Exploring New RFI Excision Techniques

Efrain Rodriguez[1] and Dr. Ryan S.Lynch[2]

[1] University of Texas at El Paso , 500 West University Avenue, El Paso, TX 79968

[2] Green Bank Observatory, 155 Observatory Road, Green Bank, WV 24944

INTRODUCTION

In the field of radio astronomy there has been a growing concern in regards to radio frequency interference (RFI) affecting astronomer's scientific data and its negative effects on ground based observatories. Not only will the human dependence on technology increase, but also the presence of RFI emission in radio astronomy observations all over the world.There are ways to promote a more suitable environment for radio observations, one of them being through RFI avoidance. This involves having a ground based radio telescope be operated in a remote and low population density area. The effects of RFI will not be completely out of the observation, but they can be greatly reduced. A famous example would be the Green Bank Observatory in Green bank, West Virginia which is located within the National Radio QuietZone (NRQZ), which occupies an enclosed area of approximately 13,000 miles. Other techniques of RFI excision are also in use today but do not provide complete protection either. These include methods that automatically identify RFI and look for samples of the data that are statistical outliers to some degree (e.g. maximum mean power). Other techniques involve the use of data analysis software tools that flag data that has been corrupted by RFI, which can then be ignored in future observations. There are a few drawbacks to these techniques, the first being that these RFI excision tools are applied after the data has been integrated and averaged for some period of time (1-10 seconds). RFI can have a spike that lasts only a fraction of a second, which contaminates the entire integration. The second drawback to using these techniques would be the simplicity in the statistical analysis of the data, which can be further improved on. Previous research towards exploring new RFI excision techniques has been developed into a machine learning algorithm that detects and removes RFI. The potential use for this algorithm is for it to be utilized during real-time observations. This algorithm provides some unique advantages over other techniques, the advantage being that it would identify RFI before the data has been integrated and averaged, which would reduce the risk of corrupting the data entirely. But before this can become an applicable tool for radio astronomers, it is important to further analyze its capabilities and verify that the quality of the data is not being corrupted. To explain this further, the algorithm simultaneously labels pixels as containing RFI and/or single pulse pulsar/FRB data (1). RFI containing pixels are replaced into either a zero or a statistical value that corresponds to the signal being detected. One downside of this would be that the replacement would be permanent and could not be undone, meaning that it could also harm the data in ways that would be irreversible, making it a rather high risk tool for astronomers if not tested properly. In order to test the robustness of this algorithm, we used GUPPI raw voltage baseband data which is useful since it provides us with the most important properties of our signal, such as frequency, amplitude, phase and polarization. This will allow us to compare our data in a more detailed way and determine whether or not the algorithm is improving the quality of our data. In this paper we will discuss the effectiveness of this RFI excision algorithm by comparing two types of data sets,unmitigated and mitigated. Ultimately, we compared some fundamental pulsar astrophysical properties within these data sets by using professional grade pulsar analysis software. These astrophysical properties include dispersion measure (DM), pulsar period and flux density to name a few.

APPROACH

The data used for the comparison analysis was recorded during the summer of 2019 using the Green BankTelescope (GBT), for a total recording of 20 TB of raw voltage baseband data which was then processed using the RFI excision algorithm. As mentioned earlier, in order to effectively compare the amount of RFI present in our recorded GUPPI data, the data was separated into two categories. These two categories being unmitigated which has no RFI excision algorithm applied and mitigated with the RFI excision algorithm applied. By separating the data into these two categories we are able to inspect the astrophysical quantities of interest more closely and accurately. Ultimately asking the question, how effective is the RFI excision algorithm in masking RFI emission compared to other types of RFI excision techniques? The data used for the comparison analysis was GUPPI raw baseband voltage data. GUPPI stands for ”Green Bank Ultimate Pulsar Processing Instrument” [2]. GUPPI is a backend that digitizes the signal being observed. One of the techniques we want to apply uses the phase of the signal which is the lowest-level recorded, this property enables us to test the algorithm more rigorously.

COMPARISON ANALYSIS

One of the first characteristics of the RFI excision algorithm that was analyzed, was the RFI masking of the data. In the unmitigated data set there was an approximate RFI masking of 37 %. On the mitigated data set there was a slight increase in the data being masked as RFI, at approximately 43 %. This provided some proof of the basic functioning of the algorithm in the flagging of data as RFI. The next step was to determine if the algorithm was flagging an excessive amount of data, which could in turn harm the quality of the astrophysical qualities we are comparing.

Promising Results

Spectrum Comparison

One of the major aspects of the data that was analysed for comparison differences would be the different spectra of the data sets. The results were promising since they provided a more robust outlook as to how the algorithm was enabling the data to be more clean of any RFI presence. The data was processed using the psrfits package in python. The unmitigated data set had a distinct high power intensity at the lower frequency channels. The mitigated data set on the other hand has a significant decrease in the power at those specific frequency channels. This provided a more solid understanding of how the data was affected by the RFI excision algorithm. In this particular case the data was affected as we predicted

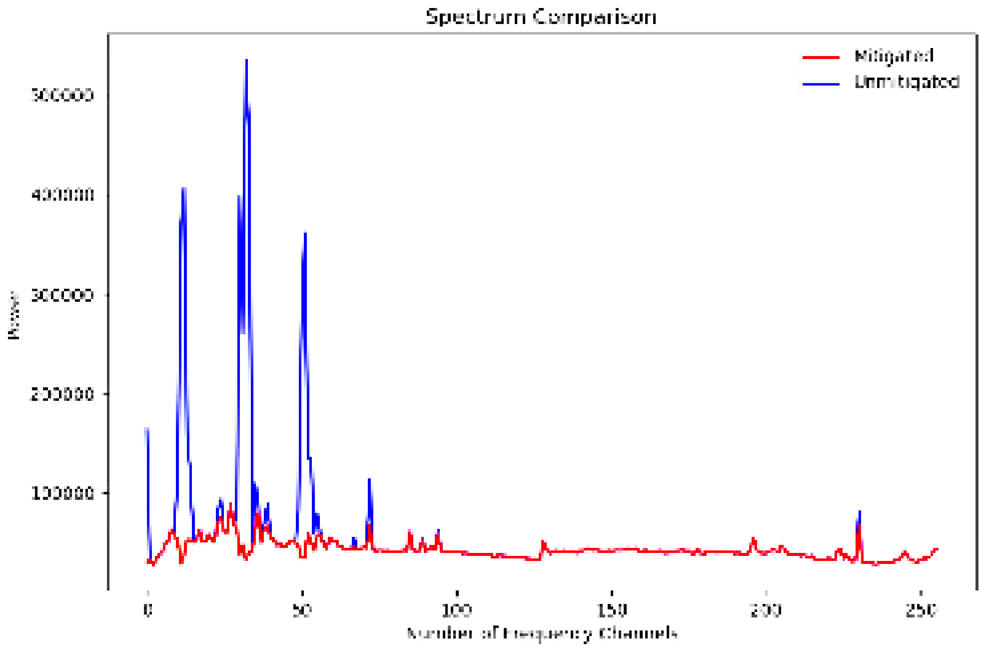

Figure 1 *: Spectrum Comparison between the two different data sets.*

Spectrum Percent Difference

The two different spectra were further analysed by comparing their percent difference, this would enable us to understand whether or not the data was affected in a negative way. In other words this specific analysis would enable us to determine how much the data was altered, in this case how much did the data change from the unmitigated data set to the mitigated data set.

