Random Forest Classifier-Aided Candidate Selection of Photometrically Variable Milky Way Halo RR Lyrae in the NGVS Data Set

Jeffrey Munsell

ABSTRACT:

Under the standard model of Big Bang cosmology, the Lambda Cold Dark Matter (**Λ**-CDM) model, astronomers have built an understanding of galactic halo structure from dwarf satellites forming fragments and streams. Dwarf satellites represent remnants of the cannibalism that is believed to have formed large galaxies such as the Milky Way. Astronomers initially identified dwarf spheroidal satellites, but observed 1-2 orders of magnitude less than expected. Observations of these satellites caused a second concern as dwarf spheroidal satellites had circular velocities lower than fitting the **Λ**-CDM model. In the last decade, studying these concerns has progressed as astronomers have discovered ultra-faint dwarf satellites, which tend to exhibit low aggregate luminosities and projected densities. Because of these features, astronomers must use RR Lyrae as stellar tracers to identify UFDs. I develop new catalogs of sparse multi-band and multi-epoch data suitable for time domain astronomy from the NGVS. The survey reaches a depth of $g \approx 25.9$ mag, meaning it reaches farther RR Lyrae in the Virgo Cluster foreground than ever before. I implement random forest classifier-aided candidate selection of RR Lyrae in the foreground of the Virgo Cluster sky area from the generated catalog. The LSST will revolutionize time domain astronomy by imaging the entire sky every night, and could detect almost all RR Lyrae to >350kpc. This project serves as a precursor to future LSST studies by developing an RR Lyrae candidate selection methodology. Additionally, it aids the research decisions and design for LSST-based studies by providing a new deep field of RR Lyrae to advance background knowledge of distant stellar tracers.

I. INTRODUCTION

1.1 Background

I use a novel candidate selection method taking feature importances from a trained random forest classifier model to combine filter standard deviations into a variability metric. The main focus of the project is to use deep time series imaging to identify RR Lyrae stars in the foreground, analyze their variability, and develop a new method of time domain analysis. RR Lyrae serve as standard candles and can be used to map the structure of the Milky Way. These techniques of RR Lyrae study are likely to discover the most distant stars in the Milky Way ever identified. This deep time series study was previously impossible, but is now done using raw data from the Next Generation Virgo Cluster Survey (NGVS). According to experts studying RR Lyrae is essential as it develops fundamental knowledge of the nature of humanity's home galaxy.

Studying outer Milky Way structure requires tracer objects. The Milky Way is surrounded by a sparse, spherical distribution of stars known as the "stellar halo". It contains the oldest stars in the Galaxy, and is believed to be made up, in part, of the shredded remains of smaller galaxies that were gravitationally captured and subsequently merged into the Milky Way (Deason 2019). What is not well-understood, however, is how far out this halo extends and the mass distribution of the Milky Way beyond even 10 Kpc (Hendel et al. 2018). RR Lyrae variable stars act as standard candles -- their distance can be calculated reliably. The goal of the project is to identify RR Lyrae stars in the Galactic halo using their colors and variations in brightness, allowing their distances to be later calculated.

RR Lyrae stars are important tracers of galactic structure. They are relatively bright stars, and are common in halo features such as globular clusters and dwarf satellite galaxies. Their light curves vary periodically on a scale of 0.2 to 1 days (Dambis et al. 2013). The period-luminosity relations are quite reliable, making RR Lyrae near standard candles (Fadeyev 2019). RR Lyrae are also relatively bright stars, so they are one of the most important objects to study for mapping the outer regions of the Milky Way.

A deep astronomical survey is required to find and analyze distant RR Lyrae in the Milky Way's stellar halo. The perfect survey for my research is The Next Generation Virgo Cluster Survey (NGVS): a multi-band (*u*griz* bandpasses) imaging survey using the Canada-France-Hawaii Telescope (CFHT) MegaCam instrument. Fields have cadence varying anywhere from hours to years. Unlike other surveys, the NGVS uses a step dithering procedure. Exposures of a given field are taken in groups rather than individually in a single sequence. Exposures of each field in the group are taken in turns (Ferrarese et al. 2012). This allows for greater analysis of object brightness as a function of time. The survey focuses on the Virgo Cluster, the closest galactic cluster to the Milky Way. RR Lyrae will be identified in the foreground in the galactic stellar halo (in front of the Virgo Cluster).

"Cutouts" of NGVS images will be created where possible RR Lyrae are detected. These cutouts will exist in different bands and points in time. They will also contain the brightness and time data to be extracted for analysis of the objects of interest. This is the primary datasource for my catalo generation and variability analysis.

