Exploring the Northern Over-density Region of the Small Magellanic Cloud with MAGIC and APOGEE

Lucas Rayder

Advisor: Prof. Steven R. Majewski

June 4, 2025

Department of Astronomy
University of Virginia

This thesis is submitted in partial completion of the requirements of the BS Astronomy-Physics Major.

Exploring the Northern Over-density Region of the Small Magellanic Cloud with MAGIC Photometry and APOGEE Spectroscopy

Lucas R. Rayder [1] Anirudh Chiti [2] Steven R. Majewski [1] Andres Almeida [1] David L. Nidever [3] and the DELVE Collaboration[4]

[1] University of Virginia
[2] University of Chicago
[3] Montana State University
[4] NSF NOIRLab

ABSTRACT

The Small Magellanic Cloud Northern Over-Density (SMCNOD), a recently discovered feature extending beyond the SMC periphery, has been explained as potentially originating as an extension of the SMC main body, potentially as a result of tidal stripping from SMC-LMC encounters, or, alternatively, the SMCNOD may represent a smaller dwarf galaxy, or the remains of one, being tidally disrupted by the SMC (A. Pieres et al. 2017). Since its discovery by A. Pieres et al. (2017), the SMCNOD has been explored in various ways in attempts to discriminate between these different origin scenarios. In particular, chemistry can be used to establish whether the SMCNOD stars share the same composition as stars in the SMC main body, or whether the SMCNOD is lower metallicity, as might be expected for a dwarf galaxy. Similarly, the kinematics of these stars may give further insights into their connection (or not) with the SMC. Here we apply to these purposes the metallicity information obtained for >800 SMCNOD stars to $g \lesssim 21.5$ using photometry in the narrow-band CaHK filter as part of early data obtained from the DECam MAGIC Survey. We find the metallicity distribution function (MDF) for the SMCNOD to be both extremely broad (ranging from [Fe/H] $<$ —2.5 to $<$ —0.5), but also broadly peaked, with a plateau in the MDF from [Fe/H] $<$ —2.0 to $<$ —1.5. We also use the relatively small sample of SMCNOD stars sampled by the APOGEE survey to help ascertain the kinematics of the region. Together, the chemical and kinematical evidence from MAGIC and APOGEE appears to lend credence to tidal stripping as a viable explanation for the origin of the SMCNOD.

1. INTRODUCTION

The prevailing theory for how galaxies grow, initially proposed on galactic scales by L. Searle & R. Zinn (1978) and further developed by S. D. M. White & M. J. Rees (1978), is by hierarchical evolution, which postulates that the merging of smaller objects to form larger objects happens at all scales. Much like how planets form from the coalescence of smaller chunks, and those chunks from even smaller particles, galaxies are believed to form from smaller galaxies, which themselves originate from yet more fundamental components. What makes discovering these dwarf constituents challenging is that the best examples are often far too faint and/or distant to be studied effectively. Because of these limitations, one of the best opportunities to test this theory lies in observing the halo of our own Milky Way and the various dwarf satellites in orbit within and around it. Many studies have been done in this region, lending more and more validity to the hierarchical formation theory (A. J. Deason & V. Belokurov 2024; V. Belokurov et al. 2006), and the continued release of data from ESA's GAIA program (Gaia Collaboration et al. 2016, 2023a) and other large surveys will only further the work being done.

The Magellanic Clouds, the two largest Milky Way satellites — large enough to be easily observed with the naked eye in the Southern Hemisphere — have long been looked to as valuable assets to understanding a variety of processes involving the evolution of galaxies, and, dwarf satellite galaxies in particular. For example, these small galaxy systems are notable for how recently, on a cosmic scale, they have fallen into the Milky Way's gravitational well. Current models, based on measurements of their proper motions (N. Kallivayalil et al. 2013), suggest that these star systems are on their first infall (G. Besla et al. 2007), which means that these irregular dwarf galaxies formed in a region far from the Milky Way; thus they provide an opportunity to study in detail the structure, stellar populations, and star formation history of galaxies that spent most of their lives in a presumably isolated, low density environment. Moreover, because of their proximity and relative lack of disruption, we are able to identify a vast array of substructures within, around, and between the Clouds. Such substructures include the Magellanic Bridge, a stream of HI gas connecting the two bodies (J. V. Hindman et al. 1963), and the Magellanic stream, a trail of HI gas left behind by the clouds that spans $\sim 200°$ on the sky (D. S. Mathewson et al. 1974; D. L. Nidever et al. 2010).

In recent years, advancements in technology have enabled more comprehensive systematic surveys, leading to the discovery of subtler, fainter stellar substructures. Among these is an object whose discovery is documented in A. Pieres et al. (2017) and is defined by a stellar overdensity in the northern periphery of the Small Magellanic Cloud (SMC). Dubbed the Small Magellanic Cloud Northern Over-Density (SMCNOD), the region is centered roughly around α = 12°, δ = -64.8°, but in total covers $6° \leq \alpha \leq 15°, -66° \leq \delta \leq -63°$ (see Figure 1).

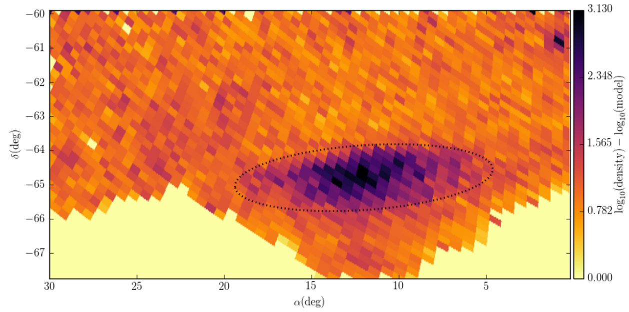

Figure 1. The SMCNOD as discovered in A. Pieres et al. (2017), located within $6° \leq \alpha \leq 15°, -66° \leq \delta \leq -63°$.

Other than the stellar overdensity itself, no other method has been found to distinguish this population from the rest of the SMC, prompting further interest and questions about the origin of this feature.

If we can identify characteristics of the overdensity that either distinguish it from those of the main body of the SMC or mark it as similar, it would aid in discriminating between a variety of origin scenarios postulated for its origin. For example, if the SMCNOD were constituted by stellar populations differing greatly in age and metallicity from those of the SMC, it would be strong evidence of hierarchical merging, offering a glimpse at the smaller subcomponents that build up dwarf galaxies — an example of "satellites of satellites" sought in empirical studies of hierarchical galaxy formation. On the other hand, if no such distinction in populations is found, the SMCNOD might instead be a tidal feature composed of material torn from the SMC by the LMC's gravitational influence. It has been found that HI gas is being stripped off the SMC by the LMC (as evidenced by the Magellanic Bridge), but there is limited evidence of stripping of stellar material at this scale.

In this study we make use of spectroscopic data from the Apache Point Observatory Galactic Evolution Experiment (APOGEE) survey (S. R. Majewski et al. 2017) as well as photometric data from DECam's Mapping the Ancient Galaxy in CaHK (MAGIC) photometric survey (Chiti et al., in prep) to understand better the SMCNOD and reveal clues to its origin. APOGEE provides reliable radial velocity measurements, which alongside proper motions from GAIA DR3 (Gaia Collaboration et al. 2023b) allow for proper kinematical analysis. If the SMCNOD region were found to have a distinct motion, it would be evidence that the SMCNOD did not develop inside the SMC main body, since for the latter we would expect a general correspondence to the SMC's general bulk motion. Meanwhile, MAGIC provides metallicities for a large number of stars in this region, allowing us enough data for a statistically robust chemical analysis. Studying the metallicity ([Fe/H]) of stars across the region gives general insights into the ages of the various relevant stellar populations, as well as the environment in which the SMCNOD formed. Older stars are generally more metal poor than younger stars as a result of forming at a time when the interstellar medium was not as enriched with heavier elements. If the SMCNOD were found to be distinctly metal poor in comparison to the larger SMC, this would indicate that the former could be an independent component that formed at a different time or in a different environment than did the SMC.

