



Machine-learned Identification of RR Lyrae Stars from Sparse, Multi-band Data: The PS1 Sample

Branimir Sesar¹, Nina Hernitschek¹, Sandra Mitrović¹, Željko Ivezić², Hans-Walter Rix¹, Judith G. Cohen³, Edouard J. Bernard⁴, Eva K. Grebel⁵, Nicolas F. Martin^{1,6}, Edward F. Schlafly^{1,7}, William S. Burgett⁸, Peter W. Draper⁹, Heather Flewelling¹⁰, Nick Kaiser¹⁰, Rolf P. Kudritzki¹⁰, Eugene A. Magnier¹⁰, Nigel Metcalfe⁹, John L. Tonry¹⁰, and Christopher Waters¹⁰

¹Max Planck Institute for Astronomy, Königstuhl 17, D-69117 Heidelberg, Germany

²University of Washington, Department of Astronomy, P.O. Box 351580, Seattle, WA 98195-1580, USA; bsesar@mpia.de

³Division of Physics, Mathematics and Astronomy, Caltech, Pasadena, CA 91125, USA

⁴Université Côte d'Azur, OCA, CNRS, Lagrange, France

⁵Astronomisches Rechen-Institut, Zentrum für Astronomie der Universität Heidelberg, Mönchhofstr. 12–14, D-69120 Heidelberg, Germany

⁶Observatoire astronomique de Strasbourg, Université de Strasbourg, CNRS, UMR 7550, 11 rue de l'Université, F-67000 Strasbourg, France

⁷Lawrence Berkeley National Laboratory, One Cyclotron Road, Berkeley, CA 94720, USA

⁸GMTO Corporation, 251 S. Lake Avenue, Suite 300, Pasadena, CA 91101, USA

⁹Department of Physics, University of Durham, South Road, Durham DH1 3LE, UK

¹⁰Institute for Astronomy, University of Hawai'i at Manoa, Honolulu, HI 96822, USA

Received 2016 November 25; revised 2017 March 7; accepted 2017 March 7; published 2017 April 7

Abstract

RR Lyrae stars may be the best practical tracers of Galactic halo (sub-)structure and kinematics. The PanSTARRS1 (PS1) 3π survey offers multi-band, multi-epoch, precise photometry across much of the sky, but a robust identification of RR Lyrae stars in this data set poses a challenge, given PS1's sparse, asynchronous multi-band light curves ($\lesssim 12$ epochs in each of five bands, taken over a 4.5 year period). We present a novel template fitting technique that uses well-defined and physically motivated multi-band light curves of RR Lyrae stars, and demonstrate that we get accurate period estimates, precise to 2 s in $>80\%$ of cases. We augment these light-curve fits with other features from photometric time-series and provide them to progressively more detailed machine-learned classification models. From these models, we are able to select the widest (three-fourths of the sky) and deepest (reaching 120 kpc) sample of RR Lyrae stars to date. The PS1 sample of $\sim 45,000$ RRab stars is pure (90%) and complete (80% at 80 kpc) at high galactic latitudes. It also provides distances that are precise to 3%, measured with newly derived period–luminosity relations for optical/near-infrared PS1 bands. With the addition of proper motions from *Gaia* and radial velocity measurements from multi-object spectroscopic surveys, we expect the PS1 sample of RR Lyrae stars to become the premier source for studying the structure, kinematics, and the gravitational potential of the Galactic halo. The techniques presented in this study should translate well to other sparse, multi-band data sets, such as those produced by the Dark Energy Survey and the upcoming Large Synoptic Survey Telescope Galactic plane sub-survey.

Key words: catalogs – Galaxy: halo – methods: data analysis – methods: statistical – stars: variables: RR Lyrae – surveys

Supporting material: machine-readable tables

1. Introduction

Studies of the Galactic halo can help constrain the formation history of the Milky Way and the galaxy formation process in general. For example, in a recent theoretical study, Hargis et al. (2014) suggested that there may be between 300 and 700 low-luminosity ($<10^3 L_\odot$) dwarf satellite galaxies orbiting the Milky Way within 300 kpc of the Sun. However, the census of such low-luminosity galaxies is currently complete only within ~ 45 kpc of the Sun (Table 3 of Koposov et al. 2008), and only at high galactic latitudes ($|b| > 25^\circ$). A deeper, wider, and more complete census of Milky Way dwarfs would be extremely valuable because it would allow us to test our assumptions about Λ CDM cosmology and galaxy formation, by comparing the observed distribution and properties of discovered dwarfs against those present in state-of-the-art

hydrodynamic simulations, such as the APOSTLE (Sawala et al. 2016) and FIRE/Gizmo simulation suites (Hopkins et al. 2014).

The Galactic halo also contains remnants of accreted satellites (i.e., dwarf galaxies and globular clusters) that were disrupted by tidal forces and stretched into stellar tidal streams and clouds (e.g., Ibata et al. 2001; Belokurov et al. 2007; Sesar et al. 2015; Bernard et al. 2016). Stellar streams are especially interesting because their orbits are sensitive to the properties of the Galactic potential (e.g., its shape and total mass of the Milky Way) and thus can be used to constrain it over the range of distances spanned by the streams (e.g., Koposov et al. 2010; Newberg et al. 2010; Sesar et al. 2013; Belokurov et al. 2014). For example, the total mass of the Milky Way is currently uncertain at a factor of two, and its more precise measurement using stellar streams may help resolve (or further aggravate) some apparent issues in the theory of galaxy formation, such as the so-called “Too-Big-To-Fail” problem (Boylan-Kolchin et al. 2011; Wang et al. 2012). A more precise measurement of the total mass requires detailed modeling of stellar streams,



Original content from this work may be used under the terms of the [Creative Commons Attribution 3.0 licence](https://creativecommons.org/licenses/by/3.0/). Any further distribution of this work must maintain attribution to the author(s) and the title of the work, journal citation and DOI.

such as the Sagittarius stream, as well as precise 3D kinematics and positions of stars in distant streams (Price-Whelan et al. 2014). While the *Gaia* mission (Perryman et al. 2001) can, and will, deliver precise proper motions of halo stars, in most cases *Gaia*'s parallax estimates will only be marginal beyond a few kiloparsecs. Studies focusing on the 6D phase space structure of the Galactic halo will need to rely on tracers with precise (spectro-)photometric distances for the foreseeable future, making “standard candles,” such as RR Lyrae stars, enormously valuable.

To measure the total mass of the Milky Way and find the faintest dwarf satellites, we need to trace the spatial and kinematic structure and substructure (i.e., stellar streams) of the Galactic halo over the greatest possible distances and with the highest possible precision in distance and velocity, and the best tracers¹¹ for this task are RR Lyrae stars.

RR Lyrae stars are old (age > 10 Gyr), metal-poor ($[Fe/H] < -0.5$ dex), pulsating horizontal branch stars with periodically variable light curves (periods ranging from 0.2 to 0.9 days; Smith 2004). They are bright stars ($M_V = 0.6 \pm 0.1$ mag) with distinct light curves, which makes them easy to identify with time-domain imaging surveys, even to large distances (5–120 kpc for surveys with a $14 < V < 21$ magnitude range; e.g., Sesar et al. 2010). These properties, and the fact that almost every Milky Way dwarf satellite galaxy has at least one RR Lyrae star (including the faintest one, Segue 1; Simon et al. 2011), open up the exciting possibility of locating very-low-luminosity Milky Way dwarf satellites by using distant RR Lyrae stars, as first proposed by Sesar et al. (2014; also see Baker & Willman 2015).

RR Lyrae stars are also precise standard candles (i.e., their intrinsic brightness is well-determined). While distances to RR Lyrae stars can be measured with 3% uncertainty using optical data (Section 3.3), thanks to a tight period–luminosity relation in the near-infrared, distances to RR Lyrae stars can be measured with 2% or better precision using, for example, *K*-band observations (Braga et al. 2015; Beaton et al. 2016). Having precise distances is crucial for measuring tangential velocities¹² and thus the Galactic potential, as the uncertainty in tangential velocity increases proportionally with the uncertainty in distance.

As made evident by several existing catalogs of Milky Way halo RR Lyrae stars (e.g., Vivas et al. 2001; Keller et al. 2008; Miceli et al. 2008; Sesar et al. 2010, 2013; Drake et al. 2013; Abbas et al. 2014), selection of RR Lyrae stars has become an almost routine procedure, as long as one has access to ~ 40 or more observations per star in a single photometric bandpass. While very useful for many Galactic studies, the above catalogs are not ideal: they are either deep with limited sky coverage (e.g., the SDSS Stripe 82 catalog covers 100 deg² and is complete up to 110 kpc, Sesar et al. 2010) or have wide coverage but are not very deep (e.g., the CRTS catalog covers 20,000 deg² and is complete up to 30 kpc, Drake et al. 2013). In addition, none of the above catalogs cover the Galactic plane, and thus cannot support studies of the old (> 10 Gyr) Galactic disk. Currently the only existing imaging survey that has the potential to overcome all of the above drawbacks, and provide a deep and wide-area catalog of RR Lyrae stars in the

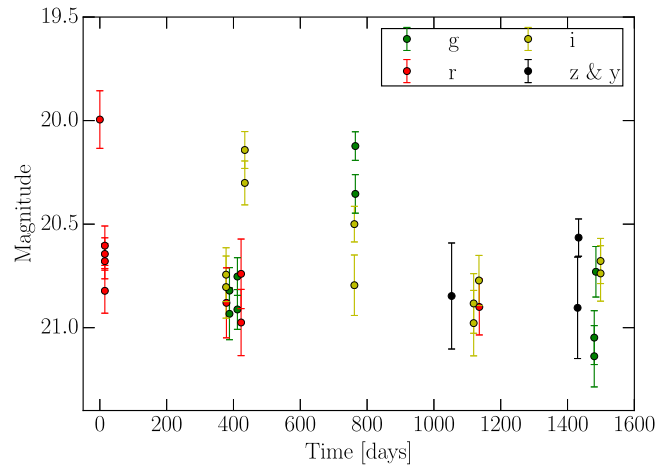


Figure 1. Multi-epoch PS1 *grizy* photometry (i.e., light curves) of a faint RR Lyrae star. Note that the observations in different bands are not synchronous, and that the light curves are sparsely covered in time: for this object, there are a total of 45 observations over four years, which spans about 3000 typical RR Lyrae periods.

northern skies (decl. $> -30^\circ$), is the Pan-STARRS1 (PS1) 3π survey.

Even though the PS1 3π survey holds great potential for Galactic studies due to its depth and sky coverage, it is a challenging data set for selection of RR Lyrae stars due to its sparse temporal coverage, cadence, and asynchronous multi-band observations (see Figure 1). As we described in our previous work (Hernitschek et al. 2016), we overcame these challenges by characterizing time-series of PS1 sources using three statistics: a χ^2 -based variability indicator, a variability amplitude (in the r_{PI} -band) ω_r , and a variability timescale τ , where the latter two were obtained by fitting a damped random walk model to observed PS1 multi-band structure functions (see Section 3.2 of Hernitschek et al. 2016). When applied to the second internal PS1 data release (PV2), our approach yielded a candidate sample of $\sim 150,000$ RR Lyrae stars covering three quarters of the sky and reaching up to 120 kpc from the Sun.

Building on the work by Hernitschek et al. (2016), in this paper, we use the final PS1 data release (PV3) to significantly increase the completeness and purity of the PS1 sample of RR Lyrae stars. Compared to Hernitschek et al. (2016), we achieve higher completeness and purity by (1) having more observation epochs per object (72 in PV3 versus 55 in PV2), (2) by excising fewer and thereby retaining more of these observations (using a machine-learning algorithm that more efficiently identifies bad photometric data), (3) by building a more detailed machine-learned model of RR Lyrae stars in data space, and (4) by developing and running a CPU-intensive multi-band light-curve fitting on PS1 time-series, thereby directly determining the RR Lyrae periods. The purer samples of RR Lyrae stars that this work delivers are especially important for studies of the Galactic halo (e.g., when searching for low-luminosity dwarf satellites), as stars incorrectly identified as RR Lyrae stars may cause the appearance of spurious halo substructures (Sesar et al. 2010).

2. Data: PS1 3π Light Curves

From an observational point of view, RR Lyrae stars are A–F-type stars with distinct, periodically variable light curves. In

¹¹ Tracers are objects whose distribution reflects the distribution of the majority of stars (hopefully, in the least biased way).

¹² Radial velocities of RR Lyrae stars are straightforward to measure (Sesar 2012).

the following sections, we describe data that capture these properties of RR Lyrae stars.

Pan-STARRS1 (Kaiser et al. 2010, PS1) is a wide-field optical/near-IR survey telescope system located at the Haleakalā Observatory on the island of Maui in Hawai‘i. The largest survey undertaken by the telescope, the PS1 3π survey (Chambers 2011), has observed the entire sky north of decl. -30° in five filter bands (Stubbs et al. 2010; Tonry et al. 2012), reaching 5σ single-epoch depths of about 22.0, 22.0, 21.9, 21.0, and 19.8 mag in g_{P1} , r_{P1} , i_{P1} , z_{P1} , and y_{P1} bands, respectively. The uncertainty in photometric calibration of the survey is $\lesssim 0.01$ mag (Schlafly et al. 2012), and the astrometric precision of single-epoch detections is 10 mas (Magnier et al. 2008).

The PS1 3π survey aimed to observe each position in two pairs of exposures per filter per year, where the observations within each so-called transit-time-interval pair were taken ~ 25 minutes apart and in the same band. Thus, the survey should have obtained about 16 observations in each band (for a total of 80), but due to bad weather and telescope downtime, fewer epochs were observed (~ 70 on average).

