Enhancing Cosmological Observations: An Automated Pipeline for Time Constant Analysis in SPT-3G Data

Ashton Davis [1]

[1] *Stephen F. Austin State University*
Nacogdoches, Texas
davisaj14@jacks.sfasu.edu

ABSTRACT

This paper introduces an automated pipeline designed for South Pole Telescope (SPT) 3G survey data, targeting the critical challenge of analyzing time constants from bolometric detectors. The main scientific goal of this project is to emphasize that time constants can vary with atmospheric conditions and the importance of accurately accounting for these variations to obtain precise time constant values for each field or subfield. The SPT-3G survey organizes observations into various fields and subfields, conducted across both summer and winter periods. The Wide field is monitored in both summer and winter, furthermore, the Wide field is further segmented into subfields from wide-a to wide-i. The pipeline integrates a series of processes to ensure precise collection, processing, and statistical evaluation of data under varying observational conditions, including weather, elevation angle, and seasonal changes. This tool not only improves data reliability but also expands the potential for future research applications.

Keywords: South Pole Telescope (SPT) — Time Constant — Millimeter-Wave Astronomy — Data Processing Pipeline — Bolometer Calibration — Modified Julian Date (MJD) — Interquartile Range (IQR) — Statistical Analysis — Weighted Linear Regression — Quartile Analysis — Elevation Angles — Cryogenic Detectors — Calibration Techniques

Contents

1. INTRODUCTION

Precise measurement of time constants is vital for enhancing the sensitivity and accuracy of data obtained in observational cosmology. The Cosmic Microwave Background (CMB) is an important observational target because it provides insights into the early universe, its composition, and subsequent evolution. SPT-3G, an influential project in millimeter-wave astronomy, implements highly sensitive detectors called bolometers to detect these faint cosmic signals. A bolometer is a device that measures the power of incident electromagnetic radiation through the heating of an absorbing material. Bolometers work by using minute temperature changes to measure incident electromagnetic radiation. The accurate characterization of their time constants, which denote the response time to these changes, is critical for maintaining the integrity of observational data. The time constants can vary with atmospheric conditions, elevation angle, and seasonal changes, affecting the detector's loading and subsequently their behavior. In order to enhance the telescope's data reliability and detection capabilities, this research presents a new automated pipeline that optimizes the calibration and analysis of these temporal constants.

The CMB is the afterglow of the Big Bang, emitted approximately 380,000 years after the event, when the universe cool enough for protons and electrons to combine into neutral atoms, allowing photons to travel freely. The CMB has a blackbody temperature of about 2.73 Kelvin, making it a faint but pervasive background radiation observable in all directions. Despite its uniformity, the CMB contains tiny temperature fluctuations, or anisotropies, on the order of microkelvins, or millionths of a degree. These anisotropies are crucial because they provide a snapshot of the early universe's density fluctuations, which eventually led to the formation of galaxies and large-scale structures. By studying the CMB, we can learn about the universe's composition, age, rate of expansion, and gain insights into the fundamental physics governing its evolution.

2. INSTRUMENTATION AND TECHNOLOGY

The SPT-3G employs a suite of advanced instruments and technologies to conduct high-precision observations of the CMB. Designed to detect faint cosmic signals and minimize noise, the telescope's equipment is crucial for cutting-edge cosmological research. Key components of the SPT-3G's instrumentation include:

- **Superconducting Detectors**: These advanced detectors are highly sensitive to millimeter-wavelength radiation and operate at cryogenic temperatures to significantly reduce thermal noise, enhancing the detection capabilities of faint astronomical signals.

- **Cryogenic Systems**: Employing sophisticated cryogenic cooling technologies, these systems maintain the detectors at extremely low temperatures, maximizing their sensitivity and performance.

