North Utah-South Idaho Weather Analysis- Area SBSSBSBB-:

0-Introduction

This is a report on the historical analysis of weather patterns in an area that approximately overlaps the area of North Utah-South Idaho .

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

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-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) The graph below shows the daily minimum and maximum temperatures for each month, as well as the total precipitation for each month.
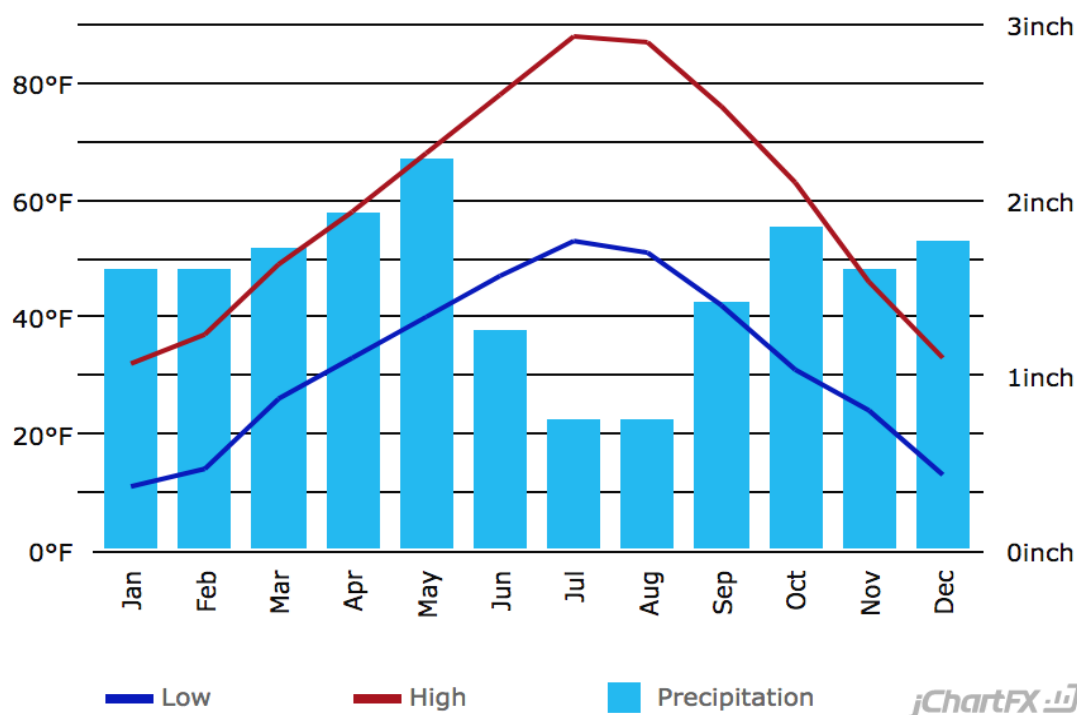
We see that the min and max daily temperature agree with the ones we got from our data, once we translate Fahrenheit to Centigrade. SAme pattern for temperature
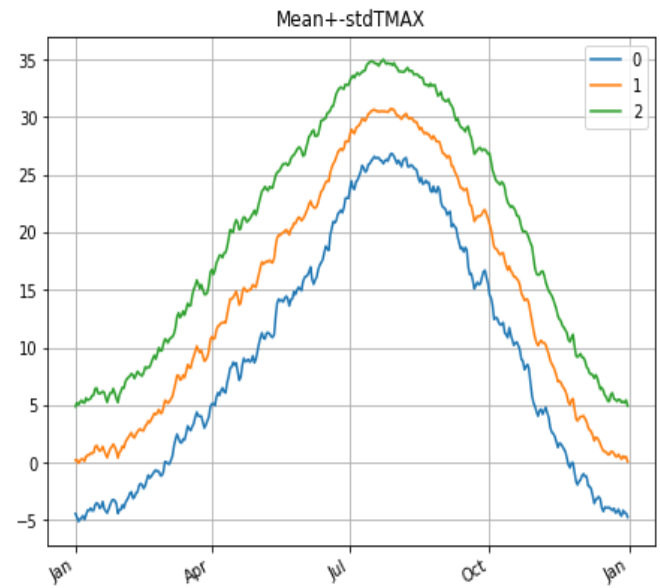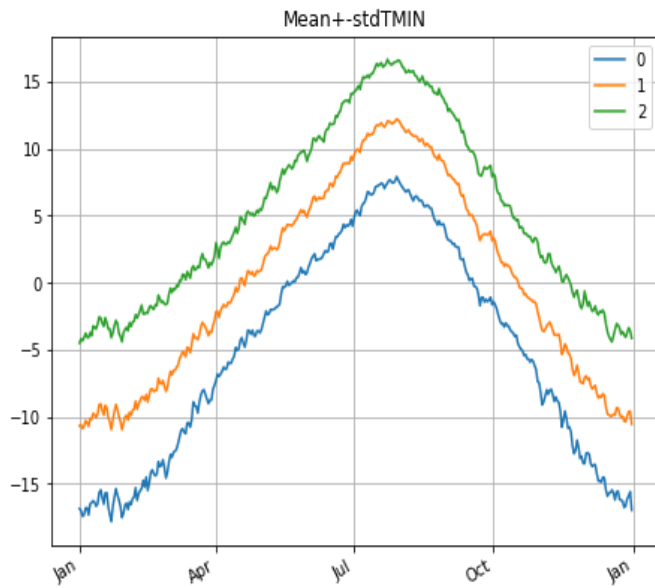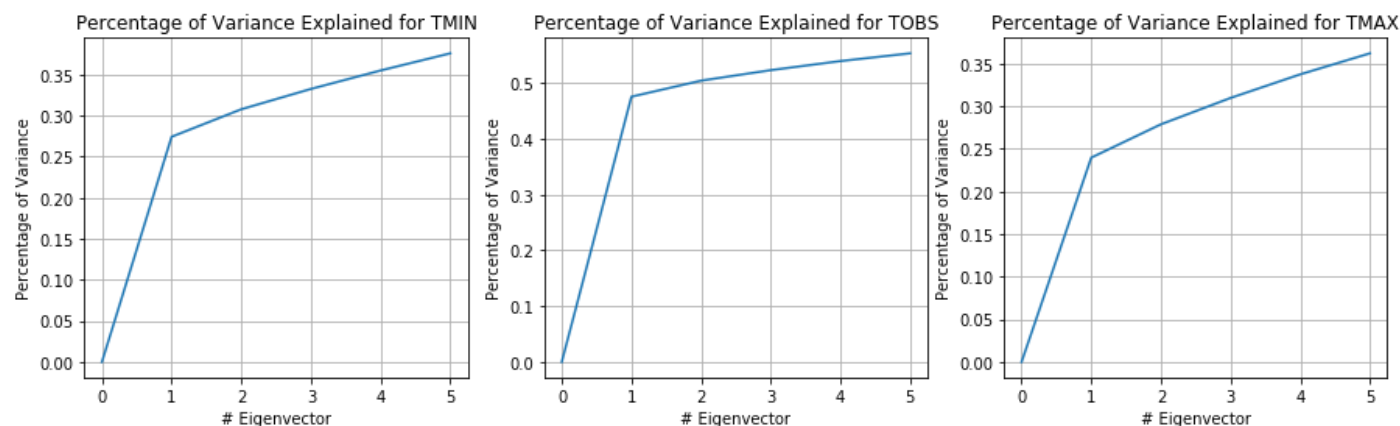
0 = "*Mean +std*" ,1 = "*Mean*", 2 = "*Mean -std*"

2-PCA analysis

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

Percentage of variance explained.

We see that the top 5 eigen-vectors explain 37% of variance for TMIN, **55%** for TOBS and 36% 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 47 % of the variance.
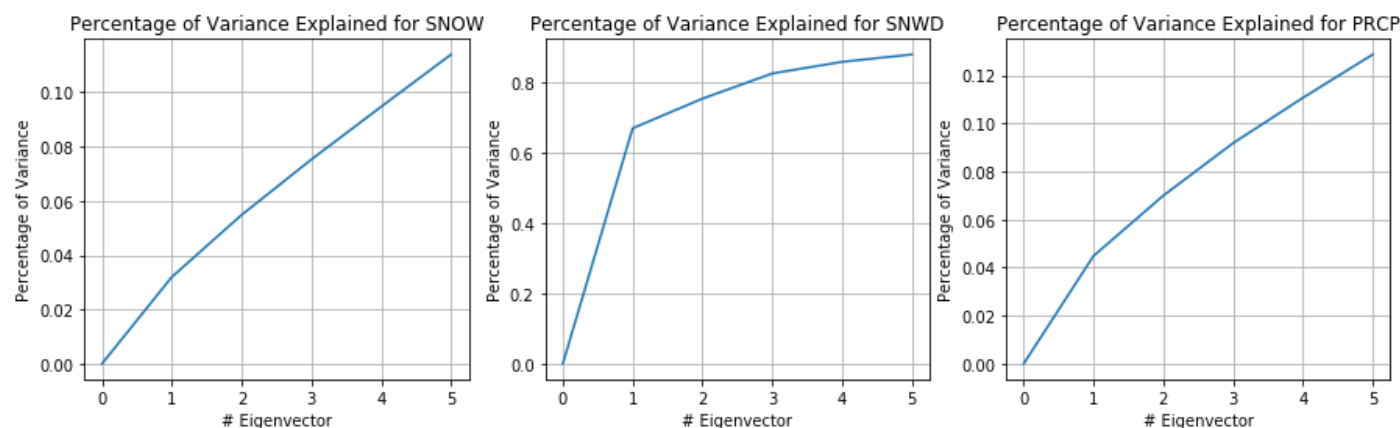
The top 5 eigenvectors explain 13% of the variance for PRCP and 11% for SNOW. Both are low values. On the other hand the top 5 eigenvectors explain **88%** of the variance for SNWD.

This means that these top 5 eigenvectors capture most of the variation in the snow signals and temperature observed. Based on that we will dig deeper into the PCA analysis for snow-depth and temperature observed .

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 diffferent years.

3- Analysis of snow depth

We choose to analyze the eigen-decomposition for snow-depth because the first 5 eigen-vectors explain 88% of the variance.

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

We observe that the snow season is from mid-november to the end of march, where the middle of February marks the peak of the snow-depth.
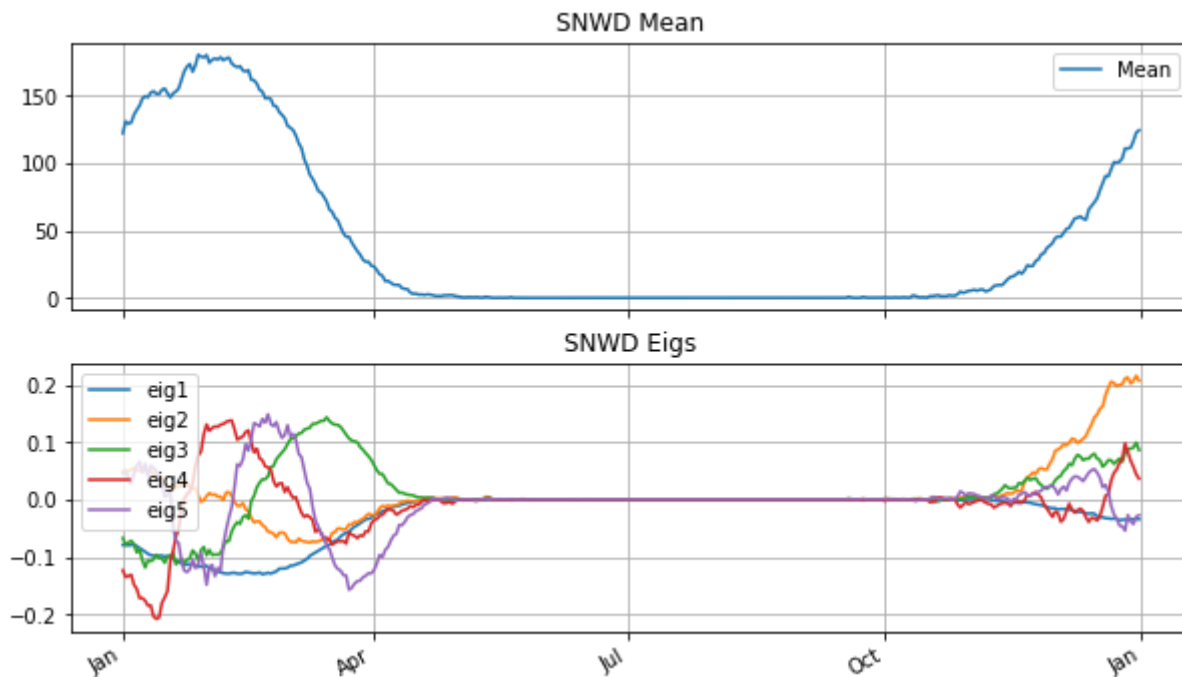
Next we interpret the eigen-functions. The first eigen-function (eig1) has a shape very similar to the mean function but with -ve sign ,main contributing stations/factors to varience have snow less than the mean

eig2,eig3 and eig4 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:** More snow November and December and March less snow from jan to April with dip in March
- **eig3:** Low snow in Jan and Feb ,high snow in March April November and December .
- **eig4:** Low snow in in Jan ,April & novemeber and high snow ib March and December
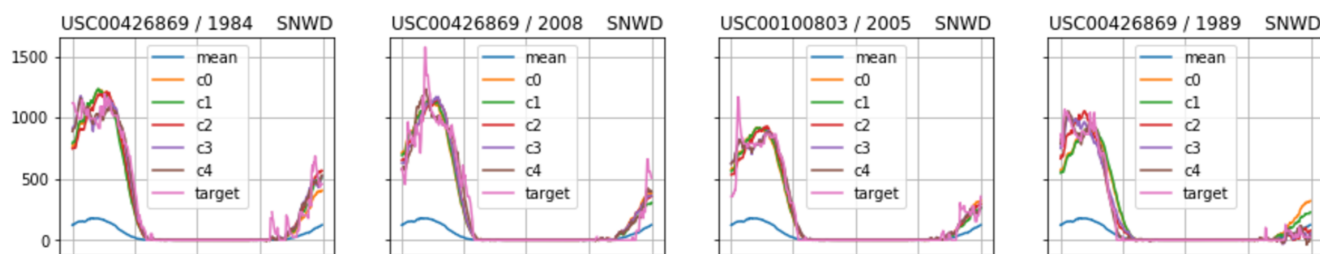- **eig5:** high snow in Jan , March and November & Low snaoe in Feb ,April and December.

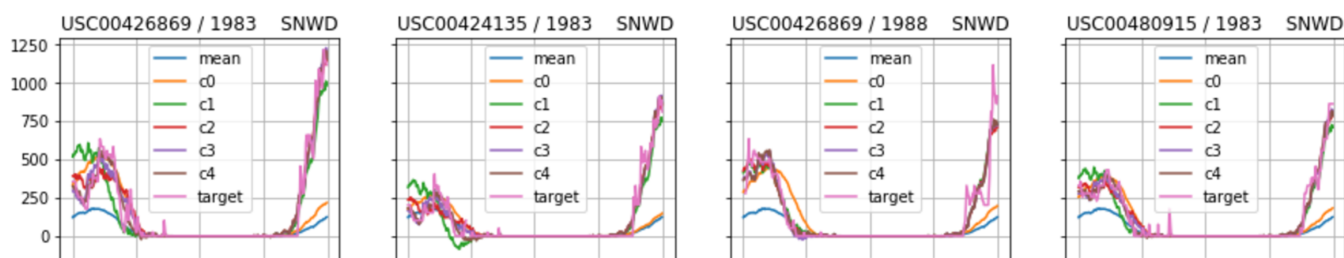# 4 - Examples of reconstructions

Coeff1
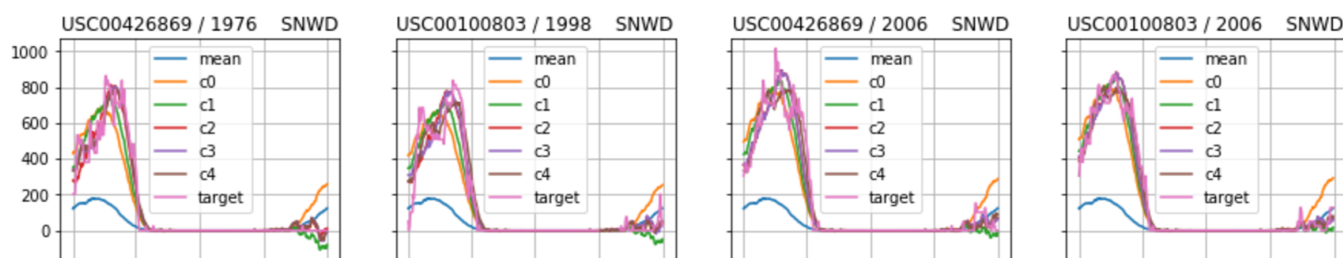
Coeff1: most positive
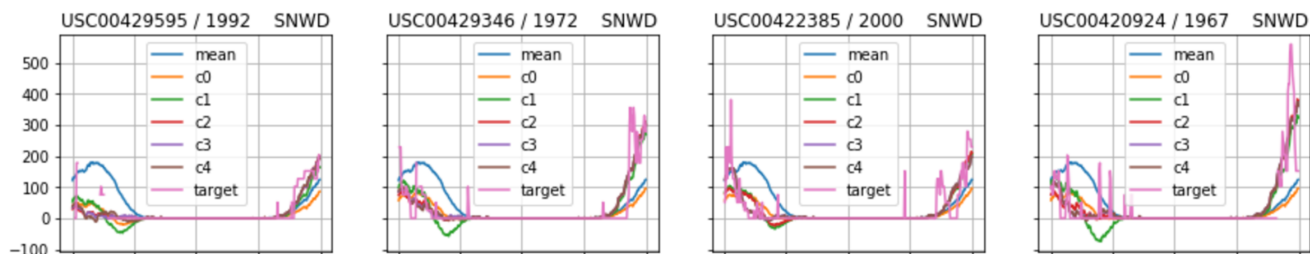
Coeff1: most negative

Large positive values of coeff1 correspond to more than average snow. Low values correspond to less than average snow.
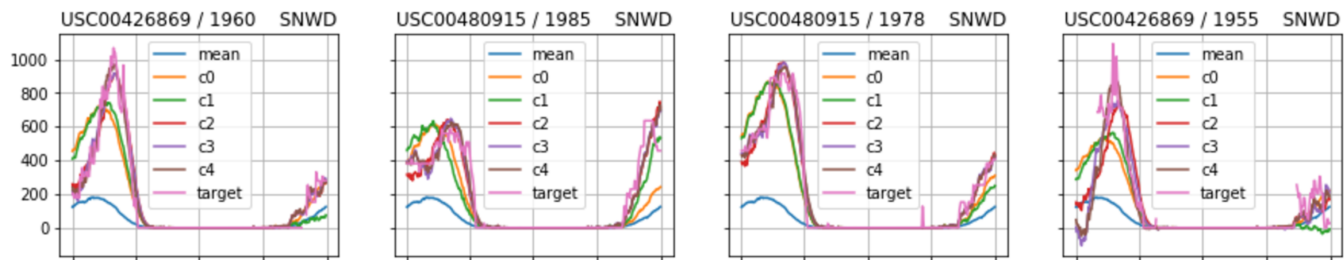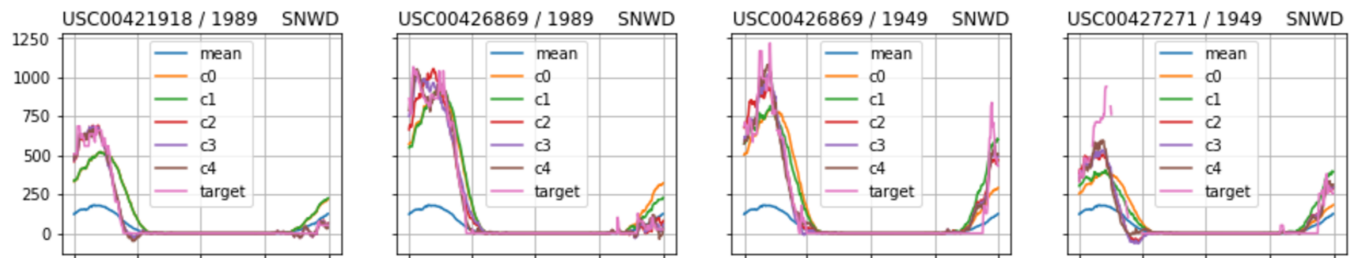
Coeff2

Coeff2: most positive

Coeff2: most negative

Large positive values of coeff2 correspond to a late snow season (most of the snowfall is after mid feb. Negative values for coeff2 correspond to an early snow season (most of the snow is before mid-feb.

Coeff3

Coeff3: most positive

Coeff3: most negative

Large positive values of coeff3 correspond to a snow season with two spikes: one in the start of january, the other at the end of february. Negative values of coeff3 correspond to a season with a single peak at the end of Jan.

5-Estimating the effect of the year vs the effect of the station

To estimate the effect of time vs. location on the first eigenvector coefficient we compute:

- The average row: `mean-by-station`
- The average column: `mean-by-year`

We then compute the RMS before and after subtracting either the row or the column vector.

- Total RMS = 601.90862572
- RMS removing mean-by-station= 575.299716846
- RMS removing mean-by-year = 346.048423438

The effect of year to year variation is more than station by station variation

6- Block Diagonal

This type of organization is called **Block Diagonal** and it typically reveals important structure such as grouping or clustering. Eigen vector 1 SNWD:

- Station 15 to 40 are highly correlated Eigen vector 2 SNWD:
- Station 0 to 28 are highly correlated Eigen vector 3 SNWD:
- Station 15 to 20 are highly correlated Eigen vector 4 SNWD:
- Station 15 to 20 are highly correlated

we can see that stationsfrom 15 to 20 are highly correlated

7- Correlation of Station per eigen vectors**
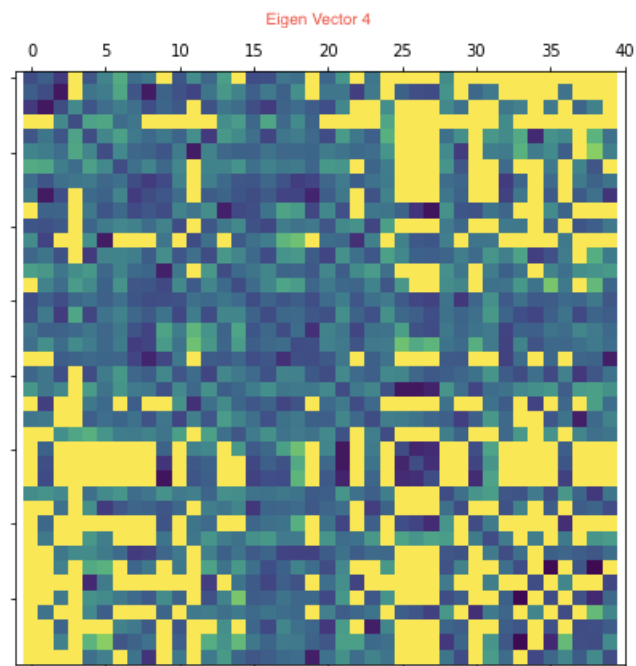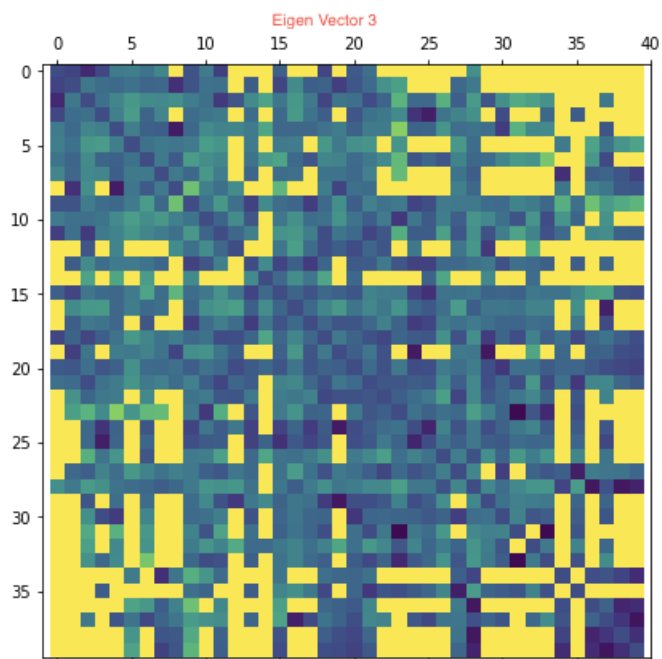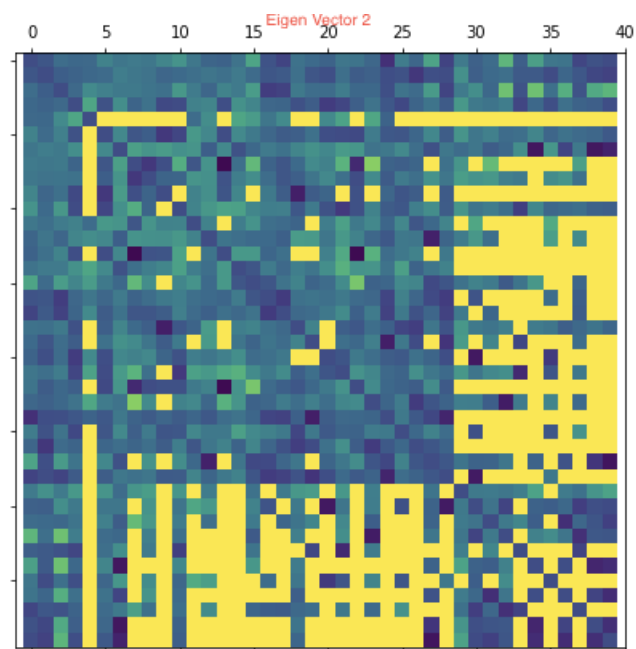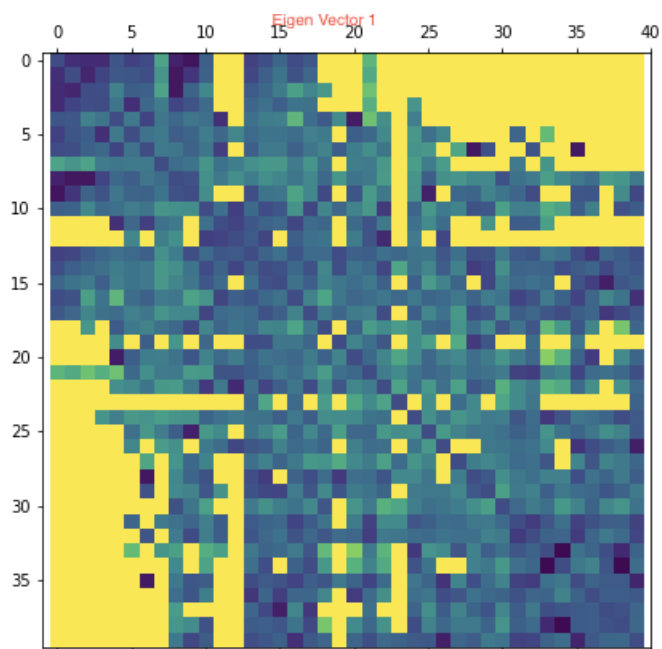
- List of Highly corellated station (15 to 20):

Station	latitude	longitude	elevation
USC00422726	41.0222	-111.9353	1335
USC00423671	41.6	-111.5667	1694.7
USC00424135	41.2403	-111.7131	1566.7
USC00424856	41.825	-111.3208	1822.7
USC00425082	41.9667	-111.8333	1366.1
USC00425182	41.7353	-111.8561	1362.5
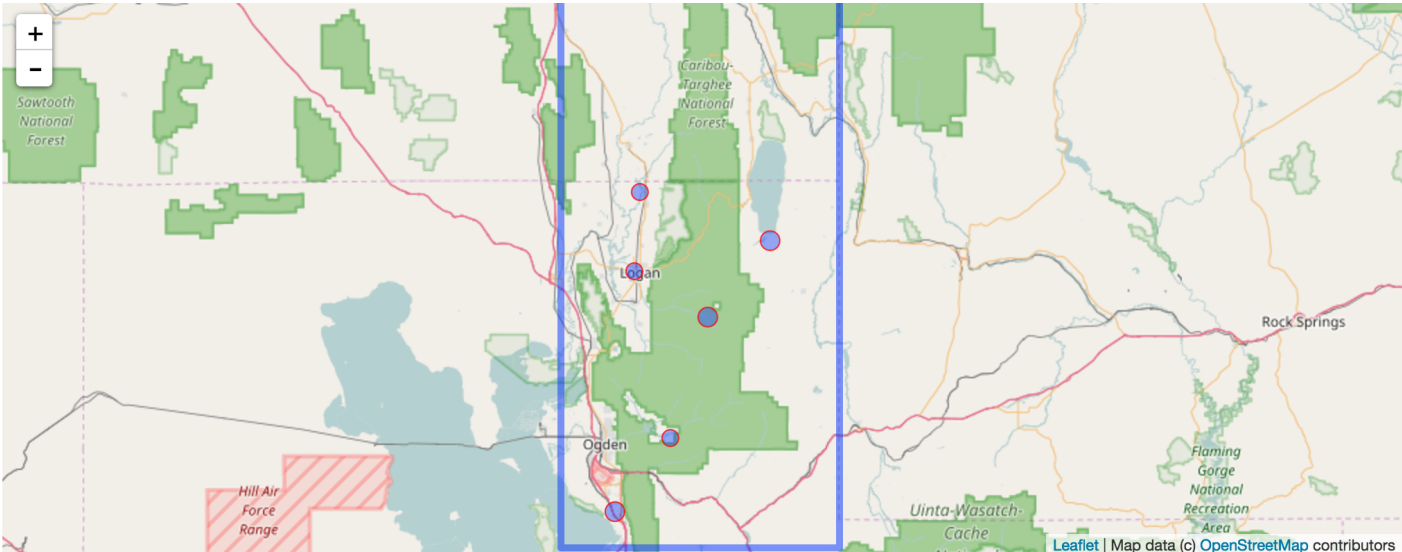
Those stations are spread on the area ,they arenot located on same elevation or location

In [ ]: