UNMASKING ECHOES: CLUTTER REMOVAL TECHNIQUES IN PASSIVE RADAR

Alex Goodman

Author Affiliation
University of Colorado, Boulder

ABSTRACT

Passive radar utilizes existing radio signals to detect and track targets without emitting its own signal. However, clutter, unwanted signals from various sources, poses a significant challenge in extracting meaningful information. This study explores techniques to enhance target detection by isolating direct paths from clutter. Data from the LWA Data Archive, specifically Project LH014, comprised of 54 sessions conducted between January 2016 and July 2017, were analyzed. The 54 sessions focus on Jupiter as the target with Tuning frequencies at 20.0 MHz and 28.0 MHz and Stokes I and V polarization. This paper explores techniques to extract the direct signal from its echoes or artifacts through a variety of techniques and models.

1. INTRODUCTION

Passively sounding celestial bodies allow us to observe their surface and subsurface without the need for costly transmitters. Jupiter's decametric radiation serves as a suitable source of radio frequency signals for sounding on geological scales of interest but its spectral structure can introduce undesired artifacts. [1] In this study, we analyze data retrieved from the LWA Data Archive, with particular emphasis on Project LH014. The investigation encompasses a total of 54 sessions conducted between January 2016 and July 2017. The LWA1 station was utilized for these observations, with each session characterized by specific parameters including tuning frequencies, bandwidth, and target celestial object. The data acquired during these sessions are dynamic spectra, capturing both the Stokes I and V parameters. With a comprehensive dataset spanning over a year and a half, this analysis aims to shed light on the radio emissions from celestial sources, particularly focusing on Jupiter, and to elucidate any temporal or spectral patterns evident in the observations.

2. DATA PREP

The transformation of data from a nested column-wide HDF5 file to a row-wise Parquet file using PySpark was motivated by several factors pertinent to implementing machine learning techniques. Parquet files offer superior performance in terms of both storage efficiency and query processing speed compared to HDF5, particularly when handling large datasets. By converting the data into a row-wise format, we optimize it for parallel processing, a crucial requirement for scalable machine learning algorithms. PySpark provides a powerful distributed computing framework that seamlessly integrates with Parquet files, allowing for efficient data manipulation and analysis across distributed computing clusters. This combination of Parquet file format and PySpark facilitates the preprocessing and feature engineering steps necessary for machine learning, enabling us to extract meaningful insights and build robust predictive models from the transformed data.

Another motivation behind PySpark lies in its scalability and the ability to test on multiple machines without the necessity of creating unique environments to run a script. However, due to limited resources—4 CPUs, 16 GB RAM, and 200 GB disk space—the computation for this project faced constraints. Consequently, data processing had to be partitioned in ways that inhibited the utilization of all multiprocessing in Spark, resulting in a trade-off between resource limitations and computational efficiency. In the appropriate environment, various aspects of the data manipulation process can be fully converted into PySpark, enabling enhanced parallel computing capabilities.

3. POLARIZATION AND FREQUENCY ANALYSIS

In The LDAdb Utility tool under the session details several key pieces of information were used in the Data Prep and in this section, we aim to find frequency emission patterns by conducting summary statistics into its spectral characteristics and polarization behavior. All sessions in the project code "LH014," had the same details except for date, time, and duration. The tuning frequencies employed were 20.000 MHz and 28.0020 MHz, with a bandwidth of 19.600 MHz. The target of observation was Jupiter. Notably, the data type consisted of dynamic spectra, with polarization recorded in both I and V channels, totaling 1024 channels. The integration time for each observation was 40.124 milliseconds. [2] Leveraging this comprehensive information, our objective in this section is to discern frequency emission patterns by generating summary statistics on its spectral characteristics and polarization behavior. This detailed session description encompasses project specifics, observational parameters, target details, and serves as a foundation for our subsequent analysis and interpretation.

In our investigation, we briefly considered the exploration of saturation data; however, due to the scope limitations of this paper, a detailed analysis of saturation effects was not pursued. Further research could delve into the impact of saturation on the observed phenomena, providing valuable insights into its influence on signal processing outcomes.

4. CLUTTER REMOVAL TECHNIQUES

Following an exploration of various methodologies, we implemented a combined approach after completing an Extract-Transform-Load (ETL) procedure, which enabled noise removal across multiple stages. Each stage involved the evaluation of several techniques, with selections or exclusions made based on their respective outcomes. Visualization emerged as the primary method for model validation in the absence of training data, providing insights into the efficacy of the clutter removal techniques.

4.1 Bandpass Filter

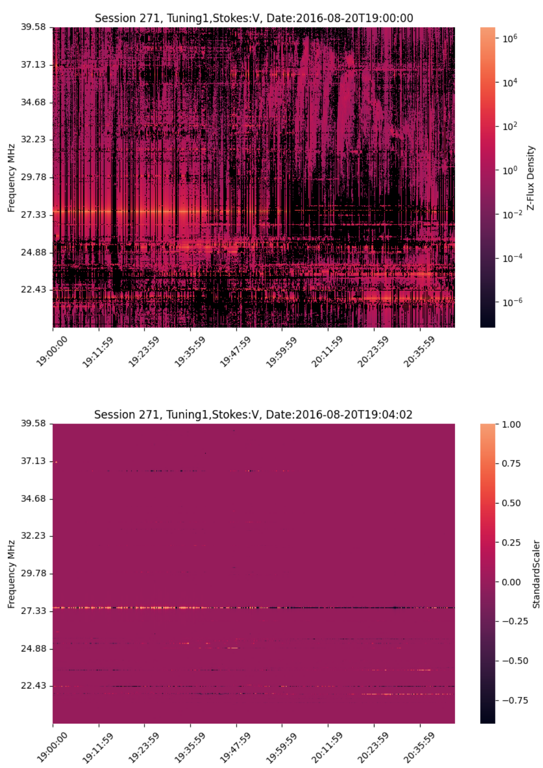


Figure. 1: Bandpass Filter - No Signal. This figure depicts Session 271, Tuning 1, Stokes I before and after applying the bandpass filter. The filter, with a conservative range set between -0.9 to 1, derived from the standard scaler. The results reveal that the entire observation consists solely of noise.

For the LWA array, which operates in the long wavelength regime, it measures the flux density of radio waves emitted by celestial objects such as stars, galaxies, and pulsars. By observing variations in flux density across different frequencies and timescales, properties and behaviors of these radio signals can be analyzed. [3]

The first method was to apply a bandpass filter to remove the majority of the clutter. Serval steps took place before the filter was applied.The flux density values were first sum-aggregated over a rounded time and frequency. This aggregation method facilitated the identification of signals more efficiently. However, due to the rounding and summation process, individual data points (such as timing and frequency) are no longer discernible.

From here the features need to be normalized to determine cutoff values. Two ways were looked at for scaling the data, MinMax Scaler and Standard Scaler. The MinMax scaler did not work since it is sensitive to outliers and the range of values is too wide.

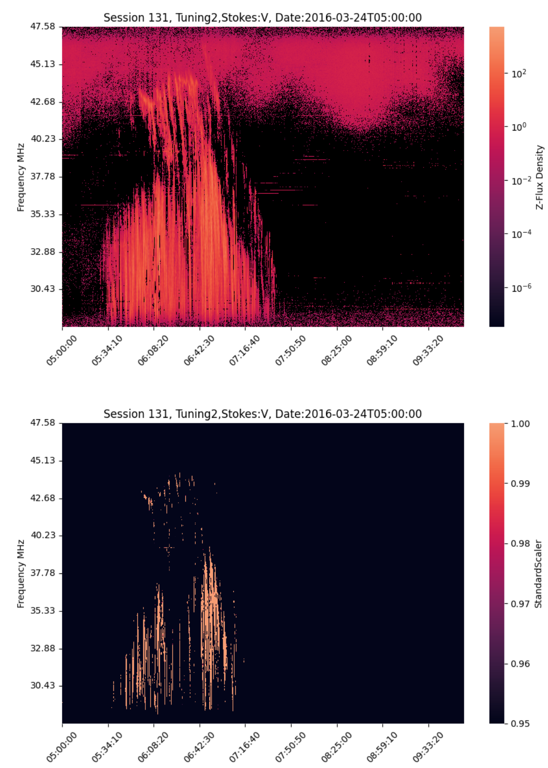


Figure. 2: Bandpass Filter - Signal Detection. This figure depicts Session 131, Tuning 2, Stokes V before and after applying the bandpass filter. The filter, with a range set between 0.95 to 1, derived from the standard scaler. The results reveal a signal between 30 and 40 MHz in the timeframe between 05:30 and 07:00.

The standard scaling process transforms the data to have a mean of 0 and a standard deviation of 1. This normalization ensures that all features are on a similar scale, preventing features with larger magnitudes from dominating the analysis. This resulted in a dataset that a Bandpass Filter could be applied to.

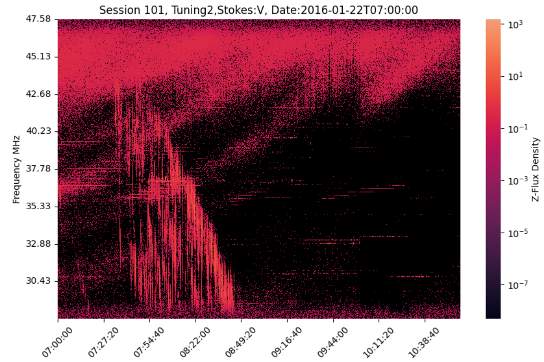
The utilization of standard scaling followed by bandpass filtering represents a synergistic approach in signal processing. Standard scaling, which normalizes data to have zero mean and unit variance, ensures uniformity in feature scales, promoting numerical stability and aiding algorithmic performance. Subsequently, applying a bandpass filter enables selective retention or amplification of specific frequency components, facilitating targeted signal extraction or feature isolation. By integrating these techniques, signal processing workflows benefit from enhanced efficiency, robustness, and interpretability, ultimately advancing various applications in signal analysis and communication systems.

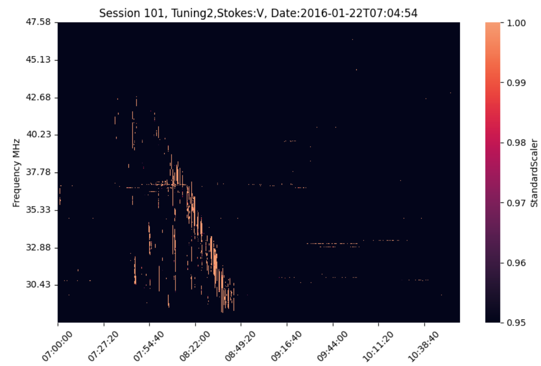
4.2 **Signal Clustering**

After the bandwidth filter is applied, there is noise or clutter left over that appears 'horizontal' or only manifests on one frequency. Moving Average, Frequency Domain Filtering, and Thresholding proved not affected as the signal does appear on the same frequency as the horizontal clutter.

Several methods were investigated to address residual clutter and 'horizontal' noise. Techniques including Principal Component Analysis (PCA), Isolation Forests, and One-Class Support Vector Machines (SVM) were ineffective in mitigating the 'horizontal' noise due to its atypical outlier characteristics. Two potentially promising approaches warrant further exploration. Firstly, Time Series Analysis, which involves examining the correlation between a signal and its lagged versions. Secondly, Fourier Transform, which entails decomposing signals into frequency components.

Density-Based Spatial Clustering of Applications with Noise (DBSCAN) was employed to identify clusters in the dataset based on the density of data points. This algorithm effectively delineates regions of high density as clusters while characterizing sparse regions as noise, providing a robust method for detecting patterns in complex datasets. DBSCAN takes two parameters, epsilon (ε), which defines the maximum distance between two points for them to be considered as part of the same neighborhood, and min_samples, which specifies the minimum number of points required to form a dense region (core point) in order to initiate a cluster.





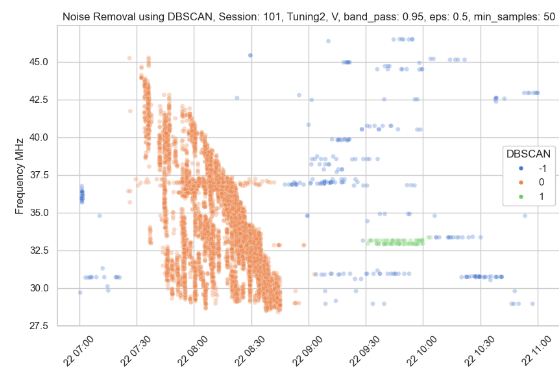


Figure. 3: DBCAN Filter - This illustration showcases Session 101 with DBSCAN effectively reducing the remaining noise from the Bandpass filter (range 0.95 to 1.0). The green cluster represents residual 'horizontal' noise. By adjusting the epsilon parameter to a smaller value, this noise cluster is eventually eliminated, leaving the predominant signal depicted in orange.

Random search followed by grid search were used for parameter tuning. Using the final grid search hyper-parameter optimization was successful in isolating the signal to a high degree. However, the hyper-parameter lead to overfitting for each Session Number. This is due to the uniqueness of each Session and upstreaming data sampling, aggregating, and scaling. Hierarchical Density-Based Spatial Clustering of Applications with Noise (HDBSCAN) was also explored as an alternative to DBSCAN.

HDBSCAN extends the capabilities of DBSCAN by automatically determining the number of clusters and providing a hierarchical clustering structure. For a semi-automated process, hyper-tuning DBSCAN was more effective than HDBCAN. Since a standard scaler was also applied, the data for the DBSCAN, epsilon has little meaning to the real world and domain knowledge of frequency and time of the signal are not helpful. Depending on further applications HDBSCAN could be useful in a fully automated scenario, but careful examination of the output would require additional steps.

4.3 **Horizontal Noise Distribution Analysis**

Examining the distribution within clusters identified by DBSCAN and HDBSCAN was employed to assess 'horizontal' noise. This noise, characterized by a uniform distribution rather than Gaussian, was scrutinized to validate the DBSCAN clustering results.

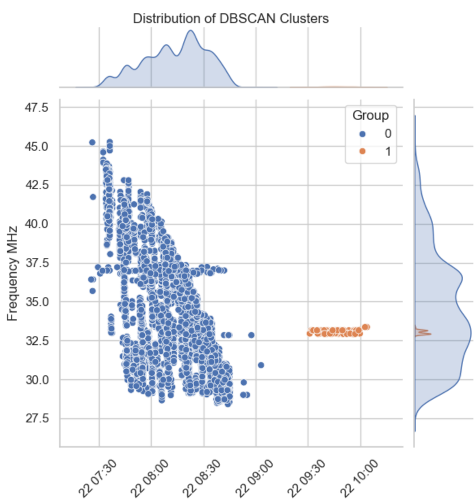


Figure. 4: DBSCAN Distribution - Extracted from the preceding Figure 3, this plot illustrates the distribution of points across the Frequency and Time domains. Though hard to fully see, the orange 'horizontal' group exhibits a more uniform distribution in the time domain.

The Kolmogorov-Smirnov test was employed to distinguish 'horizontal' noise characterized by its uniform distribution. By comparing the empirical cumulative distribution function of the noise data with the theoretical uniform distribution, the KS test quantifies the degree of discrepancy between the two distributions. This statistical method allows for the identification of 'horizontal' noise clusters, which display a distinctively uniform distribution pattern, enabling their differentiation from other signal components.

5. **CONCLUSION**

Employing an ensemble of multiple methods within an Extract-Transform-Load pipeline emerges as the most effective approach for determining the presence of a singular signal within a given session and pinpointing its location. With increased computational capabilities, bypassing data summarization or aggregation prior to bandpass filtering holds promise for achieving more precise cutoff values with meaningful interpretations, as opposed to relying on percentiles derived from a standard scaler. Regarding data delivery, uncertainties persist regarding the optimal approach tailored to the specific inquiries and research pertaining to signals within a session. While the current workflow identifies the frequency range, time range, and polarization of the signal, data processing limitations may compromise the fidelity of actual signal return values, necessitating potential restarts of downstream analyses with the original data. Nonetheless, this approach provides valuable insights into the segments of data warranting further investigation as well as identifying sessions with no discernible signal. Notably, a prevalent signal trend emerges across almost all sessions within the LH014 Project, predominantly observed in Tuning 2 and Polarization V configurations.

6. **REFERENCES**

[1] T. M. Roberts, A. Romero-Wolf, L. Bruzzone, L. Carrer, S. Peters and D. M. Schroeder, "Conditioning Jovian Burst Signals for Passive Sounding Applications," in IEEE Transactions on Geoscience and Remote Sensing, vol. 60, pp. 1-14, 2022.

[2] LDAdb Utility tool, "Observation Detail - Project: LH014" University of New Mexico.

[3] F. Schinzel and E. Polisensky, “System Equivalent Flux Density of LWA1 Beams,” v1, July 28, 2014.

[4] H. Gao, W. Zhang, F. Jia and Z. He, "The extraction method for direct path wave in the airborne passive radar system," 2016 IEEE

[5] Mauduit, E., Zarka, P., Lamy, L. et al. Drifting discrete Jovian radio bursts reveal acceleration processes related to Ganymede and the main aurora. (2023).

[6] Jayce Dowell, Gregory B. Taylor, Frank K. Schinzel, Namir E. Kassim, Kevin Stovall, The LWA1 Low Frequency Sky Survey, Monthly Notices of the Royal Astronomical Society, Volume 469, Issue 4, August 2017, Pages 4537–4550

[7] Clarke, T. E., Higgins, C. A., Imai, M., and Imai, K., “Jovian decametric emission with the Long Wavelength Array station 1 (LWA1)”, in <i>Planetary Radio Emissions VIII</i>, 2017, pp. 31–44