Martian Weather Forecasting and Anomaly Detection

Using Time Series Models on NASA REMS Data from the Curiosity Rover

Tristan Peters

May 28, 2025

Contents

1 Introduction

1.1 Project Motivation

Weather prediction is difficult even here on Earth. What about weather prediction on Mars? Even more so. Mars is plagued by seasonal planet-wide dust storms which have a tendency of eroding and hampering the ability for equipment to continue their intended functions. These are missions that require years of intense planning, research, and maintenance on top of hundreds of millions or billions of dollars. It is reasonable then to desire that our equipment should last as long as possible. Every second a mission is extended is money saved.

The primary reasons terrestrial missions on Mars end is due to dust buildup on and inside of the equipment. If a rover's solar panel gets covered in dust or if an internal mechanism gets jammed there will not be anyone to come service the equipment. It then becomes only a matter of time before the mission ends once hardware malfunctions begin. Thus, if one could gain insight into the atmospheric conditions surrounding the equipment it might be possible to predict when potentially catastrophic dust storms might arise so that measures can be taken in advance to shield sensitive equipment which could extend the duration of the mission.

While this project does not offer insight into predicting when dust storms may arise, it does offer insight into how one might go about constructing such a pipeline by predicting the temperature of the atmosphere and attempting to discover anomalies in said prediction where unusual shifts in weather may occur. After all, weather events are primarily driven by the movement of cool and warm air, so predicting the temperature is an excellent place to start our investigation.

1.2 Objective Summary

Before we begin, it is important to set expectations and outline a reasonable scope for this investigation. In this report we will attempt to address the following goals for this investigation:

- Forecast average ground temperature on Mars.
- Detect atypical atmospheric behavior using hybrid techniques.
- Enable data-driven insights for scientific monitoring or onboard diagnostics.

If we can achieve these goals we will have established solid foundations for future development of more complex monitoring systems.

2 Dataset Overview

2.1 Source and Structure

While there are a number of approaches we could take we will be using the Remote Environmental Monitoring Station (REMS) dataset [3], collected by NASA using the Curiosity rover in Gale Crater. Gale Crater is a region near the equator on the Southern Hemisphere so the weather data should be fairly well behaved and consistent year by year. While we could source the data directly from NASA archives in a raw or filtered format, for this investigation we will be using a more pre-processed dataset compiled by "DEEP CONTRACTOR" on Kaggle[1]. This dataset offers a number of features such as:

- **Solar Date (Sol)**: Martian day beginning at mission start date
- **Solar Longitude (L_S)**: Position of Mars in Sun-centered frame
- **Temperature**: Min/Max air and ground temperatures in Celsius
- **Mean Atmospheric Pressure (Pa)**: Average atmospheric pressure measured over a day
- **UV Radiation**: A categorical feature measuring the relative UV index for the day
- **Sunrise/Sunset**

While there are certainly more features we could consider, this study will be limited to these features to keep our model simple.

2.2 Temporal Framing

Understanding Martian weather requires careful handling of time-based features. Although a Martian day is similar in duration to an Earth day, minor differences accumulate over time, making standard datetime methods in Pandas impractical for analysis. To address this, our dataset uses the Martian Solar Date (Sol) as the index, starting from Sol 1 at the commencement of the Curiosity rover's mission. This provides a consistent, mission-specific reference frame for tracking Martian days.

However, indexing by sol alone doesn't convey seasonal or annual contexts clearly. Unlike Earth, Mars lacks a standardized calendar widely adopted for scientific analysis. To resolve this, we utilize Mars' solar longitude (L_S), defined as the angle between Mars and the Sun, measured from the planet's northern hemisphere spring equinox [2]. Each increment of 90° in L_S corresponds precisely to a Martian solstice or equinox, offering an intuitive and scientifically meaningful approach to understanding seasonal variation.

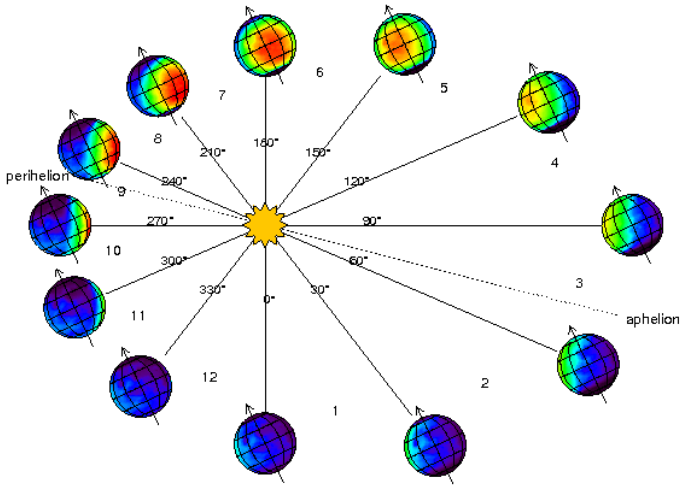

Figure 1: Martian Solar Longitude[2]

By incorporating solar longitude, we can explicitly relate observed weather patterns to Mars' seasonal cycles, enhancing both interpretability and reliability of our analyses. Because the Curiosity rover is in the southern hemisphere, the seasons are inverted from their northern counterparts. Below is a table detailing how solar longitude corresponds to seasons in the southern hemisphere:

Solar Longitude (L_S)	**Season (Southern Hemisphere)**
0° to 89°	Autumn (Fall)
90° to 179°	Winter
180° to 269°	Spring
270° to 359°	Summer

Table 1: Seasonal divisions on Mars based on solar longitude in the southern hemisphere.

3 Data Cleaning and Structuring

The dataset by "DEEP CONTRACTOR" [1], while useful, leaves much to be desired. Before this dataset is useful, a significant portion of time will need to be spent on data cleaning and basic feature engineering. Below is a list of general data cleaning tasks needed to be performed:

3.1 Cleaning Objectives

- Handle missing values and placeholder strings.
- Engineer temporal features (e.g., Martian month, season, year, day length).
- Clean feature names
- Remove or standardize unreliable or redundant features.
- Adjust datatypes

3.2 Gap Detection and Imputation

Initial data inspection revealed two primary data integrity issues: discontinuities in the sequence of solar dates (missing Sols) and missing values within recorded Sols. The specific reasons for these gaps are unknown. A comprehensive log detailing all non-continuous solar dates, including their start, end, and duration, is available in the `data/cleaned/` directory of the project's GitHub repository. While a deeper investigation into these discontinuities is beyond the current scope, it represents an area for potential future work.

For missing data within existing Sol entries, imputation was performed as follows:

- **Temperature and Pressure:** Missing numerical values in these columns were filled using bi-directional linear interpolation. This method provides reasonable estimates for missing data points based on their neighbors. Given the relatively infrequent occurrence of such missing points, the potential for this method to introduce significant bias or error into the models is considered minimal for this study's objectives.
- **UV Radiation:** As a categorical feature, missing UV radiation values were imputed using a rolling mode calculated over a 30-Sol window. This approach was chosen to realistically preserve the inherent seasonality of UV exposure without introducing excessive bias from a simpler imputation method like global mode.

3.3 Cleaning Outcome

Fields such as humidity, weather, speed, earth date time were dropped due to irrelevancy. The weather feature only contained the value "Sunny" across all entries in the dataset. Humidity and wind speed were similarly dropped due to every entry containing no data. Earth Date time was dropped due to irrelevancy. This study was concerned with only Martian weather, so the date on Earth does not matter.

We are left with the following features.

1. **Solar Longitude (L_S)**: Mars' position in its orbit.
2. **Sunrise, Sunset**: Times for sunrise and sunset.
3. **Min/Max Ground Temperature**: Ground temperature extremes.
4. **Min/Max Air Temperature**: Air temperature extremes.
5. **Mean Air/Ground Temperature**: Average air and ground temperatures.
6. **Martian Month**: Martian month based on solar longitude (30 degrees per month)
7. **Martian Year**: Martian year based on mission start (Initial year = 1)
8. **Martian Season**: Martian season based on solar longitude.
9. **Day Length**: Length of the Martian day in minutes.
10. **Mean Pressure**: Average atmospheric pressure for a given day.
11. **UV Radiation**: UV index categories.

We are left with **3,197** entries. There is no missing data or incorrectly formatted entries for any feature in any row, however there are gaps between days occasionally, with the largest gap being 29 days between sol 2,175 and 2,203. Aside from this, we are now ready to begin analysis.

4 Exploratory Data Analysis

There are certainly a plethora of interesting relationships and features to investigate when it comes to weather on Mars. By visualizing key atmospheric variables over time (Martian sols) and against solar longitude, distinct patterns were revealed which were further substantiated by statistical testing.

4.1 Atmospheric Trends

- **Air Temperature**:
  Air temperature on Mars, as recorded by Curiosity, exhibits significant variation and volatility. Distinct seasonal cycles are clearly present, with temperatures peaking during the Martian summer and reaching their lowest points in winter, which are analogous to terrestrial seasons. For instance, the temperature ranged from -100.00 °C in winter to 24.00 °C in summer.

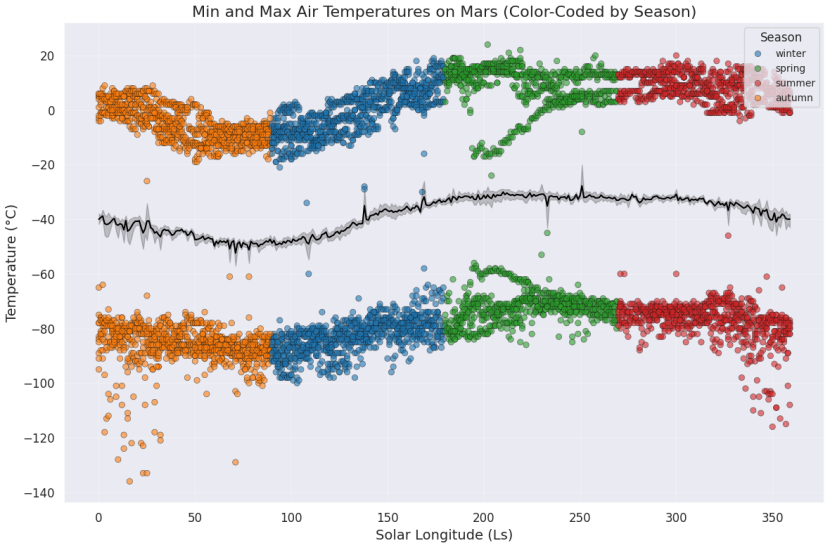

Figure 2: Annual Air Temperature by Season

By applying an Augmented Dickey-Fuller (ADF) test on `avg_air_temp` yielded a p-value of approximately 0.0014, which suggests the air temperature series might be stationary around a complex mean despite visual seasonality. The overall seasonality in air temperature is strongly indicated by the cyclic pattern illustrated by Figure 2, reinforcing the significance of solar longitude as a temporal feature.

- **Ground Temperature**: Ground temperature generally mirrors the seasonal trends found in air temperature, with clear displays of periodicity and autocorrelation as showcased in Figure 3.

