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] Dani Chao [ID],[1]
Oleksandra Razim [ID],[1] Filip Matković [ID],[5] Connor Yablonski [ID],[6] Toni Šarić [ID],[7] Tomislav Jurkić [ID],[3] 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 ex-
tinction 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 sim-
ulated 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 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 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 case of main-sequence and red giant stars, and surface gravity in 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 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 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 case of SDSS, colors include $u - g$, $g - r$, $r - i$, $i - z$, and in 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 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, distance modulus Δ can be computed from

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

or in a probabilistic form

$$p(\Delta) = p(r) - p(M_r + A_r). \tag{2}$$

Note that in practice the uncertainty of the observed magnitude, r^{obs}, is always significantly smaller than the width of $p(M_r + A_r)$ (at the bright end by an order of magnitude, $\sim$0.01 mag vs. $\sim$0.1 mag).

Therefore, $\vec{D} = (r, \vec{c})$ 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 and usually it is not explicitly computed when estimating model parameters: rather, $p(M, \vec{\theta}|D, I)$ for a given M is simply 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
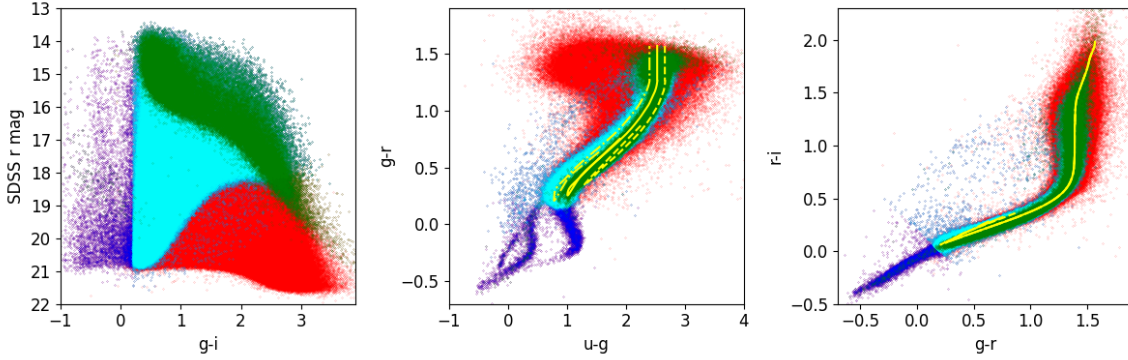
Figure 1. An illustration of multiple stellar populations. The red dots in the left panel show color-magnitude diagram for 841,000 stars from the SDSS Stripe 82 Standard Star Catalog (variable stars are excluded, Thanjavur et al. 2021) that have Gaia matches within 0.15 arcsec (after correcting for proper 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 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 other two panels. The three yellow lines in the middle panel show 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).

done with an arbitrary normalization. An important exception is model selection discussed further below, 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), 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.

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

2.2. Likelihood Computation

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

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

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

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

---

[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.

where 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

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

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 1-dimensional scalar, that is, $ln(\mathcal{L})$ is a scalar function of M_r, $[Fe/H]$, and A_r.