Unlike Hernitschek et al. (2016, see their Section 2.2), we do not use bit-flags or other ad hoc procedures to exclude detections that may appear as non-astrophysical photometric outliers in PS1 time-series data (e.g., badly calibrated data, blended objects, etc.). We define a non-astrophysical photometric outlier as a photometric measurement that deviates by more than 2.5σ from its “expected” value, where σ is the total photometric uncertainty of that detection. Instead, to identify and remove non-astrophysical photometric outliers, we employ a machine-learned model that uses other properties associated with a detection (e.g., its position on the chip, level of agreement with a point-spread function model, seeing, etc.) to predict whether a detection will be a 2.5σ outlier or not (B. Sesar et al. 2017, in preparation).

Validation tests have shown that our machine-learned outlier model identifies 80% of all true 2.5σ outliers, and only misclassifies one good observation for every true 2.5σ outlier. For comparison, the outlier rejection approach adopted by Hernitschek et al. (2016) identifies almost all of the 2.5σ outliers, but it misclassifies eight good observations for every true 2.5σ outlier.

After removing photometric outliers from the PV3 time-series (using our machine-learned outlier model), the average number of observations per object is 67 (out of the initial 72 observations). If we would have used the outlier rejection method of Hernitschek et al. (2016), the number of observations per object would have decreased to ~ 30 .

To select objects with enough epochs for multi-band light-curve fitting (Section 3.2), and signal-to-noise ratios appropriate for variability studies, we require that PS1 light curves have (after outlier rejection):

1. at least two epochs in g_{P1} , r_{P1} , and i_{P1} bands, and at least a total of two epochs in the “red” bands” (z_{P1} and y_{P1}),
2. a total of at least 23 epochs, and
3. an uncertainty-weighted mean magnitude of $15 < \langle m \rangle < 21.5$ in at least one of the PS1 g_{P1} , r_{P1} , i_{P1} bands.

Dereddened optical colors are useful because they provide a rough estimate of the spectral type, and could help with the identification of RR Lyrae stars (which are A–F-type stars). Thus, we correct observed PS1 magnitudes for extinction using the

extinction coefficients of Schlafly & Finkbeiner (2011; see their Table 6) and the Schlafly et al. (2014) dust map and calculate $\langle g \rangle - \langle r \rangle$, $\langle r \rangle - \langle i \rangle$, $\langle i \rangle - \langle z \rangle$, $\langle z \rangle - \langle y \rangle$, and $\langle g \rangle - \langle i \rangle$ colors, where $\langle \rangle$ indicates an uncertainty-weighted mean magnitude. If for some reason an object is not observed in a particular PS1 band, the value of the color involving that band is reset to 9999.99.

To extract variability information from multi-band PS1 light curves, we calculate the variability indicator $\hat{\chi}^2$ (Equation (1) of Hernitschek et al. 2016), and fit a damped random walk model to PS1 multi-band structure functions. From the best-fit damped random walk model, we measure the variability amplitude ω_r (in the r_{P1} -band), and the variability timescale τ (see Sections 3.1 to 3.3 of Hernitschek et al. 2016). As Hernitschek et al. (2016) have shown, these three parameters are very useful for separating different types of variable sources (e.g., quasars and RR Lyrae stars).

3. Light Curve and Period Fitting

In this section, we describe several approaches to fitting the multi-band light curves, which will result in a period determination and a fit to the phased light curve.

3.1. Multi-band Periodogram

A more detailed separation of variables can be obtained by studying the properties of phased (i.e., period-folded) light curves, such as the amplitude and shape (Dubath et al. 2011; Richards et al. 2011; Elorrieta et al. 2016). However, light-curve folding requires an assumed or known period.

To measure the period of variability of a PS1 light curve, we use the multi-band periodogram of VanderPlas & Ivezić (2015) as implemented in `gatspy`, an open-source Python package for general astronomical time-series analysis¹³ (Vanderplas 2015). Briefly, the algorithm of VanderPlas & Ivezić (2015) models the phased light curves in each band as an arbitrary truncated Fourier series, with the period and optionally the phase, shared across all bands.

Since the phase offsets between RR Lyrae *griz* light curves are smaller than 1% (Sesar et al. 2010), we adopt the VanderPlas & Ivezić (2015) shared-phase model when calculating the multi-band periodogram (i.e., we set `gatspy` parameters $N_{\text{base}} = 1$ and $N_{\text{band}} = 0$; see Section 5 of VanderPlas & Ivezić 2015 for details). Furthermore, since RR Lyrae stars have periods between 0.2 and 0.9 days, we limit the period search to the same range.

To test the accuracy of the VanderPlas & Ivezić (2015) period-finding algorithm on PS1 data, we use `gatspy` to calculate multi-band periodograms for 440 RR Lyrae stars previously studied by Sesar et al. (2010), but with the PS1 data at hand. From each periodogram, we select the period associated with the highest periodogram peak. We considered the periods measured by Sesar et al. (2010) from more densely sampled Sloan Digital Sky Survey (SDSS; York et al. 2000) Stripe 82 observations to be “true” periods (see Section 2 of Sesar et al. 2010 for details on SDSS Stripe 82). We define a period to be accurately measured if the selected multi-band periodogram peak is within 2 s of the true period.

We find that the VanderPlas & Ivezić (2015) period-finding algorithm accurately identifies the true period for 53% of RR

¹³ <http://www.astroML.org/gatspy/>

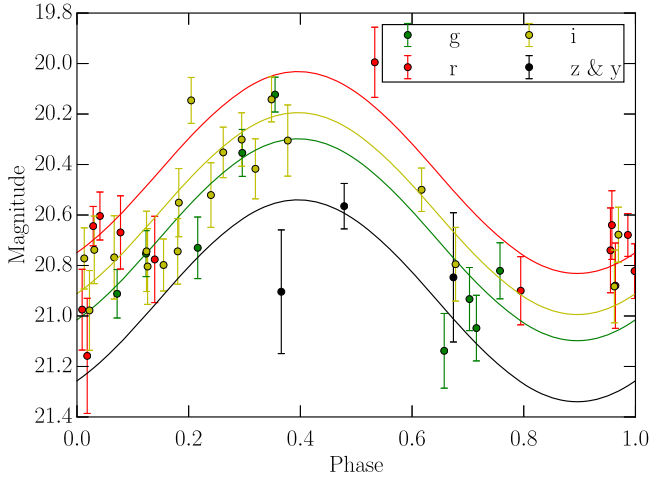


Figure 2. Unsuccessful attempt to accurately measure the period of the RR Lyrae star from Figure 1, using the multi-band periodogram of VanderPlas & Ivezić (2015). Even though the best-fit multi-band model (sine curves) agrees with the phased PS1 light curves (symbols with errorbars) in the minimum χ^2 sense, the period inferred by this modeling approach is incorrect. This happens because the algorithm permits light-curve models that are not physical: the model’s colors do not change as a function of phase.

Lyrae stars observed by PS1 (37% if the accuracy of 1 s is required). Changing the `gatspy` N_{base} and N_{band} parameters did not significantly improve this result.

Using a mathematical, not physical, multi-band light-curve model is one of the reasons why the VanderPlas & Ivezić (2015) period-finding algorithm fails to identify the true period for half of the RR Lyrae stars observed by PS1. When the light curves are sparse, there is no guarantee that the resulting best-fit multi-band model will be physical. If the light-curve model is not constrained by external information on the physics of the problem at hand, inaccurate, or not robust, period estimates may result.

An example of a best-fit, but non-physical (mathematical) multi-band model is shown in Figure 2. The $g - r$ color predicted by the best-fit model does not change as a function of phase (i.e., the difference between the green and red line), while in reality, it is well known that RR Lyrae stars have bluer $g - r$ color when they are brightest (e.g., Sesar et al. 2010). Furthermore, the model predicts $i - z \sim 0.2$ mag, while in reality $i - z \sim 0$ mag. Due to a combination of a non-physical multi-band model and sparse PS1 data, the VanderPlas & Ivezić (2015) algorithm is unable to accurately measure the pulsation period of that particular star.

Since the VanderPlas & Ivezić (2015) algorithm fails to accurately measure the period for almost half of the RR Lyrae stars with the PS1 data at hand, we cannot use it to phase the light curves. However, we still calculate and use the multi-band periodogram in this work in the subsequent classification because it improves the selection of RR Lyrae stars (see Section 4.5 below).

3.2. Multi-band Light-curve Fitting and Periods

We now show that it is possible to accurately measure periods of RR Lyrae stars observed by PS1, by using a more realistic and physically constrained multi-band light-curve model.

In principle, such a model could be obtained by extracting theoretical *griz* light curves from pulsation models, such as

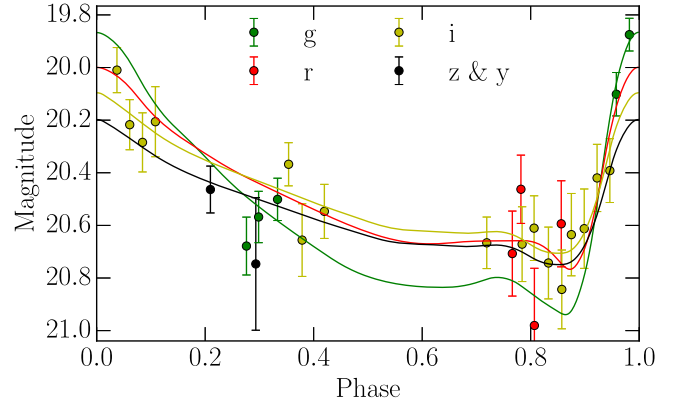


Figure 3. Phased PS1 light curves of the object shown in Figure 1, folded using the best-fit period measured from multi-band template fitting of PS1 data (see Section 3.2 for details). The best-fit multi-band template (g , r , i , z and y) is overplotted. Even though this star is very faint ($r \sim 20.5$ mag) and its light curve is sparsely sampled in PS1 (a total of 45 observations across 3000 periods), the period of ~ 0.51 days measured using multi-band template fitting agrees within 2 s with the value measured by Sesar et al. (2010) from more densely sampled SDSS Stripe 82 data.

those created by Marconi et al. (2006). However, a comparison of theoretical (Marconi et al. 2006) and empirical (i.e., observed) SDSS *ugriz* light curves by Sesar et al. (2010, see their Figure 8) has shown differences between the two light-curve sets that cannot be explained by observational uncertainties. Due to these differences, we are reluctant to use theoretical multi-band models of RR Lyrae stars when measuring periods.

Instead of using theoretical multi-band models, we adopt a set of 483 empirical *griz* models. These models consist of *griz* light-curve templates that were fitted by Sesar et al. (2010) to observed *griz* light curves of 483 RR Lyrae stars in SDSS Stripe 82. The curves in Figure 3 illustrate one of the 483 empirical multi-band models. The set contains 379 type *ab* multi-band templates, corresponding to RR Lyrae stars pulsating in the fundamental model (RRab stars), and 104 type *c* multi-band templates, corresponding to RR Lyrae stars pulsating in the first overtone (RRc stars).

We define the k th empirical multi-band model, which is a function of the pulsation phase ϕ as

$$\begin{aligned} g(\phi) &= FA_g T_g(\phi) + g_0 - r_0 + r' \\ r(\phi) &= FA_r T_r(\phi) + r' \\ i(\phi) &= FA_i T_i(\phi) + i_0 - r_0 + r' \\ z(\phi) &= FA_z T_z(\phi) + z_0 - r_0 + r', \end{aligned} \quad (1)$$

where $T_m(\phi)$ is the best-fit template light curve, A_m is the (known and fixed) amplitude of this template, and m_0 is the (known and fixed) best-fit magnitude at peak brightness (i.e., at $\phi = 0$) in the $m = g, r, i, z$ band of the k th RR Lyrae star in SDSS Stripe 82 (see Table 2 of Sesar et al. 2010 for values of A_m and m_0). Note that the free parameter r' acts as a zero-point offset in our model, since the *griz* light curves have been normalized by subtracting r_0 from each light curve. The free parameter F allows the amplitudes of model *griz* light curves to vary by up to 20% from their original values (which are listed in Table 2 of Sesar et al. 2010).

Qualitative inspection of phased PS1 z and y band light curves has shown that the two are roughly similar within photometric uncertainties. Therefore, we (can) treat all y -band

observations as z -band observations in the remainder of the analysis.

Assuming a period of P days and a phase offset ϕ_0 , we calculate the phase of each PS1 observation epoch as

$$\phi(t | P, \phi_0) = \frac{(t - 2,400,000) \bmod P}{P} + \phi_0, \quad (2)$$

where the time of observation t is in units of heliocentric Julian days, and $-0.5 \leq \phi_0 < 0.5$. The purpose of the phase offset ϕ_0 is to make sure the maximum light of the best-fit multi-band template occurs at $\phi = 0$. Note that the phase of an observation needs to be in the $0 \leq \phi < 1$ range. If it is outside of that range, one should add or subtract 1.

To find the best-fit values of F , r' , ϕ_0 , and P parameters for a given multi-band template, k , we minimize a χ^2 -like statistic calculated as

$$\chi_k^2 = \sum_{m=g,r,i,z} \sum_{n=1}^{N_{m,\text{obs}}} \left(\frac{m_{m,n} - m_k(\phi(t_{m,n} | P, \phi_0) | F, r')}{\sigma_{m,n}} \right)^2, \quad (3)$$

where $t_{m,n}$, $m_{m,n}$, and $\sigma_{m,n}$ are the time, magnitude, and the photometric uncertainty of the n th observation in the $m = g, r, i, z$ band (e.g., $t_{g,1}$, g_1 , $\sigma_{g,1}$). The best-fit parameters (period, phase offset, etc.) are measured using the Differential Evolution algorithm of Storn & Price (1997) as implemented in `scipy`, an open-source Python package for scientific computing¹⁴ (Oliphant 2007; Millman & Aivazis 2011).