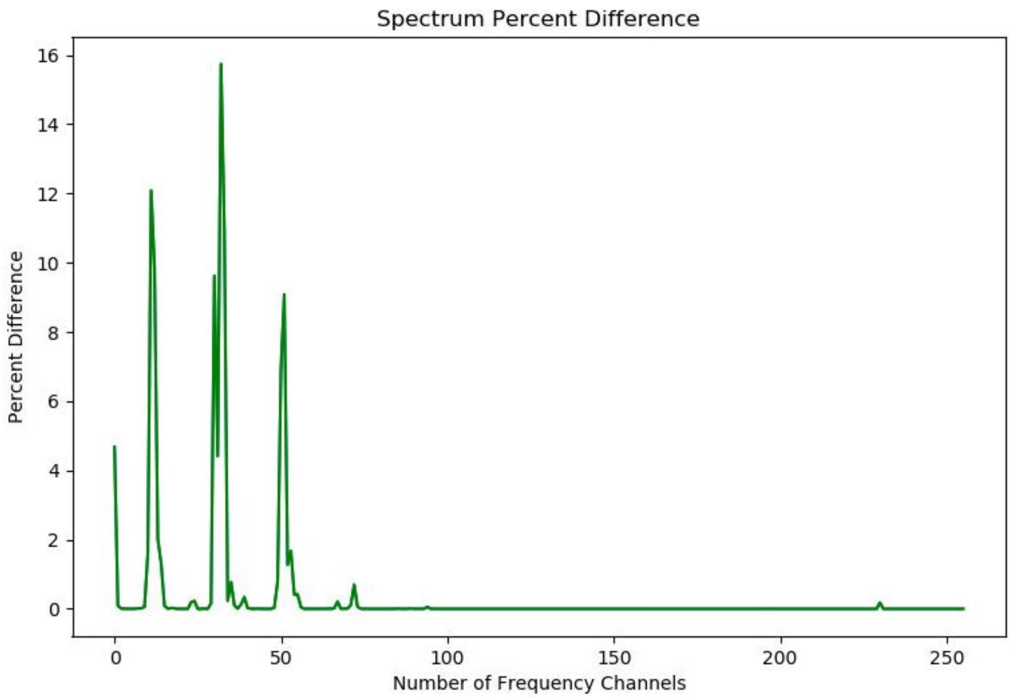

Figure 2: *Spectrum Percent difference.*

Fourier Spectrum

One of the areas that exposed another promising result in this data comparison analysis was the exploration of the fourier spectrum between the different data sets. The images presented in this report represent a small portion of the entire fourier spectrum obtained from the different data sets. The fourier spectrums presented offer a clear view as to how the overall intensity or power (y-axis) vary at certain frequencies (x-axis) between the unmitigated and mitigated dataset. This distinction can be categorized by the reduction in intensity at some RFI intense frequencies. One interesting observation would be at around 900 Hz, where there is a significant peak in the unmitigated dataset and a significantly reduced intensity at the same frequency but now in the mitigated dataset. This proves very beneficial in our analysis since it provides a more solid understanding in how the algorithm is affecting our data. In this particular case it is a positive effect in our data since there is a cleaner fourier spectrum in our mitigated data. There are still more avenues to explore since the algorithm is capable of being customized in its RFI replacing-pixel capabilities, which is something that will be explored in the future.