The existance 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 (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
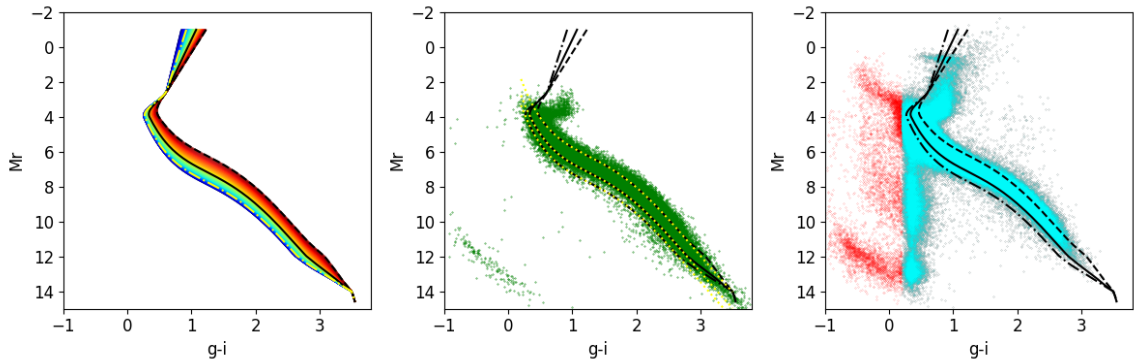
Figure 2. The left panel shows SDSS-based empirical absolute magnitude vs. color parametrization for main-sequence stars and red giants. The symbols are color-coded by metallicity, ranging from $[Fe/H] = -2.5$ to 0 (blue to red). The three lines correspond to three values of metallicity: $[Fe/H] = -2, -1, 0$ (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; same sample is shown by green dots in Figure 2). Their absolute magnitude 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 2), 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 "photo-geometric" 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 outliers (about 2.5% of the full sample) seen at $g - i = 0.4$ and $M_r > 7$ are predominantly found at the faint end ($r > 20$).

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 "photo-geometric" distances from Bailer-Jones et al. (2021) for main-sequence stars are on the "same scale". XXX ZI: provide numerical results for offsets!

We address all three problems by switching from the SDSS-based empirical isochrones that correspond to old populations to model-based isochrones that span a range of ages and a wider range of metallicities.

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 empirical SDSS-based isochrones.

Examples of 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.

These color tracks for main-sequence 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, unresolved binary stars and blue horizontal branch stars.

2.2.3. *Luminosity-Color Tracks for White Dwarfs*

There are two model families of tracks, for H and He white dwarfs, by Bergeron et al. (1995). Bob Abel will provide text and plots.

Figure 4 shows preliminary results.

2.2.4. *Luminosity-Color Tracks for Unresolved Binaries*

Following Smolčić et al. (2004), we generate luminosity-color tracks for unresolved binaries consisting of a main-sequence star and white dwarf.

Bob Abel will provide text and plots. Figure 5 shows preliminary results.

2.2.5. *Luminosity-Color Tracks for Blue Horizontal Branch Stars*

ŽI has data files from the Sirko et al. (2004) paper and needs to write a few words (and add tracks to Figure 4).

2.2.6. *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

---

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

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

the sample in a straighforward manner since they are variable (e.g., see Sesar et al. 2007). The remaining non-variable sources that are not well fit with 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 their fraction in the sample should be very small (most likely $< 1\%$).

2.2.7. *Accounting for Interstellar Dust Extinction*
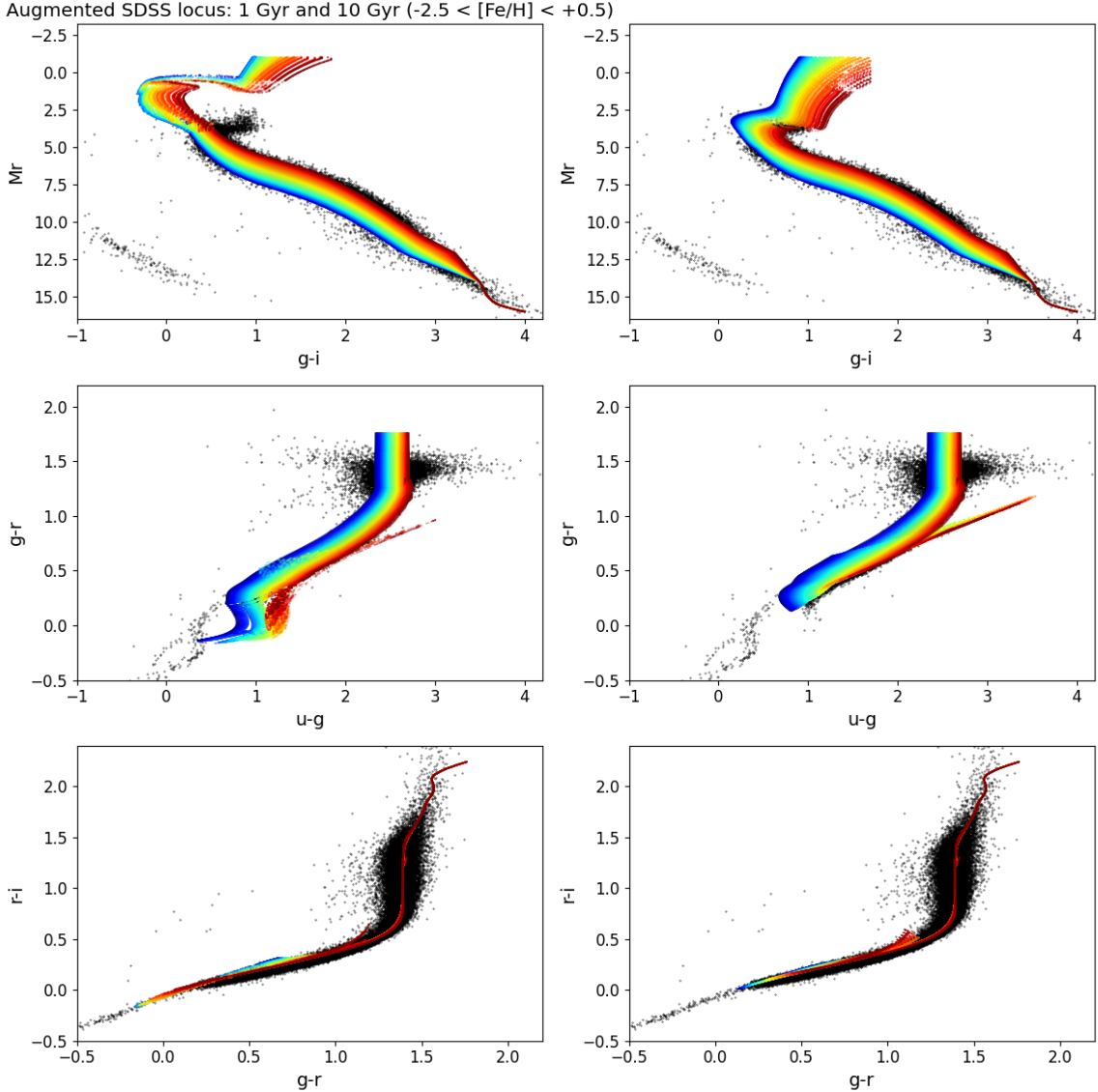
Figure 3. Augmented SDSS color tracks for two isochrone ages (left: 1 Gyr; right: 10 Gyr) and the full metallicity range ($-2.5 < [Fe/H] < +0.5$). The black dots show the same sample of 63,000 stars from the middle panel in Figure 2. As can be seen in the top right panel, sub-giant stars that could not be fit with emprical 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.
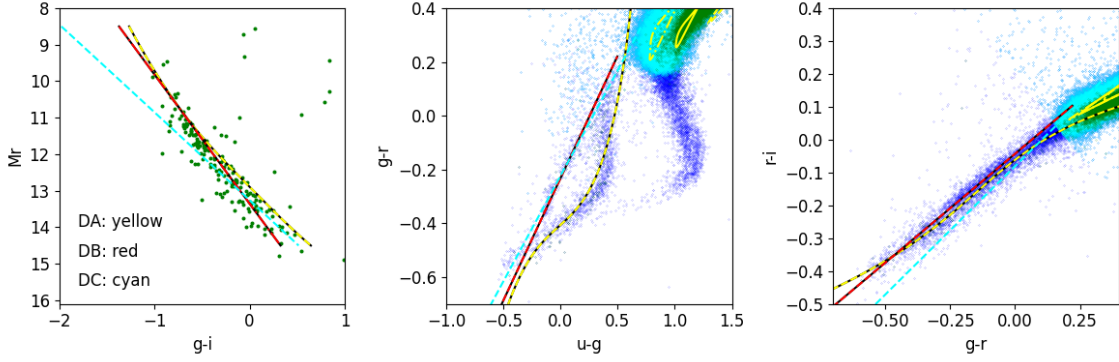
Figure 4. Preliminary results for white dwarf luminosity-color tracks. When updating this figure with Bob's improved tracks, add also tracks for BHB stars.
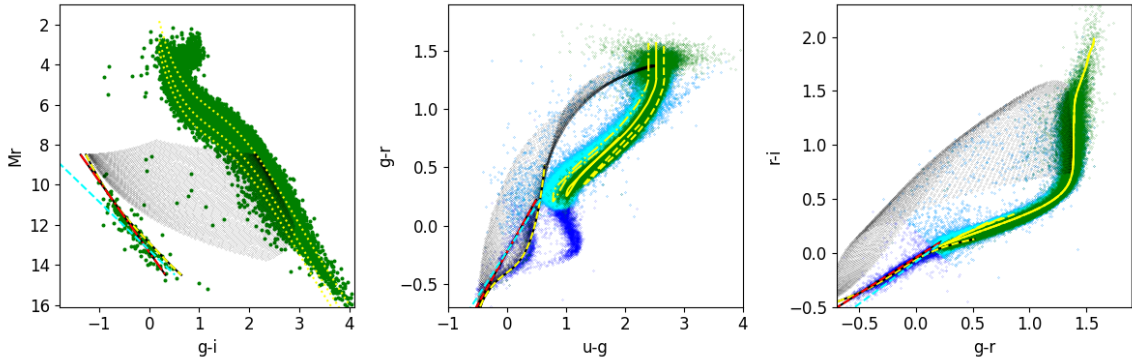
Figure 5. Preliminary results for unresolved binaries composed of a main-sequence star and white dwarf.