We perform multi-band light-curve fitting in two runs. In the first run, we fit every multi-band template to a PS1 multi-band time-series, and for each template we record the best-fit χ^2 and model parameter values. When fitting a type ab multi-band template, we constrain the minimization to periods ranging from 0.4 to 0.9 days. The minimization is constrained to periods ranging from 0.2 to 0.5 days when fitting a type c multi-band template. The above period ranges are typical of type ab (RRab) and type c (RRc) RR Lyrae stars pulsating in the fundamental or the first-overtone mode, respectively. For illustration, Figure 4 shows the 483 best-fit χ_k^2 and period values, one for each template light curve, measured for a faint RR Lyrae star and a faint non-RR Lyrae object. Clearly, there is a global minimum for the RR Lyrae light curve among the χ_k^2 values.

In a second round of light-curve fitting, we fit only type ab or type c templates, depending on the type of the best-fit template (i.e., the template with the lowest χ^2 value) found during the first run. This time, the period range is restricted to ± 2 minutes around the period associated with the best-fit template from the first run. At the end of the second fitting iteration, we save only the best-fit χ^2 and model parameter values associated with the best-fit template (of the second run).

The result of applying this procedure to the PS1 light curves of 440 RR Lyrae stars in SDSS Stripe 82 is illustrated in Figure 5. Our multi-band template fitting method accurately measures periods for 85% of RR Lyrae stars (87% of RRab and 74% of RRc stars), a 32% improvement in period recovery over the VanderPlas & Ivezić (2015) algorithm. Within 1 s, the period is recovered for 73% of RR Lyrae stars (a 36% improvement versus the VanderPlas & Ivezić 2015 algorithm). If the period fitting returns a discrepant value, this can

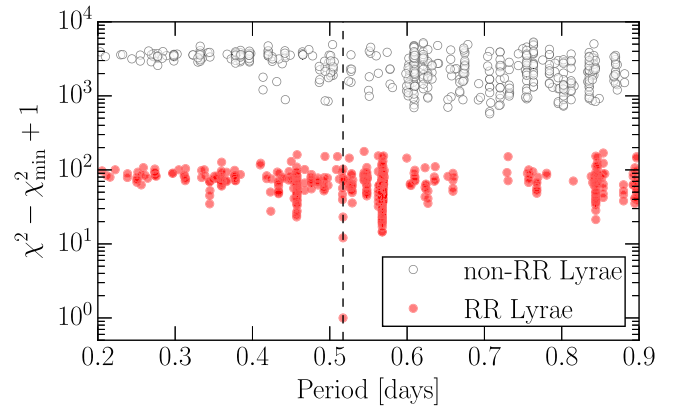


Figure 4. Symbols show the best-fit periods and associated χ^2 values obtained by fitting each of the 483 multi-band templates to PS1 light curves of a faint non-RR Lyrae object (open gray points), and the same faint RR Lyrae star shown in Figures 1–3 (red points). The vertical dashed line shows the RR Lyrae star’s true period measured by Sesar et al. (2010) from SDSS Stripe 82 data. For this RR Lyrae star, the period associated with the best-fit template (i.e., the template with the smallest χ^2 value) is consistent within 2 s with the star’s true period, indicating a successful period recovery. Note the classification power of the template fitting χ^2 statistic, because it clearly separates the RR Lyrae star from the non-RR Lyrae object, even though both objects have similarly sampled PS1 light curves and signal-to-noise ratios.

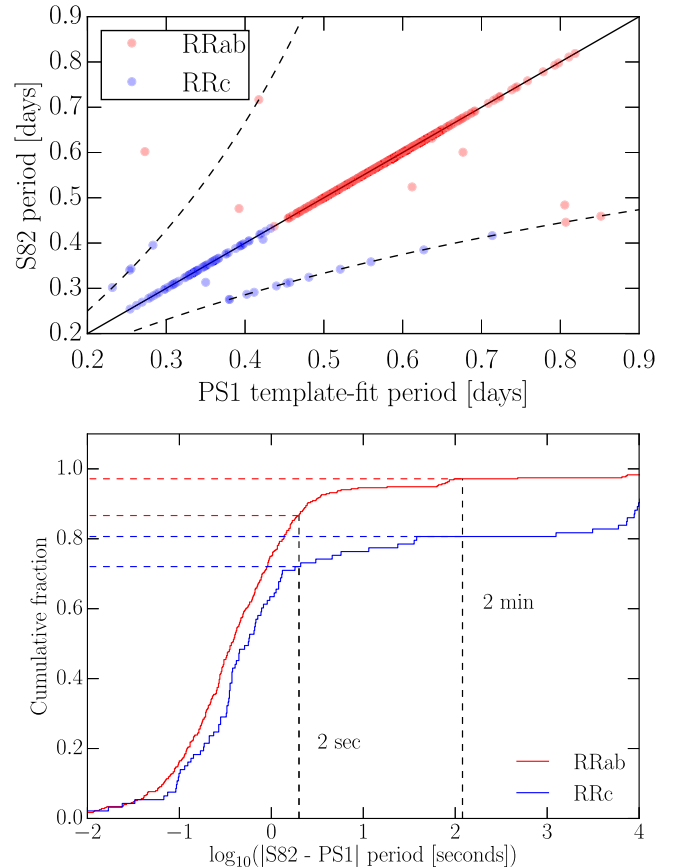


Figure 5. Accuracy, precision, and robustness of the RR Lyrae period estimates obtained using our multi-band light-curve template fitting of PS1 light curves. The top panel compares periods measured from Stripe 82 data (by Sesar et al. 2010) with those measured from PS1 data using multi-band template fitting. The dashed lines show the one-day beat frequency aliases. The bottom panel quantifies the period recovery: the period is accurately recovered (i.e., within 2 s) for 87% of RRab and 74% of RRc stars.

¹⁴ <http://www.scipy.org>

predominately be attributed to one-day beat frequency aliasing (see Figure 5).

When fitting multi-band templates to PS1 light curves of 440 RR Lyrae stars in SDSS Stripe 82, there is a possibility that some Stripe 82 RR Lyrae stars will be best fit with their own multi-band templates (recall that multi-band templates were constructed from observed SDSS light curves of individual Stripe 82 RR Lyrae stars). This “self-fitting” can be considered to be a form of overfitting,¹⁵ which, if it happens frequently, may inflate our estimate of the accuracy of period recovery. However, only 6 out of 440 Stripe 82 RR Lyrae stars are fit with their own multi-band templates, indicating that self-fitting does not happen frequently and that it does not significantly inflate our estimate of the accuracy of period recovery. The lack of self-fitting implies that many multi-band templates are quite similar to each other, and suggests that there is a potential for a computational speedup by removing redundant multi-band templates from the set.

As Figure 4 illustrates, the multi-band light-curve fitting also provides useful information for separating RR Lyrae stars and non-RR Lyrae objects. For example, the average and best-fit χ^2 values measured for the RR Lyrae star are vastly lower than the corresponding values measured for the non-RR Lyrae object, even though both objects have PS1 light curves of similar signal-to-noise ratio and sampling. This result is not unexpected, since χ^2 measures the statistical agreement between the (observed) phased PS1 light curve and a best-fit empirical multi-band light-curve model of an RR Lyrae star, and thus quantifies how “RR Lyrae-like” an object is.

In order to further characterize how RR-Lyrae-like a PS1 multi-band light curve is, we measure additional properties of phased PS1 light curves, such as the entropy of the phased light curve, the Stetson (1996) J index, as well as ~ 20 other properties (called features in machine learning), and describe them in more detail in Appendix A.

3.3. Resulting RR Lyrae Distance Precision

Along with measuring an accurate period (87% of the time for RRab stars), an important aspect of multi-band light-curve fitting is also the increased precision in estimating the star’s average flux (or magnitude). Because RR Lyrae stars follow a tight period–absolute magnitude–metallicity (PLZ) relation, we show that PS1 data constrain distances of RR Lyrae stars with 3% precision, even if the metallicity of an RR Lyrae star is unknown.

Theoretical and empirical studies (e.g., Catelan et al. 2004; Sollima et al. 2006; Braga et al. 2015; Marconi et al. 2015) have shown that the absolute magnitudes of RR Lyrae stars can be modeled as

$$M = \alpha \log_{10}(P/P_{\text{ref}}) + \beta([\text{Fe}/\text{H}] - [\text{Fe}/\text{H}]_{\text{ref}}) + M_{\text{ref}} + \epsilon, \quad (4)$$

where P is the period of pulsation, M_{ref} is the absolute magnitude at some reference period P_{ref} and metallicity $[\text{Fe}/\text{H}]_{\text{ref}}$ (here chosen to be $P_{\text{ref}} = 0.6$ days and $[\text{Fe}/\text{H}]_{\text{ref}} = -1.5$ dex), and α and β describe the dependence of the absolute magnitude on period and metallicity, respectively. The ϵ is a standard normal random variable with mean 0 and

standard deviation σ_M , that models the intrinsic scatter in the absolute magnitude convolved with unaccounted measurement uncertainties.

To constrain the PLZ relations for RRab stars in PS1 bandpasses, we use a probabilistic approach described in detail in Appendix B, where the data include metallicities and distance moduli of PS1 RR Lyrae stars in five Galactic globular clusters. The end product of this approach is a joint posterior distribution of all model parameters. To describe the marginal posterior distributions of individual model parameters, we measure the median, the difference between the 84th percentile and the median, and the difference between the median and the 16th percentile of each marginal posterior distribution (for a Gaussian distribution, these differences are equal to ± 1 standard deviation). We report these values in Table 1.

Overall, the PLZ relations behave as expected: as the bandpass moves to redder wavelengths, the dependence on the period increases (i.e., the α parameter becomes more negative), the dependence on the metallicity (i.e., the β parameter) and the scatter σ_M decrease, and the reference absolute magnitude becomes brighter. Similar trends were also observed in previous theoretical and observational studies (e.g., Braga et al. 2015; Marconi et al. 2015). Since the PLZ relation for the i_{p1} band is most tightly constrained and has low-metallicity dependence, we use it hereafter when measuring distances.

The metallicity information is not available for the vast majority of stars in PS1. To estimate the uncertainty in absolute i_{p1} magnitudes when the metallicity is unknown, we assume that RR Lyrae stars are drawn from the halo metallicity distribution function, represented with a Gaussian distribution centered on -1.5 dex and with a standard deviation of 0.3 dex (Ivezić et al. 2008). The resulting uncertainty in $M_{i_{p1}}$ is then $\sigma_{M_{i_{p1}}} = 0.06(\text{rnd}) \pm 0.03(\text{sys})$ mag, and the expression for $M_{i_{p1}}$ simplifies to

$$M_{i_{p1}} = -1.77 \log_{10}(P/0.6) + 0.46. \quad (5)$$

To calculate distance moduli of PS1 RR Lyrae stars, we use flux-averaged i_{p1} -band magnitude and Equation (5). For the uncertainty in distance modulus, we adopt $\sigma_{\text{DM}} = \sigma_{M_{i_{p1}}} = 0.06(\text{rnd}) \pm 0.03(\text{sys})$ mag. This corresponds to a distance precision of $\sim 3\%$, as long as dust extinction is not an important issue.

To validate Equation (5), we compute median distance moduli for three dwarf spheroidal galaxies, using the PS1 data for their RR Lyrae stars. We find $\text{DM} = 19.51 \pm 0.03$ mag for Draco, $\text{DM} = 19.67 \pm 0.03$ mag for Sextans, and $\text{DM} = 19.17 \pm 0.03$ mag for Ursa Minor, where the uncertainty in distance moduli is dominated by the systematic uncertainty in absolute magnitude. The values for Draco and Sextans agree well with the literature values of 19.40 ± 0.17 and 19.67 ± 0.1 (Bonanos et al. 2004; Lee et al. 2009), respectively. The $\text{DM} = 19.11 \pm 0.03$ mag we measure for Ursa Minor agrees with the 19.18 ± 0.12 mag measurement of Mighell & Burke (1999), but disagrees with the 19.4 ± 0.1 mag value measured by Carrera et al. (2002). The rms scatter of distance moduli of RR Lyrae stars in these dwarf galaxies is $\sigma_{\text{DM}} \approx 0.05$ mag. This scatter empirically verifies the intrinsic scatter of $\sigma_M = 0.05$ mag, we measured when fitting the PLZ relation in the i_{p1} band.

¹⁵ Because the correct period is more likely to be found for a star when its own template is used.

Table 1
PLZ Relations for PS1 Bandpasses

Band	α (mag dex ⁻¹)	β (mag dex ⁻¹)	M_{ref} (mag)	σ_M (mag)
g_{P1}	-1.7 ± 0.3	0.08 ± 0.03	$0.69 \pm 0.01(\text{rnd}) \pm 0.03(\text{sys})$	0.07 ± 0.01
r_{P1}	-1.6 ± 0.1	0.09 ± 0.02	$0.51 \pm 0.01(\text{rnd}) \pm 0.03(\text{sys})$	0.06 ± 0.01
i_{P1}	-1.77 ± 0.08	0.08 ± 0.02	$0.46 \pm 0.01(\text{rnd}) \pm 0.03(\text{sys})$	0.05 ± 0.01
z_{P1}^a	-2.2 ± 0.2	0.06 ± 0.02	$0.46 \pm 0.01(\text{rnd}) \pm 0.03(\text{sys})$	0.05 ± 0.01

Note.