In this paper we analyze the above combined kinematical and chemical data to create a more complete picture of the SMCNOD. With them, we test four different theories as to the origin of the SMCNOD, i.e., that it is: an extension of the SMC's main body, material pulled out of the inner SMC through tidal interactions with the LMC and/or a tidal dwarf galaxy, substructure in the SMC halo and/or debris accreted in the hierarchical formation process, or a dwarf satellite absorbed by the SMC. These origin scenarios and the kinematical measurements, metallicities, and ages expected to validate them are presented in Table 1.

To narrow down a most likely origin scenario, we characterize the region's proper motions and radial velocities, analyze metallicity values and probe the metallicity gradient, and fit isochrones to the populations of the SMCNOD region. The data sets used and the criteria employed to select members are discussed further in Section 2 and an analysis of those data is presented in Section 3. We discuss our findings in Section 4 and present our conclusions as well as suggesting directions for future research in Section 5

2. DATA AND SELECTION

2.1. *Survey Data*

The data for this project come from the APOGEE, GAIA, and MAGIC stellar surveys (S. R. Majewski et al. (2017), Chiti et al., in prep, Gaia Collaboration et al. (2023b)). APOGEE is a spectroscopic survey in Sloan Digital Sky Survey III and IV (SDSS-III/IV) (D. J. Eisenstein et al. 2011; M. R. Blanton et al. 2017) which used the 2.5-m Sloan Foundation Telescope and 1-m NMSU telescope at Apache Point Observatory (APO) in New Mexico, as well as the 2.5-m Irénée du Pont Telescope at Las Campanas Observatory (LCO) in Atacama de Chile, to obtain high resolution, high signal-to-noise (S/N), infrared spectra for hundreds of thousands of target stars across the Milky Way. Some goals of the initial project include putting constraints on dynamical models for the MW bulge, disk, and halo, deriving data for dust-obscured stars on the same level as easily accessible stars, and exploring the early galaxy by probing stellar parameters for some of the oldest stars. We use APOGEE results from SDSS Data Release 17 (Abdurro'uf et al. 2022), the final data release from the APOGEE-1 and APOGEE-2 programs, containing 657,000 stars observed over 2,660,000 visits, with 14,000 targets in the LMC and SMC. Data Release 17 contains all previous APOGEE data recorded between September 2011 and August 2018, as well new observations taken between July 2018 and November 2020 at APO, and August 2018 and January 2021 at LCO. For further information regarding the goals of APOGEE and the data contained within DR17, we refer to S. R. Majewski et al. (2017) & Abdurro'uf et al. (2022).

MAGIC (Chiti et al., in prep) is a photometric survey that takes advantage of the Dark Energy Camera (DECam) at the 4-m Blanco Telescope in Chile, and aims to cover ~5,000 deg^2 of the southern sky, obtaining

Origin Scenario	Kinematics	Metallicity	Age	Likelihood
Extension of Main Body	Similar to SMC	Follow Metallicity Gradient	Young to Intermediate	?
Material pulled from SMC by Tidal Interactions with the LMC (and/or Tidal Dwarf Galaxy)	Similar to SMC	Follow Metallicity Gradient	Young to Intermediate	?
Substructure in the SMC Halo or Accreted Debris	Similar to SMC	Follow Metallicity Gradient	Intermediate to Old	?
Dwarf Satellite	Similar OR Distinct to SMC	Differs From Metallicity Gradient	Intermediate to Old	?

Table 1. The four proposed origin scenarios for the SMCNOD and the measurements that would validate them.

photometry in the narrow-band CaHK filter (centered at ~ 3950 Angstroms) for stars in the Milky Way, the Galactic halo, and a number of dwarf satellites. The filter is particularly valuable in how it targets the Ca II H & K absorption lines, whose depths are indicative of stellar metallicity and are strongly expressed for G and early K giant stars. The survey intends to derive photometric metallicities down to very low metallicities, including $[Fe/H] \lesssim -3.0$. Deriving metallicities using photometry is less precise than via spectroscopy, but takes much less observing time due to the high resolution needed to observe iron lines, and is thus very useful in surveying large stellar samples. The survey aims to provide the necessary data for extensive Galactic archaeology studies probing the first stars, galaxy evolution, and the origins of the Milky Way and Magellanic Clouds. Observations began in Fall 2023 and are still ongoing. The latest data we access were taken in Fall 2024.

Before we perform cuts on the data in Section 2.2, we present the full coverage maps for both MAGIC and APOGEE in Figure 2 as reference. All APOGEE fields were used since they exclusively cover the SMC, while the MAGIC data have been spatially selected from $0° \leq \alpha \leq 40°$ and $-74° \leq \alpha \leq -62°$ to select all coverage of the SMC while avoiding unnecessary points beyond the periphery, which only represent the Milky Way halo. We note that the MAGIC coverage is continuous, while APOGEE is comprised of several circular fields.

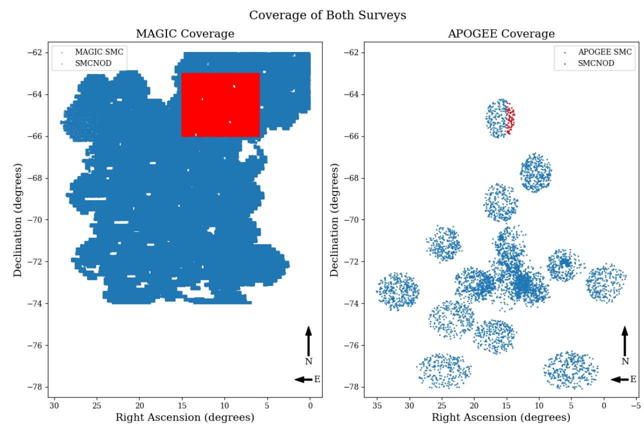

Figure 2. Spatial coverage of the MAGIC and APOGEE surveys. The SMCNOD is highlighted in red. Note that the axes are different for each panel and that APOGEE covers a wider range of α and δ values.

The SMCNOD has been plotted in red in both panels of the figure to highlight its location.

2.2. *Sample Selection*

To select stars that belong solely to the SMCNOD, we performed a number of cuts on the data from APOGEE, GAIA and MAGIC.

With the MAGIC data, we first ensure the quality of the sample by selecting stars with good metallicity measurements (i.e., within the range of values from the model used to calculate metallicity from the photometry), and a cut for 'ebv_sfd98' < 0.05, which ensures the data are not significantly affected by reddening. We then attempt to select non-variable stars by inspecting their $r-i$ and $g-r$ colors and checking that they are not listed as an entry in the GAIA DR3 variable star catalog. Next, we move to cuts based in the physical properties of the SMCNOD. We would like to select giants stars, so to do this we make cuts in surface gravity. Surface gravity is a measurement of gravitational acceleration at the surface of the object, and varies based on the size of star. When an average-mass star exhausts the hydrogen in its core, leaves the main sequence, and inflates into a red giant, its mass remains mostly the same while its radius increases by a factor of 10 to 1000 times its original size. Because they have the same mass as main sequence stars but much larger radii, the surface gravities of giant stars are orders of magnitude lower and can be used to filter giants from the main sequence. MAGIC calculates surface gravity values by combining their photometry with GAIA astrometry to generate isochrones, then matching each star to an isochrone and deriving log(g) from there. In MAGIC, we cut for $\log g < 3.5$ to select for red giant stars. This prevents us from selecting stars with radii too small to be classified as giants (mostly main sequence or dwarf stars). Further cleaning is done in proper motion space, but the implications on the properties of the SMCNOD are significant enough that we leave it until Section 3. After applying these selection criteria, we're left with 190,902 total stars and 1,892 stars in the SMCNOD spatial region. We plot these stars in Figure 3 to reference as a starting point when making our proper motion cuts.

