Lecture Notes: Data Science in Radio Astronomy II

Lecture 1: Understanding and Managing Noise in Radio Data

Introduction to Noise in Radio Astronomy

- **Thermal Noise:** Arises from the random motion of electrons in our instrumentation, quantified by the thermal noise equation:
$$P = kTB \tag{1}$$
where P is the noise power in watts, k is the Boltzmann constant ($1.38 \times 10^{-23} J/K$), T is the temperature of the system in Kelvin, and B is the observed bandwidth in Hertz.
- **Instrumental Noise:** Includes calibration errors, quantization noise, and imperfections in the hardware. This can normally be managed through careful calibration and maintenance.
- **Radio Frequency Interference (RFI):** Human made signals such as those from cell towers, satellites, and aircraft. To help avoid RFI, radio telescopes are often built in remote and shielded environments, but avoiding signals from non-terrestrial sources like satellites is difficult.
- **Cosmic Noise:** Naturally occurring signals from astronomical sources such as the Sun, Milky Way, and cosmic microwave background (CMB).
- **The Noise Floor:** The baseline level of noise power present in the frequency domain. A flat noise floor indicates white noise, which has equal power distributed across all frequencies. Deviations from flatness suggest colored noise, where power varies with frequency. For example, pink noise decreases with $1/f$, and brown noise decreases with $1/f^2$.

Visualizing Noise

- **Time-Domain Plots:** Shows variations in the signal over time but may obscure frequency-specific details.
- **Frequency-Domain Plots:** Reveals noise characteristics across frequencies, which is important for understanding RFI and instrumental effects.
- **Spectrograms:** Combine time and frequency information, providing a dynamic view of noise behavior.
- **Identifying Noise Patterns and Anomalies:** Regular, repeating patterns often indicate instrumental or RFI sources. High variability in specific frequency regions may suggest transient interference or improper calibration.
- **Tools for Plotting and Analyzing the Noise Floor:** In Python, `matplotlib` can be used for basic plots, and `scipy.signal` for spectral analysis. GNU Radio offers spectrum and waterfall sinks for real-time visualization.

Basic Statistics for Noise Analysis

- **Mean:** The average value of the noise power, calculated as:

$$\mu = \frac{1}{N}\sum_{i=1}^{N} x_i \tag{2}$$

  where x_i are the individual noise power samples, and N is the total number of samples. provides a measure of the central tendency of the noise power over a specified bandwidth or period of time. The mean is useful for understanding the overall noise level, but is sensitive to outliers such as spikes from RFI.

- **Median:** The middle value in the sorted noise power data. Offers a robust measure of the central tendency, and is less influenced by extreme outliers. Median can be used to estimate the baseline noise level in the presence of transient RFI.

- **Standard Deviation:** A measure of the spread or variability in noise power, calculated as:

$$\sigma = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(x_i - \mu)^2} \tag{3}$$

  where N is the number of samples, x_i are the individual samples, and μ is the mean. Standard deviation quantifies how much the noise fluctuates around the mean, helping to identify anomalies or regions of higher variability.

- **Signal-to-Noise Ratio (SNR):** The ratio of the power of a signal to the power of the noise, often expressed in decibels (dB) as :

$$SNR = 10 \cdot log_{10}\left(\frac{P_{signal}}{P_{noise}}\right) \tag{4}$$

  A higher SNR means the signal stands out more clearly above the noise floor, making it easier to detect and analyze, and is an important metric for deciding whether a detection is significant. SNR improves with longer integration times because noise averages out ($\propto 1/\sqrt{time}$), and decreases with wider bandwidth due to the inclusion of more noise power.

Filtering Techniques

- Filters suppress unwanted components of the signal while preserving useful data.
- **Bandpass Filters:** Allows signals within a specified frequency range to pass through, while attenuating frequencies outside this range.
- **Low-Pass and High-Pass Filters:** Attenuates high-frequency or low-frequency components, respectively.
- **Moving Average Filter:** Smoothens time-domain data to reduce variability and highlight trends.

Lecture 1 Discussion Questions

1. Why do radio telescopes often operate at very low temperatures, and how does this relate to the thermal noise equation?

   Radio telescopes operate at low temperatures to minimize thermal noise, which is directly proportional to the system temperature as described by the equation $P = kTB$. By cooling receivers to cryogenic temperatures, the random motion of electrons is reduced, resulting in lower noise power and a quieter baseline for detecting faint signals. This is particularly important in radio astronomy, where signals from space are often orders of magnitude weaker than local sources of interference. For instance, using liquid helium or liquid nitrogen cooling can significantly improve a telescope's sensitivity, allowing it to detect signals that would otherwise be lost in the noise.

2. How can statistics like mean, median, and standard deviation help distinguish between noise and potential signals?

   Statistics like mean, median, and standard deviation provide critical insights into the noise characteristics of radio data. The mean gives the average noise power and helps establish a baseline, though it can be skewed by outliers such as transient interference. The median, being less sensitive to extreme values, provides a more robust estimate of the noise floor, especially when outliers are present. The standard deviation quantifies the variability of the noise, and regions with significantly higher variability might indicate the presence of RFI or a potential signal standing out above the typical noise fluctuations. These tools allow astronomers to systematically identify and isolate signals in a noisy environment.

3. If you were designing a new radio telescope for a remote location, what steps would you take to minimize noise and interference?

   Designing a new radio telescope for a remote location would involve selecting a site far from human activity, such as a high-altitude desert or a geographically shielded valley, to reduce RFI from populated areas. The telescope would need to be equipped with cryogenically cooled receivers to minimize thermal noise and ensure high sensitivity. Shielding the electronics and using filters to block unwanted frequencies would further reduce instrumental noise. Additionally, scheduling observations during times of minimal RFI activity, such as late at night, and using real-time RFI detection and mitigation algorithms would help maintain data quality.

Lecture 1 Resources

1. Sensitivity and Noise in Radio Astronomy - University of Glasgow
2. Noise in Radio Systems - University of Toronto
3. The radio signal from source to reception - University of Manchester
4. Real-Time RFI Mitigation for Single-Dish Radio Telescopes - Green Bank Observatory
5. The Town Where Wi-Fi Is Banned: The Green Bank Telescope and the Quiet Zone
6. Radio Frequency Interference - American Astronomical Society

Lecture 2: Advanced Analysis of Spectral Features

Automated Peak Detection

- Peaks in the spectrum correspond to strong signals that stand out above the noise floor. These are often narrowband signals from spacecraft, characterized by their Doppler shifts caused by the relative motion between the spacecraft and Earth. Identifying peaks is critical for detecting and monitoring transmissions from spacecraft, as well as studying their telemetry. Techniques include:
- **Threshold-Based Detection:** Identifying points in the spectrum where the signal exceeds a predefined threshold, typically based on the noise floor plus a multiplier of the standard deviation.
- **Derivative-Based Methods:** Using changes in the slope of the spectrum to locate local maxima.
- **Matched Filtering:** Comparing the observed spectrum with a template of the expected signal to enhance detectability.
- **Algorithms for Peak Detection:**
  - **Scipy's `find_peaks` Function:** A simple, effective tool for locating peaks based on prominence, width, and height.
  - **Sliding Window Methods:** A rolling analysis to detect peaks within a localized region of the spectrum.
  - **Fourier Transform Techniques:** Enhancing peaks in the frequency domain by filtering out unwanted noise or broadening refinement

Analyzing Radio Frequency Interference

- **Sources of RFI:** RFI can originate from terrestrial sources like communication systems, satellites, and electronics. When analyzing spacecraft signals, RFI from Earth-based transmitters in the same frequency band is a significant challenge. Examples include:
  - **Satellite Downlinks:** These signals are difficult to avoid as they originate from low-Earth orbit or geostationary satellites.
  - **Ground-Based Radar and Communication Systems:** High-power transmissions that can drown out faint signals.
  - **Transmission Regulations:** Radio astronomers often operate in protected bands that are regulated by organizations such as the Federal Communications Commission (FCC), US Department of Commerce, and the International Telecommunications Union (ITU). However, transmissions from spacecraft often fall outside of these protected bands, increasing the impact of RFI.
- **Techniques for RFI Detection and Flagging:**
  - **Median Filtering and Baseline Subtraction:** Isolating narrowband interference from the noise floor.

  - **Statistical Analysis:** Techniques like spectral kurtosis can be used to identify deviations from Gaussian noise.
  - **Time-Frequency Analysis:** Spectrograms help distinguish transient interference from persistent signals.
  - **Flagging Tools:** Software like AOFlagger and custom scripts can automate the process of identifying and masking RFI in datasets.

Signal Classification and Machine Learning

- Machine learning (ML) provides a powerful set of tools for automating the classification and analysis of radio signals. With the vast amount of data generated by telescopes and spacecraft tracking systems, ML enables the rapid identification of meaningful patterns and features, such as distinguishing spacecraft signals from noise or interference.
- **Feature Engineering:** Involves selecting and extracting measurable properties from the raw data.
- **Supervised Learning:** In supervised learning, the algorithm is trained on labeled data (e.g., known examples of spacecraft signals, noise, and RFI). The trained model can then classify new, unlabeled data. Examples of supervised learning methods include decision trees, random forests, and support vector machines (SVMs).
- **Unsupervised Learning:** Unsupervised learning is used when the data lacks labels. It groups data into clusters based on similarities, which can help discover unknown signal types or identify RFI patterns. Common algorithms include k-means clustering and hierarchical clustering.
- **Applications in Spacecraft Signal Analysis:**
  - Classifying telemetry data from spacecraft, such as beacon signals, carrier tones, or modulated telemetry streams.
  - Identifying RFI contamination by training models to distinguish between natural signals and man-made interference.
  - Monitoring for anomalies in spacecraft transmissions, which could indicate system malfunctions or unexpected behaviors.
- **Challenges and Considerations:**
  - **High Dimensionality:** Radio data often contains thousands of frequency channels, making dimensionality reduction techniques (e.g., principal component analysis) essential for efficient processing.
  - **Diverse Signal Types:** Spacecraft signals can vary widely depending on the mission and transmission protocols, requiring flexible models capable of adapting to different data structures.
  - **Imbalanced Datasets:** Signals from spacecraft may be rare compared to the prevalence of noise and interference, requiring strategies to handle imbalanced datasets, such as oversampling or class weighting.

Applications to Radio Astronomy and SETI

- **RFI mitigation in SETI:** Mitigating radio frequency interference (RFI) is a cornerstone of both radio astronomy and SETI. Detecting faint extraterrestrial signals or subtle astrophysical phenomena requires effective strategies to identify and remove human-made interference.
- **How RFI Impacts Observations:** RFI sources such as cell towers, satellites, and ground-based radar often produce strong narrowband or broadband signals that overlap with frequencies used for radio astronomy or SETI. Persistent RFI can mimic the characteristics of signals from extraterrestrial sources, leading to false positive detections.
- **On-Off Observations:** When conducting radio observations, it is common to switch the pointing of the telescope from on-source to off-source. If a signal of interest originates from the source being studied, it will appear only in the on-source observations. However, if the signal is due to locally generated RFI, it will appear in both the on-source and off-source observations, allowing astronomers to distinguish between interference and genuine signals.
- **Excision Techniques:** Narrowband RFI is removed by flagging and interpolating affected frequency bins.
- **Statistical Tools:** Techniques like spectral kurtosis help identify non-Gaussian noise caused by interference.
- **Real-Time Monitoring:** Tools such as spectrograms and live SNR mapping are used to flag and exclude RFI-contaminated regions during observations.

Lecture 2 Discussion Questions:

1. Why is automated peak detection critical for analyzing signals from spacecraft like Voyager 1 or Mars orbiters?

   Automated peak detection is critical because spacecraft signals are often extremely weak compared to the noise floor, especially for distant missions like Voyager 1. Manually identifying these signals in large datasets is impractical, given the narrowband nature of most spacecraft transmissions and the vast amount of spectral data collected by modern telescopes. Additionally, Doppler shifts caused by the relative motion between the spacecraft and Earth result in the signal's frequency changing over time, requiring algorithms that can dynamically track peaks in the spectrum. Automated tools make it feasible to monitor these signals in real-time, ensuring that critical data, such as telemetry and science payload transmissions, are not missed.

2. What are the challenges of detecting spacecraft signals in the presence of RFI, and how can these be mitigated?

   Detecting spacecraft signals in the presence of RFI is challenging because terrestrial interference is often much stronger than the weak signals from spacecraft. For example, geostationary satellite downlinks and ground-based radar can produce strong, persistent interference in frequency bands near spacecraft transmissions. Mitigation strategies include observing in radio-quiet zones, using highly directional antennas to reduce interference from unwanted sources, and applying real-time RFI detection and excision algorithms. Additionally, distinguishing between RFI and legitimate spacecraft signals often requires careful statistical analysis or machine learning models trained to classify spectral features. Tools like time-frequency plots and median filtering can also help isolate transient RFI and protect the integrity of the data.

3. Why might unsupervised learning be valuable for analyzing unknown or unexpected spacecraft signals?

   Unsupervised learning is valuable because it doesn't rely on labeled data and can uncover patterns or groupings in the data that were not previously known. This makes it particularly useful for exploring unknown or unexpected spacecraft signals, such as anomalous telemetry transmissions or unexpected changes in carrier tones. Clustering algorithms like k-means can group similar signals, allowing researchers to identify new classes of signals that might correspond to different spacecraft systems or behaviors. For instance, if a spacecraft unexpectedly switches to a backup transmitter or changes its modulation scheme, unsupervised learning methods can detect these changes without prior knowledge.

Lecture 2 Resources

1. SciPy `find_peaks` Function Documentation
2. GNU Radio Peak Detector Documentation
3. Peak Finding and Measurement - University of Maryland
4. Radio Frequency Interference - NRAO
5. Introduction to RF Signal Classification

Lecture 3: Advanced Signal Processing and Applications

Understanding Signal Quality with Signal-to-Noise Ratio (SNR)

- **Recap: Signal-to-Noise Ratio:** SNR quantifies the strength of a signal relative to the noise present in the observation. It is calculated as the ratio of signal power (P_{signal}) to noise power (P_{noise})and is often expressed in decibels (dB) using:

$$SNR = 10 \cdot log_{10}\left(\frac{P_{signal}}{P_{noise}}\right) \tag{5}$$

  In radio astronomy and spacecraft signal analysis, achieving a high SNR is critical because signals from distant sources, like Voyager 1, are typically orders of magnitude weaker than the noise floor. A low SNR can render the signal indistinguishable from background noise, while a high SNR ensures the signal is detectable and analyzable.

- **Methods for Calculating SNR**
  - **Time-Domain Approach:** Compare the amplitude of the signal to the noise baseline in a time-series plot. This method is useful for detecting transient or periodic signals.
  - **Frequency-Domain Approach:** Evaluate the peak power of a signal in a power spectral density (PSD) plot and compare it to the average noise floor in adjacent frequency bands. This method is ideal for spacecraft signals, which often appear as narrowband peaks in the spectrum.

- **Enhancing SNR**
  - **Increasing Integration Time:** The noise power averages out over time a $1/\sqrt{time}$, improving SNR for faint signals.
  - **Reducing Bandwidth:** Narrower bandwidths exclude unnecessary noise, enhancing the clarity of narrowband signals like spacecraft carrier frequencies.

Noise Reduction and Signal Enhancement Techniques

- **Overview of Noise Reduction:** Noise reduction techniques are essential for improving the signal-to-noise ratio (SNR) in spacecraft signal analysis, ensuring faint signals can be detected and analyzed. These methods aim to suppress unwanted noise while preserving the integrity of the signal. Effective noise reduction is particularly important when dealing with signals like spacecraft telemetry, which are often buried in noise due to their low power and long distances.

- **Wavelet Transforms:** A wavelet transform splits a signal into high-frequency (noise) and low-frequency (signal) components. By selectively filtering the high-frequency components, noise can be reduced while retaining critical signal features. It can be used to remove broadband noise from telemetry signals, and help isolate narrowband carrier frequencies or transient modulated signals.
  - **Advantages:** Adaptive to both time-domain and frequency-domain analysis, which enables effective filtering of non-stationary signals, and preserves fine details in the signals compared to traditional Fourier-based filtering.

- **Limitations:** Can be computationally expensive for large datasets, and requires careful selection of the wavelet basis function to avoid signal distortion.

- **Savitzky-Gorlay Filters:** Smooths signals by fitting local segments to polynomial functions, reducing noise while preserving important spectral features. Works by applying a sliding window across the signal, fitting each segment to a low-order polynomial and replacing the central value with the polynomial's prediction.
  - **Advantages:** Maintains sharp features, making it ideal for spectral data with narrowband signals, while avoiding the blurring effects common with moving average filters.
  - **Limitations:** Ineffective for removing high-frequency noise if the polynomial order or window size is not chosen correctly.

- **Adaptive Filtering:** Dynamically adjust their parameters based on the noise characteristics of the signal, making them well suited for environments with non-stationary noise or interference. The filter iteratively learns the noise profile in real time and suppresses it without impacting the signal.
  - **Advantages:** Effective for removing time varying noise or interference, and can be used in real-time signal processing pipelines.
  - **Limitations:** Requires initial training or calibration, and may struggle with highly complex noise profiles.

- **Welch's Method:** Reduces noise by averaging the power spectra of overlapping segments of the signal, providing a smooth and enhanced spectral estimate. This method divides the time-domain signal into overlapping segments, applies a window function to each segment to minimize edge effects, and computes the FFT of each segment and averages the resulting power spectra.
  - **Advantages:** Reduces random noise without altering the frequency content of the signal, smoothing the power spectral density to reveal faint signals.
  - **Limitations:** Reduces spectral resolution due to segmentation, and can be sensitive to the choice of window function and segment size.

- **Spectral Excision:** Removes specific regions of the spectrum contaminated by narrowband RFI or other interference, ensuring these regions do not obscure the signal of interest. Works by identifying frequency bins with excessive power compared to the surrounding spectrum, and replaces them with interpolated values, or flags them for exclusion.
  - **Advantages:** Focused removal of noise without impacting the rest of the spectrum, which is ideal of environments with strong, localized interference.
  - **Limitations:** Ineffective for broadband interference, and has the risk of inadvertently removing legitimate signals if the thresholds are not carefully set.

- **Moving Average Filters:** Smooths signals by averaging data points within a sliding window, replacing each data point in the spectrum with the average of its neighbors within a fixed window size.

- **Advantages:** Simple and computationally efficient, and is effective for reducing noise on time-domain data.

- **Limitations:** Blurs sharp spectral features, making it unsuitable for narrowband signals, and non-stationary or frequency-specific noise.

Data Visualization for Signal Analysis

- **Overview of Data Visualization:** Data visualization is a critical tool in spacecraft signal analysis, allowing researchers to interpret complex datasets, identify patterns, and assess the impact of noise reduction techniques. Effective visualization helps highlight signal quality, detect anomalies, and compare signals before and after processing.

- **SNR Mapping:** Involves plotting the signal-to-noise ratio across the spectrum to identify regions with higher signal quality, or areas requiring further noise reduction. Python libraries such as `matplotlib` and `plotly` can be used to create dynamic and interactive SNR plots, while spectral analysis tools in GNU Radio can provide real-time SNR visualization.

- **Interactive Spectrograms:** Spectrograms provide a combined time-frequency representation of a signal, offering insights into how the signal evolves over time. Works by dividing the signal into time segments and computing the FFT for each segment, then plotting the power spectrum for each segment, where time is on the x-axis, frequency is on the y-axis, and a color scale represents power. Use `plotly` to create interactive spectrograms in Python, or use waterfall plots in GNU Radio.

- **Comparative Plots:** Used to assess the effectiveness of noise reduction techniques or signal processing algorithms. Plot the spectrum before and after applying noise reduction or filtering techniques to highlight differences in the noise floor, signal peaks, and overall spectral structure.

- **Real-Time Visualization:** Enables dynamic monitoring of signals during live observations, especially important for spacecraft signal tracking.

- **3D Visualizations:** 3D plots can represent complex relationships in the data, such as power variations over time and frequency. The `plotly` library in Python supports interactive 3D plotting, but other tools such as MATLAB or Wolfram Mathematica can generate high-quality 3D plots.

Lecture 3 Discussion Questions

1. Why is signal-to-noise ratio (SNR) an essential metric in spacecraft signal analysis, and how can it be improved during observations?

   SNR is essential because it determines the detectability of a signal amidst background noise. Spacecraft signals, such as those from Voyager 1, are often orders of magnitude weaker than terrestrial noise or cosmic background radiation, making high SNR crucial for accurate analysis. SNR can be improved by increasing the integration time, which averages out random noise, or by narrowing the observation bandwidth to focus on specific signal frequencies. Both techniques enhance the signal clarity, enabling the detection of weak, narrowband signals like telemetry or carrier tones.

2. Why do scientists use multiple noise reduction techniques instead of just one?

   Different types of noise require different reduction techniques. For example, thermal noise is best mitigated by increasing integration time, while radio frequency interference (RFI) is better handled with spectral excision or filtering. Some methods, like adaptive filtering, work well in real time, while others, like wavelet transforms, are better for post-processing. Combining multiple techniques ensures that various sources of noise are reduced without distorting or removing the actual signal. The right combination depends on the scientific goals and the characteristics of the data.

3. What ethical or philosophical implications arise from searching for and detecting signals from extraterrestrial intelligence?

   The search for extraterrestrial intelligence (SETI) raises profound ethical and philosophical questions. If we detect a signal, how should humanity respond? Should we attempt to communicate, or remain silent to avoid potential risks? Who gets to decide how we interact with an extraterrestrial civilization? Additionally, SETI challenges our perception of our place in the universe—if we find intelligent life, it could reshape our understanding of biology, evolution, and even human culture. On the other hand, if we continue to find nothing, it raises questions about the rarity of life and our uniqueness in the cosmos.

Lecture 3 Resources

1. Signal-to-Noise Ratios - University of Sheffield
2. What is Signal to Noise Ratio and How to Calculate it?
3. Signal Processing: Filtering Out The Noise
4. Wavelet Transform: A Practical Approach to Time-Frequency Analysis
5. Introduction to the Savitzky-Golay Filter: A Comprehensive Guide (Using Python)
6. Overview of Adaptive Filters and Applications
7. Welch's Method in Python (`scipy`)