DES-2018-0375
FERMILAB-PUB-18-680-AE

Draft version May 3, 2019
Typeset using LaTeX **twocolumn** style in AASTeX62

arXiv:1905.00428v1 [astro-ph.SR] 1 May 2019

Identification of RR Lyrae stars in multiband, sparsely-sampled data from the Dark Energy Survey using template fitting and Random Forest classification

K. M. Stringer,[1] J. P. Long,[2] L. M. Macri,[1] J. L. Marshall,[1] A. Drlica-Wagner,[3] C. E. Martínez-Vázquez,[4] A. K. Vivas,[4] K. Bechtol,[5,6] E. Morganson,[7,8] M. Carrasco Kind,[7,8] A. B. Pace,[1] A. R. Walker,[4] C. Nielsen,[1,9] T. S. Li,[3] E. Rykoff,[10,11] D. Burke,[10,11] A. Carnero Rosell,[12,13] E. Neilsen,[3] P. Ferguson,[1] S. A. Cantu,[1] J. L. Myron,[1] L. Strigari,[1] A. Farahi,[14] F. Paz-Chinchón,[7,8] D. Tucker,[3] Z. Lin,[2] D. Hatt,[15] J. F. Maner,[1] L. Plybon,[1] A. H. Riley,[1] E. O. Nadler,[15] T. M. C. Abbott,[4] S. Allam,[3] J. Annis,[3] E. Bertin,[16,17] D. Brooks,[18] E. Buckley-Geer,[3] J. Carretero,[19] C. E. Cunha,[10] C. B. D'Andrea,[20] L. N. da Costa,[12,13] J. De Vicente,[21] S. Desai,[22] P. Doel,[18] T. F. Eifler,[23,24] B. Flaugher,[3] J. Frieman,[3,15] J. García-Bellido,[25] E. Gaztanaga,[26,27] D. Gruen,[28,10,11] J. Gschwend,[12,13] G. Gutierrez,[3] W. G. Hartley,[18,29] D. L. Hollowood,[30] B. Hoyle,[31,32] D. J. James,[33] K. Kuehn,[34] N. Kuropatkin,[3] P. Melchior,[35] R. Miquel,[36,19] R. L. C. Ogando,[12,13] A. A. Plazas,[24] E. Sanchez,[21] B. Santiago,[37,12] V. Scarpine,[3] M. Schubnell,[38] S. Serrano,[26,27] I. Sevilla-Noarbe,[21] M. Smith,[39] R. C. Smith,[4] M. Soares-Santos,[40] F. Sobreira,[41,12] E. Suchyta,[42] M. E. C. Swanson,[7] G. Tarle,[38] D. Thomas,[43] V. Vikram,[44] B. Yanny[3]
(DES Collaboration)

[1] *George P. and Cynthia W. Mitchell Institute for Fundamental Physics and Astronomy, Department of Physics and Astronomy, Texas A&M University, College Station, TX 77843, USA*
[2] *Department of Statistics, 3143 TAMU, College Station, TX 77843, USA*
[3] *Fermi National Accelerator Laboratory, P. O. Box 500, Batavia, IL 60510, USA*
[4] *Cerro Tololo Inter-American Observatory, National Optical Astronomy Observatory, Casilla 603, La Serena, Chile*
[5] *Wisconsin IceCube Particle Astrophysics Center (WIPAC), Madison, WI 53703, USA*
[6] *Department of Physics, University of Wisconsin-Madison, Madison, WI 53706, USA*
[7] *National Center for Supercomputing Applications, 1205 West Clark St., Urbana, IL 61801, USA*
[8] *Department of Astronomy, University of Illinois, Urbana, IL 61801, USA*
[9] *Department of Physics, Purdue University, 1396 Physics Building, West Lafayette, Indiana 47907-1396*
[10] *Kavli Institute for Particle Astrophysics & Cosmology, P. O. Box 2450, Stanford University, Stanford, CA 94305, USA*
[11] *SLAC National Accelerator Laboratory, Menlo Park, CA 94025, USA*
[12] *Laboratório Interinstitucional de e-Astronomia - LIneA, Rua Gal. José Cristino 77, Rio de Janeiro, RJ - 20921-400, Brazil*
[13] *Observatório Nacional, Rua Gal. José Cristino 77, Rio de Janeiro, RJ - 20921-400, Brazil*
[14] *Department of Physics, Carnegie Mellon University, Pittsburgh, Pennsylvania 15312, USA*
[15] *Kavli Institute for Cosmological Physics, University of Chicago, Chicago, IL 60637, USA*
[16] *CNRS, UMR 7095, Institut d'Astrophysique de Paris, F-75014, Paris, France*
[17] *Sorbonne Universités, UPMC Univ Paris 06, UMR 7095, Institut d'Astrophysique de Paris, F-75014, Paris, France*
[18] *Department of Physics & Astronomy, University College London, Gower Street, London, WC1E 6BT, UK*
[19] *Institut de Física d'Altes Energies (IFAE), The Barcelona Institute of Science and Technology, Campus UAB, 08193 Bellaterra (Barcelona) Spain*
[20] *Department of Physics and Astronomy, University of Pennsylvania, Philadelphia, PA 19104, USA*
[21] *Centro de Investigaciones Energéticas, Medioambientales y Tecnológicas (CIEMAT), Madrid, Spain*
[22] *Department of Physics, IIT Hyderabad, Kandi, Telangana 502285, India*
[23] *Department of Astronomy/Steward Observatory, 933 North Cherry Avenue, Tucson, AZ 85721-0065, USA*
[24] *Jet Propulsion Laboratory, California Institute of Technology, 4800 Oak Grove Dr., Pasadena, CA 91109, USA*
[25] *Instituto de Fisica Teorica UAM/CSIC, Universidad Autonoma de Madrid, 28049 Madrid, Spain*
[26] *Institut d'Estudis Espacials de Catalunya (IEEC), 08034 Barcelona, Spain*
[27] *Institute of Space Sciences (ICE, CSIC), Campus UAB, Carrer de Can Magrans, s/n, 08193 Barcelona, Spain*
[28] *Department of Physics, Stanford University, 382 Via Pueblo Mall, Stanford, CA 94305, USA*
[29] *Department of Physics, ETH Zurich, Wolfgang-Pauli-Strasse 16, CH-8093 Zurich, Switzerland*
[30] *Santa Cruz Institute for Particle Physics, Santa Cruz, CA 95064, USA*
[31] *Max Planck Institute for Extraterrestrial Physics, Giessenbachstrasse, 85748 Garching, Germany*
[32] *Universitäts-Sternwarte, Fakultät für Physik, Ludwig-Maximilians Universität München, Scheinerstr. 1, 81679 München, Germany*

Corresponding author: Katelyn M. Stringer
kstringer@tamu.edu

[33] Harvard-Smithsonian Center for Astrophysics, Cambridge, MA 02138, USA
[34] Australian Astronomical Optics, Macquarie University, North Ryde, NSW 2113, Australia
[35] Department of Astrophysical Sciences, Princeton University, Peyton Hall, Princeton, NJ 08544, USA
[36] Institució Catalana de Recerca i Estudis Avançats, E-08010 Barcelona, Spain
[37] Instituto de Física, UFRGS, Caixa Postal 15051, Porto Alegre, RS - 91501-970, Brazil
[38] Department of Physics, University of Michigan, Ann Arbor, MI 48109, USA
[39] School of Physics and Astronomy, University of Southampton, Southampton, SO17 1BJ, UK
[40] Brandeis University, Physics Department, 415 South Street, Waltham MA 02453
[41] Instituto de Física Gleb Wataghin, Universidade Estadual de Campinas, 13083-859, Campinas, SP, Brazil
[42] Computer Science and Mathematics Division, Oak Ridge National Laboratory, Oak Ridge, TN 37831
[43] Institute of Cosmology and Gravitation, University of Portsmouth, Portsmouth, PO1 3FX, UK
[44] Argonne National Laboratory, 9700 South Cass Avenue, Lemont, IL 60439, USA

(Accepted April 30, 2019)

Submitted to The Astronomical Journal

ABSTRACT

Many studies have shown that RR Lyrae variable stars (RRL) are powerful stellar tracers of Galactic halo structure and satellite galaxies. The Dark Energy Survey (DES), with its deep and wide coverage ($g \sim 23.5$ mag in a single exposure; over 5000 deg^2) provides a rich opportunity to search for substructures out to the edge of the Milky Way halo. However, the sparse and unevenly sampled multiband light curves from the DES wide-field survey (median 4 observations in each of *grizY* over the first three years) pose a challenge for traditional techniques used to detect RRL. We present an empirically motivated and computationally efficient template fitting method to identify these variable stars using three years of DES data. When tested on DES light curves of previously classified objects in SDSS stripe 82, our algorithm recovers 89% of RRL periods to within 1% of their true value with 85% purity and 76% completeness. Using this method, we identify 5783 RRL candidates, $\sim$ 31% of which are previously undiscovered. This method will be useful for identifying RRL in other sparse multiband data sets.

Keywords: stars: variable stars: RR Lyrae — Galaxy: structure — Galaxy: halo

1. INTRODUCTION

RR Lyrae variable stars (RRL) are old (age > 10 Gyr) horizontal branch stars that pulsate with short periods (0.2 - 1.2 days). They have become one of the most widely used stellar tracers in Milky Way and Local Group studies. Thanks to the discovery of RR Lyrae itself (Pickering et al. 1901) and the subsequent studies of their pulsation (see King & Cox 1968 and Catelan & Smith 2015 for a review of the pioneering and current works in this field), these stars have well-understood period-luminosity-metallicity (*P-L-Z*) relations[1] (e.g., Cáceres & Catelan 2008; Marconi et al. 2015), making them excellent distance indicators, especially in the near-infrared bands. This, combined with their bright luminosities ($M_V \sim 0.6$) and advanced ages make RRL well-suited to trace discrete stellar populations (satellite galaxies, star clusters, and streams) within the Milky Way halo (e.g., Catelan et al. 2004; Vivas et al. 2004; Cáceres & Catelan 2008; Sesar et al. 2010; Stetson et al. 2014; Fiorentino et al. 2015).

Locating these stellar populations is crucial for testing the ΛCDM hierarchical model, which predicts that the haloes of large galaxies like the Milky Way are formed through the accretion and disruption of lower mass haloes (Bullock & Johnston 2005). Recent re-examinations of these simulations predict that the outer reaches of the stellar halo ($d \geq 100$ kpc) are primarily composed of the most recently accreted satellites and that thousands of RRL should be present in them (Sanderson et al. 2017). Once satellite galaxies and their disrupted remains are found, their distribution and properties can reveal valuable clues about the formation history, dark matter density profile, and mass of the Milky Way. While these objects are interesting in their own right, the statistical information about this sample is vital to place the Milky Way in a broader cosmological context.

[1] These are sometimes presented as Period-Luminosity-Color *(P-L-C)* relations.

Numerous Milky Way substructures have already been discovered. Eleven "classical" dwarf galaxies were known to orbit the Milky Way before 2005 (McConnachie 2012)[2]. Thanks to the advent of wide-field surveys such as the Sloan Digital Sky Survey (SDSS, York et al. 2000), the Panoramic Survey Telescope and Rapid Response System (Pan-STARRS, Chambers et al. 2016), and the Dark Energy Survey (DES, DES Collaboration 2016), over 40 new dwarf satellite candidates have been discovered (Willman et al. 2005a,b; Zucker et al. 2006a,b; Belokurov et al. 2006, 2007a, 2008, 2009, 2010; Grillmair 2006, 2009; Sakamoto & Hasegawa 2006; Irwin et al. 2007; Walsh et al. 2007; Kim et al. 2015; Bechtol et al. 2015; Koposov et al. 2015; Drlica-Wagner et al. 2015; Laevens et al. 2015a,b; Luque et al. 2016; Torrealba et al. 2016a; Luque et al. 2017; Torrealba et al. 2016b; Koposov et al. 2018; Torrealba et al. 2018b). In addition to galaxies that are still intact, tidal streams, the disrupted remains of satellite galaxies and globular clusters, have been discovered to be prevalent within the Milky Way halo (e.g., Odenkirchen et al. 2001; Newberg et al. 2002; Belokurov et al. 2006; Bernard et al. 2014; Koposov et al. 2014; Grillmair & Carlin 2016; Bernard et al. 2016; Balbinot et al. 2016; Shipp et al. 2018; Mateu et al. 2018). Besides these, additional large stellar overdensities populate the MW stellar halo with origins still unknown (e.g., Vivas et al. 2001; Newberg et al. 2002; Majewski et al. 2004; Rocha-Pinto et al. 2004; Sesar et al. 2007; Belokurov et al. 2007b; Sharma et al. 2010; Deason et al. 2014; Li et al. 2016; Pieres et al. 2017; Bergemann et al. 2018; Prudil et al. 2018).

Most of these satellites and streams were discovered as stellar overdensities in photometric catalogs (Willman 2010, and references therein). However, this detection method is biased against diffuse objects with low surface brightness ($\mu_{V,0} \gtrsim 29$ mag/arcsec^2; Baker & Willman 2015), so an alternative method is needed to locate other faint structures that may have evaded detection. RRL are sufficiently rare so as to not randomly form in pairs outside of stellar structures, so searching for groups of spatially close RRL provides an independent method to detect new structures (Ivezić et al. 2004; Sesar et al. 2014; Baker & Willman 2015; Medina et al. 2017, 2018). Indeed, at least one RRL has been found in almost every satellite galaxy with available time series data[3] (Boettcher et al. 2013; Vivas et al. 2016; Martínez-Vázquez et al. 2017, and references therein). Thus, identifying RRL in the halo can increase the census of old, metal poor satellite galaxies, streams, and overdensities, and improve our understanding of the Milky Way.

The two most common subtypes of RRL are those pulsating in the fundamental mode, RRab, and those pulsating in the first overtone, RRc. When their light curves are adequately sampled, RRab are easily identified by their short periods ($0.4 \lesssim P \lesssim 1$ d), relatively large pulsation amplitudes ($0.5 \lesssim A_g \lesssim 1.5$ mag), and a characteristic sawtooth shape. RRc have shorter periods ($0.2 \lesssim P \lesssim 0.45$ d), smaller amplitudes ($0.2 \lesssim A_g \lesssim 0.8$ mag), more sinusoidal-shaped light curves, and are generally less numerous than RRab. The fraction of RRab to RRc and the average periods of each are highly dependent on the metallicity of the stellar population in which they formed and is still not fully understood (see Catelan 2009 and references therein.) Most populations of RRL in the Milky Way are commonly subdivided into Oosterhoff I, II, and III groups based on these observational properties (named after the first dichotomy applied to globular clusters by Oosterhoff 1939. We refer the interested reader to Table 6 in Martínez-Vázquez et al. (2017) for a summary of these properties for a selection of Local Group dwarf galaxies.

Period-finding algorithms have long been used in conjunction with visual inspection to identify RRL from their time series photometry. However, with the dramatic increases in available data in recent years, the need for automated detection algorithms has grown significantly. Stetson (1996) made great strides in this regard when he introduced an automated method to identify Cepheid variables using template light curves to estimate their periods and a scoring system based on calculated variability indices. Recent studies have extended this period-finding technique to multiple filters (Mateu et al. 2012; VanderPlas & Ivezić 2015; Mondrik et al. 2015; Saha & Vivas 2017). However, even these algorithms suffer in performance when applied to extremely sparsely-sampled data. Hernitschek et al. (2016) and Sesar et al. (2017) developed separate techniques to identify RRL in the sparsely-sampled multiband Pan-STARRS data (Chambers et al. 2016) and found thousands of such variables.

We add to this census by presenting new RRL candidates discovered in the first three years of the DES data. DES is a five-year multiband ($grizY$) imaging

[2] The nature of the Canis Major Overdensity as a satellite galaxy is in doubt due to a lack of an RRL excess and a potential warp in the Milky Way disk (Mateu et al. 2009).

[3] One notable exception is the satellite galaxy candidate Carina III, which currently has no detected RRL in its vicinity (Torrealba et al. 2018a).

survey using the Dark Energy Camera (Flaugher et al. 2015) on the 4-m Blanco Telescope at Cerro Tololo Inter-American Observatory (CTIO). After the conclusion of its observations, DES will provide a deep ($\sim$ 25 mag in the final coadded images) and wide ($\sim$ 5000 deg^2) dataset near the Southern Galactic cap (DES Collaboration 2005; Diehl et al. 2016). By the end of the survey, the entire footprint will have been imaged $\sim$ 10 times in each band. While the main goal of the survey is to better constrain certain cosmological parameters, the deep and wide survey data provide an excellent test bed for probing Milky Way substructure with RRL. However, like Pan-STARRS, the DES light curves are multiband and poorly sampled. In this paper, we detail how we overcome these challenges by creating a empirically derived light curve template and a computationally efficient fitting algorithm to determine periods and other light curve parameters. We use these methods to identify 5783 RRL candidates, 31% of which are new discoveries, including three with a heliocentric distance >220 kpc. This novel technique will prove useful for other sparsely sampled multiband data sets.

The rest of the paper is organized as follows: §2 describes how we extracted star-like objects from the DES Y3 data release; §3 explains how we rescaled the photometric uncertainties and applied simple metrics to select variable objects; §4 presents the multiband RRL template, its application to DES light curves, and the construction and performance of our random forest classifier; §5 presents our RRL catalog, a comparison to overlapping surveys, and parameter uncertainties; §6 discusses possible biases, the spatial distribution of the candidates, and potential future application for LSST.

2. DATA

2.1. *DES Year 3 Quick Release*

This work is based on the DES internal Year 3 Quick Release catalog (hereafter Y3Q2), which contains all the single epoch data from years 1-3 that formed the basis for the coadded DES first public data release[4] (DES Collaboration 2018, hereafter DR1). The Y3Q2 data set was developed in the same manner as the Y2Q1 data release used for the stellar overdensity searches in Drlica-Wagner et al. (2015) (see their §2.1 for a detailed description of how the Quick Release catalogs were generated), with one major change. Instead of using stellar locus regression (SLR Ivezić et al. 2004; MacDonald et al. 2004; High et al. 2009; Gilbank et al. 2011; Desai et al. 2012; Coupon et al. 2012; Kelly et al. 2014) to determine zeropoints for the absolute photometric calibration, the Y3Q2 release utilizes the Forward Global Calibration Module (FGCM) photometric zeropoints (Burke et al. 2018). Y3Q2 contains single epoch catalogs generated from the reduced FINALCUT DES images (Morganson et al. 2018) and a cross-matched "coadded catalog" generated from these single epoch measurements. This catalog does not contain information from exposures in which an object was not detected at approximately a 5-σ level. The DES Y3Q2 coadded catalog contains nearly 2.9×10^8 unique objects and spans the entire survey footprint with $S/N \sim 10$ at a median depth of 23.5, 23.3, 22.8, 22.1, and 20.7 mag in *grizY*, respectively (DES Collaboration 2018).

As DES images are collected, the filter to be used and the location to be imaged are prioritized according to the time of the year, the sky conditions (Moon phase, seeing, weather), and how many times that particular area has already been imaged (Neilsen & Annis 2014) (see Fig. 3 in Diehl et al. 2016). While this strategy ensures uniform depth and the best use of the observing time, objects in the wide-field survey are sampled with a highly unpredictable cadence. In the Y3Q2 data set, individual objects can have from 2 to over 50 observations depending on their location. We ensure that each light curve only contains photometric observations by requiring that each observation has a SExtractor warning value FLAG $\leq$ 4, is sufficiently far away from masked regions in the images (IMAFLAG_ISO $\leq$ 4), and has a zeropoint correction available (FGCM_FLAG $\leq$ 4). After these cuts, the median number of total observations for a given object is 10, while the median number of observations in each band across the survey region is 4. The effects of the survey coverage and these cuts are discussed more in §6.1.

