

Wyoming Weather Analysis

Overview

This is a report on the historical analysis of weather patterns in an area that overlaps approximately one-fourth the area of the state of Wyoming, spanning from roughly 111 to 105 degrees W longitude and 43 to 44 degrees N latitude. The area that is the subject of this report is within the rectangle in the map below:



The data we will use comes from [NOAA](http://ftp.ncdc.noaa.gov/pub/data/ghcn/daily/), and was downloaded from the ftp site located at [ftp://ftp.ncdc.noaa.gov/pub/data/ghcn/daily/](http://ftp.ncdc.noaa.gov/pub/data/ghcn/daily/).

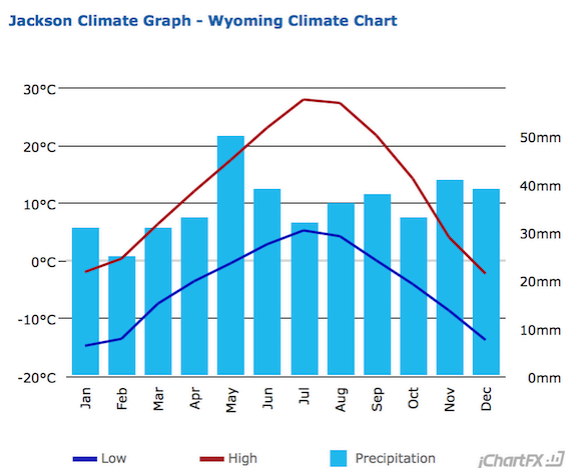
The focus of this analysis is on six measurements collected daily over several years at 66 weather stations in the above highlighted region of Wyoming:

- **PRCP:** Precipitation (tenths of mm)
- **SNOW:** Snowfall (mm)
- **SNWD:** Snow depth (mm)
- **TMIN:** Minimum temperature (tenths of degrees C)
- **TMAX:** Maximum temperature (tenths of degrees C)
- **TOBS:** Temperature at the time of observation (tenths of degrees C)

Sanity Check: Comparison with Outside Sources

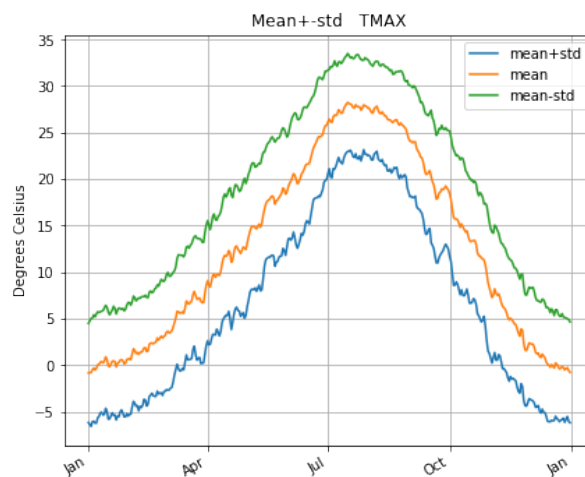
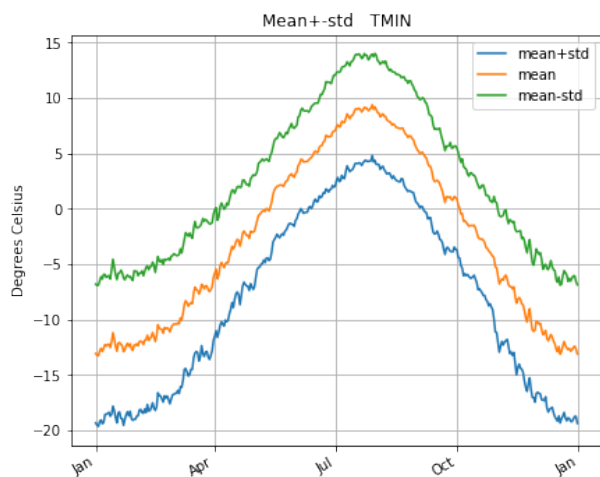
We start by comparing some weather statistics obtained from the NOAA ftp site with data obtained from a site called [US Climate Data](#), which includes measures for the average high and low temperature and total precipitation for each month.

Below is a weather graph obtained from [US Climate Data for Jackson, WY](#), which we will use to sanity check the NOAA weather data:

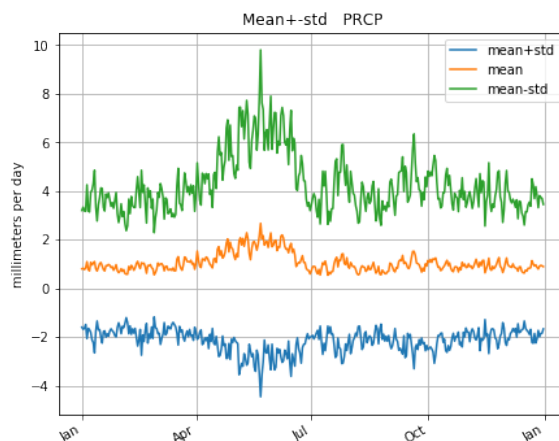


- The low temperature ranges from about -15 degrees Celsius to about 5 degrees Celsius.
- The high temperature ranges from a little below 0 degrees Celsius to almost 30 degrees Celsius.
- Monthly precipitation ranges from about 25 millimeters (mm) in February to a little over 50mm in May, but is generally between 30mm and 40mm.

These temperature ranges generally agree with the NOAA data we are using for our analysis. Below are plots of the average minimum and maximum temperature profiles for the station in Jackson, WY:



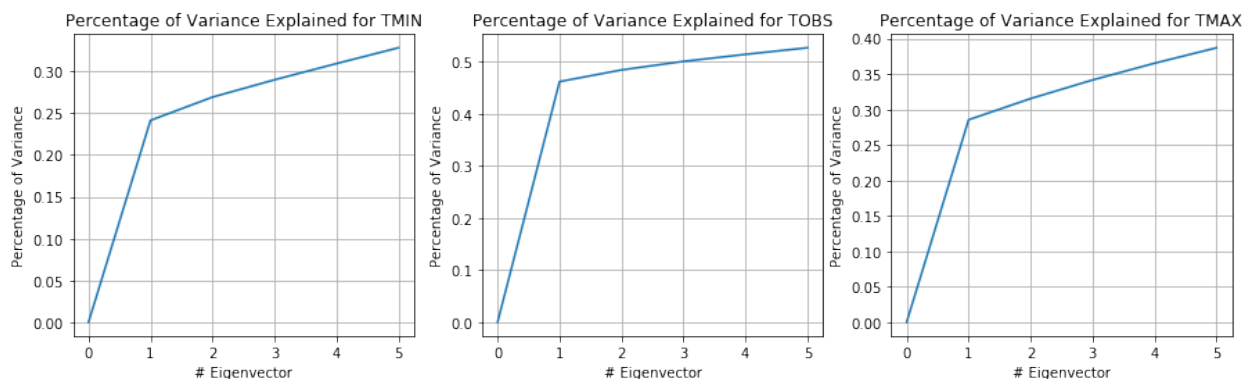
Daily precipitation values obtained from NOAA are fairly constant at about 1 millimeter (mm) per day, with an increase to about 2mm in early summer. When converted from daily to monthly totals, the NOAA data values match the monthly precipitation range of between 25 and 40 mm from the US Climate Data.



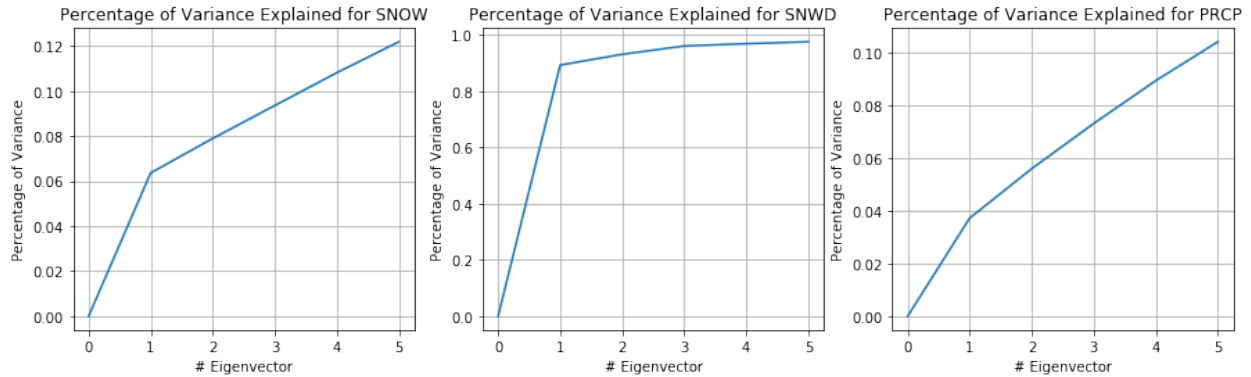
PCA Analysis

For each of the six measurements, we compute the percentage of the variance explained as a function of the number of eigenvectors used. We will focus on the measurements for which the top few eigenvectors explain a significant portion of the data variance.

Percentage of Variance Explained



We see that the top five eigenvectors explain about 33% of variance for TMIN, 53% for TOBS and 39% for TMAX. We conclude that of the three, TOBS is best explained by the top eigenvectors.

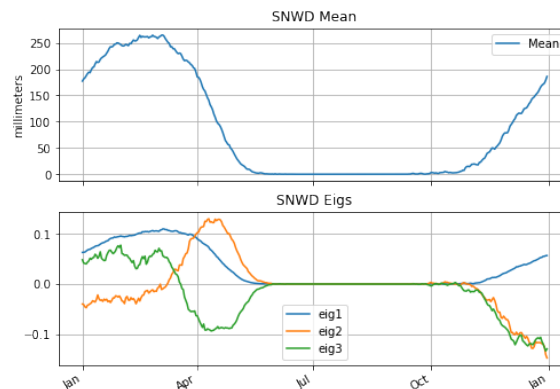


The top five eigenvectors only explain 12% of the variance for SNOW and only 10% for PRCP. On the other hand, the top five eigenvectors explain over 97% of the variance for SNWD.

Based on the percentage of variance explained by the eigenvectors for each measurement, we will perform further analysis of SNWD and TOBS.

Analysis of Snow Depth (SNWD)

The top few eigenvectors for SNWD explain nearly all of the variance within the data. The top single eigenvector explains 89% of the variance, and the top three explain 96% of the variance.



The first eigenvector (**eig1**) has a shape very similar to the mean function. We can interpret the first eigenvector as the overall amount of snow above/below the mean, without changing the distribution of snow depth over time.

The second and third eigenvectors (**eig2** and **eig3**) both oscillate between positive and negative values. They can both be interpreted as changing the distribution of the snow depth over time, without changing the total very much.

They can be interpreted as follows:

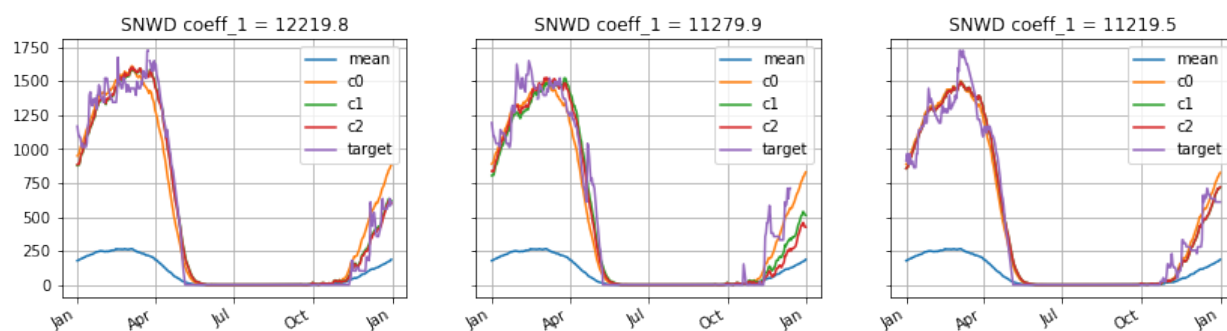
- **eig2:** Less snow depth from November through March, more snow depth from March through May/June
- **eig3:** Less snow depth from November through January, more snow depth from January to mid-March, and less from mid-March through May/June

Reconstructing Specific Station/Year Readings

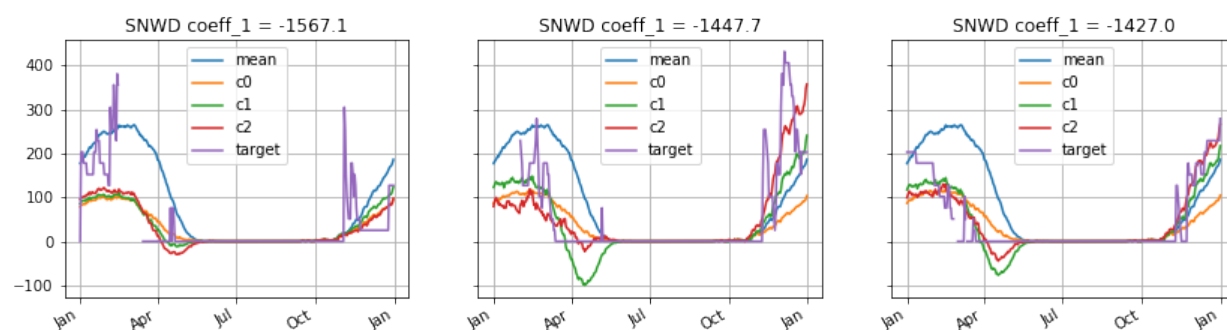
To verify our interpretation of each SWND eigenvector, we can reconstruct specific SNWD measurements from the eigenvectors and associated coefficients for specific stations and years. First, we will look at the station/year combinations for which the eigenvector coefficients are the highest and lowest.

First SNWD Eigenvector

Highest Coefficients:



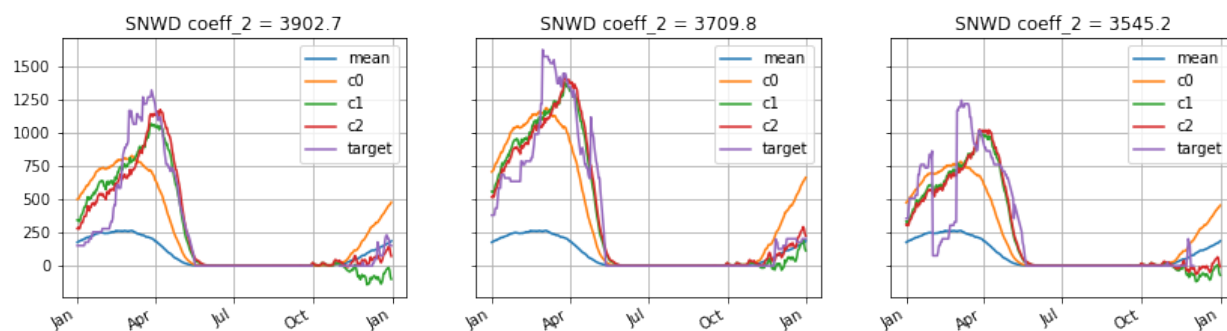
Lowest Coefficients:



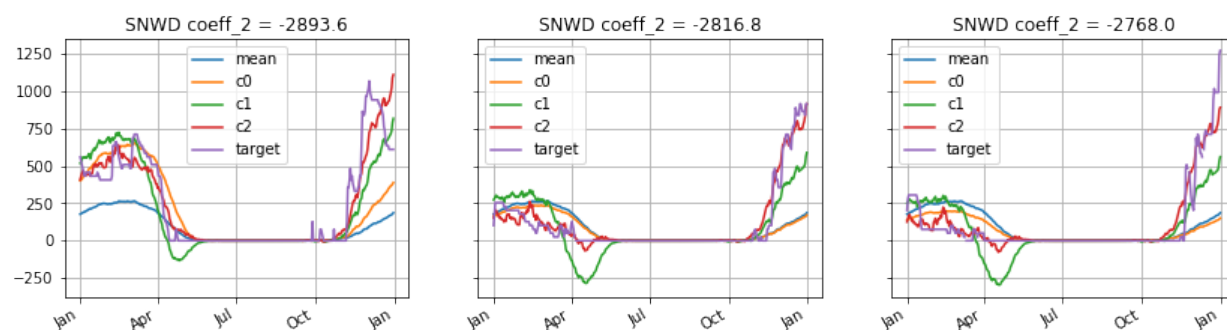
For the highest and lowest values of the first eigenvector, we see much higher and lower than average snow depth readings, respectively, confirming our interpretation of the first eigenvector for SNWD.

Second SNWD Eigenvector

Highest Coefficients:



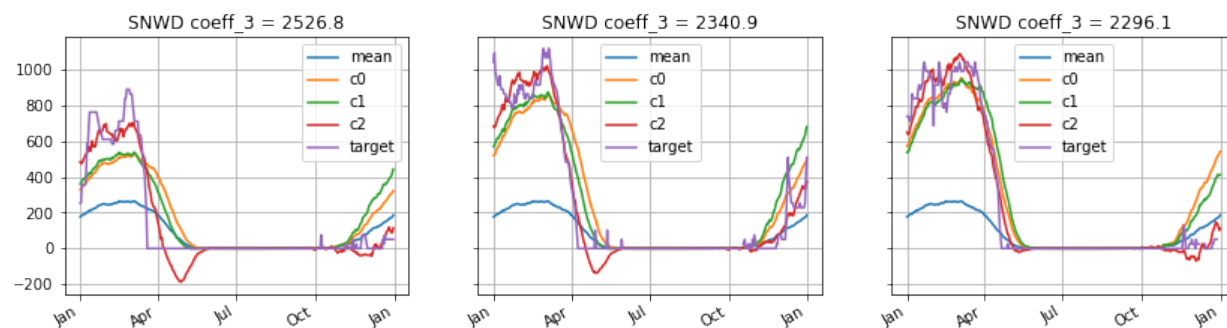
Lowest Coefficients:



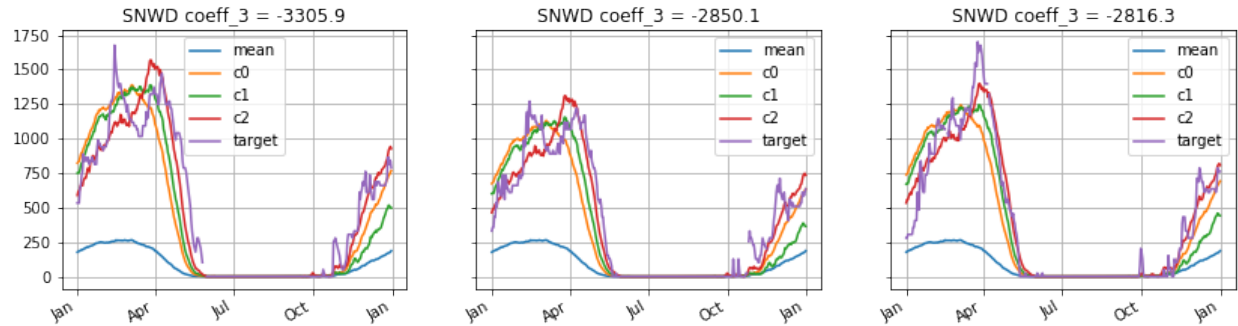
For the second eigenvector, we see that the highest coefficients correspond to stations/years for which snow depth was greatest in the March/April time frame. The lowest coefficients correspond to stations/years that experienced heavy snow accumulation much earlier in the winter season, beginning in the November/December time frame. This verifies our interpretation of the second SNWD eigenvector as impacting the seasonality of snow depth accumulation.

Third SNWD Eigenvector

Highest Coefficients:



Lowest Coefficients:



The third eigenvector behaves in a similar manner to the second eigenvector, in the sense that it captures some of the variation of SNWD in terms of seasonality. The highest coefficients correspond to higher levels of snow accumulation occurring in the January-March time frame, with less snow depth after April. When compared to the high coefficient three data points, the lowest coefficient data points experienced higher snow depths in November-December, and especially higher snow accumulation in March-April.

Analyzing Effect of Year vs. Effect of Station on SNWD Coefficients

We now assess the extent to which variation of the type captured by each SWND eigenvector is due to differences between years versus differences between individual stations. These impacts are measured by determining the fraction by which the variance is reduced when we subtract from each station/year entry the average-per-year or the average-per-station respectively. Variance is calculated using the Root Mean Square (RMS) for the eigenvector coefficients for all station/year entries.

Coefficient 1:

total RMS = 4617.16
 RMS after removing mean-by-station = 1710.13, fraction explained = 63.0%
 RMS after removing mean-by-year = 2719.15, fraction explained = 41.1%

Coefficient 2:

total RMS = 1095.87
 RMS after removing mean-by-station = 979.35, fraction explained = 10.6%
 RMS after removing mean-by-year = 678.85, fraction explained = 38.1%

Coefficient 3:

total RMS = 985.66
 RMS after removing mean-by-station = 910.57, fraction explained = 7.6%
 RMS after removing mean-by-year = 605.87, fraction explained = 38.5%

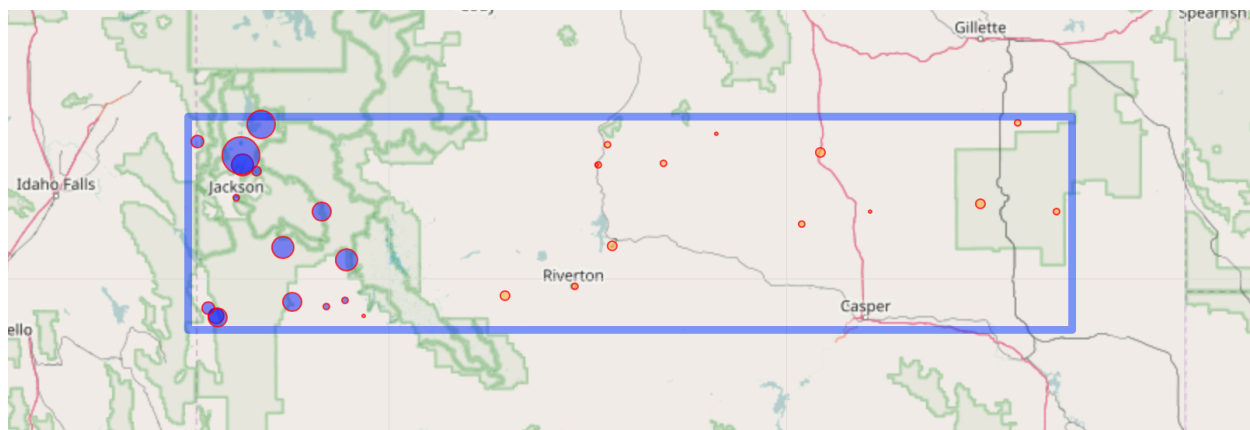
From the RMS results, we can determine that the variation of coefficient 1 (which adjusts the mean) is due more to differences in station than to differences in year. This makes intuitive sense, as we would expect certain stations to consistently have more snow depth than other stations, regardless of year.

On the other hand, coefficients 2 and 3 (which both adjust the seasonality profile of snow depth) are influenced more by year than station. This also makes intuitive sense, as we would expect seasonal variation to be more consistent from year to year than from station to station.

Geographical Analysis of SNWD Coefficients

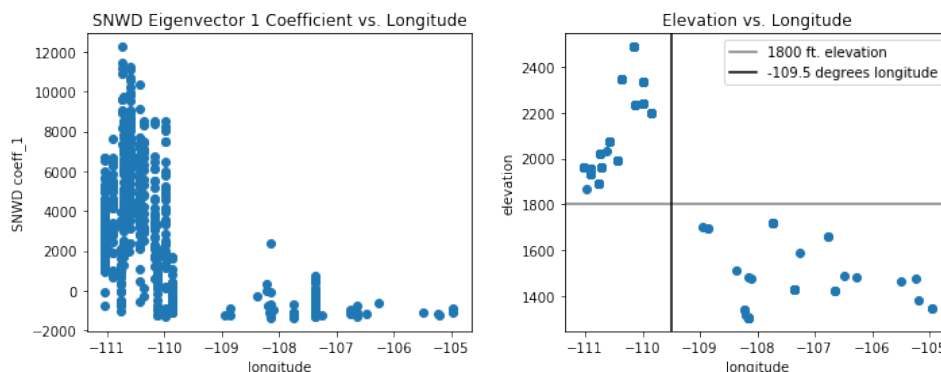
Since SNWD coefficient 1 variation is due primarily to differences in station, we can see if there is any geographical pattern to the values of coefficient 1. We will do so by plotting the average coefficient 1 value for each station on a map of the region, utilizing the latitude and longitude information for each station.

In the plot below, positive values for SNWD coefficient 1 are colored blue, while negative values are colored orange. The magnitude of the average coefficient for each station is depicted by the size of the circle at each location.



We can see clearly that stations to the west of the state tend to have positive values for SNWD coefficient 1, while those to the east tend to have negative values. Recall that a positive value for SNWD coefficient 1 corresponds to increasing snow depth from the mean, so we can say that western stations experience above average snow depth, while stations to the east experience below average snow depth.

One likely reason why snow depth tends to be higher in the western portion of our region is due to elevation. To investigate this, we plot SNWD coefficient 1 vs. longitude and elevation vs. longitude below, showing that the likely cause of higher coefficient 1 values for SNWD is the increase in elevation that occurs in the western region. In fact, we can see that all stations west of -109.5 degrees longitude are above 1,800 ft. elevation, while all stations east of -109.5 degrees longitude are below the 1,800 ft. elevation threshold.



Not surprisingly, the statistical test of equal means for coefficient 1 values for SNWD based on these longitude/elevation thresholds produce a highly statistically significant difference:

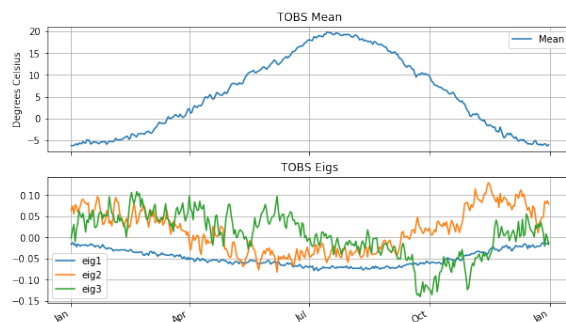
p-value for equal means east and west of -109.5 degrees longitude = $1.88\text{e-}36$

p-value for equal means above and below 1800 ft elevation = $1.88\text{e-}36$

This indicates that values for SNWD coefficient 1, which is related to the mean snow depth, is statistically different for western/high elevation regions.

Analysis of Temperature at the Time of Observation (TOBS)

The top three eigenvectors for TOBS explain over 50% of the variance within the data, and the top single eigenvector explains 46% of the variance.



The first eigenvector (**eig1**) has a shape very similar to the mean function, but all values are negative. We can interpret the first eigenvector as the overall amount of variation in observed temperature above/below the mean, without changing the distribution in temperature over time. Because all values for the first eigenvector are negative, positive coefficient values will represent a decrease in temperature from the mean, and negative coefficient values will represent an increase from the mean.

The second and third eigenvectors (**eig2** and **eig3**) both oscillate between positive and negative values. They can both be interpreted as changing the distribution of temperature over the course of the year, without changing the total very much.

They can be interpreted as follows:

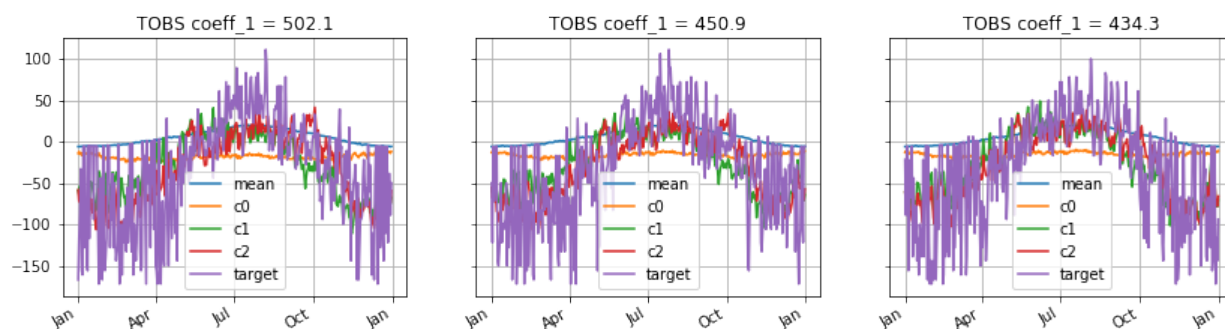
- **eig2**: Higher temperature from January through April, lower temperature from April through September, and higher temperature from October through December
- **eig3**: Higher temperature from January through July, lower temperature from July through November, with a significant drop in October, and higher temperature in December

Reconstructing Specific Station/Year Readings

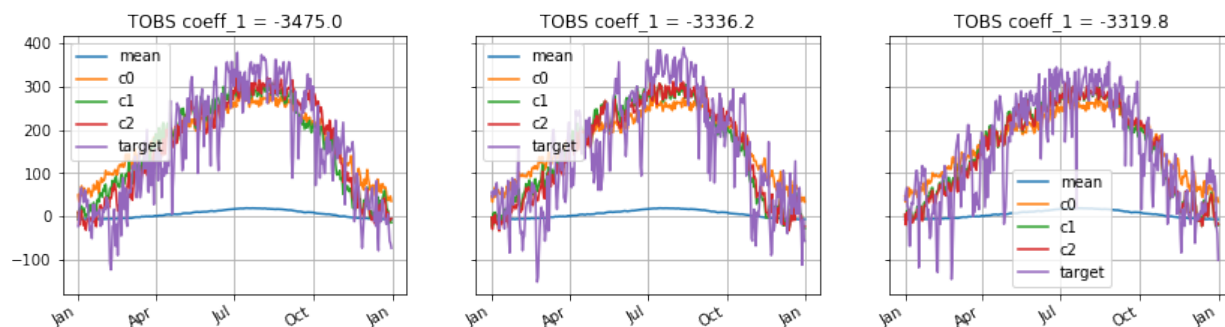
To verify our interpretation of each TOBS eigenvector, we can reconstruct specific TOBS measurements for specific stations and years. First, we will look at the station/year combinations for which the **eig1** coefficient are the highest and lowest. Note that reconstructed temperature values for TOBS here are in tenths of degree Celsius rather than whole degrees Celsius.

First TOBS Eigenvector

Highest Coefficients:



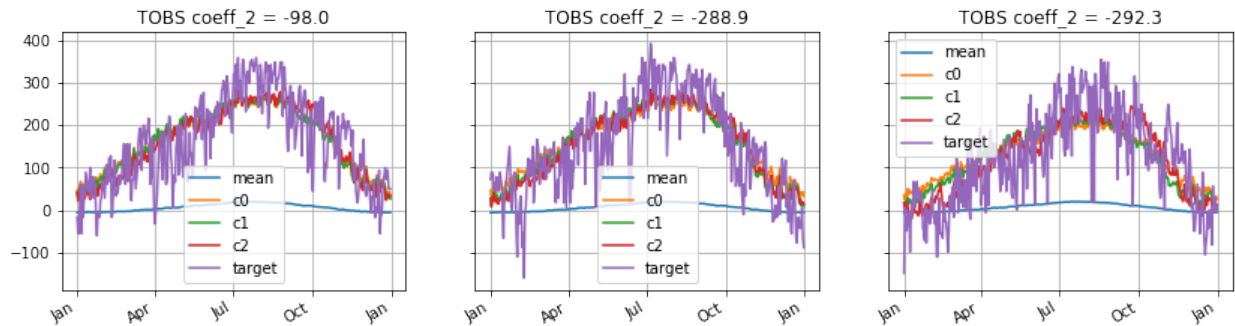
Lowest Coefficients:



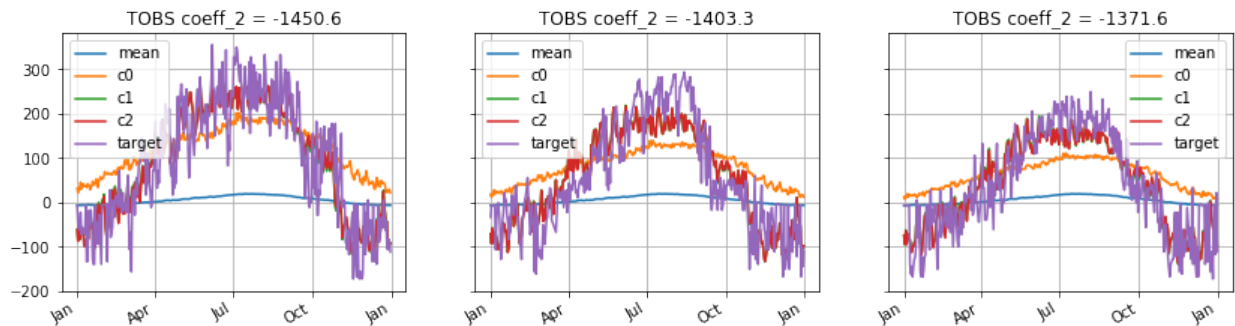
For the highest and lowest values of the first eigenvector, we see much lower and higher than average TOBS readings, respectively, confirming our interpretation of the first eigenvector for TOBS as an adjustment to the mean.

Second TOBS Eigenvector

Highest Coefficients:



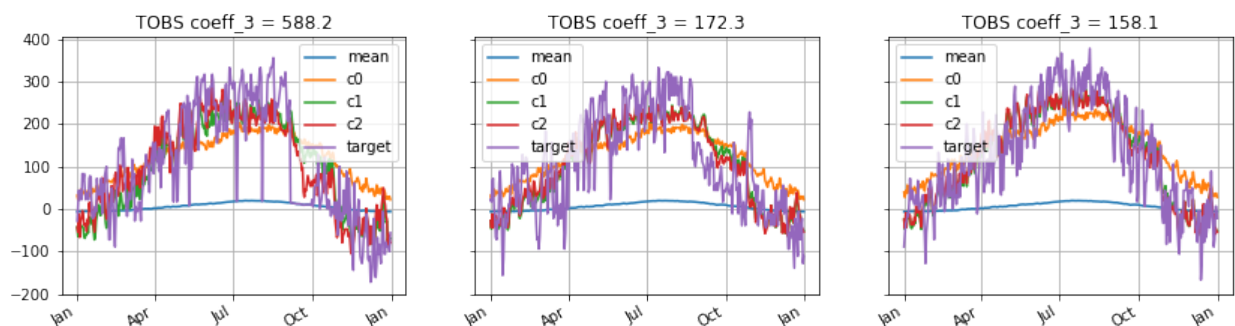
Lowest Coefficients:



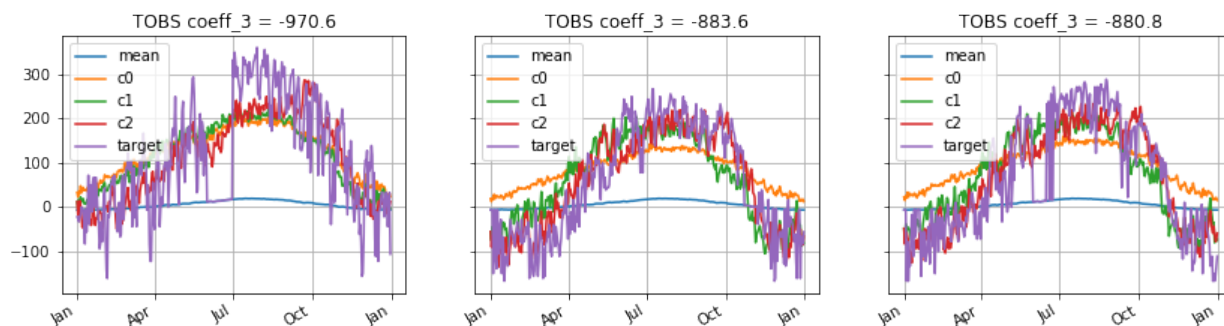
For the second eigenvector, we see that the lowest coefficients seem to have a faster increase in temperature after April, and a faster dropoff in temperature before October. The lowest coefficient plots also have a much lower temperature late in the year. Described simply in words, you might say that the lowest values of the second coefficient most correspond to a short summer period of high temperature. Our interpretation that the second TOBS eigenvector captures the seasonality of temperature seems to be confirmed.

Third TOBS Eigenvector

Highest Coefficients:



Lowest Coefficients:



Again, similar to the case with SNWD, the third TOBS eigenvector captures some of the seasonal variation in the data. The highest coefficients correspond to a faster temperature dropoff in October, while the data points with the lowest coefficient values have temperature readings that do not drop off until after October. This observation aligns to the high magnitude negative values of the third eigenvector around October.

Analyzing Effect of Year vs. Effect of Station on TOBS Coefficients

As we did with SNWD, we now assess the extent to which variation of the type captured by each TOBS eigenvector is due to differences between years versus differences between individual stations. Recall that as was done for SNWD, these impacts are measured by determining the fraction by which the variance (captured by RMS) is reduced when we subtract from each station/year entry the average-per-year or the average-per-station respectively.

Coefficient 1:

total RMS = 1651.88
 RMS after removing mean-by-station = 434.16, fraction explained = 73.7%
 RMS after removing mean-by-year = 730.79, fraction explained = 55.8%

Coefficient 2:

total RMS = 872.00
 RMS after removing mean-by-station = 152.98, fraction explained = 82.5%
 RMS after removing mean-by-year = 150.82, fraction explained = 82.7%

Coefficient 3:

total RMS = 403.71
 RMS after removing mean-by-station = 154.24, fraction explained = 61.8%
 RMS after removing mean-by-year = 116.65, fraction explained = 71.1%

From the RMS results, we can determine that the variation of coefficient 1 (which adjusts from the mean observed temperature) is due more to differences in station than to differences in year. This makes intuitive sense, as we would expect certain stations to consistently be warmer or colder than other stations, regardless of year.

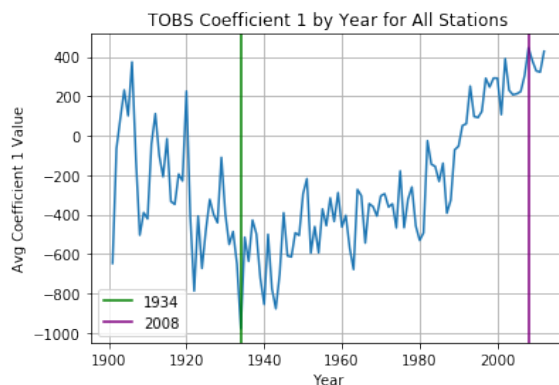
For coefficient 2, which we described above as generally capturing the length of summer, the variation seems to be due equally to differences in station location and differences in year.

Coefficient 3 is influenced more by year than station. This makes sense, as we would expect seasonal variation in temperature to be more consistent from year to year than from station to station.

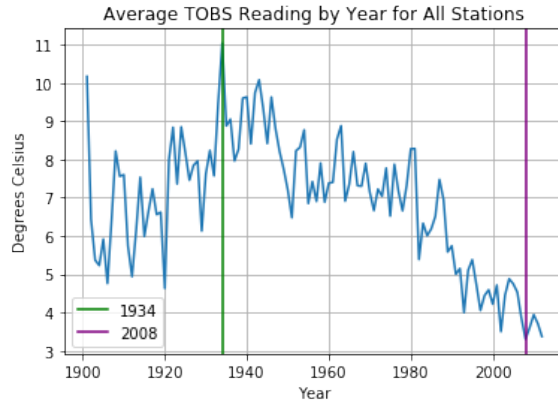
Temporal Analysis of TOBS Coefficients

Above, we interpreted the first TOBS eigenvector as adjusting the temperature from the mean. In an attempt to confirm our understanding of the first eigenvector coefficient for TOBS, we can compare the average of the first eigenvector coefficient by year to the average actual TOBS reading by year. If our interpretation of the first eigenvector is correct, we should expect to see a strong relationship between the eigenvector coefficients and the actual average TOBS values for each year.

First, we will plot the average value coefficient 1 for all stations by year. The years in which the maximum and minimum coefficient values occurred are indicated by the vertical lines.

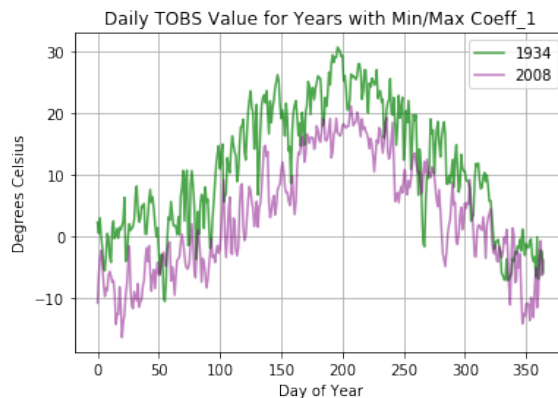


Next, we can plot the actual average TOBS value for all stations over the course of each year. Again, the years in which the minimum and maximum average TOBS values were recorded are marked by vertical lines.



Notice that the years for which the minimum and maximum values of the first eigenvector coefficient for TOBS occurred are the same years for which average TOBS readings across all stations were observed. Also notice the strong correspondence between the shapes of the coefficient 1 and average TOBS readings by year. In fact, the general graph shape of TOBS readings is nearly identical to the shape of the average yearly coefficients if either of the graphs were to be flipped vertically. Considering the fact that all values for the first TOBS eigenvector coefficient were negative, the inverse relationship between the coefficient and actual TOBS values makes sense. Thus, it seems that our initial interpretation that the first eigenvector coefficient adjusts the mean TOBS was correct.

Finally, in order to sanity check the above TOBS readings, we plot the actual average TOBS values for the years identified as having the overall highest and lowest readings. Indeed, the readings for 2008 are lower across the year than are the readings for 1934.



We can verify this difference statistically by performing a paired t-test with a null hypothesis of equal means:

$$H_0: \text{Average TOBS}_{1934} = \text{Average TOBS}_{2008}$$

$$H_A: \text{Average TOBS}_{1934} \neq \text{Average TOBS}_{2008}$$

The paired t-test p-value is 1.52×10^{-83} , so we can reject H_0 in favor of H_A . In other words, there is very strong statistical evidence that the temperature in 2008 was lower than the temperature in 1934 across our region in Wyoming.