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).

Recent 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; 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 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 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
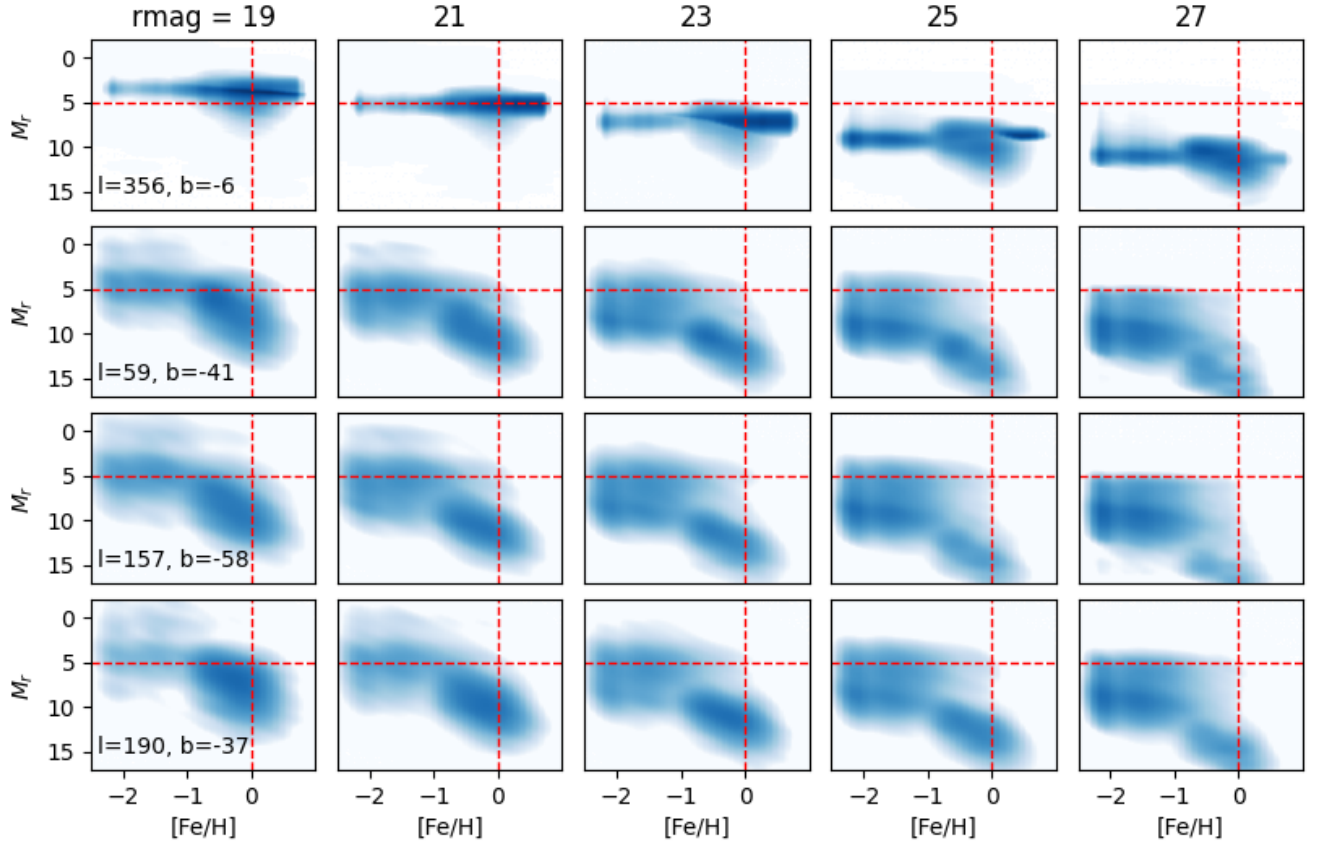
Figure 6. Examples of priors for model parameters, $p(M_r, [Fe/H])$, for main-sequence stars and red giants, for four sky positions (rows, for locations see the first column) and for several brightness levels in the LSST magnitude range (left to right: bright to faint). The integral of each map over M_r and $[Fe/H]$ is unity. The dashed lines are the same in each panel and are added to guide the eye.

