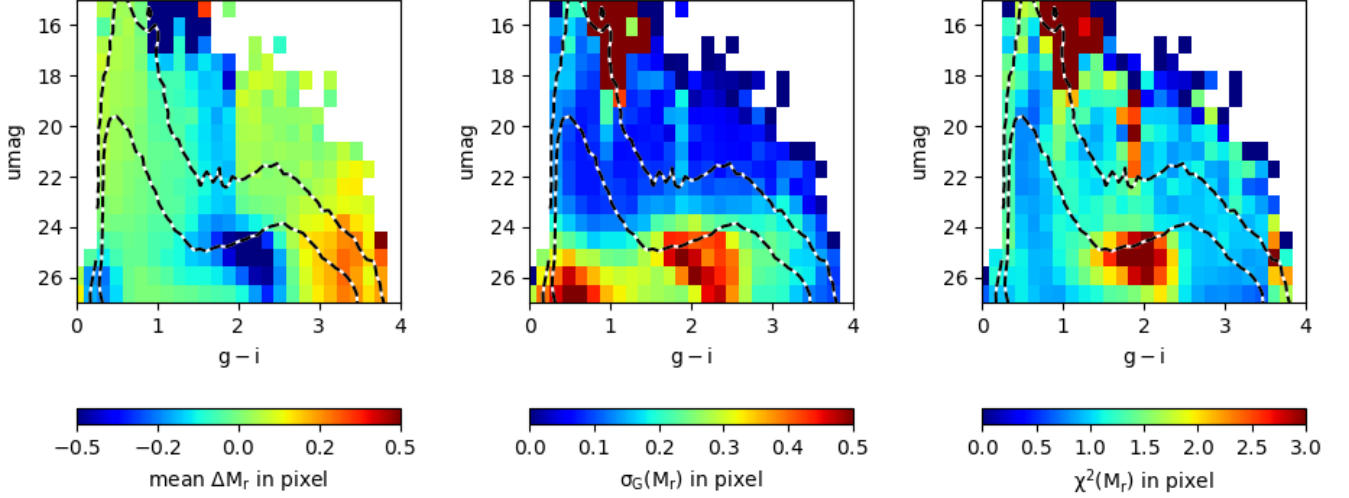
Figure 12. Statistical performance analysis for the estimates of model parameter M_r (absolute magnitude), as a function of observed u band magnitude and the $g - i$ color, for the same simulated sample as shown in Figure 11. The left column shows the mean difference per pixel between the true and estimated values, the middle column shows scatter per pixel and the right column shows the scatter normalized by estimated uncertainties (χ^2). Contours visualize the sample distribution in each diagram. Note that for main sequence stars $M_r > 4$ and for most stars $4 < M_r < 10$.
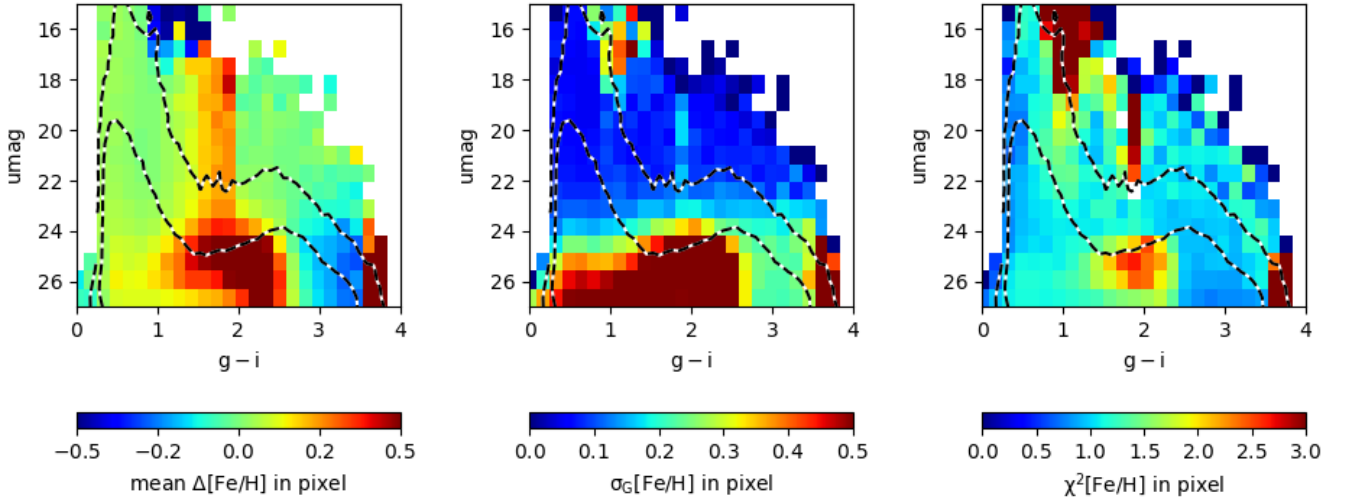
Figure 13. Analogous to Figure 12, except for estimates of $[Fe/H]$ (metallicity) instead of M_r.

As discussed in detail in Ivezić et al. (2008), the photometric metallicity is best constrained for blue main-sequence stars with $g - i < 1$ (equivalently, $0.2 < g - r < 0.6$). Therefore, we expect the best statistical behavior (smallest bias and scatter, $\chi^2 \sim 1$) in this color range, which is consistent with the behavior seen in Figures 12 and 13. An effective depth limit is about $u = 25$ (about 1 mag brighter than the 5-σ depth in the u band for coadded LSST photometry, Bianco et al. 2022); at fainter magnitudes the scatter between true and estimated values for both M_r and $[Fe/H]$ rapidly increases.

Figure 14 shows a more quantitative illustration of the bias and scatter for best-fit M_r and $[Fe/H]$ as functions of the u band magnitude, which sets the effective signal-to-noise ratio. In the bright limit ($u < 22$; approximately $r < 21$ for blue stars), M_r and $[Fe/H]$ can be estimated with uncertainties of 0.10 mag and 0.07 dex, respectively, and essentially negligible biases (0.02 mag for M_r and 0.03 dex for $[Fe/H]$). These uncertainties are consistent with simulated photometric errors and intrinsic properties of photometric parallax and photometric metallicity methods (for detailed discussion, see Ivezić et al. 2008).

This M_r uncertainty implies distance estimates for blue stars accurate to about 5% to a distance limit of 25 kpc, with coadded LSST photometry. Such unprecedented accuracy for photometric distance estimates is due to the ability of the u band to constrain $[Fe/H]$. For a subsample with $u \sim 24.73$, one magnitude brighter than 5-σ limit, the scatter increases to 0.27 mag and 0.38 dex, and further to 0.36 mag and 0.54 dex for a subsample at 5-σ limit in the u band (where the u band photometric uncertainty is 0.2 mag).

As the right panel in Figure 14 demonstrates, estimates of M_r and $[Fe/H]$ are highly anti-correlated (this is essentially a consequence of eq. A2 from Ivezić et al. 2008, also visible as the shift of stellar main sequence with metallicity in the top panels in Figure 3). An $[Fe/H]$ uncertainty of 0.3 dex, approximately the width of the halo metallicity distribution, would induce an additional uncertainty of estimated photometric M_r of 0.3 mag (as would be the case for faint stars with weak or no $[Fe/H]$ constraint coming from their u band measurement).

The bias and scatter for both M_r and $[Fe/H]$ are large in a small patch with $u < 16.5$ and $1.0 < g - i < 1.5$ where red giant stars dominate, due to strong degeneracies in color space between giants and main sequence stars (recall Figure 4). About 2/3 of all red giants are mis-identified as main sequence stars (due to priors, as they are by and large indistinguishable by colors). Nevertheless, these simulation-based tests suggest that it will be possible to select highly pure samples of red giants using best-fit M_r and a simple $M_r < 3$ cut. This selection criterion selects 28% of all red giants in the sample (which correspond to about 1% of the full sample), with a purity of 99.9%. With LSST data, it will be possible to select such red giant candidates, if they exist, to distances of several hundred kpc (note that low-$[Fe/H]$ main sequence stars at the blue edge of the stellar locus and $u = 25$, with LSST distance uncertainties of about 20%, will be detected to distances of about 100 kpc).
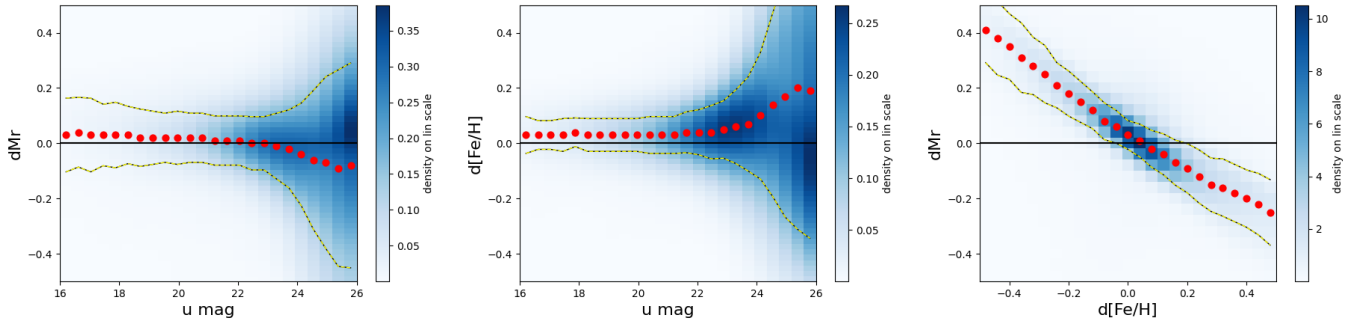

Figure 14. The variation of the difference between the true and estimated values (left: M_r, middle: $[Fe/H]$) with u band magnitude. The binned background map shows the same simulated sample as in Figure 11. The symbols show binned medians and the dashed lines show robust standard deviation around the medians. At about $u = 23$, the scatter for both M_r and $[Fe/H]$ starts increasing due to increasing u band measurement error. The dataset was generated assuming u band 5-σ limiting depth of 25.73 for coadded LSST photometry (Bianco et al. 2022). The right panel illustrates strong anti-correlation between the differences.

3.2.2. *SDSS-based Comparison to Gaia's Distance Scale*

Analysis of the method's performance in the preceding subsection was based on a simulated sample with "observed" colors generated using exactly the same luminosity-color model tracks as those used in fitting. Therefore, that analysis cannot test the validity of those tracks in an absolute sense. While these empirical luminosity-color model tracks were derived using SDSS observations of globular clusters with known distances and metallicities, here we validate them further using trigonometric distances recently obtained by Gaia.

For validation, we used the SDSS Stripe 82 Standard Star Catalog, recalibrated by Thanjavur et al. (2021), which provides the most precise available photometry in SDSS bands (due to averaging of many repeated SDSS observations). For stars listed in the catalog, we extracted Gaia measurements and "photogeometric" distances from Bailer-Jones et al. (2021). Out of 841,000 stars listed in the Standard Star Catalog, there are 415,000 stars with $r < 22$, $u < 22$ and a Gaia match within 0.15 arcsec (after correcting for proper motion using Gaia measurements). Their distribution in color-color and color-magnitude diagrams is shown in Figures 1 and 2. We estimated M_r and $[Fe/H]$ for these stars using the same procedure and assumptions as for the simulated sample from the preceding subsection. In particular,
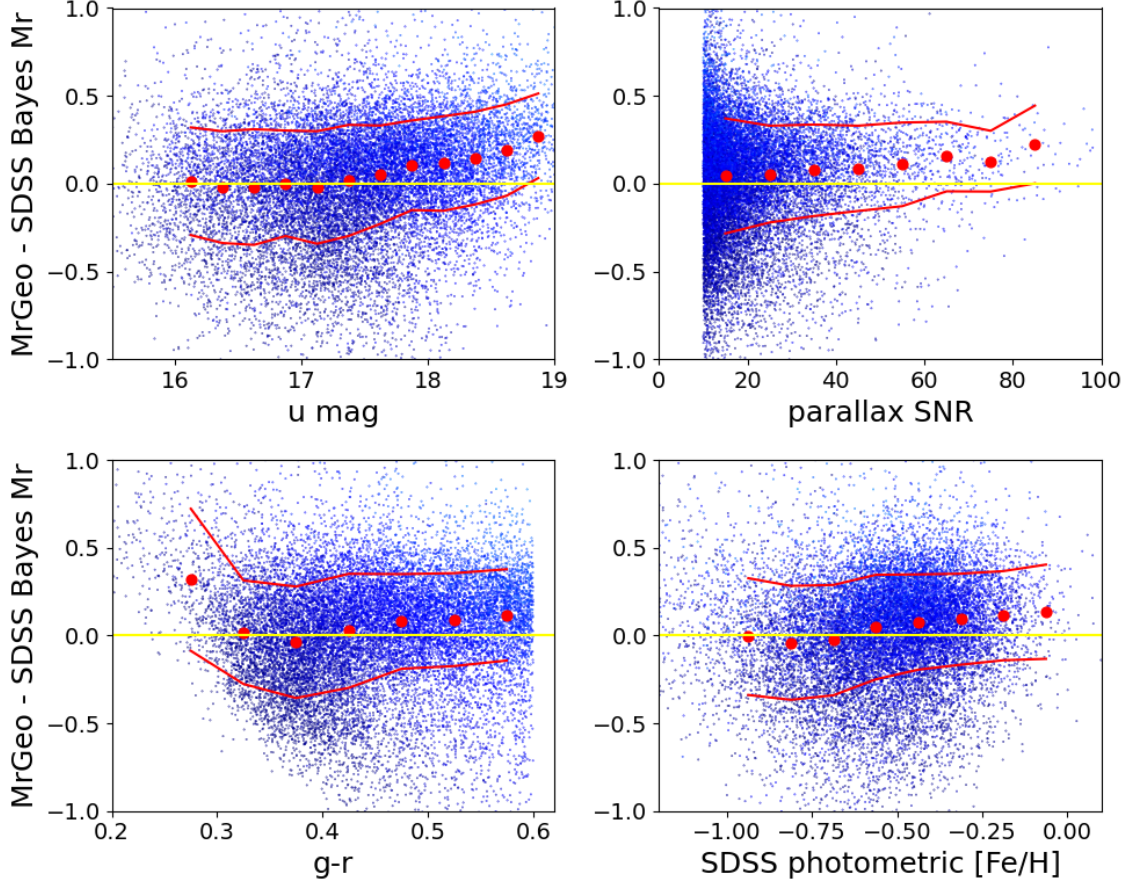
Figure 15. Analysis of the difference between Gaia's geometric (trigonometric) absolute magnitudes and photometric magnitudes derived from SDSS colors. Symbols show a sample of 33,186 stars with the parallax signal-to-noise ratio above 10, absolute magnitude MrGeo>4 and MrPho>4 , and SDSS-based $u < 21$ and $0.2 < g-r < 0.6$. The median magnitude difference is -0.01 mag, with a scatter of 0.28 mag. The four panel illustrate variation of the binned median (red circles) and scatter (red lines) with the u-band magnitude, the parallax signal-to-noise ratio, SDSS $g-r$ color, and photometric metallicity.

we assumed that all stars are beyond the dust layer and fixed A_r to its value provided by the dust maps from Schlegel et al. (1998).

We first analyze results for a subsample of stars that have $0.2 < g - r < 0.6$, a color range where photometric metallicity estimator is best constrained, and a signal-to-noise ratio for Gaia's trigonometric parallax measurement above 10. We also required $u < 21$ but that requirement is less stringent than the implied flux limit due to Gaia's signal-to-noise ratio limit (approximately, $u < 19$). There are about 18,000 stars in this subsample. We find that their M_r is estimated using SDSS photometry with a bias of 0.06 mag and root-mean-square scatter of 0.30 mag, compared to Gaia's measurements. Equivalently, photometric distances are estimated with a scatter of 15% and a bias of 3%.

The median contribution of Gaia's parallax measurement uncertainty to Gaia's absolute magnitude uncertainty for this subsample is about 0.12 mag. This estimate implies that the median uncertainty of SDSS-based M_r estimates is 0.27 mag (corresponding to about 14% distance uncertainty). On the other hand, when only stars with signal-to-noise ratio for Gaia's trigonometric parallax measurement above 50 are considered, the implied uncertainty of SDSS-based M_r estimates is 0.21 mag (corresponding to about 10% distance uncertainty). Recalling that the statistical uncertainties for M_r estimates using simulated sample in the preceding section are 0.10 mag at the bright end, these somewhat larger uncertainties may imply additional statistical effects not accounted for in our analysis, and/or imperfect luminosity color-tracks. Figure 15 demonstrates that there are no concerningly large systematic errors in photometric M_r estimates with respect to photometric noise (u band magnitude), $g - r$ color and photometric metallicity estimates (though note that the metallicity range corresponds only to disk stars).
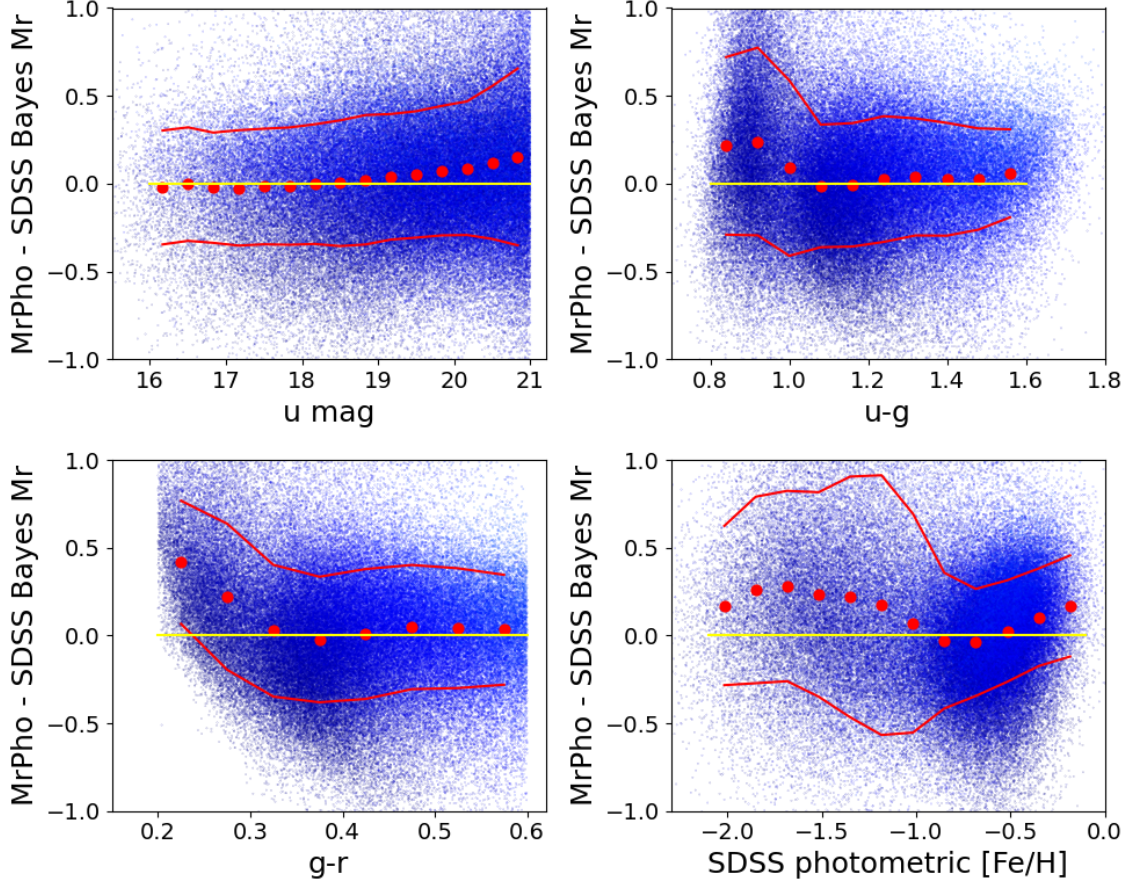
Figure 16. Analysis of the difference between Gaia's photometric (color-based) absolute magnitudes and photometric magnitudes derived from SDSS colors. Symbols show a sample of 165,834 stars with absolute magnitude MrPho>4, and SDSS-based $u < 21$ and $0.2 < g - r < 0.6$. The median magnitude difference is -0.05 mag, with a scatter of 0.32 mag. The four panel illustrate variation of the binned median (red circles) and scatter (red lines) with the u-band magnitude, SDSS $u - g$ and $g - r$ colors, and photometric metallicity.

We extended our analysis to fainter magnitude limits by replacing Gaia's magnitudes based on trigonometric distance with those based on the so-called "photogeometric" distances from Bailer-Jones et al. (2021). This comparison goes about 2 magnitudes deeper and also extends to low halo-like metallicity range. The scatter between photometric M_r based on SDSS data and M_r based on "photogeometric" distances is 0.38 mag. As shown in Figure 16, the systematic error increases to about 0.2-0.3 at the blue and low-metallicity edge of the stellar locus. When stars with $0.2 < g - r < 0.3$ are excluded, systematic errors at the low-metallicity ($[Fe/H] < -1$) end disappear[10], indicating that it is the M_r as a function of color relation that fails at the blue end, rather than the shift of M_r as a function of $[Fe/H]$. Such edge effects will have to be recalibrated when LSST photometry becomes available (both using Gaia data and globular clusters).

3.2.3. *Performance in the Galactic Plane*

As the final test, we analyze the method's performance in regions with large interstellar dust extinction (A_r). We use SDSS-SEGUE data collated and described by Berry et al. (2012). For illustration, we use a single strip perpendicular to the Galactic plane and centered on longitude $l = 110^\circ$. Figure 17 shows best-fit dust extinction along the line of

---

[10] The so-called "photogeometric" distances from Bailer-Jones et al. (2021) do not utilize [Fe/H] information. Since the [Fe/H] distribution varies with apparent magnitude, one might expect biased distances (20% or more for halo vs. disk comparison). We find no significant bias compared to our estimates, which probably indicates that their distance priors absorb the impact of unknown [Fe/H]. Hence, the only net effect of unknown [Fe/H] is increased uncertainty of estimated distances (for halo stars, with a metallicity scatter of 0.3 dex, the expected contribution to M_r scatter is about 0.2 mag).
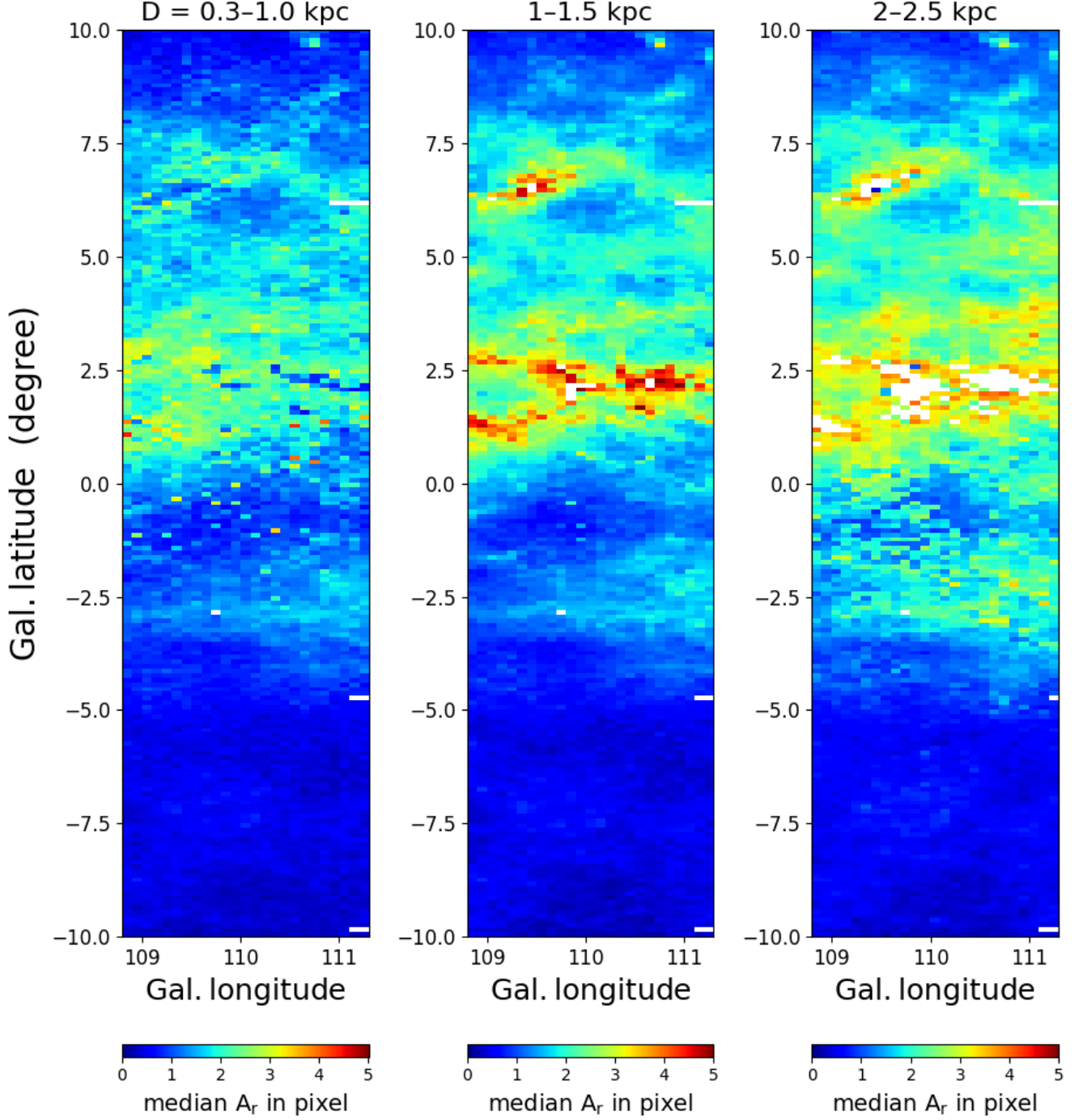
Figure 17. Color-coded maps show the best-fit A_r (dust extinction along the line of sight) per pixel for the SDSS-SEGUE strip centered on $l = 110°$ (modeled after Figure 25 from Berry et al. 2012). Each pixel, defined in Galactic coordinates, shows the median Ar according to the legend below each panel. The three panels correspond to distance slices 0.3—1 kpc (left), 1-–1.5 kpc (middle), and 2—2.5 kpc (right). Note that the median extinction increases rapidly with distance. In the right panel, there are no stars in white regions in the center because extinction makes them fainter than the sample magnitude limit.

sight for three distance slices. Clear three-dimensional structure of the dust distribution is evident and consistent with Figure 25 from Berry et al. (2012).

These tests validate numerical implementation of the Bayesian algorithm and demonstrate its performance on a single patch of data, where priors can be assumed uniform. An all-sky pipeline implementation is described in the next Section.

4. ALL-SKY PHOTOD PIPELINE IMPLEMENTATION

Implementation of the Bayesian algorithm described in preceding sections assumes that priors can be assumed constant over some sky region, hereafter a patch (see §2.3). In practice, patches need to be at least several sq. deg. large in order to have enough stars to compute priors. At the same time, they should not be too large to satisfy the assumption of constancy. Both of these size limits vary across the sky due to Galactic structure effects.

The scale of the dataset expected from LSST – O(10Bn) objects – makes this computation challenging to perform on a single core or computer. Ideally, one would wish to execute the computation in parallel on many (hundreds) of cores, likely housed on a number of separate nodes. This requires a distributed processing framework capable of orchestrating such a solution.

4.1. *LSDB and Healpix Implementation*

For our work, we've adopted the *Large Survey Database* (LSDB; Wyatt et al. (2024)) framework. LSDB is a distributed computing framework written in Python, which enables multi-node and multi-core computation on large (PB-scale), partitioned, catalog datasets. It extends the concepts of the LSD toolkit (Juric 2011) to astronomical catalogs where the distribution of objects on the sky is very uneven – such as is the case for Galactic star counts. Leveraging broadly adopted community libraries such as astropy, Pandas, and Dask, LSDB presents a user-friendly API on which it is possible to build our pipeline.

Specifically, LSDB allows us to define a piece of computation (the PhotoD function, written in Python) that should be executed either on every object in the catalog, or on a per-patch basis (where the patch has a potentially dynamically sized solid angle). For example, we compute priors for each of the patches (§2.3). LSDB then ensures that these functions are successfully run over the entire (arbitrarily sized) catalog. LSDB handles the data distribution and movement between multiple nodes in the cluster, automatically handles transient failures, and writes the output (in parallel) to another partitioned, searchable, table.

This workflow is particularly efficient and well-suited for the scientific objectives addressed by *PhotoD*, as it leverages the variable resolution of the map based on the star density within each region. For instance, in the galactic plane where star density is higher, smaller pixels provide higher resolution, while in the halo, where star density is lower, larger pixels suffice.

This approach is also computationally advantageous because it maintains consistent chunk sizes for processing. Each data chunk, regardless of its spatial resolution, is approximately the same size, optimizing the execution of the code. In our specific case, the data chunks correspond to pixels that are a few hundred megabytes in size.

4.2. *Pipeline Performance*

We have performed several benchmarks to estimate the performance of the PhotoD code. In our initial tests based on the SDSS Stripe 82 data we found that the distances for a billion stars can be estimated in about 100 days on a Mac M1 Pro laptop using one core (10 milliseconds per star).

In further testing, we used outputs from the TRILEGAL simulation and found that the parallel processing performed with LSDB scaled well with respect to the single core values. On a local cluster node with dual AMD EPYC 9474F 48-core processors we were able to achieve per-star figures similar to single-core processing, therefore reducing the required processing time to about a week. Our ongoing tests are showing that when all 10 nodes (400 cores) of the Ruđer Bošković Institute's Narval cluster are used, we may be able to estimate distances for about 10 billion stars in about two days.

The output catalog containing LSST photometry (with errors) and minimalistic auxiliary metadata for 10 billion stars are would amount to about 1 TB of data. The code outputs, including a covariance matrix of the three fitted parameters, would have approximately the same volume.

5. DISCUSSION AND CONCLUSIONS

Distances to stars are a crucial ingredient in our quest to better understand the formation and evolution of the Milky Way galaxy. Anticipating photometric catalogs with tens of billions of stars from Rubin's LSST, here we presented PhotoD, a pipeline for computing Bayesian distance estimates that can handle LSST-sized datasets.

The Bayesian procedure implemented in the PhotoD pipeline builds on previous work (e.g., Bailer-Jones 2011; Berry et al. 2012; Green et al. 2014; Green et al. 2019; Bailer-Jones et al. 2021) and improves it in several important ways: i) the use of multiple stellar populations (in addition to the dominant main-sequence stars) ii) improved color tracks for main-sequence stars and (especially) red giants, including the use of very young (<1 Gyr) populations and an extended $[Fe/H]$ range, and iii) priors based on sophisticated TRILEGAL simulations (Dal Tio et al. 2022) that include multiple stellar populations and account for all principal structural components of the Galaxy.

Extensive, although still preliminary, testing demonstrates the expected statistical behavior of implemented computations and that SDSS-based luminosity-color sequences are supported by more recent direct (trigonometric) distance measurements by Gaia. Tests of pipeline performance show that a sample of 10^{10} stars can be processed in a few days using a moderate-size cluster. We intend to process each LSST Data Release as they become available and make outputs accessible via Rubin Science Platform.

We anticipate that the accuracy of resulting distance estimates, as well as estimates of metallicity and interstellar dust extinction along the line of sight, will improve as LSST advances because:

- As more LSST data are collected, photometric depth will improve as well as photometric calibration.

- Photometric light curves will improve identification of variable populations such as quasars, RR Lyrae stars and eclipsing binary stars. When proper motion measurements become available, separation of nearby stars from distant stars will improve, too.

- Luminosity-color sequences will be recalibrated in LSST's photometric system and improved using LSST's own globular cluster data and Gaia's parallax measurements.

- The extension of photometric coverage to longer IR wavelengths (e.g., using Euclid and Roman Space Telescope survey photometry) will improve constraints on model parameters, especially in Galactic plane regions with large dust extinction (for more details, see Berry et al. 2012).

- Improvements in our understanding of the Milky Way structure will iteratively lead to improvements of Galaxy models such as TRILEGAL, and in turn to improvements of Bayesian priors.

- In the context of interstellar dust studies, it may be worthwhile to consider hiearchical Bayesian modeling (e.g., setting priors for the line-of-sight reddening profile, as discussed by Green et al. 2014).

These improvements will require substantial additional work, but given the implied transformative impact of resulting distance estimates on our understanding of the formation and evolution of the Milky Way, it seems well justified.

Facilities: Sloan Digital Sky Survey, Gaia

Software: numpy (Oliphant 2006), matplotlib (Hunter 2007), seaborn (Waskom 2021), scipy (Jones et al. 2001–), pandas McKinney (2010), astropy (Astropy Collaboration et al. 2013, 2018), astroML (VanderPlas et al. 2012), Jupyter (Kluyver et al. 2016), Python (Van Rossum & Drake 2009).

APPENDIX

REFERENCES

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

Astropy Collaboration, Price-Whelan, A. M., Sipőcz, B. M., et al. 2018, AJ, 156, 123, doi: 10.3847/1538-3881/aabc4f

Bailer-Jones, C. A. L. 2011, MNRAS, 411, 435, doi: 10.1111/j.1365-2966.2010.17699.x

Bailer-Jones, C. A. L., Rybizki, J., Fouesneau, M., Demleitner, M., & Andrae, R. 2021, AJ, 161, 147, doi: 10.3847/1538-3881/abd806

Bergeron, P., Wesemael, F., & Beauchamp, A. 1995, PASP, 107, 1047, doi: 10.1086/133661

Berry, M., Ivezić, Ž., Sesar, B., et al. 2012, ApJ, 757, 166, doi: 10.1088/0004-637X/757/2/166

Bianco, F. B., Ivezić, Ž., Jones, R. L., et al. 2022, ApJS, 258, 1, doi: 10.3847/1538-4365/ac3e72

Bressan, A., Marigo, P., Girardi, L., et al. 2012, MNRAS, 427, 127, doi: 10.1111/j.1365-2966.2012.21948.x

Cardelli, J. A., Clayton, G. C., & Mathis, J. S. 1989, ApJ, 345, 245, doi: 10.1086/167900

Dal Tio, P., Pastorelli, G., Mazzi, A., et al. 2022, ApJS, 262, 22, doi: 10.3847/1538-4365/ac7be6

Dal Tio, P., Pastorelli, G., Mazzi, A., et al. 2022, The Astrophysical Journal Supplement Series, 262, 22, doi: 10.3847/1538-4365/ac7be6

Dotter, A., Chaboyer, B., Jevremović, D., et al. 2008, ApJS, 178, 89, doi: 10.1086/589654

Dufour, P., Blouin, S., Coutu, S., et al. 2017, in Astronomical Society of the Pacific Conference Series, Vol. 509, 20th European White Dwarf Workshop, ed. P. E. Tremblay, B. Gaensicke, & T. Marsh, 3, doi: 10.48550/arXiv.1610.00986

Fitzpatrick, E. L. 1999, PASP, 111, 63, doi: 10.1086/316293

Foreman-Mackey, D., Hogg, D. W., Lang, D., & Goodman, J. 2013, PASP, 125, 306, doi: 10.1086/670067

García-Zamora, E. M., Torres, S., & Rebassa-Mansergas, A. 2023, A&A, 679, A127, doi: 10.1051/0004-6361/202347601

Gentile Fusillo, N. P., Tremblay, P. E., Cukanovaite, E., et al. 2021, MNRAS, 508, 3877, doi: 10.1093/mnras/stab2672

Gordon, K. D., Fouesneau, M., Arab, H., et al. 2016, The Astrophysical Journal, 826, 104, doi: 10.3847/0004-637X/826/2/104

Green, G. M., Schlafly, E., Zucker, C., Speagle, J. S., & Finkbeiner, D. 2019, The Astrophysical Journal, 887, 93, doi: 10.3847/1538-4357/ab5362

Green, G. M., Schlafly, E. F., Finkbeiner, D. P., et al. 2014, ApJ, 783, 114, doi: 10.1088/0004-637X/783/2/114

Hunter, J. D. 2007, Computing In Science & Engineering, 9, 90, doi: 10.1109/MCSE.2007.55

Ivezić, Ž., Connolly, A. J., VanderPlas, J. T., & Gray, A. 2020, Statistics, Data Mining, and Machine Learning in Astronomy. A Practical Python Guide for the Analysis of Survey Data, Updated Edition

Ivezić, Ž., Smith, J. A., Miknaitis, G., et al. 2007, AJ, 134, 973, doi: 10.1086/519976

Ivezić, Ž., Sesar, B., Jurić, M., et al. 2008, ApJ, 684, 287, doi: 10.1086/589678

Ivezić, Ž., Kahn, S. M., Tyson, J. A., et al. 2019, ApJ, 873, 111, doi: 10.3847/1538-4357/ab042c

Jones, E., Oliphant, T., Peterson, P., et al. 2001–, SciPy: Open source scientific tools for Python. http://www.scipy.org/

Juric, M. 2011, in American Astronomical Society Meeting Abstracts, Vol. 217, American Astronomical Society Meeting Abstracts #217, 433.19

Jurić, M., Ivezić, Ž., Brooks, A., et al. 2008, ApJ, 673, 864, doi: 10.1086/523619

Kluyver, T., Ragan-Kelley, B., Pérez, F., et al. 2016, in Positioning and Power in Academic Publishing: Players, Agents and Agendas, ed. F. Loizides & B. Schmidt, IOS Press, 87 – 90

Lallement, R., Vergely, J.-L., Valette, B., et al. 2014, A&A, 561, A91, doi: 10.1051/0004-6361/201322032

McKinney, W. 2010, in Proceedings of the 9th Python in Science Conference, ed. S. van der Walt & J. Millman, 51 – 56

Oliphant, T. E. 2006, A guide to NumPy, Vol. 1 (Trelgol Publishing USA)

Oriol, A.-P., Virgile, A., Colin, C., et al. 2023, PeerJ Computer Science, 9, e1516, doi: 10.7717/peerj-cs.1516

Queiroz, A. B. A., Anders, F., Santiago, B. X., et al. 2018, Monthly Notices of the Royal Astronomical Society, 476, 2556, doi: 10.1093/mnras/sty330

Schlegel, D. J., Finkbeiner, D. P., & Davis, M. 1998, The Astrophysical Journal, 500, 525, doi: 10.1086/305772

Sesar, B., Ivezić, Ž., Lupton, R. H., et al. 2007, AJ, 134, 2236, doi: 10.1086/521819

Smolčić, V., Ivezić, Ž., Knapp, G. R., et al. 2004, ApJL, 615, L141, doi: 10.1086/426475

Thanjavur, K., Ivezić, Ž., Allam, S. S., et al. 2021, MNRAS, 505, 5941, doi: 10.1093/mnras/stab1452

Van Rossum, G., & Drake, F. L. 2009, Python 3 Reference Manual (Scotts Valley, CA: CreateSpace)

VanderPlas, J., Connolly, A. J., Ivezić, Ž., & Gray, A. 2012, in Proceedings of Conference on Intelligent Data Understanding (CIDU), pp. 47-54, 2012., 47–54, doi: 10.1109/CIDU.2012.6382200

Waskom, M. L. 2021, Journal of Open Source Software, 6, 3021, doi: 10.21105/joss.03021

Wyatt, S., Juric, M., McGuire, S., et al. 2024, in American Astronomical Society Meeting Abstracts, Vol. 243, American Astronomical Society Meeting Abstracts, 261.09