Faint Substructure Candidate Identification with Galactic Standard Candles in Wide-Field Survey Data

Brian T. Cook, UNC-Chapel Hill

THE UNIVERSITY of NORTH CAROLINA at CHAPEL HILL

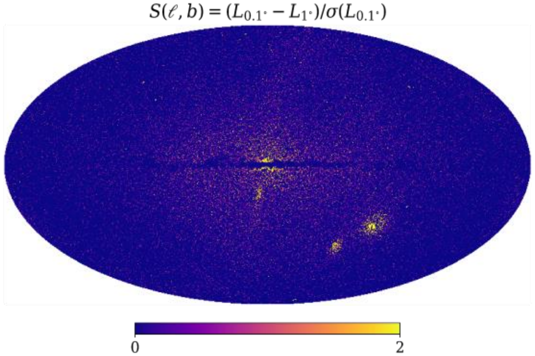
Abstract

The most recent data release by the Gaia consortium included an updated database of RR Lyrae variables (RRLs) in the Galactic environment, a noted tracer of substructure that can be used to better understand the hierarchical process by which the Milky Way formed. After a round of pre-processing to ensure that an accurate estimate of the phase space occupied by these variables could be rendered, **we employed significance mapping and hierarchical, density-based spatial clustering to detect compact satellites**. Newly discovered candidate satellites were then considered as potential progenitors of **RRL streams, which were identified with an implementation of the STREAMFINDER algorithm**. The aggregate of these routines yielded several open/globular cluster (OC/GC) and ultrafaint dwarf galaxy (UFD) candidates and their associated streams (if present). A non-zero number of tip of the red giant branch (TRGB) stars, another standard candle expected to be found in a stellar population containing RR Lyrae variables, were found in the nearby vicinity of each of the candidates. **If the open cluster candidates are confirmed with follow-up observations capable of detecting candidates' main sequence populations, they would constitute the farthest detected open clusters on record.**

Relevant Context

RR Lyrae (RRL) variable stars are reliable standard candles in the Milky Way (MW) and, once wide-field survey light curves are classified, there are comprehensive, publicly-available MW RRL databases. If such databases include astrometric and radial velocity information, a detailed picture of the MW phase space occupied by these variables emerges. RRLs are Population II stars often associated with the MW building blocks like dwarf galaxies and globular clusters. The goal of this study is to build upon previous work on RRL clustering and develop a substructure identification pipeline that could, in theory, be applied to any RRL database. **The plot below shows the significance (e.g., S=2 means the RRLs in that pixel are two sigma above the background) for the apertures we used to find substructure candidates resembling OC/GCs.**

$$S(\ell, b) = (L_{0.1^\circ} - L_{1^\circ})/\sigma(L_{0.1^\circ})$$

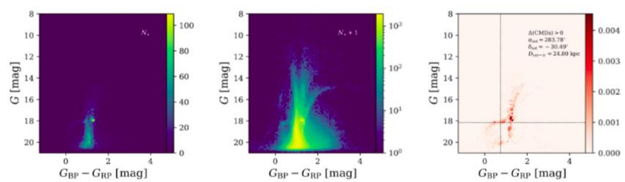


Methods

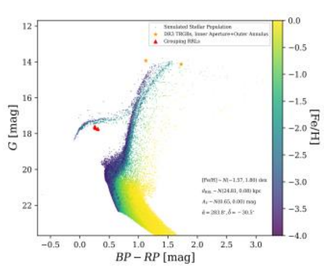
In order to find RRL groupings consistent with MW substructures, we apply the HDBSCAN algorithm, a hierarchical-clustering routine that has successfully been applied in this context. We first define two sets of inner/outer apertures: one for OC/GCs, and another for UFDs. This provides high-significance RRL subsets that can then be clustered; we cluster RRLs together that belong to the same HEALPix pixel, which reduces computational expense. Our HDBSCAN distance metric uses physical location, proper motion magnitudes, and proper motion angles (**modulated by the scaling parameters listed below**) between RRLs.

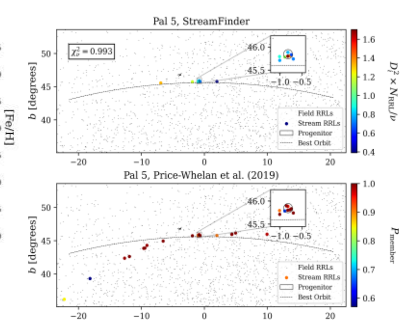
Parameter	OCs and GCs	UFDs
$\sigma_{\rm in}$ [deg]	0.1	0.3
$\sigma_{\rm out}$ [deg]	1.0	3.0
$r_{\rm scale}$ [pc]	10	120
$\mu_{\rm scale}$ [mas/yr]	0.2	0.15
$\theta_{\rm scale}$ [deg]	1.0	1.0
$\min(n_{\rm RRL})$	3	2

After some additional filtering steps, we have a set of compelling candidates that require follow-up analysis. A candidate needs to have certain proper motion, (PM) color-magnitude diagram (CMD), and integrals-of-motion (IOM) properties before we consider it a legitimate candidate. We construct a business card of sorts for each candidate, where we pull down all Gaia DR3 data in the relevant sky plane region such that we can determine the **PM/CMD overdensities of each candidates' inner aperture when compared to the background** (see the provided example for a grouping consistent with NGC 6715).



Additionally, we use dustmaps extinction data and RRL metallicity, if available, to construct a **simulated stellar population and plot the subsequent (dereddened) CMD** (see below for an example CMD using the same NGC 6715 grouping). We search for TRGB stars in the inner/outer apertures at distances consistent with the grouping, as these standard candles should also be present in a population containing RRLs. Lastly, we use an implementation of the STREAMFINDER algorithm to determine whether the RRL grouping has neighbors in IOM space consistent with a MWPotential2014 orbit, **which we tested using the Pal 5 stream.**



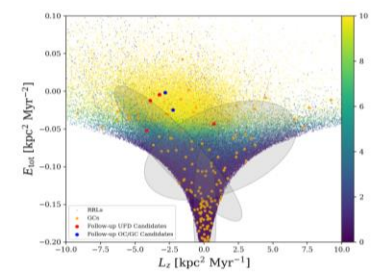


Results, follow-up with LSST

We cross-reference the outputs of our substructure identification pipeline with the SIMBAD database of known associations and clusters, as well as a bespoke catalog of OC/GC/UFDs that have been used in similar studies. We consider a grouping a true positive if a.) there is a SIMBAD object in the search inner aperture or b.) a known satellite is found sufficiently close to the candidate on the sky plane such that it would overlap with the recommended inner aperture for observations. Groupings where a known object was in the outer aperture, typically used to sample the background, were not considered when calculating the precision. Conservative estimates, before we declare any candidate as sufficiently compelling for follow-up observations, is ~0.9 for OC/GC candidates and ~0.8 for UFDs. (**see the table below**).

Total # of RR Lyrae groupings	OCs and GCs	UFDs
Initial HDBSCAN result	2067	5101
Retained candidates	90	393
SIMBAD object in Table 1 inner aperture	38	209
Known satellite < $\theta_{\rm size}$	60	82
Known satellite in inner aperture ($1.5\theta_{\rm size}$)	72	120
Known satellite in outer annulus ($10\theta_{\rm size}$)	81	192
Known true positives	82	298
Follow-up needed using §2.3-2.5 methods	3	41

The outputs of this pipeline, at present, are two OC/GC and five UFD candidates we consider sufficiently compelling to suggest follow-up observations. The main sequence populations of our candidates are below the Gaia brightness limit and each candidate is at a declination less than 10 degrees; these make them ideally suited for LSST. Additionally, this substructure identification pipeline could be applied to a future LSST RRL database and help fill in the picture of hierarchical galaxy formation further. (**The figure below show the IOM distribution of MW RRLs, known GCs, and the compelling candidates. The best-fit Gaussian mixtures of 20000 randomly-sampled RRLs may help group substructures with their associated progenitor events.**)



Acknowledgements and Selected References

This poster presentation was made possible thanks to support from NSF Grant AST2009916 to The University of North Carolina at Chapel Hill and Carnegie Mellon University. For more context related to this substructure finding pipeline, see the following publications for more details:

- Gaia Collaboration et al. (2021), A&A, 649, A7
- Cook et al. (2022), MNRAS, 513, 2509
- Iorio and Belokurov (2019), MNRAS, 482, 3868
- Koposov et al. (2008), ApJ, 686, 279
- Koposov et al. (2015), ApJ, 805, 130
- Malhan and Ibata (2018), MNRAS 477, 4063