By analyzing the novel NGVS dataset, RR Lyrae candidates are identified by random forest classifier-aided statistics, which provides evidence to further develop knowledge of the structure of the Milky Way.

1.2 Literature Review

Earlier investigations of the Milky Way halo through stellar tracers focused on RR Lyrae in spheroidal dwarf galaxies. Ivezić et al. (2005) use SDSS single-epoch data to find ~3,500 candidate RR Lyrae. They find RR Lyrae in known spheroidal dwarfs of the Milky Way halo, and overdensities of RR Lyrae in the Sagittarius tidal stream. Their methods showed it was possible to cover out to ~70 kpc and 1/4 with SDSS color analysis in the future. They suggest light curve fitting and spectroscopy follow ups to validate their candidates and to investigate metallicity.

Classifications in the last decade led to potential structural finding, but were tentative and lacking depth. Sesar et al. (2013) found 5,000 RR Lyrae in the LINEAR survey data. They identify candidates 5-30 kpc out in ~8000 deg^2 of sky. The surveyed objects reach a depth of only $g<17$ mag. They applied group finding algorithms to detect possible halo groups. However, they emphasize that these groups are only candidates, and require further investigation. They too encourage spectroscopy.

The recent move towards using sparse multi-band, many-epoch surveys to study RR Lyrae and other variability picked up with studies such as Hernitshek et al. (2016). They analyze data from Pan-STARRS1(PS1) 3π survey. Hernitshek achieves improved depth, reaching 21.5 mag (r_{p1}). They apply a random forest classifier to identify QSO and RR Lyrae candidates. This study represents a broad scan, finding ~150,000 RR Lyrae candidates between ~10-120 kpc. However, the authors stress that the training of the model is biased as the known RR Lyrae sample comes only from previous SDSS Stripe 82 classifications, which is but a specific field within the broad area covered by Hernitshek. They conclude that their work contains higher contamination than their tests suggest due to this training bias. They have no training in the galactic plane and thus find this region to perform especially poorly in their classification model.

In the most recent research, more focused studies have come to the forefront, with Medina et al. (2018) finding the most distant known RR Lyrae with Dark Energy Camera's HiTS. This survey data covers ~120 deg^2 and finds 9 RR Lyrae at >200 kpc. They find that unlike the general halo population (r<~80kpc), where OOI RR Lyrae are more prominent, ultra-faint dwarfs hold mostly OOII RR Lyrae. They call for more tracers in similar small, deep fields to make possible general analysis of Milky Way structures rather than local investigation. They note these kinds of studies are vital in the preparation for the coming revolution with the Large Synoptic Survey Telescope, which will be able to measure almost all RR Lyrae to ~350 kpc.

Just as time domain astronomy has become a hotbed of astronomical research in recent years, study of old tracer RR Lyrae has begun to shift to sparse, multi-band surveys, with an emphasis on depth. Depth is required to detect RR Lyrae that can trace ultra-faint dwarfs, which can be used to constrain the structure and formation theory of the Milky Way. Experts hope for new deep candidates that can be later confirmed by light curve fitting.

1.3 Research Question

This study aims to answer: How can RR Lyrae candidates be identified by time domain analysis using random forest classifier-aided selection on a new catalog generated from raw NGVS data?

1.4 Hypothesis

It is hypothesized that candidate RR Lyrae will be identified through color and brightness as a function of time via variability statistics tuned with random forest classifiers. Their variability will separate them from the overwhelming majority of objects, which are non variable. The brightness of the RR Lyrae is expected to vary significantly, most notably in the *u** filter and decreasingly so in the *g*, *i*, and *z* filters.

1.5 Significance

RR Lyrae are one of the most effective stellar tracer old stars. They are quite old (≥ 10Gyr) and are reasonably bright; they vary with well-known period-luminosity relationships that make them useful as standard candles. Under the standard model of Big Bang cosmology, the Lambda Cold Dark Matter (**Λ**-CDM) model, astronomers have built an understanding of galactic halo structure from dwarf satellites forming fragments and streams. Dwarf satellites represent remnants of the cannibalism that is believed to have formed large galaxies such as the Milky Way (Deason et al. 2019). Astronomers initially identified dwarf spheroidal satellites, but observed 1-2 orders of magnitude less than expected, which is known as the "missing satellites problem". Observations of these satellites caused a second concern called the "too big to fail problem" as dwarf spheroidal satellites had circular velocities lower than fitting the **Λ**-CDM model (Kunder et al. 2018).

