Utilizing Wavelet Transforms to Locate Anomalies in Time-Series Data

Tyler Turek[1], Saianeesh Haridas[2], Mathew Madhavacheril[2]

[1]University of Pennsylvania, C'26, W'26; [2]Department of Physics and Astronomy, University of Pennsylvania

Introduction

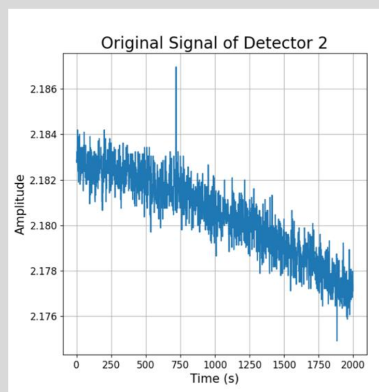

Figure 1: An example of a glitch in the raw data collected from the Simons Observatory.

- Anomalies, such as jumps and glitches (fig. 1), often occur in the **time-series data** produced by astronomical instruments (e.g., the **Simons Observatory**).
- Identifying and addressing these anomalies is crucial for creating accurate datasets, which are essential for investigating the universe's evolution.
- Currently, **Fourier transforms** are widely used for anomaly detection by decomposing signals into **sinusoidal functions**. However, this method **struggles** with signals that exhibit numerous anomalies and does not provide information about when these anomalies occur.
- **Wavelet transforms** present a promising alternative to Fourier transforms for anomaly detection by using **wavelets** localized in frequency *and* time, allowing them to better reveal localized features.
- This project focused on evaluating the feasibility and accuracy of the **Mallat-Zhong Discrete Wavelet Transform (MZ-DWT)** algorithm for detecting these anomalies.[1, 2]

Methods

- Implemented the MZ-DWT in a **python** environment.[3]
- Utilized the wavelet transform to calculate **alpha (α)** (fig. 2), which is based on the Lipschitz and Hölder Continuities.[4]
- α provides insights into signal behavior at discrete timepoints (e.g., $\alpha \approx 0 \rightarrow$ jump, $\alpha \approx -1 \rightarrow$ glitch).[4]
- Improved algorithms and runtime with **simulated data** (fig. 3).
- Utilized **real-data** to calculate **false positive and negatives**.

$$\log |Wf(u,s)| = \alpha \log s + C$$

Figure 2: Equation for calculating α, where *Wf* represents the wavelet transform and *s* denotes the scale.

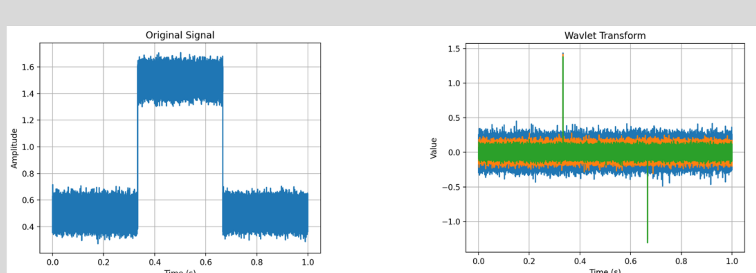

Figure 3: Example of simulated data (right) and the corresponding wavelet transform (left).

Results

This anomaly detection method relies on two inputs: the **anomaly threshold** (which determines the size of features flagged by the algorithm) and the **alpha threshold** (which specifies the types of features detected). A significant part of this project was determining the optimal combination of these to minimize false positives and false negatives.

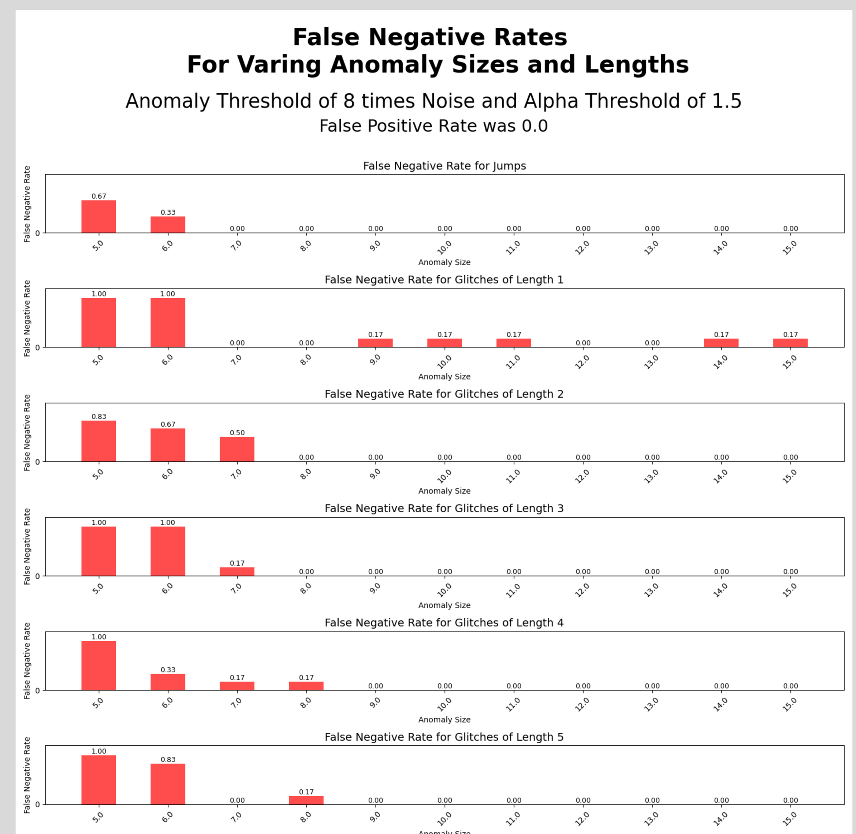

Figure 5: Performance of the anomaly detection method in identifying various types of anomalies using an optimal set of input parameters.

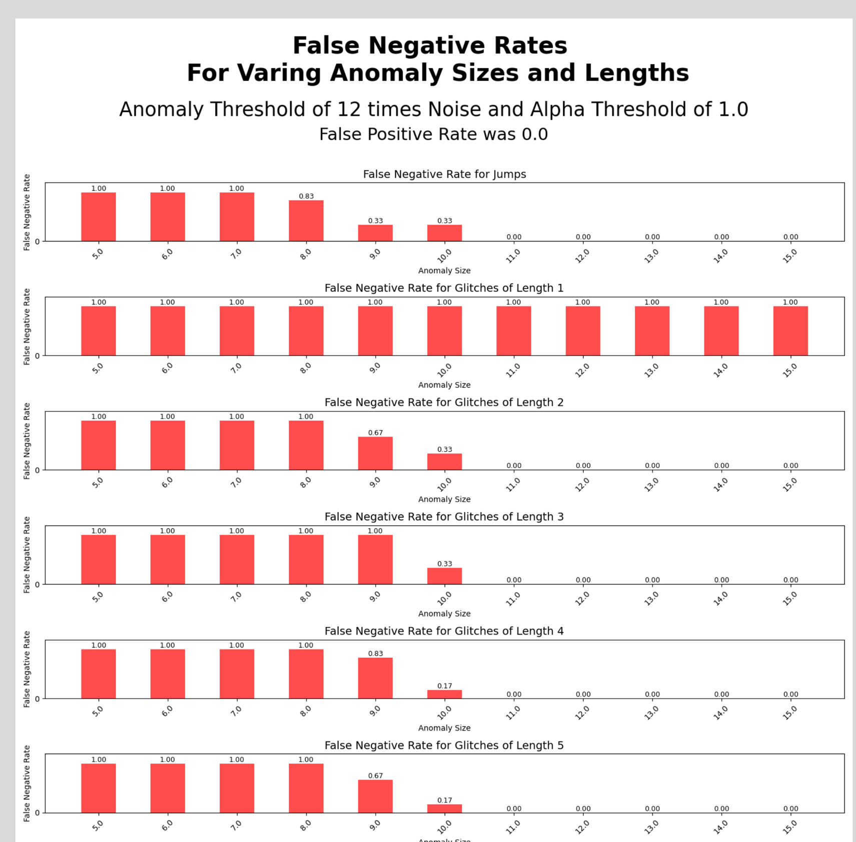

Figure 6: Performance of the anomaly detection method in identifying various types of anomalies using a suboptimal set of input parameters.

The results from Fig. 5 and 6 match what we expect, as once the anomaly size exceed the anomaly threshold, it is detected.

In Fig. 7 and 8, the blue-highlighted regions indicate the ideal combination of the anomaly threshold and alpha threshold that minimizes both false positives and false negatives.

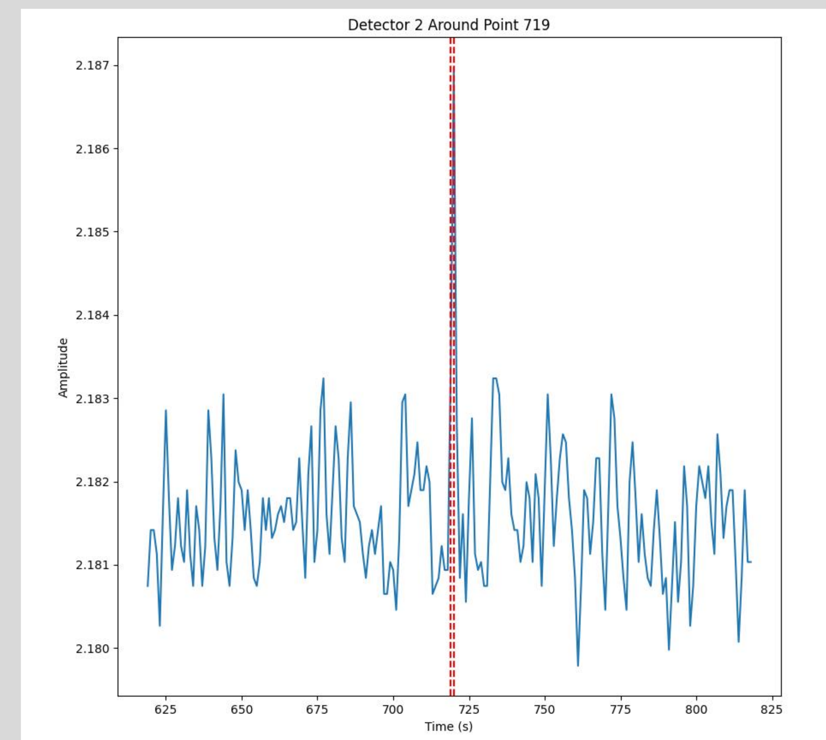

Figure 4: The anomaly detection method identifying a glitch (indicated by the red dotted line) in the dataset from Fig. 1.

Figure 7: Heatmap showing false positive rates for various combinations of input parameters.

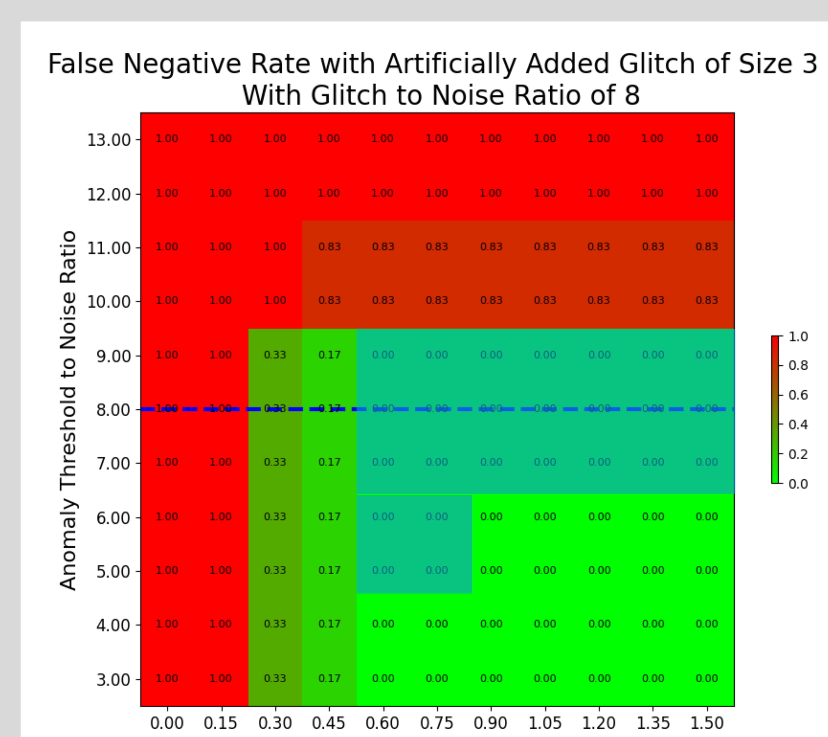

Figure 8: Heatmap showing false negative rates for a specific anomaly across various combinations of input parameters.

Conclusion

- **This project has demonstrated that wavelet transforms are a promising tool for detecting anomalies in time-series data.**
- These transforms can identify anomalies with a high degree of accuracy, even those as small as 8 times the white noise level.
- Future steps for this project include optimizing runtime, using transform data to quantify the size of anomalies, and comparing the performance of wavelet transforms with existing Fourier transform methods.
- Ultimately, wavelet transforms could become an invaluable tool in the data processing pipeline for astrophysics instruments, helping researchers uncover the secrets of the universe's evolution.

Acknowledgements

This project was made possible through the generous support of the Penn Undergraduate Research Mentoring Program, organized by the Center for Undergraduate Research and Fellowships. Special thanks to Dr. Mathew Madhavacheril and Saianeesh Haridas for their invaluable guidance and assistance.

References

1. Mallat S. & Zhong S. (1992). Characterization of signals from multiscale edges. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 4(7) 710-732. doi: 10.1109/34.142909.
2. Mallat S. & Hwang W. L. (1992). Singularity detection and processing with wavelets. *IEEE Transactions on Information Theory*, 38(2), 617-643. doi: 10.1109/18.119727.
3. Sadler B. M. & Swami A. (1999). Analysis of multiscale products for step detection and estimation. *IEEE Transactions on Information Theory*, 45(3), 1043-1051. doi: 10.1109/18.761341.
4. Damerval C. (2012). Study of Lipschitz regularity. *JRC Technical Reports*. doi:10.2788/46849