- **High Angular Resolution**: The telescope's 10-meter dish provides high angular resolution, enabling the detection of small-scale fluctuations in CMB temperature and polarization patterns. This high resolution allows SPT-3G to measure angular separations as small as 3.6 arcminutes, crucial for resolving fine structures in the cosmic microwave background. The 3.6 arcminutes value is derived from the telescope's design, specifically its 10-meter dish size, and the wavelengths used in the SPT-3G survey, particularly around the 150 GHz frequency band. Angular resolution is determined by the diffraction limit, calculated using the formula $\theta \approx \frac{1.22\times\lambda}{D}$, where λ is the observing wavelength and D is the diameter of the dish.

- **Observation Frequency Bands**: Operating across 90 GHz, 150 GHz, and 220 GHz frequency bands, SPT-3G optimizes the detection of CMB signals while effectively minimizing interference from galactic dust and other foreground sources. These bands were specifically chosen to balance the detection of distinct CMB features against the mitigation of astrophysical noise and contamination. Additionally, the selection of these frequencies strategically avoids the water vapor absorption line at 183 GHz, prevalent in the Earth's atmosphere, ensuring clearer and less attenuated signals.

This combination of advanced technologies ensures that SPT-3G not only meets but exceeds the requirements for precision in cosmological observations, providing unparalleled insights into the universe's earliest moments.

3. ATMOSPHERIC INFLUENCES ON BOLOMETER TIME CONSTANTS

Located at the South Pole, the SPT offers unique observational advantages including atmospheric stability, minimal atmospheric absorption, prolonged periods of darkness during the Antarctic winter, and reduced radio frequency interference, all of which significantly enhance the clarity and quality of the data collected. These geographical benefits are crucial for the bolometers' optimal performance, allowing precise measurement of faint cosmic signals.

A time constant, in the context of bolometer performance, measures the speed of a bolometer's response to changes in incident radiation, defined as the time required for the bolometer's response to reach about 63% of its final value following a sudden change in power. This parameter critically affects the bolometer's capability to accurately track rapid variations in the observed signal, with shorter time constants crucial for capturing high-frequency variations in cosmic phenomena such as the CMB and other astronomical events. Conversely, longer time constants can cause signal smearing, resulting in loss of information on smaller time scales.

Changes in elevation angle and season cause variations in time constants because the loading on the detectors changes is primarily due to the atmosphere. The amount of atmosphere the detectors observe varies with both the season and the elevation angle. During different seasons, atmospheric conditions like temperature and humidity change, affecting the atmosphere's transparency and, in turn, the detector loading. Similarly, at lower elevation angles, the detectors have to observe through a thicker layer of the Earth's atmosphere, leading to increased loading compared to higher elevation angles. These variations affect the detectors' response, necessitating careful planning of observations.

4. METHODS

This section details the methodologies employed in the automated pipeline developed for the SPT-3G survey data. Each method is designed to ensure the accurate collection, processing, and statistical evaluation of time constant data. The comprehensive approach integrates advanced computational techniques, robust statistical analysis, and clear visualizations to provide insights into bolometer performance.

4.1. *Data Collection*

The system initiates by traversing specified directories to collect time constant data files. It recursively searches through these directories and subdirectories to locate all relevant files. Files are identified using a regular expression pattern (`time_const_(\d+)_el_(\d+)deg.pkl`), designed to extract observation IDs (`\d+`) and elevation angles (`el_(\d+)deg`) from the filenames. This pattern ensures that only files corresponding to time constant data are collected. Each matched file's observation ID is converted to Modified Julian Date (MJD) format using the `obsid_to_g3time_mjd` function, facilitating precise temporal alignment with field observations. The elevation angle is also extracted and recorded along with the corresponding MJD date and file path, enabling organized data retrieval based on observational conditions.

4.2. *Grouping Bolometers*

Bolometer properties are loaded from a specified calibration file (e.g., `60000000.g3`). These properties include essential characteristics such as frequency band, wafer, and other identifiers. A wafer, in this context, refers to a thin slice of silicon based semiconductor material on which the bolometers are constructed. Each wafer typically contains an array of bolometers, and the characteristics of the wafer can influence the performance of the detectors. The grouping is defined using a function (e.g., `calibration.template_groups.get_template_groups`), which organizes bolometers into meaningful categories. Bolometers with incomplete properties or specific patterns (e.g., `'lc_resistor'`) are handled appropriately to ensure accurate grouping and subsequent analysis.

4.3. *Matching Observation Dates*

The pipeline matches field observation dates with the collected time constant data within a twenty-six minute window, ensuring temporal relevance. The collected time constant data is sorted by MJD, and for each field observation date, the closest matching time constant data is identified. This matching ensures that the most relevant observations are included in the analysis, minimizing temporal discrepancies. The matched data is organized by elevation angles, and any unmatched files are noted for further consideration.

4.4. *Processing Time Constant Data*

Each matched file is processed to extract time constant values for bolometers within the defined groups. The extraction involves loading the data from pickle files and accessing the first index of each `fit_val` values, which represent the time constants in seconds. For each bolometer group, the median time constant is calculated. The median is chosen as it is less sensitive to outliers compared to the mean, providing a robust measure of central tendency. The extracted median values are aggregated by frequency and elevation angle, enabling the calculation of overall statistical measures across different observational conditions.

4.5. *Statistical Analysis*

The statistical analysis involves calculating quartile values Q1 and Q3, as well as the Interquartile Range (IQR) for each frequency and elevation angle. These quartile values provide insights into the distribution and variability of time constants. The IQR, defined as the difference between Q3 and Q1, represents the range within which the central 50% of the data lies, offering a measure of variability. Error bars are derived from the IQR values, representing the uncertainty in the median time constants.

4.6. *Weighted Linear Regression*

To quantify the dependence of time constants at differing elevation angles, weighted linear regression is performed. The dependent variable (median time constant) and the independent variable (elevation angle) are defined, along with error bars derived from the IQR. Weights are computed as the inverse of the squared errors ($\frac{1}{y_{err}^2}$), emphasizing data points with smaller uncertainties. The regression calculates the slope and intercept of the best-fit line, providing a measure of the rate of change of time constants with elevation angles. The uncertainties in the slope and intercept are also calculated, providing confidence intervals for the regression parameters.

4.7. *Visualization*

Visualization plays a pivotal role in the analysis, utilizing graphical techniques to enhance the interpretation of complex datasets. The generation of combined histograms and quartile plots is performed using Python's Matplotlib library.

- **Combined Histograms**: These histograms depict the distribution of time constants across different frequencies, helping to identify patterns and outliers in the data. Bins are uniformly set to 20 bins to ensure detailed representation across the range of observed values. The x-axis is capped at five times the median value of the time constants for each frequency to focus on the most relevant range and to avoid distortion by extreme outliers. Histograms are annotated with median values and interquartile ranges, and distinct colors—chosen from a colorblind-friendly palette—differentiate each frequency, enhancing visual clarity and accessibility.
- **Quartile Plots**: These plots illustrate the median time constants against elevation angles, incorporating error bars derived from the IQR values to depict variability. The slope of the best-fit line from the weighted linear regression, included in the plot, indicates how time constants change with elevation, providing crucial insights into environmental or instrumental effects on measurement accuracy.

5. RESULTS

Based on the generated histograms and quartile plots, we can draw several conclusions about the time constants and their correlation with elevation angles:

- **Time Constant Values**: The histograms reveal distinct distributions of time constants across different frequencies, with median values and interquartile ranges serving as robust indicators of central tendency and variability. The relatively narrow distributions across frequencies support the decision to use a single representative time constant per subfield, suggesting that this simplification adequately captures the performance characteristics of the bolometers within each category. Specifically, the median time constants are 5.38 ms at 90 GHz, 5.84 ms at 150 GHz, and 3.55 ms at 220 GHz, highlighting notable variations across the different frequencies.
- **Elevation Angle Correlation**: The quartile plots and weighted linear regression analysis reveal a frequency-dependent correlation between time constants and elevation angles. For instance, at 150 GHz, the slope of the best-fit line was -0.00244 $\pm$ 0.0432 ms/degree, suggesting minimal to no influence, indicated by the wide confidence interval encompassing zero. Conversely, at 220 GHz, the slope was more pronounced at 0.0301 $\pm$ 0.0299 ms/degree, demonstrating a more substantial influence, as the confidence interval does not include zero. This observed variability implies that elevation angle effects must be considered in the calibration and data analysis processes, particularly when setting observational strategies for different frequency bands.