In the last decade, astronomers have discovered ultra-faint dwarfs, which tend to exhibit low aggregate luminosities and projected densities. Because of these features, astronomers must use RR Lyrae to identify them, and RR Lyrae may be the only way to study faint galactic satellites (Kunder et al. 2018). RR Lyrae have been found in all ultra-faint dwarfs searched for variability. Discovering new distant RR Lyrae will reveal these old structures as well as whether the Milky Way halo extends farther than ever previously measured, which contribute to formation theory and the **Λ**-CDM model's application to galaxies, especially the Milky Way (Stringer et al. 2019; Belokurov et al. 2018). These old structures allow astronomers to gaze into the past of the Milky Way as the remnants of galactic formation that still exist today in the dark matter-dominated outer halo.

II. MATERIALS & METHODS

Astronomy experts have never been able to use the NGVS for time domain analysis, so my methodology works from raw data and creates new catalogs for time domain astronomy, which I use to find undiscovered RR Lyrae candidates. RR Lyrae are standard candles that can map the shape of our galaxy, and reveal its outer limits that have never been measured. My project separates rare candidate variable objects from all other objects around the Virgo Cluster region of the sky. These objects lie in the foreground of the images, in the outer halo of the Milky Way.

The Next Generation Virgo Cluster Survey (NGVS) was designed primarily to study distant galaxies. Its data is publicly accessible, but it was not intended for time domain astronomy. NGVS composited all the data for each object to generate a catalog of magnitudes in the *u**, *g*, *r*, *i*, and *z* bands. Experts used these data, but were not able to use the survey for time domain astronomy, because they were lacking the necessary methodology. This is what I build in this project. I take raw NGVS image data to build a framework for time domain astronomy in NGVS. The NGVS dithering process took exposures in a region, with consecutive exposures taken from different locations in the sky. Each group of measurements was repeated many times over the several years of the survey. This layout of raw data leads to an uneven cadence in the measurements of each object, with epochs anywhere from an hour apart to a year apart. This spread of the raw data allows for time domain analysis. Experts had never previously realized this advantage of the NGVS, and as a result have never used the data for a project like mine. Adapting the NGVS for time domain astronomy makes it the deepest dataset around the Virgo Cluster. Focused studies rather than broad surveys are needed for effective variable identification (Szabados 2019). Sesar's previously identified objects, though largely incomplete, provide a reliable way to examine expected RR Lyrae behavior in the data (Sesar et al. 2013).

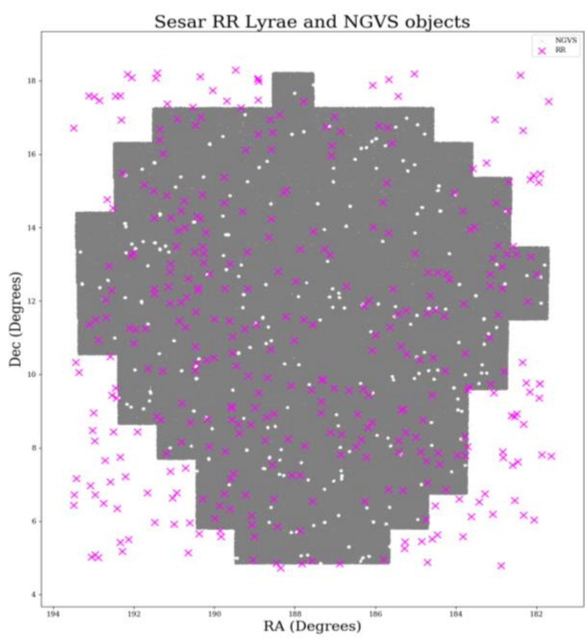

Figure 1. Sky positions of Sesar RR Lyrae (pink crosses) around the NGVS footprint (in gray). Note smalls holes at contamination sources.

Known objects in the region are required to tune the selection parameters of new candidate RR Lyrae. Sesar's high confidence RRab (confidence>0.8) and RRc (confidence>0.55) objects are matched by position to the NGVS catalogue. There are 85 such matched objects (See Figure 1). These previously identified objects, which represent a small percentage of the total RR Lyrae in the region, will produce a benchmark for RR Lyrae used to calibrate new selection.

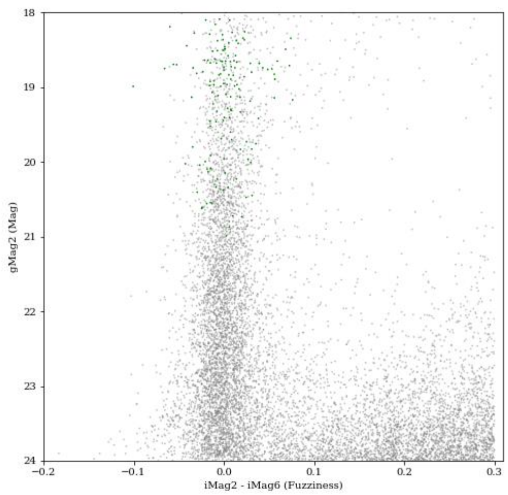

Figure 2. Fuzziness vs. g-band magnitude of Sesar RR Lyrae (green), and 100,000 random NGVS sources.

To bring out RR Lyraes' variable behavior, the precision of the data must be optimized. The accuracy of the model is very sensitive to the precision of the data. The raw data is not suitable for time domain analysis as variability due to atmospheric and instrumental error will obscure the trends of the objects. These sources of error invalidate the model, and experts would not have been able to use the data in this state for time domain analysis of variability if they tried. The fundamental idea of my methodology that makes the model valid is using non variable reference stars. These reference stars must first be selected from outside the RR Lyrae region of C-C space, so that they do not overlap with the RR Lyrae. The reference objects chosen are also bright (g < 22 mag) and selected within the fuzziness range |iMag6-iMag2| < 0.1 (see Figure 2), so they are point-like. This allows for more reliable data of the reference objects, which will be used for correction of exposures.

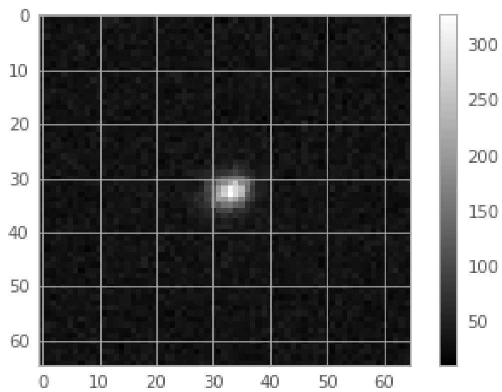

Figure 3. Example cutout of an NGVS exposure.

Data extraction is performed to gather NGVS images containing objects to be analyzed. Script queries of Stephen Gwyn's cutout tool collect images and epoch descriptors for all exposures containing the R.A. and Dec. of objects. This algorithm was developed in the last few years to provide access to raw NGVS data around specific positions. Access to this new data format facilitates my research by searching the NGVS database system for all exposures containing a given position. Then the algorithm generates new cutouts that are 64x64 pixels centered on the objects' catalogued positions (See Figure 3). The tool generates an HTML table and displays a webpage with photo header data and links to cutouts. This is done for all objects of interest as well as reference objects and the generated web pages are scraped for the images and header data, providing the raw data to be processed.

Standard methodologies are applied to measure the apparent brightness and error of the objects in the raw data. Each image is screened to have a valid object and complete image, including a check of roundness threshold. The valid images are recentered on the object, so that the following aperture photometry can be performed at the correct location. Fixed aperture photometry is performed, which measures a fixed aperture around the target object and make local background corrections by sampling an annulus around the aperture. Astronomical apparent magnitude is m = $-2.5log(\frac{F_x}{F_{x,0}})$, where F_x is the observed flux density in filter x and $F_{x,o}$ is the flux zero point of the exposure, as reported by the NGVS. The error for each measurement is calculated as $m_{err} = 1.0857/(S/N)$, where S/N is the signal to noise ratio of the measurement. This custom data will be used for analysis, but first it must be adjusted with an analysis of reference objects to validate the model.

I next use reference objects to greatly reduce atmospheric and instrumental variations in the data, so that the true variability of the RR Lyrae will differentiate them from other objects. I construct a new catalog incorporating this adjustment. The reference stars are grouped by exposure ID for each band. In each exposure and band group, the objects' deviations from their respective catalogued magnitudes are collected. The median of these deviations in each group is compiled to form a correction catalogue for all NGVS exposures by band and exposure ID. This correction catalog is applied to all objects, both target and reference, to generate new catalogs of the objects' measurements. Without this catalog, experts would not be able to study variability in NGVS. This new catalog of precise measurements can now be analyzed for candidate selection.

Candidate selection begins with cuts from NGVS catalogued magnitudes. The known Sesar RR Lyrae are used to evaluate selection criteria, and are plotted against a large random sample of about 100,000 NGVS objects. Of the 85 known RR Lyrae, 80 have at least three images per band, the cutoff for inclusion in analysis. I examine color-color space to find the regions containing high probability density of RR Lyrae, using the Sesar RR Lyrae data. The chosen fuzziness range selects for point-like sources, |iMag6-iMag2| < 0.1, as fitting the RR Lyrae. Examining color-color diagrams, the RR Lyrae are around the turn-off, as expected. The C-C selections from *u-g* vs *g-i* and *g-i* vs *i-z* are aided through gaussian color-color maps. Gaussians for each color axis are computed for each Sesar RR Lyrae on a grid in the area containing the RR Lyrae. The grids of all RR Lyrae are summed and the two axis are multiplied to produce a density plot. The standard deviations (sigma) of each band used to generate the gaussians are sigma-clipped for $\sigma > 3$ and weighted by $\frac{1}{(error+0.015)^2}$, where error is the photometric error of each epoch. The small constant (0.015) is added to prevent tiny errors from causing massive weights. This process shows the percentage of total density falling within selection boxes drawn in C-C space. This allows me to choose selections in C-C space. After this first group of selections, I move on to selections using the new catalog I created.

Band sigmas for each object are combined by weighted mean to create a single statistic. A machine learning approach is applied to accurately determine the weight of each variable. The weights of bands are determined by random forest classifier importance, but the classification itself is not used. The random forest classifier itself is a standard algorithm. However, I take the unique approach of training it for the purpose of determining feature importances, so that the sigma statistic will resolve RR Lyrae from all other objects. There are 85 RR Lyrae and tens of thousands of reference objects. The huge discrepancy in class size would cause the model to be trained to fit only the larger class. Therefore, undersampling is used, and I select ten times the number of RR Lyrae data points for Reference data points. The reference star and RR Lyrae band sigmas are shuffle split into train and test data. The random forest classifier is trained to determine from the band sigmas of an object whether that object is an RR Lyrae or non-RR Lyrae. The feature importances are then checked, yielding weights for the average sigma statistic.

I plot this weighted mean of sigmas from each band against the *g*-band magnitude of RR Lyrae and reference objects. Binning the sigma statistic by g-band magnitude, the central tendency (mean, median, and standard deviation) of each bin are computed for reference star data. The median is a better measure of central tendency due to large outliers skewing the mean, so one standard deviation above the median is used to make the candidate selection cutoff.

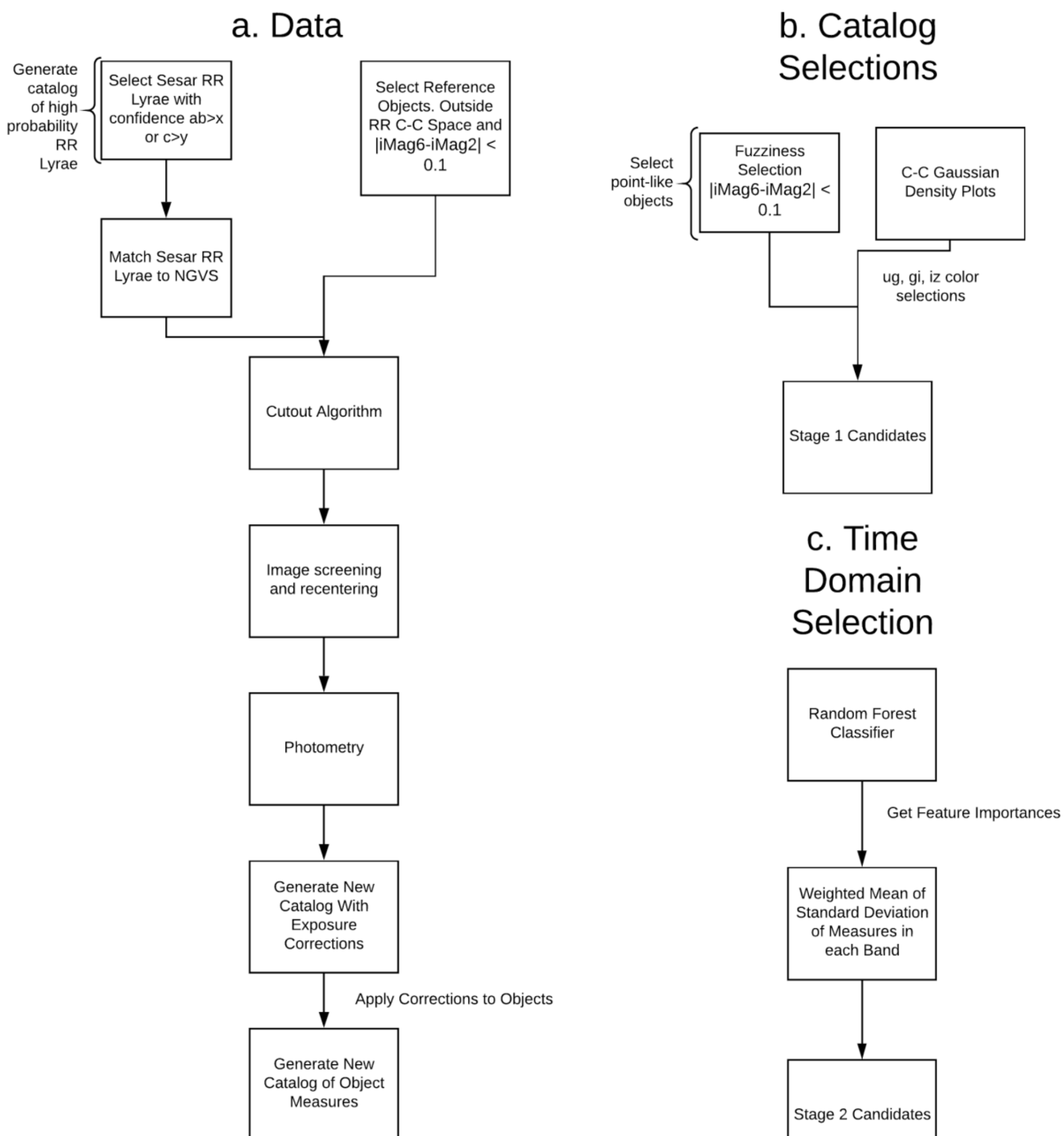

Figure 4a. Catalog generation process.
Figure 4b. Existing catalog-based initial selection process.
Figure 4c. Time domain selection of candidates.

III. RESULTS

2.1 Plots

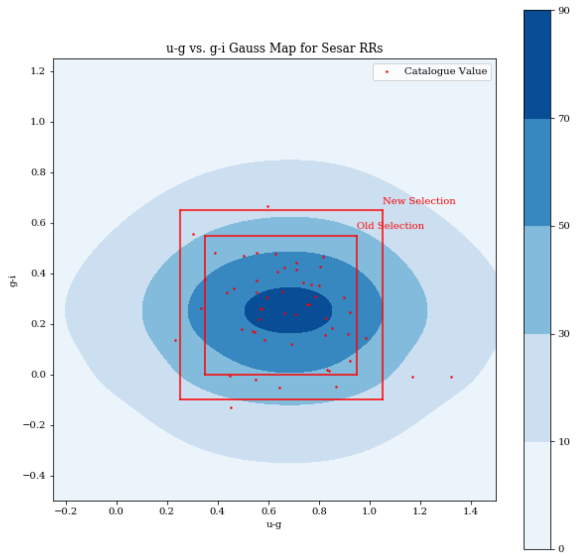

Figure 5. *u-g* color vs. *g-i* color gaussian map showing Sesar RR Lyrae density in C-C Space.

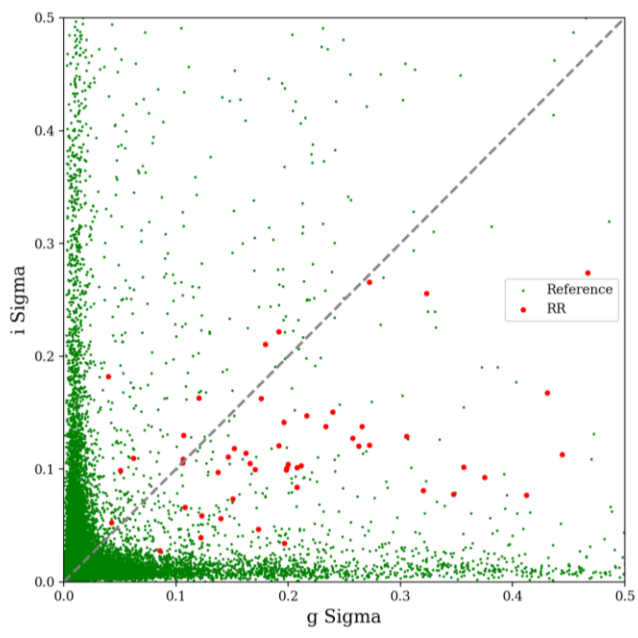

Figure 6. *i*-band sigma vs. *g*-band sigma of Sesar RR Lyrae (red) and reference objects (green).

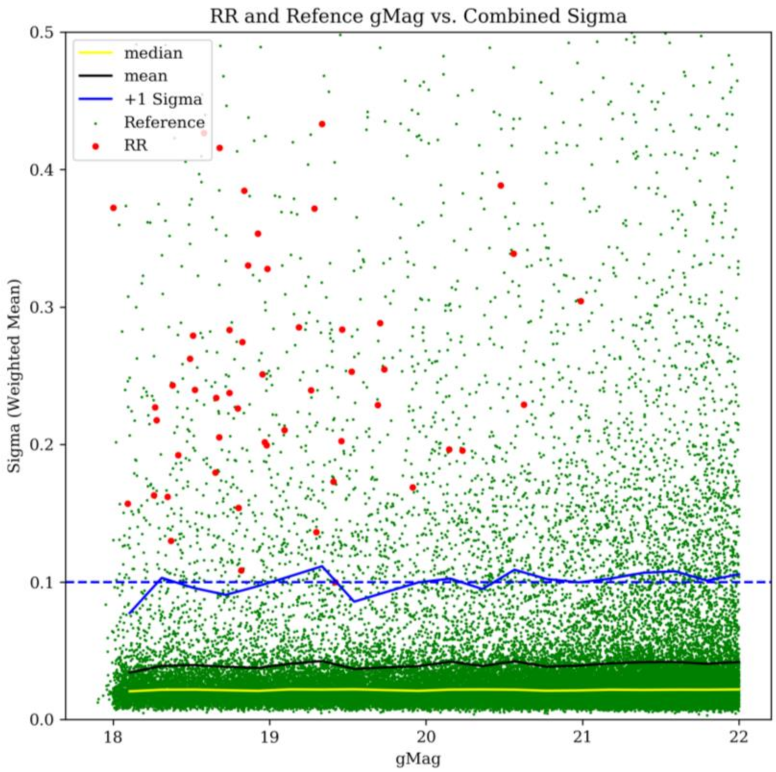

Figure 7. Sigma statistic of band sigmas vs. g-band magnitude of Sesar RR Lyrae and reference objects.

2.2 Analysis

Gaussian density plots show the C-C distribution of known Sesar RR Lyrae (see Figure 5). Initial selections are made at 0.25<*u-g*<1.05, -0.1<*g-i*<0.65, and -0.3<*i-z*<0.25. In the case of *g-i* vs. *u-g*, 47.90% of the 2-D probability distribution falls in this area, and 90.74% of Sesar RR Lyrae. This shows excluding these areas of color-color space from the reference star sample decreases the likelihood of reference stars being RR Lyrae on top of the innate rareness of RR Lyrae. The gaussians derived from color sigmas, in turn derived from band sigmas, which are high due to RR Lyrae variability show said variability in the large discrepancy between the ~91% of RR Lyrae by quantity and ~48% of RR Lyrae PDF falling in the selection.

The random forest classifier trained to differentiate RR Lyrae from non-RR Lyrae do so with 98.07% accuracy. I tested the model with a number of estimators (trees) ranging from five to thirty, and found effectively no improvement after n=10, so the final model uses ten estimators. This evaluation is determined via k-fold cross examination with k=3. However, this also determines class accuracies of 98.5% and 77.5% for reference objects and known RR Lyrae, respectively. This shows that the machined learned classification is not sufficient to resolve RR

Lyrae itself, as expected. The use of the random forest is instead, as earlier explained, to determine feature importances. From most important to least, I expect *u**, *g*, *i*, then *z* to predict RR Lyrae variability, as RR Lyrae vary more towards the blue side of the light spectrum. Consistently, the random forest method determines weights of ~0.65, 0.14, 0.12, and 0.09, for the *u**, *g*, *i*, and *z* bands, respectively.

This relationship between the bands is also demonstrated through plots comparing sigmas of different bands. Figure 6, illustrating *i* sigma vs. *g* sigma, displays a disproportionate concentration of RR Lyrae below a slope of one, demonstrating again the greater variability of RR Lyrae in *g* than in *i*. The same pattern is evident for the other bands pairs when plotted sigma vs. sigma.

When the weights determined by the random forest classifier are used to take a weighted mean of the sigmas of each band for all objects, candidates can be selected. The cutoff based on one standard deviation above the median of the weighted sigma binned by cataloged *g* magnitude is 0.1 mag (see Figure 7). This cutoff results in an RR Lyrae completeness of 98.15% and a reference object completeness of 6.79%. This means the candidate selection reduces computation for RR Lyrae identification by 93.17%, while only throwing out 1.85% of RR Lyrae.

IV. DISCUSSION

I develop a catalog of NGVS data suitable for time domain analysis. After assembling and organizing raw NGVS imaging data as cutouts, I screen and recenter the images. I then perform photometry. By comparing this raw photometric catalog to the NGVS composite catalog, I determine epoch corrections, which I apply to all measurements in the raw catalog.

The time domain analysis portion of my methodology provides computationally efficient candidate selection that is broadly applicable. Steps in the catalog creation from NGVS were the only tasks that were not feasible to perform on a personal computer. Cutout queries, download, image checks and corrections, and photometry were performed on UCSC servers. Now that the new catalog is complete, I or other researchers can apply my time domain analysis procedure to the NGVS at little computational expense.

My research answers the call for deeper RR Lyrae data and new deep fields, and the development of methodologies for future LSST work. I develop new catalogs of multi-band, sparse data suitable for time domain astronomy from the NGVS. The NGVS was not designed for time domain astronomy, and experts have therefore not used it to study RR Lyrae variability with multi-epoch data (Ferrarese et al. 2012, see "The NGVS: Science Objectives"). I develop tools to process raw NGVS exposure data into a new catalog and to generate variability statistics for time domain analysis. The NGVS is the most modern survey generate of the virgo cluster sky region and reaches g ≈ 25.9 mag, containing new potential accessible through my new catalog and methodology (Ferrarese et al. 2012). These tools are tuned to NGVS and RR Lyrae, but apply to other surveys such as LSST and other objects such as QSOs.

This project selects candidate RR Lyrae, setting up future validation. Validation requires mass computing power and is, based on expert work, not feasible on personal computer-scale computing. In order to test the validity of my methodology, I asked an independent expert to apply it to his dataset, and it immediately led to discovering five new RR Lyrae. Light curve fitting was used to confirm the candidate objects. I suggest light curve fitting be performed on all candidates for validation as a follow up to this project, and the expert mentioned is currently working on that project.

My variability selection methods do not rely on a unique property of RR Lyrae. These methods can therefore be applied to other variable objects such as QSOs. I recommend QSOs especially as Hernitshek et al. (2016) have proved feasible. This would require adapting color-color selections from the 2D-gaussian plots for the *u-g* vs. *g-i*, and *g-i* vs. *i-z*, retuning the random forest classifier, and adjusting the sigma statistics.

This study serves as a precursor to the Large Synoptic Survey Telescope (LSST). The focus of astrophysical research has shifted recently from the standard universe to the study of the universe as a function of time. The LSST will image the entire night sky every few days starting in 2022. This will push forward time domain astronomy in revolutionary fashion. But in order for this revolution to occur, tools and methodologies must be developed in preparation. One of the specific goals of LSST is to study variable stars such as RR Lyrae, which my work directly applies to. However, my project is not only applicable to RR Lyrae. The methodology developed in this study is tuned to RR Lyrae, but can be applied to identify other variable objects such as QSOs in the future, especially with the much improved data of studies such as LSST. Thus, this study creates a basis for future research utilizing the LSST data set and enables future discoveries in time domain astronomy.

References

Belokurov, V., Deason, A. J., Koposove, S. E., et al. 2018, Monthly Notices of the Royal Astronomy Society. doi:10.1093/mnras/sty61.5

Dambis, A. K., Berdnikov, L. N., Kniazev, A. Y., et al. 2013, Monthly Notices of the Royal Astronomical Society. doi:10.1093/mnras/stt1514

Deason, A. J., Belokurov, V., Sandars, J. L. 2019, Monthly Notices of the Royal Astronomical Society. doi:10.1093/mnras/stz2793

Fadeyev, Y. A., 2019, Astronomy letters. doi:10.1134/S1063773719060021

Ferrarese, L., Côté, P., Cuillandre, J., et al. 2012, The Astrophysical Journal Supplement. doi:10.1088/0067-0049/200/1/4

Hendel, D., Scrowcroft, V., Johnston, K. V., et al. 2018, Monthly Notices of the Royal Astronomy Society. doi:10.1093/mnras/sty1455

Hernitschek, N., Schlafly, E. F., Sesar, B., et al. 2016, The Astrophysical Journal. doi:10.1088/0004-6256/146/2/21

Ivezić, Ž., Vivas, A. K., Lupton, R. H., et al. 2005, The Astronomical Journal. doi:10.1086/427392

Kunder, A., Valenti, E., Dall'Ora, M., et al. 2018, Space Science Reviews. doi:10.1007/s11214-018-0519-0

Medina, G. E., Muñoz, R. R., Vivas, A. K., et al. 2018, The Astrophysical Journal. doi:10.3847/1538-4357/aaad02

Sesar, B., Ivezić, Ž., Scott, S. J., et al. 2013, The Astronomical Journal. doi:10.1088/0004-6256/146/2/21

Sesar, B., Hernitscheck, N., Sandra, M., et al. 2017, The Astronomical Journal. doi:10.3847/1538-3881/aa661b

Stringer, K. M., Long, J. P., Macri, L. M., et al. 2019, The Astronomical Journal. doi:10.3847/1538-3881/ab1f46

Szabados, L. 2019, Contributions of the Astronomical Observatory Skalnaté. arxiv:1902.06620