Final Project Update
April 8, 2024

Anomaly Detection with Convolutional Autoencoders

Exploratory Analysis

We have completed the preliminary analysis for the datasets, providing us with a comprehensive understanding of the nature of the data and the outcomes achieved. Our next goals are to build a convolutional autoencoder network and run it on these datasets. We want to play with different combinations of hyperparameters so that we can get the most optimal trained model that will learn to detect anomalies with high accuracy, low bias, and without overfitting.

Dataset Descriptions:

Bus Voltage: The dataset represents a univariate time series containing approximately 1.8 million data points. Each data point reflects a bus voltage reading captured at a 5-minute interval. To enhance computational efficiency, the data was down-sampled to include readings captured every 8 hours, resulting in a dataset of 18,807 rows. Rows with missing values were excluded from the analysis. The data exhibits a predictable annual voltage dip. Additionally, the trend remains constant until encountering a change point around the beginning of 2008. This change point is characterized by a spike in voltage followed by a subsequent downward trend.

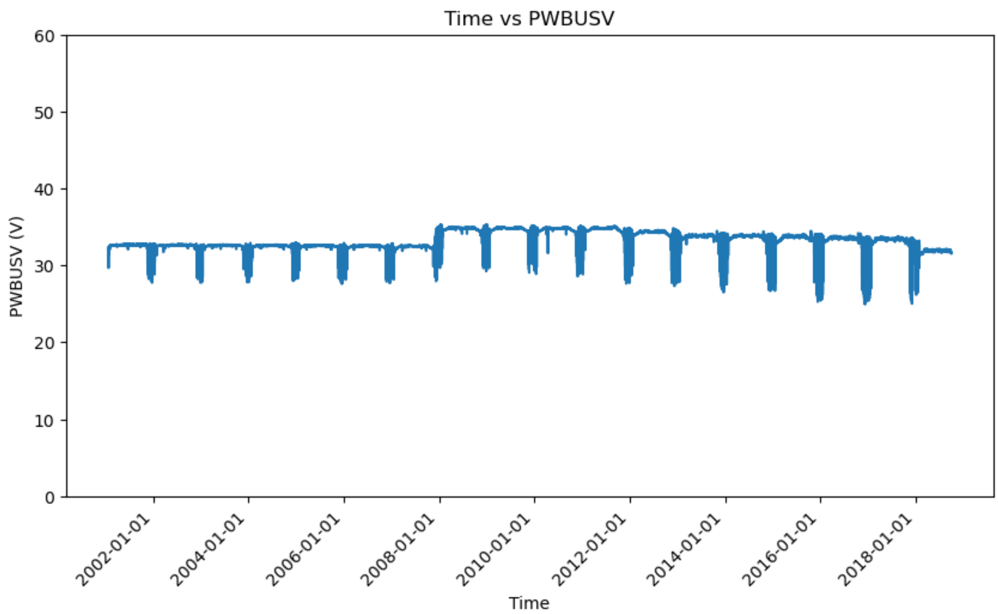


Figure: Time vs PWBUSV plot

Exploratory Data Analysis

Volatility analysis

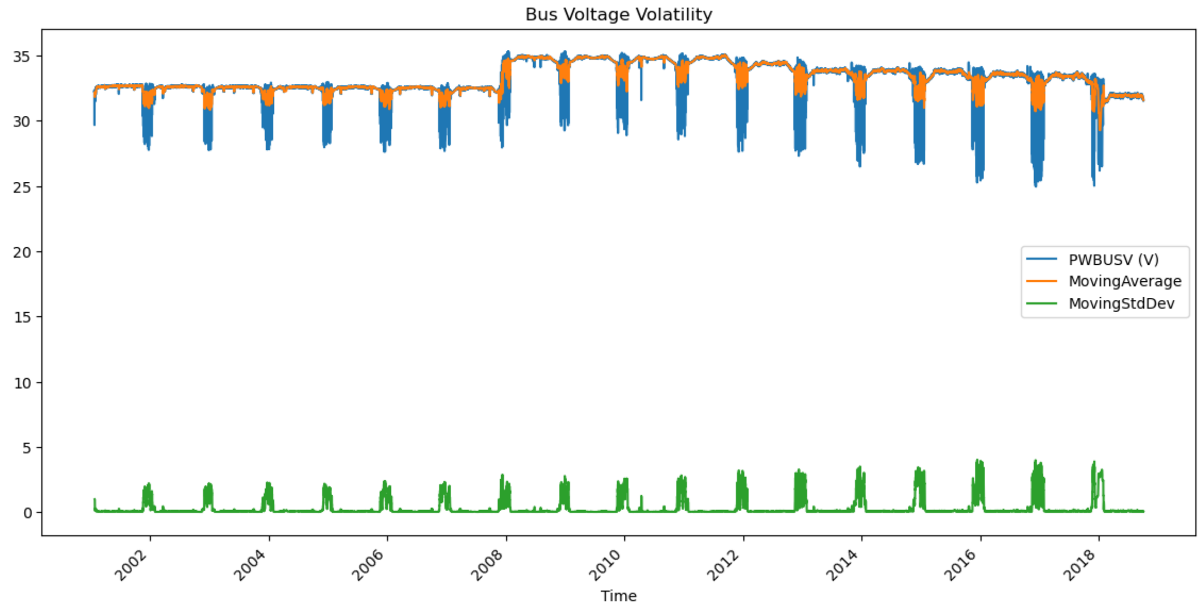


Figure: Bus Voltage Volatility

Bus Voltage (Blue Line): This represents the actual voltage readings. The line seems quite steady, with occasional dips which might be due to normal operational events or could indicate potential issues or anomalies.

Moving Average (Orange Line): This line is smoother than the actual voltage readings, indicating it's averaging out the fluctuations to show the underlying trend. It seems to closely follow the actual voltage readings, suggesting the voltage doesn't have dramatic changes over the long term.

Moving Standard Deviation (Green Line): This line represents the volatility of the voltage readings – how much they vary over time. Spikes in the moving standard deviation indicate periods where the voltage readings were more variable. The relatively low and steady line, interspersed with spikes, suggests that the bus voltage is generally stable, with occasional periods of volatility.

From this plot, one could infer that the Bus Voltage is generally stable, with predictable behaviour most of the time and occasional periods of higher variability. The times at which the green spikes occur may be of particular interest for further investigation as they represent periods where the voltage varied more than usual.

Seasonal Decomposition:

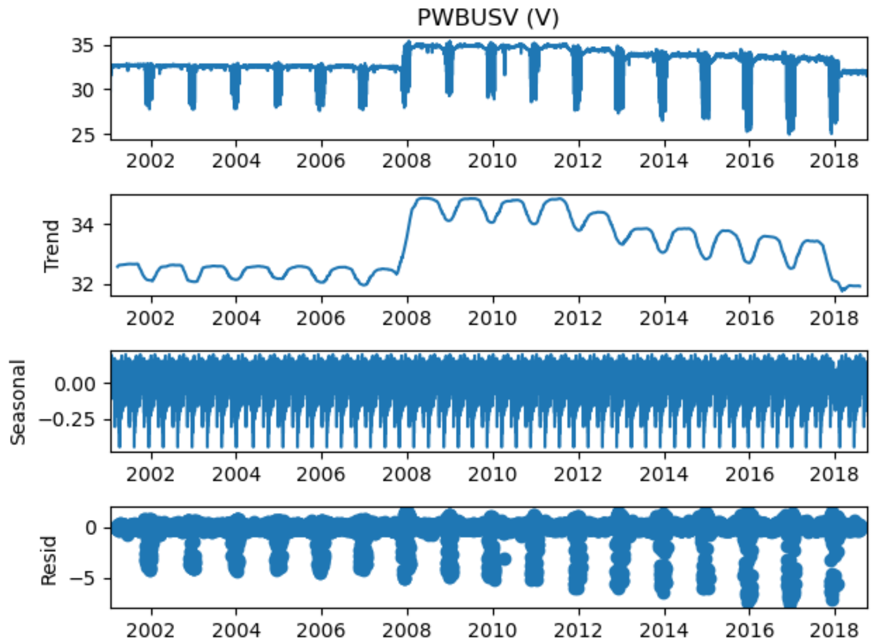


Figure: Seasonal decomposition of Bus Voltage data

Trend: The trend component of the graph shows a gradual increase in bus voltage over time. This could be due to a number of factors, such as increasing demand for electricity or improvements in the power grid.

Seasonal: The seasonal component of the graph shows a cyclical pattern that repeats over a year. The voltage is highest in the summer and lowest in the winter. This suggests that there is a seasonal effect on bus voltage.

Residual: These are the irregularities or 'noise' in the data after removing the trend and seasonal components. It represents the randomness or anomalies that are not explained by the seasonality or trend. The 'Residual' subplot shows that there are some fluctuations with occasional spikes and dips, which may represent anomalies, unexpected variations, or could be noise.

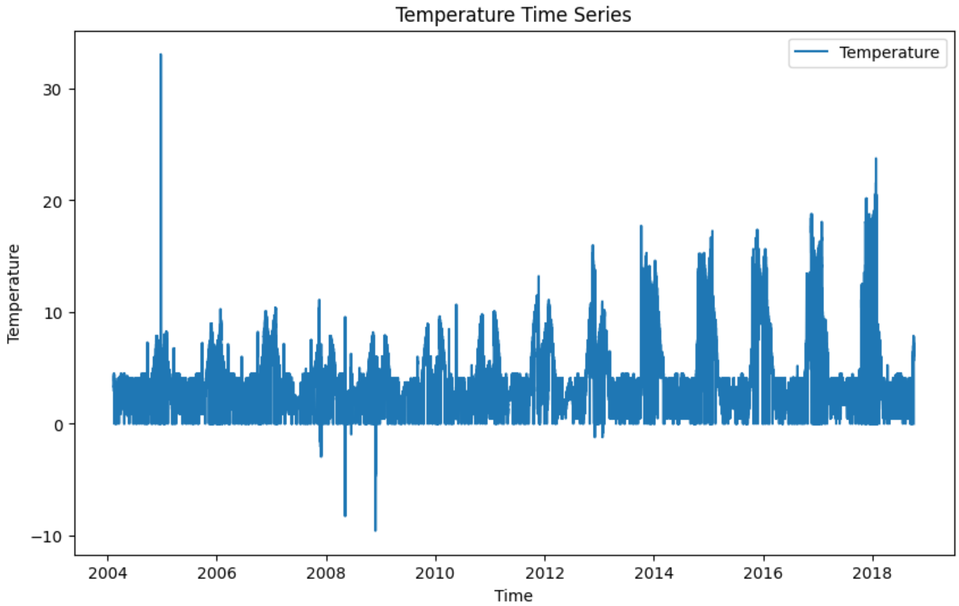
Dataset Descriptions:

Battery Temperature: The dataset represents a univariate time series containing a little over 1.5 million temperature readings, recorded between February 2004 to October 2018, in 5 minutes interval. In this context, the batteries power the satellite's systems and are critical component of spacecraft health monitoring for efficiency, safety, and operational decisions. A deviation from the expected temperature range might indicate an anomaly in the satellite and might require maintenance – it could also just as well be an expected behavior for other reasons but deemed anomalous because the observation is too unique. All important information to gather!

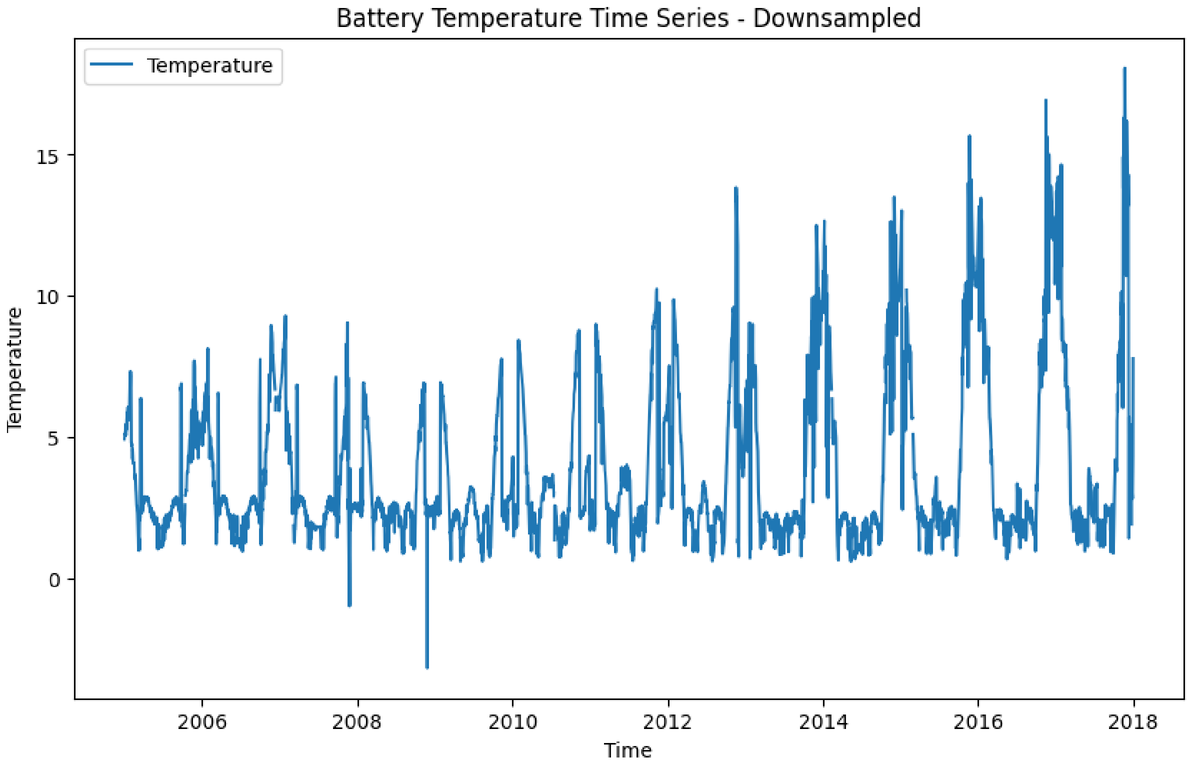
Exploratory Data Analysis

```
       TCPV6T (C)
count  1.518180e+06
mean   3.802520e+00
std    3.055154e+00
min   -9.578091e+00
25%    1.708204e+00
50%    2.911223e+00
75%    5.102928e+00
max    3.306738e+01
```

Time series data that displays the temperature readings, recorded every 5 minutes.

