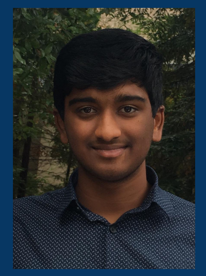
Machine-Based Spectral Classification of Weak CN & Carbon Stars in M31

UNIVERSITY OF CALIFORNIA SANTA CRUZ
SIP SCIENCE INTERNSHIP PROGRAM



Suhas Kotha[1], Rachel Raikar[2], Puragra GuhaThakurta[2]

1. Evergreen Valley High School, San Jose, CA 2. University of California Santa Cruz, Santa Cruz, CA

Introduction

The Importance of Studying Star Evolution

Stars come in a vast range of sizes and temperatures, and their physical characteristics vary greatly over the course of their evolution. Studying the existing variation within a population of stars in a certain galaxy can therefore reveal a significant amount of information about that galaxy's age and initial formation. Although rare and short-lived, luminous stars are particularly useful for such studies, as their brightness allows them to be viewed from greater distances than stars of average brightness, thereby enabling the study of more distant galaxies.

Andromeda as a Test Bed

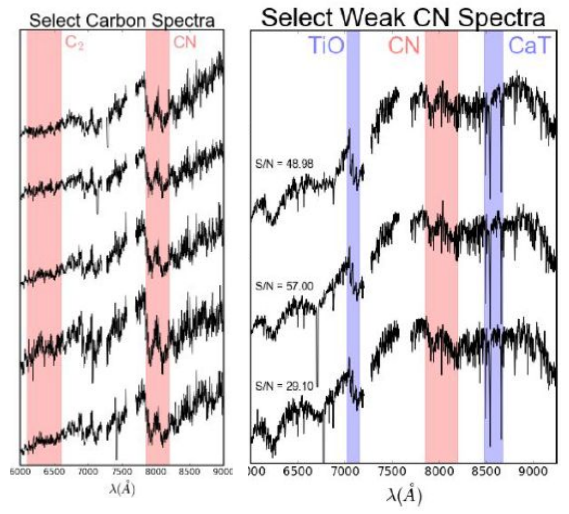
The proximity of the Andromeda Galaxy (M31) to our own Milky Way Galaxy, as well as the vast size of its disk, create near-optimal survey conditions for luminous stars.

Image 1: M31 with the area of interest outlined.

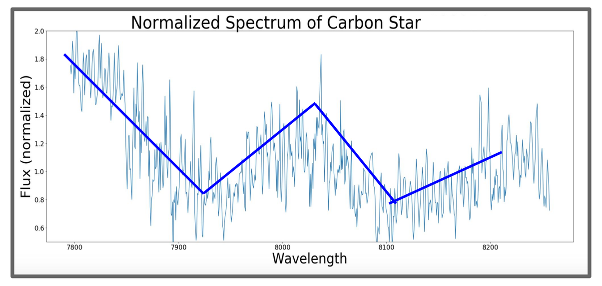
Identifying Evolved Stars in M31

M31 contains a population of stars referred to as the "Asymptotic Giant Branch" (AGB) that represent a later, more luminous stage of star evolution. Previous work such as that of Katherine Hamren et. al has revealed that AGB stars are primarily characterized by the ratio (C/O) of carbon and oxygen in their atmospheres. If the C/O ratio of a particular star exceeds or falls below 1, carbon or oxygen molecules will, respectively, dominate the star's atmosphere. Prior analysis of spectroscopic and photometric data of about 10,000 individual stars in the disk of M31 has indicated that the galaxy contains AGB stars both with C/O > 1 (referred to as carbon stars) and C/O < 1 (referred to as "other" or "normal" stars).

Discovery of "Weak CN" Stars

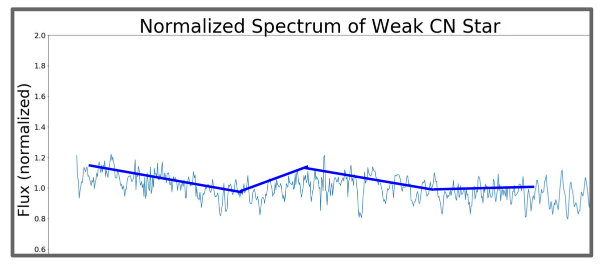


Further analysis of the M31 data also revealed an unusual population of about 168 AGB stars whose spectra display evidence of both the carbonaceous molecule cyanogen (CN) and oxygen-based molecules such as titanium oxide (TiO). With a C/O ratio of ~1, these stars do not fit into the previously defined AGB categories of carbon or "other," and they were temporarily given the name of "weak CN" stars after their most prominent characteristic, a diluted version of the CN feature. As this feature is also present, albeit more strongly, in carbon star spectra, but absent from the spectra of "other" stars, we hypothesized that an association may exist between the weak CN and carbon populations.