^a The PLZ relation for the z_{P1} band was derived using z_{P1} and y_{P1} band observations, since a qualitative inspection of phased PS1 z and y band light curves has shown that the two are roughly similar within photometric uncertainties.

3.4. WISE Data

Quasars (QSOs) are some of the biggest sources of contamination when selecting RR Lyrae stars, especially at faint magnitudes (i.e., as the probed volume of the universe increases). They overlap with RR Lyrae stars in $g - r$ and redder optical colors (e.g., Figure 4 of Sesar et al. 2007), and may look as variable as RR Lyrae stars when observed in sparse data sets (such as PS1, Figure 3 of Hernitschek et al. 2016). Because most QSOs have a hot dust torus, they show an excess of radiation in the mid-infrared part of the spectrum, and have the *WISE* (Wright et al. 2010) mid-infrared color $W12 = W1 - W2 > 0.5$ mag (Figure 2 of Nikutta et al. 2014). RR Lyrae stars, on the other hand, have $W12 \sim 0$ mag.

To better separate QSOs and RR Lyrae stars, we supplement PS1 data with the $W12$ color provided by the all-sky *WISE* mission by matching PS1 and *WISE* positions using a $1''$ radius. If a PS1 object does not have a *WISE* $W1$ or $W2$ measurement, or those measurements have uncertainties greater than 0.3 mag (i.e., the *WISE* detection is less than 5σ above the background), we reset its $W12$ color to 9999.99 (this happens for about 50% of objects). We also calculate the $\langle i \rangle - W1$ color, and set its value to 9999.99 if one of its magnitudes is missing.

4. RR Lyrae Identification

We wish to build a model that returns the probability of an object to be an RR Lyrae star, given the data from Section 2 and the light-curve fits from Section 3. We will address this problem with a supervised machine-learning approach, where we use a training set (labeled or classified objects and their data) to infer a function that determines the class of unlabeled objects from their data. Since we have a reliable training set (see Section 4.1), supervised learning techniques (Section 4.2) represent a natural choice for building a classification model for a selection of RR Lyrae stars. The light-curve fitting and its χ^2 value (Section 3.2) play a crucial role in this process.

4.1. Training Set

To “learn” how to classify objects, supervised algorithms need to be trained using a subset of the data in which each object is labeled, i.e., their class is known in advance.

Our main training set consists of 1.9 million PS1 objects located in the SDSS Stripe 82 region ($310^\circ > \text{R.A.} < 59^\circ$, $|\text{decl.}| < 1.25^\circ$) that are brighter than 21.5 mag, have at least 23 observations (Section 2), and are at least $24'$ (two tidal radii; Harris 1996 (2010 edition)) away from the center of globular cluster NGC 7089. To label the objects in the training set, we match them to Sesar et al. (2010) and Süveges et al. (2012)

catalogs of RR Lyrae stars. If the position of an RR Lyrae star in one of these two catalogs matches the position of the closest PS1 object within $1''$, we label the PS1 object as an “RR Lyrae” (class 1; there are 462 such matches, and only three RR Lyrae stars do not have a PS1 match). The remaining PS1 objects are labeled as “non-RR Lyrae” (class 0). Out of 465 matched RR Lyrae stars, we know that 364 are of type ab (RRab) and 98 are of type c (RRc).

Since the SDSS Stripe 82 observations are slightly deeper and have six times more epochs than PS1 observations, we consider the Sesar et al. (2010) and Süveges et al. (2012) RR Lyrae catalogs to be 100% pure and complete up to the adopted faint PS1 magnitude limit (and likely beyond). Consequently, we consider the labels of PS1 objects in SDSS Stripe 82 to be the “ground truth” when measuring the efficiency of our selection method (i.e., the selection completeness and purity).

In SDSS Stripe 82, the majority of previously identified RR Lyrae stars are located within ~ 30 kpc of the Sun (400 stars or 83% of the sample, see Figure 10 of Sesar et al. 2010), and are thus fairly bright ($r_{P1} < 18.5$). This distribution is the result of Galactic structure, and is not a selection effect. To enhance the training set with fainter, and thus more distant RR Lyrae stars, we use RR Lyrae stars in the Draco dwarf spheroidal galaxy, located at a heliocentric distance of ~ 80 kpc (Kinemuchi et al. 2008). If the position of an RR Lyrae star in the Kinemuchi et al. (2008) catalog matches the position of the closest PS1 object (with at least 23 observations) within $1''$, we label the PS1 object as an “RR Lyrae” (there are 261 such matches, and 5 RR Lyrae stars do not have a PS1 match). Out of 261 matched RR Lyrae stars, 205 are of type ab (RRab), 30 are of type c (RRc), and 25 are d -type RR Lyrae stars (RRd) that pulsate simultaneously in the fundamental mode and first overtone.

4.2. Supervised Learning

To build this classification model, we use XGBoost¹⁶ (Chen & Guestrin 2016), an open-source implementation of the gradient tree boosting supervised machine-learning technique (Friedman 2001).

We use gradient tree boosting because the technique produces a prediction model in the form of an ensemble of decision trees, and because tree-based models are robust to uninformative features¹⁷ (Hastie et al. 2009; Dubath et al. 2011; Richards et al. 2011). This fact supports the usage of a

¹⁶ <https://github.com/dmlc/xgboost>

¹⁷ In machine learning, a feature is an individual measurable property of a phenomenon being observed (e.g., period of variability, color, light-curve amplitude, etc.).

large number of features when building the classification model, even when some of them may not be useful. By permitting the classification algorithm to consider even seemingly uninformative features, we allow it to consider potential correlations between data that may improve the classification in the end.

Given the resilience of gradient tree boosting to uninformative features, and the improvement in classification that additional features may bring, the best approach seems to be to train the classifier using the full set of features. However, this is impractical for the data set at hand. While calculating mean optical colors, low-level variability statistics (Section 2), and the multi-band periodogram takes less than a second per object, multi-band light-curve fitting takes ~ 30 min per object. Given that our training set contains about 1.9 million objects, calculating all of these features for all objects in the training set would be computationally prohibitive.

Instead of training a single classifier using the full set of features, we build three progressively more detailed classifiers using progressively smaller, but purer training sets (purer in the sense that the fraction of RR Lyrae stars in the training set increases). We describe these classifiers in Sections 4.4–4.6, but first give an overview of how a classifier is trained in Section 4.3.

4.3. Overview of Classifier Training

In brief, the steps in training a classifier are to

1. select training objects and input features,
2. tune XGBoost hyperparameters and train the classifier, and
3. measure the classification performance with a purity versus completeness curve.

While the first step is fairly self-explanatory, the remaining steps require further explanation.

The gradient tree boosting technique produces a prediction model in the form of an ensemble of decision trees. The number of trees in the ensemble, the maximum depth of a tree, the fraction of features that are considered when constructing a tree, and many other parameters that affect the manner in which the trees are grown and pruned, can be controlled via parameters¹⁸ exposed by the XGBoost package. By properly tuning these model hyperparameters, we can ensure that the classification produced by the model is not sub-optimal (e.g., not overfitted).

Before tuning the hyperparameters, we select input features, and shuffle and split the training set into two equal-sized sets, which we call development and evaluation sets. We use stratified splitting, i.e., we make sure that the ratio of RR Lyrae and non-RR Lyrae objects is equal in both sets.

To find the optimal hyperparameters, we use the GridSearchCV function in the scikit-learn open-source package for machine learning¹⁹ (Pedregosa et al. 2011). GridSearchCV selects test values of hyperparameters from a grid, and then measures the performance of the classification model (for the given hyperparameters) using tenfold stratified cross-validation on the development set. In detail, the development set is split into 10 subsets (using stratified

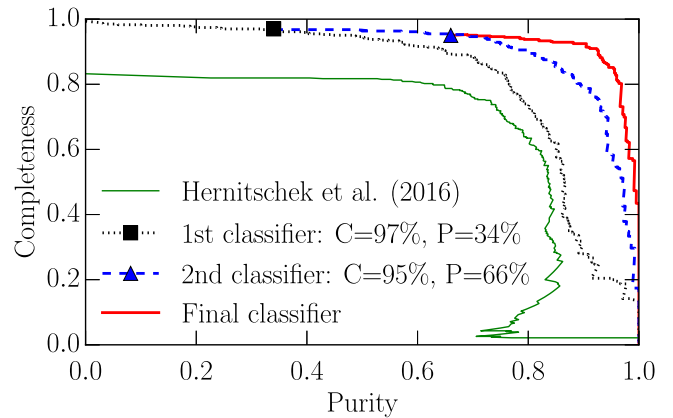


Figure 6. Power of the multi-band light-curve fitting in the classification of RR Lyrae stars. The figure shows purity vs. completeness curves produced by progressively more detailed classifiers described in Sections 4.4–4.6. The ideal classifier should approach the top right corner of the diagram. The square and star symbols show the purity and completeness of the classification with the adopted choice of scores returned by the first and second classifier (score₁ > 0.01 and score₂ > 0.13, respectively). The initial completeness is 99% due to initial data quality cuts (Section 2). Using the final classifier, we can select samples of RR Lyrae stars that are, for example, 90% complete and 90% pure.

splitting, see above), the model is trained on 9 subsets, and the probability of being an RR Lyrae star²⁰ (hereafter, the classification score²¹) is obtained from the trained model for objects in the tenth (i.e., withheld) subset. The performance of the classification is evaluated on the withheld set using some metric (see Sections 4.4–4.6 for details), and the whole procedure is repeated nine more times, each time with a different withheld set. The average of the 10 performance evaluations is stored, and the set of hyperparameters with the best average performance is used when training the classifier (step 3).

To verify whether the choice of the development set significantly affects the tuning of hyperparameters, we evaluate the performance of the tuned model on the evaluation set (which was not used by GridSearchCV during hyperparameter optimization), and then repeat the tuning process, but this time we use the evaluation set for tuning and the development set for evaluation. We find that the tuning procedure returns similar values of hyperparameters, regardless of the choice of the development set, indicating that the tuning of hyperparameters is not significantly biased by our choice of the development set.

Once the hyperparameters are tuned and the classifier is trained, we evaluate the performance of the classification using a purity versus completeness curve (see Figure 6 for examples). To measure this curve, we use the classification scores of objects in the Stripe 82 part of the training set (see Section 4.1 for details on this set). The classification scores of these objects were calculated using the tenfold cross-validation on the full (Stripe 82 and Draco) training set. For any threshold on the score and knowing the true class of each object in SDSS Stripe 82, we obtain the fraction of recovered RR Lyrae stars

¹⁸ See <https://xgboost.readthedocs.io/en/latest//parameter.html> for the full list.

¹⁹ <http://scikit-learn.org>

²⁰ Computed as the mean predicted class probabilities of the trees in the forest. Given a single tree, the probability that an object is of the RR Lyrae class (according to that tree) is equal to the fraction of training samples of the RR Lyrae class in the leaf in which the object ends up.

²¹ In this work, we use three classifiers and label their scores as score_j, where $j = 1, 2, 3$.

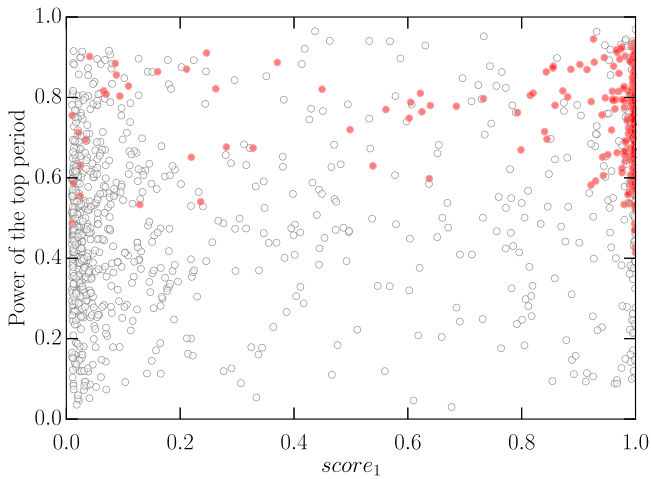


Figure 7. This plot shows that the multi-band periodogram contains useful information for separating RR Lyrae stars (solid circles) from non-RR Lyrae objects (open circles). Even though some true RR Lyrae stars may have low $score_1$ values (i.e., are not recognized by the first classifier as likely RR Lyrae stars), the power of the top period clearly separates them from non-RR Lyrae objects (e.g., $power_0 \gtrsim 0.4$).

(completeness), and the fraction of RR Lyrae stars in the selected sample (purity).

4.4. First Classification Step: Optical/IR Colors and Variability

To train the first classifier, we use the full training set of 1.9 million objects (Section 4.1) and adopt their variability statistics, as well as average PS1 and *WISE* colors as input features for the classifier (for a total of 10 features, see Sections 2 and 3.4 for details); we do not use the light-curve fitting of Section 3. When tuning the classifier, we select the values of hyperparameters that maximize the area under the purity versus completeness curve. The black dotted line in Figure 6 characterizes the performance of the trained classifier.

Our first classification outperforms the one obtained by Hernitschek et al. (2016) for all choices of sample purity and completeness (i.e., for all thresholds on the classification score). This is attributable perhaps foremost to a substantially greater number of observations per object in our data set (67 versus 35 in Hernitschek et al. 2016). The hyperparameter tuning, and use of a different machine-learning algorithm (XGBoost versus *scikit-learn* Random Forest) may also contribute to better performance.

Using a cut on the classification score of $score_1 > 0.01$ ($score_1$), we are able to reduce the number of objects under consideration by more than three orders of magnitude (from about 1.9 million to ~ 1500), while losing only 2% of RR Lyrae stars. However, the purity of the selected sample is still unacceptably low (only 34%). In order to improve the purity of the selected sample, we need to train the classifier using additional features, such as the multi-band periodogram (Section 3.1).

4.5. Second Classification Step: Multi-band Periodogram

For 1568 objects that pass the first classification cut ($score_1 > 0.01$), we calculate multi-band periodograms and extract the top 20 periods from each periodogram. Along with the periods, we also extract the power of each period (i.e., the height of the periodogram at that period).

As Figure 7 illustrates, the multi-band periodogram contains useful information for separating RR Lyrae stars from non-RR Lyrae objects. In principle, we could improve the purity of the selection by simply keeping all objects with $power_0 > 0.4$, without a loss of completeness. On the other hand, we may achieve even better classification if we provide the entire set of periods and their powers to XGBoost, and let the algorithm decide which features to use.

To improve the classification of RR Lyrae stars, we create a new feature set by combining 10 features used by the first classifier, with the top 20 periods and their powers obtained from the multi-band periodogram (for a total of 50 features). As the training set, we use ~ 1500 objects remaining from the initial training set after the $score_1 > 0.01$ cut. When tuning hyperparameters, we adopt values that optimize the area under the purity versus completeness curve. The blue dashed line in Figure 6 characterizes the performance of the second classifier.

We find that the addition of multi-band periodogram data improves the selection of RR Lyrae stars, as evidenced by higher sample purity at a given completeness (blue dashed line in Figure 6). For example, at 90% completeness, adding multi-band periodogram data increases the purity of the selected sample by 15% (with respect to the purity delivered by the first classifier, the black dotted line).

4.6. Final Classification Step: Multi-band Light-curve Fitting

Given the information in hand, a nearly optimal classification may be obtained by also including features extracted from phased multi-band light curves (Section 3.2 and Appendix A), but first, we need to fit multi-band light curves to objects under consideration.

Since multi-band light-curve fitting is computationally quite expensive (~ 30 minute per CPU and per object), we only do it for 910 training objects that have $score_2 > 0.13$, where $score_2$ is the classification score produced by the second classifier. According to Figure 6, this selection cut returns a sample with 66% purity and 95% completeness. By using this cut, we avoid fitting objects that are not likely to be RR Lyrae stars (i.e., we do not waste CPU time), but at the same time, do not reject many true RR Lyrae stars (i.e., the completeness decreases by 2% due to this cut, with respect to the 97% completeness obtained after the $score_1 > 0.01$ cut). In principle, we could have reached the same sample purity using a cut on the first classification score ($score_1$), but the decrease in completeness would have been much greater (a 6% decrease).

In Sections 4.4 and 4.5, we have trained classifiers using non-RR Lyrae objects (class 0) and RR Lyrae stars (class 1), that is, we have performed binary classifications. The reason for this two-step procedure was practical. In order to make the multi-band template fitting computationally feasible, we had to reduce the number of objects under consideration to a manageable level (by increasing the purity of the selected sample), while retaining as many true RR Lyrae stars as possible (i.e., by keeping the selection completeness as high as possible). Doing this required only knowing whether an object is likely an RR Lyrae star or not. By using cuts on $score_1$ and $score_2$ (binary) classification scores, we were able to reduce the number of objects from about 1.9 million to 900, while retaining 95% of RR Lyrae stars.

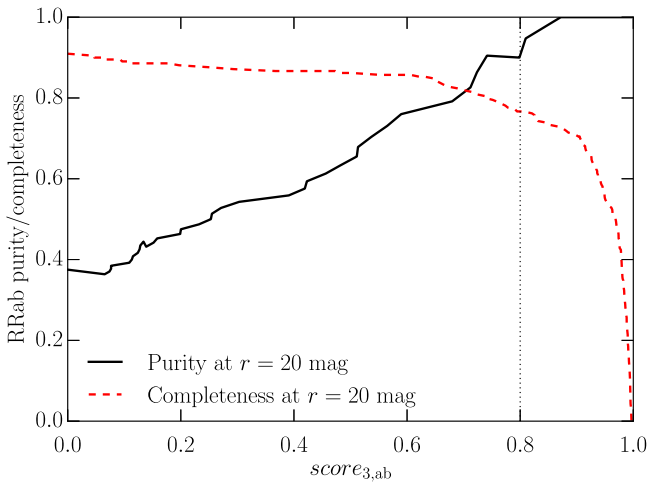


Figure 8. Expected purity and completeness for faint RRab stars, shown as a function of the threshold on the RRab classification score, $\text{score}_{3,ab}$. The initial purity is 38% due to the $\text{score}_2 > 0.13$ requirement (Section 4.6). A threshold of 0.8 (i.e., $\text{score}_{3,ab} > 0.8$, vertical dotted line) returns an RRab sample that is 91% pure and 77% complete at ~ 80 kpc ($\langle r \rangle \sim 20$ mag).

We now take a further step, determining whether an object is a non-RR Lyrae object (class 0), a type *ab* (class 1), or a type *c* or *d* RR Lyrae star (class 2) through a multiclass classification.

To train the final (multiclass) classifier, we use 910 training objects that have $\text{score}_2 > 0.13$ (and, of course, $\text{score}_1 > 0.01$). Of these 910 objects, 541 are RRab stars, and 144 are RRC or RRd stars (based on Sesar et al. 2010 and Kinemuchi et al. 2008 classifications). The remaining 225 objects are non-RR Lyrae objects. The feature set consists of 50 features employed by the second classifier, and 20 features extracted from phased multi-band light curves (Appendix A). Since we are training a multiclass classifier, when tuning hyperparameters we adopt values that minimize the logistic (or cross-entropy) loss (Bishop 2006). The thick red line in Figure 6 characterizes the purity and completeness of the selection as a function of the threshold on $\text{score}_{3,ab} + \text{score}_{3,c}$, where $\text{score}_{3,ab}$ and $\text{score}_{3,c}$ are RRab and RRC classification scores, respectively.

5. Verification and Analysis of the RR Lyrae Selection at High Galactic Latitudes

The purity versus completeness curve obtained using the full classifier (the thick red line in Figure 6) shows that we can select samples of RR Lyrae stars that are 90% complete and 90% pure; we deem this a gratifying and impressive success. However, it is important to keep in mind that the purity and completeness shown in Figure 6 are integrated over RRab and RRC stars, and over a range of distances (or magnitudes; roughly, 5–120 kpc, $14.5 < \langle r \rangle / \text{mag} < 21$). Since we have reasons to expect variations in purity and completeness as a function of type and distance (e.g., because the classification becomes more uncertain as objects get fainter and light curves become noisier), and because the knowledge of such variations is important for studies of Galactic structure (e.g., when measuring the stellar number density profile), below we present a more detailed analysis of the selection of RR Lyrae stars at high galactic latitudes ($|b| > 15^\circ$).

Table 2

Expected RRab Selection Purity and Completeness at ~ 80 kpc ($\langle r \rangle \sim 20$ mag)

Threshold on $\text{score}_{3,ab}$	Purity	Completeness
0.00	0.38	0.91
0.05	0.37	0.90
0.10	0.39	0.89
0.15	0.44	0.89
0.20	0.47	0.88
0.25	0.50	0.88
0.30	0.54	0.87
0.35	0.55	0.87
0.40	0.56	0.87
0.45	0.61	0.87
0.50	0.65	0.86
0.55	0.72	0.86
0.60	0.76	0.86
0.65	0.78	0.84
0.70	0.81	0.82
0.75	0.90	0.80
0.80	0.91	0.77
0.85	0.98	0.74
0.90	1.00	0.71
0.95	1.00	0.55

Note. A machine readable version of this table with a 0.01 step in threshold, is available in the electronic edition of the journal.

(This table is available in its entirety in machine-readable form.)

5.1. Purity and Completeness in Detail

The solid line in Figure 8 shows the purity of the RRab selection at the faint end (at ~ 80 kpc or $r \sim 20$ mag), given a threshold on $\text{score}_{3,ab}$. To make this curve, we use 80 labeled objects from the SDSS Stripe 82 training set with $19.7 < \langle r \rangle < 20.3$ and $\text{score}_2 > 0.13$, and calculate the fraction of true RRab stars in selected samples (given a threshold on $\text{score}_{3,ab}$).

To quantify the completeness of the RRab selection at the faint end, we use 242 RRab stars from the Draco dSph training set (see Section 4.1) that have $\text{score}_2 > 0.13$. The dashed line in Figure 8 shows the completeness of the selection (i.e., the fraction of recovered RRab stars) as a function of the threshold on $\text{score}_{3,ab}$. This completeness includes all losses due to initial data quality cuts, and classification cuts (i.e., $\text{score}_1 > 0.01$ and $\text{score}_2 > 0.13$). For convenience, we tabulate the purity and completeness in Table 2.

Above, we have used the Stripe 82 sample to measure the purity, and the Draco sample to measure the completeness. We did so because the S82 sample covers a large area and thus contains a more representative sample of contaminants that we may expect to encounter elsewhere on the sky. The Draco sample was used because it contains more faint ($r \sim 20$ mag) RR Lyrae stars than the Stripe 82 sample, and thus the estimate of the completeness has a lower Poisson noise.

Given the sparseness and the multi-band nature of PS1 data, it is remarkable that our selection method can deliver samples of RRab stars that are $\sim 90\%$ pure and $\sim 80\%$ complete (e.g., for $\text{score}_{3,ab} > 0.8$), even at distances as far as ~ 80 kpc from the Sun. This raises hope for even better performance at the bright end.

To measure the purity and completeness at the bright end, we select objects from the SDSS Stripe 82 training set with $\text{score}_2 > 0.13$ and $\langle r \rangle < 18.5$ mag (i.e., within ~ 40 kpc from

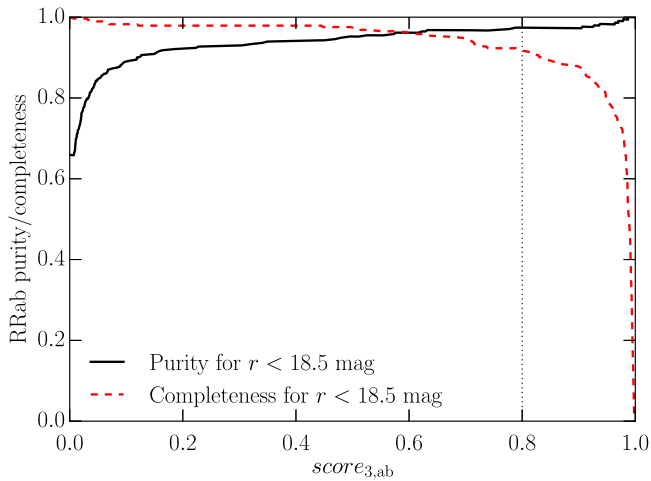


Figure 9. Expected purity and completeness of selected samples of bright RRab stars, as a function of the threshold on the RRab classification score, $score_{3,ab}$. A threshold of 0.8 (vertical dotted line) returns an RRab sample that is 97% pure and 92% complete within ~ 40 kpc ($\langle r \rangle \lesssim 18.5$ mag).

Table 3
Expected RRab Selection Purity and Completeness within ~ 40 kpc
($\langle r \rangle < 18.5$ mag)

Threshold on $score_{3,ab}$	Purity	Completeness
0.00	0.66	1.00
0.05	0.85	0.99
0.10	0.89	0.98
0.15	0.91	0.98
0.20	0.92	0.98
0.25	0.93	0.98
0.30	0.93	0.98
0.35	0.94	0.98
0.40	0.94	0.98
0.45	0.94	0.98
0.50	0.95	0.97
0.55	0.96	0.97
0.60	0.96	0.96
0.65	0.97	0.95
0.70	0.97	0.95
0.75	0.97	0.92
0.80	0.97	0.92
0.85	0.97	0.89
0.90	0.97	0.88
0.95	0.98	0.80

Note. A machine readable version of this table, with a 0.01 step in threshold, is available in the electronic edition of the journal.

(This table is available in its entirety in machine-readable form.)

the Sun). We use the $\langle r \rangle = 18.5$ mag brightness cut because the vast majority of halo RR Lyrae stars are located within that magnitude range (Sesar et al. 2010). The relevant curves are plotted in Figure 9 and tabulated in Table 3.

Finally, the purity and completeness curves characterizing the selection of bright RRc stars are shown in Figure 10 and tabulated in Table 4. Figure 10 shows that the selection of pure and complete samples of RRc stars is more challenging, both due to the lower amplitude of the pulsation and to the contamination by contact binaries with similar sinusoidal light curves. Nonetheless, it is still possible to produce samples that

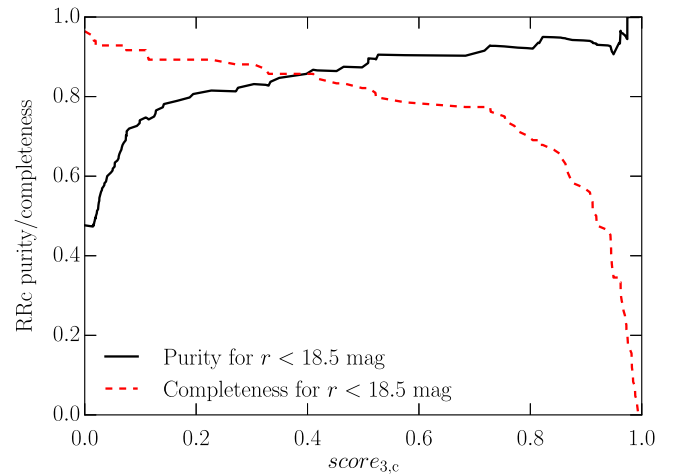


Figure 10. Expected purity and completeness of selected samples of bright RRc stars, as a function of the threshold on the RRc classification score, $score_{3,c}$. In comparison with RRab stars (Figure 9), the selection of pure and complete samples of RRc stars is more challenging.

Table 4
Expected RRc Selection Purity and Completeness within ~ 40 kpc
($\langle r \rangle < 18.5$ mag)

Threshold on $score_{3,c}$	Purity	Completeness
0.00	0.48	0.96
0.05	0.61	0.93
0.10	0.74	0.92
0.15	0.78	0.89
0.20	0.81	0.89
0.25	0.81	0.89
0.30	0.83	0.88
0.35	0.85	0.86
0.40	0.86	0.86
0.45	0.86	0.83
0.50	0.88	0.82
0.55	0.90	0.79
0.60	0.90	0.78
0.65	0.90	0.78
0.70	0.91	0.77
0.75	0.93	0.75
0.80	0.92	0.70
0.85	0.95	0.66
0.90	0.94	0.57
0.95	0.91	0.35

Note. A machine readable version of this table, with a 0.01 step in threshold, is available in the electronic edition of the journal.

(This table is available in its entirety in machine-readable form.)

are over 80% complete and pure within ~ 40 kpc from the Sun. We do not discuss RRc stars further because they are less numerous than RRab stars (by a factor of three), and thus are of lesser importance for Galactic studies.

5.2. RRab Selection Function

Given a position on the sky and the flux-averaged r -band magnitude of an RR Lyrae star, what is the probability of selecting that star using the PS1 data at hand? Characterizing this selection function is of obvious importance for studies of the Galactic structure, especially when modeling the number density distribution of stars (e.g., Bovy et al. 2012;

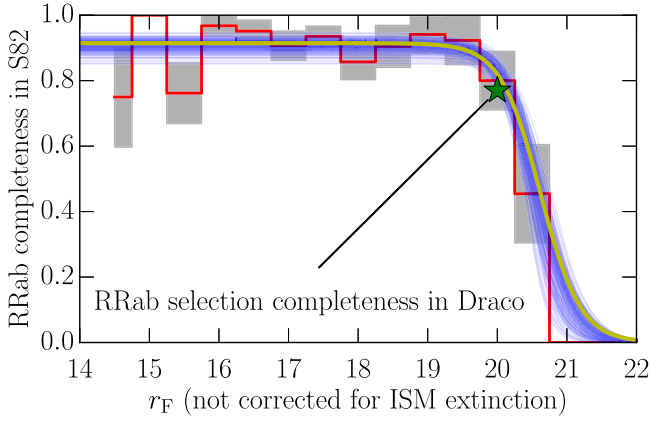


Figure 11. RRab selection function, or the completeness of the selection of RRab stars at high galactic latitudes ($|b| > 20^\circ$), as a function of the flux-averaged r -band magnitude (not corrected for interstellar matter extinction). The red curve shows the ratio of the number of selected (n_{sel} , $\text{score}_{3,\text{ab}} > 0.8$) and all RRab stars from the SDSS Stripe 82 training set (n_{all}), in 0.5 mag wide bins. The shaded region shows the standard deviation of the bin height, $\sqrt{n_{\text{sel}}(1 - n_{\text{sel}}/n_{\text{all}})/n_{\text{all}}}$, computed based on the binomial distribution. For comparison, the star symbol shows the fraction of recovered RRab stars in the Draco dSph (see Figure 8). The thick yellow line shows the best-fit logistic curve (Equation (7), $L = 0.91$, $k = 4.1$, $x_0 = 20.57$ mag), and the thin blue lines illustrate the uncertainty of the fit (see Section 5.2 for details).

Xue et al. 2015). In this section, we restrict ourselves to characterizing the selection function of RRab stars at high galactic latitudes ($|b| > 20^\circ$), because (1) they are three times more numerous than RRC stars, and (2) at a given purity, they can be recovered at a much higher rate than RRC stars (compare Figure 9 versus Figure 10). Characterizing the selection function at low galactic latitudes would require an appropriate training set (Stripe 82 and Draco are both located at $|b| > 20^\circ$).

We assume that the selection function S depends only on the flux-averaged r -band magnitude of an RRab star (not corrected for interstellar extinction), and not on its position (i.e., $S(r_{\text{F}})$). This is a reasonable assumption given the uniformity of dust extinction away from the Galactic plane, and the uniformity of PS1 multi-epoch coverage. The selection function will also depend on the threshold imposed on the final classification score, $\text{score}_{3,\text{ab}}$. For the sake of simplicity, we only consider the case when $\text{score}_{3,\text{ab}} > 0.8$, as this selection cut returns a sample that is appropriate for many studies (90% purity and 80% completeness, even at the faint end; see Section 5.1). By assuming spatial independence, we can now use the SDSS Stripe 82 and Draco training sets to determine the PS1 3π selection function of RRab stars at high galactic latitudes. The result is illustrated in Figure 11.

We find that the RRab selection function is approximately constant at $\sim 90\%$ for $r_{\text{F}} \lesssim 20$ mag, after which it steeply drops to zero at $r_{\text{F}} \sim 21.5$ mag. To characterize the selection function, we construct a simple probabilistic model.

There are 577 RRab stars in our training set (in SDSS Stripe 82 and Draco), of which 483 pass the $\text{score}_{3,\text{ab}} > 0.8$ selection cut. With each RRab star, we associate a $(r_{\text{F},n}, s_n)$ pair of values, where $s_n = 1$ if the star is selected, otherwise $s_n = 0$ ($r_{\text{F},n}$ is the star's extinguished flux-averaged r -band magnitude). We denote the full data set of 577 pairs of values as $\mathbf{d}_n = \{r_{\text{F},n}, s_n\}$.

The likelihood of this data set is given by

$$p(\mathbf{d}_n|L, k, x_0) = \prod_{n=1}^{577} p(s_n|S(r_{\text{F},n}|L, k, x_0)), \quad (6)$$

where $p(s_n|S(r_{\text{F},n}|L, k, x_0))$ is the Bernoulli probability mass function with success probability given by the selection function, $S(r_{\text{F},n}|L, k, x_0)$.

To model the selection function, we use the logistic curve

$$S(r_{\text{F},n}|L, k, x_0) = \frac{L}{1 + \exp(-k(r_{\text{F},n} - x_0))}, \quad (7)$$

where L is the curve's maximum value, k is the steepness of the curve, and x_0 is the magnitude at which the completeness drops to 50%.

The probability of this model given data \mathbf{d}_n is then

$$p(L, k, x_0|\mathbf{d}_n) = p(\mathbf{d}_n|L, k, x_0)p(L, k, x_0), \quad (8)$$

where $p(L, k, x_0)$ is the prior probability of model parameters. We impose uniform priors such that $0.4 \leq S(r_{\text{F}} < 18.5) \leq 1.0$ (i.e., completeness within 40 kpc is between 40% and 100%) and $S(r_{\text{F}} > 22) = 0$.

We explore the probability of various model parameters using the Goodman & Weare (2010) Affine Invariant Markov chain Monte Carlo (MCMC) ensemble sampler, as implemented in the `emcee` package²² (v2.2.1, Foreman-Mackey et al. 2013). The most probable model of the selection function (yellow curve) is shown in Figure 11, with the best-fit logistic curve (Equation (7)) being $L = 0.91$, $k = 4.1$, $x_0 = 20.57$ mag. To illustrate the uncertainty in the model, we also plot the curves associated with 200 randomly selected models from the posterior distribution (thin blue lines).

6. PS1 Catalog of RR Lyrae Stars

We have applied the above multi-step-selection procedure to about 500 million PS1 objects that pass PS1 data quality cuts (Section 2), and have calculated final RRab and RRC classification scores ($\text{score}_{3,\text{ab}}$, and $\text{score}_{3,\text{c}}$) for 240,000 objects. We report their positions, distances, PS1 photometry, and classification scores in Table 5. A total of $\sim 400,000$ CPU hours of super-computing time was used to process all of the data and calculate the final classification scores. Below, we illustrate some properties of this sample and leave a more detailed analysis of the distribution of RR Lyrae stars in the Galactic halo for future studies.

To illustrate the coverage of the PS1 catalog of RR Lyrae stars, we have selected a sample of $\sim 45,000$ highly probable RRab stars ($\text{score}_{3,\text{ab}} > 0.8$, expected purity of 90% and completeness of $\sim 80\%$ at 80 kpc), and have plotted their angular distribution in Figure 12.

The leading arm of the Sagittarius tidal stream (Ibata et al. 2001) and four Milky Way satellite galaxies are most easily discernible features in Figure 12. However, another notable feature is an almost complete absence of high probability RRab stars (i.e., those that have $\text{score}_{3,\text{ab}} > 0.8$) in regions with high ISM extinction (e.g., $E(B - V) > 1$). Improperly dereddened photometry is the most likely reason for this lack of high probability RRab stars at low galactic latitudes.

²² <http://dan.iel.fm/emcee/current/>

Table 5
PS1 Catalog of RR Lyrae Stars

R.A.	Decl.	score _{3,ab} ^a	score _{3,c} ^a	DM ^b	Period	ϕ_0 ^c	$A'_g \dots A'_z$ ^d	$g' \dots z'$ ^e	$T_g \dots T_z$ ^f	$g_F \dots z_F$ ^g	$E(B - V)$ ^h
180.39736	-0.23480	0.57	0.02	15.44	0.671302	-0.40301	0.22...0.11	15.96...15.77	100...100	16.09...15.82	0.020
179.98457	-0.00105	0.99	0.00	15.90	0.471807	-0.18385	1.32...0.68	15.72...16.12	120...113	16.75...16.58	0.012

Notes. For collaborations on projects and earlier access to the PS1 catalog of RR Lyrae stars, please contact the first author.

^a Final R Rab and R Rc classification scores.

^b Distance modulus calculated using the flux-averaged i_{p1} -band magnitude and Equation (5). The uncertainty in distance modulus is $0.06(\text{rnd}) \pm 0.03(\text{sys})$ mag for R Rab stars. This distance modulus may be biased and more uncertain for R Rc stars.

^c Phase offset (see Equation (2)).

^d Best-fit amplitude (e.g., $A'_g = FA_g$; see Equation (1)).

^e Best-fit magnitude at $\phi = 0$, corrected for dust extinction using extinction coefficients of Schlafly & Finkbeiner (2011) and the dust map of Schlafly et al. (2014; e.g., $g' = g_0 - r_0 + r'$; see Equation (1)).

^f Best-fit template ID number (see Section 3.1 and Table 2 of Sesar et al. 2010).

^g Flux-averaged magnitude, corrected for dust extinction using extinction coefficients of Schlafly & Finkbeiner (2011) and the dust map of Schlafly et al. (2014).

^h Reddening adopted from the Schlafly et al. (2014) dust map.

Briefly, when dereddening photometry, we assume that all sources are located behind the dust layer. At low galactic latitudes, this may not always be true because sources may be embedded in the dust layer. After dereddening, the photometry of such sources will be overcorrected for extinction, and their optical PS1 colors will be shifted blueward from their dust-free values. In addition, improperly dereddened light curves will not be well-fit by multi-band templates (Section 3.2). As a result of these effects, true RR Lyrae stars may look like non-RR Lyrae objects, and may end up with low score_{3,ab} values.

The lack of high probability R Rab stars at low galactic latitudes also demonstrates the resilience of the classifier to contamination. Due to the increase in stellar number density, and the fact that some fraction of stars will be incorrectly tagged as RR Lyrae stars, one would naively expect for the density of objects tagged as RR Lyrae stars to increase toward the Galactic plane. However, no such increase is observed in Figure 12. The features extracted during multi-band template fitting are most likely responsible for this resilience, as even a significant increase in the number of contaminants is not sufficient to produce objects that match multi-band light-curve characteristics of RR Lyrae stars.

Finally, to illustrate the efficiency of the final multiclass classifier (Section 4.6) at separating R Rab and R Rc stars, we show their distribution in the period versus the r_{p1} -band amplitude diagram (Figure 13).

7. Discussion and Summary

In this paper, we have explored with what fidelity RR Lyrae stars can be identified in multi-epoch, asynchronous multi-band photometric data. We have done this for the specific case of the PS1 3π light curves, which are very sparse; $\lesssim 12$ epochs per band over 3000 typical RR Lyrae pulsation periods (i.e., four years). To identify RR Lyrae stars, we have employed, in particular, the fitting (and period phasing) of very specific empirical RR Lyrae light curves, and have utilized supervised machine-learning tools. While we have applied our selection only to this specific data set, many of the approaches described here will be applicable to other sparse, asynchronous multi-band data sets, such as those produced by the Dark Energy Survey (DES; Dark Energy Survey Collaboration et al. 2016) and the Large Synoptic Survey Telescope (LSST; Ivezić et al. 2008). For example, for its Galactic plane sub-survey, LSST is

currently planning to obtain 12 observations over four years in each of its *ugrizy* bandpasses (i.e., 30 observations per band over 10 years; Ž. Ivezić 2017, private communication), making its data set very similar to the one produced by the PS1 3π survey. Compared to PS1, however, LSST will go deeper by at least 2 mag in *izy* bands, allowing us to select RR Lyrae stars and study old stellar populations close to the Galactic plane, even at the far side of the Galaxy.

