Wyoming Weather Analysis

This is a report on the historical analysis of weather patterns in an area that approximately overlaps the area of the state of Massachusets.

The data we will use here comes from [NOAA](https://www.ncdc.noaa.gov/). Specifically, it was downloaded from This FTP site.

We focused on six measurements:

* **TMIN, TMAX:** the daily minimum and maximum temperature.
* **TOBS:** The average temperature for each day.
* **PRCP:** Daily Percipitation (in mm)
* **SNOW:** Daily snowfall (in mm)
* **SNWD:** The depth of accumulated snow.

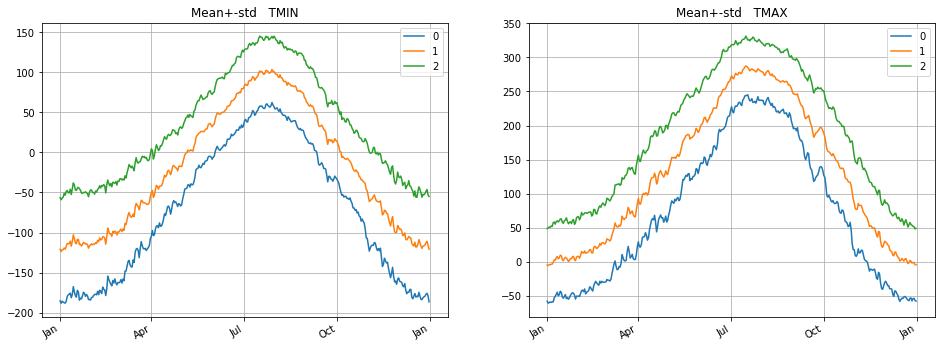
Wyoming has a characteristic with July high temperatures averaging between 29 and 35 °C in most of the state. In few location it drops rapidly to around 21 °C, these regions have good amount of snow. Snowfall in some of the regions is high. This difference in climate changes within in a state explains the cause of tornado in the south eastern parts of the state.

### **Data validation**

First step in data analysis is to make sure that data being used for analysis is correct. In this section, we provide comparison with general statistics from <http://www.wrcc.dri.edu/cgi-bin/cliMAIN.pl?id2260> to validate that data is good for further analysis.

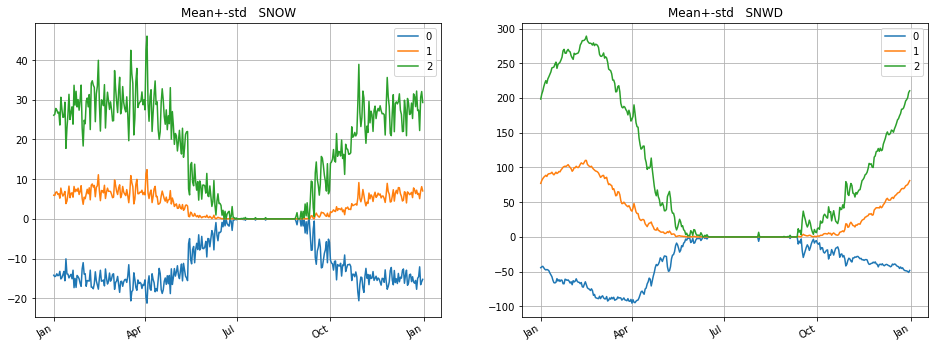
#### **Comparison of Temperature (TMIN, TMAX) data**

In this section, we compare TMIN, TMAX measurements from the data being used for analysis and external source. Comparing these data, our data used for analysis agrees with the external source.



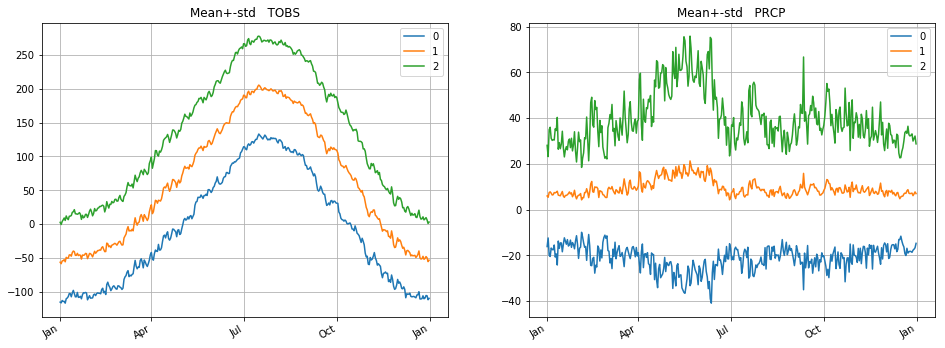
#### **Comparison of SNOW fall and Snow depth data data**

In this section, we compare SNOW fall and SNOW depth measurements from the data being used for analysis with external source. Peak of average snow depth is observed during months of January and February.



#### **Comparison of Precipitation data**

In this section, we compare precipication data.

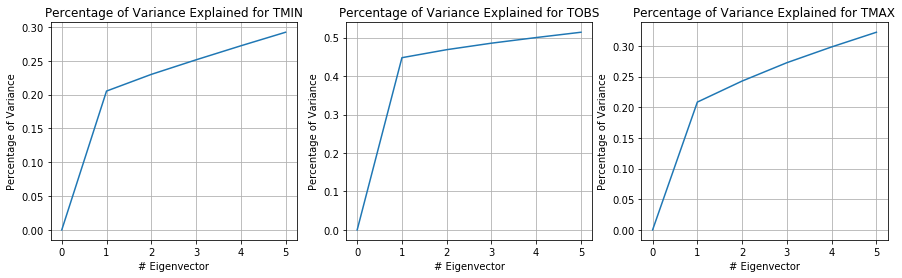


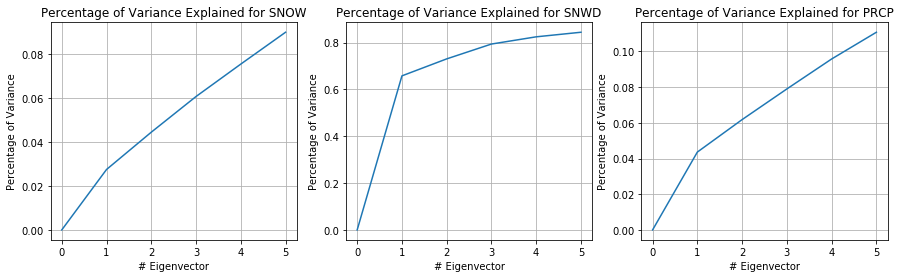
## **PCA analysis**

For each of the six measurement, we compute the percentage of the variance explained as a function of the number of eigen-vectors used.

We see that the top 5 eigen-vectors explain 30% of variance for TMIN, 50% for TOBS and 35% for TMAX.

We conclude that of the three, TOBS is best explained by the top 5 eigenvectors. This is especially true for the first eigen-vector which, by itself, explains 45% of the variance.





The top 5 eigenvectors explain 11% of the variance for PRCP and 9% for SNOW. Both are low values. On the other hand the top 5 eigenvectors explain 90% of the variance for SNWD. This means that these top 5 eigenvectors capture most of the variation in the snow signals. Based on that we will dig deeper into the PCA analysis for snow-depth.

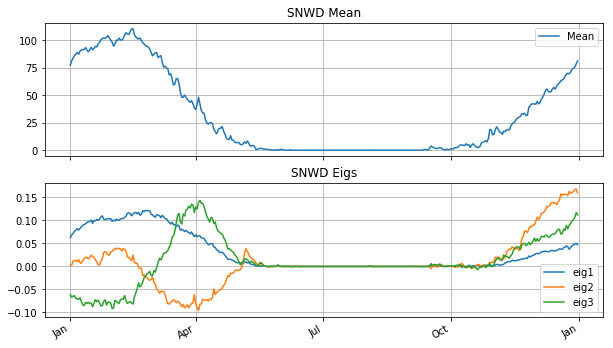
It makes sense that SNWD would be less noisy than SNOW. That is because SNWD is a decaying integral of SNOW and, as such, varies less between days and between the same date on different years. Especially in the months of January and February when the snow fall is more.

## **Analysis of snow depth**

We choose to analyze the eigen-decomposition for snow-depth because the first 3 eigen-vectors explain 80% of the variance.

First, we graph the mean and the top 3 eigen-vectors.

We observe that the snow season is from start of november to the end of march, where around the end of February marks the peak of the snow-depth.



Next we interpret the eigen-functions. The first eigen-function (eig1) has a shape very similar to the mean function. The main difference is that the eigen-function is close to zero during october-november while the mean is not. The interpretation of this shape is that eig1 represents the overall amount of snow above/below the mean, but without changing the distribution over time.

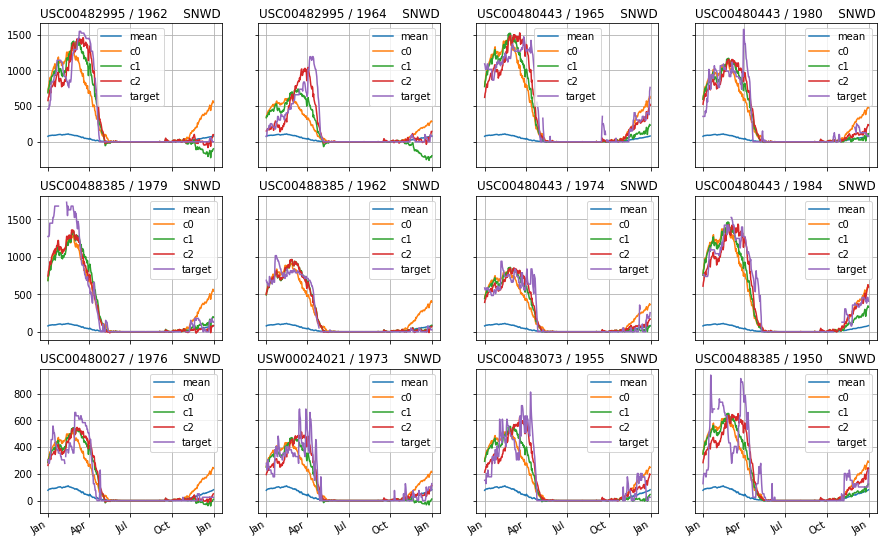
**Eig2 and eig3** are similar in the following way. They all oscilate between positive and negative values. In other words, they correspond to changing the distribution of the snow depth over the winter months, but they don't change the total (much).

They can be interpreted as follows:

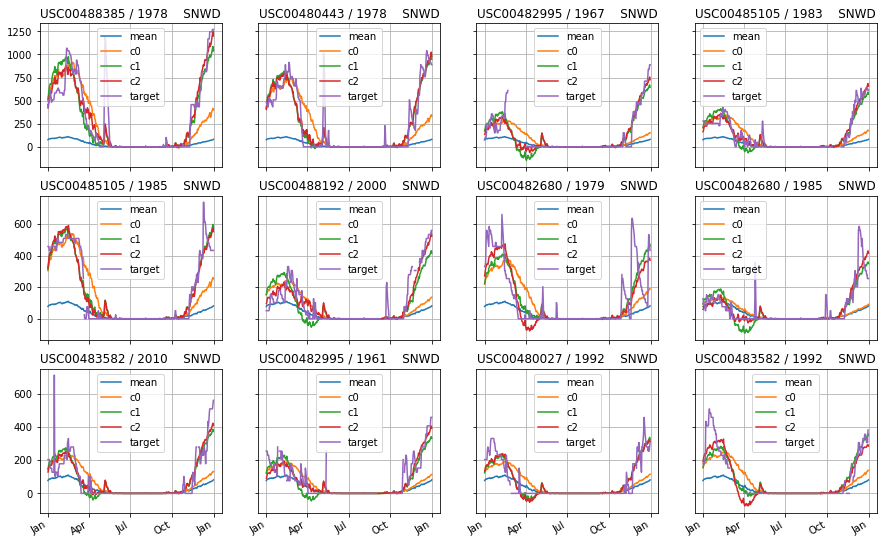
* **eig2:** less snow in jan - mid feb, more snow in start of march and very less snow towards end of march.
* **eig3:** less snow in jan, less snow in feb, more snow in march.

### **Examples of reconstructions**

Coeff1 postive



**Coeff2 postive**



**Coefficient1 values increases from jan to feb to early march pointing to the increase in the snowfall.**

**Coefficeint 2 values show that after january and feburary, towards the march snowfall decrease**

### **Conclusions**

**Based on our statistical significance tests and PCA analysis, we have noticed few observations.**

* **There is not much change in precipitation through out the year. Hence larger dimensions are required even in projected coordinate system.**
* **Snow depth has a strong correlation with minimum temperature**
* **Performing statistical analysis using PCA technique, we observe that SNWD is described by first few principle components. Around 65% of the SNWD is captured with just one eigen vector. With 3 eigen vectors, 80% of the information is captured. Precipitation on the contrary required large number of dimensions.**