Visual Weak CN Classification

Carbon stars contain a <u>strong</u> double-peaked "W" absorption feature around 7800-8200 Å. This spectral feature is distinct for carbon stars; it is not present in "other" stars.



Weak CN stars contain a <u>weak</u> double-peaked "W" absorption feature around 7800-8200 Å. The spectral feature was used to visually classify stars as weak CN.

Data

Spectroscopic Landscape of Andromeda's Stellar Halo (SPLASH): Spectroscopic data collected with DEIMOS (DEep Imaging Multi-Object spectograph) at Keck Observatory.

Image 2 : Keck Observatory

Panchromatic Hubble Andromeda Treasury (PHAT): Photometric data measuring light magnitudes across wavelength bands ranging from ultraviolet to infrared.

Image 3: Hubble Space Telescope

Summary and Objectives

Upon identifying a previously undiscovered population of weak CN stars in the Andromeda Galaxy with spectral features that were similar, but not identical, to the features of carbon star spectra, we hypothesized that weak CN stars may constitute their own distinct stellar population. Our goal was to use the SPLASH and PHAT datasets to develop an automated algorithm capable of objectively classifying stars as "carbon," "weak CN," or "other" based on spectral features. Given the pre-identified similarity between weak CN and carbon stars, especially on the 7800-8200 Å range, the main basis of this classification consisted of comparisons between test spectra and a high-fidelity carbon star template spectrum on this limited wavelength range. Using these comparisons, we hoped for the algorithm to classify stars that were most similar to the template as carbon stars, stars that were less similar but still matched the template fairly well as weak CN stars, and stars that did not match at all as "other" stars. Ultimately, our objective is to create a machine-classified set of weak CN stars that will provide more reliable data for further studies aimed at obtaining a better understanding of the weak CN stage in stellar evolution and the properties of weak CN stars in relation to other types of stars.

Methods

Polygon Identification

To identify which star belonged to what class, we leveraged the apparent clustering of the stars on certain graphs. We identified polygons on these graphs that best symbolized the clustering. For certain graphs, there are inner and outer polygons to identify degrees of significance. There are 0 points assigned for falling in no polygon, 0.5 points for the outer polygon, and 1 point for the inner polygon. Higher point values correspond to higher similarity.

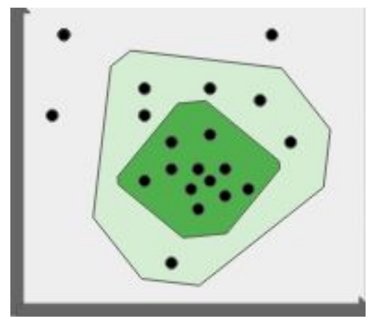


Fig. 1: Sample graph to demonstrate Polygon Identification method

Metric 1: Score Tests

We used chi-squared statistics to observe how a star's spectrum compared against the average star, or "template". We computed a score against a carbon template and Weak CN template. A lower score signifies higher correspondence

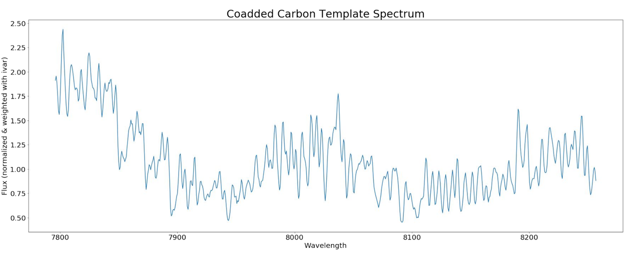


Fig. 2: Carbon Template, or average spectrum of a carbon star

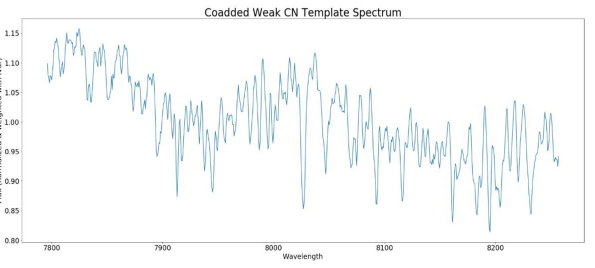


Fig. 3: Weak CN Template, or average spectrum of a Weak CN star

Metric 2: Dilution Factor

In a different approach to quantifying the similarity between test weak CN spectra and the carbon template, we gave the template two degrees of freedom – tilting and dilution – so that it could be machine-adapted to best match the test spectrum. The chi-squared "score" was once again used to determine how close the match between the template and test spectrum was, and the template was tilted and diluted in increments until the dilution factor that produced the lowest score (a.k.a. the optimal dilution factor) was found. Higher optimal dilution factors indicated that the template had to be modified more, meaning that the test spectrum was a worse match.

Metric 3: Slope Tests

Since the goodness-of-fit tests work well in analyzing fine-scale differences between spectra, we used another metric to focus analysis on determining the broader shape. We defined four slopes in the spectral region of the W, as shown below, such that higher magnitude slopes meant stronger W shapes.

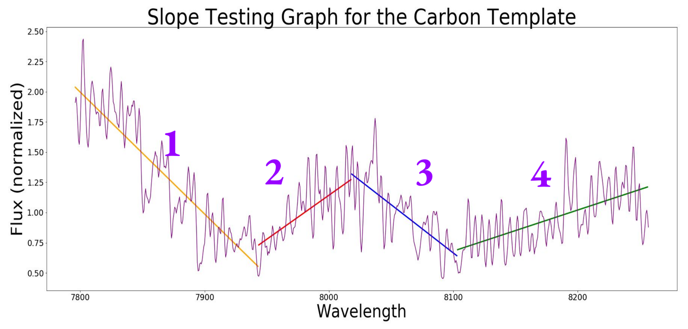


Fig. 4: The four slopes calculated for the W-shaped spectral region.

Metric 4: 1st U vs. Double U

The presented "Double U" can be broken down into corresponding U's. Observation has yielded that in most cases, the first U is more consistent then both U's for Weak CN stars. Therefore, when isolating the Weak CN stars, we used the first U, and when isolating the carbon stars, we used both U's.

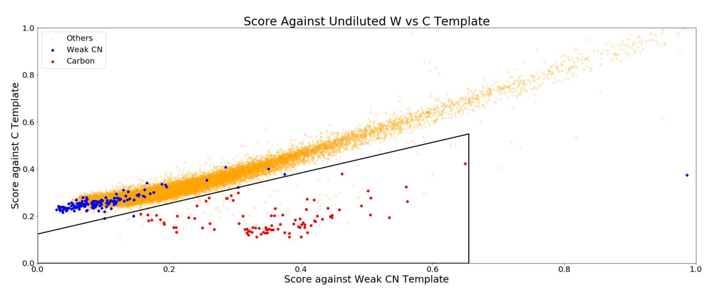
Classification Sequence

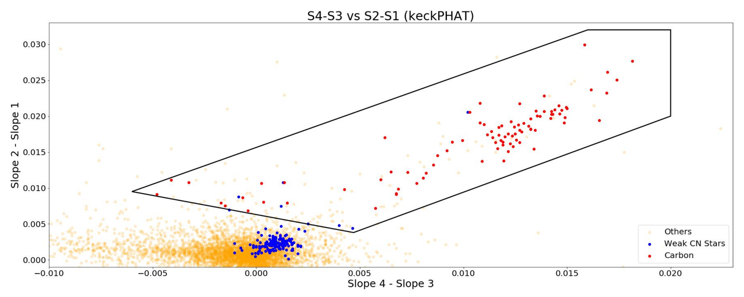
We first separated the carbons from the stars then went on to separate the Weak CN stars. We used the following plots to see how the stars fared against various metrics of correspondence

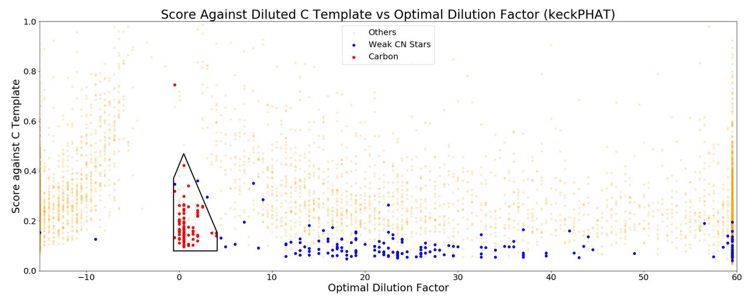
Carbon	Weak CN
Scores: Carbon vs Weak CN Template	Scores: Carbon vs Weak CN Template (1U)
⇩	⇩
Slope 4 - Slope 3 vs Slope 2 - Slope 1	Slope 4 - Slope 3 vs Slope 2 - Slope 1
⇩	⇩
Scores: Diluted C Score vs Dilution Factor	Scores: Diluted W Score vs Dilution Factor

Score Against Undiluted W vs C Template | Slope 4 - Slope 3 vs. Slope 2 - Slope 1 | Diluted C Score vs. Optimal Dilution Factor

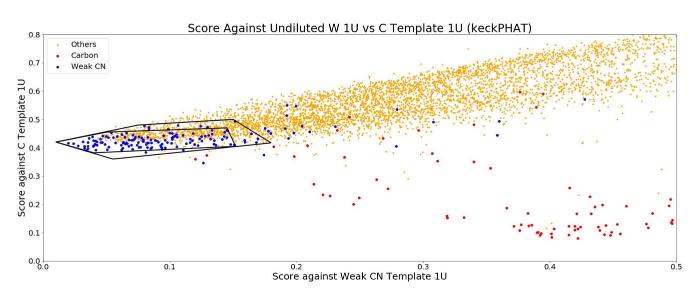
Carbon

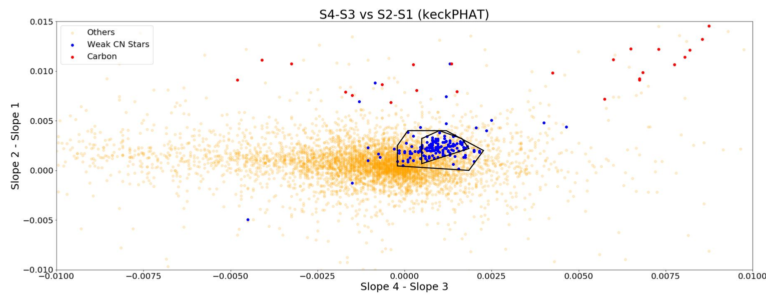


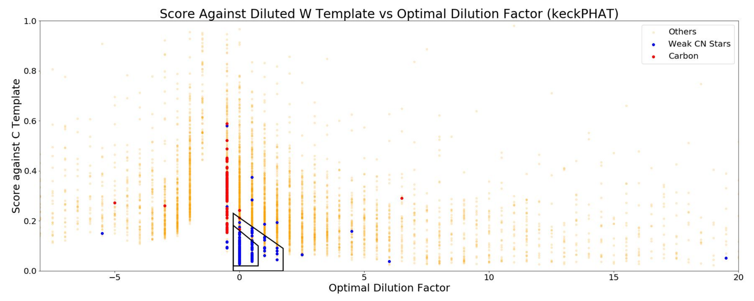




Weak CN







Score Against Undiluted W vs C Template (1st U only) | Slope 4 - Slope 3 vs. Slope 2 - Slope 1 | Diluted W Score vs. Optimal Dilution Factor

Results

Color-Magnitude Diagrams

	Carbon	Not Carbon
carbon	89	3
kphOthers	12	5574
wNm	0	162

Carbon Classification Results: A star is carbon if it falls into all polygons

	0	0.5	1	1.5	2	2.5	3
kphOthers	4015	777	594	126	29	17	28
wNm	14	5	9	9	12	26	86

Weak CN Results: A star is Weak CN if it has a score of 2.5 or 3, as indicated by the purple box

Fig. 5: Color-magnitude diagrams displaying the positions of weak CN stars. The left pictures visual classification and right pictures machine classification

These color magnitude diagrams show how the machine classification has improved clarity by producing a much tighter and well defined grouping of the rares-stars

Conclusions

Based on the results produced by the various methods used in our automated spectral analysis, we can conclude that while carbon stars represent a completely distinct population, weak CN stars do not. Instead, weak CN stars seem to be better described as a well-defined tail of the "others" population. Plots of spectral metrics consistently show a tight grouping and distinct separation of the carbon star population, with weak CN stars inhabiting a region that is also well-defined but not distinctly separated from "other" stars. However, the fact that the weak CN stars were separated from the "others" using the classifier indicates that this is a significant population of interest. Further investigation into the physical properties of this mysterious population of stars would be instrumental in confirming or denying this possibility.

Future Work

1. Work to remove the human bias by implementing a Support Vector Machine (supervised machine learning algorithm)
2. Use techniques to separate the W structure from the fine-scale fluctuations to see what metrics help best in identification
3. Once an objective, machine-classified sample is created, attempt to obtain a physical understanding of the weak CN phase of stellar evolution: create stellar models, determine chemical composition based on spectral features, and explore other theories behind the empirical data.

Literature Cited

Hamren, Katherine M., et al. "A Spectroscopic And Photometric Exploration Of The C/M Ratio In The Disk Of M31." *The Astrophysical Journal*, vol. 810, no. 1, 2015, p. 60.

Kamath, Anika., et al., "A Mysterious Population of Stars With "Weak CN" Absorption in the Disk of M31." *American Astronomical Society Abstract*, 2017.

Kamath, Anika., "A Mysterious Population of Stars With "Weak CN" Absorption in the Disk of M31." *Siemens Paper*, 2016.

Note: Literature cited above and data used has been obtained and supplemented by private communication with Anika Kamath and Puragra GuhaThakurta in 2018.

Acknowledgments

We thank Anika Kamath, Katie Hamren, Jon Hays, Atmika Sarukkai, Alyssa Sales, Sumedh Guha, Alexandra Masegian, Ary Maheshwari, Allison Chang, and Antara Bhattacharya for their work in the early stages. Suhas Kotha participated in this research under the auspices of the Science Internship Program (SIP) at UC Santa Cruz. This research was funded in part by the National Science Foundation and the National Aeronautics and Space Administration