We apply a different series of cuts to the APOGEE data. The only pure quality cut necessary is for stars with a signal-to-noise $(S/N) \geq 20$. We then cut for stars with $\log g < 3$. We use slightly different $\log g$ cuts for APOGEE and MAGIC because of the different methods of target selection between the surveys. APOGEE generally targets high luminosity, low $\log g$ stars (though it still suffers from some dwarf contamination due to the use of color-magnitude criteria), whereas MAGIC is not specifically searching for giants. The next step would usually be to make cuts in radial velocity, but much like

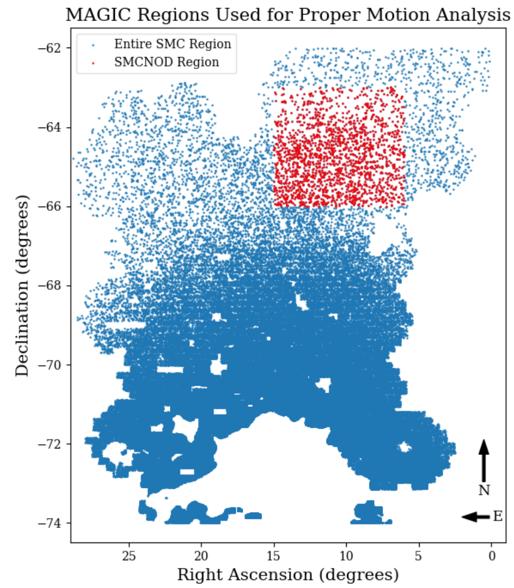

Figure 3. Spatial distribution of the remaining MAGIC data after quality, reddening, and $\log g$ cuts. All SMC points are plotted in blue, with stars in the SMCNOD region $(6^\circ < \alpha < 15^\circ, -66^\circ < \delta < -63^\circ)$ highlighted in red.

with proper motions, the implications require us to leave this until Section 3. After the cuts in S/N and $\log g$ we're left with 3,762 total APOGEE stars and 18 stars in the SMCNOD spatial region. We plot these stars in Figure 4.

3. ANALYSIS

3.1. *Proper Motions*

Proper motions can often distinguish an object from its background. For example, globular clusters within the SMC are centered at a much different location in proper motion space (proper motion in declination, μ_δ, vs. proper motion in right ascension, μ_α) than the SMC, and the SMC proper motions themselves can be separated from the Milky Way background. The MAGIC data have proper motions attached via the GAIA DR3 catalog (Gaia Collaboration et al. 2023b), and we use them to first examine the proper motions for the entire SMC field. Referencing the left plot in Figure 5, we observe the density of the proper motion distributions for all the stars in the MAGIC SMC coverage. The very dense cluster centered around $(\mu_\alpha, \mu_\delta) \approx (0.79, -1.22)$ milliarcseconds/year (mas/yr) represents the SMC (ac-

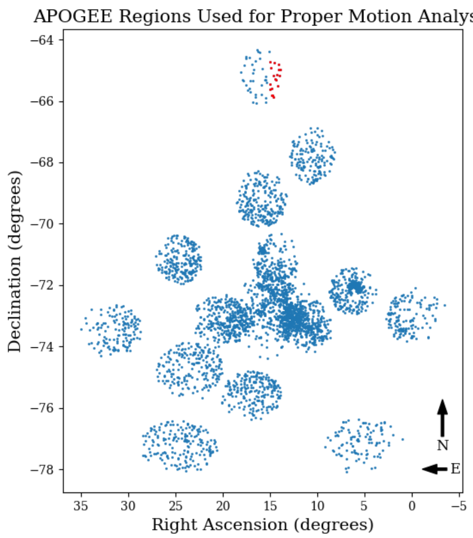

Figure 4. Spatial distribution of the remaining APOGEE data after S/N and $\log g$ cuts. All SMC points are plotted in blue, with stars in the SMCNOD region $(6° < \alpha < 15°, -66° < \delta < -63°)$ highlighted in red.

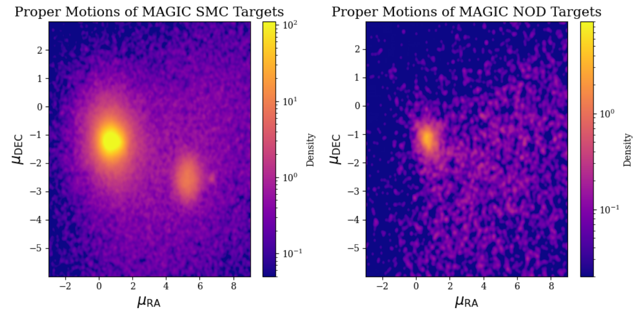

Figure 5. 2D histograms for both the entire SMC (left) and the SMCNOD (right) fields from the MAGIC catalog. We refer to Figure 3 for more information on the spatial bounds of these regions. This density map allows us to see clearly the centers of the proper motion distributions for a number of objects in the left plot. The SMC is the largest elliptical concentration at left, 47 Tuc the smaller one in the middle, and the globular cluster NGC 362 the much smaller circular concentration at right. The color of the two plots do not indicate the same density as can be observed by comparing their color bars. There are significantly less points in the SMCNOD and as such the color scale of the right panel has been adjusted to more clearly highlight its proper motions.

cording to D. L. Nidever et al. 2020). The other two densities are globular clusters, with 47 Tucanae shown at $\mu_\alpha \approx 5.27$, $\mu_\delta \approx -2.54$ mas/yr, and NGC 362 at $\mu_\alpha \approx 6.54$, $\mu_\delta \approx -2.51$ mas/yr [close to literature values (5.25, −2.55) and (6.69, −2.54) respectively from E. Vasiliev & H. Baumgardt 2021]. We then look to the SMCNOD field to compare its proper motions to those of the SMC. Considering only stars in the region in space where the SMCNOD resides, $(6° < \alpha < 15°, -66° < \delta < -63°)$, one observes that their proper motions are largely the same as that of the main body of the SMC (see Figure 5). This is also true of the SMCNOD stars we have from the APOGEE data. Aside from the SMC proper motions, we also see a large amount of contamination from the Milky Way, which is to be expected as the SMCNOD resides on the very periphery of the SMC.

We conclude that proper motions do not seem to distinguish the SMCNOD from the main body of the SMC, as they are largely similar to one another. This indicates that the overdensity is moving along with the SMC in the plane of the sky. We will discuss this further in Section 4. For now, we assume that we can cut all of the data according to SMC proper motions and correctly select SMCNOD stars. Therefore, we cut out an ellipse in proper motion space centered at $(\mu_\alpha, \mu_\delta) \approx (0.79, -1.22)$ with a semimajor axis length of 0.9 along the μ_α axis and a semiminor axis length 0.6 along the μ_δ axis. The lengths of these axes correspond to those chosen by D. L. Nidever et al. (2020) in their analysis of SMC giant stars. We apply this cut to both the MAGIC and APOGEE data. This leaves us with 3,151 total APOGEE points but only 8 stars in the SMCNOD, and 149,762 total MAGIC stars with 1,047 in the SMCNOD.

3.2. *Radial Velocities*

Because APOGEE contains radial velocity (RV) measurements, they can be used for further kinematical analysis. If the SMCNOD were a smaller body being absorbed by the SMC, it may have a unique motion as it is pulled towards the SMC center and perhaps disrupted by tidal forces. There is a slight problem with the RVs, as their values are distorted by the bulk motion of the SMC. We apply a kinematical model based on P. Zivick et al. (2021) which corrects for this motion and its impacts on RV values. Using the RVs from this model, we find the mean value to be 0.38 km/s, and cut within 3σ from -85.23 to 85.99 km/s. This selection brings our total APOGEE sample down to 3,135 stars, with 8 of them being in the SMCNOD. Seeing as none of the SMCNOD stars present before the cut based on mean RV are removed, it appears that the SMCNOD is once again kinematically consistent with the main body of the SMC.

3.3. *Metallicities*

The metallicity distribution function (MDF) of a galaxy is critical to understanding its development as it reveals information about the time period and environment in which star formation first occurred. Our initial analysis of the SMCNOD was motivated by the metallicity of stars in the general region, primarily because of the known metal-poor population there that, albeit with rather few data points, displayed an apparently bimodal MDF. First inspection of the MAGIC data revealed a similar bimodality, but after discovering an error in the catalog that biased some metallicities towards higher values, this bimodality disappeared, and the distribution is now peaked at a low mean metallicity (see Figure 6 for the MDFs from both datasets). The metallicity distribution for the SMCNOD in the MAGIC catalog is still quite interesting. It spans a wide range of metallicities from $-3.0 \lesssim [Fe/H] \lesssim -0.1$, including 8 very metal poor stars with $[Fe/H] < -2.5$.

In the top panel of Figure 7 we match the metallicities in the two data sets to look for the degree of consistency between the two catalogs' values. To ensure that we are comparing the same stars, we used a match radius for the coordinates between the catalogs of 0.5 arcseconds. The red line running through the data is the function y=x. If the data had a 1-to-1 relationship they would be perfectly aligned with y=x, which they are not. We attempt to find the source of this misalignment in the bottom panel of this same figure, where we plot the difference between the APOGEE and MAGIC metallicity

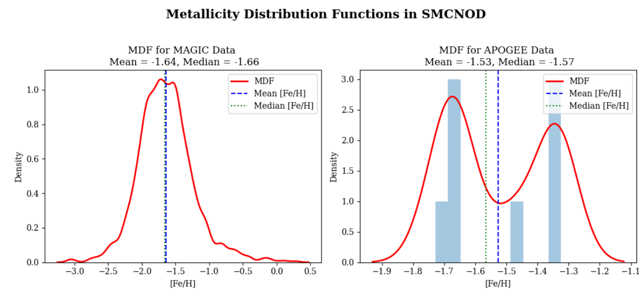

Figure 6. Metallicity distribution functions (MDFs) for the SMCNOD from our two different data sets, visualized as Gaussians using a Kernel Density Estimate (KDE). (Left) The MDF made using the MAGIC data for stars in the SMCNOD. This dataset is much larger, so a Gaussian is used to display it better. The mean [Fe/H] lies at −1.61 and the median at −1.64. (Right) The MDF from the APOGEE data for stars in the NOD. Since there are very few APOGEE data points we also plot the histogram of data behind the KDE. The mean [Fe/H] lies at −1.53 and the median at −1.57.

measurements versus the APOGEE measurements. We find that APOGEE tends to measure a systematically higher [Fe/H] by about +0.2 dex, though there seems to be more agreement at lower metallicity values. We continue with our analysis keeping this in mind.

To ascertain whether the lower metallicities in the SMCNOD represent a distinct population rather than an expected feature of the SMC we inspect the radial metallicity gradient of the SMC with and without the presence of the SMCNOD. Metallicity gradients arise in galactic stellar populations because the deepest part of the gravitational potential well is able to collect and retain the products of star formation events most effectively, thereby enabling it to experience more frequent star deaths and births, enriching the interstellar medium there with more continuous opportunity and ultimately achieving a broader MDF, in contrast to the outer periphery and stellar halos of galaxies, which see little self-enrichment. (Indeed, much of the variety of metallicity seen in the outer parts of galaxies is due not from self-enrichment, but from the accretion of smaller bodies, which themselves do not enrich to very high metallicities due to their smaller potential wells.) Figure 8 shows the MAGIC metallicities plotted as a function of distance from SMC center on the sky. We use the spherical law

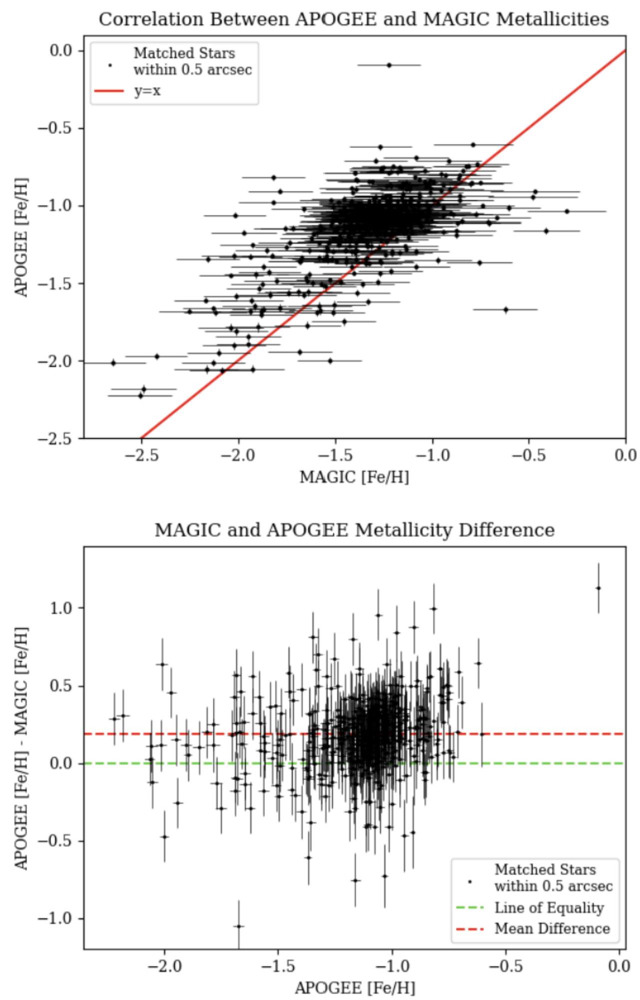

Figure 7. (Top panel) Correlation between MAGIC metallicity values (x-axis) and APOGEE values (y-axis). Error bars are plotted for both measurements. The red line running through the points is y=x, testing whether the data are 1-to-1. (Bottom panel) The difference in metallicity values between APOGEE and MAGIC (y-axis) versus the APOGEE values (x-axis). The red dashed line indicates the mean difference, and the green dashed line indicates where APOGEE and MAGIC would find equal values.

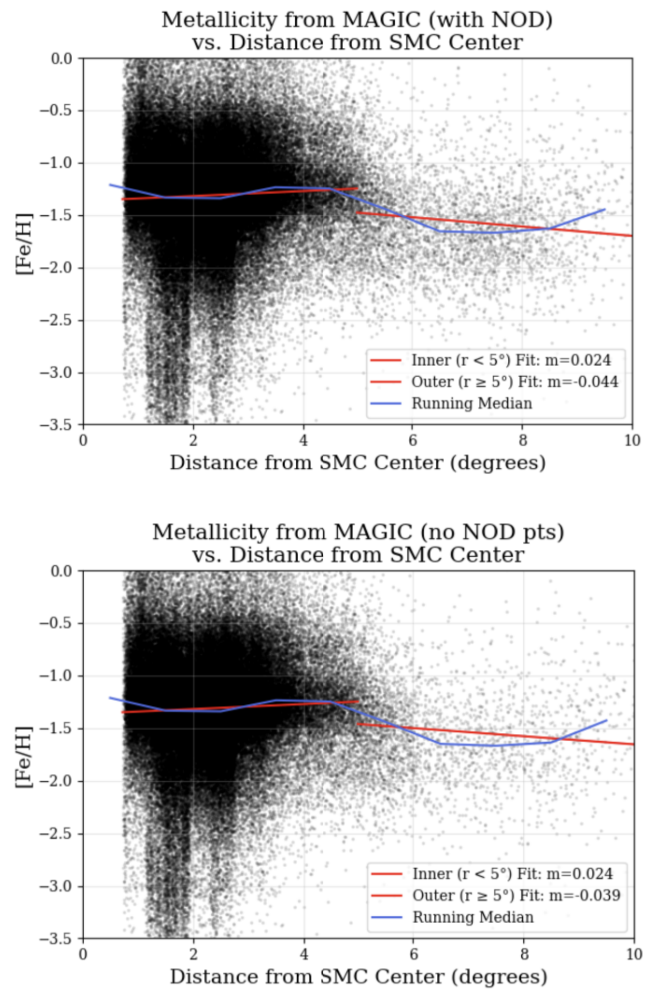

Figure 8. Metallicity as a function of angular distance from SMC center (in degrees on the sky). A linear fit and running median are plotted over the data. The linear fit is segmented at 5 degrees radius to account for the large difference in the metallicity distribution at lower and higher radii. The top panel shows the SMC population with SMCNOD points included. On the panel below we remove the SMCNOD stars to see how the linear fit and median are effected.

of cosines to find this angular distance. A linear fit and running median are plotted over the points to show how the metallicity varies over the whole of the SMC. The linear fit is segmented at 5 degrees to account for the differing metallicity distributions between the inner and outer regions of the SMC. It seems that the SMCNOD is not largely affecting the linear fit or the running median, which might suggest that the SMCNOD is part of a larger metallicity gradient. We do see a notable dip in median metallicity roughly 5 degrees from SMC center, indicating that the SMCNOD population may be more characteristic of the SMC's outskirts than its interior.