We demonstrated that we can precisely and accurately measure the periods (to within 2 s) for the vast majority of RR Lyrae within PS1 3π , extending to distance moduli of ~ 20 , or ~ 100 kpc. The high precision of the period determination may seem surprising at face-value, but is owed to the long time baseline: a 2 s/period difference causes a cumulative light-curve shift of 90 minutes across 3000 pulsation periods. Accurate periods are crucial for calculating the phase of spectroscopic observations, and for transforming the observed radial velocity to the center-of-mass velocity needed for kinematic studies (Sesar 2012). The ephemerides (i.e., periods and phase offsets) provided by our catalog can thus be readily used to turn RR Lyrae stars observed by current (e.g., SDSS-IV/TDSS; Ruan et al. 2016) and upcoming multi-object spectroscopic (MOS) surveys (e.g., *Gaia*, WEAVE, DESI; Perryman et al. 2001; Levi et al. 2013; Dalton et al. 2014) into precise kinematic tracers of the halo structure and substructure (i.e., stellar streams). With a density of 1 deg^{-1} , PS1 RR Lyrae stars represent a unique “piggyback” project for MOS surveys, with a potentially high impact and certainly low cost (~ 1 target per MOS field).

Using these light-curve fits as one (crucial) feature in a supervised classification of RR Lyrae, we showed that we can—at least at high Galactic latitudes—construct a sample of $\sim 45,000$ R Rab stars that has 90% purity and 80% completeness, even at 80 kpc from the Sun. In comparison with previous catalogs, our sample is deeper than the SDSS Stripe 82 sample of Sesar et al. (2010), while covering more of the sky than the CRTS sample of Drake et al. (2013). The PS1 3π data and the classification presented here even allow for a quite reliable separation of R Rab and R Rc type of RR Lyrae stars, as shown by the period-amplitude diagram (Figure 13)

All this opens up many avenues in exploring the Galactic halo. With its second data release (DR2) expected in April 2018, the *Gaia* astrometric mission will provide unprecedented

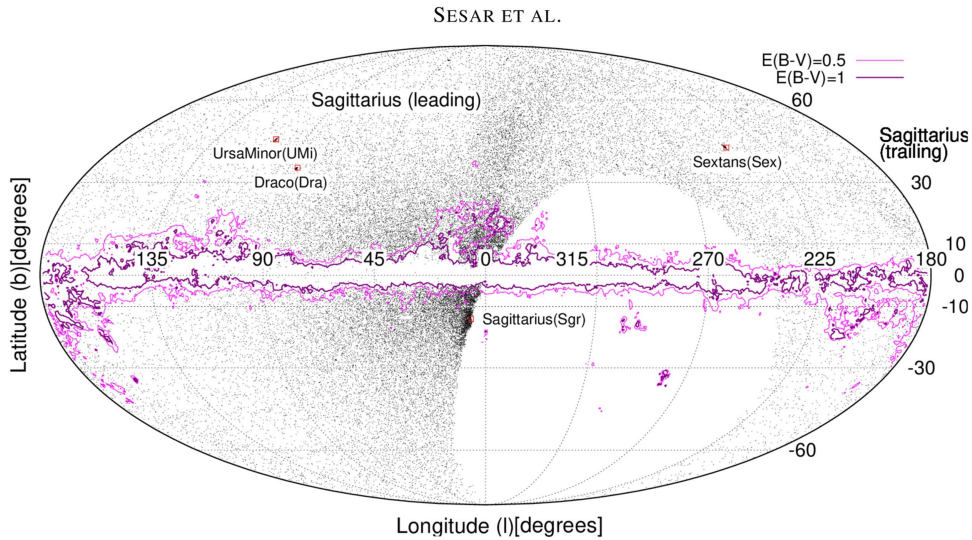


Figure 12. Distribution of $\sim 45,000$ highly probable RRAb stars ($\text{score}_{3,ab} > 0.8$, expected purity of 90% and completeness of $\sim 80\%$ at 80 kpc), shown in Mollweide projection of Galactic coordinates. A contour plot of the reddening-based $E(B - V)$ dust map (Schlafly et al. 2014) is overlaid, as well as the positions of four Milky Way dwarf satellite galaxies. The locations of the leading and trailing arms of the Sagittarius tidal stream are also indicated.

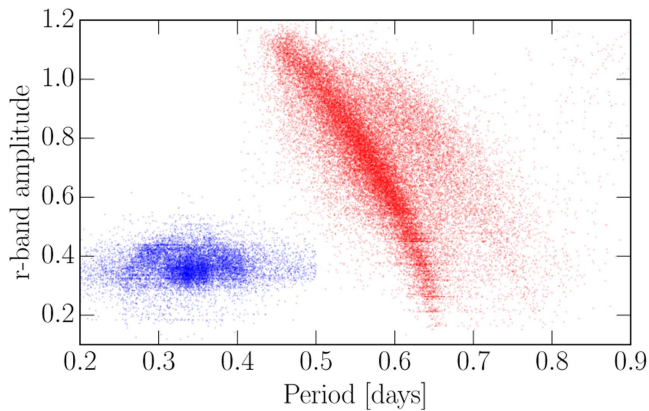


Figure 13. Distribution of highly likely RRAb (red; $\text{score}_{3,ab} > 0.8$) and RRc stars (blue; $\text{score}_{3,c} > 0.55$) in the period vs. r_{P1} -band amplitude diagram. Note the well-defined Oosterhoff I locus of RRAb stars (Oosterhoff 1939; Catelan 2009), and the less populated Oosterhoff II locus shifted to longer periods (along the lines of constant amplitude). The apparent clumps of RRc stars are likely caused by period aliasing (e.g., the one-day beat frequency alias, see the top panel of Figure 5).

proper motions for PS1 RR Lyrae stars brighter than $V \sim 20$ mag, but no competitive distance information (beyond a few kiloparsecs). Having precise distances is crucial for measuring tangential velocities,²³ and thus the Galactic potential, as the uncertainty in tangential velocity increases proportionally with the uncertainty in distance. Therefore, it is particularly remarkable that, using PS1 data and a period–absolute magnitude relation, we can measure distances to RR Lyrae stars with a precision of $\sim 3\%$, even for stars at 100 kpc from the Sun.

An important avenue to explore with the resulting RR Lyrae catalog is the question of RR Lyrae at low latitudes (covered by PS1). These objects open up the possibility to explore the oldest portion of the Galactic disk. At low latitudes, the selection function of the sample will, however, be considerably

more complicated, warranting careful testing and characterization beyond the scope of this paper.

B.S., N.H., and H.-W.R. acknowledge funding from the European Research Council under the European Unions Seventh Framework Programme (FP 7) ERC Grant Agreement n. [321035]. H.-W.R. acknowledges support of the Miller Institute at UC Berkeley through a visiting professorship during the completion of this work. We thank the anonymous referee for the thorough review, positive comments, and constructive remarks on this manuscript. The Pan-STARRS1 Surveys (PS1) have been made possible through contributions by the Institute for Astronomy, the University of Hawaii, the Pan-STARRS Project Office, the Max-Planck Society and its participating institutes, the Max Planck Institute for Astronomy, Heidelberg, and the Max Planck Institute for Extraterrestrial Physics, Garching, The Johns Hopkins University, Durham University, the University of Edinburgh, the Queen’s University Belfast, the Harvard-Smithsonian Center for Astrophysics, the Las Cumbres Observatory Global Telescope Network Incorporated, the National Central University of Taiwan, the Space Telescope Science Institute, and the National Aeronautics and Space Administration under Grant No. NNX08AR22G issued through the Planetary Science Division of the NASA Science Mission Directorate, the National Science Foundation Grant No. AST-1238877, the University of Maryland, Eotvos Lorand University (ELTE), and the Los Alamos National Laboratory.

Appendix A

Features Extracted From Phased Multi-band Light Curves

In this section, we describe features extracted from phased multi-band light curves.

1. *rrab*: this flag has a value of 1 if the best-fit template is of type *ab*, and 0 otherwise.
2. *period*: the best-fit period associated with the best-fit template (i.e., the multi-band template with the smallest χ^2 value).
3. χ^2 and χ^2_{dof} : goodness-of-fit χ^2 and χ^2 per degree of freedom associated with the best-fit template.

²³ Radial velocities of RR Lyrae stars are straightforward to measure (Sesar 2012).

4. L_{\max} and $L_{\max, \text{dof}}$: to calculate these two values, we sort χ^2 and χ^2_{dof} values (measured during the first fitting run) into 1 s wide bins in period (e.g., see Figure 3), and for each bin calculate $L(\text{bin}) = \sum \exp(-\chi^2/2)$ and $L_{\text{dof}}(\text{bin}) = \sum \exp(-\chi^2_{\text{dof}}/2)$, where the sum goes over all χ^2 and χ^2_{dof} in a bin. Then, $L_{\max} = \max(L(\text{bin}))$ and $L_{\max, \text{dof}} = \max(L_{\text{dof}}(\text{bin}))$.
5. Stetson J index: we phase a PS1 multi-band light curve using its best-fit period, and calculate normalized residuals $\delta(\text{phase}) = (m_n - m(\text{phase}(t_n, \text{period}, \phi)))/\sigma_{m_n}$, where m_n and σ_{m_n} are the magnitude and uncertainty of the n th observation in the $m = g, r, i, z$ band, and $m(\text{phase}(t_n, \text{period}, \phi))$ is the magnitude predicted by the best-fit model at the phase of the n th observation. The δ values are then sorted by phase, and used to calculate the Stetson J index (see Section 2 of Stetson 1996).
6. Entropy of the phase light curve S_{phased} : We phase a PS1 multi-band light curve using its best-fit period, bin observations into 20 phase bins (ignoring the band), and calculate $S_{\text{phased}} = -\sum_{\text{bin}} (N_{\text{bin}}/N_{\text{tot}}) \ln(N_{\text{bin}}/N_{\text{tot}})$, where N_{bin} is the number of observations in a phase bin, and N_{tot} is the total number of observations in the light curve.
7. ΔP : the difference between the best-fit period and the period associated with the best-fit multi-band template (i.e., the corresponding RR Lyrae star in SDSS Stripe 82; listed in Table 2 of Sesar et al. 2010).
8. $\Delta(g-r)_0$, $\Delta(r-i)_0$, $\Delta(r-z)_0$: the difference between the color measured from the best-fit model at $\text{phase} = 0$, and the color measured from the best-fit template. Since the fitting parameter F allows the amplitudes in each band to vary by up to 20% with respect to amplitudes associated with the best-fit multi-band template, the above differences do not have to be zero.
9. $(g-r)_0$, $(r-i)_0$, $(r-z)_0$: color at $\text{phase} = 0$, measured from the best-fit model.
10. FA_g , FA_r , FA_i , FA_z : best-fit amplitudes of *griz* bands, where A_m is the amplitude associated with the $m = g, r, i, z$ band of the best-fit multi-band template, and F is a fitting parameter.

Appendix B

Constraining Period–Absolute Magnitude–Metallicity Relations for PS1 Bands

To constrain Equation (4), we use average (apparent) PS1 *griz* magnitudes of 55 RR Lyrae stars located in five Galactic globular clusters that have metallicities ranging from -1.02 dex to -2.37 dex (on the Carretta et al. 2009 scale and taken from the Harris 1996, 2010 catalog; NGCs 6171, 5904, 4590, 6341, and 7078). The distance moduli of these globular clusters were measured by Sollima et al. (2006, see their Table 4) using a period–luminosity–metallicity (PLZ) relation whose zero point of -1.05 ± 0.13 mag was constrained using the Hubble Space Telescope (HST) parallax of the star RR Lyrae. Using HST parallaxes of five RR Lyrae stars, Benedict et al. (2011) measured the zero point of the Sollima et al. (2006) PLZ relation to be -1.08 ± 0.03 mag. Thus, for distance moduli of 55 RR Lyrae stars, we adopt the distance moduli of their globular clusters (see Table 4 of Sollima et al. 2006), shifted by 0.03 mag to match the more precise zero point measured by

Benedict et al. (2011). As the systematic uncertainty in distance moduli of calibration RR Lyrae stars, we assume a value of 0.03 mag.

We measure average (apparent) *griz* magnitudes of RR Lyrae stars by first converting their best-fit *griz* model light curves (e.g., see sold lines in Figure 3) from magnitude to flux units. The model curves in flux units are then integrated over a pulsation cycle, and the integrated fluxes are expressed in units of magnitude. Hereafter, we call these flux-averaged magnitudes (g_F , r_F , etc.), and distinguish them from uncertainty-weighted average magnitudes ($\langle g \rangle$, $\langle r \rangle$, etc.).

Given the above data, we wish to calculate the probability $p(\theta^b | \mathcal{D}^b)$ of some set of model parameters $\theta^b = \{\alpha^b, \beta^b, M_{\text{ref}}^b, \sigma_M^b\}$ for the $b = g, r, i, z$ PS1 band, where $\mathcal{D}^b = \{\mathbf{d}_k^b\}$ is the full data set of 55 stars, and $\mathbf{d}_k^b = \{\text{DM}, [\text{Fe}/\text{H}], \log_{10}(P), b_F\}$ is the data set associated with the k th star that contains its distance modulus and metallicity (assumed to be the same as that of the cluster), period, and flux-averaged PS1 apparent magnitudes.

Using the relations of conditional probability, the probability $p(\theta^b | \mathcal{D}^b)$ can be written as

$$p(\theta^b | \mathcal{D}^b) \propto p(\mathcal{D}^b | \theta^b) p(\theta^b), \quad (9)$$

where $p(\theta^b)$ is the prior probability of model parameters, and

$$p(\mathcal{D}^b | \theta^b) = \prod_k p(\mathbf{d}_k^b | \theta^b) \quad (10)$$

is the likelihood of the full data set given some values of model parameters, θ^b .