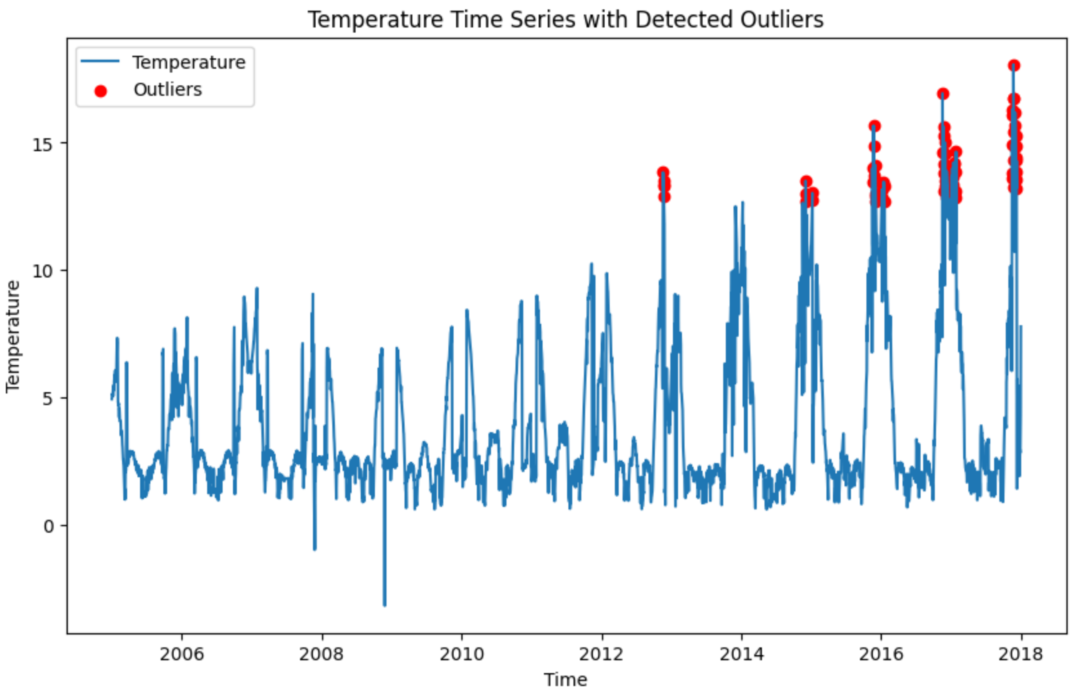


Since this is a big dataset, we down sampled it and got only one observation, daily, calculated by getting the mean of all the readings of that day. We also included the years, inclusively, between 2007 and 2017 because 2006 and 2018 have data for only few months of the year.

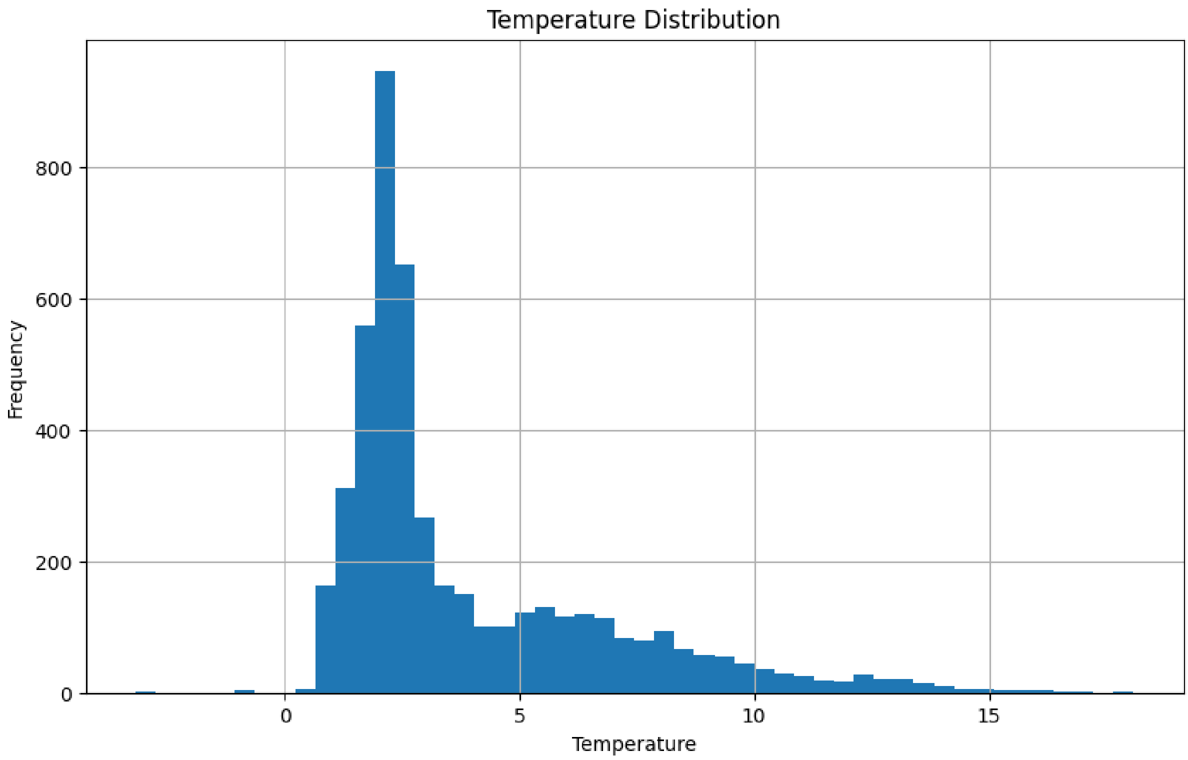


By getting the z-score of each value in the series, where z-score = (x - mean) / std, we get 83 observations that are outliers, assuming z-score > 3 is an outlier.