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 $a = 1.3$, $b = 0.1$. 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).

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
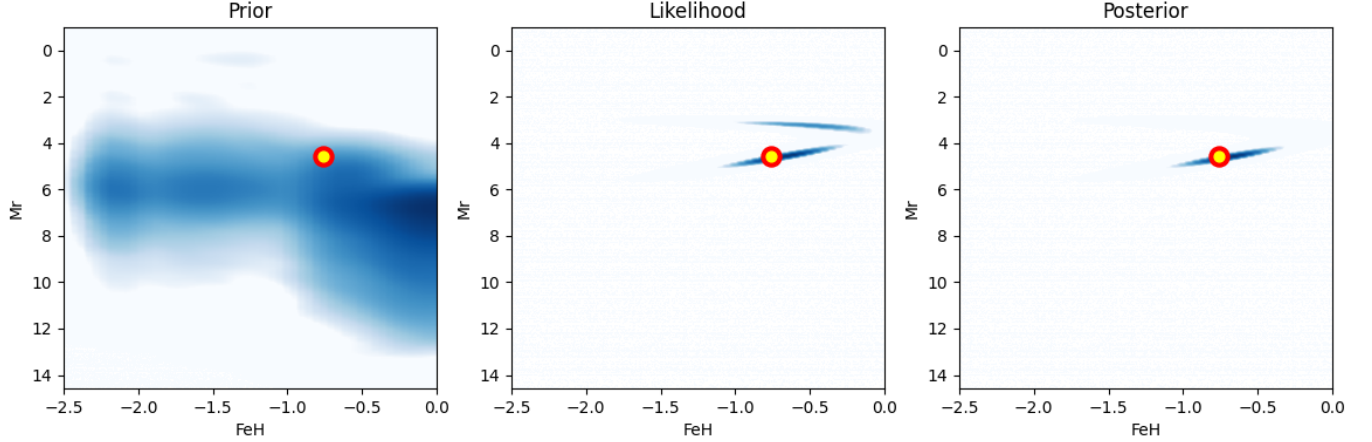
Figure 7. Maps of the prior (left), likelihood (middle; computed using eq. 6) and posterior (right; computed using eq. 3) for a simulated main-sequence star with $r = 21.2$, $(g - r) = 0.48$, $M_r = 4.58$, $[Fe/H] = -0.76$ and $A_r = 0.37$. The prior in the A_r direction is uniform. The circles mark the values of input M_r and $[Fe/H]$. The posterior marginal distributions for all three model parameters are shown in Figure 8.