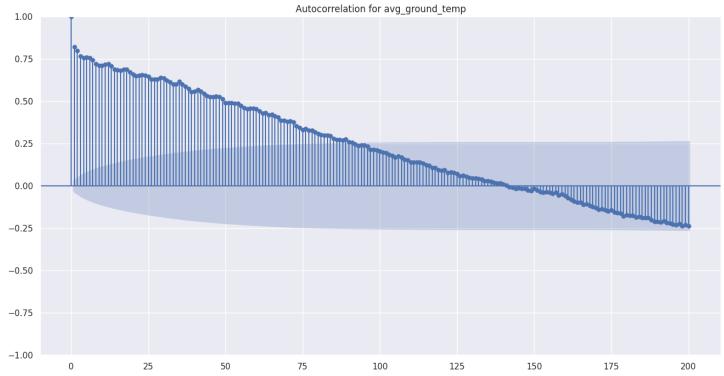

Figure 3: Autocorrelation of Average Ground Temperature

Key similarities include timing of dips and peaks, seasonal trends, and outlier phenomena among other visually distinct features showcased in Figure 4. However, it appears that ground temperature is considerably more stable. A Levene test comparing the variance of average air and ground temperatures yields a p-value of approximately 0.027. This result confirms that their variance is considerable different with ground temperature showcasing significantly lower variance overall. This additionally aligns with observations of fewer outliers in the ground temperature data compared to that of air temperature. Average ground temperature yielded 1 outlier event compared to 8 from average air temperature based on values beyond 1.5 times the interquartile range from the 25th and 75th percentiles. Performing an ADF test for `avg_ground_temp` results in a p-value of approximately 0.085, which indicates we cannot reject the null hypothesis of non-stationarity, which is consistent with apparent visual seasonality and trends. This means that appropriate handling of time series models will need to be taken. These indications of relative stability and direct relevance to surface conditions suggest `avg_ground_temp` to be a good candidate for our target variable.

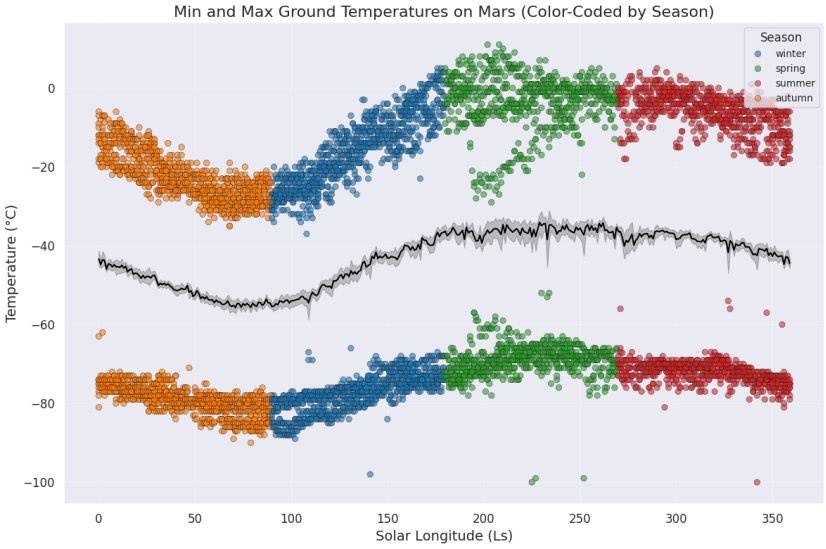

Figure 4: Annual Ground Temperature by Season

- **Atmospheric Pressure**: Atmospheric pressure exhibits distinct and interesting annual bimodal structure when plotted against solar longitude as showcased in Figure 5.

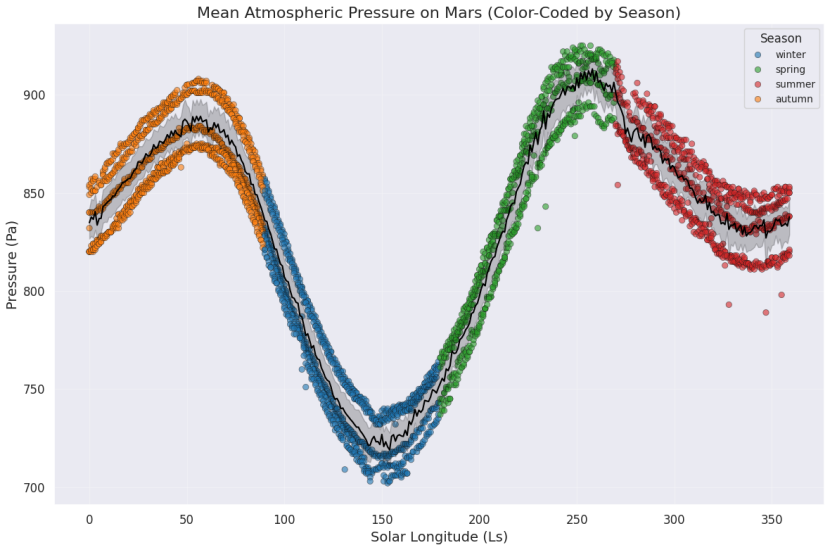

Figure 5: Annual Atmospheric Pressure by Season

This pattern suggests multiple large-scale phenomena influencing Martian atmospheric conditions resulting these shifts in pressure levels. The likely driver for these conditions is seasonal sublimation and condensation of CO2 at the Martian polar ice caps. Performing an ADF test on `mean_pressure` yields a very small p-value of approximately 4.03×10^{-11}, indicating that the pressure series as a whole is stationary. Interestingly, initial correlation testing (Pearson, Spearman, and Kendall) applied to the entire dataset showed no direct correlation between pressure and either air or ground temperatures. For instance, the Pearson correlation between ground temperature and pressure was approximately -0.041 with a p-value of approximately 0.019. Intuitively, this seeming non-relationship does not make sense. A more complex relationship must be at play and is explored further in the "Temporal Cycles" section.

- **UV Radiation**: The UV radiation index as a categorical feature also showcased clear seasonal patterns. For example, "High" UV days were predominantly observed during summer months, while "Low" UV days were typically seen in the spring.

