Summer 2020 Project Write-up: Searching for Laser Technosignatures in APF Spectra

Anna Zuckerman
Brown University, BSRC Breakthrough Listen
(Dated: August 15, 2020)

Laser emission provides a potentially powerful method of interstellar communication. Past works have presented searches for technosignatures in the form of laser signals in stellar spectra. In this project, a new method of searching for laser technosignatures is developed and preliminarily implemented using Automated Planet Finder (APF) spectra of a variety of targets. In addition, properties are derived for a catalog of stars using the empirical Specmatch-Emp algorithm (Yee et al. 2017) on spectra from APF observations, and calibrated against a library of previously published values. Various other intermediate data products, including a set of residuals between the target and matched spectra, a set of APF blaze functions derived from the spectra of B-stars, and a set of normalized, deblazed, and registered APF spectra are also described. This write-up should serve as documentation of the progress made this summer, in the case that another researcher takes on this project in the future. It could also be useful in preparing a paper based on future work on this project.

I. INTRODUCTION AND BACKGROUND

Many SETI searches focus on analyzing radio frequencies for unusual signals, while fewer searches have been conducted at optical wavelengths (for example, Tellis and Marcy 2017, Lipman et al. 2019). The potential of optical lasers to provide a powerful form of interstellar communication was first proposed by Townes and Schwartz (1961). Laser technosignatures offer a few clear advantages from the perspectives both of information transmission and of detectability: high intensity in a narrow bandwidth means a laser can outshine a host star in a narrow region (Lipman 2019); lasers are energy efficient and present a relatively low risk of eavesdropping (Tellis and Marcy 2017); extinction in the ISM is lower at optical (compared with shorter) wavelengths (Rather 1991); and less noise is present in the optical (compared with radio) regions both from Earth and cosmic interference sources (Ross Kingsley 2011). In addition, it has been shown that even under the limitations of current human technology, transmission and detection of laser signals over interstellar distances would be possible (Tellis 2017, Lipman 2017, Clark and Cahoy 2018).

Past laser searches have analyzed stellar spectra (in either raw or reduced form) directly for laser lines. For example, Tellis and Marcy (2017) focus on FGKM stars from Keck/HIRES observations. This work presents a new method for identifying laser emission. The goal of this work is to conduct a laser line search by matching a target APF spectrum to a catalog of known spectra, and subtracting away the linear combination of best matching spectra to produce a residual. This residual is then free of any features present in both the target spectrum and matched spectra, such as stellar absorption features, which might otherwise obscure the desired signals (Nunez 2019). Searching the residual, rather than the target directly, will provide a more sensitive laser signal search than has been conducted in the past.

Yee et al. 2017 describe the model SpecMatch-Emp. This model takes as input a target stellar spectrum, and matches it to a library of spectra which correspond to stars with known properties (temperature, radius, metallicity, etc.). It then makes a linear combination of the closest matching library spectra, with coefficients that represent how close of a match each spectrum is. The corresponding properties of the closest matching spectra are then weighted by the set of coefficients in order to derive a set of properties for the target star. Yee et al. performed this algorithm on 404 stars using Keck/Hires spectra.

This project built on the work of several past Breakthrough Listen interns. Cameron Nunez took the SpecMatch-Emp model and wrote a new script to allow it to run on APF target spectra instead of Keck/Hires spectra. Several important modifications were made to C. Nunez's script.

A calibration was then performed of the APF version of SpecMatch-Emp by comparing the results to those published by Yee et al. First, the set of stars which were analyzed by Yee et al. and also had APF observations in the Breakthrough Listen database was determined. Note that only stars which returned a HIP name in the Simbad database were used, so some stars were not included in the comparison. A signal-to-noise ratio (SNR) for each set of observations (night of observing) corresponding to each star was then calculated, and the set with the highest SNR for each star selected. SpecMatch-Emp was then run on these 101 stars in order to determine how close the SpecMatch-Emp derived results were to the library values.

Data products produced during this project which might be useful for future work include:

1. Derived properties from SpecMatch-Emp for each of the 101 target stars, using the highest SNR set of APF spectra for each.

2. APF blaze functions derived from nine rapidly rotating B-stars for the orders of interest of SpecMatch-Emp (order 30 through 51).

3. Normalized and deblazed spectra corresponding to the highest SNR set of observations of each of the 101 target stars.

4. A residual between the input target spectrum and the linear combination of best match spectra for each of the 101 target stars.

Data products (3) and (4) are output in the same fits file for each star. See Section V for locations of these data products.

The second part of the project focused on searching the residuals for potential laser features. This project began the process of implementing and refining an algorithm to search the residuals and identify peaks with certain characteristics, based on the algorithm described in David Lipman's 2017 project write-up. First, peaks from representative residuals were plotted to gain an intuition for the usual and unusual characteristics of peaks, then a residual was injected with simulated gaussians in an attempt to recover the injected signals.

II. METHODS

In this section I describe the scripts, inputs and output files, and data that I used (in more detail that would usually be necessary, but which might be helpful to read before picking up the project).

The SpecMatch-Emp model is described fully in Yee et al. 2017. The original algorithm takes a Keck/Hires spectrum as input, and compares it against a library of spectra which correspond to stars with 'known' (ie. determined in some experimental way, rather than by modeling) properties. It then makes a linear combination of the five best matching spectra, with weights corresponding to the closeness of the match. The properties of the best matching stars are weighted by the coefficients of the linear combination, to produce derived properties for the target star. See https://specmatch-emp.readthedocs.io and https://github.com/samuelyeewl/specmatch-emp for further details. The model is located at the BL datacenter in the directory /home/mattl/.specmatchemp but can also be obtained from github.

Past BL intern Cameron Nunez wrote a Python script to allow SpecMatch-Emp to be run on APF spectra. This script, smemp.py, first processes the spectra (combines multiple spectra for one star, normalizes, registers, and deblazes the spectrum, resamples into the given wavelength scale), and then calls the methods of SpecMatch-Emp (including shift, match, and lincomb). See https://github.com/stevecroft/bl-interns/tree/master/cameronn for details of C. Nunez's work.

Past intern Jackie Telson wrote a deblazing function that returns a normalized, registered, and deblazed spectrum given an input spectrum. See https://github.com/stevecroft/bl-interns/tree/master/jackietel for details of J. Telson's work.

This work began with a determination of the list of stars which both have at least one APF spectrum (as listed in all_apf_non_i2.csv), and were included in the library of known spectra in Yee et al. This was complicated by the fact that many stars are listed under multiple names both within and between the two catalogs, and some stars have multiple spectra listed by different names in the APF catalog. See the script find_apf _and_yee _overlap.ipynb. This produced the list of 101 stars which were then used in the rest of the project.

Next, the signal-to-noise ratio was calculated for each set of observations of each star (a set of observations is marked by the same initial four letters in the APF fits file name of the spectrum, and corresponds to one night of observations). This was done by taking the square root of the median of the spectral order containing the Na-D lines, or spectral order 45 in these APF spectra. Where a set of observations included multiple spectra, the spectra were summed element-wise prior to the SNR calculation. Then, only the set of observations with the highest SNR was used for each star.

Before running the spectra through SpecMatch-Emp, certain modifications were made to smemp.py. The most minor involved changing the way system output is sent to a null file in the rescale.py function (potentially due to incompatibility with python 3.7), changing python scripts to python notebooks for ease of use in Jupyterlab, adding imports of pandas and import-ipynb, adding a log file (smemp_output.txt), and changing the results file name (specmatch_results.csv). Other modifications are described below.

The test to check that all spectra in each input directory correspond to the same star was also modified, and was moved to a separate script (which should be run prior to running smemp.py, or alternatively the logic it contains could be copied to smemp.py in place of the original test). This was done because the original test relied on fits files corresponding to the same star to all have the same name in the header. This was not always the case due differences in labeling (upper/lowercase, different names used for same star, occasional mislabeling of files, etc). The new test script (run in check_file_labeling.ipynb) has methods to both check that the header name in each fits file is the same as the name listed for that file in all_apf_non_i2.csv to check for mislabeling, and to check that all spectra in each subdirectory in fact correspond to the same star using the coordinates of the stars as listed in the Simbad catalog. The coordinates as listed in the fits file headers could not be used as they differ slightly even for the same star.

Additionally, smemp.py was modified to properly process a directory of spectra containing both individual spectra files, and subdirectories containing multiple spectra of a single star. Subdirectories should be labeled by the HIP name of the star (ie.HIPXXXX_spectra),

and only contain spectra of one star. Currently, the input spectra are contained in the directory structure APF_spectra/apf_spectra_highest_SNR at the BL datacenter, with the directory APF_spectra being located in the same directory as the script smemp_multifile.ipynb. They can also be found in the github repository described in section V.

The deblazing function was changed from C. Nunez's deblaze.py, written in R (which uses the AFS algorithm from Xu et al. 2019), to instead use the python deblazing script written by J. Telson. Initially, minor modifications were made to this script in order to use a spectrum instead of a fits file as input, and to input the order used for the deblazing. However, this function was written to work for the order containing the H-alpha features (order 53), and the masking of that feature led to a poor blaze function for any other order. A first attempt at fixing this was to modify the existing function in three ways: to smooth the blaze function further by using a wider median fit window, to force the ends of the blaze function to align with those of the spectrum to prevent spikes at the ends as a results of deblazing, and to manually mask out the Na-D and Mg-D features in orders 45 and 34 respectively. However, a better blaze function was found to result from a different method. Spectra of rapidly rotating B-stars (which have few spectral lines) were combined and smoothed to produce the blaze function for each order. This new method is used in bstar_deblaze.ipynb, which is called by smemp_multifile.ipynb. See Figure 1.

After the call to the deblazing function, an option to remove cosmic rays from the spectrum was added. Set remove_cosmic_rays to False for any run other than the calibration, as this may remove actual signals. The cosmic ray removal function is very simple currently and could be improved in the future.

Finally, a modification was made in order to return the residual between the target spectrum and the linear combination of best matched spectra. The normalized, deblazed target spectrum and registered wavelength scale are also stored in the same fits file. This is saved for each input star. In addition, for the most recent run the entire SpecMatch object created for each star has also been saved. This is to facilitate future analysis without a new run through all stars (which can take 24 hours). See the Results directory in the github repository described in Section V.

An attempt has been made to return the number of pixels by which each spectrum has been shifted during the call to sm.shift(), in order to identify and reject known atmospheric lines in the laser line search algorithm, and this has been saved in the Supporting_files section of the same repository.

The new verison of smemp.py is called smemp_multifile.ipynb, and can be found in the Spectmatch_scripts directory in the same repository.

The laser search algorithm was implemented following the description of David Lipman's algorithm in his 2017 project write-up, with some modification of the thresholds and search criteria to account for the input of normalized, deblazed, residual spectra rather than target spectra. This algorithm is a work in progress, and results are preliminary. The primary goal of the algorithm is to identify peaks that meet the following characteristics:

1. The feature is narrow in wavelength yet with high intensity. The usefulness of lasers for interstellar communication stems from their ability to produce signals of very high intensity in very narrow bands.

2. The feature has a gaussian-like shape. Any signal that traveled through the telescope optics must follow the point-spread function the telescope (which is a function of telescope configuration and local seeing), otherwise the signal is likely a cosmic ray or faulty pixel.

3. The feature does not correspond to known atmospheric or stellar emission lines.

These characteristics are addressed via four tests in the algorithm. The first criteria imposed on candidate signals is a height minimum. This first tier of candidates is identified as indices within the residual such that the value of the residual at that index and the two adjacent indices is greater than three standard deviations above the median. Next, an initial test of width rejects all candidate indices whose values are less than that of any of the two adjacent pixels on each side. This test helps remove 'compound' peaks (for example, very close cosmic rays) that are individually only a few pixels wide. Candidates in this second tier are then modeled as guassians of the same height above the median and Full-Width at Half-Maximum (FWHM) as the candidate peak. Another test rejects all candidates with FWHM greater than 10 pixels or less than 3 pixels (20 and 6 pixels in the interpolated gaussian) to reject peaks too wide or too narrow to be lasers and form the third tier of candidates. Note that pixels in the interpolated peak and gaussian fit represent 1/4 the wavelength range of the original pixels. A fourth and final tier of candidates is formed by requiring the normalized root mean squared deviation between the identified peak and simulated Gaussian to be less than 0.15. This algorithm is located in the github repository described in Section V.

III. RESULTS AND ANALYSIS

A. Calibration of SpecMatch-Emp for APF Spectra

The model was run on the 101 stars in the calibration target set (those both in the Yee et al. library and with at least one APF observation on the BL database) using the script smemp_multifile.ipynb. The resulting properties were then compared to the catalog values as shown in Figures 2 and 3. No clear trend in the magnitude of these differences as a function of input spectrum SNR

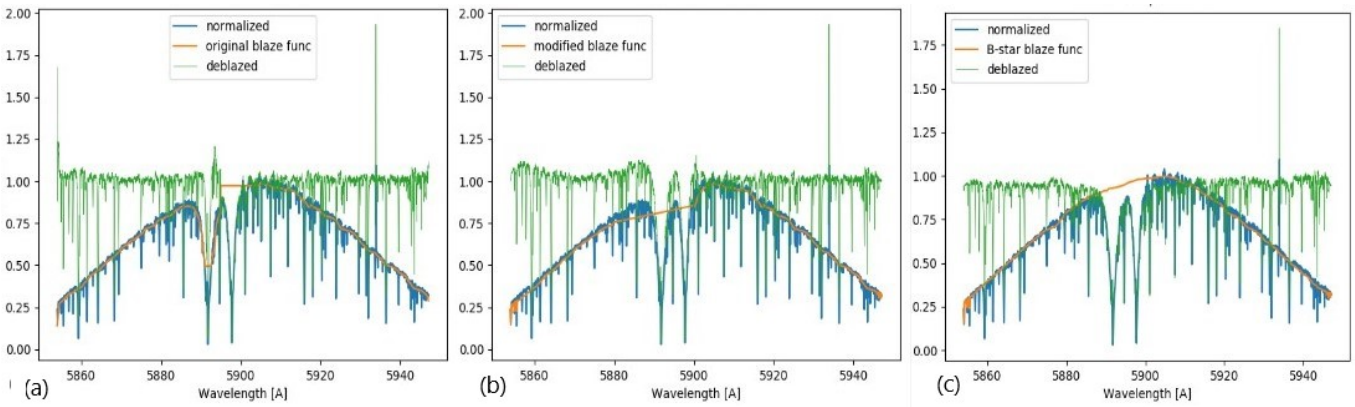

FIG. 1. Comparison of deblazing methods on order 45 of HIP13642. (a) The initial function, which uses a median filter to smooth the spectrum and masks out the H-alpha features, is not designed to deblaze any order other than that containing the H-alpha features. (b) A deblazing function with a wider median filter window and masking of a greater list of broad features performs somewhat better. (c) A new deblazing method which derives the blaze functions for each order from the spectra of rapidly rotating B-stars performs well for most orders.

is apparent, indicating that in general the algorithm is accurate even at low values of SNR.

The RMS of the differences between the derived and catalog values is shown in Table I. These values can be compared to Table 3 in Yee et al. A division was apparent in the differences between derived and catalog values at both $R \approx 1.0R_*$ and $T \approx 4500K$, with the RMS for both temperature and radius being greater at greater values of temperature and radius. The RMS of the differences in metallicity was greater at lower temperatures and greater radii. This division may be due in part to the fact that the library stellar properties are less well constrained for hotter stars due to their relative lack of spectral features.

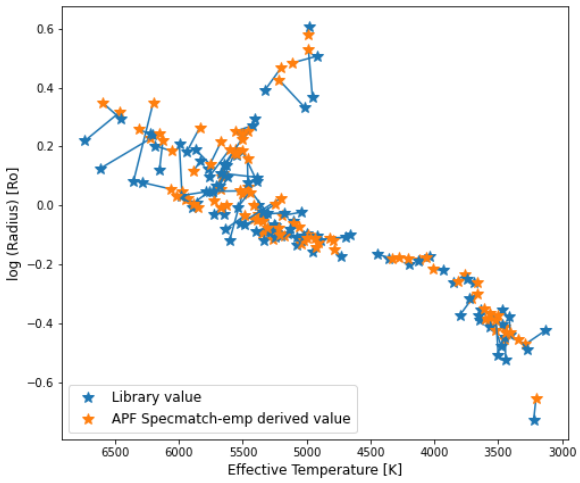

FIG. 2. A visualization of the comparison between library and derived stellar properties.

	$\sigma(\Delta T_{eff})$ K	$\sigma(\Delta R_*/R_*)$ %	$\sigma([Fe/H])$ dex
All stars	144.33	16.05	0.091
$T_{eff} < 4500K$	71.48	10.74	0.111
$T_{eff} \geq 4500K$	125.82	17.52	0.083
$R < 1.0R_o$	89.11	10.52	0.083
$R \geq 1.0R_o$	145.72	22.16	0.102

TABLE I. RMS of differences between catalog and SpecMatch-Emp derived properties for 101 stars with APF spectra. Two regions are apparent in both temperature (above and below $T = 4500K$) and radius (above and below $R_* = 1.0R_o$

B. Laser Line Search

A preliminary laser line search algorithm as been implemented based on the work of David Lipman. Future work should improve upon this algorithm.

1. *Signal Injection and Recovery*

An initial test of the search algorithm was performed by injection of 100 simulated gaussian signals of FWHM between 1 and 10 pixels and height between 0.1 and 1 (arbitrary normalized intensity units) above the median into a sample residual. This was performed by inserting 10 gaussians at a time into a sample residual, and searching the residual to attempt to recover the gaussians.

No gaussians of FWHM = 1 pixel were recovered. This is consistent with expectation, because features of 1 pixel width should be rejected as likely cosmic rays or faulty

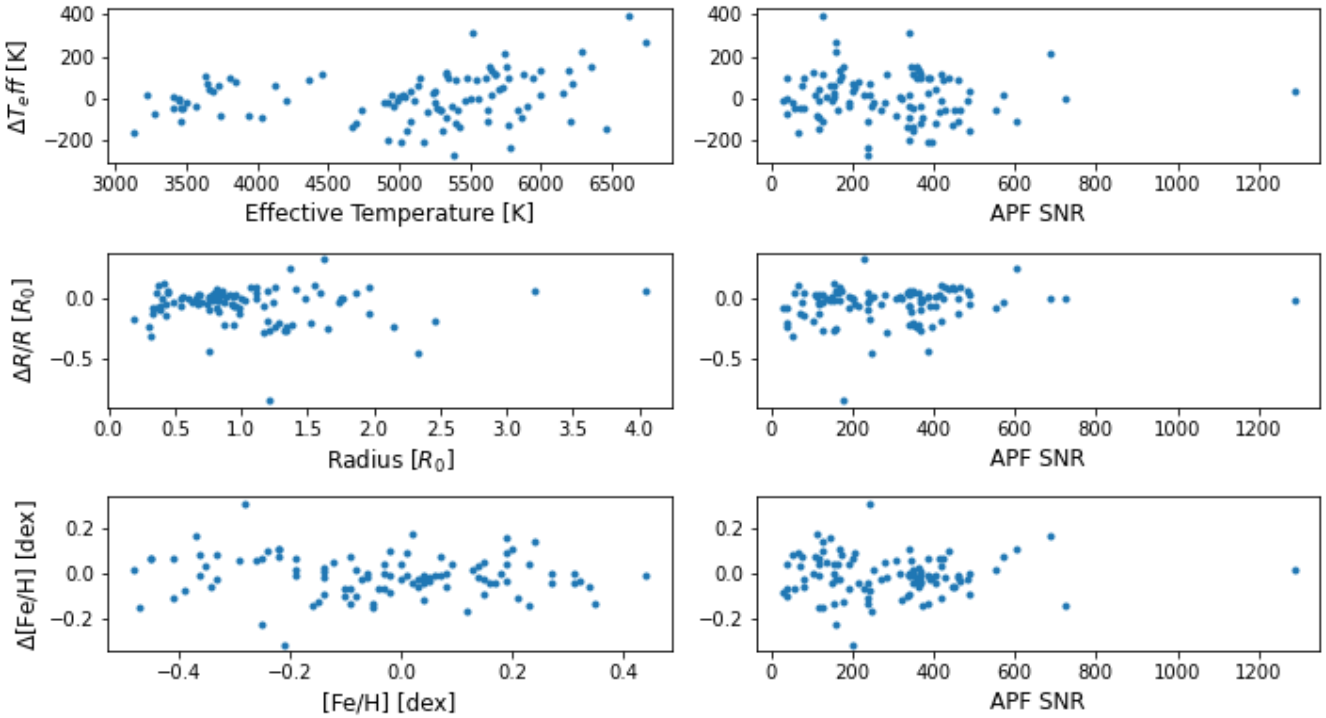

FIG. 3. Differences between library and derived value for effective temperature, radius, and metallicity as a function of library values (right), and as a function of the SNR of the target spectrum (left). No clear trend in the scatter is evident with SNR.

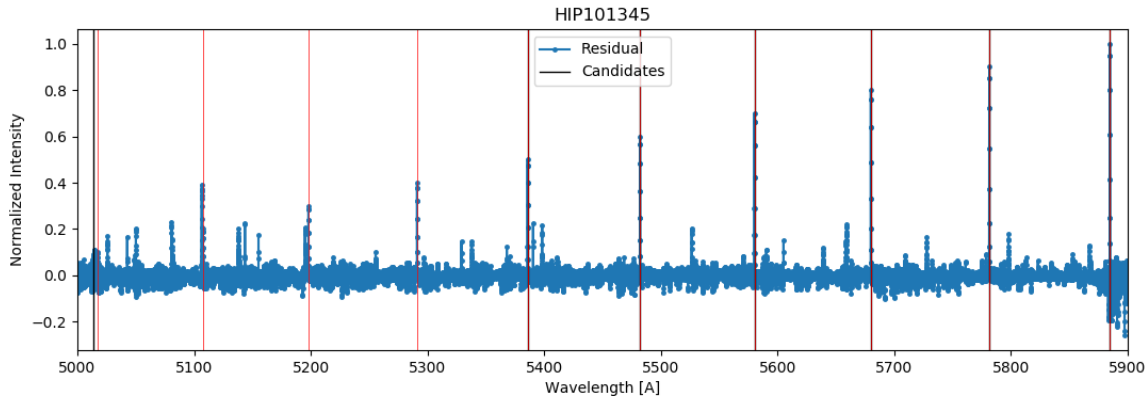

FIG. 4. Example residual with 10 injected gaussians of heights between 0.1 and 1, all with FWHM of 4 pixels. Red vertical lines mark the injected signals. Black vertical lines mark recovered signals. Note that some guassians were too short to be recovered, and some real candidate signals from the residual were caught.

pixels.

For gaussians of FWHM = 2 pixels, all but the gaussians of height 0.1 and 0.2 were included in the first and second tier of candidate signals (see above), and none were included in subsequent tiers. This is consistent with signals of this width being too narrow to have traversed the telescope optics. The gaussians of heights 0.1 and 0.2 did not meet the intensity requirement on pixels adjacent to the index pixel.

For gaussians of FWHM = 3 pixels, all were included in the first, second, and third tiers of candidates, due to all having sufficient height and width, but none passed the final test of 'gaussian-ness'. This highlights a current issue with the algorithm: shorter and narrower peaks (even true gaussians) are not fit well by the gaussian. Typically, the algorithm's gaussian fit is too wide. This

	Injected gaussians passing test			
FWHM [pixels]	Height of peak	Height of neighboring points	Width of peak	Closeness of gaussian fit
1				
2	h >0.2	h >0.2	h >0.2	
3	all	all	all	
4	all	all	all	all
5	all	all	all	h >0.4
6	all	all	all	
7	all	all	all	
>7	all	all		

TABLE II. Results of injection and attempted recovery of simulated gaussians. Ideally, all simulated gaussians of width between 3 and 10 pixels would pass through to the last tier of candidates.

could be improved by using a method such as scipy's interpolate module to find a best fitting gaussian to each peak, rather than forcing the fit gaussian to have the determined FWHM and height of the peak.

For gaussians of FWHM = 4 pixels, all gaussians were caught in all tiers of candidates.

For gaussians of FWHM = 5 pixels, all gaussians reached the third tier of candidates, but those with heights of 0.1 through 0.4 did not reach the final tier due to the issue with gaussian fitting described above.

All gaussians of FWHM = 6 and 7 pixels were caught as candidates up to tier three and then rejected due to the same fitting issue.

All gaussians of FWHM > 7 pixels were caught as candidates up to tier two and then rejected due to width (the fit being wider than the injected signal).

2. Search of Sample Residuals

Each of 10 sample residuals was searched for laser lines, and candidate signals identified (and rejected) in tiers as described above, in order to understand the current state of the algorithm. 1633 peaks were identified in the first tier of candidates, 367 in the second, 122 in the third, and 12 in the fourth. The method and reason for rejection of candidates at each tier is describe above. Candidates reaching the fourth tier of classification were then examined by eye.

C. Data Products

Data products produced for this project include:

1. A table of SpecMatch-Emp derived properties for 101 stars with APF spectra, based on the highest SNR set of observations of each star. The properties can be found in the results directory of the github repository.
2. A signal-to-noise ratio for each of the 101 stars, which can be found in the Calibration directory of the github repository.
3. A residual between the target spectrum and linear combination of best match spectra for 101 stars with APF spectra, based on the highest SNR set of observations of each star. Residuals are stored as fits file where the PrimaryHDU is the target spectrum, the first ImageHDU is the residual, and the second ImageHDU is the registered wavelength scale. The header is a copy of the header of one of the input spectra for each star, which the addition of the keyword 'RESID' set to 'Yes', 'SPECT' set to the names of the input spectra used for that star, and 'NDR' (indicating normalized, deblaed, and registered) set to 'YES'. These are located at the BL datacenter in /home/azuckerman/APF_spectra/NDRR.
4. A normalized, deblazed spectrum and registered wavelength solution for each of the 5869 APF spectra from the directory /mnt_blpd0/datax/apf at the BL datacenter. The PrimaryHDU is the normalized, deblazed spectrum, and the ImageHDu contains the shifted wavelength scale. Note that the headers in the files copy the original header with the addition of the keyword 'NRD' set to 'Yes', with the comment 'Normalized, registered, deblazed'. These are located at the BL datacenter in /home/azuckerman/APF_spectra/NDR.

IV. FUTURE WORK

A. Current 'To-do List'

The next steps in this project include:

1. Improve the residual search algorithm by updating the gaussian fitting of signals, automatically rejecting known atmospheric or stellar emission lines (try using the saved pixel shift to find shifted lines, though this may not yet return the correct value), and incorporating a better treatment of very high intensity signals (these would currently be too wide if reasonably gaussian). An improved method of fitting gaussians to the peaks is currently partially implemented in the search script.
2. Run the search algorithm on all 101 stars in the calibration target set to further refine useful threshold values for each test.
3. Run the algorithm on new residuals not in the calibration set.
4. Attempt to reproduce past search results, for example Tellis and Marcy 2015 and 2017, Lipman 2019.
5. Derive limits on detectability.

B. Extensions

Possible future extensions of this work include using machine learning techniques to search for very unusual signals, extending the analysis to look for other types of optical signals, and using repeat observations of the same star to add a time component to the analysis.

V. LOCATIONS OF DATA, CODE, AND RESULTS

All scripts and files mentioned above are located on github at:
https://github.com/annazuckerman/specmatch-emp-apf and also on the BL datacenter at /home/azuckerman. The former is recommended because it includes documentation listing where each file can be found in the repository. The locations of data and results files are also noted in the documentation of this repository.

As mentioned above, the model SpecMatch-Emp itself can be found at

https://github.com/samuelyeewl/specmatch-emp

and at the BL datacenter in the directory /home/mattl/.specmatchemp.

Please do not hesitate to contact me at anna_zuckerman@brown.edu if there is any difficulty in locating files, or if I can help answer any questions about picking up this project.

[1] James R. Clark and Kerri Cahoy. Optical Detection of Lasers with Near-term Technology at Interstellar Distances. Astrophys. J., 867(2):97, November 2018.

[2] Michael Hippke. Interstellar communication: Short pulse duration limits of optical SETI. Journal of Astrophysics and Astronomy, 39(6):74, December 2018.

[3] David Lipman, Howard Isaacson, Andrew P. V. Siemion, Matt Lebofsky, Danny C. Price, David MacMahon, Steve Croft, David DeBoer, Jack Hickish, Dan Werthimer, Greg Hellbourg, J. Emilio Enriquez, and Nectaria Gizani. The Breakthrough Listen Search for Intelligent Life: Searching Boyajian's Star for Laser Line Emission. , 131(997):034202, March 2019.

[4] Shin-ya Narusawa, Tatusya Aota, and Ryo Kishimoto. Which colors would extraterrestrial civilizations use to transmit signals?: The ;magic wavelengths; for optical SETI. , 60:61–64, April 2018.

[5] John D. G. Rather. Lasers revisited - Their superior utility for interstellar beacons, communications, and travel. Journal of the British Interplanetary Society, 44:385–392, August 1991.

[6] Monte Ross and Stuart Kingsley. Optical SETI: Moving Toward the Light, page 147. 2011.

[7] Nathaniel K. Tellis and Geoffrey W. Marcy. A Search for Optical Laser Emission Using Keck HIRES. , 127(952):540, June 2015.

[8] Nathaniel K. Tellis and Geoffrey W. Marcy. A Search for Laser Emission with Megawatt Thresholds from 5600 FGKM Stars. , 153(6):251, June 2017.

[9] Samuel W. Yee, Erik A. Petigura, and Kaspar von Braun. Precision stellar characterization of fgkm stars using an empirical spectral library. The Astrophysical Journal, 836(1):77, Feb 2017.