For further comparison, we inspect two different regions of the SMC, one eastern region without the SMCNOD and one western region with the over-density included. We split the two regions along $\alpha = 15°$ at the eastern boundary of the SMCNOD, as shown in Figure 9. We plot the counts as a function of radius for various metallicity populations in Figure 10, presenting the

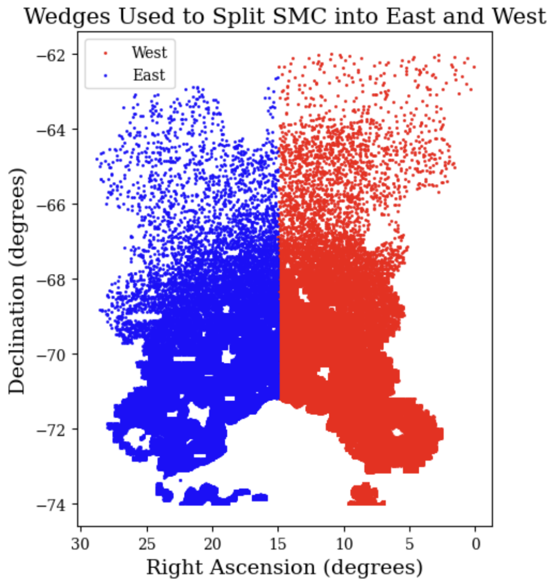

Figure 9. We split the SMC along $\alpha = 15°$, dividing it into a western "wedge" (red), containing the SMCNOD, and an eastern wedge (blue).

region without the SMCNOD below, and the region containing the SMCNOD above. These plots better illustrate the magnitude of the overdensity and the metallicity of the stars that comprise it. We see that the power law approximation, plotted in red, has a shallower slope in the region containing the SMCNOD (above) than the region without the SMCNOD (below) for metallicity values from $-1.25 > [Fe/H] > -1.75$. This is the result of an overdensity of stars at these metallicities at further distances from SMC center, particularly at the location of the SMCNOD.

3.4. *Isochrones*

Isochrones are functions in color-magnitude space that describe the primary loci of a single metallicity population of stars of different masses evolved to a specific age. This is especially useful in analyzing the SMCNOD as if its origin were as a dwarf satellite, because such systems are expected to host a more metal-poor, older population with a distinct isochrone. Isochrone-fitting could also support the "material pulled out of the inner SMC by the LMC" theory if the age and metallicity of the SMCNOD match that of the central SMC.

First, we generate a color magnitude diagram (CMD) for a region spanning $5° < \alpha < 17°$ and $-67° < \delta < -63°$, which contains the SMCNOD. We convert the g and r band measurements into absolute magnitudes assuming the SMCNOD is at the same distance as the SMC main body. We use a distance of ∼61.94 kpc, which was also used in A. Pieres et al. (2017) based on the distance modulus calculated in R. de Grijs & G. Bono (2015). We then plot g vs. $g - r$ (assuming negligible reddening) and obtain the leftmost panel of Figure 11. Repeating this process for a comparison field adjacent to the SMCNOD, which spans the exact same area on the sky (spanning $17° < \alpha < 29°$ and $-67° < \delta < -63°$), we obtain the control field in the middle plot. Finally, to further isolate the SMCNOD population, we filter for points in the SMCNOD CMD which are further than 0.012 magnitudes in both bands from the control field. What this means is that we calculate the Euclidean distance between SMCNOD and control points on the CMD, which we can do since both g and $g-r$ are in the same units. The value 0.012 is chosen specifically because this selects enough points that the newly filtered SMCNOD sample contains the same amount of stars as the comparison sample.

We generate our isochrones with [Fe/H] as our main fixed value, since we have more reliable values for this parameter. We use [Fe/H] $= -1.64$ and -1.57 as our two metallicities to represent both the MAGIC and APOGEE samples. We then allow our age to vary from 2.33 to 10.5 Gyr. The lower end of these ages is motivated by J. T. Povick et al. (2023), whose radial [Fe/H] distribution for SMC stars around this age (presented in their Figure C4, here as Figure 12) resembles ours in form. The upper end of test isochrone ages is motivated by previous Magellanic Cloud star formation histories, which extend their limits back to around this time (such as D. L. Nidever et al. (2020)). If the SMCNOD population has an age corresponding to a previously studied SMC starburst, we gain further insight into its origin. We include within this range a 6 Gyr isochrone to match

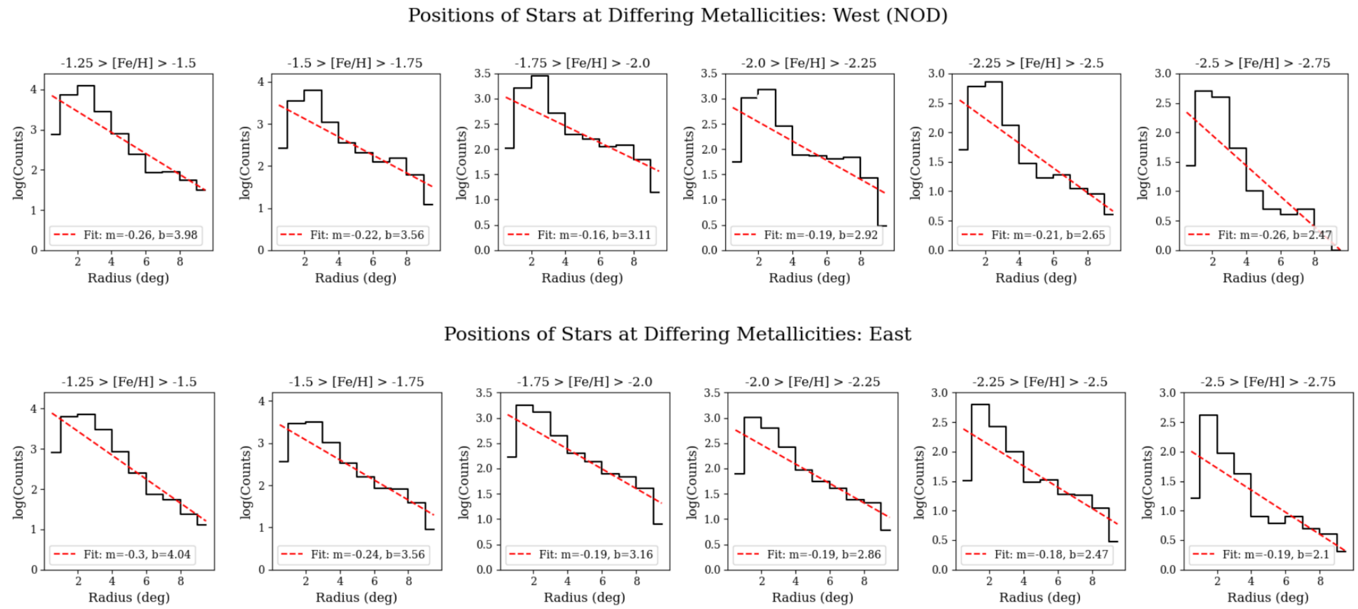

Figure 10. The radial density profile of stars for populations of various metallicity. Below we show the eastern SMC field, which does NOT contain the SMCNOD, and above we show the western field which does. Higher metallicity population are at top and lower metallicity at bottom. The red dashed line is a power law approximation for the radial gradient at each metallicity.

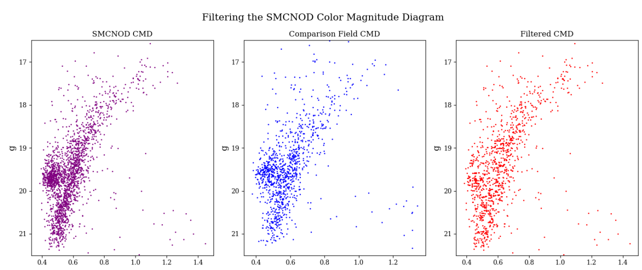

Figure 11. (Left) Color magnitude diagram of the SMCNOD field before subtraction. (Middle) CMD for an adjacent field that does not contain the SMCNOD, used as a comparison for the SMCNOD field. (Right) The SMCNOD CMD with points less than 0.012 magnitude from those in the comparison field subtracted out of the plot. This "distance" is calculated using $d = \sqrt{(g-r)^2 - (g)^2}$.

the one fit to the SMCNOD by the original discovery paper (A. Pieres et al. 2017).

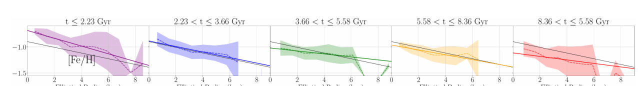

Figure 12. Radial metallicity ([Fe/H]) gradient from J. T. Povick et al. (2023), which inspired the age range for our isochrones and is later used for comparison to our own metallicity gradient presented in Figure 8.

With both the isochrones and the CMD prepared, we then plot the isochrones over the CMD and calculate the distance between the isochrone and every point in the CMD. We do this for each of the nine values in our age range. Because the units of g and $g-r$ are the same, we can calculate this distance using the standard Euclidean distance. We use a distance of 0.050 magnitudes from the isochrone to define "closeness" since the SMCNOD's branch on the CMD is roughly this wide, and it allows us to avoid selecting stars on the apparent blue horizontal branch, which is not included in these models. In Tables 2 and 3 we list the percentage of points from the isochrone that fell within the distance required for the isochrone to be marked "close". From the isochrones we tested, the 6.0 Gyr, [Fe/H] $= -1.57$ plot appears to fit the best. This combination of age and metallicity happens to be in agreement with those found in the SMCNOD discovery paper, and would place these stars among the intermediate age population in the SMC.

3.5. *Using GMM Clustering*

In an attempt to discover further connections between the different stellar populations in the SMC, and intend-

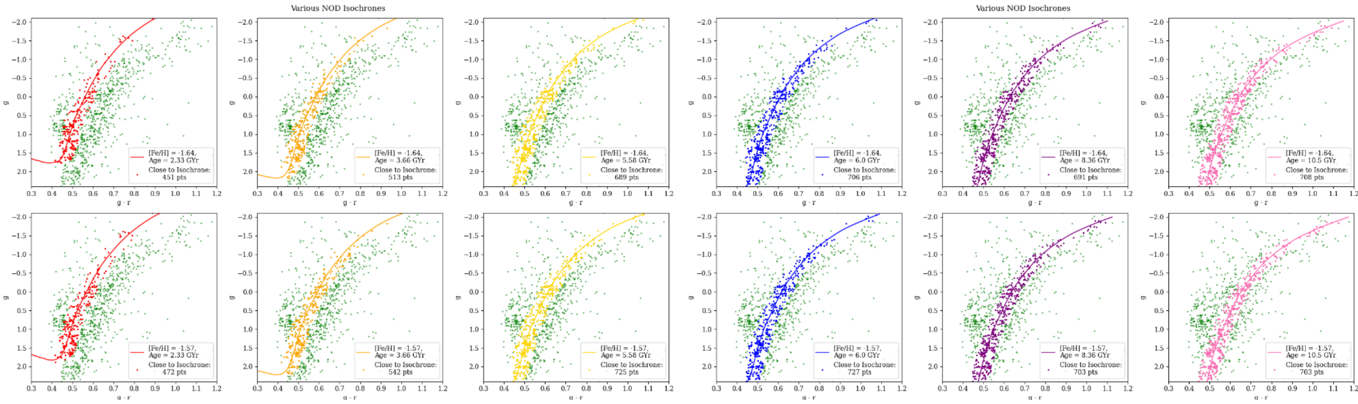

Figure 13. The range of test isochrones, varying from 2.33 to 10.5 Gyr in age. The top plots have [Fe/H] = −1.64 and the bottom −1.57. Green points represent the color magnitude diagram for all the stars in the SMCNOD, whereas the colored points represent stars that are close to the fitted isochrone.

Age (Gyr)	Metallicity	Points Near Isochrone
2.33	-1.64	21.9%
3.66	-1.64	24.9%
5.58	-1.64	33.4%
6.0	-1.64	34.2%
8.36	-1.64	33.5%
10.5	-1.64	34.3

Table 2. Table of parameters for [Fe/H] = -1.64 stars

Age (Gyr)	Metallicity	Points Near Isochrone
2.33	-1.57	22.9%
3.66	-1.57	26.3%
5.58	-1.57	35.1%
6.0	-1.57	**35.2**%
8.36	-1.57	34.1%
10.5	-1.57	34.1

Table 3. Table of parameters for [Fe/H] = -1.57 stars

ing to characterize the SMCNOD as belonging to one of these populations, we apply Gaussian Mixture Modeling. Gaussian Mixture Modeling (GMM) is a probabilistic method used to model subpopulations, or clusters, within a larger set of data. GMM represents the data as a series of Gaussian distributions, each characterized by statistical parameters like the mean, covariance, and mixing probability. It then assigns each data point a probability that it belongs to a certain cluster. The model then uses what is called an "Expectation-Maximization" algorithm to calculate the probability that each point belongs to a certain cluster and then updates the statistical parameters of the Gaussians to better fit the data. This optimization repeats until the model finds the best fit.

We use the built-in GaussianMixture() function from scikit-learn (F. Pedregosa et al. 2011), a Python package designed for machine learning, to reduce all these steps into a single line of code. Its inferences are based on the kinematical features of the data (right ascension, declination, and proper motions in both RA and DEC), as well as metallicity. We have also applied Bayesian and Akaike Criterion (BIC & AIC respectively) techniques to determine the optimal number of clusters, which is found to be three. In Figure 14, we display the distribution of all three clusters in the top right, with the rest of the panels displaying each cluster individually.

From the clusters the model identifies, clusters 0 and 1 cover a large area across the SMC, while cluster 2 is much denser around SMC center. To further identify the properties of each cluster, we look to Figure 15, where we plot the metallicity distribution functions (above) and proper motions in right ascension (below). Cluster 2 spans a very wide range of metallicities, with peaks at both extremes, while clusters 0 and 1 are both peaked around the same value. This indicates that cluster 2 contains a much more diverse stellar population. We then use kinematics to distinguish clusters 0 and 1 since their proper motions are quite different. Cluster 0 has an average μ_α, which is ~ 0.5 mas/yr greater than

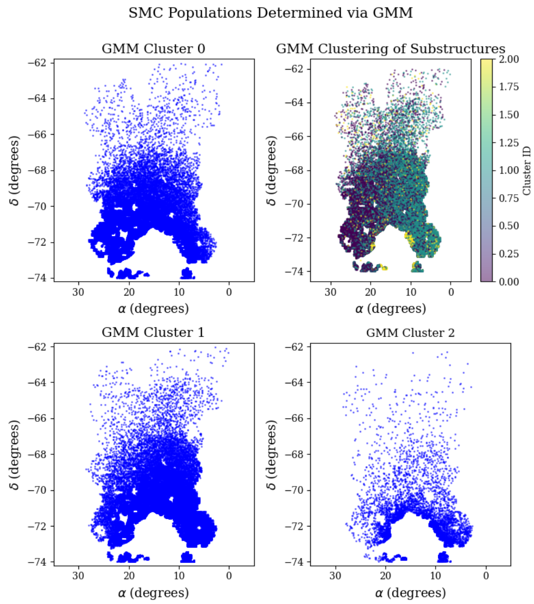

Figure 14. Plotted at the top right are all 3 clusters on the sky, colored by cluster according to the bar at right. The remaining panels display each cluster plotted in RA/DEC, with the index of the cluster in the title.

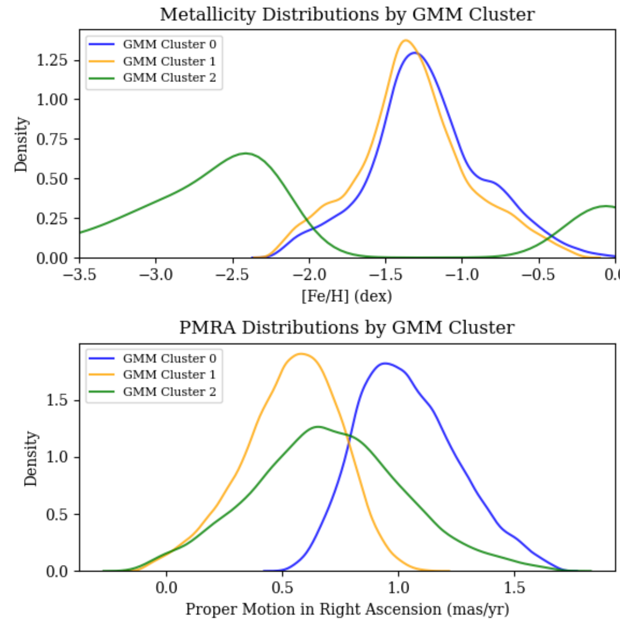

Figure 15. (Above) Metallicity distributions of all three clusters. While clusters 1 and 2 have similar mean metallicities, cluster 3 spans a wide range. (Below) Distributions of proper motion in right ascension for each cluster. Cluster 0's mean proper motion is higher than both 1 and 2, a possible sign of LMC gravitational influence.

cluster 1. A higher μ_α indicates greater motion to the east, the direction of the LMC, which may be a sign of gravitational interaction. With all this in mind, we speculate that because clusters 0 and 1 span the entire SMC, and are characterized by a single metallicity (as opposed to the diversity we see in cluster 2), that they may represent a halo population in which one half is experiencing more tidal influence from the LMC. The potential of SMCNOD stars belonging to the SMC halo would be a key component in determining its origin.

4. DISCUSSION

Below we discuss how our data inform the distinction of the SMCNOD from the SMC, as well as how they weigh in on the possible origin scenarios and which ones are now most favored.

4.1. *SMCNOD in Comparison to the SMC*

We are fortunate to have both proper motions and radial velocities (however limited) for our analysis of the SMCNOD's kinematics. Our ultimate purpose is to infer the origin of the overdensity, and so we compare its kinematical data to those of the SMC main body. In proper motion space, we reference Figure 5 once again to see that the proper motions of the SMCNOD are almost identical to the SMC. No other densities exist in this figure, only a scattering of background Milky Way stars. As far as radial velocities, we already know that all SMCNOD points present after quality and proper motion cuts are within 3σ of the mean RV of the SMC. We add Figure 16 here to demonstrate just how well the SMCNOD RVs comply, with 7/8 points visibly very close to that of the mean. What this means for the SMCNOD is that its motion is coupled to that of the SMC. This does not rule out any SMCNOD origin scenarios, but it makes it slightly more difficult to justify the dwarf satellite theory.

Looking to metallicities, we find that the SMCNOD [Fe/H] values correspond to expected radial gradients for the SMC main body from the literature. Figure 8 clearly displays that when we remove the SMCNOD population from the data, the running median and linear fit for the overall SMC hardly change. We compare our plot to

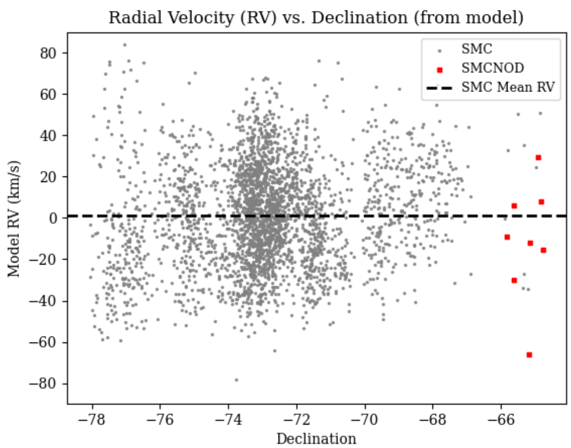

Figure 16. Radial velocities for all stars in the SMC, plotted against declination as to make the SMCNOD clearly visible on the north (right) end. All stars in the SMC field are in gray and SMCNOD stars are marked as big, red squares.

Figure C4 of J. T. Povick et al. (2023) (our Figure 12 where they show a similar [Fe/H] gradient for populations of various ages. Our radial metallicity distribution displays the same drop in metallicity at ~ 6 kpc from SMC center (at the distance of the SMC 1 kpc $\approx$ 1 degree) as shown in their figure for stellar populations with ages $3.66 < t < 8.36$. While the drop-off may indicate some characteristic unique to the outer regions of the SMC, the SMCNOD is once again not distinguished from the SMC as a whole. A low metallicity population is expected in the stellar halo of a dwarf galaxy like the SMC.

The isochrone fit to the SMCNOD is peculiar in the context of the larger SMC. The best-fit isochrone indicates an age of 6 Gyr, placing it among the intermediate-age population in the SMC. However, if the SMCNOD does represent a stellar halo population, it is expected that this region should contain older stars as the lack of gas in the outskirts prevents recent star births. It is also interesting that the MAGIC data is characterized by an isochrone of this age, since the metallicity used for the fit comes from the APOGEE data.

4.2. *The Origin of the SMCNOD*

To list them again, the proposed origin scenarios of the SMCNOD are as follows:

1. Simply an extension of the SMC's main body (i.e., "disk").
2. Material pulled out of the inner SMC by tidal interactions with the LMC (and/or tidal dwarf galaxy).
3. Substructure in the SMC's halo and/or accreted debris as part of its hierarchical formation.
4. A dwarf satellite absorbed by the SMC.

The analysis of our data provides evidence for against both for and against some of these options.

As to the first theory, the drop-off in metallicity at a ~ 6 kpc radius is the primary point against it, showing that the SMCNOD does not largely follow the metallicity profile of the central SMC. Otherwise, the kinematics of the SMCNOD align with the main body and there is an intermediate aged population that exists in the central SMC containing stars 6 Gyr in age P. Massana et al. (2022).

On the other hand, the SMCNOD's alignment with the established overall metallicity gradient of the SMC from J. T. Povick et al. (2023) serves as strong evidence against the "dwarf satellite" hypothesis. The 6 Gyr age of the SMCNOD is also a bit young for a primordial dwarf satellite that would have formed in metal-poor intergalactic space. The fact that its proper motions and radial velocities resemble the SMC is not a point against this theory, but if they were distinct it could have been more definitive evidence for this theory.

The "substructure in the SMC halo" theory is hampered by the age we determine for the SMCNOD. Our ~ 6 Gyr determination is slightly younger than the 7-8 Gyr and 10 ± 2 Gyr halo population age estimates found by A. E. Piatti (2015) and D. Hatzidimitriou & L. T. Gardiner (1992) respectively. This may mean that the SMCNOD is a distinct population, but it is also possible that the younger age is a result of later accumulation.