To make the fitting of PLZ relations robust to possible outliers (e.g., due to an incorrectly measured period, which may happen for 13% of RRab stars; Section 3.2), we model the likelihood of the k th data set using a mixture model (see Section 3 and Equation (17) of Hogg et al. 2010)

$$p(\mathbf{d}_k^b | \theta^b) = (1 - A_{\text{out}}^b) p_{\text{in}}(\mathbf{d}_k^b | \theta_{\text{in}}^b) + A_{\text{out}}^b p_{\text{out}}(\mathbf{d}_k^b | \theta_{\text{out}}^b), \quad (11)$$

where p_{in} and p_{out} are the likelihoods of drawing the data set \mathbf{d}_k^b from the inlier or outlier distribution, respectively, and A_{out}^b is the mixing proportion.

The likelihood of drawing the data set \mathbf{d}_k^b from the inlier distribution is

$$p_{\text{in}}(\mathbf{d}_k^b | \theta_{\text{in}}^b) = \mathcal{N}(b_F - \text{DM} | M^b, \sigma_M^b), \quad (12)$$

where $\theta_{\text{in}}^b = \{\alpha^b, \beta^b, M_{\text{ref}}^b, \sigma_M^b\}$, b_F is the apparent flux-averaged magnitude in the $b = g, r, i, z$ PS1 band, M^b is the absolute magnitude in the b band

$$M^b = \alpha^b \log_{10}(P/P_{\text{ref}}) + \beta^b ([\text{Fe}/\text{H}] - [\text{Fe}/\text{H}]_{\text{ref}}) + M_{\text{ref}}^b, \quad (13)$$

and

$$N(x | \mu, \sigma) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}. \quad (14)$$

The likelihood of drawing the data set \mathbf{d}_k^b from the outlier distribution is defined as

$$p_{\text{out}}(\mathbf{d}_k^b | \theta_{\text{out}}^b) = \mathcal{N}(b_F - \text{DM} | M_{\text{out}}^b, \sigma_{\text{out}}^b), \quad (15)$$

where M_{out}^b has the same form as Equation (4), but different parameters, $\theta_{\text{out}}^b = \{\alpha_{\text{out}}^b, \beta_{\text{out}}^b, M_{\text{ref, out}}^b, \sigma_{\text{out}}^b\}$. Most importantly, the distribution of outliers is assumed to be wider than the distribution of inliers, that is, $\sigma_{\text{out}}^b > \sigma_M^b$.

Before we can calculate probability $p(\theta^b | \mathcal{D}^b)$, we need to define the prior probabilities of model parameters, $p(\theta^b)$. For the α^b parameter and r_{p1} and i_{p1} bands, we adopt informative Gaussian priors based on the slopes of PLZ relations in the globular cluster M4 (Braga et al. 2015): $p(\alpha^r | r_{p1}) = \mathcal{N}(\alpha^r - 1.5, 0.1)$ and $p(\alpha^i | i_{p1}) = \mathcal{N}(\alpha^i - 1.72, 0.07)$. For σ_M^b and σ_{out}^b , we adopt Jeffreys log-uniform priors (Jaynes 1968), and require that $\sigma_{\text{out}}^b > \sigma_M^b$. Uniform priors are adopted for the remaining model parameters and bands.

To efficiently explore the parameter space, we use the Goodman & Weare (2010) Affine Invariant MCMC Ensemble sampler as implemented in the `emcee` package²⁴ (v2.2.1, Foreman-Mackey et al. 2013). We use 200 walkers and obtain convergence²⁵ after a short burn-in phase of 300 steps per walker. The chains are then evolved for another 2000 steps, and the first 300 (burn-in) steps are discarded.

To describe the marginal posterior distributions of individual model parameters, we measure the median, the difference between the 84th percentile and the median, and the difference between the median and the 16th percentile of each marginal posterior distribution (for a Gaussian distribution, these differences are equal to ± 1 standard deviation). We report these values in Table 1 (Section 3.3). We do not characterize parameters of the outlier distribution, and instead simply marginalize over them (i.e., we consider them to be nuisance parameters).

References

- Abbas, M., Grebel, E. K., Martin, N. F., et al. 2014, *AJ*, 148, 8
- Baker, M., & Willman, B. 2015, *AJ*, 150, 160
- Beaton, R. L., Freedman, W. L., Madore, B. F., et al. 2016, *ApJ*, 832, 210
- Belokurov, V., Evans, N. W., Bell, E. F., et al. 2007, *ApJL*, 657, L89
- Belokurov, V., Koposov, S. E., Evans, N. W., et al. 2014, *MNRAS*, 437, 116
- Benedict, G. F., McArthur, B. E., Feast, M. W., et al. 2011, *AJ*, 142, 187
- Bernard, E. J., Ferguson, A. M. N., Schlafly, E. F., et al. 2016, *MNRAS*, 463, 1759
- Bishop, C. M. 2006, *Pattern Recognition and Machine Learning* (Berlin: Springer)
- Bonanos, A. Z., Stanek, K. Z., Szentgyorgyi, A. H., Sasselov, D. D., & Bakos, G. Á. 2004, *AJ*, 127, 861
- Bovy, J., Rix, H.-W., Liu, C., et al. 2012, *ApJ*, 753, 148
- Boylan-Kolchin, M., Bullock, J. S., & Kaplinghat, M. 2011, *MNRAS*, 415, L40
- Braga, V. F., Dall’Ora, M., Bono, G., et al. 2015, *ApJ*, 799, 165
- Carrera, R., Aparicio, A., Martínez-Delgado, D., & Alonso-García, J. 2002, *AJ*, 123, 3199
- Caretta, E., Bragaglia, A., Gratton, R., D’Orazi, V., & Lucatello, S. 2009, *A&A*, 508, 695
- Catelan, M. 2009, *Ap&SS*, 320, 261
- Catelan, M., Pritzl, B. J., & Smith, H. A. 2004, *ApJS*, 154, 633
- Chambers, K. C. 2011, *BAAS*, 43, 113.01
- Chen, T., & Guestrin, C. 2016, arXiv:1603.02754
- Dalton, G., Trager, S., Abrams, D. C., et al. 2014, *Proc. SPIE*, 9147, 91470L
- Dark Energy Survey Collaboration, Abbott, T., Abdalla, F. B., et al. 2016, *MNRAS*, 460, 1270
- Drake, A. J., Catelan, M., Djorgovski, S. G., et al. 2013, *ApJ*, 763, 32
- Dubath, P., Rimoldini, L., Süveges, M., et al. 2011, *MNRAS*, 414, 2602
- Elorrieta, F., Eyheramendy, S., Jordán, A., et al. 2016, *A&A*, 595, A82
- Foreman-Mackey, D., Hogg, D. W., Lang, D., & Goodman, J. 2013, *PASP*, 125, 306
- Friedman, J. H. 2001, *AnSta*, 29, 1189
- Goodman, J., & Weare, J. 2010, *Commun. Appl. Math. Comput. Sci.*, 5, 65
- Hargis, J. R., Willman, B., & Peter, A. H. G. 2014, *ApJL*, 795, L13
- Harris, W. E. 1996, *AJ*, 112, 1487
- Harris, W. E. 2010, arXiv:1012.3224
- Hastie, T., Tibshirani, R., & Friedman, J. 2009, *The Elements of Statistical Learning: Data Mining, Inference and Prediction* (2nd ed.; Berlin: Springer)
- Hernitschek, N., Schlafly, E. F., Sesar, B., et al. 2016, *ApJ*, 817, 73
- Hogg, D. W., Bovy, J., & Lang, D. 2010, arXiv:1008.4686
- Hopkins, P. F., Kereš, D., Oñorbe, J., et al. 2014, *MNRAS*, 445, 581
- Ibata, R., Irwin, M., Lewis, G. F., & Stolte, A. 2001, *ApJL*, 547, L133
- Ivezić, Ž., Sesar, B., Jurić, M., et al. 2008, *ApJ*, 684, 287
- Ivezic, Z., Tyson, J. A., Abel, B., et al. 2008, arXiv:0805.2366
- Jaynes, E. T. 1968, *ITSMC*, 4, 227
- Kaiser, N., Burgett, W., Chambers, K., et al. 2010, *Proc. SPIE*, 7733, 77330E
- Keller, S. C., Murphy, S., Prior, S., Da Costa, G., & Schmidt, B. 2008, *ApJ*, 678, 851
- Kinemuchi, K., Harris, H. C., Smith, H. A., et al. 2008, *AJ*, 136, 1921
- Koposov, S. E., Belokurov, V., Evans, N. W., et al. 2008, *ApJ*, 686, 279
- Koposov, S. E., Rix, H.-W., & Hogg, D. W. 2010, *ApJ*, 712, 260
- Lee, M. G., Yuk, I.-S., Park, H. S., Harris, J., & Zaritsky, D. 2009, *ApJ*, 703, 692
- Levi, M., Bebek, C., Beers, T., et al. 2013, arXiv:1308.0847
- Magnier, E. A., Liu, M., Monet, D. G., & Chambers, K. C. 2008, in *IAU Symp. 248, A Giant Step: from Milli- to Micro-arcsecond Astrometry*, ed. W. J. Jin, I. Platais, & M. A. C. Perryman (Cambridge: Cambridge Univ. Press), 553
- Marconi, M., Cignoni, M., Di Criscienzo, M., et al. 2006, *MNRAS*, 371, 1503
- Marconi, M., Coppola, G., Bono, G., et al. 2015, *ApJ*, 808, 50
- Miceli, A., Rest, A., Stubbs, C. W., et al. 2008, *ApJ*, 678, 865
- Mighell, K. J., & Burke, C. J. 1999, *AJ*, 118, 366
- Millman, K. J., & Aivazis, M. 2011, *CSE*, 13, 9
- Newberg, H. J., Willett, B. A., Yanny, B., & Xu, Y. 2010, *ApJ*, 711, 32
- Nikutta, R., Hunt-Walker, N., Nenkova, M., Ivezić, Ž., & Elitzur, M. 2014, *MNRAS*, 442, 3361
- Oliphant, T. E. 2007, *CSE*, 9, 10
- Oosterhoff, P. T. 1939, *Obs*, 62, 104
- Pedregosa, F., Varoquaux, G., Gramfort, A., et al. 2011, *Journal of Machine Learning Research*, 12, 2825
- Perryman, M. A. C., de Boer, K. S., Gilmore, G., et al. 2001, *A&A*, 369, 339
- Price-Whelan, A. M., Hogg, D. W., Johnston, K. V., & Hendel, D. 2014, *ApJ*, 794, 4
- Richards, J. W., Starr, D. L., Butler, N. R., et al. 2011, *ApJ*, 733, 10
- Ruan, J. J., Anderson, S. F., Green, P. J., et al. 2016, *ApJ*, 825, 137
- Sawala, T., Frenk, C. S., Fattahi, A., et al. 2016, *MNRAS*, 456, 85
- Schlafly, E. F., & Finkbeiner, D. P. 2011, *ApJ*, 737, 103
- Schlafly, E. F., Finkbeiner, D. P., Jurić, M., et al. 2012, *ApJ*, 756, 158
- Schlafly, E. F., Green, G., Finkbeiner, D. P., et al. 2014, *ApJ*, 789, 15
- Sesar, B. 2012, *AJ*, 144, 114
- Sesar, B., Banholzer, S. R., Cohen, J. G., et al. 2014, *ApJ*, 793, 135
- Sesar, B., Bovy, J., Bernard, E. J., et al. 2015, *ApJ*, 809, 59
- Sesar, B., Ivezić, Ž., Grammer, S. H., et al. 2010, *ApJ*, 708, 717
- Sesar, B., Grillmair, C. J., Cohen, J. G., et al. 2013, *ApJ*, 776, 26
- Sesar, B., Ivezić, Ž., Lupton, R. H., et al. 2007, *AJ*, 134, 2236
- Simon, J. D., Geha, M., Minor, Q. E., et al. 2011, *ApJ*, 733, 46
- Smith, H. 2004, *RR Lyrae Stars*, Cambridge Astrophysics (Cambridge: Cambridge Univ. Press)
- Sollima, A., Cacciari, C., & Valentini, E. 2006, *MNRAS*, 372, 1675
- Stetson, P. B. 1996, *PASP*, 108, 851
- Storn, R., & Price, K. 1997, *Journal of Global Optimization*, 11, 341
- Stubbs, C. W., Doherty, P., Cramer, C., et al. 2010, *ApJS*, 191, 376
- Süveges, M., Sesar, B., Váradi, M., et al. 2012, *MNRAS*, 424, 2528
- Tonry, J. L., Stubbs, C. W., Lykke, K. R., et al. 2012, *ApJ*, 750, 99
- Vanderplas, J. 2015, *gatspy: General tools for Astronomical Time Series in Python*, Zenodo, doi:10.5281/zenodo.14833
- VanderPlas, J. T., & Ivezić, Ž. 2015, *ApJ*, 812, 18
- Vivas, A. K., Zinn, R., Andrews, P., et al. 2001, *ApJL*, 554, L33
- Wang, J., Frenk, C. S., Navarro, J. F., Gao, L., & Sawala, T. 2012, *MNRAS*, 424, 2715
- Wright, E. L., Eisenhardt, P. R. M., Mainzer, A. K., et al. 2010, *AJ*, 140, 1868
- Xue, X.-X., Rix, H.-W., Ma, Z., et al. 2015, *ApJ*, 809, 144
- York, D. G., Adelman, J., Anderson, J. E., Jr., et al. 2000, *AJ*, 120, 1579

²⁴ <http://dan.iel.fm/emcee/current/>

²⁵ We checked for convergence of chains by examining the auto-correlation time of the chains per dimension.