- **Consistency of Findings**: These conclusions indicate that while a single representative time constant per subfield is generally sufficient due to the narrow distributions across frequencies, the influence of elevation angle on time constants is more pronounced at higher frequencies like 220 GHz. This means that, although the overall stability of time constants allows for some simplification, adjustments must still be made based on elevation angles, especially at higher frequencies. It's important to note that there was only sufficient data to analyze three subfields: Wide-F, Wide-G, and Wide-I. This limitation suggests that while the findings are consistent within these subfields, further analysis across additional subfields would be necessary to generalize these results more broadly.

These findings highlight the need for frequency-specific approaches in handling bolometer data, considering the significant impact of observational conditions on measurement accuracy. The relatively stable time constants across subfields suggest that a generalized approach can be effective, but the noted variations with elevation angles, particularly at higher frequencies, underscore the importance of incorporating these factors into calibration and analysis processes. Future work should focus on further quantifying these effects and exploring their implications for the design and operation of millimeter-wave astronomical instruments.

6. FIGURES

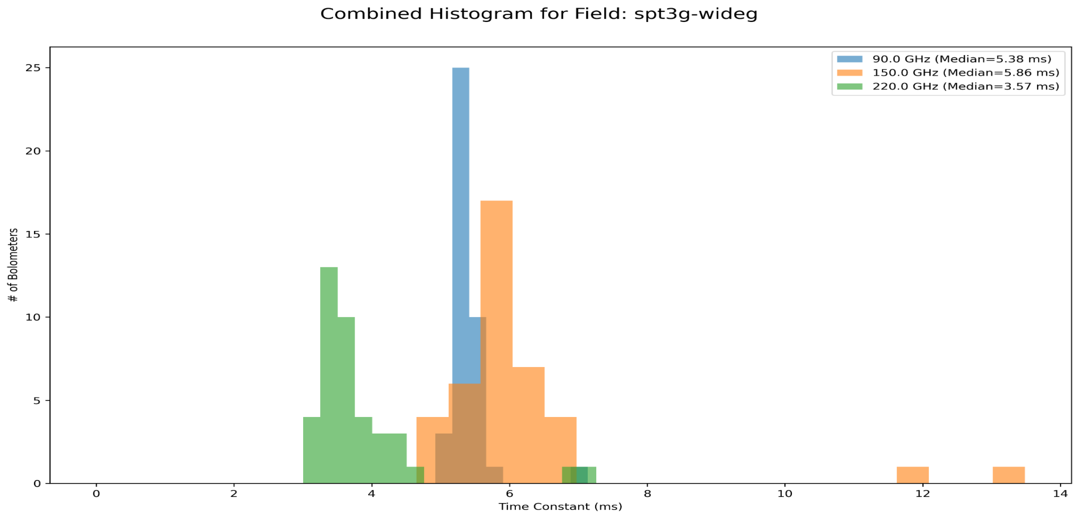

Figure 1. Combined Histogram for the wide-g subfield. This histogram shows the distribution of time constants for bolometers at different frequencies (90 GHz, 150 GHz, and 220 GHz) in the Wide-G subfield. The median time constants for each frequency are annotated, indicating the central tendency of the time constant values for each frequency band.

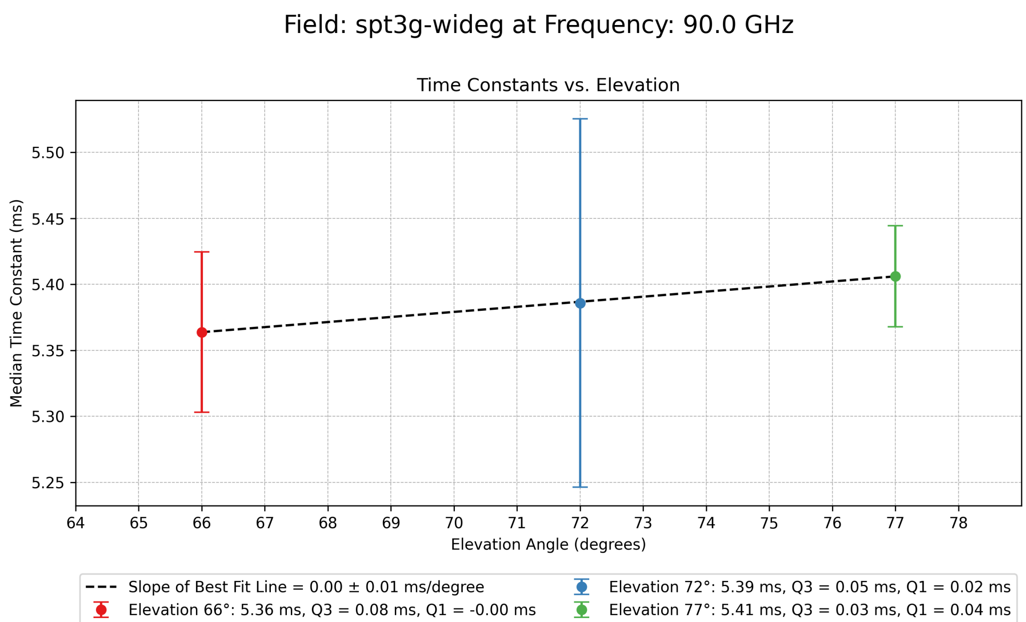

Figure 2. Wide-G subfield at 90.0 GHz. This quartile plot illustrates the median time constants at different elevation angles for the 90 GHz frequency band. The error bars represent the IQR, and the dashed line shows the best-fit slope, indicating the relationship between elevation angle and median time constant.

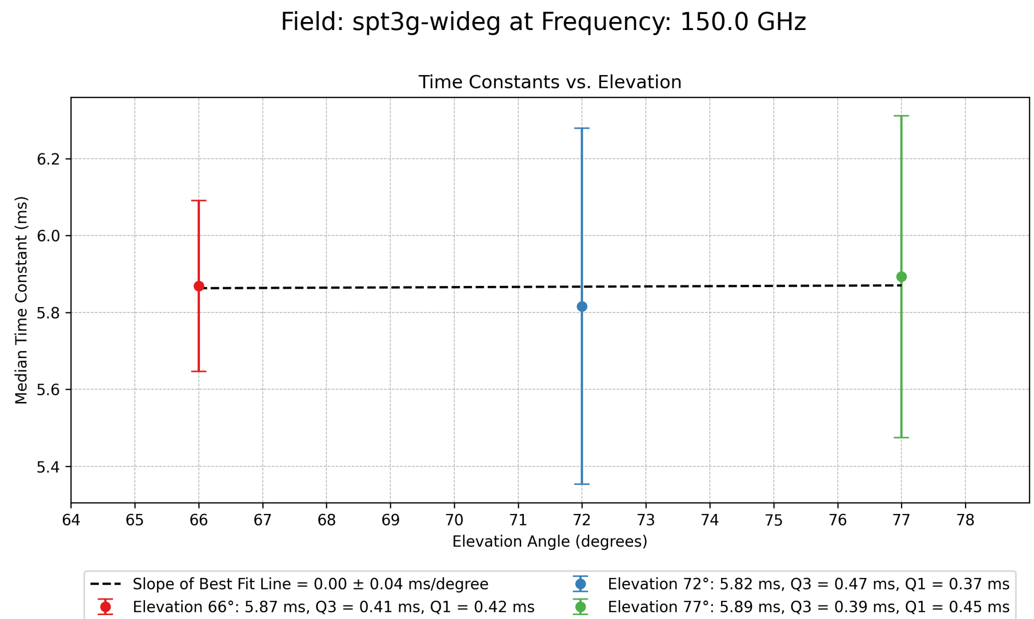

Figure 3. Wide-G subfield at 150.0 GHz. This quartile plot shows the median time constants at different elevation angles for the 150 GHz frequency band. The error bars display the IQR, and the dashed line represents the best-fit slope, illustrating the correlation between elevation angle and median time constant at this frequency.

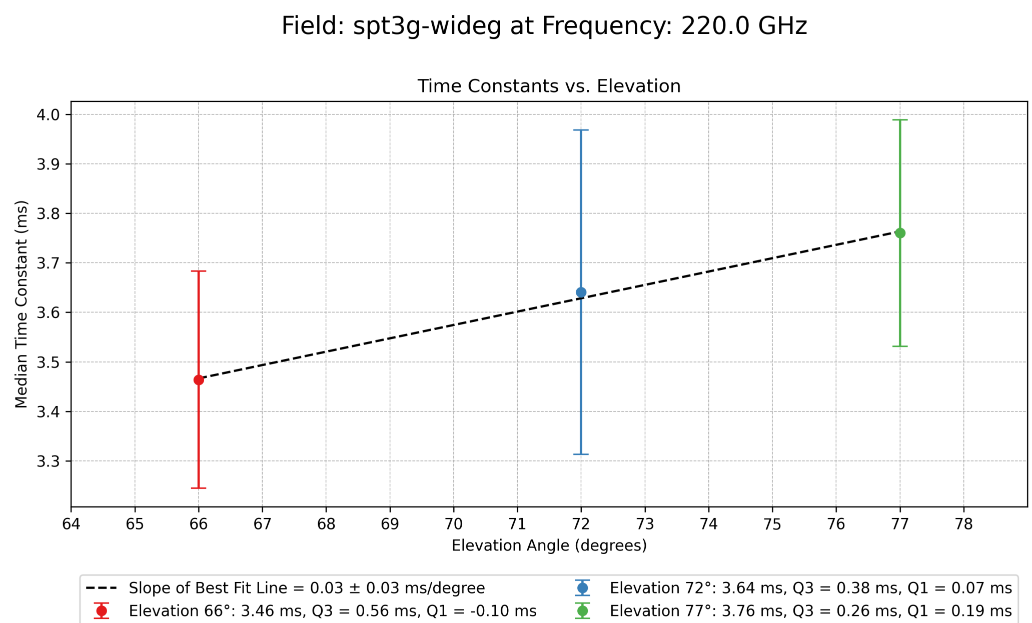

Figure 4. Wide-G subfield at 220.0 GHz. This plot highlights the impact of elevation angle on median time constant for the 220 GHz frequency, with the IQR and best-fit line slope annotated.

7. CONCLUSION

The automated pipeline developed for the SPT-3G survey data represents a significant advancement in the calibration and analysis of time constants. A crucial feature of the pipeline is its conversion of observation IDs to MJD and matching them with field observation dates within a stringent twenty-six minute window, ensuring temporal relevance and accuracy crucial for high-quality data analysis. The detailed grouping of bolometers based on their properties allows for analysis that accounts for variations in detector configurations, significantly enhancing the reliability of results.

The statistical techniques employed, such as quartile analysis and weighted linear regression, provide insights into the dependencies of time constants on elevation angles. These methods offer a robust measure of central tendency and variability, allowing for a comprehensive understanding of bolometer performance under different observational conditions. The visualization techniques, including combined histograms and quartile plots, are annotated to facilitate a thorough understanding of underlying patterns and trends, crucial for addressing any anomalies in bolometer performance.

Overall, this pipeline not only streamlines the data handling and analysis process but also significantly enhances the accuracy and reliability of bolometer calibration in cosmological observations. The robust statistical summaries and detailed visualizations generated are instrumental in advancing our understanding of bolometer behavior, ultimately contributing to the refinement of calibration techniques and improving the sensitivity of cosmological observations.

Future work could explore the application of this pipeline to other cosmological surveys and potentially to other fields within astronomy and astrophysics. Further refinements could involve integrating predictive analytics to anticipate bolometer performance, potentially reducing calibration time and improving data throughput. These advancements will pave the way for even more precise and comprehensive studies, underscoring the critical role of sophisticated computational tools in advancing modern astrophysical research.

8. ACKNOWLEDGMENTS

I would like to express my gratitude to everyone who contributed to this research and the development of the automated pipeline. Special thanks to Tom Crawford and Kyra Fichman for their invaluable support and feedback. As well as a thank you to many others including but not limited to my colleagues and coworkers, Juliet Crowell, Samar Alqatari, Sidney Nagel, Eileen Sheu, and a special thanks to the University of Chicago REU program and the NSF for funding this research.

Facilities: SPT, South Pole Telescope (SPT-3G)

Software: numpy (Harris et al. 2020), matplotlib (Hunter 2007), spt3g (Benson et al. 2014), Visual Studio Code

APPENDIX

A. SUPPORTING INFORMATION

A.1. *Methodological Details*

- **Weighted Linear Regression**: This analysis leverages weighted linear regression to account for variations in measurement reliability across elevation angles. By applying weights inverse to the square of measurement uncertainties, the regression model prioritizes more reliable data, providing a robust estimation of trends.

- **Quartile Analysis**: Quartile analysis is used to derive insights into the distribution of time constants, highlighting the median and interquartile range. This method is chosen for its robustness to outliers, offering a clear view of the central tendency and variability within the data.

A.2. *Data Structures*

- **TimeConstantData**: This custom data structure is pivotal for managing the collection of time constant measurements. It includes fields for storing the MJD, elevation angle, and the file path of the data source, facilitating precise data management and retrieval for analysis.

- **Bolometer Grouping**: Bolometers are grouped based on specific properties such as frequency bands and wafer locations. This structured grouping is crucial for analyzing variations across different configurations and enhancing the reliability of the analysis by comparing like-with-like wherever possible.

A.3. *Operational Flow*

The pipeline operates through several key phases:

- **Data Collection**: Data is systematically collected and vetted for completeness and accuracy.
- **Temporal Matching**: Time constant data is matched with field observations within a strict twenty-six minute window, ensuring the relevance of data points for analysis.
- **Data Processing**: Data is processed through statistical methods tailored to identify and quantify patterns in the time constants.

A.4. *Statistical Methods*

- **Error Analysis**: The analysis method calculates the standard error of the mean for each set of bolometer readings, which is then used to weight the linear regression, ensuring that data with lower variance has a greater influence on the model.
- **Regression Analysis**: The linear regression model is used to understand how time constants change with elevation, with the regression slope providing insight into the rate of change. While specific p-values are not calculated, confidence intervals derived from the standard errors offer insights into the statistical robustness of the observed trends.

A.5. *Visualization Techniques*

- **Histograms**: Histograms are employed to visualize the distribution of time constants across different setups, providing a straightforward depiction of data spread and central values.
- **Quartile Plots**: These plots display the median and interquartile ranges of time constants against elevation angles, enriched with best-fit lines from regression analysis to illustrate underlying trends.

A.6. *Future Work*

- **Extending the Pipeline**: Future modifications could include the integration of machine learning algorithms to predict time constants based on historical data, potentially increasing predictive accuracy and efficiency.
- **Addressing Limitations**: Further work could also explore the scalability of the pipeline to handle larger datasets or extend its application to other observational instruments and settings, broadening its utility in the field of astrophysics.

REFERENCES

Benson, B. A., Ade, P. A. R., Ahmed, Z., et al. 2014, in Millimeter, Submillimeter, and Far-Infrared Detectors and Instrumentation for Astronomy VII, Vol. 9153, 91531P, doi: 10.1117/12.2057305

Harris, C. R., Millman, K. J., van der Walt, S. J., et al. 2020, Nature, 585, 357, doi: 10.1038/s41586-020-2649-2

Hunter, J. D. 2007, Computing in Science & Engineering, 9, 90, doi: 10.1109/MCSE.2007.55