2.2. *Object Selection*

We selected our objects using the coadded catalog before examining the time series data, since the former contained most of the information needed to identify candidates (such as the number of times each object was imaged in each band and the star-galaxy classification). We further restricted the sample to stellar-like objects by following a prescription similar to Bechtol et al. (2015), based on the SExtractor (Bertin & Arnouts 1996) SPREAD_MODEL parameter which selects stars well down to $r \sim 23$ (Drlica-Wagner et al. 2015). Lastly, we required at least five total observations to be able to search for variability.

We selected objects that are bright enough to be detected in multiple images by requiring the coadded PSF (WAVG_MAG_PSF) or the aperture magnitudes

[4] https://des.ncsa.illinois.edu/releases/dr1

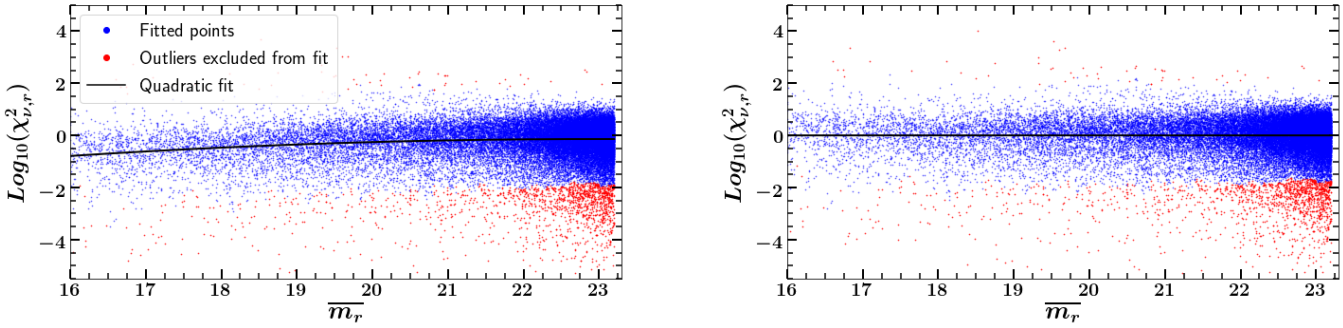

Figure 1. *Left:* Variation of $\log_{10}(\chi^2_{\nu,r})$ vs. median r magnitude, $\overline{m_r}$, demonstrating that photometric errors are slightly overestimated for brighter objects in the DES pipeline. Red points were excluded using an iterative 3-σ clipping procedure. The black curve shows the quadratic fit that was used to rescale the errors. *Right:* Distribution after the photometric errors were rescaled.

in the exposure with the best seeing in that band (MAG_AUTO) to be brighter than the median depth of the Y3Q2 single-epoch exposures across the entire survey region (see §2.1). We excluded all objects for which the coadded photometry errors exceed 0.3 mag in each of *griz* to reduce the number of spurious detections.

To ensure that we did not discard stars with missing data in a single band, we considered these quantities separately for each of *griz*. We did not use the Y data for these initial selections because those exposures are generally taken under worse seeing conditions than the other bands (Diehl et al. 2016) and are thus a poor choice to use for star-galaxy separation. The star-galaxy separation we used performs best in *riz* due to the better seeing conditions for those observations, as discussed in the previous section. Including objects which passed this cut in g likely allowed some extended sources into our sample, which we discuss further in §5.1.

Although RRL have well-characterized colors, we did not employ a color cut in this early stage of the analysis because we did not want to exclude any potential RRL with poor coverage across filters or pulsation phase. Simultaneous colors were not available for some objects in the DES footprint, so we would have to calculate colors using coadded magnitudes or magnitudes from arbitrary phases in the star's variation to calculate colors, which would expand the range of possible colors for RRL in our data. The RRL template we describe in §4.1 provides the color information we need to identify RRL candidates. In the future, when more epochs of DES data are available, color cuts will be a more reliable RRL indicator prior to the template fitting.

In summary, our combined selection criteria were:

- ≥ 2 observations in *g,r,i,* or *z*
- $|\text{SPREAD_MODEL}| \leq (0.003 + \text{SPREADERR_MODEL})$ in *g,r,i,* or *z*
- $(16 \leq \text{WAVG_MAG_PSF} \leq \text{median depth})$ or $(16 \leq \text{MAG_AUTO} \leq \text{median depth})$ in *g,r,i,* or *z*
- $(\text{WAVG_MAGERR_PSF} < 0.3)$ or $(\text{MAGERR_AUTO} < 0.3)$ in *g,r,i,* or *z*

A sample of $\sim 1.5 \times 10^8$ objects passed all of these combined selection criteria. We used their time series data instead of their coadded values for the remainder of our analysis.

3. VARIABILITY ANALYSIS

3.1. *Error Rescaling*

Photometric uncertainties can have a large impact on the success of our variability classification. We first account for both the photometric uncertainties reported by the DES pipeline and the uncertainties in the FGCM zeropoint solution[5] for each exposure by adding them in quadrature. Since photometric uncertainties can be over- or under-estimated for different magnitude ranges (Kaluzny et al. 1998), we calculated the reduced chi-squared statistic, $\chi^2_{\nu,b}$, from the median magnitude $\overline{m_b}$ in a given band b for each light curve:

$$\chi^2_{\nu,b} = \frac{1}{N_b - 1} \sum_{1}^{N_b} \frac{(m_{i,b} - \overline{m_b})^2}{\sigma^2_{i,b}}. \quad (1)$$

where N_b is the number of observations for a unique object in band b, $m_{i,b}$ is the ith observation in that band, and $\sigma_{i,b}$ is the photometric uncertainty combined with the zeropoint uncertainty for that observation. As this statistic measures the goodness-of-fit to a constant value of $\overline{m_b}$, one would expect $\chi^2_{\nu,b} \approx 1$ for a non-variable source and $\chi^2_{\nu,b} > 1$ for variable sources.

[5] At the time of this analysis, only a pre-release version of the FGCM zeropoints was available. A later version of these zeropoints was used for other DES Year 3 analyses.

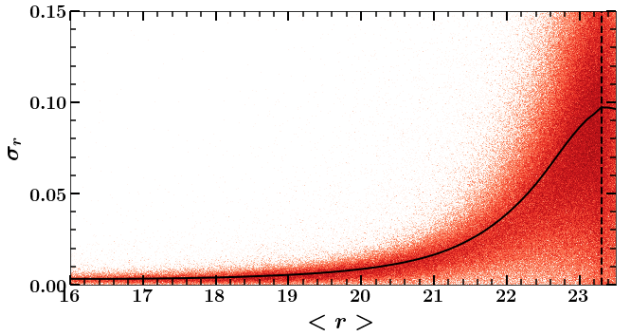

Figure 2. Uncertainties in the r-band magnitudes for the entire survey region after the error rescaling described in §3.1. The dashed black line denotes the median survey depth in this band and the solid black line shows the 3rd degree polynomial fit to the uncertainties.

Since the majority of objects within a given field will have constant (non-varying) light curves, any overall trend of $\chi^2_{\nu,b}$ vs. magnitude will be indicative of incorrect estimations of photometric uncertainty. For ease of calculations, we subdivided the single epoch data by HEALPix (nside=32) (Górski et al. 2005)[6] and filter. This resulted in 1772 unique DES HEALPix regions. We fit a quadratic function:

$$\log_{10}(\chi^2_{\nu,b}) = c_{0,b} + c_{1,b}(\overline{m_b} - 20) + c_{2,b}(\overline{m_b} - 20)^2 \quad (2)$$

for each filter b in each of these regions, excluding variables and outliers by applying an iterative 3-σ clip from the median value using the `sigma_clip` function in `astropy.stats`. We show the initial trend in $\chi^2_{\nu,r}$ for the objects in HEALPix 11678 in the r band in the left portion of Figure 1. We multiplied the reported photometric uncertainties of each object by scaling factors based on the best-fit value of $\chi^2_{\nu,b}$ for each of its magnitudes in a given band. Once the uncertainties were suitably rescaled, we repeated the calculation to verify that no trends remained. The right portion of Figure 1 shows the resulting lack of trend in $\chi^2_{\nu,r}$ after the rescaling procedure. Table 1 lists the coefficients used to rescale the errors in each band for this example region (the full version is available online). Figure 2 shows the final rescaled photometric uncertainties for the entire survey region as a function of magnitude for the r band.

3.2. *Variability Cuts*

Once photometric uncertainties were rescaled, we assessed the variability of each light curve using two simple metrics. The first was $\chi^2_{\nu,b}$ described in §3.1. The second was a metric we called "significance", consisting of a weighted range of the magnitudes in one band that acts as a proxy for light curve amplitude[7].

$$\text{significance}_b = \frac{(m_{\text{max},b} - m_{\text{min},b})}{\sqrt{\sigma^2_{\text{max},b} + \sigma^2_{\text{min},b}}} \quad (3)$$

To test the effectiveness of these metrics and determine the threshold values to separate variables from constant stars, we assembled a labeled training set of previously classified objects in SDSS stripe 82 region (hereafter, "S82"). S82 is a $\sim 300\,\text{deg}^2$ area spanning $300° \lesssim \alpha \lesssim 60°$ and $|\delta| \lesssim 1.25°$ that was observed 70-90 times by SDSS in *ugriz* over a period of 10 years. Numerous authors used the resultant well-sampled multi-band light curves to identify thousands of variable stars in the region with high confidence (Ivezić et al. 2007; Sesar et al. 2010; Süveges et al. 2012). These labeled objects are extremely useful for studies of variables from both hemispheres thanks to their equatorial location. Although the magnitude range of DES is deeper than that of SDSS, there is sufficient overlap to create a well-populated training set for our study. Using the calibration and variable catalogs from Ivezić et al. (2007), we cross-matched 641,710 "standard" (i.e., constant) stars and 16,752 variables in common between SDSS and DES objects. We also identified 296 RRL in common between DES and either Sesar et al. (2010) or Süveges et al. (2012), consisting of 238 and 58 objects of subtype RRab and RRc, respectively.

As an example, we show the cumulative distributions of $\chi^2_{\nu,b}$ values and "significance" values for the cross-

Table 1. Error rescaling coefficients for Equation 2

HEALPix	Band b	$c_{0,b}$	$c_{1,b}$	$c_{2,b}$
	g	-0.2672	0.0816	-0.0152
	r	-0.2572	0.0793	-0.0134
11678	i	-0.1180	0.0395	-0.0156
	z	-0.2160	0.0827	-0.0198
	Y	-0.1942	0.0505	-0.0250

NOTE—The full version of this table is available online.

[6] These HEALPix indices are also provided in the DES DR1 products (DES Collaboration 2018).

[7] While other metrics such as the Welch-Stetson I Welch & Stetson (1993) and Stetson J (Stetson 1996) indices are widely used and very effective at detecting variability, due to the sparsity of our observations, we chose to use metrics that were agnostic of observation time.

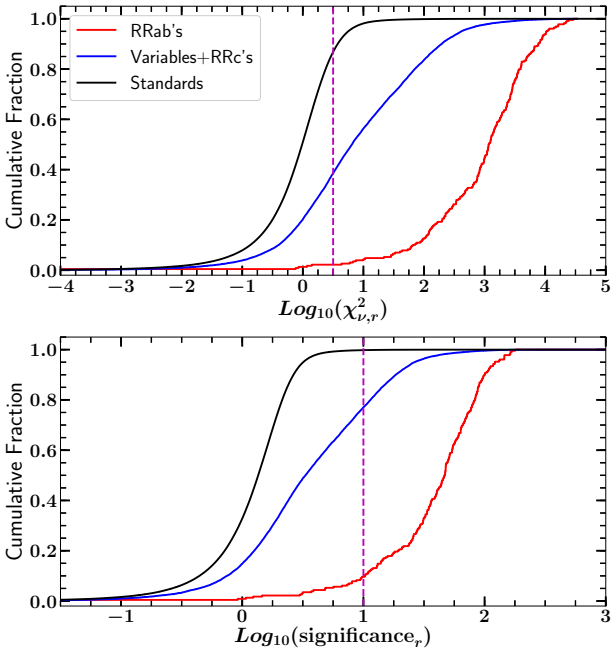

Figure 3. Initial variability metric values for previously identified objects in S82. *Top:* Cumulative distribution of $\log_{10}(\chi^2_{\nu,r})$. The magenta dashed line denotes the chosen threshold value of 0.5; objects with larger values in any band are considered variable. *Bottom:* Histogram of $\log_{10}(\text{significance}_r)$. The vertical magenta dashed line denotes the chosen threshold value of 1.0; objects with larger values in any band are considered variable. Note that although some real RRab are excluded by this cut, most of the non-RRab variables are also excluded.

matched objects in the r band in Figure 3. For both metrics, the threshold values were chosen to minimize the number of non-variable stars that would be subject to subsequent analysis. Any objects that showed $\log_{10}(\chi^2_{\nu,b}) \geq 0.5$ and significance ≥ 1 in any one of $griz Y$[8] were kept for subsequent analysis. When these cuts were applied across all five filter bands, 234 ($\sim 98\%$) RRab, 57 ($\sim 98\%$) RRc, 5196 ($\sim 31\%$) variable, and 3004 ($\lesssim 0.05\%$) standard light curves from S82 met these criteria. These results for our training set are summarized in Table 2. Over the entire survey region, approximately $\sim 7 \times 10^5$ light curves passed these variability cuts. We caution that passing this criterion is simply an initial cut and does not imply that these sources are truly variable.

Despite these cuts, a small but non-negligible fraction of the objects identified as "standard" by Ivezić et al. (2007) were still selected. It is possible that some of these objects are truly variable sources that did not display significant changes in the previous studies. Another possibility is erroneous photometry that, while rare, occurs sometimes in the Y3Q2 dataset due to incorrectly attributing observations from separate sources to one object or imperfect masking of observations obtained in very poor weather conditions. While these objects may have passed the initial variability cuts, their light curves fit the RRL template poorly, and most of them were rejected later in our analysis.

[8] Unlike the initial cuts we applied to the coadded catalog, we included the Y band in these variability cuts because the Y band values were weighted by the photometric uncertainties.

Table 2. Training set of cross-matched S82 objects

SDSS Label	Present in DES	Passed Cuts
RRab	238	234
RRc	58	57
Variables	16,752	5196
Standards	641,710	3004

Note—Objects were originally identified in Ivezić et al. (2007); Sesar et al. (2010); Süveges et al. (2012).

All of the selected light curves were corrected for extinction using reddening values from the maps of Schlegel, Finkbeiner & Davis (1998) multiplied by filter coefficients derived from the Fitzpatrick (1999) reddening law (for $R_V = 3.1$) and the adjustments to the Schlegel, Finkbeiner & Davis (1998) map presented by Schlafly & Finkbeiner (2011) (see §4.2 in DES Collaboration (2018) for more details). We then used the extinction-corrected light curves as the input for our template fitting algorithm.

4. CANDIDATE IDENTIFICATION

4.1. *RR Lyrae Template*

Our current work introduces a novel method of identifying RRab by fitting an empirically derived periodic model to the sparsely sampled multiband light curves. The model has the form:

$$m_b(t) = \mu + M_b(\omega) + a\gamma_b(\omega t + \phi) \quad (4)$$

where m_b is the measured magnitude in a given band b at a given time t, μ is the distance modulus, M_b is the average absolute magnitude in that band, $\omega = 1/P$ is the frequency of the variability (inverse of the period P), a is the g-band amplitude (the amplitudes of the curves for the other bands are proportional to a), γ_b is a periodic shape function, and ϕ is the phase. Only the four parameters μ, a, ω, ϕ are estimated during

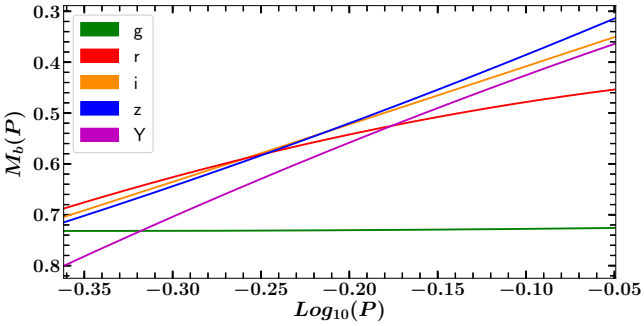

Figure 4. $P-L$ relations used in our template fitting procedure. RRL have nearly constant absolute magnitudes in g regardless of period. See §4.1 for more details.

the fitting process while the forms of γ_b and $M_b(\omega)$ are fixed. These were derived using well-sampled RRab light curves from S82 (Sesar et al. 2010) and shifted to adjust for slight differences between SDSS and DES filters. A more detailed description of the template construction is included in Appendix A.

Using the same reasoning as Sesar et al. (2017), we chose to exclude RRc from our study because: a) their sinusoidal light curves are difficult to distinguish from light curves from other variable objects such as eclipsing binaries, b) their small amplitudes would make them difficult to identify in our sparse data, and c) searching over a larger period range to recover their short periods ($\sim$ 0.3d) would introduce additional common period aliases into our sample. While excluding RRc weakens our sample size for tracing substructure, it is not prohibitive since RRc are usually less numerous than RRab. Furthermore, this approach allowed us to use only one generalized RRab shape for our template instead of an ensemble of shapes as in Sesar et al. (2017). While they were able to recover a more diverse group of RRL by fitting multiple shapes, our approach makes our algorithm more computationally efficient.

The *P-L-Z* relations were implemented in our template fitting procedure as *P-L* relations evaluated at a starting metallicity of [Fe/H] $= -2.85$. However, the values of these offsets between template filter curves were later adjusted to fit the S82 RRL light curves, so this metallicity should not be treated as the true value of the template. Hence, the metallicity value of our template is somewhat ambiguous. We discuss the effects of the metallicity further in §5.4. The final template *P-L* relations are shown in Figure 4. More details on how these parameters were derived are presented in Appendix B.

4.2. *Template Fitting*

In this section, we include a high-level overview of the template fitting procedure. A more detailed description of this process can be found in Appendix C.

To fit the template to a light curve, the algorithm performs a grid search over a specified range of frequencies. To prevent misestimated and aliased periods outside of the true range of periods known for RRab, we restricted this range to values corresponding to periods of 0.44 days to 0.89 days following the period-amplitude relation shown in Figure 16 of Sesar et al. (2010). At each frequency gridpoint, the algorithm first calculates $M_b(\omega)$ (see Equation 4 and Figure 4) and subtracts these values from the light curve magnitudes in the appropriate band. Then, with the frequency fixed, the model alternates between estimating the best-fitting ϕ using Newton's method, and (a, μ) using weighted least squares. The (ϕ, a, μ) values which minimize the weighted Residual Sum of Squares (RSS) at each frequency gridpoint are chosen as the best-fitting parameters, and act as the starting point for the parameter search at the next gridpoint. The (ω, ϕ, a, μ) at the global minimum RSS value over the entire frequency grid are chosen as the best-fitting parameters[9].

A major strength of this algorithm is that the template shape is fit simultaneously to the light curve data in all five bands combined. Since there are only four free parameters that must be fit for the entire light curve, unique solutions can be found for sparse light curves with very few measurements in any single band. An example RSS curve for an RRab from the labeled training set is shown in Figure 5. Because the local minima in the RSS curve are sometimes very close in value, we include the best-fitting values for the top three minima of RSS in our data products for completeness, but do not discuss the results of the 2nd and 3rd minima further in this work.

Our algorithm is also effective at estimating distances. At the best fit parameter estimates, the template fitting algorithm correctly estimated $\sim$81% of the S82 RRL distances to within 3% of the values obtained by Sesar et al. (2010) and Süveges et al. (2012) (if available) with an overall standard deviation of 2.8% (see §5.4 for an extended discussion of the uncertainties in distance modulus). The accuracy of the template estimates of both

[9] In his study of Cepheid variables, Stetson (1996) also developed a template fitting method based on least squares. However, instead of using a string-length minimization technique in a single band, we use all the bands simultaneously and used a fixed shape instead of a family of derived template curves calculated for each trial period.

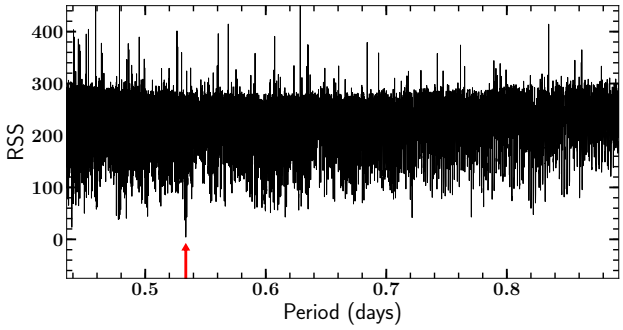

Figure 5. Residual Sum of Squares (RSS) curve for an RRab originally discovered by Sesar et al. (2010). The red arrow denotes the global minimum of the RSS which corresponds to the true period of 0.5336 days.

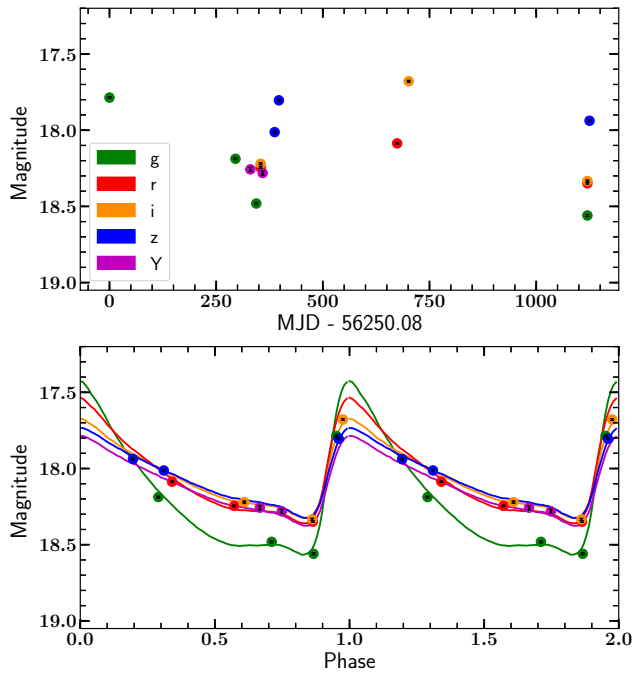

Figure 6. *Top*: Poorly sampled DES light curve of an RRab originally discovered by Sesar et al. (2010) (same star as Figure 5). *Bottom*: Phased light curve of the same source with the period correctly estimated by our algorithm. *Note:* Photometric uncertainties are smaller than the plotting symbols.

the period and the distance for the training set of RRab light curves is summarized in Table 3).

Our algorithm is computationally efficient and only takes $\sim 3-5$ minutes per light curve on an Intel Xeon E5420 processor. The template fitting code returned the estimated parameters of the top three best-fitting templates as well as the features used in the random forest classification detailed in §4.4 and Table 4. The computation time for fitting the template and calculating features for $\sim 7 \times 10^5$ light curves was $\sim$ 44K CPU hours. For comparison to a similar analysis, our algorithm is $> 9\times$ faster than the template fitting methods used by Sesar et al. (2017), which required $\sim$ 30 minutes per star.

Table 3. Period and distance estimation accuracy

Parameter	% of RRab within		$\sigma_{\text{parameter}}$
	1%	3%	
$\Delta P/P_{\text{prev}}$	88.89	89.74	6.81%
$\Delta D/D_{\text{prev}}$	44.64	81.11	2.83%

Note—P_{prev} and D_{prev} are the values reported in Sesar et al. (2010) and Süveges et al. (2012).

There are several other well-documented methods available in the literature for identifying RRL in multiband data (e.g. VanderPlas & Ivezić 2015, Hernitschek et al. 2016 and Sesar et al. 2017), which yield excellent results for data sets with an average of 35 or more observations per light curve. We present this algorithm as an alternative to these other methods for especially sparse multiband data sets (see §5.3 for a discussion of the observational limitations.) The template and the associated fitting code are available at https://github.com/longjp/rr-templates and further described in Appendix D.

4.3. *Feature Selection*

While it is possible to identify RRL by visually inspecting their phase-folded light curves, the sheer volume of light curves in our data set makes this classification method unfeasible. Instead, we computed numerical features to describe the behavior of the light curves. To assess the specific parameter space occupied by RRab, we compiled a training set consisting of all the cross-matched labeled objects from S82 which passed the initial variability cuts (discussed in §3.2). This left us with a training set of 234 RRab, 57 RRc, 5196 other variable objects, and 3004 "standard" sources. Since we only aimed to identify RRab, we chose a simple identification scheme with two classes: RRab and non-RRab. This resulted in an RRab class with 234 objects and a non-RRab class with 8257 objects.

With the goal of identifying RRab, we chose features which were motivated by how well the light curves fit the RRab template and other observed properties of RRab. As demonstrated in Figure 3, RRab have relatively large $\log_{10}(\chi^2_{\nu,b})$ compared to most of the other objects in our sample. So, to quantify the base variability of the light curve while ignoring spurious signals or missing observations in any particular band, we included the median

Table 4. Random Forest Features

Feature Name	Description	Importance
lchi_med	median $\log_{10}(\chi^2_{\nu,b})$ value across all bands	0.2232
amp_rss_0	Best-fitting Amplitude divided by the best-fitting RSS/dof	0.1942
f_dist1_0	Closest distance of best-fitting period/amplitude to Oosterhoff I relation from Sesar et al. (2010)	0.1591
rss_dof_0	RSS of the best-fitting template per degree of freedom	0.1571
f_dist2_0	Closest distance of best-fit period/amplitude to Sesar et al. (2010) curve dividing the Oosterhoff I and II groups	0.1025
amp_0	Best-fitting amplitude	0.0792
rss_lchi_med	RSS of the best-fitting template divided by the median $\log_{10}(\chi^2_{\nu,b})$ value across all bands	0.0502
κ_0	Best-fitting von Mises-Fisher concentration parameter of phases in the folded light curve	0.0345

NOTE—All of these feature values for each candidate are included in the electronic version of Table 6.

of this value calculated across all five filters as a feature. Additionally, most true RRab should fit the template with small residuals, so we quantified the quality of the best template fits using the RSS per degree of freedom, $\mathrm{RSS_{dof}} = \mathrm{RSS}/(N_{obs} - 4)$. Then, to amplify the separation provided by these two characteristics, we divided the $\mathrm{RSS_{dof}}$ by the median $\log_{10}(\chi^2_{\nu,b})$ to form another feature.

To take advantage of the distinctive amplitude ranges of RRab, we also selected the amplitude of the best-fitting template as a feature. We then created a new feature by dividing this amplitude by the $\mathrm{RSS_{dof}}$, expecting that the large amplitudes and excellent template fits of RRab would clearly distinguish them from other objects. We can take advantage of these amplitudes again by evaluating how closely each object matches the observational trends shown by RRab in the first two Oosterhoff groups (see the introduction for a brief description). To measure how closely the objects' estimated template parameters matched these trends for RRab, we calculated the distance of the object in period-amplitude space from the Oosterhoff I relation measured in Sesar et al. (2010) and their shifted curve which separates the Oosterhoff I and II populations (see their Figure 16 and our Figure 12.)

Our last feature attempted to quantify the phase distribution of the observations in each light curve. Period-finding algorithms often recover periods at common aliases, sometimes resulting in light curves with many of their observations clustered near a particular phase. Because light curve phases are periodic, the two-dimensional case of the von Mises-Fisher distribution (Fisher 1953; Jupp & Mardia 1989) is a good approximation. This distribution can be written as:

$$f(x) = \frac{e^{\kappa cos(x-\mu)}}{2\pi I_0(\kappa)} \tag{5}$$

where κ is the concentration parameter, μ is the mean, and $I_0(\kappa)$ is the modified Bessel function of the first kind at order 0. The von-Mises Fisher distribution is akin to a Gaussian distribution wrapped around a circle, where the κ parameter is analogous to the inverse of the variance (see §2.2 in Sra 2016 for a more detailed description). We calculated this concentration parameter κ for each light curve folded over the best-fitting template period. Light curves with observations highly clustered in phase will have very large values of κ, aiding in the rejection of objects with aliased periods.

Although our choice of classifier is generally robust against non-informative features, we limited our features to these to make the classifier results easier to interpret. The features are summarized and ranked by their importance, or how much they contributed to splitting the data across all the decision trees in our classifier[10], in Table 4. These features are shown in Figure 7. The development of additional features to further separate the classes will be explored in future work.

4.4. *Random Forest Classifier*

To identify likely RRab automatedly and consistently, we trained a random forest classifier (Amit and Geman 1997; Breiman 2001) using these features. The random forest is a machine learning algorithm that predicts classes for data by combining results from a "forest" of decision trees. Each decision tree consists of a series of nodes where the data is split into subgroups based on the values of a random subset of their features, or characteristics. Before the random forest can make accurate predictions, it must be trained to recognize the trends in features that correspond to different classes. Thus, one needs a labeled training set to build the random forest. Each decision tree uses the labels to build

[10] See §1.11.2.5 in the `scikit-learn` documentation for more details.

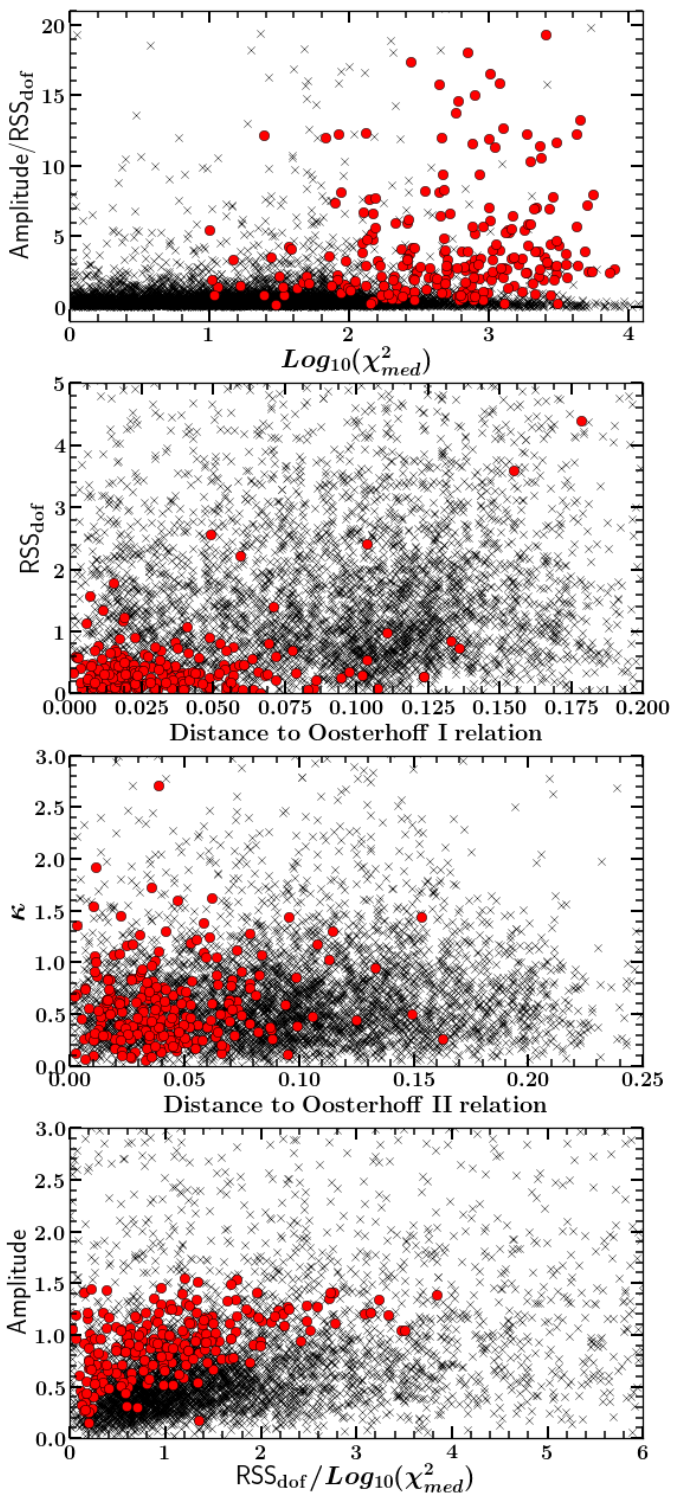

Figure 7. Features used to identify RRab plotted for the training set. Red points denote previously identified RRab while black X's are non-RRab. While the RRab clearly occupy a specific region in this feature space, they are not linearly separable.

Table 5. Estimated RRab selection purity & completeness

Score Threshold	Purity	Completeness
0.00	0.043	0.983
0.05	0.402	0.928
0.10	0.567	0.886
0.15	0.661	0.852
0.20	0.727	0.840
0.25	0.773	0.819
0.30	0.815	0.798
0.35	0.845	0.756
0.40	0.877	0.722
0.45	0.896	0.693
0.50	0.922	0.651
0.55	0.955	0.630
0.60	0.953	0.596
0.65	0.956	0.554
0.70	0.968	0.525
0.75	0.973	0.470
0.80	0.971	0.432
0.85	0.988	0.348
0.90	0.982	0.231
0.95	1.000	0.071

Note—The full version of this table is available in the online data products.

a series of boundaries in feature space that divide the data into their correct classes. Once trained, the random forest algorithm assigns a score to unlabeled data based on the proportion of trees that identify them as a particular class. Random forest classifiers have been extremely successful in variable star classification (see Richards et al. 2011 for a comparison with other machine learning techniques), even in the case of sparsely sampled Pan-STARRS PS1 light curves (Hernitschek et al. 2016). Thus, the random forest was a natural choice of classifier for this study.

We created the classifier using the `RandomForest` package available in `scikit-learn` (Pedegrosa et al. 2011). To prevent overfitting to our small training set and ensure repeatability, the classifier contained 500 trees with a maximum depth of 5, and used a random seed of 10.

We assessed the performance of our classifier by estimating the purity (the percentage of objects classified as RRab that were truly so) and the completeness (the percentage of real RRab that were identified as such) as a function of the class score reported by the random forest. The purity and completeness were estimated using a five-fold cross validation technique, where the data were divided into five test groups and classified based on

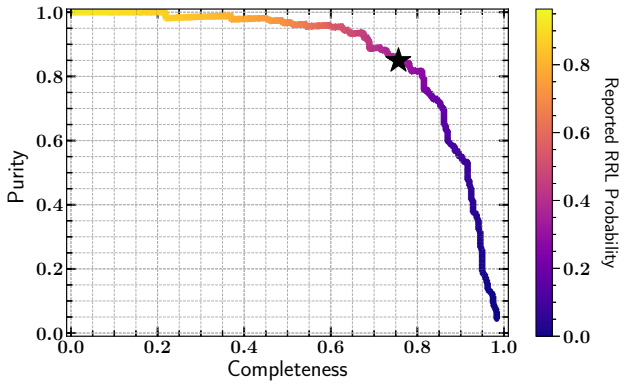

Figure 8. Purity/Completeness curve for the random forest classifier trained on cross-matched objects in S82. The black star denotes a classifier-reported score of 0.35, where the purity is $\sim 85\%$ and the completeness is $\sim 76\%$. The area under the curve is 0.864.

a classifier trained on the other four groups. The classifier correctly identified 190 of the 234 RRab used to train the random forest as such with a score threshold of $\geq 35\%$. As shown in Figure 8, defining RRab as objects with a score $\geq 35\%$ yields a purity of 85% and a completeness of 76%. A common metric used to assess the performance of a classifier is the area under the curve (AUC) shown in Figure 8, which we find to be 0.864. The purity and completeness calculated at other score thresholds are listed in Table 5. We include all other objects with lower scores in our catalog so that other score thresholds can be specified by interested readers. Although an incorrect period estimate led to a worse RSS value for the fit, some RRab with incorrect period estimates were still correctly identified as such by the classifier.

Since our training set was mostly composed of nearby RRab confined to a small region in the survey footprint, it is imperative to assess our classifier performance using other samples of RRL. To this end, we cross-matched our sample with external surveys in §5.2 and applied our method to simulated RRab light curves at fainter magnitudes in §5.3.

5. CATALOG DESCRIPTION

5.1. *Visual Validation*

We applied the classifier to the $\sim 7 \times 10^5$ objects with template fits and found 8026 RRL candidates with a score ≥ 0.35. Although most of our candidates were indeed RRL found in other surveys, there were still some non-stellar interlopers in the sample due to the lenient initial cuts on the shape of the photometric point-spread function (§2.2). Thus, we visually inspected all RRab candidate light curves and their DR1 coadded images. After visually validating the candidates and removing any objects with non-RRL classifications in the Simbad database (Wenger et al. 2000), 1786 objects were discarded and 5783 RRL candidates remained in the catalog, with the rest too ambiguous to confirm. A sample of visually verified candidates with high ($p > 0.94$) and low ($p < 0.36$) classifier scores are shown in Figures 9 and 10, respectively. The most typical contaminants in the sample were extended sources. Given our lenient selection criteria described in §2.2, it is not surprising that some of these objects made it into our final sample. Examples of candidates that were rejected by visual inspection as being extended sources are shown in Figure 11. The full catalog of candidates, their best-fit parameters, and their features are available at https://des.ncsa.illinois.edu/releases/other/y3-rrl. A sample view of this catalog is shown in Table 6. Appendix E contains a detailed description of these data products.

Although the light curves have been visually inspected, further photometric observations of some extremely poorly sampled candidate light curves would be useful to confirm their classification. One may wish to remove these more uncertain candidates from their analysis by only considering objects with a larger minimum number of observations. Some of these candidates have gaps in observations near their maximum and minimum brightness, providing poor constraints on their estimated amplitudes and mean magnitudes. Therefore, we assigned a flag to each object based on how its phase-folded light curve is sampled. An object with fewer than two observations near its minimum brightness ($0.55 \leq$ phase < 0.87, which we chose to encompass both the near-constant portion of the light curve where other authors chose their minimum phase (e.g., Vivas et al. 2017) and the 10% quantile of template magnitudes) will receive a "flag_minmax" value of 1, while an object with < 2 observations near its maximum brightness ($0.96 \leq$ phase < 1 or $0 \leq$ phase < 0.05 corresponding to the 90% quantile in template magnitudes) receives a "flag_minmax" value of 2. Objects missing observations near both of these receive a flag value of 3.