Lab[4]. 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 using 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 parameters (e.g., M_r and $[Fe/H]$ in 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 6 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[5] 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 6 to illustrate the Bayesian model parameter estimation. Figure 7 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[6] in the likelihood map (the two "islands").
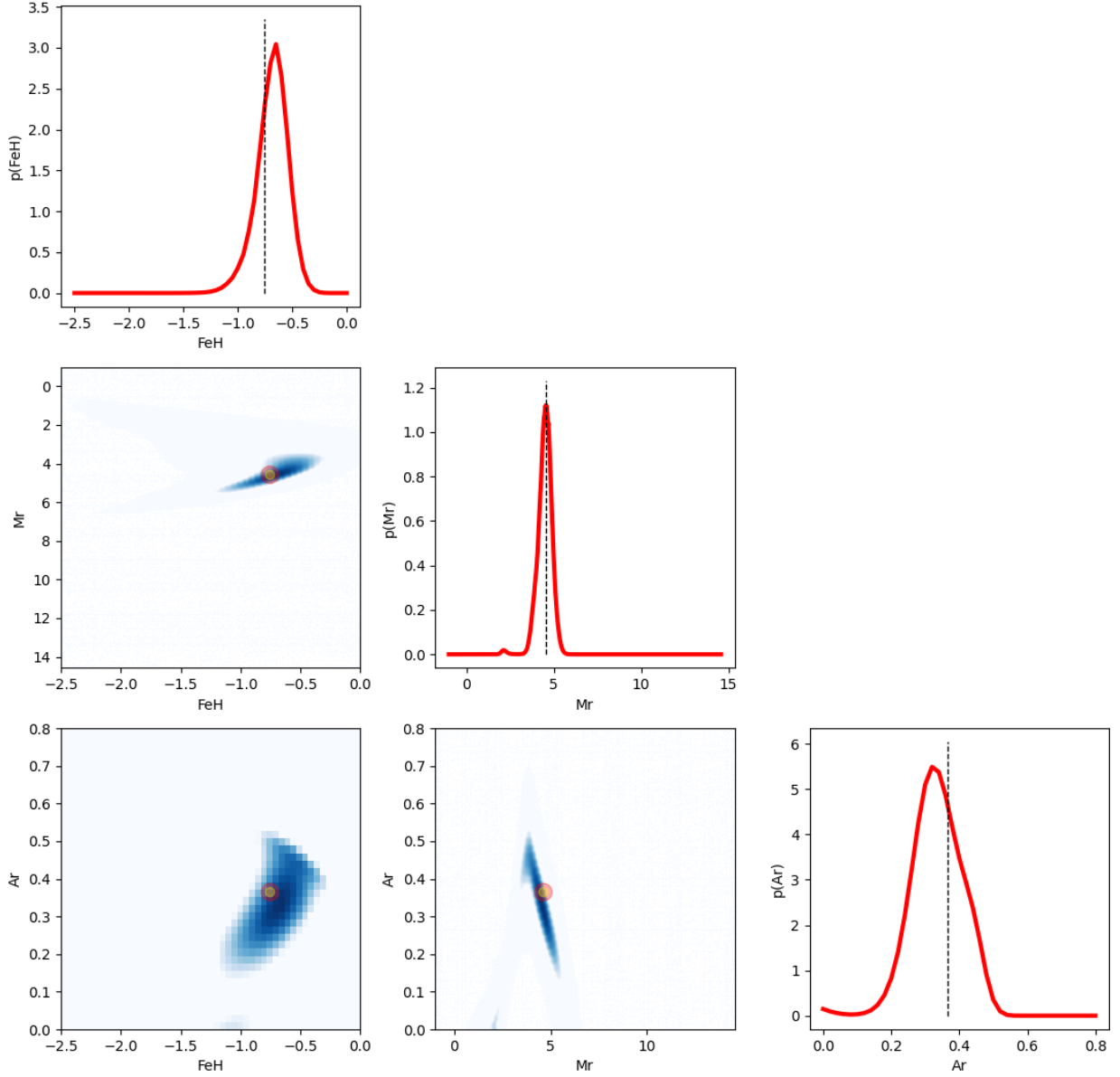
Figure 8 shows two-parameter covariances and marginal distributions for this case. Note that true values are recovered within expected uncertainties, as well as non-vanishing covariances between the parameters. The marginal distributions produced with the prior, likelihood and posterior maps, shown in 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 Bayesian model selection method for assigning posterior probabilities to each stellar population used to interpret observations.

---

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

[5] We use HEALPix geometry, see https://healpix.jpl.nasa.gov.

[6] 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. Two-parameter covariances and marginal distributions for the posterior map from Figure 7. The symbols and dashed lines show true input values.

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
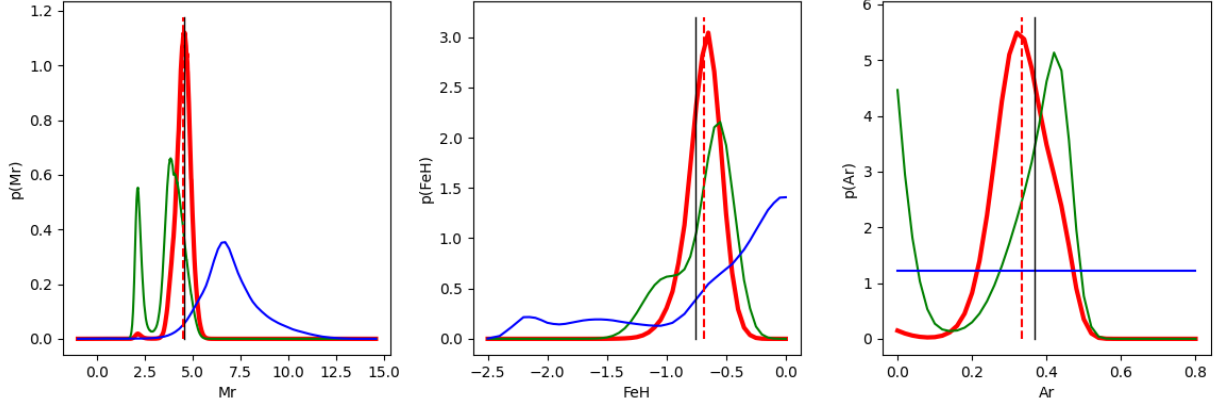
Figure 9. An illustration of the improvement in the "knowledge" of model parameters between prior and posterior, brought by measurements via the likelihood map. The lines show marginal distributions produced with the prior (blue), likelihood (green) and posterior (red) maps from Figure 7. The dashed line shows the expectation value for the marginalized posterior and the vertical solid line marks the true input value.

$\vec{\theta}$. 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 7). If the chosen model color tracks cannot explain the observed colors, the likelihood (eq. 6) will never be very 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; 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

$$B_{21} = \frac{\int p(D|M_2,\vec{\theta_2},I)\,p(\vec{\theta_2}|M_2,I)\,d\vec{\theta_2}}{\int p(D|M_1,\vec{\theta_1},I)\,p(\vec{\theta_1}|M_1,I)\,d\vec{\theta_1}}. \tag{15}$$

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 ($O_{11} = 1$). We will illustrate model selection with multiple stellar populations in the next Section.

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

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

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

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., 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/

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

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)

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

Sirko, E., Goodman, J., Knapp, G. R., et al. 2004, AJ, 127,
    899, doi: 10.1086/381483

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