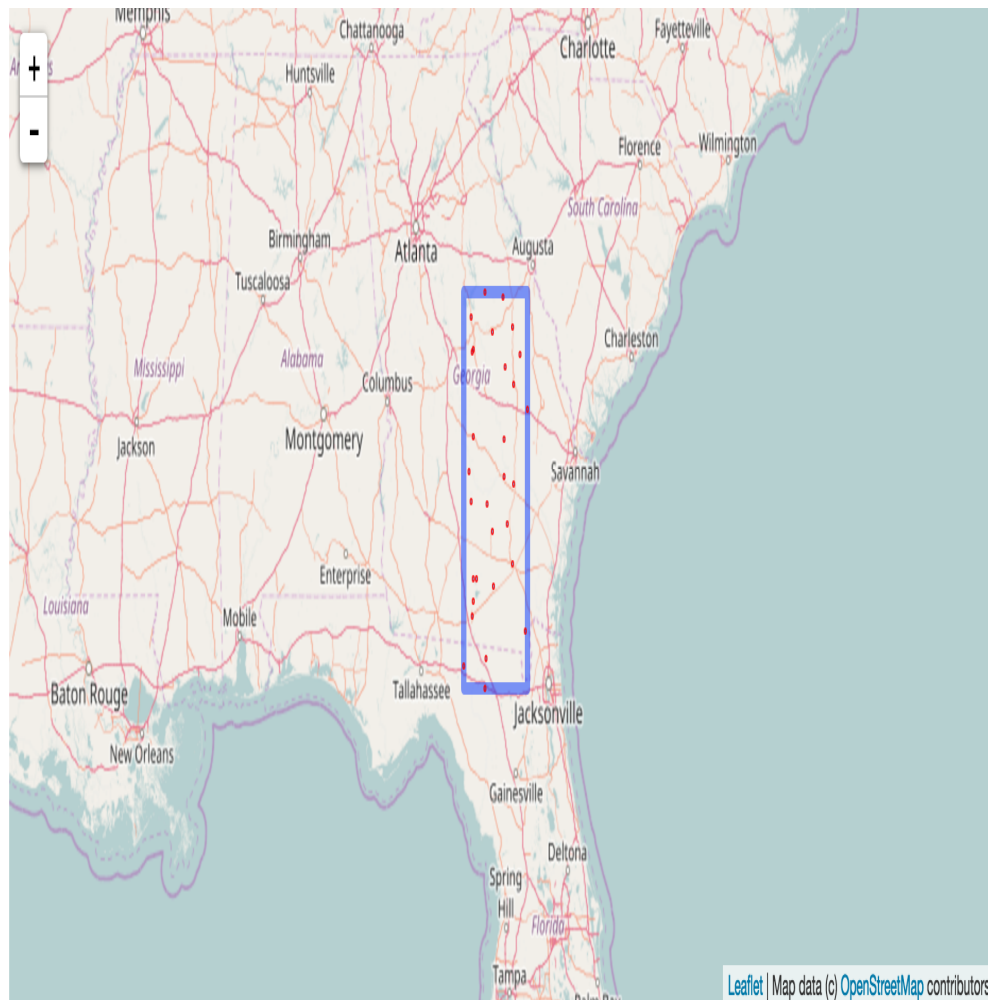


1 Southeast Georgia and North Florida Weather Analysis

This is a report on the historical analysis of weather patterns in an area that overlaps the area of the state of Georgia and partially the state of Florida, with latitudes and longitudes spanning from 33.326 to 30.185 and -83.468 to -82.037. The entire state, including the North Georgia mountains, receives moderate to heavy rain, which varies from 45 inches (1,100 mm) in central Georgia. Most of Georgia has a sub-tropical climate tempered some by occasional climax air masses in the winter, hot/humid summers are typical, except at the highest elevations.



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:

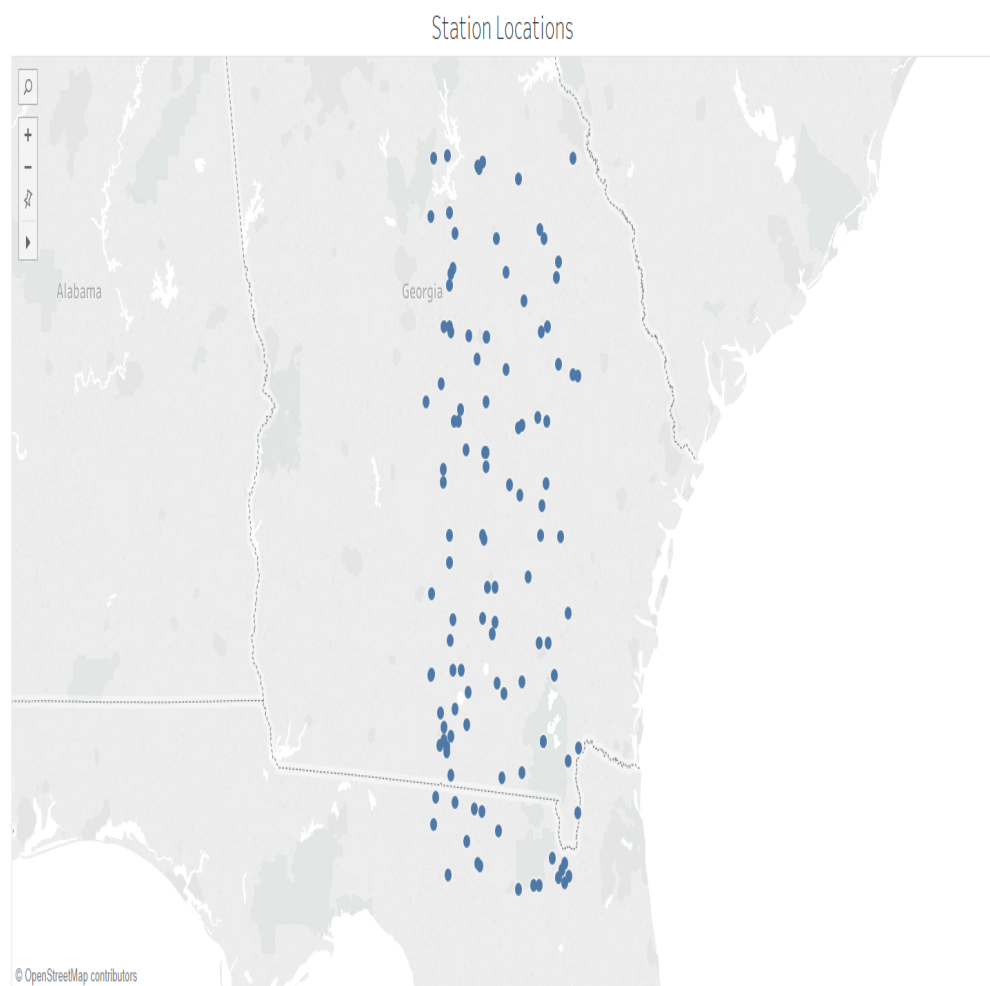
- **TMIN, TMAX:** the daily minimum and maximum temperature.
- **TOBS:** The average temperature for each day.
- **PRCP:** Daily precipitation (in mm)
- **SNOW:** Daily snowfall (in mm)

- **SNWD:** The depth of accumulated snow (in mm).

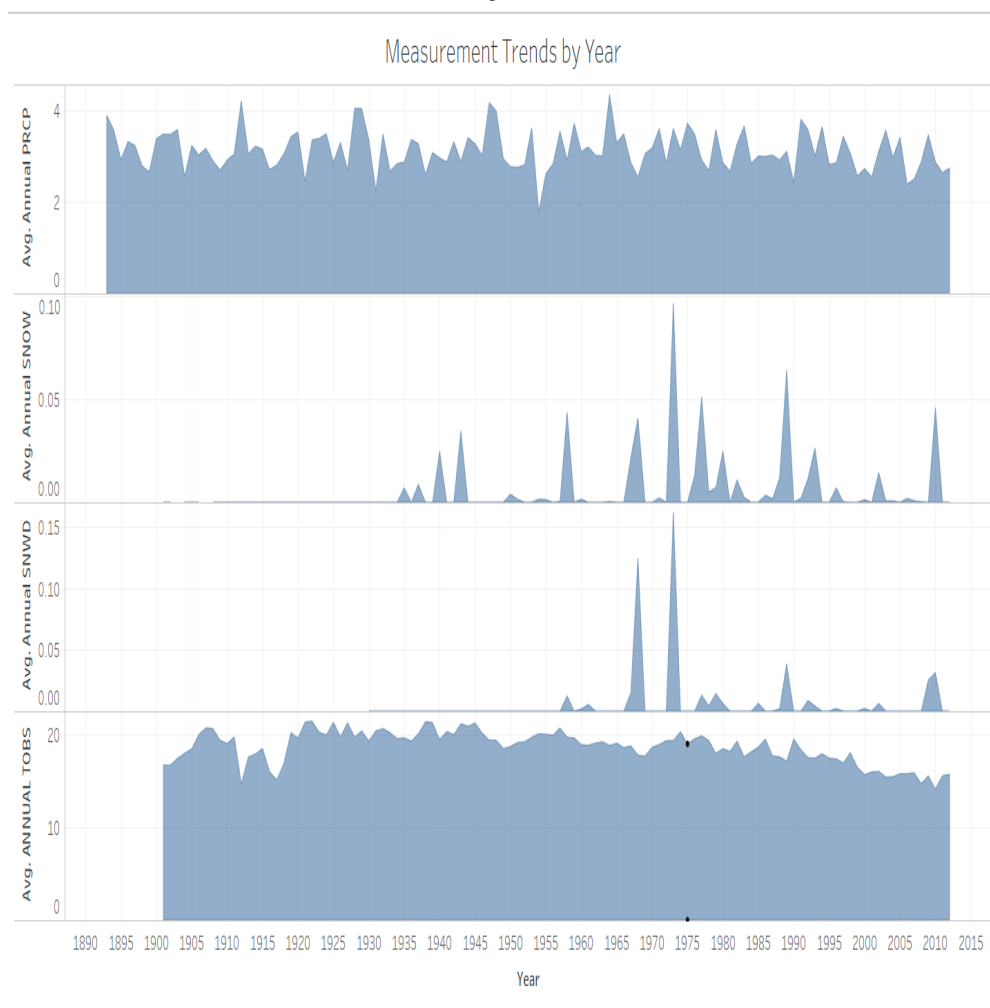
1.1 The data was initially loaded into Tableau for simple data exploration.

The vectorized values for each of the 6 values were extracted from the parquet file, the average measurement by station and year was calculated, and the data was outer joined forming a shape of (3006, 6) to be inputted into Tableau. Tableau is a business intelligence tool which allows for quick, informative, and interactive visualization (see supplementary link for full Tableau dashboard).

The stations were plotted according to their corresponding latitudes and longitudes, there were 161 unique stations in the dataset.



Next, general year to year trends for the average PRCP, SNOW, SNWD, and TOBS (since TMIN and TMAX followed similar trends) were plotted.



Dates of data span:

PRCP: 1893 to 2012

SNOW: 1901 to 2012

SNWD: 1930 to 2012

TOBS: 1901 to 2012

Interesting observation by measure:

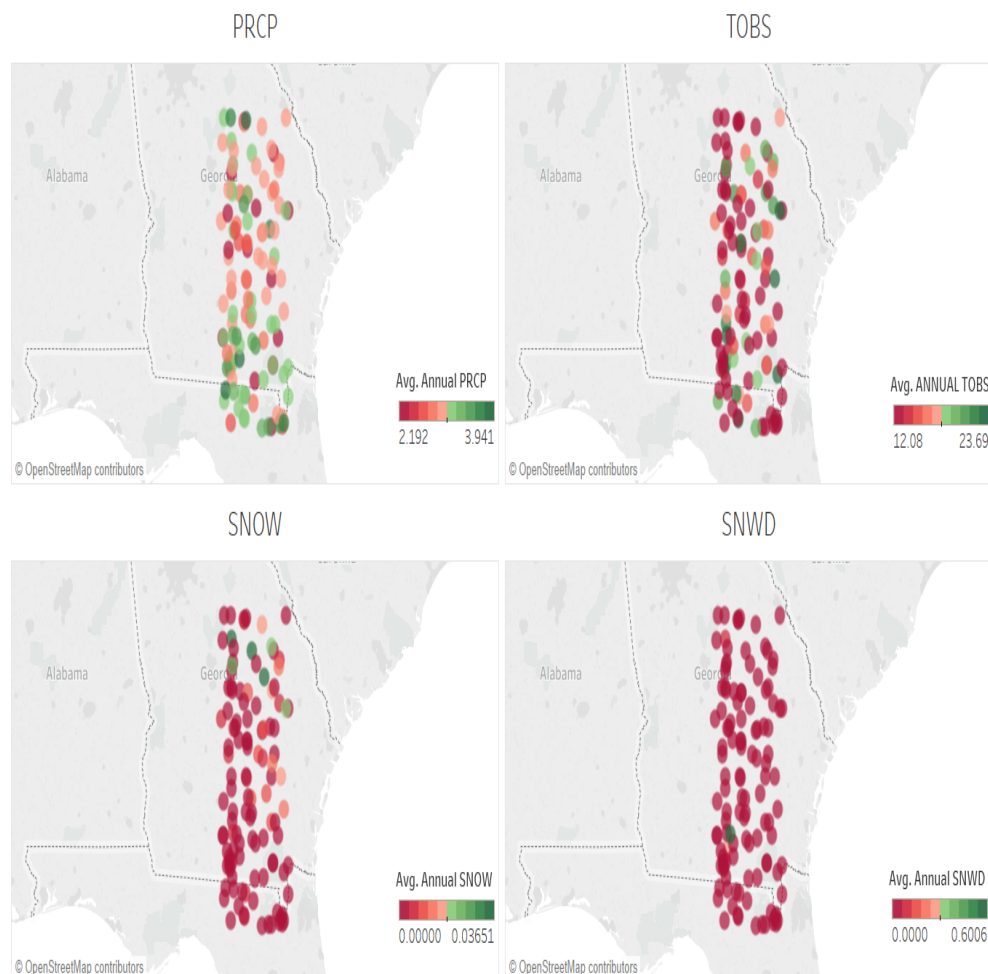
PRCP: The average PRCP over the years is 3.113 mm. The maximum average PRCP occurred in 1964 with 4.336 mm. and the minimum average PRCP occurred in 1954 with 1.781 mm.

SNOW: The average SNOW over the years is 0.0054 mm. The maximum average SNOW occurred in 1973 with 0.09619 mm. and the minimum average SNOW (0 mm.) occurred for the majority of the years.

SNWD: The average SNWD over the years is 0.0058 mm. The maximum average SNWD occurred in 1973 with 0.1624 mm. and the minimum average SNWD (0 mm.) occurred for the majority of the years. SNWD, as expected, follows a similar year-to-year trend to SNOW.

TOBS: The average TOBS over the years is 18.645 (in Celsius). The maximum TOBS occurred in 1922 with a temperature of 21.441 deg. C and the minimum TOBS occurred in 1912 with a temperature of 14.711 deg. C.

Finally, Tableau was used to plot the average PRCP, TOBS, SNOW, and SNWD for each of the station locations.



Interesting observations by measure:

PRCP: Visual inspection shows that the majority of northern weather stations are below 3.0665 mm.

TOBS: -

SNOW: The northern most weather stations had the higher average annual SNOW.

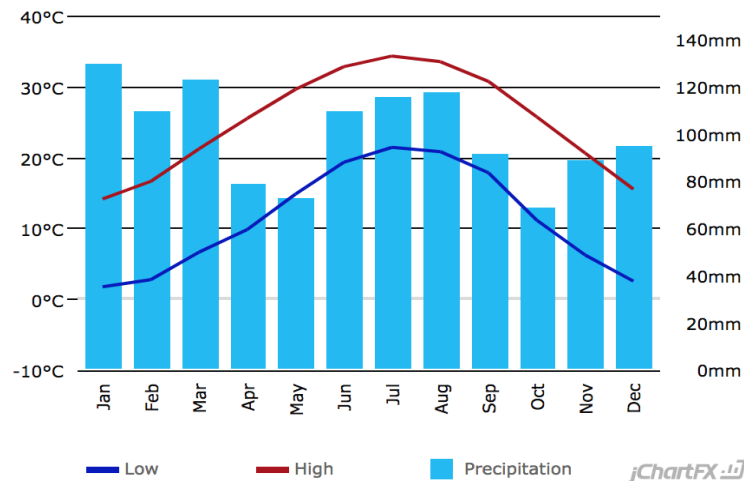
SNWD: The majority of the weather stations had average annual SNWD below 0.3003 mm., only station US1GALN0001 had an average annual SNWD of 0.6006 mm.

1.2 Sanity-check: comparison with a representative weather station

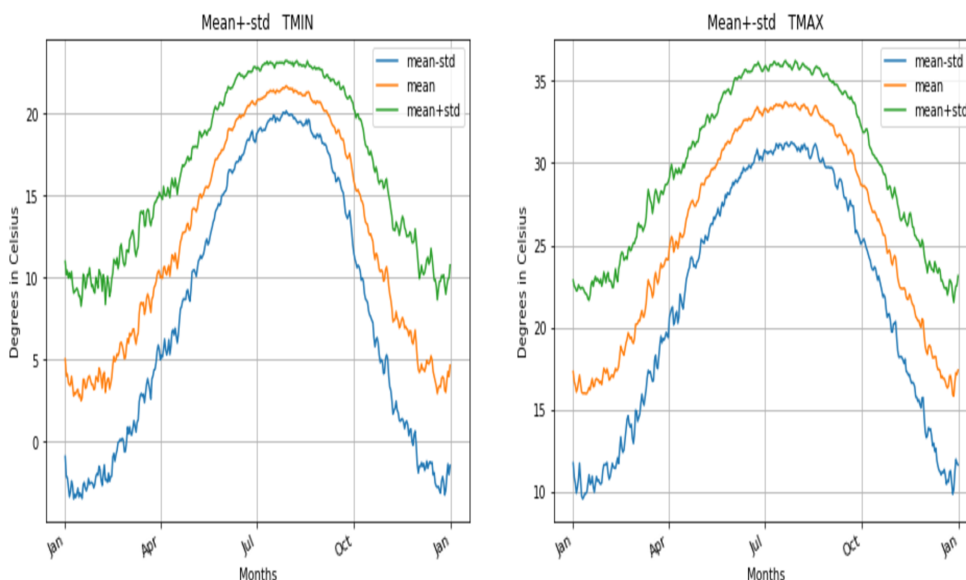
Using the average values for TOBS, PRCP, SNOW, and SNWD found in Tableau a representative weather station is located, USC00092839. The weather station is located in Dublin, GA (32.557, -82.904) and has an average annual PRCP of 3.0678 mm., average annual TOBS of 16.077 deg. Celsius, average annual SNOW of 0.01746 mm., and average annual SNWD of 0 mm.

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/dublin/georgia/united-states/usga0182) (<http://www.usclimatedata.com/climate/dublin/georgia/united-states/usga0182>) The graph below shows the daily minimum and maximum temperatures for each month, as well as the total precipitation for each month. Note precipitation highs occur in June and August.

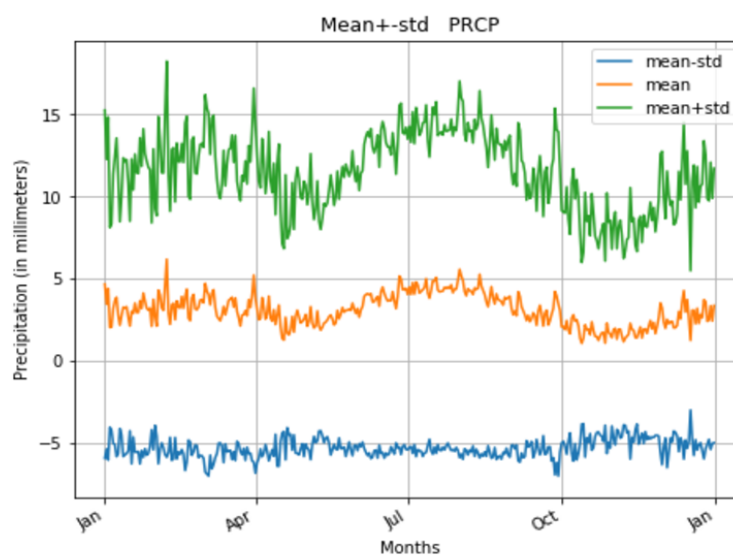
Dublin Climate Graph - Georgia Climate Chart



We see that the min and max daily temperature agree with the ones we got from our data, once we translate Fahrenheit to Centigrade.



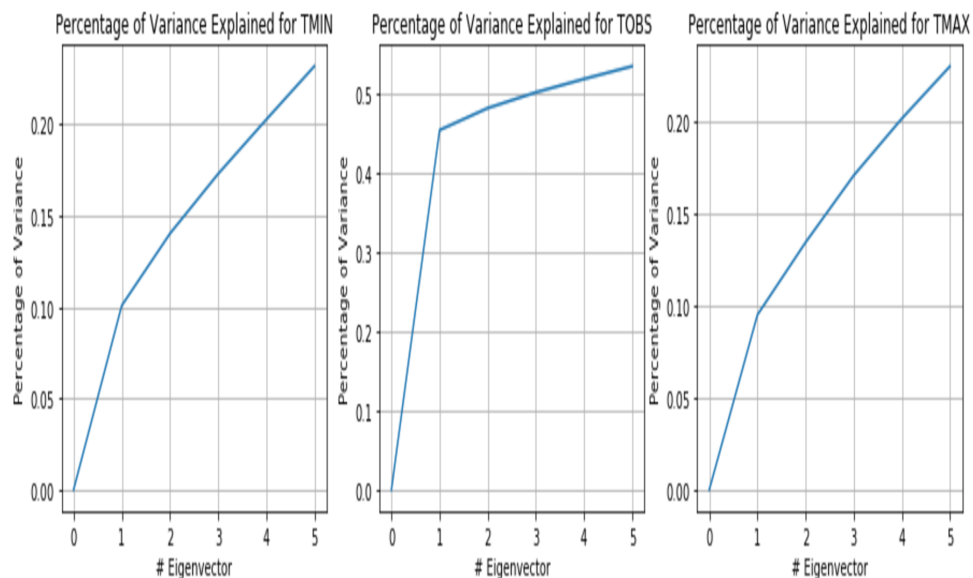
According to our analysis the average annual precipitation for our region of interest is 3.09 mm./day, which is notably smaller than the average annual precipitation in Dublin, GA, which is 3.29 mm./day.



1.3 PCA analysis

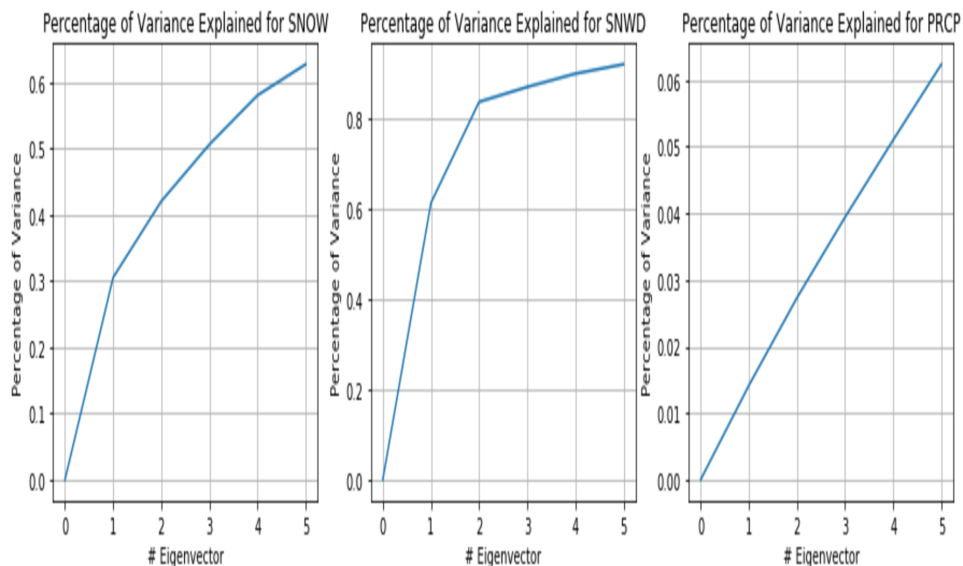
For each of the six measurement, we compute the percentage of the variance explained as a function of the number of eigen-vectors used.