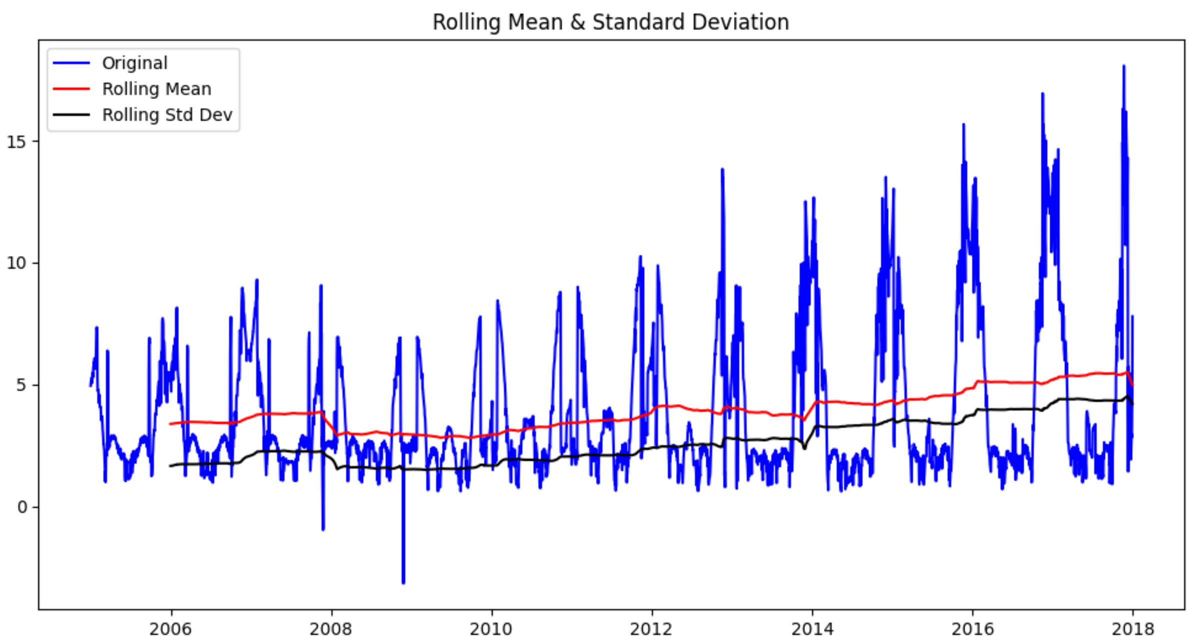
Following shows the time series data with the outliers marked by red dots.



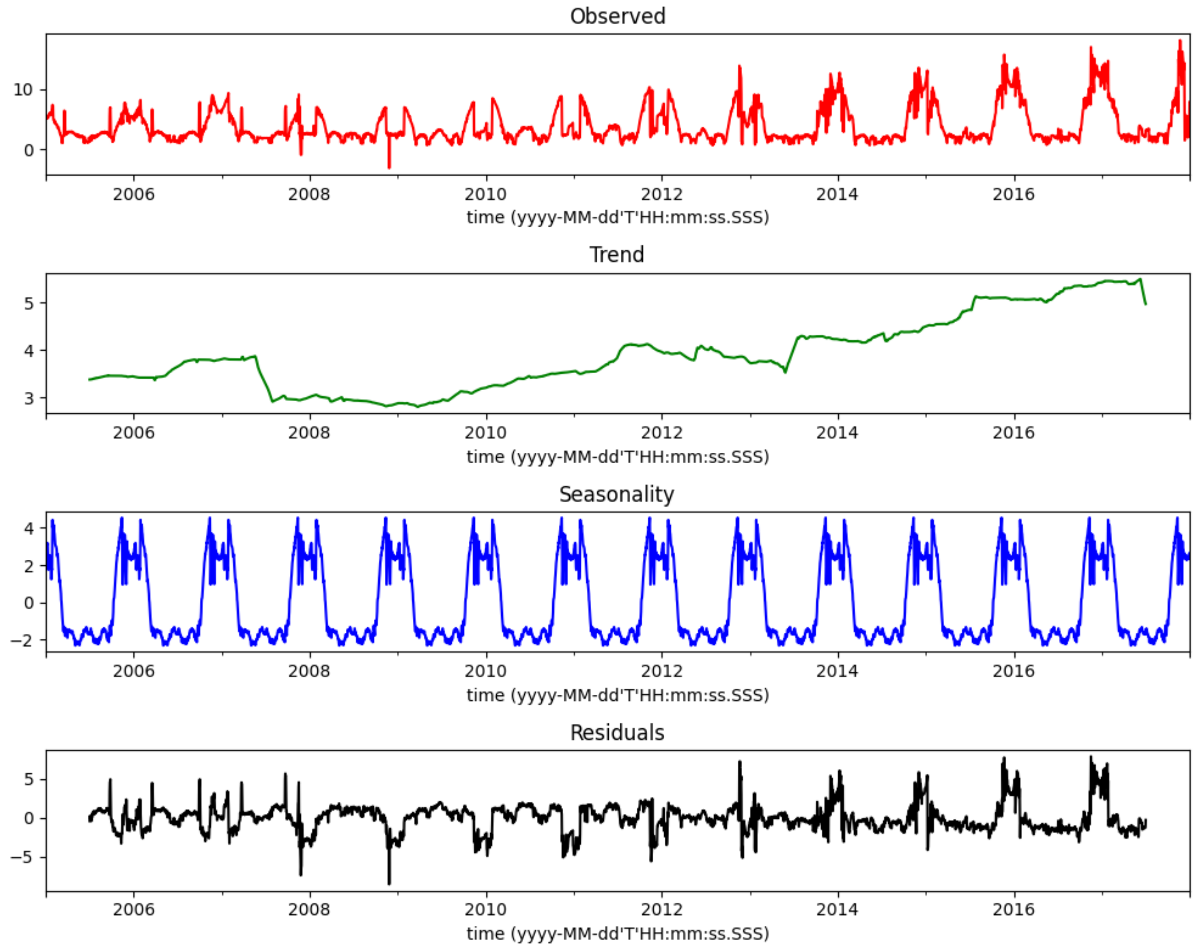
Temperature Distribution for understanding:



The dataset's rolling mean and standard deviation against the original time-series plot, with a window size of 365 for trends on an annual basis:

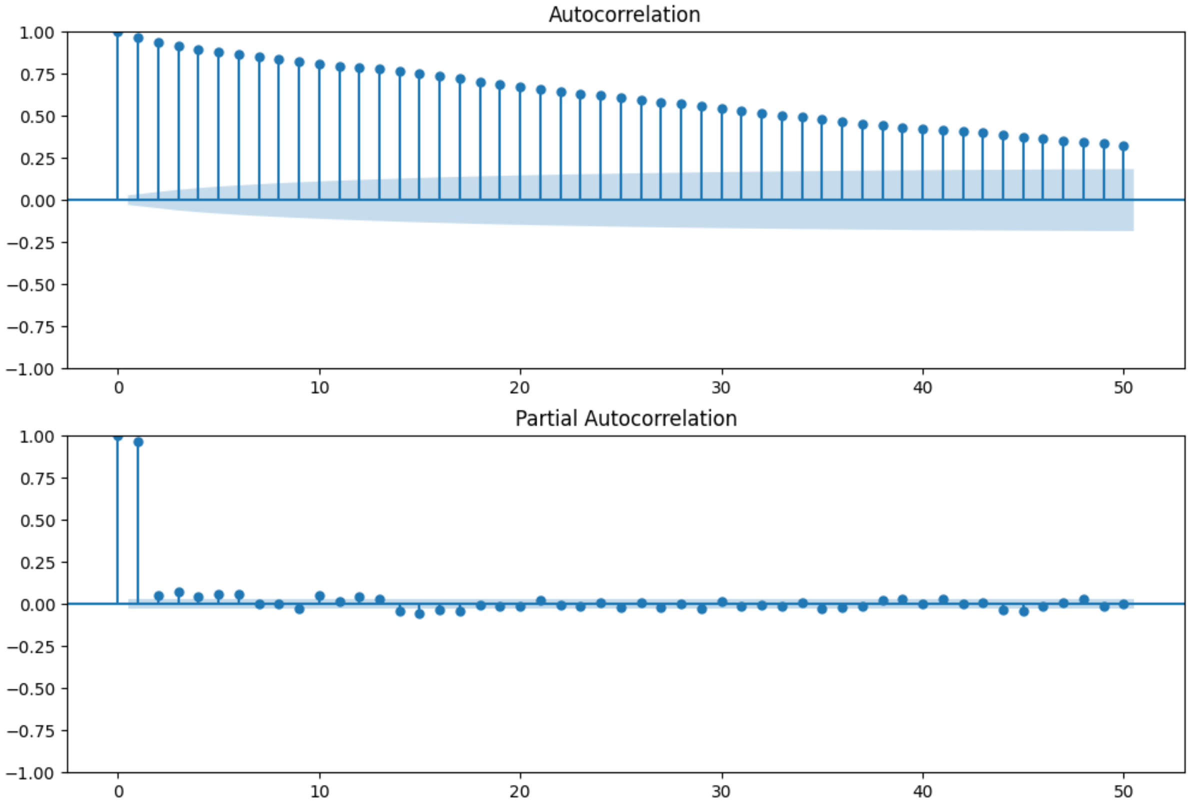


Seasonal trends are displayed below for further insights on the data.



To further understand the dataset, we get the Augmented Dickey-Fuller (ADF) statistics to determine if the time series data is stationary. We found that our dataset indicates stationarity as it is -6.41, suggesting that the null hypothesis (the time series is non-stationary) can be rejected. Moreover, our p-value of approx. 1.91×10^{-8} is very small, further suggesting that we can reject the null hypothesis that the dataset is non-stationary. We can reject the null hypothesis at all the conventional significance levels (1%, 5%, and 10%) of critical values, and the series does not appear to have a unit root based on the test. The ADF statistic is less than all the critical values and thus, it strongly suggests that the time series is stationary and fit for modeling using techniques that assume stationarity.

Following displays the Autocorrelation Function (ACF) and the Partial Autocorrelation Function (PACF) plots - they show the correlation of the series with its lags to identify seasonality in the data and the partial correlation of each lag to identify the order of an autoregressive model, respectively.

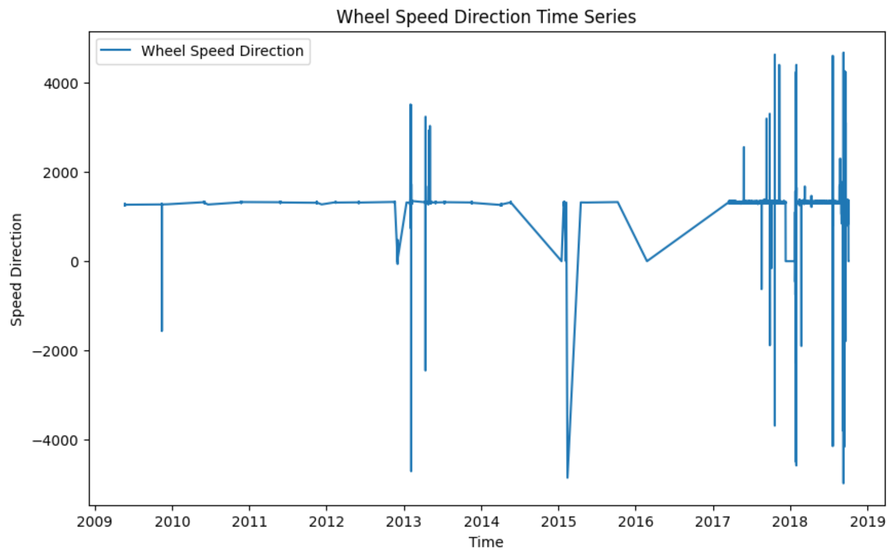


Dataset Description:

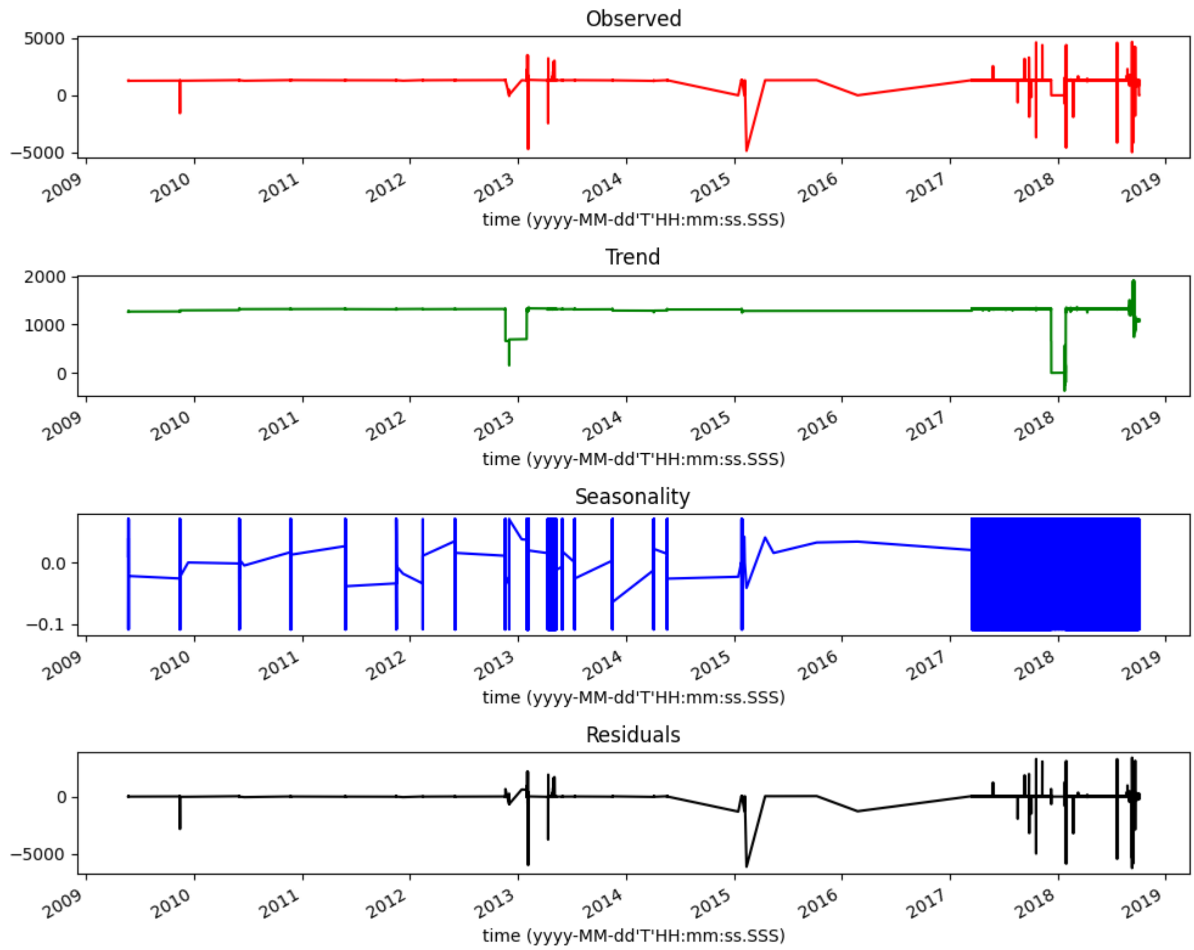
Reaction Wheel RPM (ADW1SDIR (RPM)): This dataset constitutes a comprehensive univariate time series consisting of approximately 48,865,494 RPM readings, observed over a span of roughly 9 years from May 2009 to October 2018. Each entry captures the rotation per minute of a reaction wheel within a spacecraft, a critical component for attitude control. The dataset showcases a high-resolution data collection with a very dense sampling rate.

Exploratory Data Analysis

Following shows the timeseries chart of the dataset.



Seasonal trends are displayed below for further insights on the data.



Dataset Descriptions:

Reaction Wheel Temperature: Temperature Data (TRW1MT (C)): This univariate time series dataset comprises approximately 1.5 million data points, each representing the temperature measured at a 5-minute cadence following mean resampling from an original 1-second cadence. After cleaning to exclude missing values and removing data points where the temperature was recorded as zero (indicating inactivity of the reaction wheels), the dataset still presents a substantial amount of data for analysis. The temperature readings are highly variable and show no apparent signs of a predictable pattern or seasonality.

Exploratory Data Analysis

Volatility analysis

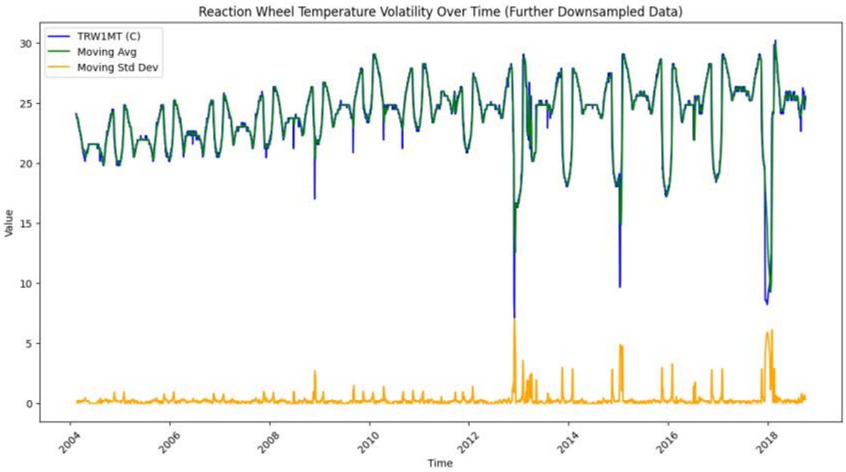


Figure: Time vs Value plot

Temperature Data (Blue Line): This line represents the actual temperature readings, fluctuating around a certain range most of the time. The plot exhibits periodic spikes and sharp drops that punctuate an otherwise consistent pattern, potentially corresponding to specific operational scenarios or anomalies.

Moving Average (Green Line): The smoother green line demonstrates the moving average of the temperature, which serves to soften the fluctuations and more clearly illustrate the longer-term trends or cycles in the data.

Moving Standard Deviation (Orange Line): This line indicates the level of volatility in the temperature readings. A spike in the moving standard deviation suggests a period where the temperature readings showed greater variation. The plot shows these spikes are intermittent, implying that while the temperature is generally stable, there are episodes of significant fluctuation that may be of interest for further exploration.

From this visualization, we can deduce that the temperature is generally consistent, with intermittent periods of increased volatility. These periods of greater variation, denoted by the peaks in the orange line, could be indicative of noteworthy operational events or system responses that require further investigation.

Seasonal Decomposition:

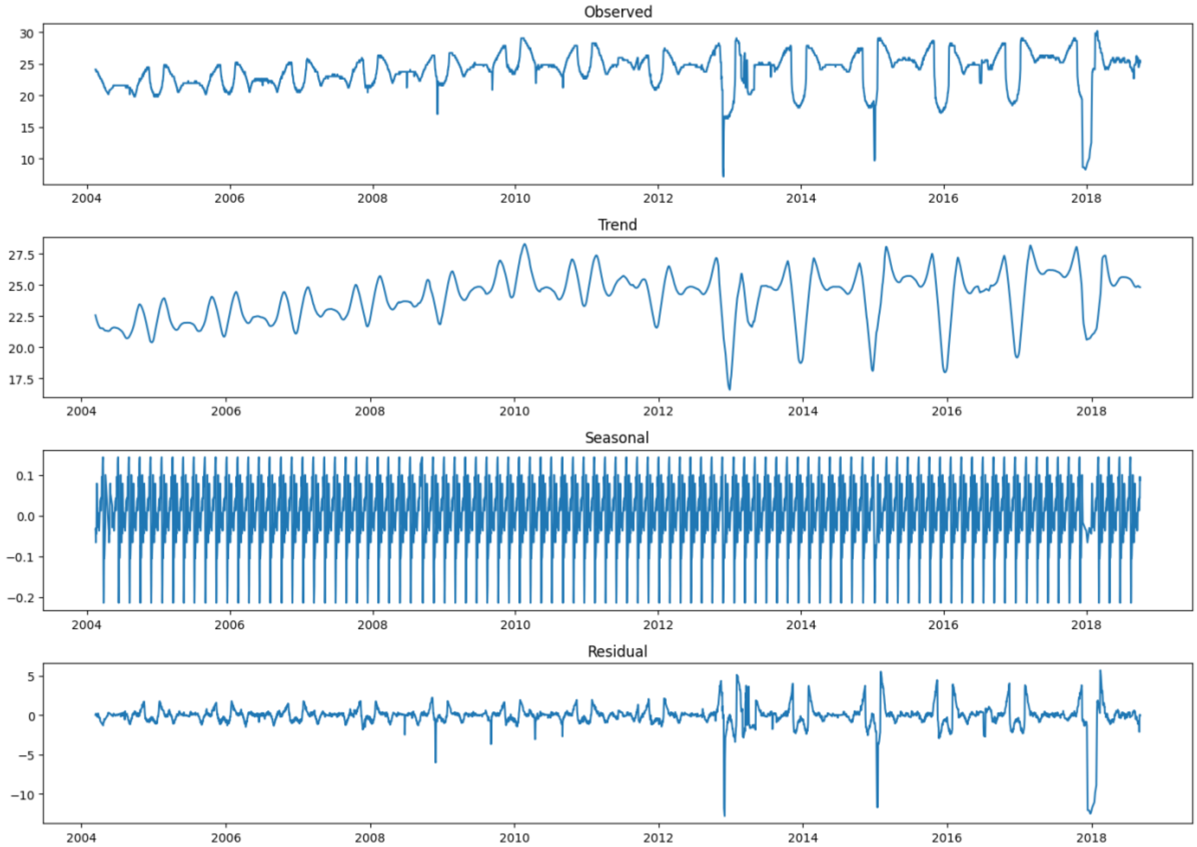


Figure: Seasonal decomposition of Reaction Wheel Temperature

Observed (Top Plot): This plot shows the actual temperature data as it was recorded. It exhibits fluctuations with a relatively consistent range, interspersed with sharp spikes and dips, which could be indicative of specific events affecting the temperature.

Trend (Second Plot): The trend component illustrates a long-term movement in the temperature readings. Despite some fluctuations, there's no clear long-term upward or downward trend, indicating general stability in the temperatures over the years.

Seasonal (Third Plot): The seasonal plot shows a consistent pattern repeating at regular intervals, indicative of a strong seasonal influence on temperature. This regularity suggests predictable behavior in temperature changes that could be linked to seasonal variations in operational activities or environmental factors.

Residual (Bottom Plot): Residuals represent the noise or random fluctuations once the trend and seasonal components have been accounted for. In this plot, the residuals are generally small, indicating that the trend and seasonal components explain much of the behavior of the temperature data. However, there are occasional spikes in residuals which might be due to anomalies or irregular events that the model cannot explain.

From this decomposition, we can infer that while the temperature data has a significant seasonal pattern and generally does not exhibit a clear long-term trend, there are sporadic anomalies that could be of interest for further investigation. These findings can guide maintenance decisions, inform operational adjustments, or direct further inquiry into specific periods of interest.

Dataset Descriptions:

Total Spacecraft Bus Current: This dataset, comprising roughly 1.5 million entries, represents a univariate time series of bus current readings captured every 5 minutes. For analysis efficiency, the data was down sampled to include only readings every 8 hours, resulting in a more manageable dataset. Missing values and zero readings, which may represent inactive periods, have been excluded to focus on active operation times. The data reflects a stable pattern of current usage with occasional surges that could signify operational adjustments or system anomalies.

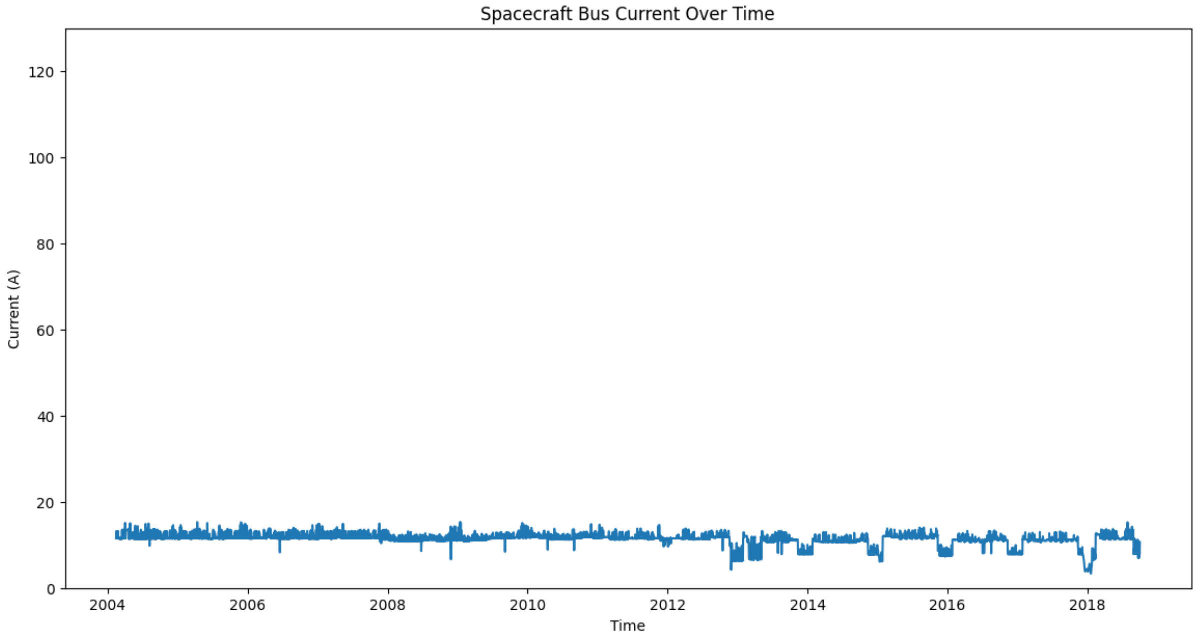


Figure: Spacecraft Bus Current Over Time

Exploratory Data Analysis

Volatility analysis:

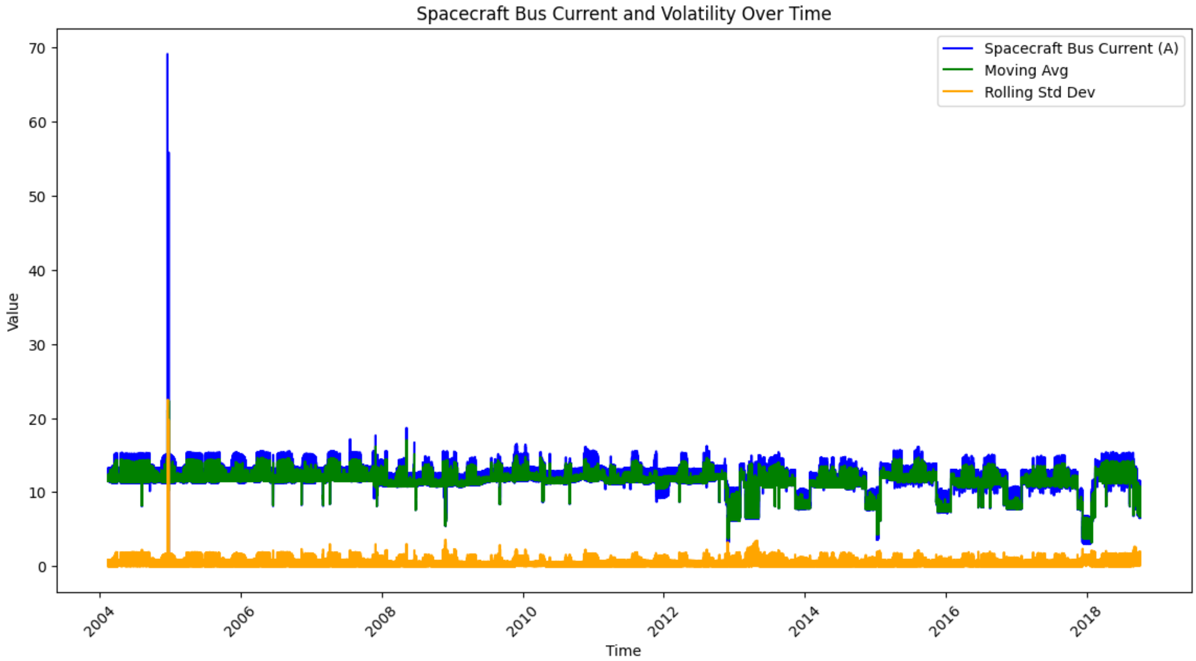


Figure: Spacecraft Bus Current Volatility

Spacecraft Bus Current (Blue Line): The primary data line shows the spacecraft's bus current readings, which appear relatively steady and consistent, apart from a few significant spikes. The graph demonstrates a baseline current usage, punctuated by occasional surges in power draw.

Moving Average (Green Line): Contrary to the initial statement, the moving average is depicted on the graph. The green line, which is smoother than the blue line of raw data, indicates an averaging out of the short-term fluctuations to reveal the underlying, more stable pattern of current usage.

Rolling Standard Deviation (Orange Line): The orange line represents the rolling standard deviation, serving as a measure of volatility. The spikes in this line signify periods when the current readings showed greater variability**, highlighting instances of increased electrical activity or potential anomalies.**

Observations:

The overall pattern suggests a spacecraft bus system characterized by a relatively predictable and stable operational mode, as evidenced by the moving average which follows closely with the actual current readings.

The periodic spikes in the rolling standard deviation (volatility) may correlate with specific operational events, such as maneuvers or the activation of certain equipment, which temporarily increase the current demand.

The most significant volatility spikes, particularly the one at the beginning of the time series, could be the result of a singular event or anomaly and might warrant a deeper investigation into the system's logs or any external influences at those times.

Seasonal Decomposition:

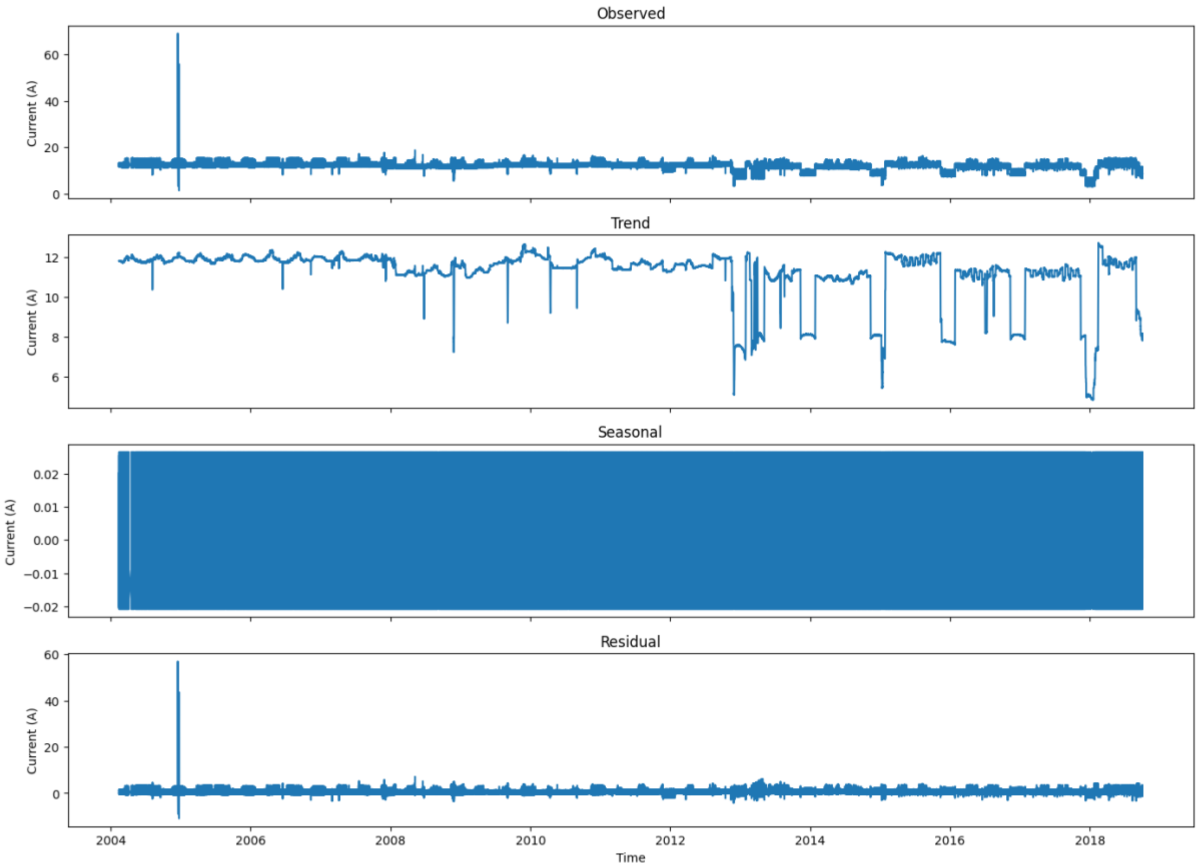


Figure: Seasonal Decomposition of Spacecraft Bus Current Data.

Observed (Top Plot): The observed plot reflects the raw spacecraft bus current data. There is one prominent spike at the beginning, but otherwise, the current remains relatively consistent over time.

Trend (Second Plot): The trend component shows more variation than the observed data. This indicates that when seasonal effects are accounted for, there are underlying changes in the current usage over time. The trend line dips and peaks periodically, which may correspond to long-term operational cycles or system changes.

Seasonal (Third Plot): This component should represent any regular pattern that repeats at fixed intervals. However, the plot appears to be a solid block of color, which suggests the seasonal fluctuations are too small relative to the scale of the plot to be discernible. This could indicate that the bus current does not have strong seasonal variability, or the chosen period may not match the actual seasonal cycles present in the data.

Residual (Bottom Plot): The residual plot shows the noise or irregularities after both the trend and seasonal patterns have been removed. These are minimal and scattered, indicating that the trend and any undetected seasonal components account for most of the variability in the bus current. The few spikes that do appear could be anomalies or unusual operational events.

From this decomposition, one might infer that the spacecraft bus current does not exhibit a strong seasonal pattern but does have some long-term trends and occasional anomalies. The relatively quiet residual component suggests a stable system with few unexplained fluctuations, which is a good sign for spacecraft operations. However, the seasonal plot's lack of visible patterns warrants a review of the decomposition's period selection to ensure it aligns with the expected seasonal cycles. If a different period does not yield a clearer seasonal pattern, it might confirm that bus current behavior is not heavily influenced by seasonality.