Maunakea Summit Temperature

Analysis and Prediction

Version 1.1

5 October 2023

Author:

Cole Odegard

Introduction

Temperature prediction on the summit of Maunakea is utilized by the W.M. Keck Observatory to set inner dome temperatures as close to the outside as possible. Equalizing inner and outer dome temperatures reduces airflow at opening time, which minimizes the disturbance to the instruments within the dome. Keck currently utilizes a model based solely on historical temperature data to determine what temperature to set the inner dome to, which is accurate, on average, to the nearest degree.

Deliverables

- Background information on Maunakea's temperature trends and correlations between predictive variables and temperature. Predictive variables will include historical temperature, time of year, humidity, barometric pressure, and dewpoint.
- A model for temperature at the summit of Maunakea; created based on the information collected during the analysis process.
- Descriptive statistics for model accuracy including average and median prediction error in °C.
- Plots that overlay the model on top of daily temperature data.

Project Purpose

To predict the outside temperature at the telescope at opening time.

Proposed Solutions

- Using historical temperature data without a model
- Using a linear model for temperature
- Using a polynomial model for temperature
- Focusing solely on predicting temperature at opening time
- Using a weighted model of temperature that takes historical data.
- Using a combination of temperature models
- Using a neural network/machine learning algorithm

Chosen Solution

- Using a trapezoidal integration model

Data Import & Cleaning

The data for this project is stored in two csv files that each contain 11 columns:

- Date: The year, month, and day of each entry

- UT/Time: The time in HR: MIN: SEC of each entry
- OutTemp: The recorded outside temperature (°C)
- OutHumidity: The recorded outside humidity (% water vapor)
- OutPressure: The recorded outside barometric pressure (mb)
- OutDewpoint: The recorded outside dewpoint (°F)
- InTemp: The recorded inside temperature (°C)
- InHumidity: The recorded inside humidity (% water vapor)
- InDewpoint: The recorded inside dewpoint (°F)
- primTemp: The recorded temperature on the primary mirror inside (°C)
- secTemp: The recorded temperature on the secondary mirror inside (°C)

Date and UT columns are merged to create a timestamp column. InHumidity and InDewpoint have been removed since inside conditions will not be useful in predicting outside temperature.

Each of the columns must be cleaned to remove erroneous data points. OutTemp is filtered to only include values below 12C and above -11C, which are the record high and low temperatures recorded on Mauna Kea. OutPressure values below 600 and above 630 are removed since atmospheric conditions in Hawaii make barometric pressure values above and below those thresholds impossible. Values below 0% for OutHumidity have been removed since humidity cannot fall below 0%; values above 100% have been left since humidity can climb above 100% in periods of high moisture such as during heavy rainfall. Using the filtered temperature and humidity data, it is possible to calculate the range of dew point values, -60.3 to 55.7 degrees. These values are converted from Fahrenheit to Celsius to match the temperature data.

Variable Characterization

All the "Out" variables can be used to predict temperature. Each variable changes over both short- and long-term periods of time. Temperature tends to rise in the late spring through mid-fall months, and throughout the day when the sun warms the earth. Barometric Pressure is correlated with temperature, with lower barometric pressure values during the nighttime and winter months, but also during periods of atmospheric instability occurring in weather fronts and extreme weather. Humidity will rise during the nighttime and when there is moisture in the air from clouds, rain, or snow. For increased model accuracy it is important to consider the short term and long-term changes in each of the chosen predictive variables.

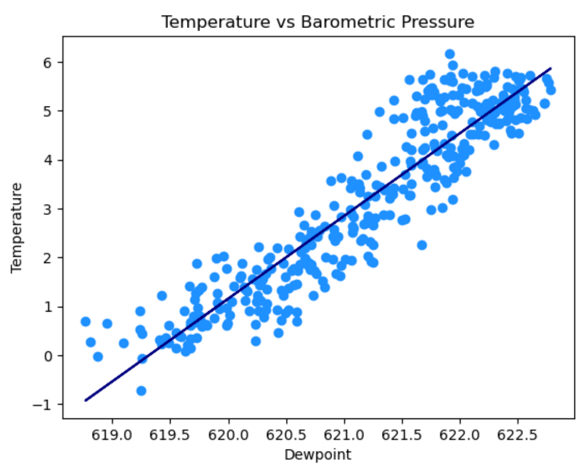
Research Questions

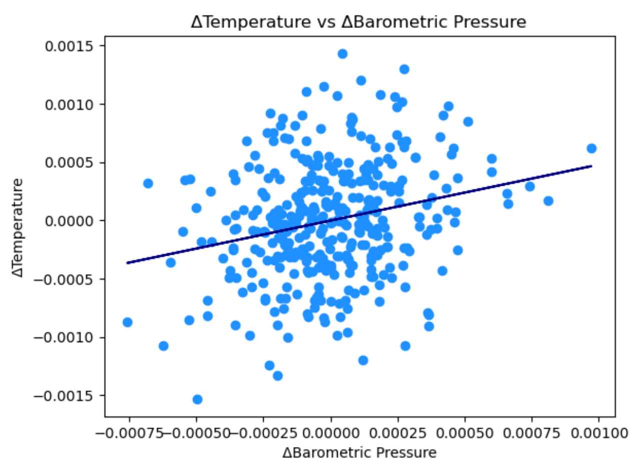
- How does changing barometric pressure (both rapid and gradual changes) affect temperature?
- Is there a pattern to warming and cooling cycles? (Ex: usually takes 4 days to cool and then 3 days to warm)

- Can we aggregate cooling by month? (Ex: Average daytime temperature variation is 9°C in September but only 5°C in January)
- How do Linear and Polynomial models fair in temperature prediction? Is there any usefulness in keeping models that simple?
- Can we predict opening temperatures solely based on data from past opening times?
- How do humidity and dewpoint affect the temperature? Are they predictive or reactive variables?
- What is the correlation between daytime high and nighttime low temperatures?

Exploration Results

Barometric Pressure Correlation





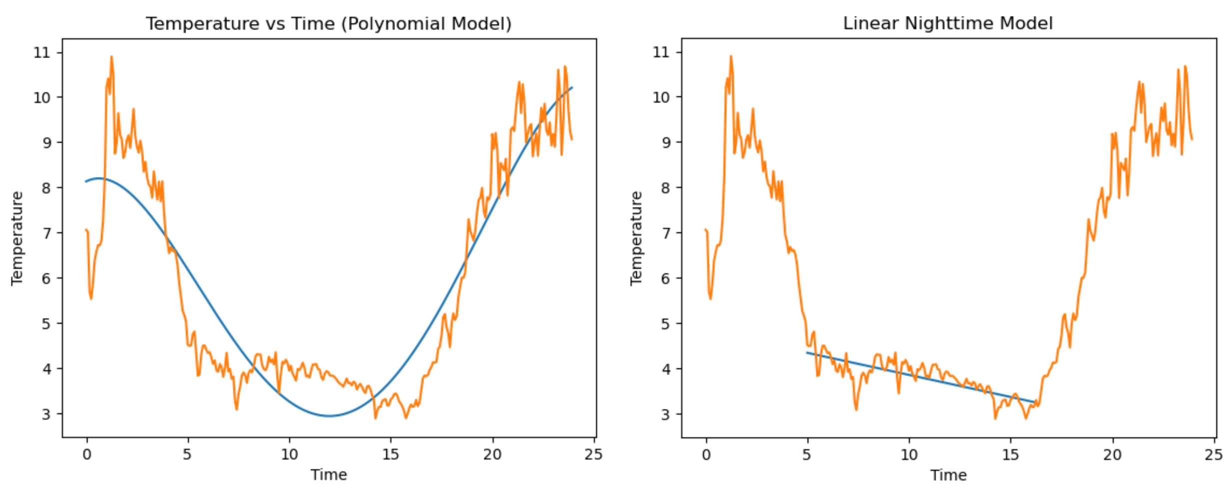
Figures 1.1&1.2: Aggregated daily average of temperature vs barometric pressure from 2000-2019 and Δtemperature vs Δbarometric pressure from 2000-2019

Figure 1.1 displays the correlation between barometric pressure and temperature. When the barometric pressure rises air molecules are pressed closer together and heat is formed, warming the air. While this extremely oversimplified description of barometric pressure is accurate, it does not explain the relationship between pressure and temperature shown in Figure 1.2. Figure 1.2 shows the relationship between daytime highs and higher barometric pressure, but it also shows an equal rise in pressure during the night, which is usually followed by a brief spike in temperature. It appears that during the daytime hours both barometric pressure and temperature can act on each other, but when barometric pressure rises during the nighttime it leads to a clearer rise in temperature. This complex relationship can be used most effectively when modeling more extreme barometric pressure changes that lead to more extreme temperature swings.

Month	AvgTempMin	AvgTempMax	TempChange
1	-1.31030	5.27069	6.58099
2	-2.04135	4.97636	7.01770
3	-1.98979	5.37632	7.36611
4	-0.71832	7.59464	8.31295
5	0.29319	8.76330	8.47012
6	1.57027	10.18875	8.61849
7	1.76744	10.40157	8.63413
8	2.00032	10.36694	8.36661
9	2.26205	10.47828	8.21623
10	1.33602	9.26793	7.93191
11	0.03851	7.20055	7.16204
12	-0.84199	5.80582	6.64780

Figure 1.3: Average high and low temperatures by month with temperature change included.

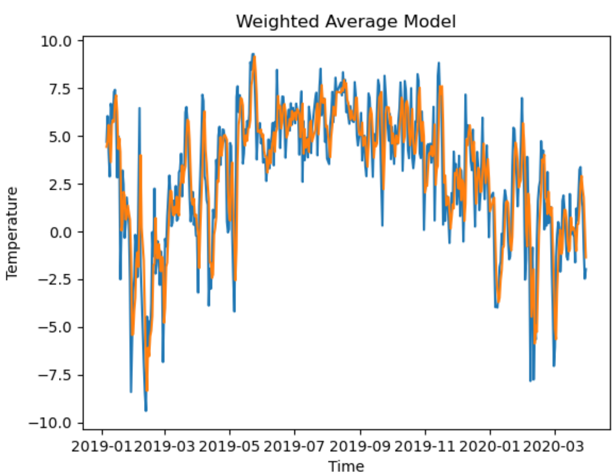
The second question deals with temperature variability throughout the year. By aggregating monthly data from 2000-2019 there is a clear trend of larger daily temperature swings during the summer months, which can potentially be useful in model building that takes into account historical trends in the data. A model like this could be very effective in predicting stable weather patterns or returns to stable weather patterns.

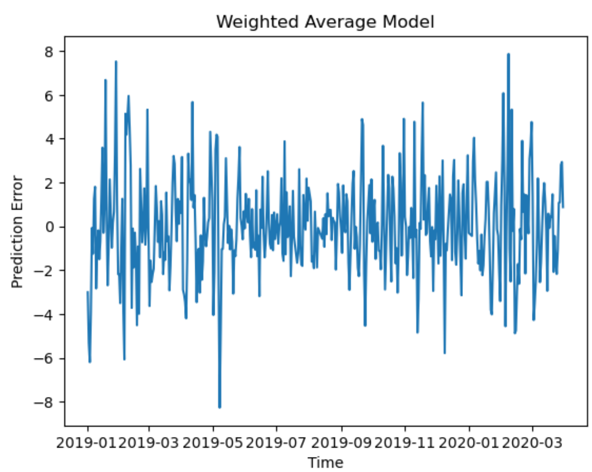


Figures 1.4 & 5: Polynomial model for daily temperature and Linear model for nighttime temperature

Monthly and yearly temperature swings are largely stable over the long term, which is also a trend that applies to daily temperature swings. Figure 1.4 shows what happens when a 4th degree polynomial model is fitted to temperature on a random day. Polynomials will not be used since they do not accurately capture the steepness of temperature rises during the day and falls at night. This phenomenon is one that is relatively unique to Mauna Kea, with other locations around the world often taking much longer to warm in the morning and cool at night. Linear models are also extremely inaccurate when fitted to an entire 24hr temperature cycle, but they are one of the most accurate ways to predict nighttime cooling trends. Figure 1.5 shows a linear model fitted to just nighttime cooling.

Weighted Average Model





Figures 1.6 & 7: Model for the temperature at sundown using previous 4 days

Figures 1.6 and 1.7 show the results of fitting a model to just temperature at opening time. The simplest approach to predicting temperature at opening time is by focusing on just historical opening time temperature data. Figure 1.7 demonstrates that this simplistic approach actually works with an error of <1°C 42% of the time. The problem with the weighted average of historical temperatures is that it will never be able to accurately capture the effect of daytime warming, barometric pressure changes, humidity or dewpoint changes, and its aggregate error of 1.97°C is unacceptable.

Model Ideas

Sundown Prediction

- A weighted average of last 3 sundown temperatures
  - 15%--3 days ago; 25%--2 days ago; 60%--1 day ago
  - "Sundown" is defined as 6:30pm

24 Hour Prediction

- Use previous 3/5/7 days of temperature data along with 24hr pressure data.

- Use previous 1/3 days of temperature data along with a historical data average (to predict that temperature extremes will eventually return to normal)
- Use previous high/low temperature swings to adjust our 1/3/5/7-day model (Ex: depending on how different the previous night was from the nights before weight it more heavily. So, if we are entering a cooling pattern, we should extrapolate more recent temperature data)

Model Exploration

1 Day Temperature Model

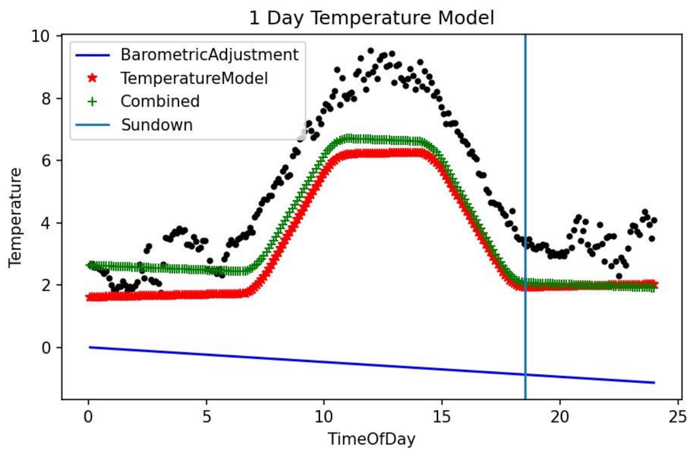


Figure 2.1: A graph showing the model for temperature based on the previous day's temperature (red), the model for 24hr barometric pressure change (blue) and the combination of these models (green)

The easiest historical model for temperature and barometric pressure combined can be made using the previous 24 hours of temperature and barometric pressure. Using just one day of temperature data is inaccurate because the temperature does not usually behave uniformly over two consecutive days even when adjusted for barometric pressure changes. To improve model accuracy there must be more historical data considered.

3 Day Temperature Model

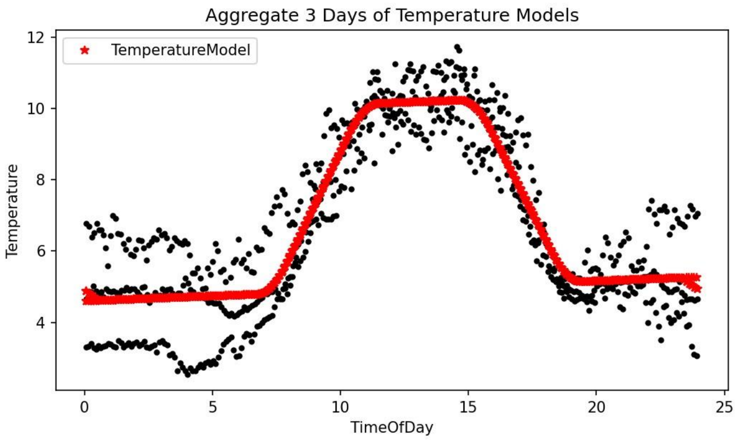


Figure 2.2: The model from combining three days of temperature data.

A three-day temperature model provides the necessary historical temperature range to capture both medium term warming and cooling trends along with past 24-hour swings. Figure 2.2 shows three separate days of temperature data and the aggregated model they produce. Once adjusted to the current temperature, this model will be, on average, much more accurate than a simple one-day model. Figure 2.3 shows the final temperature model once adjusted for barometric pressure changes and current temperature. Once adjusted for barometric pressure changes the model accurately captures the more stable nighttime temperatures than the 3-day temperature model alone, which predicts a rising temperature at night. Finally, Figure 2.4 shows the final adjustment to the 3-day temperature model, which smooths the raw temperature data itself. The model is less affected by the input noise of raw temperature data once that noise has been smoothed.

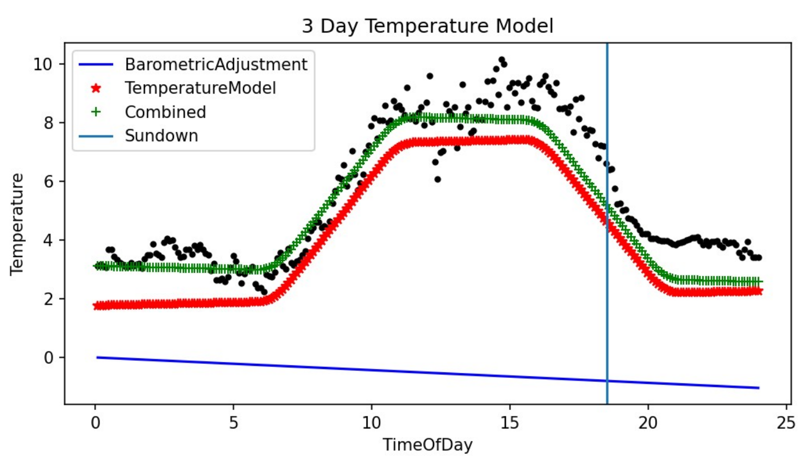


Figure 2.3: The model from Figure 2.2 combined with a 24-hour model for barometric pressure.

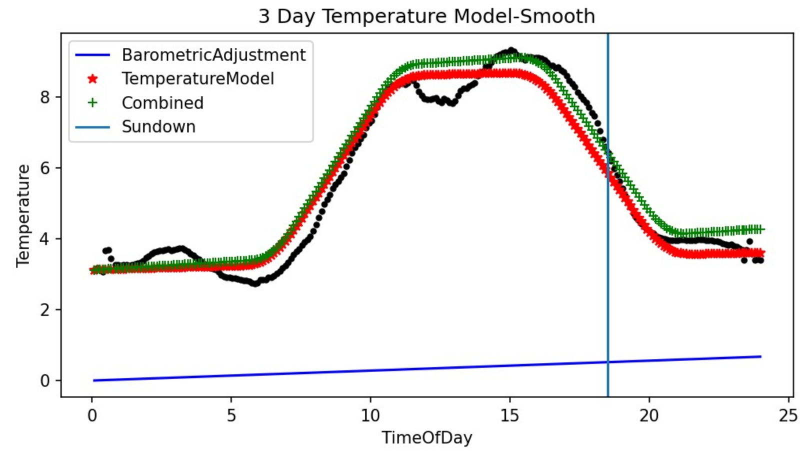


Figure 2.4: The model from Figure 2.3 after smoothing temperature data in 75-minute increments.

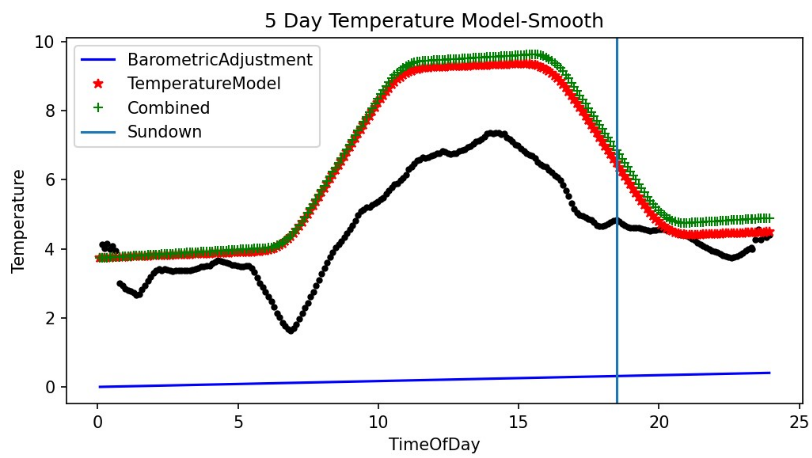


Figure 2.5: The model from Figure 2.4 expanded to include 5 days of historical temperature data.

However, more data is not always a good thing. While accurate generalizations about a location's climate can be made over time, weather is highly dependent on a variety of real time conditions. Figure 2.5 is an example of how zooming out to include temperature data from more than 3 days in the past opens models up to unwanted generalizing. When too much historical data is considered, the model ends up relying too much on the past and will also take longer to even out after periods of extreme weather.

Aggregate Model Results

Chosen Format

The final model was trained on 3-day intervals of temperature data from 2019-2020. The model was constructed using the trapezoidal rule of integration, which takes the following parameters:

- X: The range of time values we want to model for
- Amp: The amplitude of our function
- Mu: The average of our function
- Top: The width of the top of our curve
- Bottom: The width of the bottom of our curve
- Slope: The slope of the barometric pressure adjustment over 24 hours
- Bg: The intercept of the barometric pressure adjustment
- Start: The temperature at the start of the day (00:00) to be predicted

The model parameters were all tuned for accuracy, including X, which was adjusted to weight each of the prior 3 days equally at 33.3% each. This ratio represents the tendency of temperatures at the summit to follow both medium (3-5 day) and short term (1-3 day) trends. Including three days of historical temperature data is the most accurate, and barometric pressure inclusion provides a ~6% increase in model accuracy. Importantly, barometric pressure slope was used instead of barometric pressure change. This is because barometric pressure is cyclical and naturally rises when the sun and the moon are overhead. A single lunar cycle is ~24 hours and 45 minutes, though, so barometric pressure will either slightly rise or slightly fall naturally over 24 hours. When barometric pressure trend is modeled, these natural changes are no longer included in the slope estimate. This minor detail increases the percentage of days that are accurate within 1°C by 3.8%.

Model Parameters

- X: *Dataset with the prior 3 days of temperature data weighted equally (33%-33%-33%)
- Amp: maximum value of X – minimum value of X
- Mu: 12.1875
- Top: 1.4
- Bottom: 4.4314
- Slope: 0.07644953
- Bg: 3.748
- Start: *The temperature at the start of the day (00:00) to be predicted

The model's weighting for each day of historical data was updated once parameters were chosen. As the weight on the most recent 24 hours of temperature data increases, there is no increase in model accuracy. The most accurate model takes 24-hour barometric pressure trends and gives the previous day equal the weight of the 2^{nd} and 3^{rd} previous days.

Model Statistics

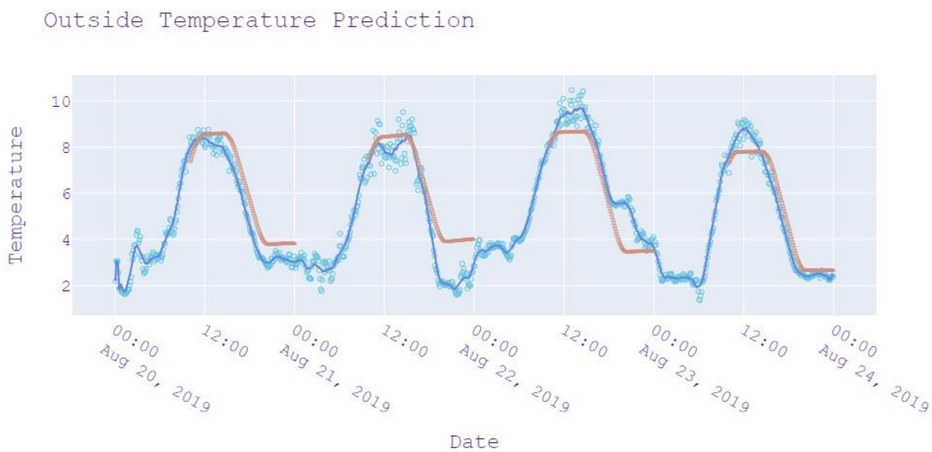


Figure 3.1: Temperature vs Date with the final temperature model overlayed from August 19th through August 25th, 2019.

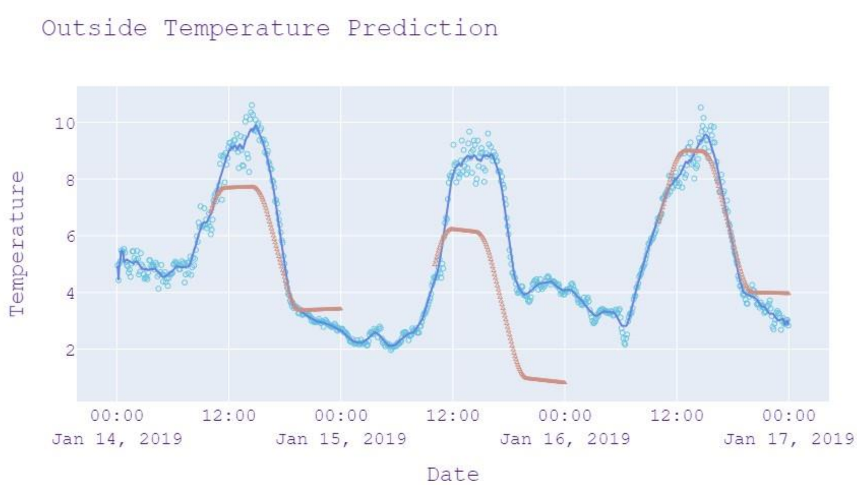


Figure 3.2: Temperature vs Date with the final temperature model overlayed from January 14th through January 17th, 2019.

In general, the temperature model performs best when the temperature is stable and has been stable for at least two days and worse when one of the three days included has a large temperature swing. The period from August 20, 2019, through August 24, 2019, as shown in Figure 3.1, was stable and follows a string of successful predictions.

However, days of stable temperatures do not always guarantee successful temperature predictions. Figure 3.2 shows another stable three day stretch of temperature trends, but inconsistent morning warming patterns, which affect the model heavily. The model begins its predictions at 10am which means a day like January 15th, which took much longer to warm than the prior three days, will have a very inaccurate prediction. This type of error will only persist for one or two prediction days as the model adjusts itself completely over just a three-day stretch.

By reducing the temperature adjustment at 10am, we can reduce the effect morning temperature swings have on the model. This means instead of fully adjusting the model to start at the temperature at 10am (or whatever time the model is run), the model will only adjust a percentage of the way. The reason this reduces average temperature error by more than 8% is that large adjustments also correspond with pattern changes. If the temperature is predicted to be 5°C at 10am and it is already 9°C, then a 4°C would need to be made to the *entire* model.

The temperature at 10am is not a perfect indicator of the temperature at sunset, and making this full adjustment may cause egregious modeling errors like the one made on January 15th in Figure 3.2. When a 75% or even 50% adjustment is made instead, the result is an adjustment towards the 10am temperature while retaining the more stable trends of the prior three days. A 70% adjustment is the most accurate adjustment in this case, leading to a 9% reduction in average temperature error and a 7% reduction in median temperature error.

Further Exploration

Accurate temperature modeling is limited to the amount of weather data collected and model complexity. For the prediction purposes of the W.M. Keck Observatory it is not necessary to predict any weather conditions other than the temperature at opening time. More advanced weather prediction needs other information than just temperature or barometric pressure, taking wind speed, direction, humidity, weather fronts, and other global weather patterns into account. When predicting temperature there is some usefulness to adding more variables, but only to a certain point. The weather will always be unpredictable. Model expansion could include adding more variables, using machine learning, or even updating Keck's weather data collection. Another major improvement would be integrating the temperature model or outside temperature with the inside thermostat so that slight changes in temperature throughout the day could be reflected inside the dome.

Conclusion

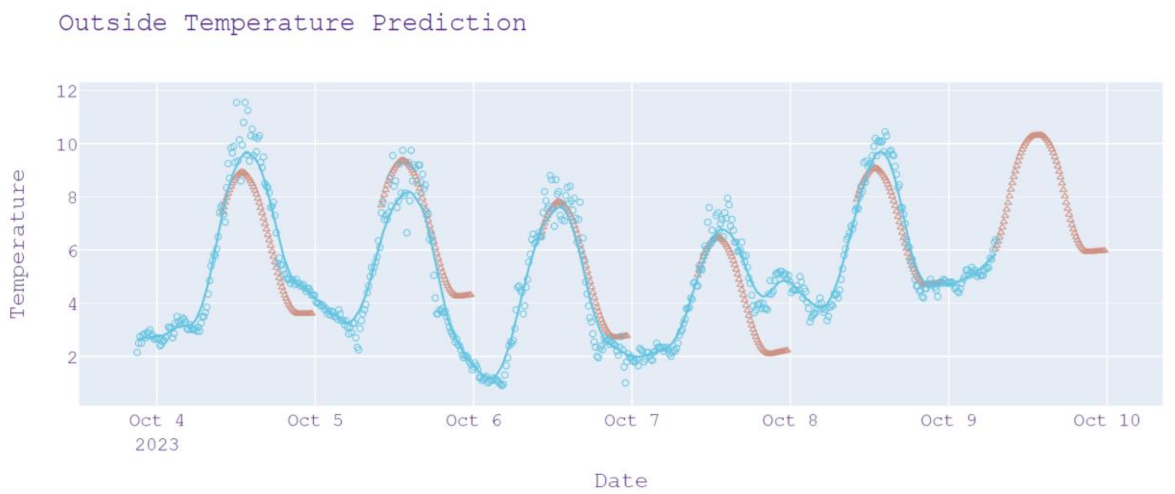


Figure 4.1: The final temperature model for Oct 4-9, 2023, updated in real time.

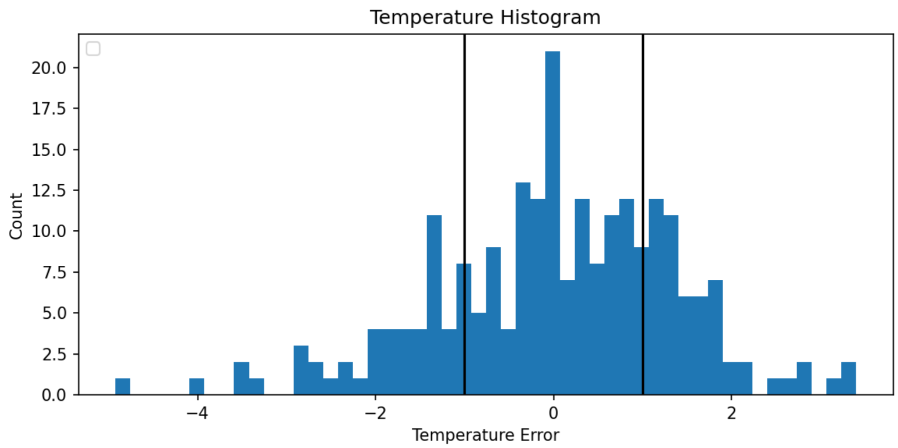


Figure 4.2: Histogram of the temperature error from the final model

The final temperature model was turned into a class, and then packaged, allowing for easier temperature modeling across platforms. The class will plot any temperature data available up to the current time and then display the model for the rest of the day. The class allows the option to adjust sunset time, the "adjustor," which is the amount that the model considers the temperature swing of the current day, and the time that the model should start predicting from. The model historically has an RMSE of 1.36 °C and is within 1°C on 55.5% of days. This model is a ~4% improvement when barometric pressure is added, and at least a 2% improvement when considering the current day's temperature depending on when the model is run. Finally, the histogram of the model's error shows a normal distribution centered around 0, which is confirmed by the model's average error of 0.06°C (Figure 4.2).