1.3.1 Percentage of variance explained.



We see that the top 5 eigen-vectors explain 24% of variance for TMIN, 52% for TOBS and 23% 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 6% of the variance for PRCP, which is low due to the noisy variance present for precipitation. The top eigenvectors explain 61% of the variance for SNOW. On the other hand the top 5 eigenvectors explain 92% 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.

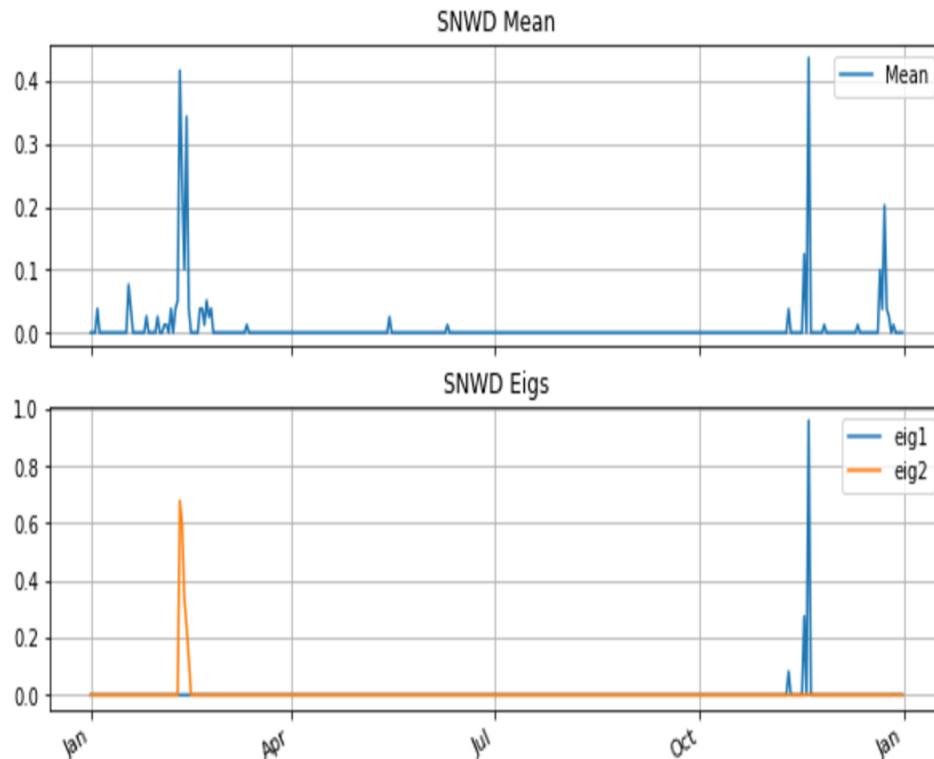
It makes sense that SNWD would be less noisy than SNOW since SNWD is a decaying integral of SNOW.

1.4 Analysis of snow depth (SNWD)

We choose to analyze the eigen-decomposition for snow-depth because the first 2 eigen-vectors explain 82% of the variance.

First, we graph the mean and the top 2 eigen-vectors.

We observe that the snow season in the region of interest experiences 2 different snow seasons described by the top 2 eigen-vectors. The first eigen-vector describes the snow depth accumulation between November and December and the second eigen-vector describes the snow depth accumulation between February to March.

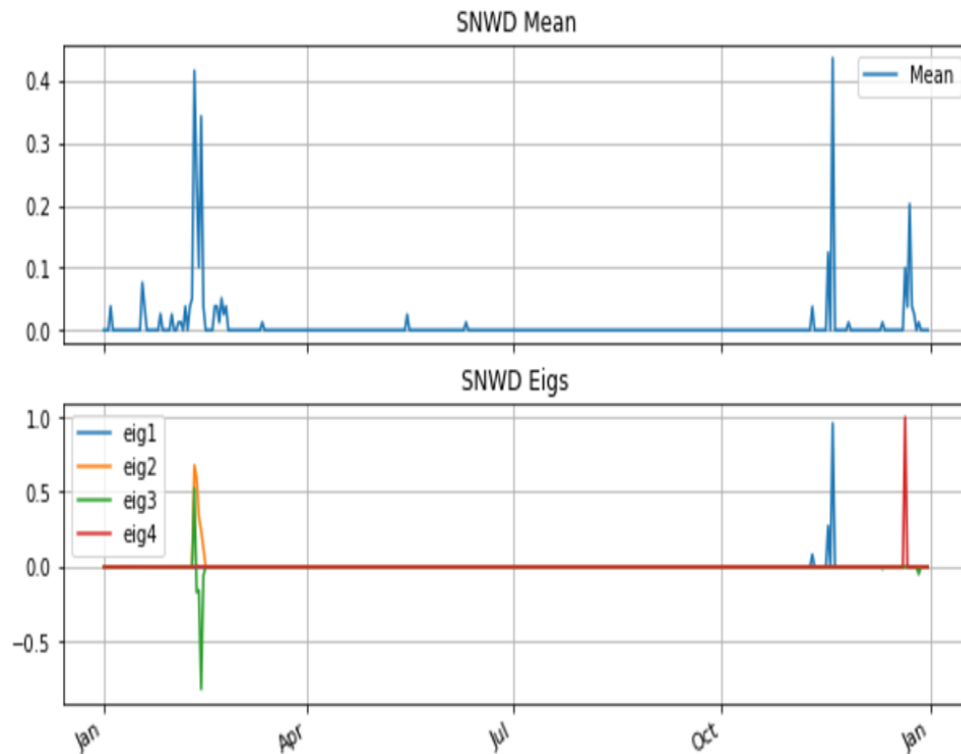


Next we interpret the eigen-functions. Both eigen-functions are similar to the mean function, the first eigen-function (eig1) has the same shape as the mean snow depth increase between November and December, while the second eigen-function (eig2) has a similar shape to the mean snow depth increase between February and March but capture the mean increase as 1 main peak as the sum increase of the 2 mean peaks between February and March.

Examining the data more closely, only 2.33% or 48/2057 of the total instances have records for SNWD (at least 1 day out of the year for that station has a SNWD value > 0). Considering, that the presence of SNWD in the area of interest, it makes sense that 2 eigen-vectors are enough to explain the majority of the variance.

In order to capture additional peaks, we will focus on analyzing the eigen-decomposition for snow-depth based on the first 4 eigen-vectors as to capture additional variance (up to 91% variance). We observe that the top 4 eigen-vectors capture additional information. The combination of eig2 and eig3

describe more thoroughly the snow depth between February and March while the combination of eig1 and eig4 describe more thoroughly the snow depth between November and January. The first 4 eigen-vectors will be used in further analysis of SNWD.

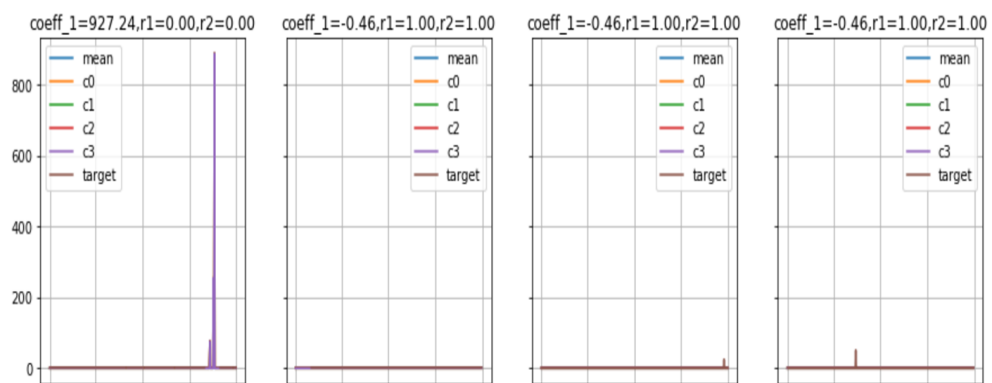


Next we interpret the eigen-functions. The first eigen-function (eig1) describe the increased snow depth between November and December, the second eigen-function (eig2) describe the increased snow depth between February and March, the third eigen-function (eig3) describe the negative of the mean peak between February and March, and the fourth eigen-function (eig4) describe the mean SNWD peak between December and January.