We're then left with Theory 2. The idea that the SMCNOD was created by material pulled out of the inner SMC by tidal interaction with the LMC seems the most likely. This would not mean that stars themselves were pulled out of the central SMC, but rather that HI gas has been stripped from the SMC by the LMC and created a localized starburst. This process is what led to the large 2 Gyr stellar population in the Magellanic Bridge. However, if the SMCNOD truly was created during a starburst generated by stripped SMC gases, it represents a somewhat unique population. Aside from localized regions like the bridge's 2 Gyr population, there's no indication of any particular increase in star formation around this time. The star formation history of the SMC plotted in P. Massana et al. (2022) shows no indication of major activity around 6 Gyr, and J. Harris & D. Zaritsky (2004) even reports a "long quiescent period" from $3 < t < 8.4$ Gyr ago where there was little star formation occurring. The SMCNOD does display some properties of what is called a "dwarf tidal galaxy", an object created from gases pulled out of another object into intergalactic space (P. A. Duc & I. F. Mirabel 1999; P.-A. Duc 2012), though these are primarily observed in mergers of more massive galaxies. The SMCNOD is composed of younger stars that may originate from expelled HI clouds, has a metallicity inherited from the outer regions of its parent galaxy, and appears to still be gravitationally bound.

5. CONCLUSIONS

In conclusion, we find that we are likely able to disregard the "dwarf satellite" theory, that there's tentative evidence for the "extension of the main body" and "substructure in the SMC halo" theories, but that the SMCNOD origin scenario most in line with our evidence is that of it being "material pulled from the SMC by tidal interactions with the LMC" and/or a tidal dwarf galaxy. In Table 4 we color code the original Table 1 that displays each theory and their properties by how well our data aligns with that property.

Our findings are in no way definitive, and future studies could help in increasing our understanding of the SMCNOD. We currently work with very little radial velocity data in the NOD, as well as very few chemical element abundances aside from [Fe/H]. Another low surface-brightness spectroscopic survey in the same realm as APOGEE, to follow up on the photometric work done by MAGIC, could dramatically increase the amount of data in the region.

ACKNOWLEDGMENTS

This thesis makes use of data from the APOGEE and MAGIC surveys, cataloged and processed by members of the SDSS team and DELVE collaboration. The author would like to thank the Department of Astronomy at the University of Virginia, the DELVE Collaboration, and the APOGEE team, as well as Steven Majewski, Anirudh Chiti, David Nidever, and graduate student Andres Almeida.

REFERENCES

Abdurro'uf, Accetta, K., Aerts, C., et al. 2022, ApJS, 259, 35, doi: 10.3847/1538-4365/ac4414

Belokurov, V., Zucker, D. B., Evans, N. W., et al. 2006, ApJL, 642, L137, doi: 10.1086/504797

Besla, G., Kallivayalil, N., Hernquist, L., et al. 2007, ApJ, 668, 949, doi: 10.1086/521385

Blanton, M. R., Bershady, M. A., Abolfathi, B., et al. 2017, AJ, 154, 28, doi: 10.3847/1538-3881/aa7567

de Grijs, R., & Bono, G. 2015, AJ, 149, 179, doi: 10.1088/0004-6256/149/6/179

Deason, A. J., & Belokurov, V. 2024, NewAR, 99, 101706, doi: 10.1016/j.newar.2024.101706

Duc, P.-A. 2012, in Astrophysics and Space Science Proceedings, Vol. 28, Dwarf Galaxies: Keys to Galaxy Formation and Evolution, ed. P. Papaderos, S. Recchi, & G. Hensler, 305, doi: 10.1007/978-3-642-22018-0_37

Duc, P. A., & Mirabel, I. F. 1999, in IAU Symposium, Vol. 186, Galaxy Interactions at Low and High Redshift, ed. J. E. Barnes & D. B. Sanders, 61

Origin Scenario	Kinematics	Metallicity	Age	Likelihood
Extension of Main Body	Similar to SMC	Follow Metallicity Gradient	Young to Intermediate	Somewhat Likely
Material pulled from SMC by Tidal Interactions with the LMC (and/or Tidal Dwarf Galaxy)	Similar to SMC	Follow Metallicity Gradient	Young to Intermediate	Likely
Substructure in the SMC Halo or Accreted Debris	Similar to SMC	Follow Metallicity Gradient	Intermediate to Old	Somewhat Likely
Dwarf Satellite	Similar OR Distinct to SMC	Differs From Metallicity Gradient	Intermediate to Old	Unlikely

Table 4. The four proposed origin scenarios for the SMCNOD and the measurements that would validate them, now colored by how well our findings align. Red means that a property is not consistent with our findings, yellow means somewhat consistent, green means consistent.

Eisenstein, D. J., Weinberg, D. H., Agol, E., et al. 2011, AJ, 142, 72, doi: 10.1088/0004-6256/142/3/72

Gaia Collaboration, Brown, A. G. A., Vallenari, A., et al. 2016, A&A, 595, A2, doi: 10.1051/0004-6361/201629512

Gaia Collaboration, Vallenari, A., Brown, A. G. A., et al. 2023a, A&A, 674, A1, doi: 10.1051/0004-6361/202243940

Gaia Collaboration, Vallenari, A., Brown, A. G. A., et al. 2023b, A&A, 674, A1, doi: 10.1051/0004-6361/202243940

Harris, J., & Zaritsky, D. 2004, AJ, 127, 1531, doi: 10.1086/381953

Hatzidimitriou, D., & Gardiner, L. T. 1992, in IAU Symposium, Vol. 149, The Stellar Populations of Galaxies, ed. B. Barbuy & A. Renzini, 157

Hindman, J. V., Kerr, F. J., & McGee, R. X. 1963, Australian Journal of Physics, 16, 570, doi: 10.1071/PH630570

Kallivayalil, N., van der Marel, R. P., Besla, G., Anderson, J., & Alcock, C. 2013, The Astrophysical Journal, 764, 161, doi: 10.1088/0004-637x/764/2/161

Majewski, S. R., Schiavon, R. P., Frinchaboy, P. M., et al. 2017, AJ, 154, 94, doi: 10.3847/1538-3881/aa784d

Massana, P., Ruiz-Lara, T., Noël, N. E. D., et al. 2022, MNRAS, 513, L40, doi: 10.1093/mnrasl/slac030

Mathewson, D. S., Cleary, M. N., & Murray, J. D. 1974, ApJ, 190, 291, doi: 10.1086/152875

Nidever, D. L., Majewski, S. R., Butler Burton, W., & Nigra, L. 2010, ApJ, 723, 1618, doi: 10.1088/0004-637X/723/2/1618

Nidever, D. L., Hasselquist, S., Hayes, C. R., et al. 2020, ApJ, 895, 88. https://arxiv.org/abs/1901.03448

Pedregosa, F., Varoquaux, G., Gramfort, A., et al. 2011, Journal of Machine Learning Research, 12, 2825, doi: 10.48550/arXiv.1201.0490

Piatti, A. E. 2015, MNRAS, 451, 3219, doi: 10.1093/mnras/stv1179

Pieres, A., Santiago, B. X., Drlica-Wagner, A., et al. 2017, MNRAS, 468, 1349. https://arxiv.org/abs/1612.03938

Povick, J. T., Nidever, D. L., Massana, P., et al. 2023, arXiv e-prints, arXiv:2310.14299, doi: 10.48550/arXiv.2310.14299

Searle, L., & Zinn, R. 1978, ApJ, 225, 357, doi: 10.1086/156499

Vasiliev, E., & Baumgardt, H. 2021, MNRAS, 505, 5978, doi: 10.1093/mnras/stab1475

White, S. D. M., & Rees, M. J. 1978, MNRAS, 183, 341, doi: 10.1093/mnras/183.3.341

Zivick, P., Kallivayalil, N., & van der Marel, R. P. 2021, ApJ, 910, 36, doi: 10.3847/1538-4357/abe1bb