

1 Alabama - Florida 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 Massachusetts.

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

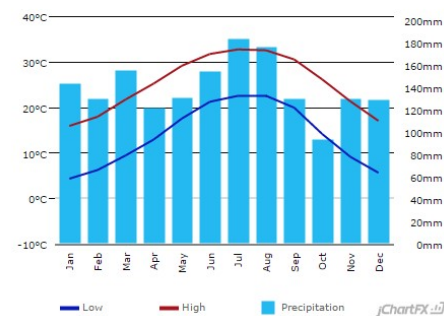
We focused on six measurements:

- **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.

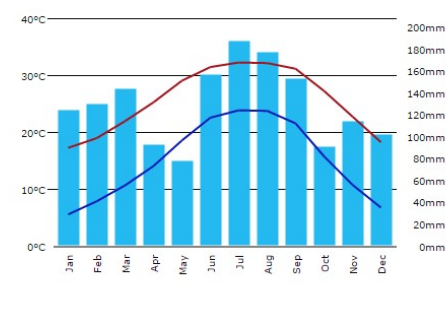
1.1 Sanity-check: comparison with outside sources

We start by comparing some of the general statistics with graphs that we obtained from a site called [US Climate Data \(http://www.usclimatedata.com/climate/boston/massachusetts/united-states/usma0046\)](http://www.usclimatedata.com/climate/boston/massachusetts/united-states/usma0046). **Mobile, AL** and **PanamaCity, FL** are the two major cities in the give region(of BSSBSBS indexed data). The following graph below shows the daily minimum and maximum temperatures per month, as well as the total precipitation per month, for these two cities.

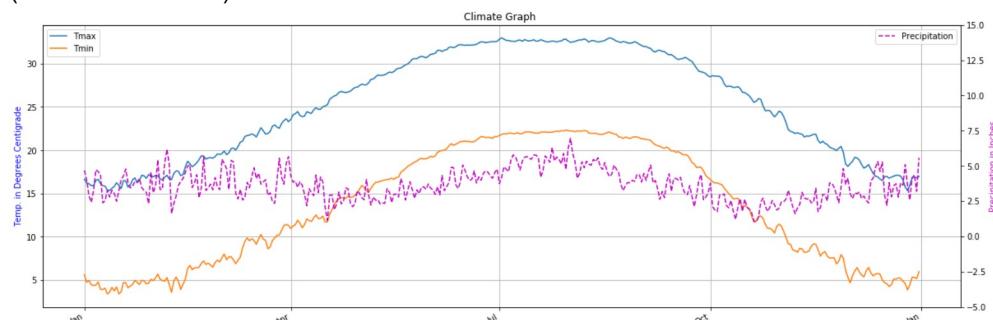
Mobile Climate Graph - Alabama Climate Chart



Panama City Climate Graph - Florida Climate Chart



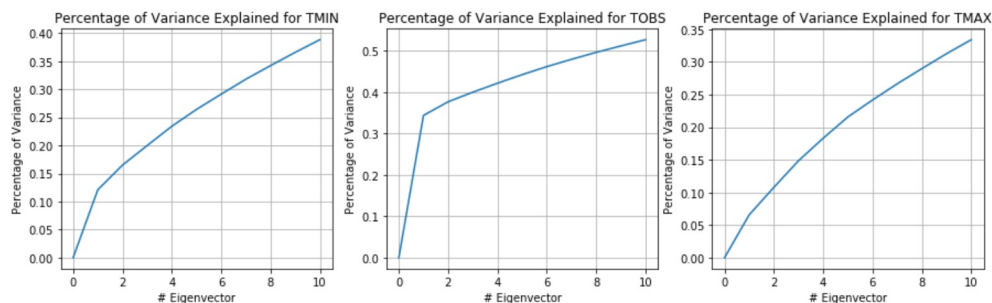
The following Climate graphs shows the mean of Tmax , Tmin and Precipitation for the current BSSBSBS region. We see that the min and max daily temperature agree with the ones we got from our data as shown below. We can see the temperatures are ranging from 32°C to 5°C. Constant Precipitation through out the year ranging between 4.7 - 6 inches (120mm to 160 mm)