Figure 12 shows a Bailey (period-amplitude) diagram of the candidates rejected by the classifier or our visual checks plotted in black, visually accepted candidates shown in red, and ambiguous candidates in blue. We plot the Oosterhoff I (Oosterhoff 1939) relation and the curve dividing groups I and II from (Sesar et al. 2010) which we used in the calculation of our features in solid and dashed black lines. Many of these ambiguous candidates are likely RRab, but cannot be classified as such with high confidence in this work. Due to the sparse

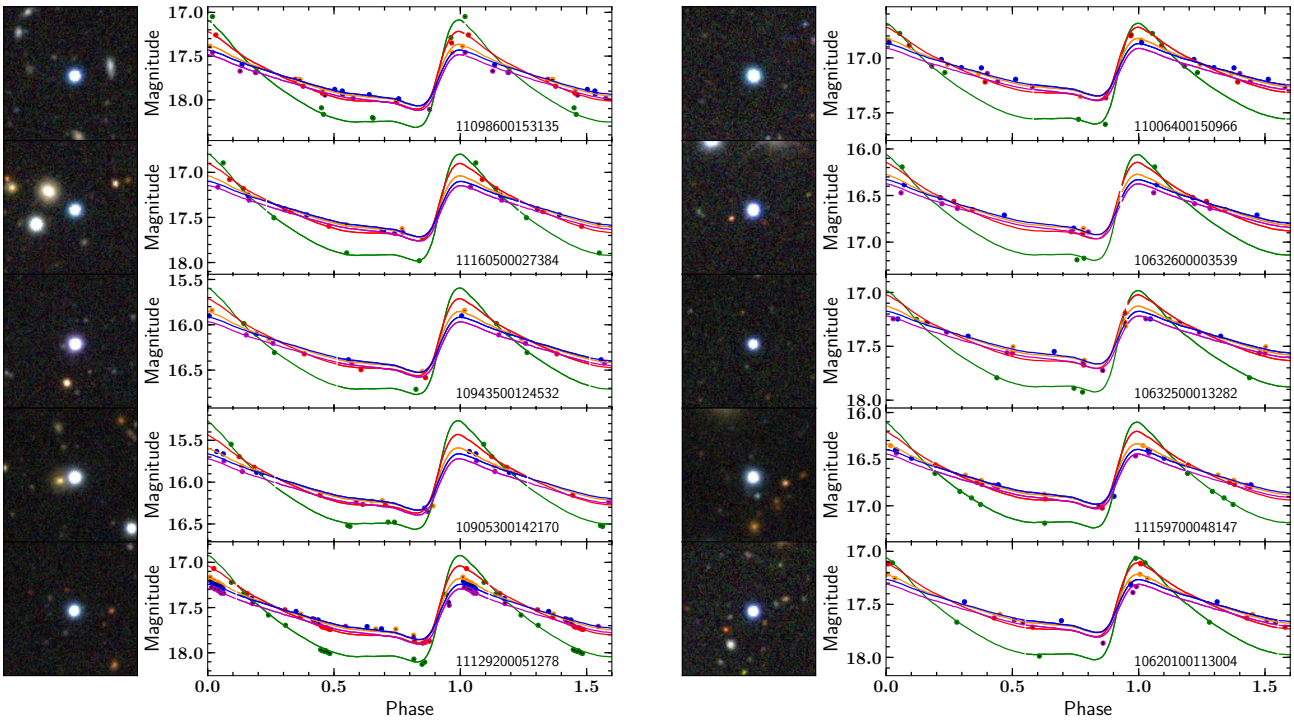

Figure 9. Sample of DES coadded images and representative light curves of visually accepted RRL candidates with classifier scores exceeding 0.94, labeled with their Y3Q2 ID number. The observations and templates are colored by filter using the same convention as Figure 6.

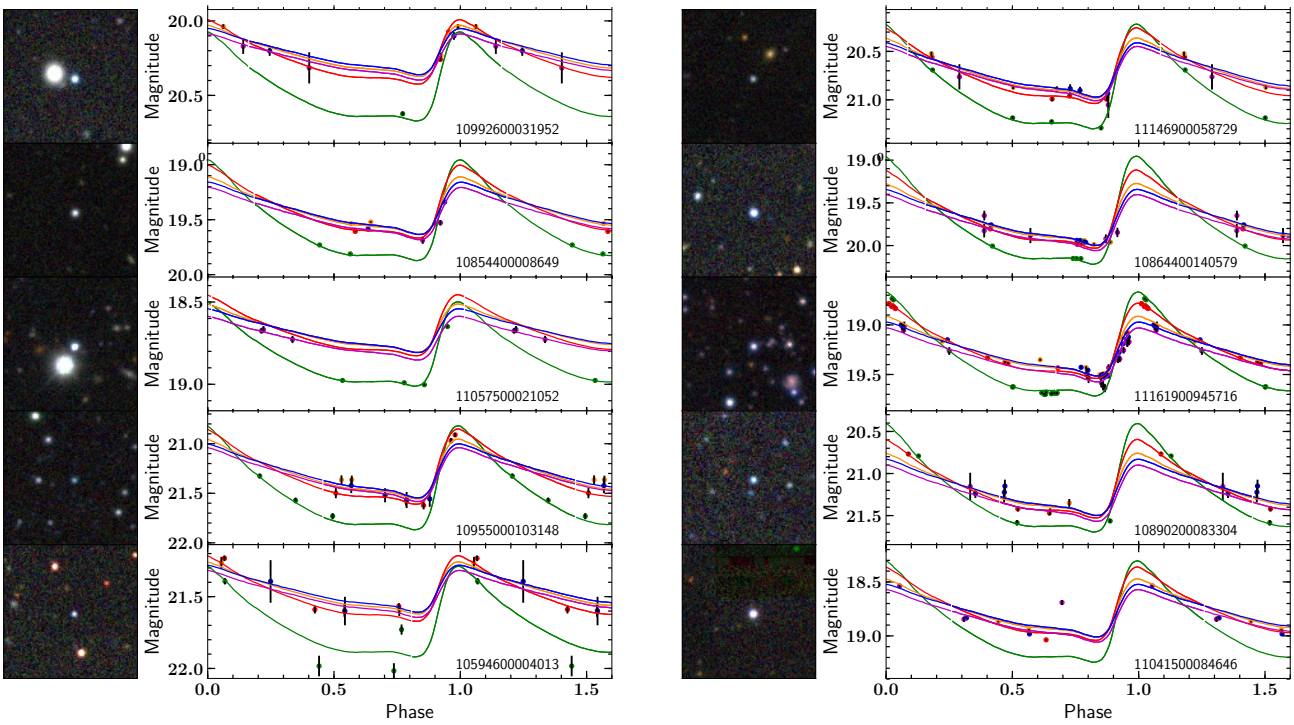

Figure 10. Sample of DES coadded images and representative light curves of visually accepted RRL candidates with classifier scores below 0.36, labeled with their Y3Q2 ID number. The observations and templates are colored by filter using the same convention as Figure 6.

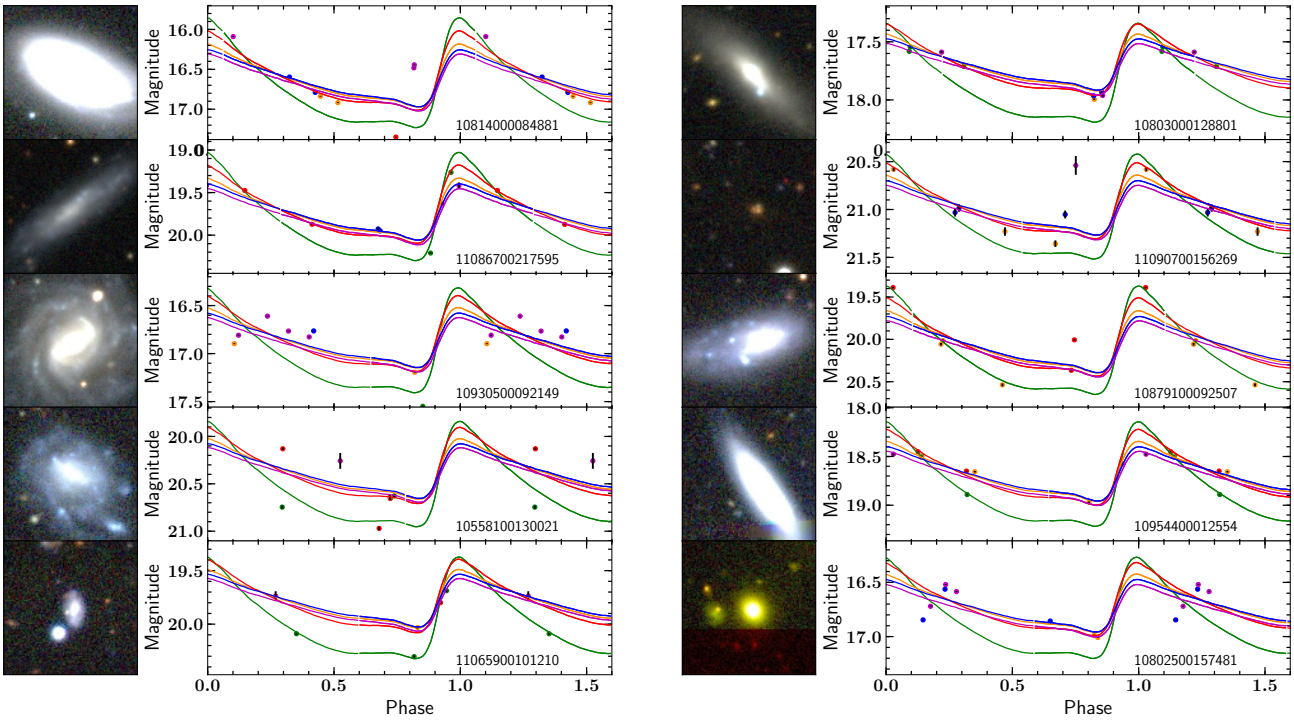

Figure 11. Sample of DES coadded images and representative light curves of visually-rejected candidates (extended sources or possible supernova) that passed the classifier score threshold, labeled with their Y3Q2 ID number. The observations and templates are colored by filter using the same convention as Figure 6.

Table 6. DES RRab Candidates

DES Y3Q2 ID	α	δ	$\langle g \rangle$	$\langle r \rangle$	$\langle i \rangle$	$\langle z \rangle$	$\langle Y \rangle$	P	A_g	μ	p
	[deg, J2000]				[mag]			[d]	[mag]		
11136400113264	0.000107	-59.559187	17.657	17.500	17.547	17.535	17.588	0.6415	0.551	16.90	0.436
10646400013129	0.013042	-2.430057	17.234	17.052	17.099	17.103	17.252	0.4938	1.151	16.66	0.733
11004800140792	0.131498	-41.482218	19.229	19.096	19.125	19.195	19.145	0.5893	1.137	18.69	0.568
11108800089990	0.134204	-54.295118	15.754	15.625	15.557	15.554	15.571	0.6698	0.469	14.97	0.659
10595200009863	0.283437	1.178535	17.986	17.922	17.908	17.868	17.869	0.5480	1.033	16.96	0.886

NOTE—DES Y3Q2 ID: DES Y3Q2 QUICK_OBJECT_ID number. α: Right Ascension. δ: Declination. $\langle grizY \rangle$: Mean extinction-corrected magnitude. P: Best-fit period. A_g: Best-fit amplitude in DES g. μ: Best-fit distance modulus. p: RRab score assigned by the classifier. The full version of this catalog, including feature values and cross matching information, is available in the online data products at https://des.ncsa.illinois.edu/releases/other/y3-rrl.

nature of our observations, we cannot detect amplitude modulations such as those arising from the Blažko effect (Blažko 1907), although we did recover five out of twelve Blažko RRL previously identified in the Catalina Surveys (Drake et al. 2014, 2017).

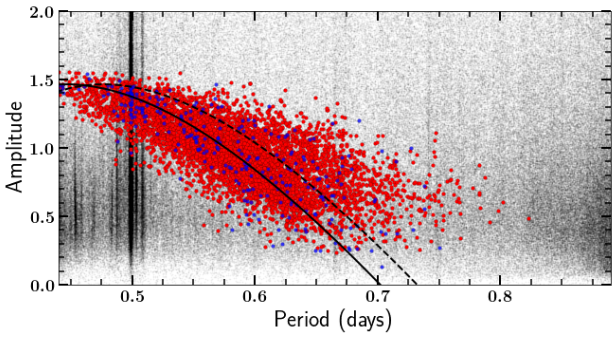

Figure 12. Bailey diagram of template-estimated amplitudes and periods for objects that passed the initial variability cuts in black and visually accepted RRab identified by our classifier in red. Ambiguous candidates that could not be visually accepted are plotted in blue. We overplot the Oosterhoff I relation and the curve dividing the Oosterhoff I and II populations from Sesar et al. (2010) in black solid and dashed lines, respectively. The abundance of objects with periods of $P = 0.5$ day denotes a common alias of the 1 day rotation period of the Earth.

5.2. *Comparison with Overlapping Catalogs*

The DES footprint has significant overlap with other surveys, such as *Gaia* (Gaia Collaboration et al. 2016), Pan-STARRS (Chambers et al. 2016), the Catalina Surveys (Drake et al. 2009), and the Asteroid Terrestrial-impact Last Alert System (ATLAS: Tonry et al. 2018). We used our cross-matches with these external RRab catalogs to independently assess the performance of our algorithm at the different magnitude ranges probed by these surveys. We used the `SkyCoord` package in `astropy` (Astropy Collaboration 2018) to select matches within $1''$ of DES objects while removing duplicates. Details for each individual survey are presented in the following paragraphs and summarized in Table 7, while Figure 13 shows the respective overlaps with DES.

Clementini et al. (2018) found over 1.4×10^5 RRL in *Gaia* DR2, including $\sim 5 \times 10^4$ that were previously unknown. These variables were identified from multiband (G, G_{BP}, G_{RP}) light curves that had at least 12 observations in G (see Figure 10 in Holl et al. 2018). While the *Gaia* temporal coverage is very uneven, their RRL catalog spans the entire sky (see Figure 26 in Clementini et al. 2018) and has high purity ($\sim 91\%$), making it an excellent independent check of our method at brighter magnitudes. 4609 of the *Gaia* DR2 RRabs were present in our initial stellar catalog (§2.2) and 3227 ($\sim$70%) were identified as such. To assess this recovery another way, if we create a purity vs. completeness curve from these cross-matches like the one shown in Figure 8, we find an AUC of 0.727. As we have significantly fewer single-band observations than *Gaia* DR2, it is not surprising that we do not recover all of their RRab.

We also searched for RRab discovered in Pan-STARRS PS1 by Sesar et al. (2017). Like DES, Pan-STARRS has sparsely-sampled multiband light curves and Sesar et al. (2017) employed a similar template fitting method to identify these variables. However, Sesar et al. (2017) used the final data release of PS1 with an average of 67 observations per object (compared to our median of 18). We adopted their suggested *ab_score* cut of 0.8 to select only RRab. As Pan-STARRS primarily surveyed the Northern hemisphere, we found just 1021 RRab in our initial stellar catalog, but we identified 805 ($\sim$79%) as such, with an AUC of 0.681. As the Pan-STARRS light curves are the most similar to the DES Y3 ones out of all the external catalogs under consideration, our similar classification results show that our approach is similarly effective as the methods used by Sesar et al. (2017).

The Catalina Surveys RRL catalog (Drake et al. 2013a,b, 2014; Torrealba et al. 2015; Drake et al. 2017) is based on a wide-field (26,000 deg^2) time series survey that probes the variable sky to a depth of $V \sim 19 - 20$ mag. The observations are unfiltered and collected in sequences of four images equally spaced over 30 minutes in each pointing (Drake et al. 2009). After several years of operation, the Catalina Surveys have over 200 observations for most of their variables (Drake et al. 2014), which makes the catalog largely complete. Given the limited magnitude overlap between the Catalina Surveys and DES, we only found 1463 of their 32775 RRab in our initial stellar catalog, but we identified 1185 ($\sim$81%) as such, with an AUC of 0.733.

ATLAS, a planetary defense initiative with a high cadence well suited for variability studies, recently released its first catalog of variable stars (Heinze et al. 2018). Thus far, ATLAS has at least 200 observations across two filters (c, o) over one-fourth of the sky. We select RRab stars from the ATLAS DR1 variable star catalog using the suggested CasJobs query in Appendix 10.2 of Heinze et al. (2018). As ATLAS is based in the Northern hemisphere and quite shallow compared to DES ($r \approx 20$ mag), we only have 484 of their 21061 RRab in our initial stellar catalog but identify 391 ($\sim$ 81%) as such, with an AUC of 0.635. This recovery rate is quite similar to the ones for Pan-STARRS and the Catalina Surveys.

Table 7. Description of Selected External RRL Catalogs and their Overlap with DES

Survey	Area [sq deg]	Filters	Depth	Observational cadence	$N_{\rm obs}$	RRab total	RRab in DES	% found
SDSS stripe 82	∼300	*ugriz*	g≤21	most observed every 2 days	70-90	447	238	75
Gaia DR2	all sky	G_{BP}, G, G_{RP}	G∼21	uneven, follows Gaia scanning law	12-240	140,000	4609	70
Pan-STARRS PS1	∼30000	*grizY*	r≤21.5	2 same-band obs. sep. by 25 min	∼67	35,000	1021	79
Catalina Surveys	∼9000	unfiltered	V≤19-20	4 obs. within 30 min	≳200	32,775	1463	81
ATLAS	∼13000	c, o	r∼20	4× per night	∼200	21061	484	81
DES single epoch	∼5000	*grizY*	g∼23.5	irregular	∼ 50*	5783	5783	–

NOTE—The details of DES are listed for comparison. (*): by the end of the survey (Y6)

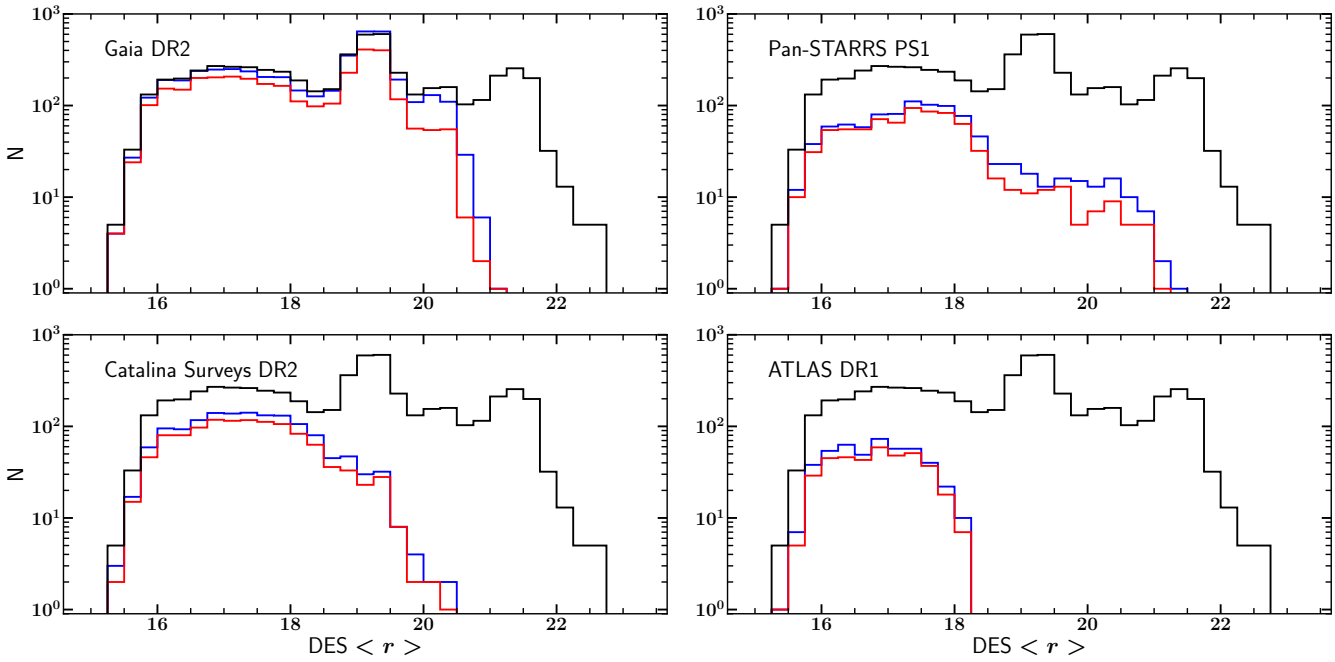

Figure 13. Histograms of magnitudes of RRab stars from external catalogs cross-matched with our DES initial stellar catalog, as a function of the extinction-corrected weighted average coadded DES r magnitude. *Top left: Gaia* DR2. *Top right*: Pan-STARRS PS1. *Bottom left*: Catalina Surveys DR2. *Bottom right:* ATLAS. Blue curves show the RRab from each catalog that were present in our sample before applying any cuts, while red curves show those that were identified as RRab in our analysis. Black curves show the distribution of DES RRab candidates and are the same in all panels. The overdensities at $r \approx 18.8$ and $r \approx 21.2$ correspond to the LMC outskirts and the Fornax dSph. Our catalog is deeper than the others by $\sim 1, 1, 2$ and 4.5 mag, respectively.

In addition to searching for RRab candidates with previous identifications from the aforementioned wide-field surveys, we also checked for overlaps near the Magellanic Clouds (Soszyński et al. 2016), the Fornax dSph (Bersier & Wood 2002), the Sculptor dSph (Martínez-Vázquez et al. 2016), in the General Catalogue of Variable Stars (Samus' et al. 2017), and in the SIMBAD database (Wenger et al. 2000). To the best of our knowledge, and based on publicly available catalogs, 1795 (nearly 31% of our sample) are newly-discovered RRab candidates. Although the external catalogs under consideration are not complete, the fraction of their RRab recovered by our analysis is consistent with our estimate of $\sim 75\%$ completeness. Our method is just as effective (if not more so) at recovering RRL from sub-optimally sampled data than the methods used in comparable surveys.

Although we recover most of the RRab in the aforementioned overlapping catalogs, we can see from the AUC of each of these that there is a marked degradation in our algorithm's performance when applied to light curves outside our S82 training set. Thus, we use their AUC values to construct a confidence interval for

the performance of our classifier. With the AUC of the training set and all four of these external cross-matches, we find a mean AUC of 0.728 with a standard deviation of 0.077. From this, we can determine that our classification methods have a lower effiency for fainter RRab. Unfortunately, we do not have well-characterized training data in a comparable filter system for fainter RRab, so we tested this with simulated light curves.

Table 8. Coefficients for Parameter Uncertainties

Value	p_0	p_1	p_2	p_3
$\sigma(\Delta P/P)$	4.8585×10^{-1}	-4.5912×10^{-2}	1.5636×10^{-3}	-1.7574×10^{-5}
$\sigma(\Delta a)$	1.5333	-1.5101×10^{-1}	5.3006×10^{-3}	-6.1034×10^{-5}

Note—The best fit 3rd degree polynomial is of the form $\sigma(\text{Value}) = p_0 + p_1N + p_2N^2 + p_3N^3$.

5.3. *Estimated Recovery Rates and Uncertainties from Simulated Data*

To estimate the robustness of our results for the noisier photometry at fainter magnitudes, we followed a method similar to Medina et al. (2018) and applied our method to simulated light curves with known light curve parameters in the DES filter system. We created the simulated light curves by sampling the smoothed templates of Sesar et al. (2010) in `gatspy` (VanderPlas & Ivezić 2015) with the DES cadence from different areas of the survey. We shifted these light curves to various distances by adding the appropriate distance modulus and inserting scatter in the observations based on the magnitude-dependent uncertainty relations we found in §3.1 (shown in Figure 2). Appendix F contains further details on the construction of these simulated light curves.

Figure 14 shows the recovery rates of both the classifier and the period as a function of magnitude and total number of observations. As expected, the recovery rate of our algorithm decreases significantly with increasing distance modulus. This is mostly due to the larger photometric uncertainties and fewer observations due to the brighter limiting magnitudes for the redder bands (see §2.1). We see that the accuracy of the period estimation decreases following the trend of increasing photometric errors shown in Figure 14, and dramatically improves with increasing total number of observations up to $N \sim 20$. As expected, the RRab classification accuracy follows a similar trend. We find that our template fitting recovers the true period to within 1% for 95% of the simulated light curves with N=20 observations.

Beyond assessing our classifier performance with these simulated light curves, we can also use them to estimate the uncertainties of the best fitting template parameters. To make sure we treat light curves with especially poor phase coverage separately, we divided the simulated light curves into groups based on their "flag_minmax" values (described in §5.1). Then, we subdivided those into bins of two observations and 0.5 magnitude wide in N and μ, respectively. In each of these bins, we calculate the fraction of light curves with period estimates within 1% of their input values for each phase sampling group. To quantify the uncertainty of the period estimates, we calculated the standard deviation of $\Delta P/P = P_{est} - P_{true}/P_{true}$, where the "est" subscript represents the parameter estimate from the template fitting and "true" represents the input value of the simulated light curve. Likewise, we calculated $\Delta a = a_{est} - a_{true}$ to quantify the uncertainty of the amplitude estimates. The number of light curves included in each bin differs widely, so we estimate the spread of these uncertainty values within each subgroup with jackknife resampling. These results are shown in Figure 14.

Other than fluctuations due to the small sample sizes in some of the bins, these values follow expected trends. When there are fewer observations to constrain the parameter values during the template fitting, both the period and the amplitude are more uncertain, with these values beginning to stabilize around N=20 observations. In distance space, the parameter estimates are generally low until $\mu \approx 20$, where the brighter detection limits of the redder filters decrease the number of observations in the light curves. We have very few simulated light curves that are missing observations near their maximum only or both their maximum and minimum (the blue and orange points in Figure 14), so we cannot draw any definitive conclusions about the effect of phase sampling on the estimation of these parameters. We assign these parameter uncertainties to the real RRab candidates based on the best fitting 3rd degree polynomial to the trends in N observations for all simulated light curves. We do not assign uncertainties to objects with $N > 43$ observations due to a lack of simulated data with that sampling. We also do not report these uncertainties for objects not identified as RRab by the classifier since these simulated light curves do not accurately represent the behavior of non-RRab. The coefficients of the best fitting polynomials are included in Table 8 and the uncertainties are included in the full catalog described in Appendix E.

The uncertainty of the remaining parameter ϕ is significantly more difficult to constrain. Phases for individual observations in the folded light curves are calculated using phase = (MJD/P) mod 1. Any small offset in

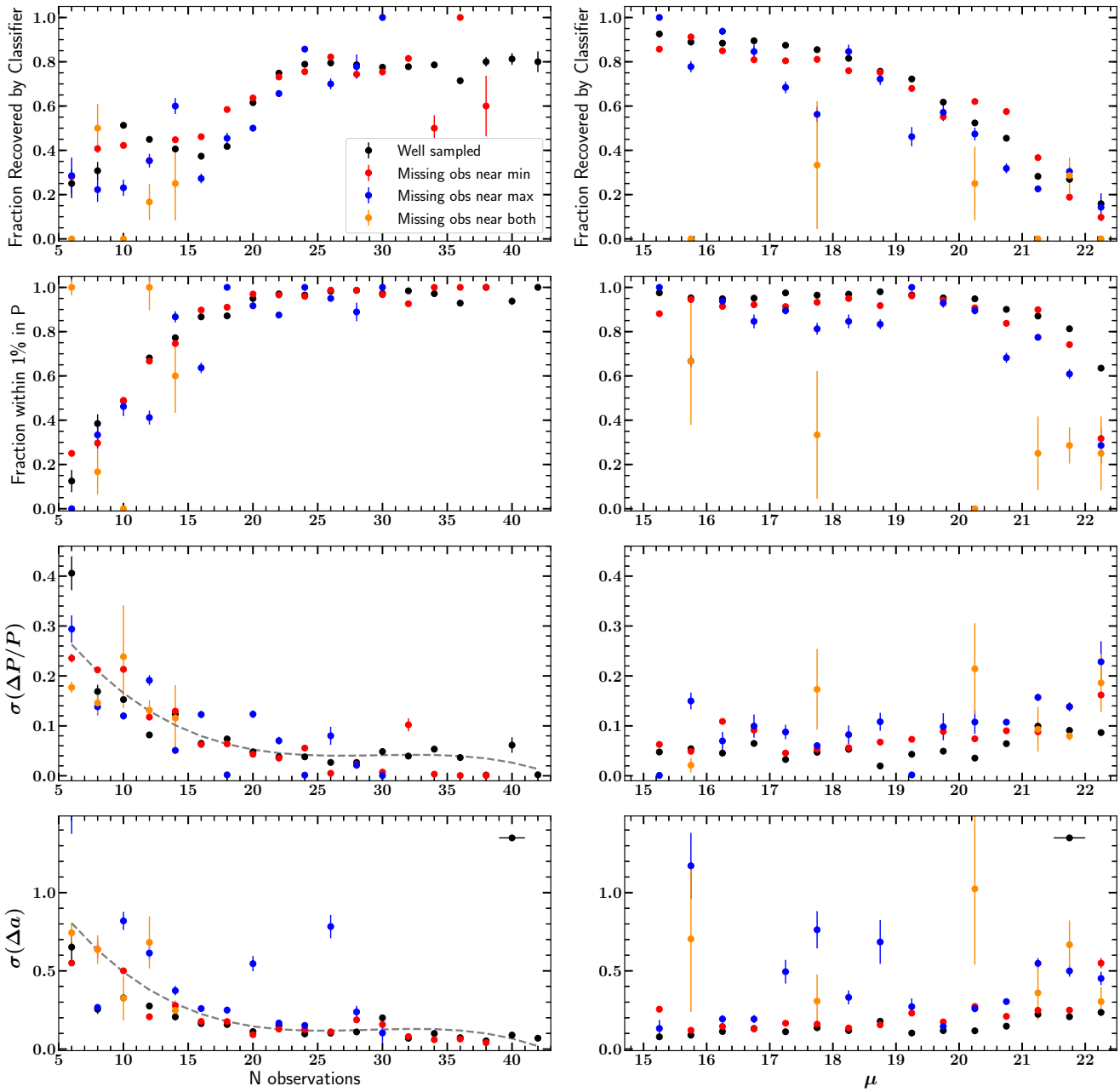

Figure 14. Recovery rates and parameter uncertainties as a function of the number of observations in the light curves and distance modulus μ. Colored points denote the behavior of the recovery fractions or parameter offsets for light curves with the phase sampling flags described in §5.1. Uncertainties on these values were estimated using jackknife resampling. The black points in the upper right corner of the bottom panels show the representative width of the bins in each column. Dashed grey lines show the best fitting 3rd degree polynomial to the trends shown by the combined simulated data used to assign uncertainties in the RRab catalog. The coefficients for these fits are listed in Table 8. *First row:* Fraction of simulated RRab light curves which received a classifier score ≥ 0.35. *Second row:* Fraction of estimated periods within 1% of the true input values. *Third row:* Standard deviation of the percent difference in period. *Bottom row:* Standard deviation of the offset in amplitude. *Note:* If not visible, uncertainties are smaller than the plotting symbols.

the period will compound over successive pulsations and result in a phase offset that varies over time. Even simulated light curves with $\Delta P/P < 0.0005$ (a difference < 1 minute) can yield $\Delta\phi \approx 0.5$ after three years when compared to the phases calculated using the input period. Thus, we do not report these uncertainties in ϕ as they require a level of period precision we do not attain even in light curves with $N > 20$. We caution against using the phases reported here for purposes other than plotting the template curves.

5.4. *Uncertainties in the Distance Moduli*

Since the absolute magnitudes of RRL depend on their metallicities (see Figure 14 in Marconi et al. 2015), the fixed [Fe/H] of our RRL model contributes systematic uncertainty to our distance estimates. Although the abundances of the individual RRL in our catalog are unknown, we can approximate the size and direction of this effect by comparing our results to those from an external catalog with metallicity measurements. The Catalina Surveys catalog of Torrealba et al. (2015) (hereafter T15) is convenient for this purpose because it has photometric estimates of [Fe/H] and a significant overlap with the DES survey footprint (although it has a brighter magnitude limit, see §5.2. We calculated the difference in distance moduli for 521 RRL in common between both catalogs, $\Delta\mu = \mu_{DES} - \mu_{T15}$, which we plot as a function of [Fe/H] in in Figure 15.

We split the sample into bins of 0.1 dex in metallicity and perform an iterative 3-σ clip from the median value using the `sigma_clip` function in `astropy.stats`. We fit a linear relation between $\Delta\mu$ and [Fe/H]:

$$\begin{aligned}\Delta\mu(\mathrm{DES} - \mathrm{T15}) &= (-0.058 \pm 0.003) \\ &+ (0.168 \pm 0.009)\ ([\mathrm{Fe/H}] + 1.5).\end{aligned} \tag{6}$$

The root-mean-square error (RMSE) of the fit is 0.06 mag, consistent with the standard deviation of $\Delta D/D$ between this work and Sesar et al. (2010) listed in Table 3. Thus, we estimate our statistical uncertainty in distance moduli to be $\sigma_{\mathrm{stat}} = 0.06$ mag.

As is evident from Figure 15, our algorithm systematically underestimates distances for very metal-poor RRL and overestimates distances for more metal-rich RRL. While this cross-matched sample illustrates how much of an effect an RRab's metallicity has on the accuracy of our distance estimates, we caution against using this sample to derive a metallicity correction to our distance estimates. The T15 RRL cover most of the spatial DES footprint, but our RRab catalog extends to fainter magnitudes than the magnitude limit probed by the T15 sample, which means we cannot assume the metallicity distribution of this sample accurately reflects that of the entire catalog. However, if we use this subsample and assume that the stellar halo metallicity distribution function is represented by a Gaussian with a mean of [Fe/H] $= -1.5$ dex and standard deviation of $\sigma = 0.3$ dex (Ivezić et al. 2008b) as in Sesar et al. (2017), we find a 1-σ systematic uncertainty of $\sigma_{[\mathrm{Fe/H}]} \approx 0.05$ mag in distance modulus. If we follow Medina et al. (2018) and estimate distance offsets for a metallicity shift of ± 0.5 dex and ± 1.0 dex in this subsample, we find a change in distance modulus of 0.08 mag and 0.17 mag, respectively. Again, we caution that the true distribution of metallicities in our full catalog is unknown, so these values are merely representative of the systematic uncertainty that would apply to particular stellar populations in the Milky Way halo. However, given our lack of metallicity information, we cannot quantify these systematic offsets without making such assumptions.

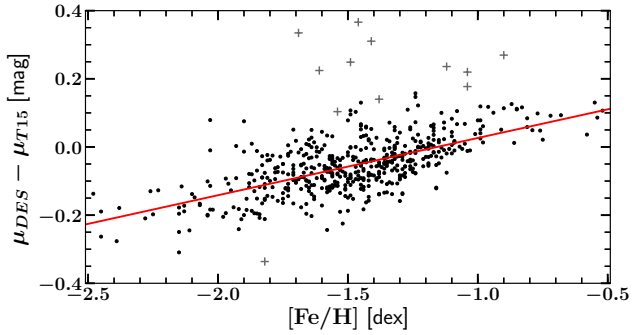

Figure 15. Difference in distance moduli between our work and Torrealba et al. (2015, T15) plotted against their photometric metallicity estimates. Stars that passed the 3-σ clip are plotted as black points, while the points that were removed are plotted as grey crosses. The best-fit linear relation fit to the clipped data is plotted in red. The RMSE of the fit is 0.06.

Another contribution to the systematic uncertainty we have not previously considered is the RRL evolution off the horizontal branch. We adopt the value $\sigma_{<V>} = 0.08$ mag, which Vivas & Zinn (2006) estimated from RRL in globular clusters (see their §4 and Figure 4). Adopting the halo metallicity distribution from Sesar et al. (2017), we add both sources of uncertainty in quadrature to arrive at $\sigma_{\mathrm{sys}} = (\sigma^2_{[\mathrm{Fe/H}]} + \sigma^2_{<V>})^{1/2} \approx 0.09$ mag.

We verify this estimate of systematic uncertainty by comparing our estimated distance moduli for various MW satellites with previously-published results. For the Fornax dSph, which has a horizontal branch [Fe/H]≈ -1.8 dex (Rizzi et al. 2007), our median distance modulus is 0.05 ± 0.04 mag closer than the values of $\mu = 20.72 \pm 0.04$ and $\mu = 20.72 \pm 0.06$ mag found by Greco et al. (2006) and Rizzi et al. (2007), respectively. In the

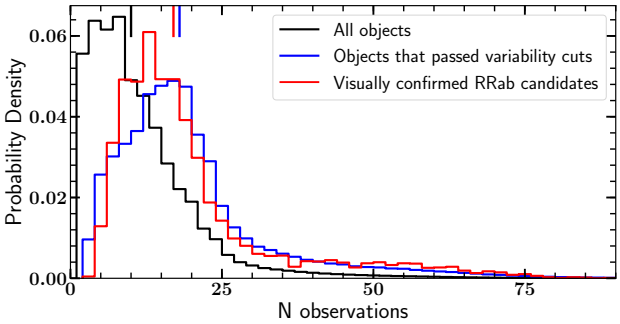

Figure 16. Histograms of the total number of observations across all bands for objects in the stellar sample (see §2.2) (black), objects which passed the variability cuts (blue), and the objects identified as RRab (red). Short line segments denote the median number of observations for each group (10, 18, and 17, respectively). Note that many of the RRab candidates have very few observations and would benefit from follow-up observations to confirm their nature.

case of the Sculptor dSph, we find a median distance modulus 0.13 ± 0.04 mag closer than the value of $\mu = 19.62 \pm 0.04$ mag from Martínez-Vázquez et al. (2016). This larger difference is likely due to the large spread in metallicity exhibited by Sculptor's stellar populations, $-2.3 \lesssim [Fe/H] \lesssim -1.5$ dex (Martínez-Vázquez et al. 2016). Our distance estimate to the LMC, based on the RRL we found in its outskirts, is 0.12 ± 0.09 mag closer than the $\mu = 18.52 \pm 0.09$ mag found by Clementini et al. (2003). The LMC also has a spread of metallicities for HB stars, centered on [Fe/H]≈ -1.5 dex with a dispersion of 0.4 dex (Clementini et al. 2003; Gratton et al. 2003). We expect that replacing the template's $M_b(P)$ relation with a calibrated *P-L-Z* or *P-L-C* relation in the DECam filters (K. Vivas et al., in prep.) will significantly reduce these offsets.

In summary, our distance moduli have 1-σ statistical and systematic uncertainties of 0.06 and 0.09 mag, respectively. The equivalent distance uncertainties are $\sim$2.8% (stat) and $\sim$4.2% (sys).

6. DISCUSSION

6.1. *Detection Biases*

The strength of DES lies in its wide-field coverage and depth, but the results presented here are limited by the low number of multiband observations. Figure 16 displays histograms of the total number of observations for all objects in the stellar sample (black), all objects passing variability cuts (blue), and all RRab candidates (red). The median number of total observations for each group, marked by a short colored line segment, is 10, 18, and 17, respectively. Note that most of our RRab have fewer observations than the N$\sim$20 observation threshold we saw from the simulation results in Figure 14. As future DES data releases will have an increased number of observations, we expect to find more RRL and have a more robust classification of the candidates presented here.

We note that the light curves used in this analysis typically had fewer total observations than the number expected from three years of DES data. We suspect that the total number of observations for the "stellar" sample is skewed by objects near the detection limits of DES, which suffer from noisy photometry and likely have few overlapping observations across all five filters. This low number of observations is also a result of the stringent quality cuts we applied on the single epoch photometry in §2.2. In future work, we aim to be more judicious in applying our photometric quality cuts so that we do not discard observations unnecessarily.

To verify that our sample is not affected by spatial fluctuations in the number of observations, we calculated the median number of total observations in each HEALPix of our Y3Q2 stellar data set. We show the median and the standard deviation of the total number of observations of light curves in each HEALPix in Figure 17. As expected, regions with the lowest number of observations fall near the edges of the survey footprint. Regions which have a median number of observations ≥ 25 correspond to the Science Verification region, in which 50 observations were made in the first year to demonstrate year 5 depth, and the DES Supernova fields, which are observed roughly weekly (e.g. DES Collaboration 2016). The linear patterns of constant Right Ascension are a result of the survey observation strategy (DES Collaboration 2018). Beyond these patterns, the DES photometry suffers in photometric completeness in crowded stellar fields near the central regions of nearby dSph galaxies and globular clusters, thus our catalog also suffers in completeness near those regions. Otherwise, the survey coverage is fairly uniform and we do not expect large scale trends in RRL detection outside of these fields of larger-than-average observation counts and dense stellar populations. We expect the addition of DES Y4-Y6 data to increase our detections of RRL considerably.

Some additional biases in our RRab sample are results of choices made to exclude non-RRab from our analysis. While we weighted our initial variability cuts by the photometric errors to make the cuts robust against spurious observations (see §3.2), using these error-weighted metrics biased our variable sample against RRab with smaller amplitudes located at larger distances. We also excluded some real RRab from our sample by limiting

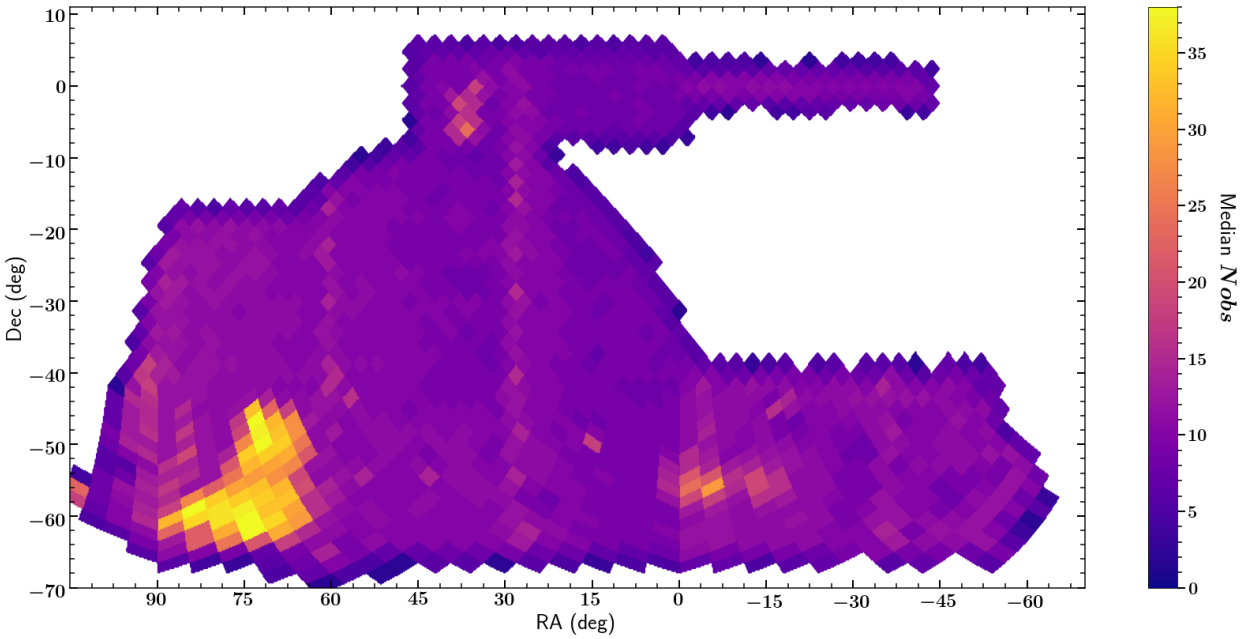

Figure 17. Spatial map of the median number of observations across all bands in each HEALPix. As expected, the regions with the fewest observations are near the edges of the survey region. The regions with a relatively large number of observations correspond to Science Verification or supernova fields. Outside these regions, the DES coverage is relatively uniform, but suffers from a small number of time series observations. Future studies of RRL will benefit from additional years of DES data.

the period range to 0.44 d $\leq P \leq$ 0.89 d to avoid the common 1-day alias.

6.2. *Spatial Distribution of the Candidates*

The spatial distribution of the 5783 visually validated RRab candidates is shown in Figure 18. We also plot these candidates as a function of their heliocentric distance in Figure 19. In both figures, the overdensities of RRab candidates associated with (in order of decreasing heliocentric distance): the Fornax dSph, the Sculptor dSph, and the the outskirts of the Large Magellanic Cloud are easily visible. We expect that the inclusion of a metallicity term into the model combined with additional epochs of DES observations in the next release of this catalog will enable further characterization of these and other substructures.

One of the largest strengths of the DES data set is its depth (see our comparisons to other wide-field surveys in Figure 13). This is extremely valuable for our understanding of the outer halo as the current census of RRL known beyond 100 kpc falls short of the thousands predicted by simulations (Sanderson et al. 2017). In this work, we identified 800 RRab candidates beyond 100 kpc (most of which have been previously discovered) and eight RRab candidates beyond 200 kpc, all of which are new discoveries. The coadded images and light curves for the candidates beyond 200 kpc are shown in Figure 20.

The three most distant visually verified RRab candidates in our sample have heliocentric distances of $\sim$ 231.6, 223.0, and 221.3 kpc. While these three stars are the most distant to-date RRab in the Milky Way, they are not the most distant RRL. Medina et al. (2018) recently found two RRc with larger distances (232.9 and 261.2 kpc) using data from the HiTS Survey (Förster et al. 2016). Even though the candidates in our RRab sample suffer from a small number of observations and require additional follow-up for confirmation, the fact that there are so many RRab beyond 100 kpc and three RRab beyond 220 kpc provide reasonable evidence that the Milky Way stellar halo extends at least out to 220 kpc. Future DES data releases and other upcoming deep surveys such as LSST will increase the census of known RRL at this distance, enabling further characterization of the outer halo.

6.3. *Applicability to LSST*

The next-generation large ground-based Large Synoptic Survey Telescope (LSST) (Ivezić et al. 2008) is set to begin full science operations in early 2023[11]. The most current LSST "Baseline Cadence" for its Wide-Fast-

[11] https://www.lsst.org/about/timeline

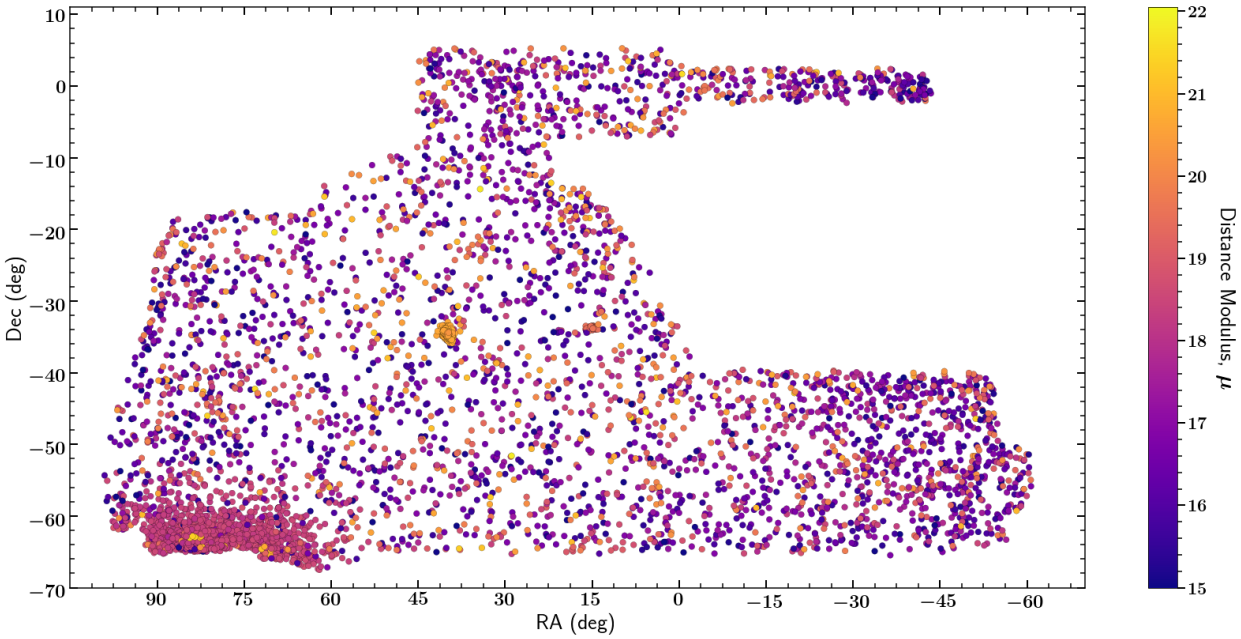

Figure 18. Map of 5783 visually accepted RRab candidates across the DES wide-field survey footprint. The RRab are marked by dots colored by distance modulus. Large MW satellite galaxies are easily distinguishable by their overdensities of RRab. The outskirts of the LMC are located near $(80°, -62°)$, the Fornax dSph is located near $(41°, -34°)$, and the Sculptor dSph is located near $(15°, -34°)$.

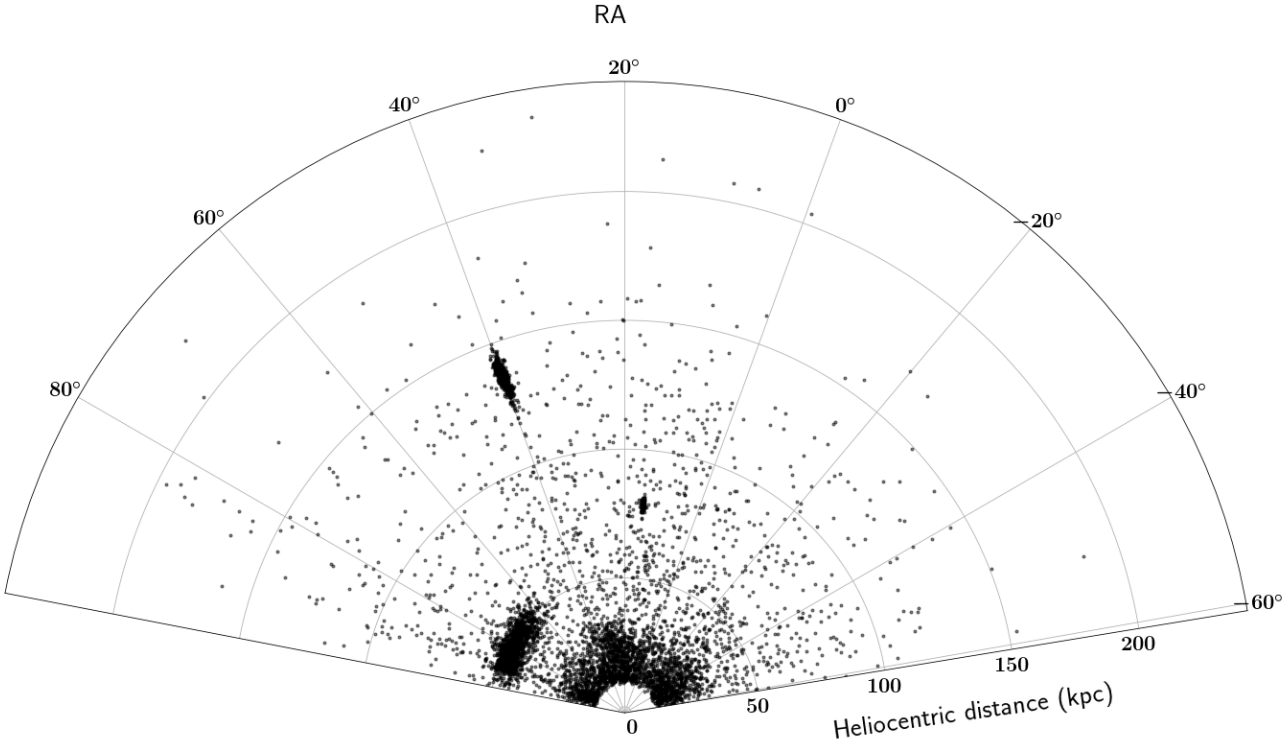

Figure 19. Radial Distribution of DES RRab stars. The overdensities associated with (in order of decreasing heliocentric distance) the Fornax dSph, the Sculptor dSph, and the periphery of the Large Magellanic Cloud are easily distinguishable. Note: The Sculptor and Fornax galaxies appear elongated due to uncertainties in the RRab distance moduli.

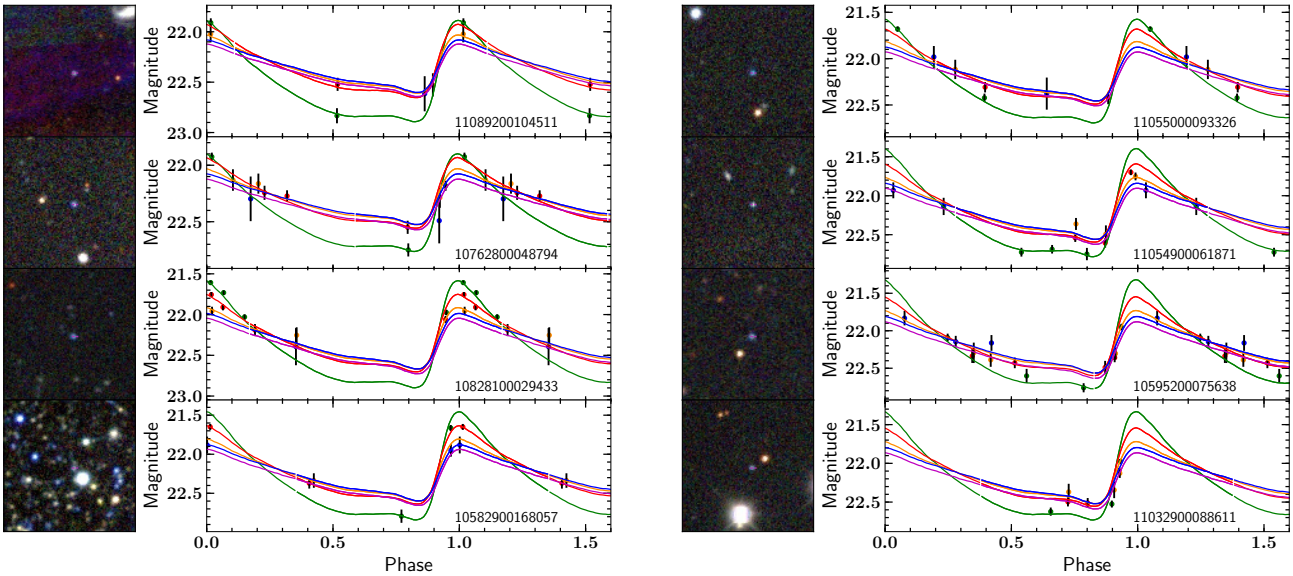

Figure 20. DES coadded images and representative light curves of visually-accepted candidates beyond 200 kpc, labeled with their Y3Q2 ID number. The observations and templates are colored by filter using the same convention as Figure 6. These very distant candidates have very few observations and will benefit from future DES data releases.

Deep Survey (WFD), which covers 18,000 deg^2 of the sky and comprises ∼85% of its total allocated observing time, is to image each field twice 40 minutes apart once every three days in a different filter. After 10 years of operation, each field is expected to have a median of (62, 88, 199, 201, 180, 180) visits in *ugrizy* with a single epoch depth of (23.14, 24.47, 24.16, 23.40, 22.23, 21.57) mag. Assuming these observations are spaced uniformly over 10 years, one can expect most light curves to have ∼80 multiband observations within the first year (LSST Science Collaboration et al. 2017)[12].

Oluseyi et al. (2012) found from their analysis of simulated LSST data that reliable RRL period estimation will require several years of operation, however, several multiband techniques have been developed since their publication. For instance, VanderPlas & Ivezić (2015) estimated accurate periods 64% of the time on downsampled S10 light curves with ∼55 observations and Sesar et al. (2017) accurately estimated periods for 85% of their PS1 training set with ∼67 observations. Our simulations show that our template fitting method is capable of estimating the correct periods to within 1% for 95% of the light curves with 20 total observations. Thus, our algorithm would be effective to identify potential RRab candidates (which would need followup for confirmation) within the first year of LSST operations. After the first year, the light curves in the WFD survey will be adequately sampled to use other multiband methods available in the literature.

[12] See the most current version of the draft of the LSST Observing Strategy white paper located at https://github.com/LSSTScienceCollaborations/ObservingStrategy.

7. SUMMARY

We have presented a new physically-motivated general multiband RRab template and a computationally-efficient fitting procedure. We combined this method with a random forest classifier to create a powerful technique that can robustly identify these variables even when fewer than 20 observations are available. Despite the poor cadence and sampling of DES data, we detected 5783 RRab candidates, 1795 (31%) of which are previously undiscovered to the best of our knowledge. The large quantity of RRL we recovered in common with overlapping external surveys such as *Gaia* DR2, Pan-STARRS, Catalina Surveys, and ATLAS provide strong evidence of the effectiveness of this algorithm. Although the number of observations is relatively uniform across the survey footprint in the DES Year 3 data, time series analyses like these will benefit immensely from the additional observations in future data releases. We make the template, these catalogs, and the light curves of the RRab candidates and the training sample available to the scientific community for future studies. Our method is especially useful for other multiband data sets which were not specifically designed for time series analysis.

The authors would like to thank the anonymous referees for their insightful and constructive comments. The authors acknowledge the Texas A&M University Brazos HPC cluster that contributed to the research re-

ported here: brazos.tamu.edu. KMS would like to thank Chris Stringer and Ryan Oelkers for helpful conversations. Part of this work was completed while coauthors ZL and JPL were research fellows in the Astrostatistics program at the Statistics and Applied Mathematical Sciences Center (SAMSI) in Fall 2016. ZL and JPL gratefully acknowledge SAMSIs support. C. Nielsen and J. L. Myron would like to acknowledge support from the NSF grant AST-1263034, "REU Site: Astronomical Research and Instrumentation at Texas A&M University." This research has made use of the SIMBAD database, operated at CDS, Strasbourg, France.

Funding for the DES Projects has been provided by the U.S. Department of Energy, the U.S. National Science Foundation, the Ministry of Science and Education of Spain, the Science and Technology Facilities Council of the United Kingdom, the Higher Education Funding Council for England, the National Center for Supercomputing Applications at the University of Illinois at Urbana-Champaign, the Kavli Institute of Cosmological Physics at the University of Chicago, the Center for Cosmology and Astro-Particle Physics at the Ohio State University, the Mitchell Institute for Fundamental Physics and Astronomy at Texas A&M University, Financiadora de Estudos e Projetos, Fundação Carlos Chagas Filho de Amparo à Pesquisa do Estado do Rio de Janeiro, Conselho Nacional de Desenvolvimento Científico e Tecnológico and the Ministério da Ciência, Tecnologia e Inovação, the Deutsche Forschungsgemeinschaft and the Collaborating Institutions in the Dark Energy Survey.

The Collaborating Institutions are Argonne National Laboratory, the University of California at Santa Cruz, the University of Cambridge, Centro de Investigaciones Energéticas, Medioambientales y Tecnológicas-Madrid, the University of Chicago, University College London, the DES-Brazil Consortium, the University of Edinburgh, the Eidgenössische Technische Hochschule (ETH) Zürich, Fermi National Accelerator Laboratory, the University of Illinois at Urbana-Champaign, the Institut de Ciències de l'Espai (IEEC/CSIC), the Institut de Física d'Altes Energies, Lawrence Berkeley National Laboratory, the Ludwig-Maximilians Universität München and the associated Excellence Cluster Universe, the University of Michigan, the National Optical Astronomy Observatory, the University of Nottingham, The Ohio State University, the University of Pennsylvania, the University of Portsmouth, SLAC National Accelerator Laboratory, Stanford University, the University of Sussex, Texas A&M University, and the OzDES Membership Consortium.

Based in part on observations at Cerro Tololo Inter-American Observatory, National Optical Astronomy Observatory, which is operated by the Association of Universities for Research in Astronomy (AURA) under a cooperative agreement with the National Science Foundation.

The DES data management system is supported by the National Science Foundation under Grant Numbers AST-1138766 and AST-1536171. The DES participants from Spanish institutions are partially supported by MINECO under grants AYA2015-71825, ESP2015-66861, FPA2015-68048, SEV-2016-0588, SEV-2016-0597, and MDM-2015-0509, some of which include ERDF funds from the European Union. IFAE is partially funded by the CERCA program of the Generalitat de Catalunya. Research leading to these results has received funding from the European Research Council under the European Union's Seventh Framework Program (FP7/2007-2013) including ERC grant agreements 240672, 291329, and 306478. We acknowledge support from the Australian Research Council Centre of Excellence for All-sky Astrophysics (CAASTRO), through project number CE110001020, and the Brazilian Instituto Nacional de Ciência e Tecnologia (INCT) e-Universe (CNPq grant 465376/2014-2).

This manuscript has been authored by Fermi Research Alliance, LLC under Contract No. DE-AC02-07CH11359 with the U.S. Department of Energy, Office of Science, Office of High Energy Physics. The United States Government retains and the publisher, by accepting the article for publication, acknowledges that the United States Government retains a non-exclusive, paid-up, irrevocable, world-wide license to publish or reproduce the published form of this manuscript, or allow others to do so, for United States Government purposes.

Facilities: CTIO: Blanco (DECam) (Flaugher et al. 2015)

Software: astropy (Astropy Collaboration 2018), easyaccess (Carrasco Kind et al. 2018), gatspy (VanderPlas & Ivezić 2015), Matplotlib (Hunter 2007), scikit-learn (Pedegrosa et al. 2011), SIMBAD (Wenger et al. 2000)

APPENDIX

A. RRL MODEL ASSUMPTIONS

The form for the RRL model is

$$m_b(t) = \mu + M_b(\omega) + R_b E(B-V) + a\gamma_b(\omega t + \phi) \tag{A1}$$

where the population parameters, common to all RRL, are:

$$\begin{aligned} M_b(\omega) &= \text{absolute magnitude in band } b \text{ for an RRL with frequency } \omega \\ R_b &= \text{total-to-selective extinction coefficient for band } b \\ \gamma_b &= \text{shape of the RRL light curve in band } b \end{aligned}$$

The object specific parameters, different for each RRL, are

$$\begin{aligned} \mu &= \text{distance modulus} \\ E(B-V) &= \text{reddening} \\ a &= \text{amplitude} \\ \omega &= \text{frequency } (1/P) \\ \phi &= \text{ phase} \end{aligned}$$

For one RRL, the time-series photometry can be written as $\{\{t_{bi}, m_{bi}, \sigma_{bi}\}_{i=1}^{n_b}\}_{b=1}^{B}$ where m_{bi} is the observed magnitude at time t_{bi} in filter b measured with (known) uncertainty σ_{bi}. The bands are indexed $1, \ldots, B$ instead of typical letters e.g., *ugriz*. The model and data are related by

$$m_{bi} = m_b(t_{bi}) + \epsilon_{bi}$$

where $\epsilon_{bi} \sim N(0, \sigma_{bi}^2)$, meaning the noise parameter ϵ_{bi} can be viewed as a random normal variable with mean=0 and $Var(\epsilon_{bi}) = \sigma_{bi}^2$.

This model assumes all RRL share a common shape by band and that RRL are strictly singly periodic functions. These assumptions are an approximation. For example, our model does not account for the amplitude and phase modulations which vary according to an additional period caused by the Blažhko effect (Blažko 1907). Rather than construct a perfectly accurate model, the goal is to construct a model with few free parameters that provides a better approximation to RRL variation than existing methods. For example, a simple sinusoid model fit to 5 filters has a total of 16 free parameters (5 means, 5 amplitudes, 5 phases, and 1 frequency) while providing only a very rough approximation to the steep rise and slow decline in brightness observed in RRL light curves. In contrast, this model provides a significantly better approximation while fitting for 5 free parameters (or 4 if light curves are corrected for extinction prior to fitting).

B. DETERMINING TEMPLATE POPULATION VALUES

We estimated the population parameters common to all RRL ($M_b(P)$,R_b, and γ_b) using a combination of theory and existing data sets.

For the filter-dependent extinction coefficients R_b, we assumed a Galactic reddening value of $R = 3.1$ from Fitzpatrick (1999). These extinction values R_b for both SDSS and DES filters are summarized in Table 9. Note that these values are only used if the templates are fit with light curves uncorrected for extinction. In general, it is better to fit with dust-corrected light curves because the model has one fewer free parameter and the uncertainty on distance is greatly reduced. In this work, we corrected the light curves for extinction prior to fitting the template (see §3.2).

To develop the $P-L$ relation for this work, we determined the values of $M_b(P)$ for *ugriz* using version 3.2 of the `BaSTI` synthetic horizontal branch generator[13], based on the evolutionary tracks of Pietrinferni et al. (2004, 2006). We

[13] Available at albione.oa-teramo.inaf.it/BASTI/WEB_TOOLS/HB_SYNT/

Table 9. Extinction Coefficients

Band	R_b	
b	SDSS	DES
u	4.799	N/A
g	3.737	3.665
r	2.587	2.464
i	1.922	1.804
z	1.430	1.380
Y	N/A	1.196

NOTE—Based on Fitzpatrick (1999). Note that these model dust coefficients differ from those listed in Schlegel, Finkbeiner & Davis (1998) and DES Collaboration (2018).

Table 10. $P-L$ Coefficients

Band b	$c_{\rm SDSS}$	$p_{1,\rm SDSS}$	$p_{2,\rm SDSS}$	$c_{\rm DES}$	$p_{1,\rm DES}$	$p_{2,\rm DES}$
u	1.889	-0.049	-0.319	—	—	—
g	0.767	0.167	-0.595	0.730	-0.020	-0.065
r	0.550	-0.637	-0.353	0.542	-0.739	0.997
i	0.505	-1.065	-0.202	0.522	-1.136	-0.057
z	0.510	-1.308	-0.231	0.520	-1.292	-0.535
Y	—	—	—	0.558	-1.392	0.657

generated synthetic absolute magnitudes for RRL spanning $-0.48 \leq \log P \leq 0.08$ with a metallicity of $[Fe/H] = -2.85$ to use as our starting values for $M_b(P)$. Then, we shifted the template curve in each filter to match the magnitude offsets shown in real SDSS and DES light curves. We parametrized these empirical $M_b(P)$ by using a quadratic period-absolute magnitude ($P-L$) relation at a fixed metallicity of [Fe/H] ≈ -1.5 (see §5.4 for an extended discussion of the template metallicity):

$$M_b(P) = c_{0b} + p_{1b}(log_{10}(P) + 0.2) + p_{2b}(log_{10}(P) + 0.2)^2 \tag{B2}$$

where the c, p_1 and p_2 values for the SDSS and DES filters are shown in Table 10.

The light curve shape in filter b γ_b is a function of phase that covers one pulsation period. We use RRL found by Sesar et al. (2010) to estimate γ_b for the SDSS filters. We assume the same shapes for DES *griz* and assume the DES Y band shape is the same as the z band shape.

To infer the shape, we first "fold" the well-sampled S82 RRL light curves from Sesar et al. (2010) into phase coordinates by taking the modulus of the Modified Julian Dates of the observations with respect to the pulsation period of each individual RRL. We "phase-align" the light curves by shifting them so that they all reach their maximum brightness at phase=0. We smooth the light curves by removing observations with photometric errors ≥ 0.2 and linearly interpolating them in equally spaced phase bins. Then, we shift all of the light curves so that the curves in each filter *ugriz* have an average magnitude value $m_b = 0$. We sample each light curve on a grid in phase space so that a single RRL is denoted by X_{tb} for $t = (1, \ldots, T)$ and $b = (1, \ldots, B)$ where t indexes the phase (the grid has $T = 100$ equally spaced phases) and b indexes the filter (total B filters).

Let X_{itb} be the magnitude for the ith RRL, at phase t in band b. Let $\gamma_b \in \mathbb{R}^T$ be the template in filter b. The template matrix of all five bands is thus defined as $\Gamma = (\gamma_1, \ldots, \gamma_B) \in \mathbb{R}^{T\times B}$. Let $a \in \mathbb{R}^n$ be the amplitudes for the n RRL in the SDSS S82 sample. Let $X_{i\cdot\cdot} \in \mathbb{R}^{T\times B}$ represent the phase-folded, shifted, and normalized photometry

described in the previous paragraph for the ith RRL. To determine the Γ matrix of the template shapes, we solve the following optimization problem:

$$\min_{\Gamma,a} \sum_{i=1}^{n} ||X_{i\cdot\cdot} - a_i\Gamma||_F^2 \tag{B3}$$

where $||a||_2 = 1$ (the Euclidean norm) for identifiability and F denotes the Frobenius norm, or the square root of the absolute squares of its elements. The resulting Γ matrix are the template shapes in each filter. To reflect the dependence of the RRL amplitudes on the filter in which they were observed, we rescale the templates so the peak–to–peak g-band amplitude is 1 and the amplitudes of the other filter shapes are fractions of the g-band amplitude. These shapes and population parameter values form the set of templates that we use for fitting in our analysis.

C. FITTING THE MODEL

In this section, we describe how to fit the model to the data. Directly using the inverse of the observation uncertainties as weights is known to be suboptimal when the templates are an approximation, see Long et al. (2017). We estimate a model error term σ_{me} which is then used in the least square fitting. To compute σ_{me}, we fit the template to all-well sampled SDSS RRL light curves and compute the difference between the squared residuals and the squared photometric error σ_{bi}^2. σ_{me} is the square root of the average of these differences. The value across all of the SDSS bands $ugriz$ is 0.0547.

The model is fit by minimizing a weighted sum of squares ("χ^2 minimization"). There are at most five free parameters μ, $E[B-V]$, a, ω, ϕ. The dust can be turned off in the fitting in which case $E(B-V)$ is set to 0. We perform a grid search across the frequency because the objective function is highly multimodal. At frequency ω in the grid, we solve for the four parameters $\mu, E(B-V), a, \phi$ using:

$$\min_{\mu,E(B-V),a,\phi} \sum_{b,i} \frac{(m_{bi} - M_b(\omega) - \mu - E(B-V)R_b - a\gamma_b(\omega t_{bi} + \phi))^2}{\sigma_{bi}^2 + \sigma_{me}^2} \tag{C4}$$

where b is the filter index and i is the epoch index. We use a block–relaxation method in which we alternate between minimizing across the $(\mu, E(B-V), a)$ parameters and minimizing across the ϕ parameter. The number of iterations can also be specified.

When minimizing across $(\mu, E(B-V), a)$ at fixed ϕ, the model is linear in $(\mu, E[B-V], a)$, so we find the closed-form weighted least squares solution. Occasionally the update will result in a negative amplitude. In this case, we do a random phase update in the next step (i.e. draw phase uniformly in [0,1]), rather than the Gauss–Newton method described below.

When minimizing across ϕ, with fixed $(\mu, E[B-V], a)$, we cannot analytically solve for ϕ. Instead we use a Gauss–Newton. Define

$$m_{bi}^* \equiv m_{bi} - M_b(\omega) - \mu - E(B-V)R_b \tag{C5}$$

and

$$\gamma_{bi}(\phi) \equiv a\gamma_b(\omega t_{bi} + \phi) \tag{C6}$$

Then the objective function which we seek to minimize is

$$g(\phi) = \sum_{b,i} \frac{(m_{bi}^* - \gamma_{bi}(\phi))^2}{\sigma_{bi}^2 + \sigma_{me}^2}. \tag{C7}$$

With $\phi^{(m)}$ as our current phase estimate, the Newton update has the form:

$$\phi^{(m+1)} = \phi^{(m)} - H(g)^{-1}(\phi^{(m)})\nabla(g)(\phi^{(m)}) \tag{C8}$$

where $\nabla(g)$ and $H(g)$ are the first and second derivatives of g. We have

$$\nabla(g) = \frac{\partial g}{\partial \phi} = -2\sum_{b,i} \frac{(m_{bi}^* - \gamma_{bi}(\phi))\gamma_{bi}'(\phi)}{\sigma_{bi}^2 + \sigma_{me}^2}. \tag{C9}$$

and

$$H(g) = \frac{\partial^2 g}{\partial \phi^2} = 2 \sum_{b,i} \frac{\left(\gamma'_{bi}(\phi)^2 - (m^*_{ib} - \gamma_{bi}(\phi))\gamma''_{bi}(\phi)\right)}{\sigma^2_{bi} + \sigma^2_{me}}. \tag{C10}$$

The Gauss–Newton update approximates $H(g)$ with

$$H^*(g) = 2 \sum_{b,i} \gamma'_{bi}(\phi)^2. \tag{C11}$$

where we substitute H^* for H in Equation C8, rather than using $H(g)$ in Equation C10. This is a standard approach in non–linear regression which avoids computation of γ''_{bi} and ensures that the second derivative is positive (see Section 14.4 in Lange (2010)). We approximate γ'_{bi} by storing numerical derivatives of the γ_b templates. At each new ω in the grid of frequency we obtain a warm start for the $\mu, E[B-V], a, \phi$ parameters by using estimates from the last frequency. We choose the frequency at which the RSS is minimized to be the parameters of the best-fitting template to the data.

D. TEMPLATE CODE PRODUCTS

The RRab template for both SDSS and DES filters and the fitting algorithm presented in this work are available at https://github.com/longjp/rr-templates. The template is originally implemented in R, but can be accessed in Python via the rpy2 module as was done in this analysis. Examples of the template usage are available in both R and Python 3, though it is also compatible with Python 2.

When using the template fitting functions, there are two options available to the user that impact the values returned by the fits. As described in the previous section, the template model includes an optional dust term. The model already includes extinction coefficients appropriate for both the DES and SDSS filter systems and can estimate the amount of extinction affecting the light curve as one of the parameters. However, if the light curves to be fit are already corrected for dust extinction prior to fitting, this parameter can be turned off to reduce the number of estimated parameters from 5 to 4 and thus improve the quality of the fits. This option is included in the code to allow the user to choose the option most appropriate for their data.

The other option determines whether or not to use the uncertainties associated with the individual observations in the light curve when performing the fits. In this analysis, we rescaled the uncertainties and used them when fitting the template to our data. However, if one suspects that the magnitude uncertainties in the light curves are misestimated or they simply aren't available, this option can be turned off and the uncertainties will not be used. We leave this option open to the user.

The repository contains examples explaining how to fit the template to both DES and SDSS light curves using all combinations of these options. To make this code accessible to a variety of users, these examples are included as R and Python Jupyter notebooks.

E. DATA PRODUCTS

To enable further work with similar data, we provide all of the RRab candidate and training light curves at https://des.ncsa.illinois.edu/releases/other/y3-rrl and an extended version of Table 6. These light curves have already had their photometric uncertainties rescaled as described in §3.1 and include both the dust-corrected and uncorrected magnitudes in each band. The light curves are indexed by their DES Y3Q2 QUICK_OBJECT_ID numbers, which are included as a column in the full data table.

A description of the columns in the included data table is shown in Table E and example sample selection criteria are shown in Table 11. All of this information is also available in the documentation at https://des.ncsa.illinois.edu/releases/other/y3-rrl.

F. SIMULATING DES RRAB LIGHT CURVES

Because DES is deeper than most overlapping surveys and much deeper than our training set in S82, we use simulated light curves to estimate the recovery rates at fainter magnitudes. We create the simulated light curves as follows:

1) We generate light curve shapes from all 379 smoothed RRab light curve templates from Sesar et al. (2010) with `gatspy`'s `RRLyraeGenerated` function (VanderPlas & Ivezić 2015). These generated template light curves already

Table 11. Description of Data Columns

Column Name	Description
QUICK_OBJECT_ID	DES Y3Q2 ID Number
COADD_OBJECT_ID	DES DR1 ID Number
RA	Right Ascension in degrees (J2000)
DEC	Declination in degrees (J2000)
p_ab	Classifier RRab score
EBV	Extinction value from Schlegel, Finkbeiner, & Davis 1998
$< g > (rizY)$	mean $g(rizY)$ magnitude measured in light curves (not extinction corrected)
nobs_$g(rizY)$	Number of observations in DES $g(rizY)$ in object's light curve
nobs	Total number of observations in final light curve
period_0(1,2)	Period of 1st(2nd,3rd) best fitting template (days)
sigma_dp_p	Uncertainty in $\Delta P/P$ of the best fit period
amp_0(1,2)	Amplitude of 1st(2nd,3rd) best fitting template (mag)
sigma_da	Uncertainty in Δa of the best fit amplitude
mu_0(1,2)	Distance Modulus of 1st(2nd,3rd) best fitting template (mag)
phase_0(1,2)	Phase offset of 1st(2nd,3rd) best fitting template
rss_0(1,2)	Residual Sum of Squares (RSS) of 1st(2nd,3rd) best fit template
chi2_g(rizY)	Reduced chi squared of light curve from constant value in DES $g(rizY)$
sig_g(rizY)	"Significance" of light curve in DES $g(rizY)$
rss_dof_0(1,2)	RSS per degree of freedom of 1st(2nd,3rd) best fitting template
lchi_med	Median Log(reduced chi squared) across DES grizY
rss_lchi_med	(RSS/dof)/ Median Log(reduced chi squared) across DES $grizY$
amp_rss_0(1,2)	Amplitude/(RSS/dof) for 1st(2nd,3rd) best fitting template
f_dist1_0(1,2)	Distance of 1st(2nd,3rd) best fit Amplitude/Period from Sesar et al. (2010) Oosterhoff I relation
f_dist2_0(1,2)	Distance of 1st(2nd,3rd) best fit Amplitude/Period from Sesar et al. (2010) division between the Oosterhoff I and II groups
kappa_0(1,2)	von Mises-Fisher concentration parameter of phases for 1st(2nd,3rd) best fitting template
flag_minmax	Light curve sampling flag, 0:$\geq$ 2 obs at max and min, 1:<2 obs at min, 1:<2 obs at max, 3:<2 obs at max and min
GaiaDR2_ID	Gaia DR2 source_id for cross-matched RRab in Clementini et al. (2018)
PS1_RRab	Present in Sesar et al. (2017) RRab catalog? 0 = no, 1 = yes
CSDR2_ID	ID number from associated Catalina Surveys DR2 RRab catalogs
ATLAS_ID	ID from Heinze et al. (2018) ATLAS RRab catalog
SDSS_ID	SDSS DR7 ID
SDSS_class	classification from S82 studies from Sesar et al. (2010) or Ivezić et al. (2007)
Simbad_ID	Object identifier in `SIMBAD` database (Wenger et al. 2000)
Simbad_class	Object classification in `SIMBAD` database (Wenger et al. 2000)
OGLE_ID	OGLE ID from LMC and SMC RRAb catalogs from Soszyński et al. (2016)
CEMV_ID	ID from Martínez-Vázquez et al. (2016)
GCVS5_ID	Cross matched ID with the General Catalogue of Variable Stars v5.1 (Samus' et al. 2017)
GCVS5_Type	Variable type for cross matched object as listed in the General Catalogue of Variable Stars v5.1 (Samus' et al. 2017)
hpx32	Object's Healpix (nside=32) used to subdivide the light curves
comments	Comments from first author during visual validation followed by a reason code
train	Identifies objects used to train the random forest classifier. 0 = no, 1 = yes
rrab	Identified as a high confidence RRab in this study. 0 = no, 1 = yes
filepath	filepath to the object's light curve

Note—All of these features are included in the table included in the data products available at https://des.ncsa.illinois.edu/releases/other/y3-rrl.
For the "comments" column, the possible comments and their meanings are as follows:
confirmed – convincing RRab
no_: rejected followed by a reason code
maybe_: ambiguous candidate followed by a reason code
missing_image missing coadd image in SkyViewer
Reason codes:
galaxy – DES image showed galaxy
bad_fit – poor template fit
n_points – few observations and/or poor phase coverage
crowding – object is close to another object
misc – miscellaneous reasons

Table 12. Example Selection Criteria

To select:	Choose:
Objects that were identified as RRab by the classifier	p_ab ≥ 0.35
Objects found in Clementini et al. (2018) Gaia DR2 RRab catalog	GaiaDR2_ID > 0
Objects found in S82	SDSS_class $!= 0$
Objects found in Sesar et al. (2017) Pan-STARRS catalog	PS1_RRab > 0
Objects found in Catalina Surveys RRab catalog	CSDR2_ID $!= 0$
Objects found in ATLAS RRab catalog	ATLAS_ID $!= 0$

include the measured period and amplitude from their real observed counterpart light curves.

2) We shift the light curves into the DES filter system using the DES-SDSS filter transformation relations[14]:

$$g_{\rm DES} = g_{\rm SDSS} + 0.001 - 0.075(g-r)_{\rm SDSS} \qquad RMS = 0.021 \text{ per star}$$
$$r_{\rm DES} = r_{\rm SDSS} - 0.009 - 0.069(g-r)_{\rm SDSS} \qquad RMS = 0.021 \text{ per star}$$
$$i_{\rm DES} = i_{\rm SDSS} + 0.014 - 0.214(i-z)_{\rm SDSS} - 0.096(i-z)^2_{\rm SDSS} \qquad RMS = 0.023 \text{ per star}$$
$$z_{\rm DES} = z_{\rm SDSS} + 0.022 - 0.068(i-z)_{\rm SDSS} \qquad RMS = 0.025 \text{ per star}$$
$$Y_{\rm DES} = z_{\rm SDSS} + 0.045 - 0.306(i-z)_{\rm SDSS} \qquad RMS = 0.030 \text{ per star}$$

3) We use the distance estimates from Sesar et al. (2010) in the distance modulus equation to shift the light curves to a distance of 10 pc so that the light curves reflect the absolute magnitudes of the stars.

4) We then sample the photometric measurements in the light curve at phases corresponding to the real DES cadence. The cadence is randomly selected from 1808 distinct fields in the DES wide-field footprint with unique observation times. This results in a light curve that is sampled in the same manner as the light curve for a real object somewhere in the DES footprint.

5) To simulate the effects of distance on our recovery, we shift the downsampled light curve magnitudes to the apparent magnitudes they would have at a specified distance within the range that DES could detect in the single epoch images. Once the magnitudes are shifted, we remove any magnitudes that are fainter than the median magnitude depth for that filter in the DES single epoch images: $g \sim 23.5$, $r \sim 23.3$, $i \sim 22.8$, $z \sim 22.1$, $Y \sim 20.7$ (DES Collaboration 2018). Thus, the light curves reflect the magnitude limits of each band at fainter magnitudes.

6) Last, we assign a photometric uncertainty to each observation in the light curves. Following a method similar to Medina et al. (2018), we calculate the standard deviation of error-rescaled light curves in the survey region as a function of their mean magnitudes in each band as shown in Figure 2. We apply a shift in magnitude to the simulated observations using a Gaussian distribution with the appropriate standard deviation for that magnitude and band.

Following this procedure, we created 5685 simulated RRab light curves. Since the photometric uncertainties were sampled from the rescaled uncertainties we applied to the real data, there was no need to rescale the errors using the method described in Section 3.1. Because we did not also simulate non-variable light curves to analyze alongside the simulated RRab light curves, rescaling the errors using the same procedure would have removed real variable objects. We fit the template to all 4751 simulated light curves which passed the initial variability cuts and had at least 5 observations and used the results to determine our detection efficiency. Results of this analysis are detailed in Section 5.3.

[14] http://www.ctio.noao.edu/noao/node/5828#transformations

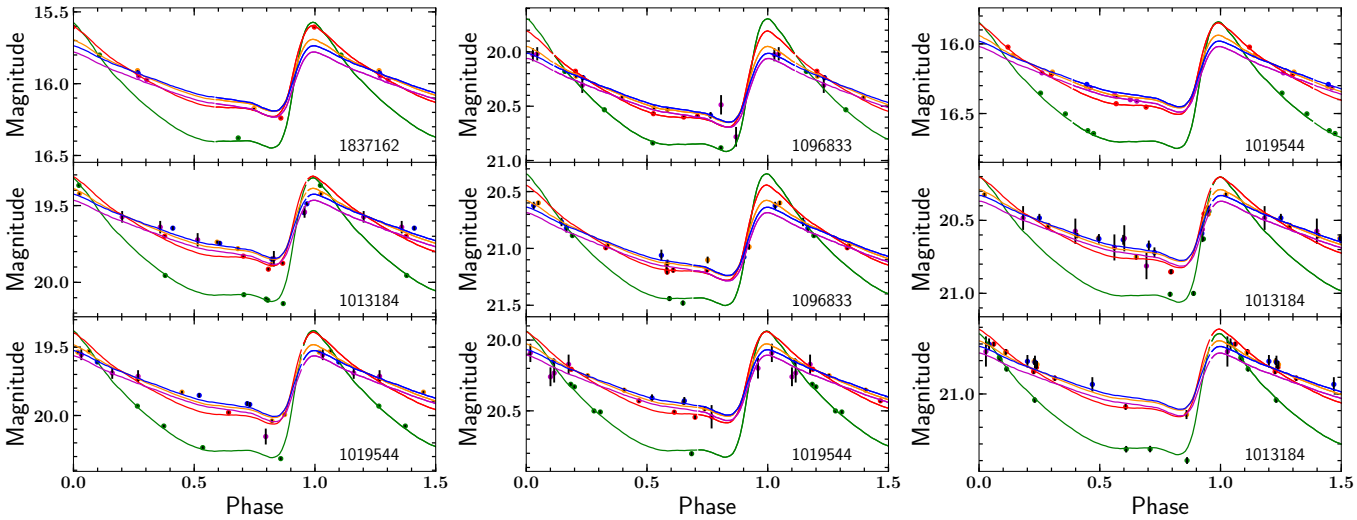

Figure 21. Examples of simulated RRL light curves labeled with the ID number of the SDSS light curve used to generate them. The observations and templates are colored by filter using the same convention as Figure 6.

REFERENCES

Abbott, T. M. C., Abdalla, F. B., Allam, S., et al. 2018, arXiv:1801.03181

Amit, Y. and Geman, D., Neural Computation, 9, 1545

The Astropy Collaboration: Price-Whelan, A. M., Sipőcz, B. M., et al. 2018, arXiv:1801.02634

Baker, M., & Willman, B. 2015, AJ, 150, 160

Balbinot, E., Yanny, B., Li, T. S., et al. 2016, ApJ, 820, 58

Bechtol, K., Drlica-Wagner, A., Balbinot, E., et al. 2015, ApJ, 807, 50

Belokurov, V., Zucker, D. B., Evans, N. W., et al. 2006, ApJL, 642, L137

Belokurov, V., Zucker, D. B., Evans, N. W., et al. 2006, ApJL, 647, L111

Belokurov, V., Zucker, D. B., Evans, N. W., et al. 2007, ApJ, 654, 897

Belokurov, V., Evans, N. W., Bell, E. F., et al. 2007, ApJL, 657, L89

Belokurov, V., Walker, M. G., Evans, N. W., et al. 2008, ApJL, 686, L83

Belokurov, V., Walker, M. G., Evans, N. W., et al. 2009, MNRAS, 397, 1748

Belokurov, V., Walker, M. G., Evans, N. W., et al. 2010, ApJL, 712, L103

Bergemann, M., Sesar, B., Cohen, J. G., et al. 2018, Nature, 555, 334

Bernard, E. J., Ferguson, A. M. N., Schlafly, E. F., et al. 2014, MNRAS, 443, L84

Bernard, E. J., Ferguson, A. M. N., Schlafly, E. F., et al. 2016, MNRAS, 463, 1759

Bersier, D., & Wood, P. R. 2002, AJ, 123, 840

Bertin, E., & Arnouts, S. 1996, A&AS, 117, 393

Blažko, S., 1907, AN, 175, 325

Boettcher, E., Willman, B., Fadely, R., et al. 2013, AJ, 146, 94

Breiman, L. 2001, Machine Learning, 45, 5, doi: 10.1023/A:1010933404324

Bullock, J. S., & Johnston, K. V. 2005, ApJ, 635, 931

Burke, D. L., Rykoff, E. S., Allam, S., et al. 2018, AJ, 155, 41

Cáceres, C., & Catelan, M. 2008, ApJS, 179, 242-248

Carrasco Kind M., Drlica-Wagner A., Koziol A. M. G., Petravick D., 2018, arXiv e-prints, arXiv:1810.02721

Catelan, M. 2009, Ap&SS, 320, 261

Catelan, M., Pritzl, B. J., & Smith, H. A. 2004, ApJS, 154, 633

Catelan, M., & Smith, H. A. 2015, Pulsating Stars (Wiley-VCH), 2015,

Chambers, K. C., Magnier, E. A., Metcalfe, N., et al. 2016, arXiv:1612.05560

Clementini, G., Gratton, R., Bragaglia, A., et al. 2003, AJ, 125, 1309

Clementini, G., Ripepi, V., Molinaro, R., et al. 2018, arXiv:1805.02079

Coupon, J., Kilbinger, M., McCracken, H. J., et al. 2012, A&A, 542, A5

Deason, A. J., Belokurov, V., Hamren, K. M., et al. 2014, MNRAS, 444, 3975

The Dark Energy Survey Collaboration 2005, arXiv:astro-ph/0510346

Dark Energy Survey Collaboration, Abbott, T., Abdalla, F. B., et al. 2016, MNRAS, 460, 1270
Desai, S., Armstrong, R., Mohr, J. J., et al. 2012, ApJ, 757, 83
Diehl, H. T., Neilsen, E., Gruendl, R., et al. 2016, Proc. SPIE, 9910, 99101D
Drake, A. J., Djorgovski, S. G., Mahabal, A., et al. 2009, ApJ, 696, 870
Drake, A. J., Catelan, M., Djorgovski, S. G., et al. 2013, ApJ, 763, 32
Drake, A. J., Catelan, M., Djorgovski, S. G., et al. 2013, ApJ, 765, 154
Drake, A. J., Graham, M. J., Djorgovski, S. G., et al. 2014, ApJS, 213, 9
Drake, A. J., Djorgovski, S. G., Catelan, M., et al. 2017, MNRAS, 469, 3688
Drlica-Wagner, A., Bechtol, K., Rykoff, E. S., et al. 2015, ApJ, 813, 109
Fiorentino, G., Bono, G., Monelli, M., et al. 2015, ApJL, 798, L12
Fisher, R. 1953, Proceedings of the Royal Society of London Series A, 217, 295
Fitzpatrick, E. L. 1999, PASP, 111, 63
Flaugher, B., Diehl, H. T., Honscheid, K., et al. 2015, AJ, 150, 150
Förster, F., Maureira, J. C., San Martín, J., et al. 2016, ApJ, 832, 155
Gaia Collaboration, Prusti, T., de Bruijne, J. H. J., et al. 2016, A&A, 595, A1
Gilbank, D. G., Gladders, M. D., Yee, H. K. C., & Hsieh, B. C. 2011, AJ, 141, 94
Górski, K. M., Hivon, E., Banday, A. J., et al. 2005, ApJ, 622, 759
Gratton, R. G., Bragaglia, A., Carretta, E., et al. 2003, A&A, 408, 529
Greco C.et al., 2006, in Valls-Gabaud D. Chavez M., eds, ASP Conf. Ser. Vol., Resolved Stellar Populations. Astron. Soc. Pac., San Francisco, in press (0507244)
Grillmair, C. J. 2006, ApJL, 645, L37
Grillmair, C. J. 2009, ApJ, 693, 1118
Grillmair, C. J., & Carlin, J. L. 2016, Tidal Streams in the Local Group and Beyond, 420, 87
Heinze, A. N., Tonry, J. L., Denneau, L., et al. 2018, arXiv:1804.02132
Hernitschek, N., Schlafly, E. F., Sesar, B., et al. 2016, ApJ, 817, 73
High, F. W., Stubbs, C. W., Rest, A., Stalder, B., & Challis, P. 2009, AJ, 138, 110
Holl, B., Audard, M., Nienartowicz, K., et al. 2018, arXiv:1804.09373
Hunter, J. D. 2007, CiSE, 9, 3
Irwin, M. J., Belokurov, V., Evans, N. W., et al. 2007, ApJL, 656, L13
Ivezić, Ž., Lupton, R., Schlegel, D., et al. 2004, Satellites and Tidal Streams, 327, 104
Ivezić, Ž., Lupton, R. H., Schlegel, D., et al. 2004, Astronomische Nachrichten, 325, 583
Ivezić, Ž., Smith, J. A., Miknaitis, G., et al. 2007, AJ, 134, 973
Ivezić, Ž., Kahn, S. M., Tyson, J. A., et al. 2008, arXiv:0805.2366
Ivezić, Ž., Sesar, B., Jurić, M., et al. 2008, ApJ, 684, 287
Jupp, P. E., Mardia, K. V. 1989, Annals of Statistics, 7, 3
Kaluzny, J., Stanek, K. Z., Krockenberger, M., et al. 1998, AJ, 115, 1016
Kelly, P. L., von der Linden, A., Applegate, D. E., et al. 2014, MNRAS, 439, 28
Kim, D., Jerjen, H., Mackey, D., Da Costa, G. S., & Milone, A. P. 2015, ApJL, 804, L44
King, D. S., & Cox, J. P. 1968, PASP, 80, 365
Koposov, S. E., Irwin, M., Belokurov, V., et al. 2014, MNRAS, 442, L85
Koposov, S. E., Belokurov, V., Torrealba, G., & Evans, N. W. 2015, ApJ, 805, 130
Koposov, S. E., Walker, M. G., Belokurov, V., et al. 2018, MNRAS, 479, 5343
Laevens, B. P. M., Martin, N. F., Ibata, R. A., et al. 2015, ApJL, 802, L18
Laevens, B. P. M., Martin, N. F., Bernard, E. J., et al. 2015, ApJ, 813, 44
Lange, K. *Numerical Analysis for Statisticians*. Springer
Li, T. S., Balbinot, E., Mondrik, N., et al. 2016, ApJ, 817, 135
Long, J. P. 2017, *EJS*, 11, 1464, doi:10.1214/17-EJS1255
LSST Science Collaboration, Marshall, P., Anguita, T., et al. 2017, arXiv:1708.04058, https://github.com/LSSTScienceCollaborations/ObservingStrategy
Luque, E., Queiroz, A., Santiago, B., et al. 2016, MNRAS, 458, 603
Luque, E., Pieres, A., Santiago, B., et al. 2017, MNRAS, 468, 97
MacDonald, E. C., Allen, P., Dalton, G., et al. 2004, MNRAS, 352, 1255
Majewski, S. R., Ostheimer, J. C., Rocha-Pinto, H. J., et al. 2004, ApJ, 615, 738
Marconi, M., Coppola, G., Bono, G., et al. 2015, ApJ, 808, 50
Martínez-Vázquez, C. E., Stetson, P. B., Monelli, M., et al. 2016, MNRAS, 462, 4349

Martínez-Vázquez, C. E., Monelli, M., Bernard, E. J., et al. 2017, ApJ, 850, 137
Mateu, C., Vivas, A. K., Zinn, R., Miller, L. R., & Abad, C. 2009, AJ, 137, 4412
Mateu, C., Vivas, A. K., Downes, J. J., et al. 2012, MNRAS, 427, 3374
Mateu, C., Read, J. I., & Kawata, D. 2018, MNRAS, 474, 4112
McConnachie, A. W. 2012, AJ, 144, 4
Medina, G. E., Muñoz, R. R., Vivas, A. K., et al. 2017, ApJ, 845, L10.
Medina, G. E., Muñoz, R. R., Vivas, A. K., et al. 2018, ApJ, 855, 43
Mondrik, N., Long, J. P., & Marshall, J. L. 2015, ApJL, 811, L34
Morganson, E., Gruendl, R. A., Menanteau, F., et al. 2018, PASP, 130, 074501
Neilsen, E., & Annis, J. 2014, Astronomical Data Analysis Software and Systems XXIII, 485, 77
Newberg, H. J., Yanny, B., Rockosi, C., et al. 2002, ApJ, 569, 245
Odenkirchen, M., Grebel, E. K., Rockosi, C. M., et al. 2001, ApJL, 548, L165
Oluseyi, H. M., Becker, A. C., Culliton, C., et al. 2012, AJ, 144, 9
Oosterhoff, P. T. 1939, The Observatory, 62, 104
Pedegrosa, F., Varoquaux, G., Gramfort, A., et al. 2011, JMLR, 12, 2825
Pickering, E. C., Colson, H. R., Fleming, W. P., & Wells, L. D. 1901, ApJ, 13, 226P
Pieres, A., Santiago, B. X., Drlica-Wagner, A., et al. 2017, MNRAS, 468, 1349
Pietrinferni, A., Cassisi, S., Salaris, M., & Castelli, F. 2004, ApJ, 612, 168
Pietrinferni, A., Cassisi, S., Salaris, M., & Castelli, F. 2006, ApJ, 642, 797
Prudil, Z., Grebel, E. K., Dékány, I., & Smolec, R. 2018, MNRAS, 480, 669
Richards, J. W., Starr, D. L., Butler, N. R., et al. 2011, ApJ, 733, 10
Rizzi, L., Held, E. V., Saviane, I., Tully, R. B., & Gullieuszik, M. 2007, MNRAS, 380, 1255
Rocha-Pinto, H. J., Majewski, S. R., Skrutskie, M. F., Crane, J. D., & Patterson, R. J. 2004, ApJ, 615, 732
Saha, A., & Vivas, A. K. 2017, AJ, 154, 231
Sakamoto, T., & Hasegawa, T. 2006, ApJL, 653, L29
Samus', N. N., Kazarovets, E. V., Durlevich, O. V., Kireeva, N. N., & Pastukhova, E. N. 2017, Astronomy Reports, 61, 80
Sanderson, R. E., Secunda, A., Johnston, K. V., & Bochanski, J. J. 2017, MNRAS, 470, 5014
Schlafly, E. F., & Finkbeiner, D. P. 2011, ApJ, 737, 103
Schlegel, D. J., Finkbeiner, D. P. and Davis, M. 1998, ApJ, 500, 525
Sesar, B., Ivezić, Ž., Lupton, R. H., et al. 2007, AJ, 134, 2236
Sesar, B., Ivezić, Ž., Grammer, S. H., et al. 2010, ApJ, 708, 717
Sesar, B., Banholzer, S. R., Cohen, J. G., et al. 2014, ApJ, 793, 135
Sesar, B., Hernitschek, N., Mitrović, S., et al. 2017, AJ, 153, 204
Sharma, S., Johnston, K. V., Majewski, S. R., et al. 2010, ApJ, 722, 750
Shipp, N., Drlica-Wagner, A., Balbinot, E., et al. 2018, arXiv:1801.03097
Soszyński et al., 2016, AcA, 66, 131
Sra, S. 2016, arXiv e-prints , arXiv:1605.00316.
Stetson, P. B. 1996, PASP, 108, 851
Stetson, P. B., Fiorentino, G., Bono, G., et al. 2014, PASP, 126, 616
Süveges, M., Sesar, B., Váradi, M., et al. 2012, MNRAS, 424, 2528
Tonry, J. L., Denneau, L., Heinze, A. N., et al. 2018, PASP, 130, 064505
Torrealba, G., Catelan, M., Drake, A. J., et al. 2015, MNRAS, 446, 2251
Torrealba, G., Koposov, S. E., Belokurov, V., & Irwin, M. 2016, MNRAS, 459, 2370
Torrealba, G., Koposov, S. E., Belokurov, V., et al. 2016, MNRAS, 463, 712.
Torrealba, G., Belokurov, V., Koposov, S. E., et al. 2018, MNRAS, 475, 5085
Torrealba, G., Belokurov, V., Koposov, S. E., et al. 2018, arXiv:1811.04082
van Albada, T. S., & Baker, N. 1973, ApJ, 185, 477
VanderPlas, J. T., & Ivezić, Ž. 2015, ApJ, 812, 18
Vivas, A. K., Zinn, R., Andrews, P., et al. 2001, ApJL, 554, L33
Vivas, A. K., Zinn, R., Abad, C., et al. 2004, AJ, 127, 1158
Vivas, A. K., & Zinn, R. 2006, AJ, 132, 714
Vivas, A. K., Olsen, K., Blum, R., et al. 2016, AJ, 151, 118
Vivas, A. K., Saha, A., Olsen, K., et al. 2017, AJ, 154, 85
Walsh, S. M., Jerjen, H., & Willman, B. 2007, ApJL, 662, L83
Welch, D. L., & Stetson, P. B. 1993, AJ, 105, 1813
Wenger, M., Ochsenbein, F., Egret, D., et al. 2000, A&AS, 143, 9

Willman, B., Blanton, M. R., West, A. A., et al. 2005a, AJ, 129, 2692

Willman, B., Dalcanton, J. J., Martinez-Delgado, D., et al. 2005b, ApJL, 626, L85

Willman, B. 2010, Advances in Astronomy, 2010, 285454

York, D. G., Adelman, J., Anderson, J. E., Jr., et al. 2000, AJ, 120, 1579

Zucker, D. B., Belokurov, V., Evans, N. W., et al. 2006a, ApJL, 643, L103

Zucker, D. B., Belokurov, V., Evans, N. W., et al. 2006b, ApJL, 650, L41