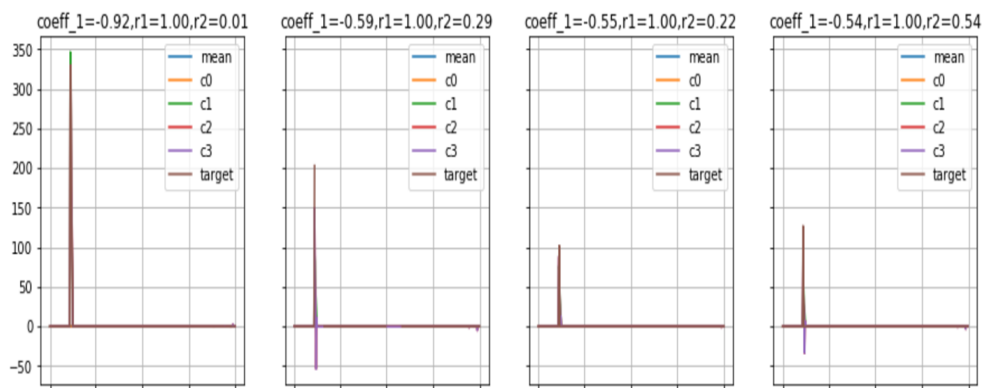
1.4.1 Examples of reconstructions

1.4.1.1 Coeff1

Coeff1: most positive



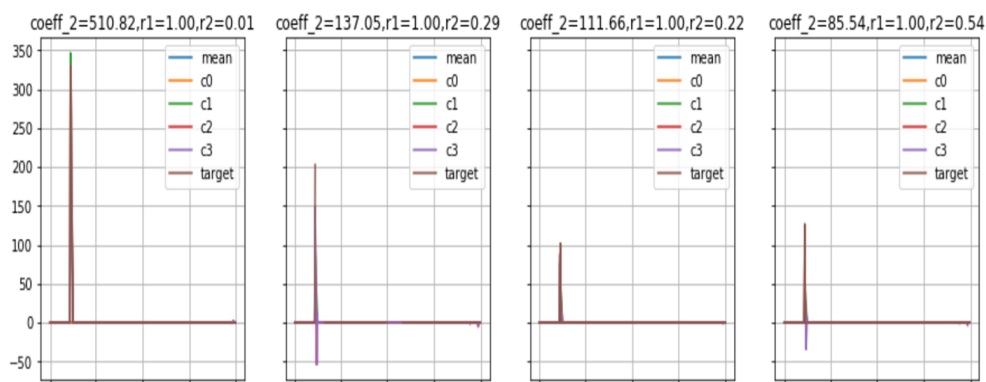
Coeff1: most negative



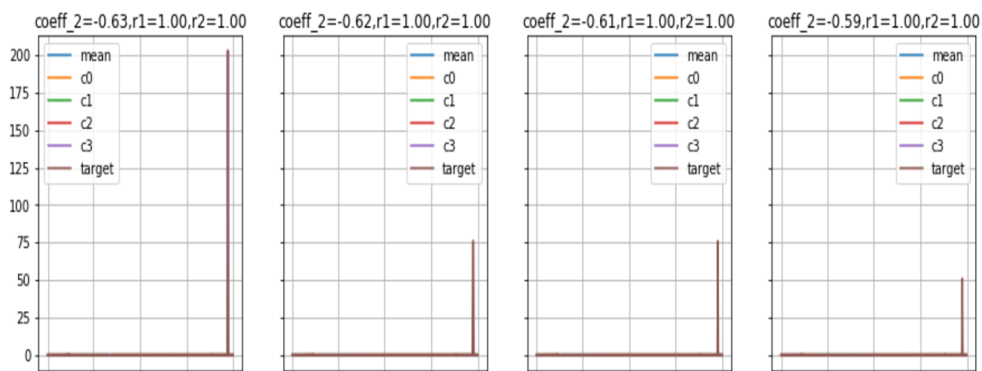
Large positive values of coeff1 correspond to more than average snow and low values correspond to less than average snow depth between November and December.

1.4.1.2 Coeff2

Coeff2: most positive



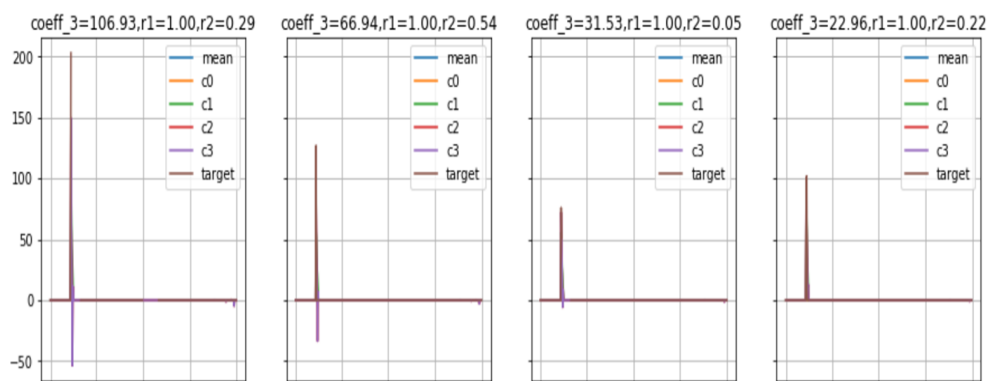
Coeff2: most negative



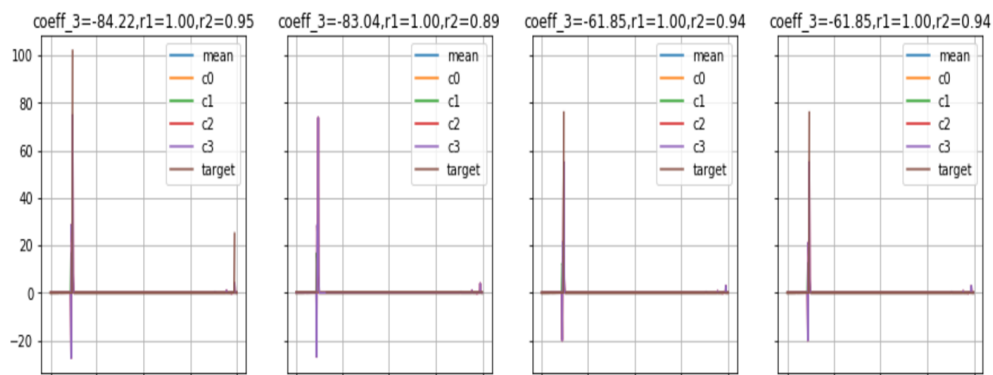
Large positive values of coeff1 correspond to more than average snow and low values correspond to less than average snow depth between February and March.

1.4.1.3 Coeff3

Coeff3: most positive



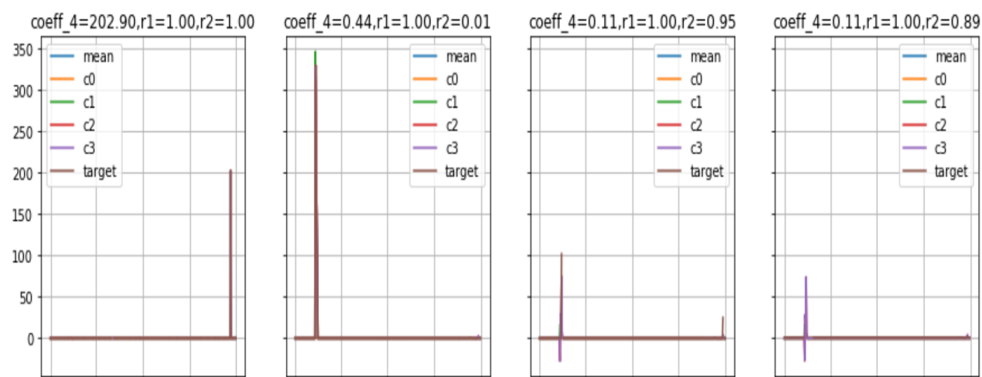
Coeff3: most negative



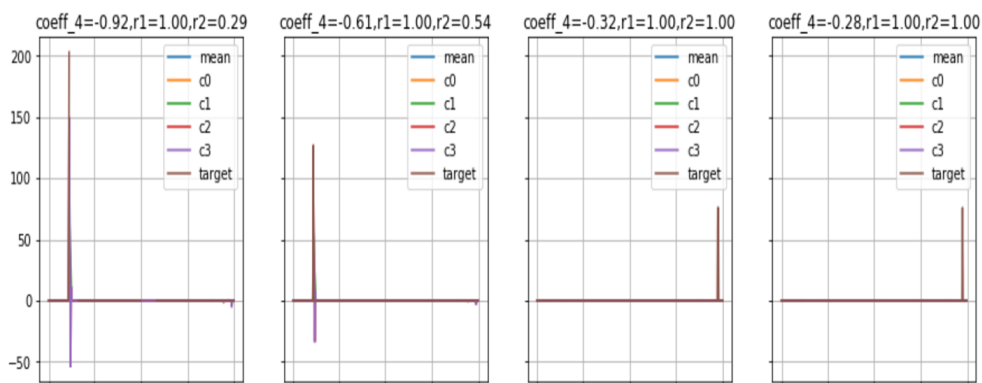
Large positive values of coeff1 correspond to more than average snow and low values corresponds to the negative mean peak between February and March.

1.4.1.4 Coeff4

Coeff4: most positive

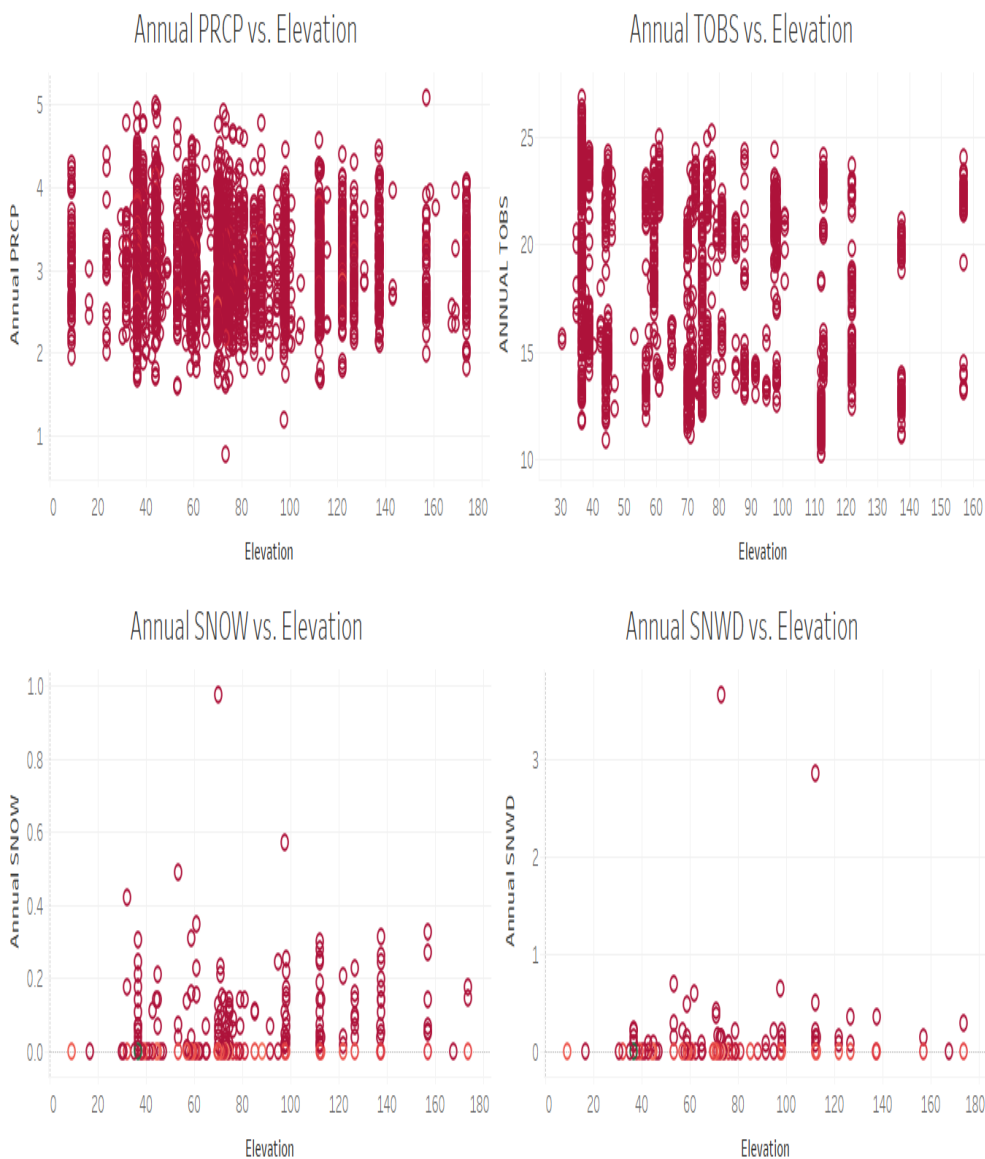


Coeff4: most negative



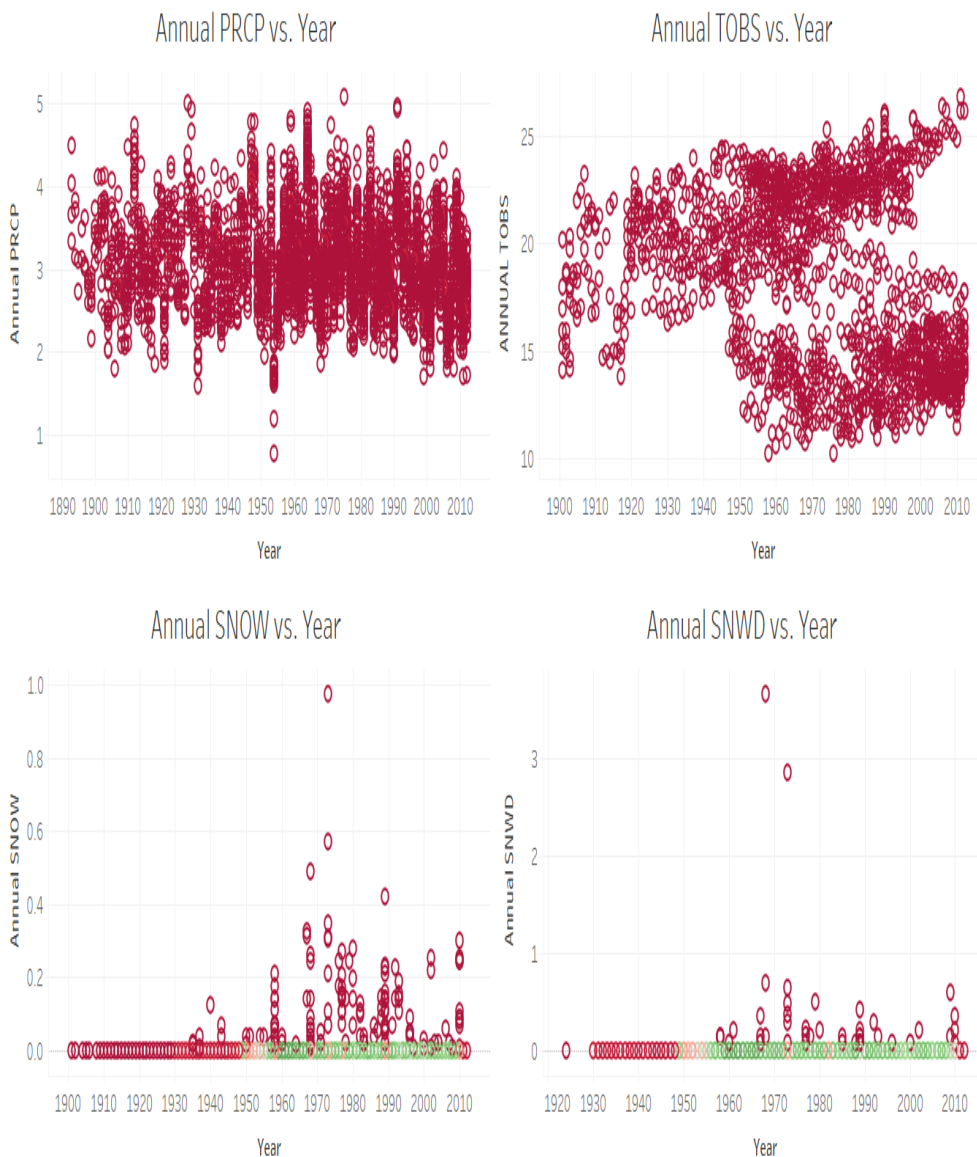
Large positive values of coeff_1 correspond to more than average snow and low values corresponds to less than average snow depth between December and January.

The assumption that instances of SNWD is correlated with elevation was tested, along with PRCP, SNOW, and TOBS. The number of instances per year for each metrics were plotted and the points were colored according to number of instance (red to green corresponds to low to high instances, not as obvious here).



As seen, there are no obvious correlations seen in the graphs, in part this could be due to not much a difference in elevation across the weather stations.

Furthermore, scatter plots were also generated for PRCP, TOBS, SNOW, and SNWD vs. Year to view if any relation was seen. The number of instances per year for each metrics were plotted and the points were colored according to number of instance (red to green corresponds to low to high instances).



As seen above, there are no obvious correlations seen in the graphs.

1.4.2 Temporal Variation Contributes to Snow Depth

We now estimate the relative importance of location-to-location variation relative to year-by-year variation, by interpreting the relative RMSs. Here are the results:

coeff_1

```
total RMS                = 139.787060801
RMS removing mean-by-station= 98.8948101876
RMS removing mean-by-year  = 114.19180867
```

coeff_2

```
total RMS = 83.7824184311
RMS removing mean-by-station= 70.4438076129
RMS removing mean-by-year = 63.0150971492
```

coeff_3

```
total RMS = 32.2335045915
RMS removing mean-by-station= 14.5318583353
RMS removing mean-by-year = 24.961756426
```

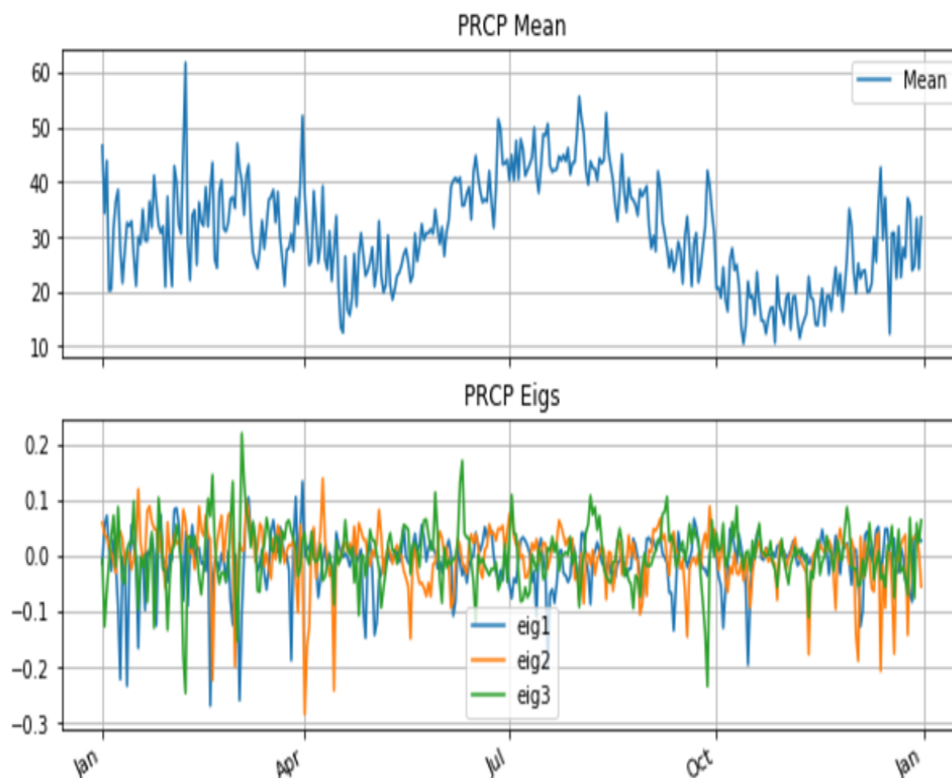
coeff_4

```
total RMS = 30.588638776
RMS removing mean-by-station= 0.0894169101026
RMS removing mean-by-year = 21.6407752978
```

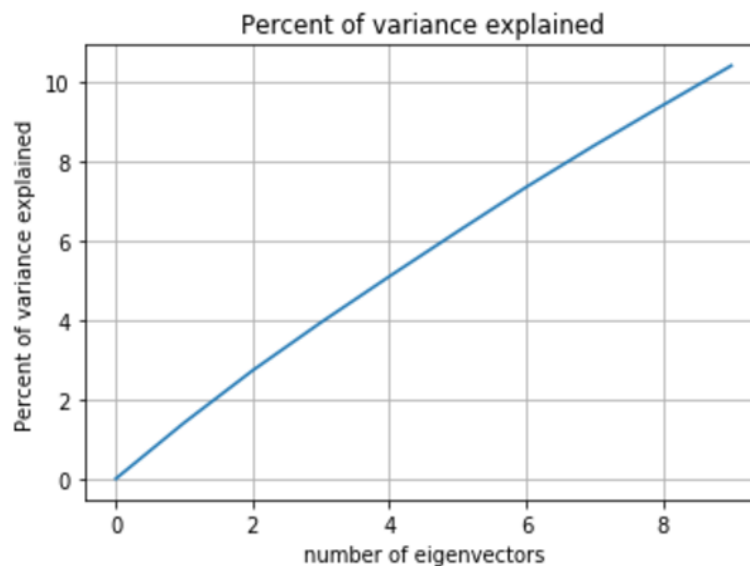
For coeff_1, the variation is temporal for snow depth since the RMS after removing mean-by-year is greater than the RMS after removing mean-by-station. For coeff_2, the variation is spatial for snow depth since the RMS after removing mean-by-station is greater than the RMS after removing mean-by-year. For coeff_3, the variation is temporal for snow depth since the RMS after removing mean-by-year is greater than the RMS after removing mean-by-station. For coeff_4, the variation is temporal for snow depth since the RMS after removing mean-by-year is greater than the RMS after removing mean-by-station. Therefore, temporal variation primarily contributes to snow depth.

1.5 Analysis of the precipitation (PRCP)

Next, we choose to analyze the eigen-decomposition for the average precipitation with the top 3 eigen-vectors considered.



The precipitation data is very noisy and this noisiness can be explained due to the % variance explained vs. number of eigenvectors. Therefore, further analysis for precipitation is void and will not be benefitted by eigen-decomposition.

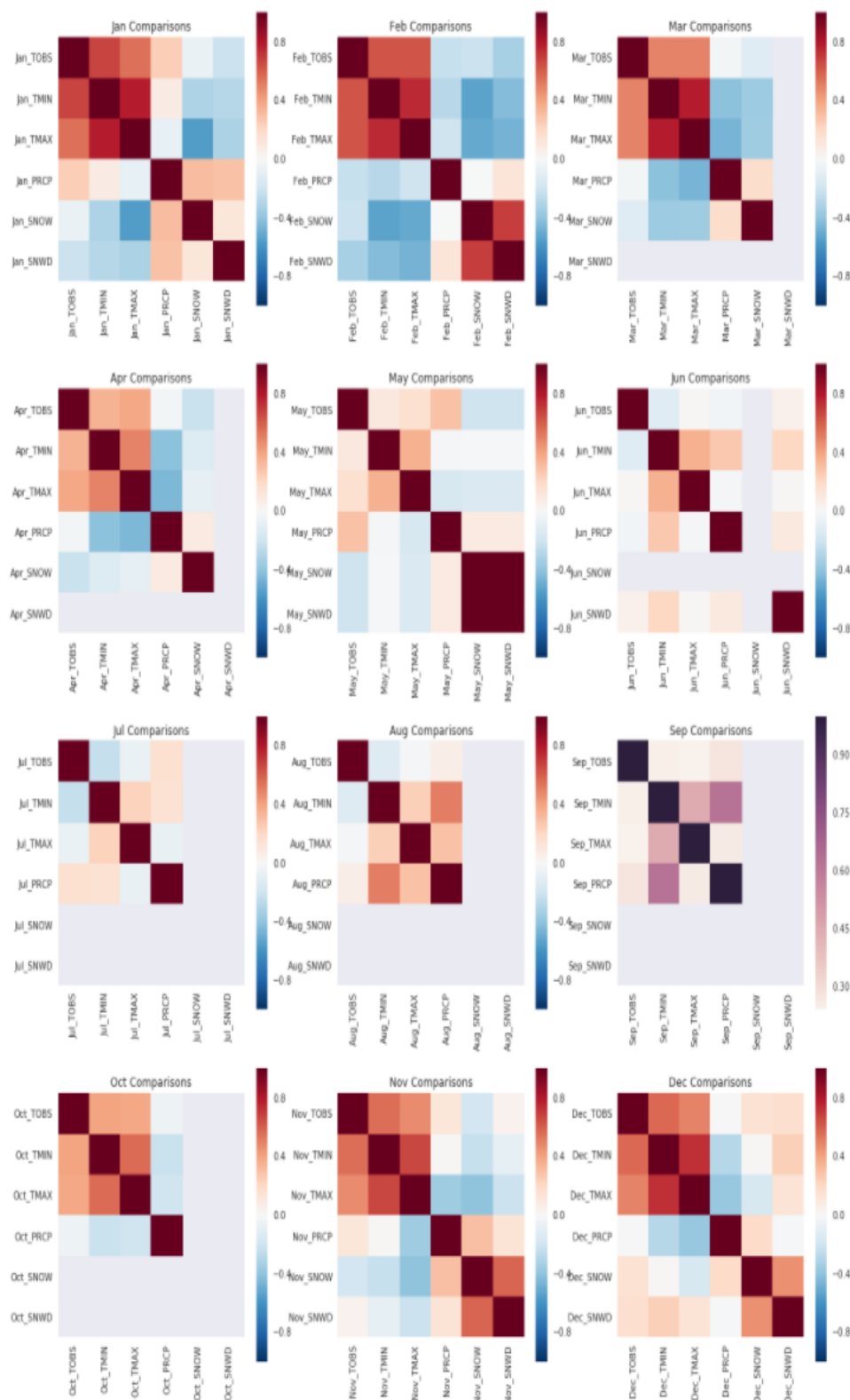


1.6 Examining the Correlation and Covariance between Weather Measurements

Vectors for each metric were extracted from the parquet file, the associated measurement array was incremented by month to determine the average month measurement for TOBS, TMIN, TMAX, SNWD, SNOW, and PRCP

grouped by station, and stations which were present for each of the measurements were considered (48 Stations in total). Correlation and covariance heatmaps were generated to compare the weather specific measurements by month to one another to determine the relationship between TOBS, TMIN, TMAX, SNWD, SNOW, and PRCP by month.

Month to Month Weather Metric Analysis Grouped By Weather Station - Correlation



Interesting observations:

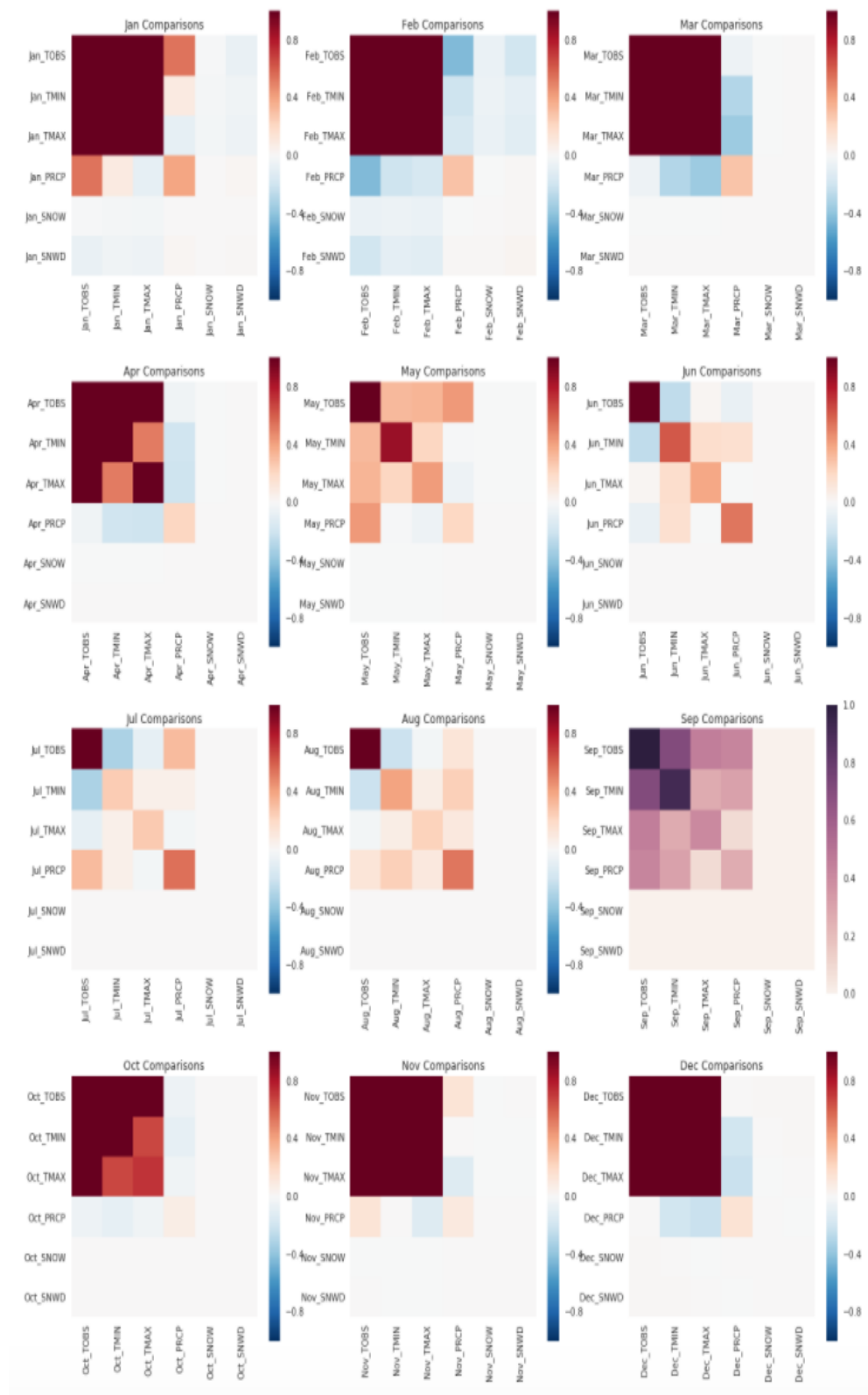
January: SNOW and TMAX are is negatively correlated (-0.7)

February: SNOW and TMAX are is negatively correlated (-0.7). SNOW and SNWD are positively correlated (+0.8), also reflected by the eigenvectors associated with SNOW and SNWD

September: PRCP and TMIN are positively correlated (+0.75)

November: SNOW and SNWD are positively correlated (+0.8), also reflected by the eigenvectors associated with SNOW and SNWD

Month to Month Weather Metric Analysis Grouped By Weather Station - Covariance



Interesting observations:

- January: PRCP and TOBS vary positively (+0.5)
- February: PRCP and TOBS vary negatively (-0.5)
- May: PRCP and TOBS vary positively (+0.6)
- July: PRCP and TOBS vary positively (+0.5)
- September: PRCP and TOBS vary positively (+0.7)

2 Summary

The following analysis was completed:

1. Weather metric trend analysis by year: Showed the average, minimums, and maximums metric values by year. This can explain years of drought, heat waves, and strange data (ie. irregular SNOW or SNWD)
2. Weather average metrics by station location: Visually allowed for some weather location ideas
3. Comparing TMIN, TMAX, and PRCP to a weather station representative
4. Percentage of variance explained vs. number of eigenvectors
5. SNWD and PRCP eigen-decompositions and reconstructions
6. Weather metric vs. elevation and vs. year analysis
7. Spatial vs. temporal analysis for SNWD
8. Correlation and covariance heatmap generation for weather metrics by months of year

3 Supplementary Links

https://public.tableau.com/profile/orysya.stus#!/vizhome/Weather_Analysis/Station_Representative
[\(https://public.tableau.com/profile/orysya.stus#!/vizhome/Weather_Analysis/Station_Representative\)](https://public.tableau.com/profile/orysya.stus#!/vizhome/Weather_Analysis/Station_Representative)

```
In [1]: %%HTML
<div class='tableauPlaceholder' id='viz1494916944366' style=

Station_locations
PRCP_TOBS_SNOW_SNWD_by_WeatherLocation
All_by_year
Metrics_vs_Elevation
Metrics_vs_Year
```

```
In [ ]:
```