Fourier Spectrum Comparison

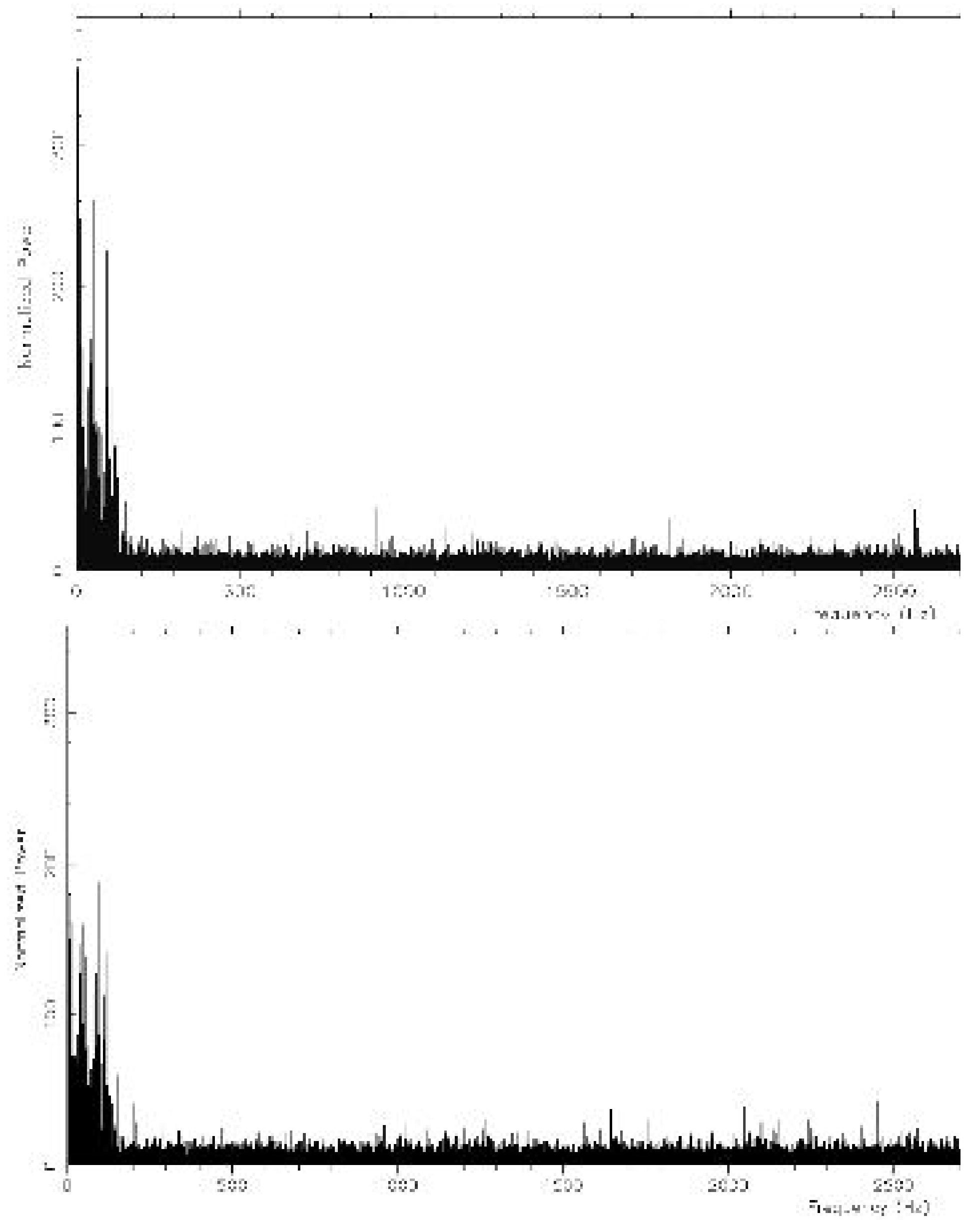

Figure 3: *Unmitigated fourier spectrum (above) and the Mitigated fourier spectrum (below), processed using the PRESTO package. The y-axis labeled as the Normalized power (unit less), and the x-axis labeled as Frequency (Hz).*

Pulsar Properties Comparison

One of the main areas of interest in the data sets was the fundamental astrophysical properties of the Pulsar data recorded. The properties compared in this analysis would be the pulsar spin period in (ms), the dispersion measure (DM) and the signal to noise ratio (S/N). These properties were obtained from the raw data by using accelsearch, which is part of the PRESTO package. Accelsearch works by making a blind search, without knowing any previous parameters of the pulsar and it gives the most significant candidates of the data. In our analysis we only used the top candidates accelsearch obtained. One of the first properties that became fundamental in our analysis was the pulsar spin period. As seen in the table below, there is the clear result that both the data sets had identical periods, which is a promising result since the data was not altered negatively. On a negative note, the dispersion measure was completely off, The pulsar being studied in this case has a true DM of 26.7641[1], which is nothing compared to either result obtained from the unmitigated or mitigated datasets. There could be a few explanations to this phenomenon. For the mitigated dataset the reason why the DM is very incorrect could be due to some of the parameters in the RFI excision algorithm that are causing some effect in the overall statistics of the data. In either case of the dataset, one common issue would be the increased noise and RFI at lower dispersion measures. More exploration in this area and it is something to look more carefully in future work. Lastly, the signal to noise ratio (S/N) was one of the properties that was favorable to our predictions since accelserach reported a candidate with a higher significance in the mitigated dataset, this means that the algorithm influenced the data in sense that we have the true candidate be shown with a higher significance, which is ideal and a minimal improvement front the unmitigated dataset.

ATNF Pulsar Catalogue

Pulsar Name	Period (s)	DM (cm^-3 pc)
B0329+54	*0.714519699726*	*26.7641*

[1] https://www.atnf.csiro.au/research/pulsar/psrcat/

Table 1: *ATNF catalog pulsar properties for pulsar B0329+54.*

Unmitigated vs Mitigated

Pulsar Properties	Unmitigated	Mitigated
Period (ms)	714.513607	714.513607
Dispersion Measure (DM)	39.33	46.73
Best S/N	244.36	249.75

Table 2*: Properties obtained by accelsearch from the PRESTO package, comparing the two different datasets.*

Negative Results

RFI Masks

Unfortunately, the comparison analysis revealed issues with the data after it was processed with the RFI excision algorithm. After the data has been processed with the algorithm we performed an RFI masking technique to remove further RFI contamination from the data. Our intuition was that the unmitigated dataset would have a higher percentage of the data being flagged by rfifind since it was not processed with any rfi excision technique prior to our analysis. The mitigated dataset consequently had a higher percentage of the data be masked by rfifind, which was not expected since the data should be more clean in terms of RFI presence. The unmitigated dataset had 37% of the data be flagged by rfifind and the mitigated dataset had 43% of the data be flagged. There is a potential reason as to why this occurred. As mentioned earlier the algorithm has certain specific parameters that the observer can specify for the data. In our case we had the rfi containing pixels be replaced with zeros. This specific alteration in the data could have caused some anomaly in the statistical nature of the data that caused rfifind to mask the mitigated dataset even more than the unmitigated data. Future work in this area will be performed with a different approach in the parameters of the RFI excision algorithm in order to determine if this was the cause of the amount of data being flagged in the mitigated data.

RFI Masks Comparison

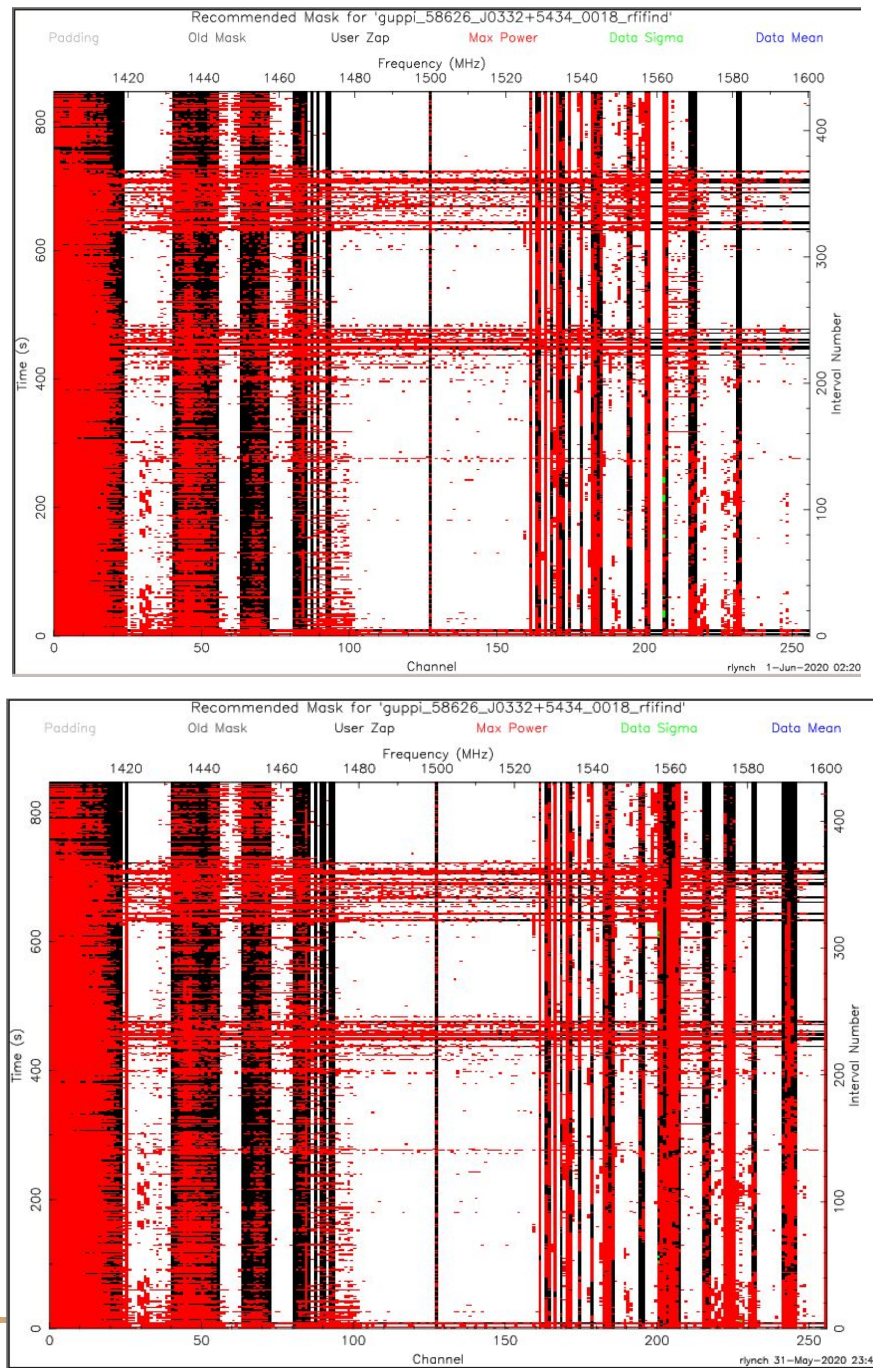

Figure 3: RFI masks using rfifnd from the PRESTO package, unmitigated (top), mitigated (bottom). Percent of data flagged by rfifnd. Unmitigated (37%) Mitigated (43%)

Integrated Pulse Profiles

Further analysis of the data was performed in the area of the pulse structure of the pulsar being analysed. This analysis was performed using psrfits, which gave us extremely close pulse structures of both of the data sets. Nevertheless, in order to be as concise as possible we performed a pulse structure difference where the unmitigated pulse profile was subtracted from the mitigated pulse which unfortunately provided a more serious view in how the data is being altered. In this case it is being altered in a negative way.

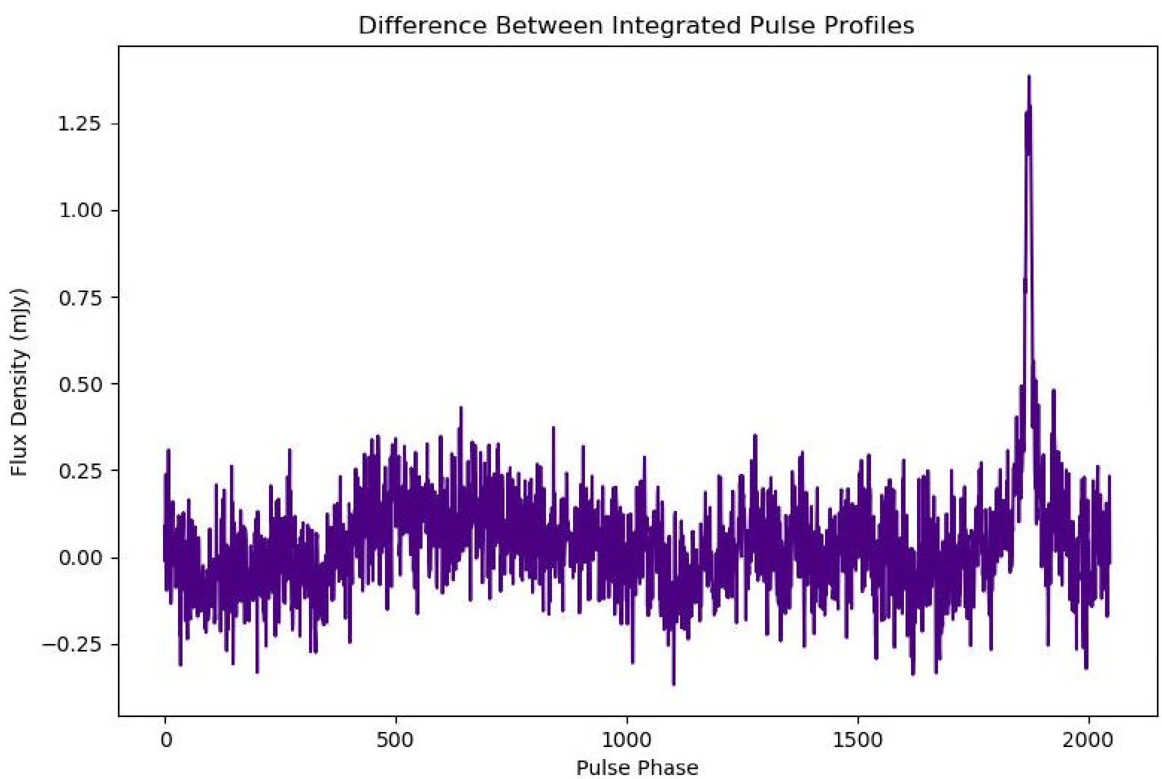

Figure 4: *DIfference between the two integrated pulse profiles visualized by using python and the matplotlib library.*

ACKNOWLEDGEMENTS

This paper and research experience would not have been possible without the incredible patience, enthusiasm, expertise, and mentorship of my mentor Dr. Ryan S.Lynch. His incredible attention to detail and passion for radio astronomy has motivated me to continue my studies and pursue my goals for one day to be part of the Green Bank Observatory scientific staff. Thank you NRAO/GBO for allowing us, all of the summer interns continue with our projects during this difficult time. I enjoyed getting to know everyone from the different sites, and the science lectures offered a tremendous amount of learning. Thank you for remaining patient and understanding. The learning potential was massive and very rewarding, I could not have asked for a better summer program.

REFERENCES

1. Ransom, S. M., Eikenberry, S. S., Middleditch, J. 2002. Fourier Techniques for Very Long Astrophysical Time-Series Analysis. The Astronomical Journal 124,1788–1809.

2. Ransom, S. 2011. PRESTO: PulsaR Exploration and Search TOolkit. Astro-physics Source Code Library.

3. Hotan, A., van Straten, W., Manchester, R. N. (2004). PSRCHIVE and PSRFITS: an open approach to radio pulsar data storage and analysis.

4. Van Straten, W., Bailes, M. 2011. DSPSR: Digital Signal Processing Software for Pulsar Astronomy. Publications of the Astronomical Society of Australia 28,1–14.

5. Hawkins, M. (2019). *High Time-Resolution Radio Frequency Interference Mitigation and Single Pulse Pulsar and FRB Detection using Machine Learning Semantic Segmentation* (pp. 1-12, Rep.). Green Bank, WV.

6. Balser, D., Braatz, J., Frayer, D., Ghigo, F., Kepley, A., Kobelsky, A., . . . Skipper, J. (2020, February 5). Observing With The Green Bank Telescope. Retrieved August 30, 2020, from https://www.gb.nrao.edu/scienceDocs/GBTog.pdf