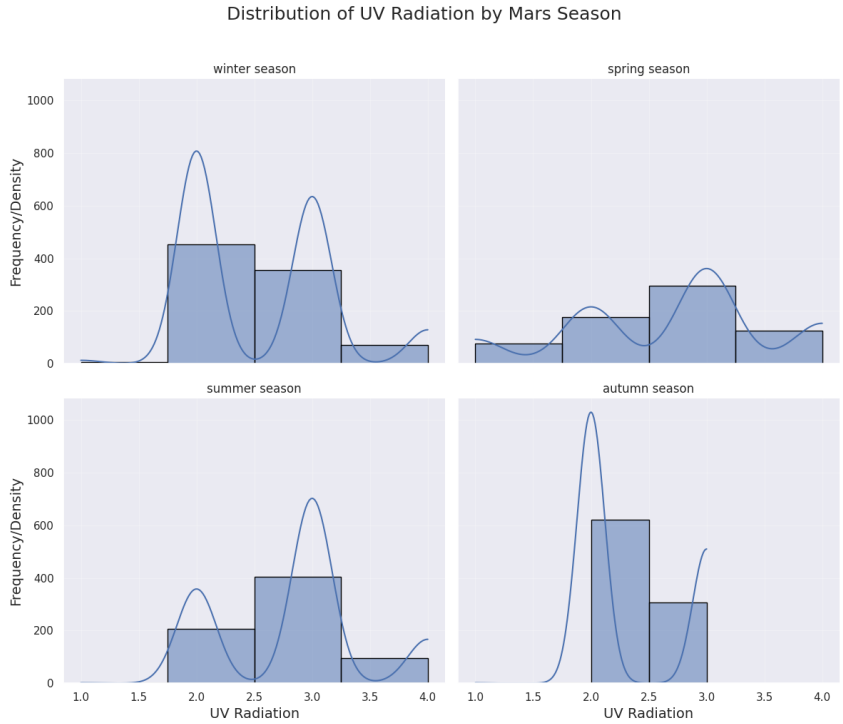

Figure 6: Seasonal UV Index Distributions

The significance of UV radiation's relationship with other atmospheric variables was confirmed through various statistical tests. An ANOVA test between UV radiation categories and temperature yielded an F-statistic of approximately 271.0 with a p-value of around 1.09×10^{-156}. Similarly, an ANOVA test between UV and pressure yielded an F-statistic of 650.1 with a p-value of about 2.04×10^{-138} and a Kruskal-Wallis test showed an H-statistic of approximately 21.59 with a p-value of about 7.96×10^{-5}. These results strongly suggest that UV radiation levels are significantly associated with both variation in temperature and pressure.

4.2 Temporal Cycles

It is important to understand the interplay between various temporal markers and atmospheric conditions.

- **Day Length**: The length of a Martian sol varies slightly throughout the year due to the eccentricity of Mars's orbit. By taking the difference between sunset and sunrise, we can engineer the "Day Length" feature. Day length can vary throughout the year, forming a sinusoidal pattern when plotted against solar longitude. In Figure 7 we can see an interesting trend where the hottest and coolest months follow daylight maximums and minimums respectively. This suggests that the increased or decreased energy deposited into the atmosphere by the sun influences atmospheric conditions with some degree of lag.

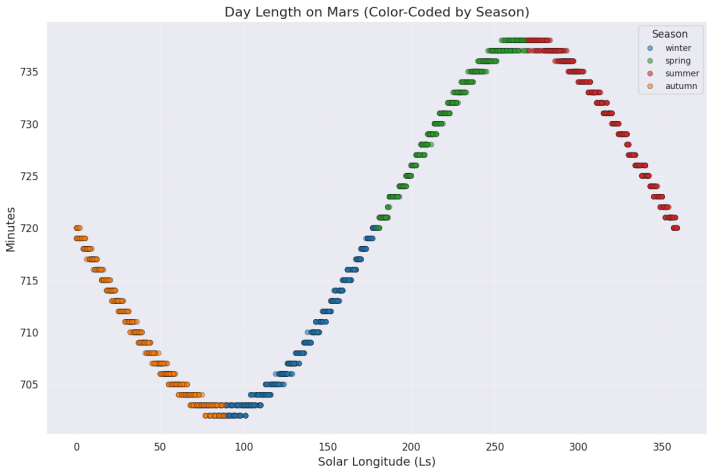

Figure 7: Day Length by Season

- **Pressure and Temperature Correlations by Season**:
  - During **winter**, a very strong negative Pearson correlation was discovered between pressure and ground temperature (≈ -0.80) and air temperature (≈ -0.71).
  - In **Autumn**, a moderate negative Pearson correlation was observed between pressure and ground temperature (≈ -0.51) and air temperature (≈ -0.31).
  - **Summer** showed a moderate positive Pearson correlation between pressure and ground temperature (≈ 0.43) and air temperature (≈ 0.30).
  - **Spring** exhibited a weaker positive correlation for ground temperature (≈ 0.19) and a very weak, non-significant correlation for air temperature.

  These results, all with p-values near zero where correlations were notable highlight that pressure and temperature are strongly linked. The direction and strength of this relationship depend heavily on season. Cooling seasons show negative correlations while warming seasons tend towards positive or weak correlations. The particularly strong correlation during the winter suggests the cooling process has a more pronounced or direct impact on local pressure.

- **Granger Causality (Ground Temperature and Pressure)**: Granger causality tests were performed to investigate if past values of ground temperature could predict current values of pressure. The results indicated that ground temperature does Granger-cause pressure, with statistical significance emerging and strengthening with an increasing number of lags. For example, with 3 lags, the p-value was approximately 0.0007, and with 4 to 6

lags, the p-values were effectively 0.0000. This suggests that changes in ground temperature have a predictive influence on atmospheric pressure, typically manifested over several Sols.

- **Solar Longitude as Primary Driver**: Across multiple variables, including temperature, pressure, and UV radiation, solar longitude proved to be the most indicative feature for capturing seasonality. Plots of these variables have consistently revealed cyclical patterns distinctly align with Martian seasons.

4.3 Insights

Through exploratory data analysis and statistical testing, several key insights were discovered which directly informed subsequent preprocessing and modeling strategies:

- **Target Variable Selection**: The relative stability of **average ground temperature**, confirmed by a significantly lower variance compared to air temperature (Levene test p-value of approximately 0.027) and fewer identified outliers (1 vs. 8), solidified its suitability as the primary target variable for forecasting.
- **Feature Importance & Seasonality**:
  - **Solar Longitude**: Its strong correlation with all major atmospheric variables makes it a critical feature. Its cyclical nature necessitates sinusoidal encoding for use in machine learning models.
  - **Seasonality**: Strong annual seasonality is present in temperature, pressure, and UV radiation. This is both visually and statistically evident and is supported by varying correlation across seasons.
- **Data Characteristics (Stationarity & Autocorrelation)**:
  - Autocorrelation in temperature series data was confirmed by Figure 3, which guides parameter selection in ARIMA-based models.
  - ADF tests indicated that `avg_ground_temp` is likely non-stationary ($p \approx 0.085$), while `avg_air_temp` ($p \approx 0.0014$) and `mean_pressure` ($p \approx 4 \times 10^{-11}$) appear stationary or trend stationary, possibly around their seasonal components. These statistics highlight the need for appropriate transformations or models inherently designed for seasonal and non-stationary data.
- **Outliers and Potential Weather Event Indicators**: While few outliers were identified in ground temperature, air temperature had more. An interesting observation was the potential for instances where the standard deviation of air temperature shifts dramatically without a corresponding shift in ground temperature, possibly indicating localized atmospheric events. Investigating periods where both air and ground temperature standard deviations spike concurrently could also be a focus for future anomaly detection related to more significant events. These observations motivate the anomaly detection component of this project.
- **Interactions and Predictive Relationships**:
  - The relationship between pressure and temperature is strongly season-dependent.
  - Granger causality results suggest that ground temperature changes can predict subsequent pressure changes with a lag of several days. This finding could be leveraged in multivariate forecasting models.
- **Need for Scaling**: The varying scales across each feature necessitate the need for appropriate scaling when training and testing neural network models.

5 Preprocessing

This section details the necessary steps taken to further process the data into a suitable format for time series forecasting and anomaly detection.

5.1 Feature Engineering

Guided by Exploratory Data Analysis (EDA) insights, new features were engineered to enhance models' ability to capture temporal patterns and relevant physical phenomena:

- **Day Length**: Calculated as the difference between sunset and sunrise times for each sol. This feature provides a measure of daily solar insolation duration, which intuitively correlates to temperature.
- **Martian Seasons**: Derived from solar longitude and defined in Table 1. Explicit handling of seasons as a categorical feature helps models account for distinct seasonal regimes.
- **Cyclical Encoding of Solar Longitude**: To ensure that the cyclical nature of Mars' orbit (and thus L_S) is correctly interpreted by the models, solar longitude was transformed into two continuous features using sine and cosine functions.

$$L_{S,sin} = \sin\left(\frac{2\pi L_S}{360^\circ}\right) \tag{1}$$

$$L_{S,cos} = \cos\left(\frac{2\pi L_S}{360^\circ}\right) \tag{2}$$

- **Categorical Feature Encoding**: Features such as `mars_season` and `UV_Radiation` were one-hot encoded to ensure model compatibility. These created binary columns for each category, allowing models to learn distinct effects without imposing an ordinal relationship between categories.

These engineered features were designed to provide more explicit signals to the models, improving their ability to understand and forecast Martian weather patterns.

5.2 Dataset Construction

To prepare the data for training and evaluation, the following steps were taken:

- **Data Splitting**: The dataset was divided into training and testing sets using a chronological split to preserve the temporal dependencies inherent in time series data. The first 80% of the dataset (ordered by Sol) was allocated for training the models, while the remaining 20% was reserved as an unseen test set for evaluating their out-of-sample forecasting performance. This approach simulates a real-world scenario where models predict future values based on past data.
- **Data Scaling**:
  - **Unscaled Data**: Traditional time series models like SARIMA and SARIMAX are often applied to unscaled data, as their formulation can directly handle the original units and magnitudes of the time series.
  - **Scaled Data**: Neural network models, such as LSTMs generally perform better and converge faster when input features are scaled to a common range. Therefore, a separate dataset was created where numerical features (including the target variable) were scaled to a range of $[0, 1]$ using `MinMaxScaler` from the `scikit-learn` library. The scaler was fit only to the training data so as to prevent data leakage from the test set, with the scaler object saved to allow for later inversion of the predictions back to their original scale for interpretation and metric calculation.

6 Forecast Modeling

This section describes the development and evaluation of three distinct time series models for forecasting average ground temperature on Mars: **SARIMA, SARIMAX, and Long Short-Term Memory (LSTM) neural network**. The primary goal was to identify a model that accurately captures the complex seasonality and trends observed in Martian weather data. All models were trained to predict the average ground temperature, however the pipeline could realistically be applied towards predicting any of the atmospheric features.

6.1 SARIMA (Seasonal Autoregressive Integrated Moving Average)

The SARIMA model was chosen for its ability to explicitly model both non-seasonal (p, d, q) and seasonal (P, D, Q, s) autoregressive, differencing, and moving average components. These model features are well-suited for the periodicity observed in the Martian temperature data as seen in Figure 3.

The optimal hyperparameters were all determined through a grid search optimizing for the Akaike Information Criterion (AIC) with the aim to balance model fit and complexity. The best performing SARIMA model, detailed in Table 2, utilized an order of $(2, 0, 2)$ and a seasonal order of $(0, 1, 1)$ with a seasonal period (s) of 25 Sols. The selection of $d = 0$ suggests the original series (or the component being modeled after seasonal differencing $D = 1$) did not require non-seasonal differencing to achieve stationarity for the model.

While SARIMA provides a strong baseline for seasonal time series, its performance (MAE: 3.57, RMSE: 4.52) indicated room for improvement, particularly in capturing more complex relationships or the influence of other atmospheric variables.

Parameter	**Value**
Seasonal Period	25
Order	(2, 0, 2)
Seasonal Order	(0, 1, 1) + (25)
AIC	10,317.75
MAE	3.57
RMSE	4.52
Directional Accuracy	43.98%

Table 2: Evaluation Metrics and Hyperparameters for Best SARIMA Model

6.2 SARIMAX (Seasonal Autoregressive Integrated Moving Average with Exogenous Variables)

The SARIMAX model extends SARIMA by incorporating exogenous variables, allowing external factors to influence our forecast. This was considered as a natural progression given the observed interplay between temperature and other variables like atmospheric pressure and UV radiation during EDA.

Similar to SARIMA, hyperparameters were tuned via grid search. The best SARIMAX model, outlined in Table 3, showed a notable improvement in forecasting accuracy (MAE: 2.10, RMSE: 2.91) and directional accuracy (50.24%) over the univariate SARIMA model. The model parameters (order (0,0,1), seasonal order (1,1,1) with $s = 5$) suggest a simpler underlying ARIMA structure once exogenous variables and seasonality were accounted for. The exogenous variables were: `avg_air_temp`, `mean_pressure`, `day_length`, $L_{S,sin}$, $L_{S,cos}$, Martian season (one-hot encoded), and UV Radiation (one-hot encoded), accounting for 11 exogenous features. This improvement highlights the benefit of including relevant external factors in the forecasting process.

Parameter	Value
Seasonal Period	5
Order	(0, 0, 1)
Seasonal Order	(1, 1, 1) + (5)
AIC	10,526.86
MAE	2.10
RMSE	2.91
Directional Accuracy	50.24%

Table 3: Evaluation Metrics and Hyperparameters for Best SARIMAX Model

6.3 LSTM (Long Short-Term Memory) Neural Network

An LSTM neural network was developed to capture potentially non-linear patterns and long range dependencies in the temperature time series which might be missed by traditional statistical models. LSTMs are well-suited for sequence data due to their recurrent nature and gating mechanisms that regulate information flow.

The architecture outlined in Table 4 consisted of an LSTM layer with 20 units, followed by a Dense output layer. The model was trained on sequences of 60 past temperature readings to predict the next Sol's temperature through a univariate approach. Training was performed for up to 30 epochs with a batch size of 32, utilizing the Adam optimizer and Mean Squared Error (MSE) as our loss function. Early stopping with a patience of 3 epochs was implemented on a validation split of 20% of the training data to prevent overfitting.

The LSTM model achieved the lowest MAE (1.768) and sMAPE (4.16%) among the tested models, indicating strong performance in terms of average error magnitude and percentage error.

Parameter	Value
Feature Set	Univariate (Average Ground Temperature)
Sequence Length	60
Batch Size	32
LSTM Units	20
Epochs	30
Early Stopping	Yes (patience = 3)
Validation Split	0.2
Optimizer	Adam
Loss Functions	MSE (forecast), CCE (direction)
Evaluation Metrics	MSE (forecast), Accuracy (direction)
MAE	1.768
RMSE	2.939
sMAPE	4.16%
Directional Accuracy	49.66%
Selection Score	2.775

Table 4: Evaluation Metrics and Hyperparameters for Best LSTM Model

6.4 Results Summary

A comprehensive comparison of key performance metrics for each model on our test set is summarize in Table 5.

Model	MAE	RMSE	sMAPE	Directional Accuracy
SARIMA	3.570	4.525	8.326%	43.98%
SARIMAX	2.097	**2.914**	4.913%	**50.24%**
LSTM	**1.768**	2.939	**4.162%**	49.66%

Table 5: Comparison of Model Performance

The **LSTM model** demonstrated the best overall performance in terms of Mean Absolute Error (MAE: 1.768°C) and Symmetric Mean Absolute Percentage Error (sMAPE: 4.162%), suggesting its predictions were, on average, closest to the actual values and had the lowest percentage error.

The **SARIMAX model**, however, achieved the lowest Root Mean Squared Error (RMSE: 2.914°C), indicating it was slightly better at penalizing larger errors, and also achieved the highest Directional Accuracy (50.24%), meaning it was marginally better at predicting whether the temperature would rise or fall compared to the previous Sol.

The **SARIMA model** served as a good baseline but was outperformed by both SARIMAX and LSTM, highlighting the benefits of either including exogenous variables (SARIMAX) or employing more complex non-linear models (LSTM) for this dataset.

Overall, while the LSTM showed superior accuracy in average error magnitudes, the SARIMAX model provided a competitive alternative, especially considering its slightly better RMSE and directional accuracy. The choice of the "best" model could depend on the specific application: LSTM for minimizing average forecast error, and SARIMAX if penalizing large errors or predicting the direction of change is more critical. For the purpose of generating residuals for anomaly detection, the model that best captures the predictable patterns (potentially SARIMAX or LSTM) would be most suitable.

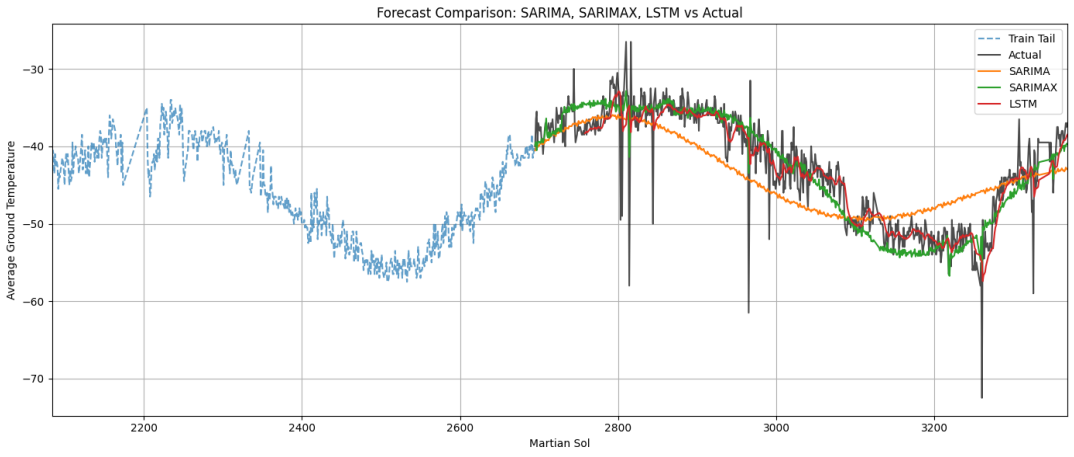

Figure 8: Comparison of Model Forecasting Ability

7 Anomaly Detection

Following the forecasting of average ground temperature, this section focuses on detecting atypical atmospheric behavior. The strategy involves identifying deviations from the expected patterns captured by the best forecasting model.

7.1 Strategy

The core strategy for anomaly detection is to analyze the residuals from the temperature forecasts. Residuals represent the difference between the actual observed temperatures and the values predicted by a forecasting model (Actual - Predicted). The rationale is that a well-fitted forecasting model captures the regular, predictable components of the time series (like seasonality and trend). Therefore, its residuals should ideally be small and random (noise). Anomalies, or unexpected atmospheric events, would manifest as unusually large residuals, indicating that the observed temperature significantly deviated from what the model predicted based on learned patterns.

For this study, an LSTM-based autoencoder was trained on the residuals generated by the SARIMA and SARIMAX forecasting models. This hybrid approach aims to learn the distribution of ”normal” forecast errors and then flag residuals that have a high reconstruction error when passed through the autoencoder, signifying they don't conform to this learned normality.

7.2 Detection Method

An LSTM autoencoder was chosen for its proficiency in learning representations of sequential data. The architecture consisted of an encoder (LSTM layers that compress the input sequence into a lower-dimensional latent representation) and a decoder (LSTM layers that attempt to reconstruct the original sequence from this latent representation).

- **Input**: The autoencoder was trained on sequences of residuals obtained from the SARIMA and SARIMAX models.
- **Training**: The autoencoder was trained to minimize the reconstruction error between input residual sequences and their reconstructed output.
- **Anomaly Scoring**: After training, an anomaly score for each new sequence of residuals was calculated as its reconstruction error. Higher reconstruction errors indicate that the autoencoder struggled to reproduce that particular sequence of residuals, suggesting it was ”abnormal” compared to what it learned during training.
- **Thresholding**: A threshold for anomaly detection was determined based on the distribution of reconstruction errors on a validation set (or a portion of the training residuals if no separate validation set was used for the autoencoder). For instance, the 95th percentile of these reconstruction errors was used as a threshold (as indicated in Table 6 for the autoencoders trained on SARIMA and SARIMAX residuals respectively), meaning any new residual sequence with a reconstruction error above this value was flagged as an anomaly.

7.3 Results

The performance of the LSTM autoencoder when trained on residuals from the SARIMA model (SARIMA AE) versus the SARIMAX model (SARIMAX AE) is compared in Table 6. These metrics refer to the reconstruction errors of the autoencoder itself.

Metric	SARIMA AE	SARIMAX AE
Mean Error	**0.404**	0.526
Median Error	**0.362**	0.467
95th Percentile	**0.805**	0.904
Max Error	0.942	**1.129**
Std Deviation	**0.179**	0.205

Table 6: Comparison of Autoencoder Performance

The **SARIMA AE** (autoencoder trained on SARIMA residuals) generally exhibited lower mean (0.404 vs. 0.526) and median (0.362 vs. 0.467) reconstruction errors, as well as a lower 95th percentile threshold (0.805 vs. 0.904) and standard deviation (0.179 vs. 0.205). This suggests that the autoencoder was able to more consistently reconstruct the "normal" patterns present in the SARIMA residuals. Paradoxically, this might imply that SARIMA residuals contained less complex "noise" or that the SARIMAX residuals, by having more of the predictable signal removed by a better base model, were inherently more "noise-like" and harder for the autoencoder to model consistently. However, the **SARIMAX AE** had a higher maximum error (1.129 vs. 0.942). This could mean that when true anomalies (or harder-to-model residuals) were present, they resulted in a more pronounced reconstruction error from the autoencoder trained on SARIMAX residuals, which could be beneficial for detection if the goal is to find more distinct deviations.

The choice between using SARIMA or SARIMAX residuals for anomaly detection would depend on further analysis of the flagged anomalies. If the SARIMA AE flags more meaningful anomalies despite its generally lower error profile, it might be preferred. Conversely, if the SARIMAX AE, with its potentially "cleaner" residuals (as SARIMAX was a better forecaster), leads to more distinct spikes for true anomalies, it could be more effective.

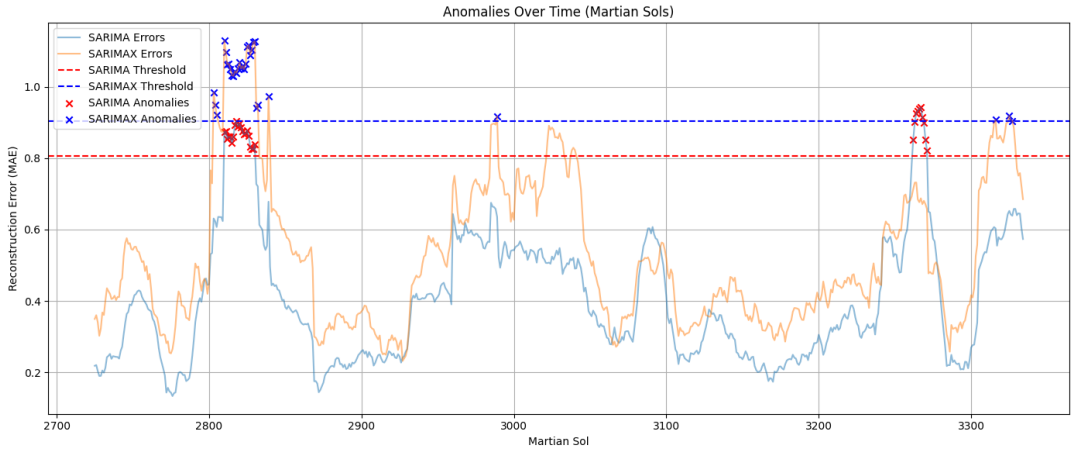

Figure 9: Anomalies Over Time

It is interesting to note that each of the autoencoder models detected 31 anomalous sol's, with 21 of them being shared. Further investigation of these anomalies and their characteristics will be left for a later study.

8 Conclusion

This study successfully developed and evaluated time series models for forecasting Martian ground temperature and a hybrid system for detecting atmospheric anomalies using NASA REMS data from the Curiosity rover.

8.1 Key Takeaways

- **Forecasting Performance**: A combination of classical (SARIMA, SARIMAX) and machine learning (LSTM) models were explored. The LSTM model achieved the lowest MAE and sMAPE, while the SARIMAX model, incorporating exogenous variables, demonstrated strong performance with the best RMSE and directional accuracy. This highlights that while LSTMs can capture complex non-linearities, well-specified classical models with relevant external information remain highly competitive.
- **Anomaly Detection Efficacy**: A hybrid approach using an LSTM autoencoder trained on the residuals of forecasting models proved viable for identifying atypical atmospheric behavior. The characteristics of the residuals from different forecasting models (SARIMA vs. SARIMAX) influenced the autoencoder's performance and the nature of detected anomalies.
- **Value of EDA**: Thorough Exploratory Data Analysis was crucial in identifying seasonality, stationarity properties, key relationships (like the season-dependent pressure-temperature correlation and Granger causality), and the suitability of ground temperature as a stable forecasting target. These insights directly informed feature engineering and model selection.
- **Methodological Blend**: The project demonstrated the flexibility and utility of blending classical statistical time series techniques with modern machine learning approaches to tackle complex environmental data analysis.

8.2 Next Steps

While this study provides a solid foundation, several avenues for future research and development exist:

- **Expanded Feature Set**: Incorporate additional environmental sensor data from Curiosity (e.g., dust opacity, wind speed/direction if more complete data becomes available) into multivariate forecasting models like SARIMAX or multivariate LSTMs. This could improve forecast accuracy and the ability to detect a wider range of atmospheric phenomena.
- **Refined Anomaly Detection**:
  - Explore more sophisticated anomaly detection algorithms beyond reconstruction-based autoencoders.
  - Conduct a qualitative analysis of flagged anomalies by cross-referencing them with Martian daily science reports or imagery to validate their significance and potentially categorize different types of anomalies.
  - Investigate the use of residuals from the LSTM forecast model for the autoencoder input.
- **Direct REMS Archive Integration**: Refine the data processing pipeline to work directly with the continuously updated NASA REMS archived data (currently 4,197 Sols at the time of writing and growing), rather than a static, preprocessed Kaggle dataset. This would enable a more dynamic and operationally relevant system.

- **Dust Storm Precursors**: While not the primary focus of this project, future work could explore if the developed anomaly detection framework, particularly if expanded with dust sensor data, can identify subtle precursors to localized or regional dust storms.

8.3 Recommendations

Based on the results of this analysis, we propose the following applications of the developed pipeline:

1. **Integrate residual-based anomaly detection into environmental monitoring routines.** The hybrid LSTM residual scoring method can reliably detect deviations from modeled atmospheric behavior. Flagged anomalies, particularly those identified using residuals from the best-performing forecast model, may assist scientists in identifying environmental shifts or instrument irregularities not immediately evident in raw sensor outputs, especially during critical seasonal transitions.

2. **Use SARIMAX or LSTM forecasts to estimate short-term atmospheric trends in support of autonomous scheduling.** The developed SARIMAX and LSTM models provide stable and relatively accurate short-range forecasts of ground temperature. These forecasts could help future rover operations plan time-sensitive activities (e.g., drilling, sample analysis, or long traverses) to coincide with more favorable environmental conditions, enhancing mission efficiency and safety.

3. **Incorporate this pipeline into simulation or mission rehearsal environments.** The modular and interpretable nature of the modeling and detection system makes it suitable for use in simulated mission scenarios. This can aid in synthetic data generation for testing autonomous systems, rehearsing responses to anomalous environmental readings, or debugging onboard diagnostic tools.

References

[1] DEEP CONTRACTOR. *Mars Rover Environmental Monitoring Station (REMS) Data*. Kaggle. https://www.kaggle.com/datasets/deepcontractor/mars-rover-environmental-monitoring-station/data.

[2] Laboratoire de Météorologie Dynamique (LMD). *Mars Solar Longitude*. https://www-mars.lmd.jussieu.fr/mars/time/solar_longitude.html.

[3] NASA PDS Atmospheres Node, New Mexico State University. *Mars Science Laboratory REMS Data Archive*. https://atmos.nmsu.edu/data_and_services/atmospheres_data/MARS/curiosity/rems.html.