1.2 PCA analysis

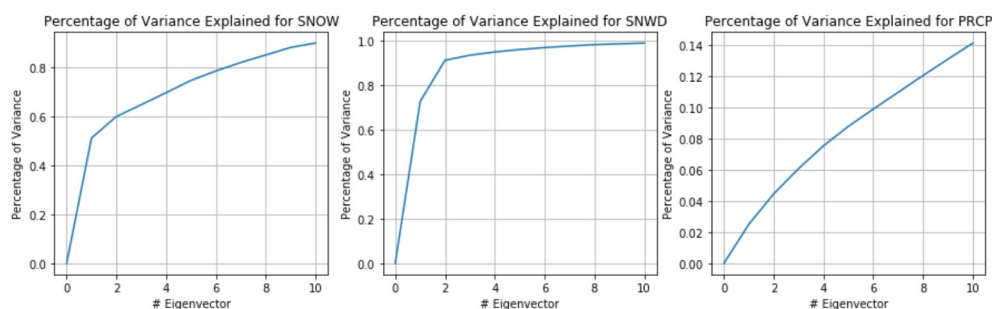
For each of the six measurement, we compute the percentage of the variance explained as a function of the number of eigen-vectors used.

1.2.1 Percentage of variance explained.



We see that the top 5 eigen-vectors explain 27% of variance for TMIN, 45% for TOBS and 20% 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 35% of the variance.



The top 5 eigenvectors explain 9% of the variance for PRCP. But for SNOW and SNWD, the top 5 eigenvectors explain 75% and 95% of the variance. 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.

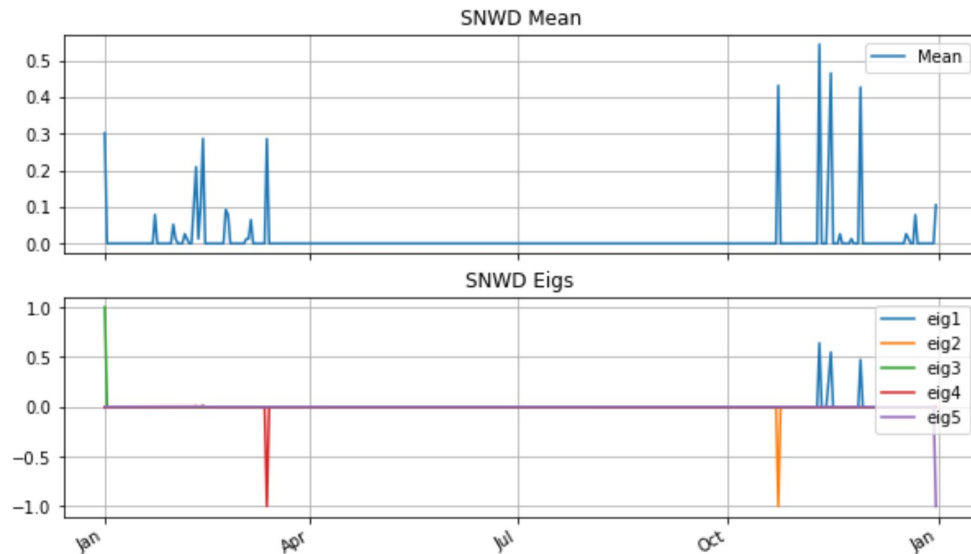
It makes sense that SNWD would be less noisy than SNOW. That is because SNWD is a decaying integral of SNOW and, as such, varies less between days and between the same date on different years.

1.3 Analysis of snow depth

We choose to analyze the eigen-decomposition for snow-depth because the first 3 eigen-vectors explain 90% of the variance.

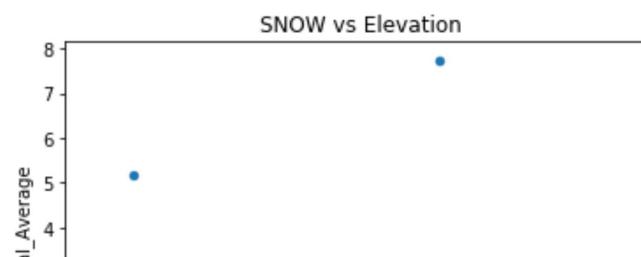
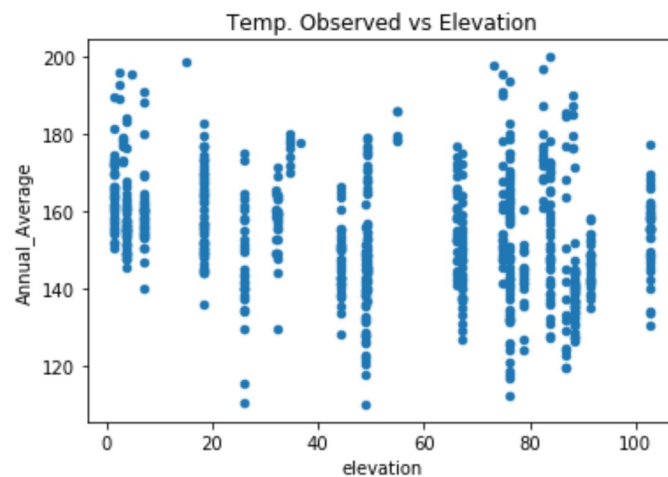
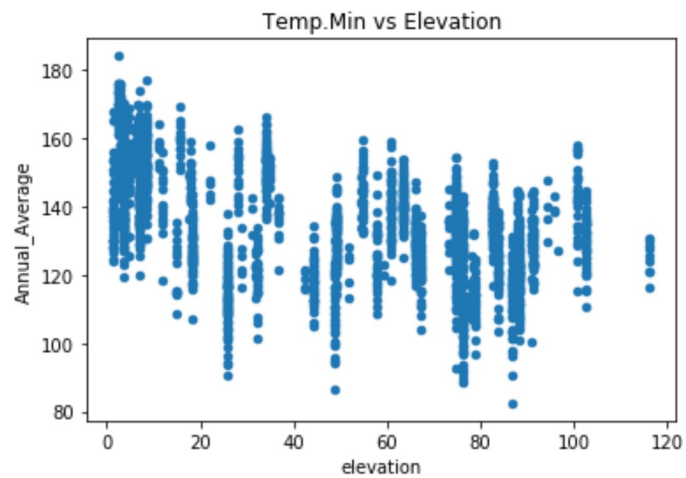
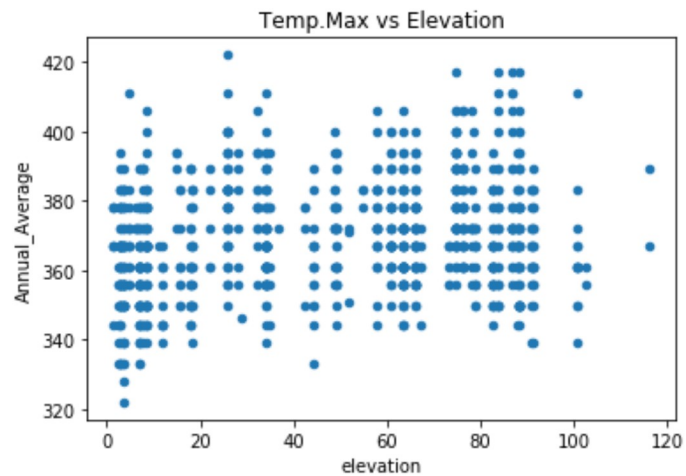
First, we graph the mean and the top 5 eigen-vectors.

By looking at the SNWD mean , we understand that there is no snow at all, in this region. And there is only little variation in the SNWD and the top eigen vectors are trying to capture this little variation.



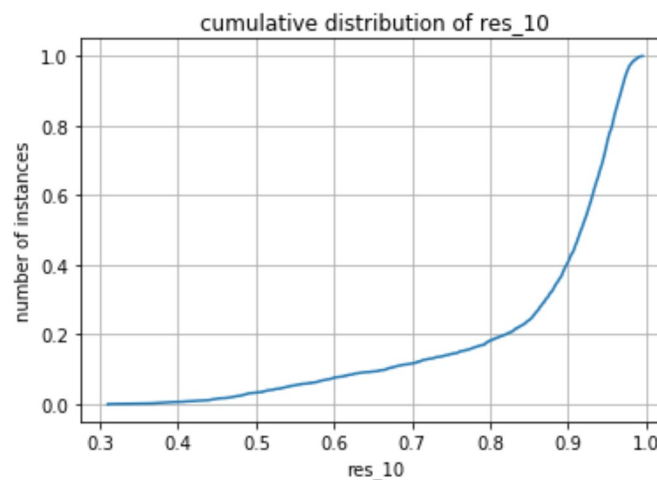
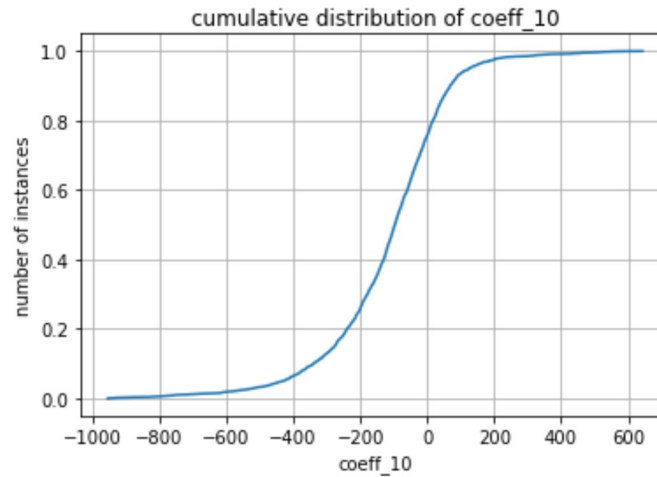
1.4 Analysis with respect to Elevation

We try to understand if there is any correlation between the elevation and measurements observed. Here are different graphs which are just scatter plots of Annual average of measurements with respective elevations

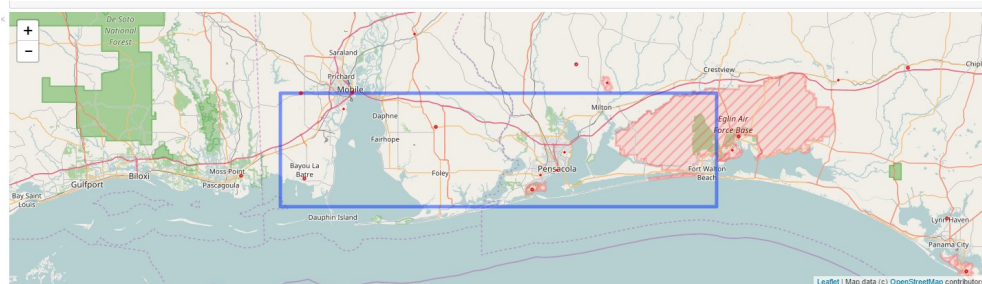


1.5 Analysis with reconstruction of PRCP

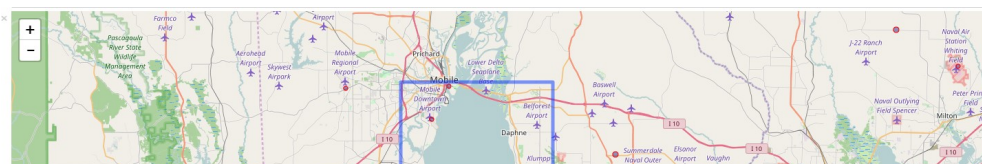
From "percentage of variance explained vs no. of eigen vectors" graph, it is understood that there is lot of variance in the precipitation data. Lets consider reconstructing the data with considering top 10 eigen vectors which will explain 14% of variance. The distribution of **residual** and **coefficient** after subtracting mean and projection on top 10 eigen vectors is as follows :



Lets try plotting the regions with highest coeff_10, when the corresponding residue 'res_10' is small

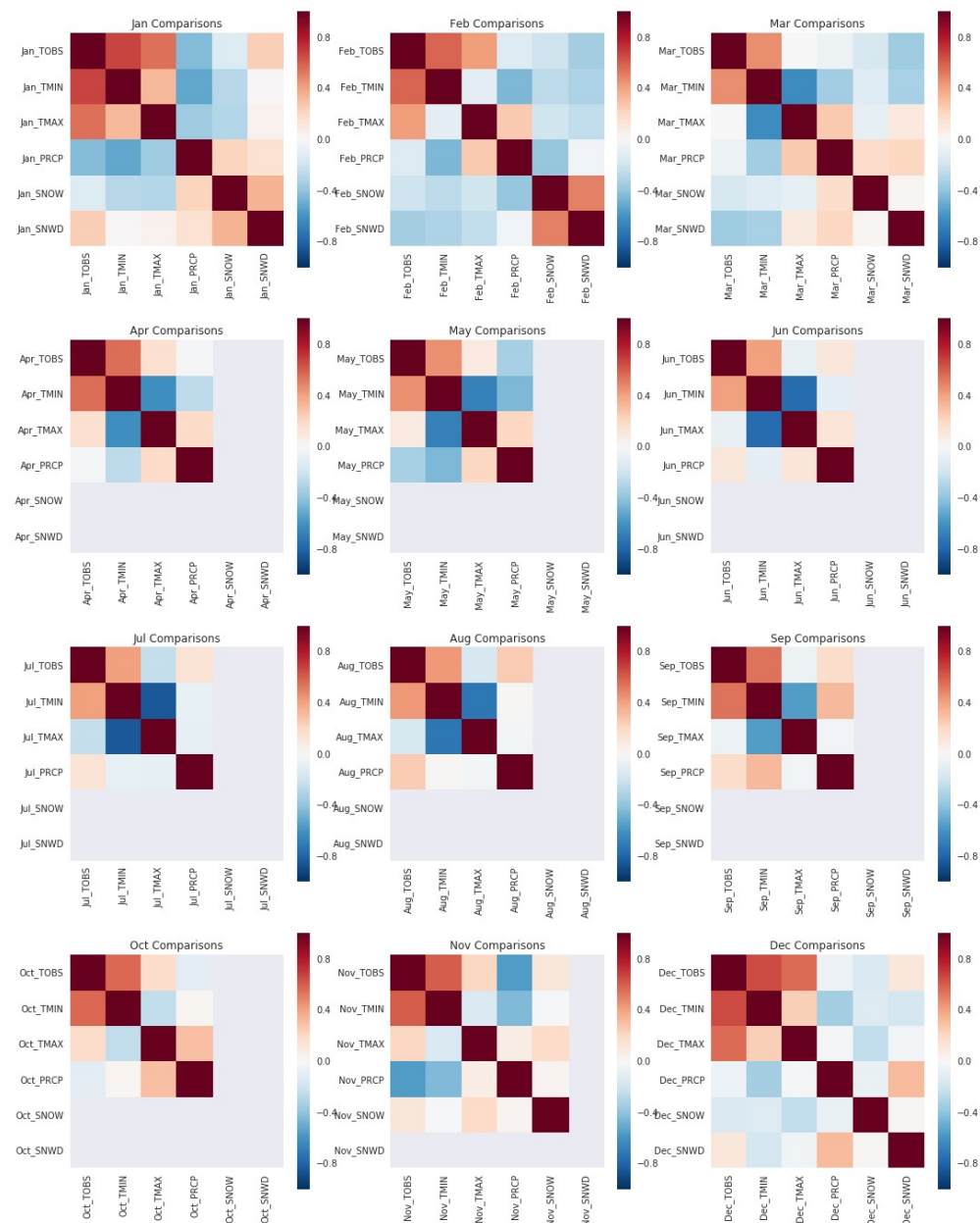


And this is the region with lowest coeff_10, when the corresponding 'res_10' is small



1.6 Heatmap of monthly observations

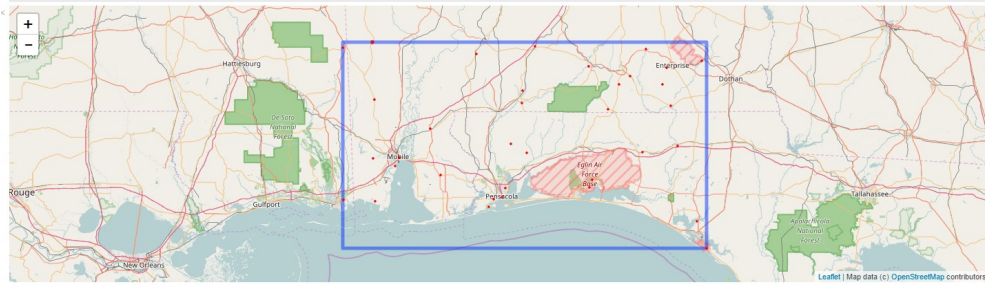
Let's try analyzing heat map of monthly observations , to observe any correlation between observations.



As we can see the heat map does not convey much information, there is no correlation in annual observations

1.7 Conclusion

For the given region:



- The temperature and precipitations are high in the summers, suggesting a tropical climate.
- There is no strong correlation/explanation for the distribution/variation of Precipitation data.

In []: