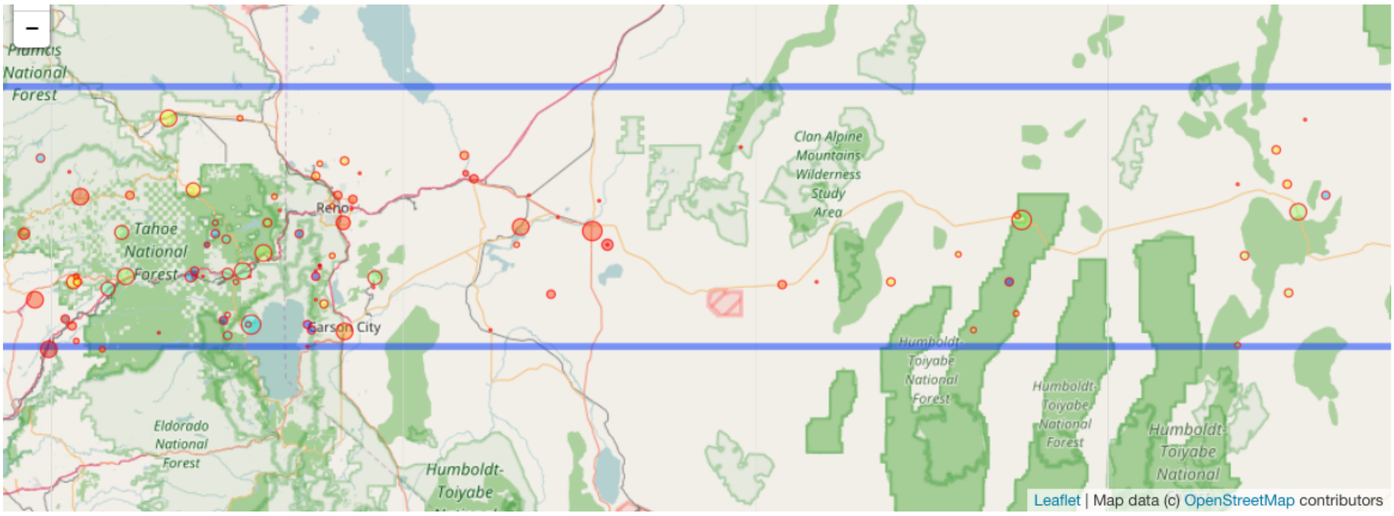
Weather Analysis for area for file indexed SSSBSBBB

Visually area of interest



This is a report on the historical analysis of weather patterns in an area above.

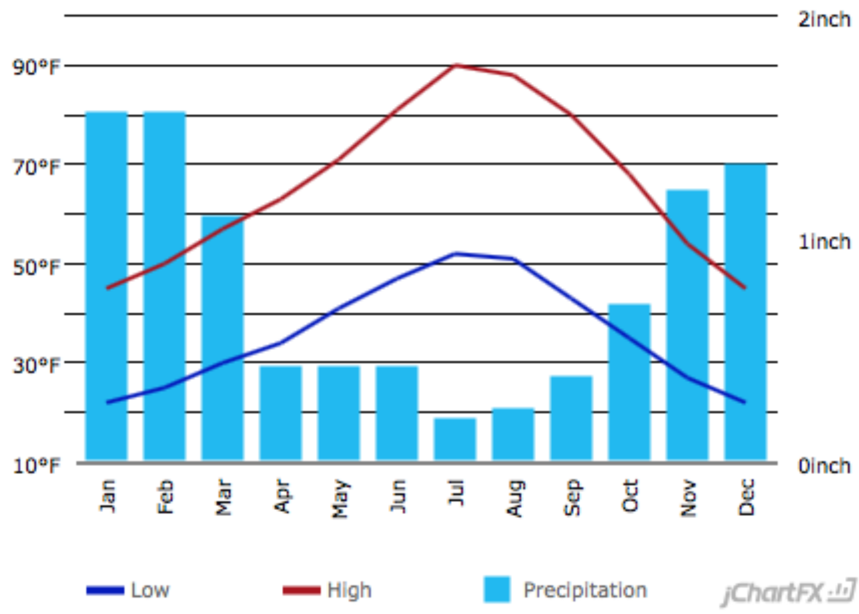
The data we will use here comes from NOAA (https://www.ncdc.noaa.gov/). Specifically, it was downloaded from This FTP site.

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.

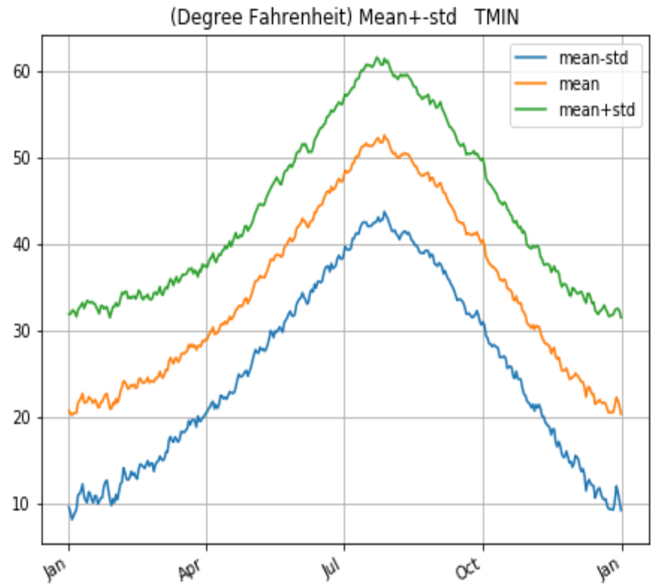
Highlights about our area of interest

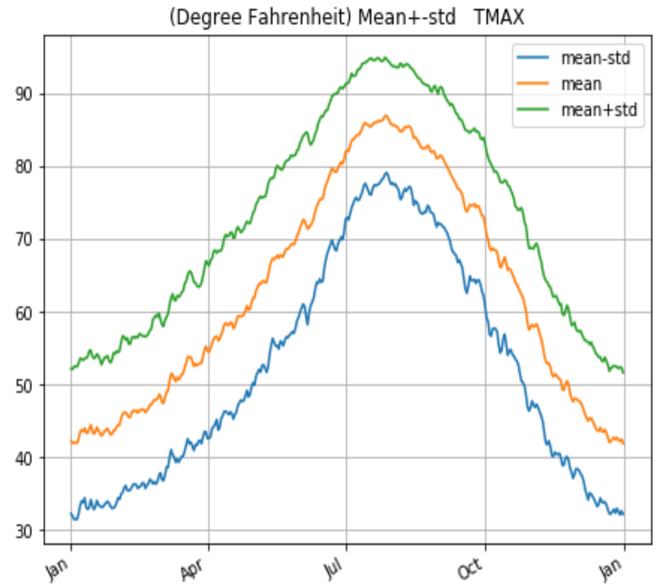
We start by comparing some of the general statistics with graphs that we obtained from a site called US Climate Data (http://www.usclimatedata.com/climate/carson-city/nevada/united-states/usnv0108) The graph below shows the daily minimum and maximum temperatures for each month, as well as the total precipitation for each month.



TMIN and TMAX

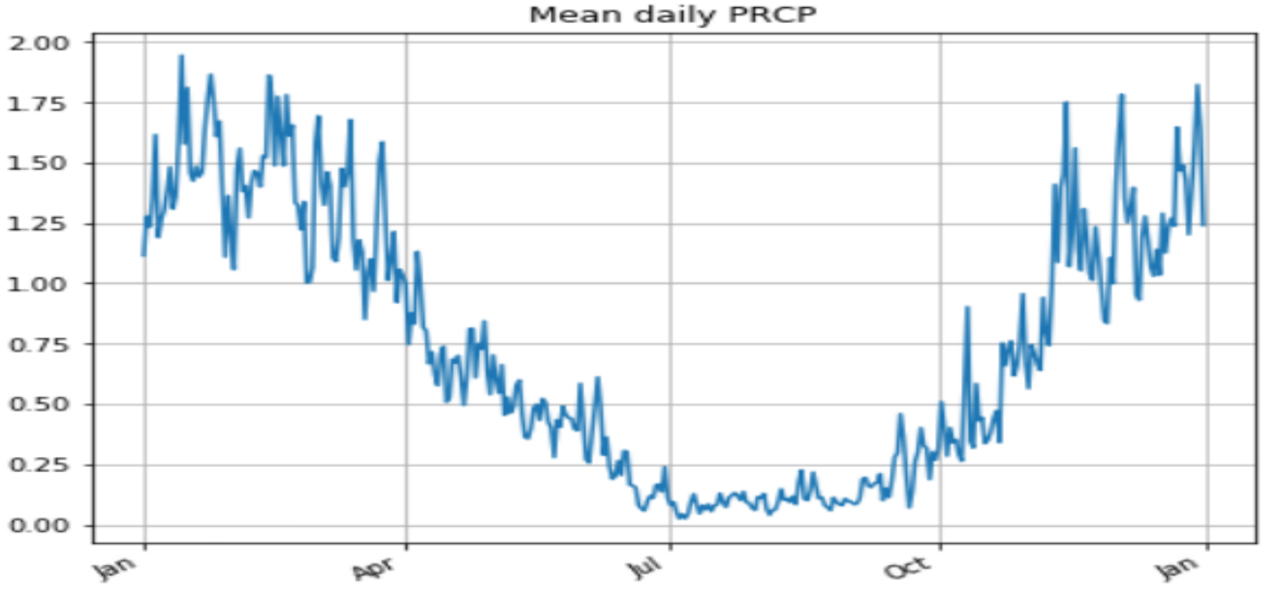
We see that the min and max daily temperature agree with the ones we got from our data.





PRCP

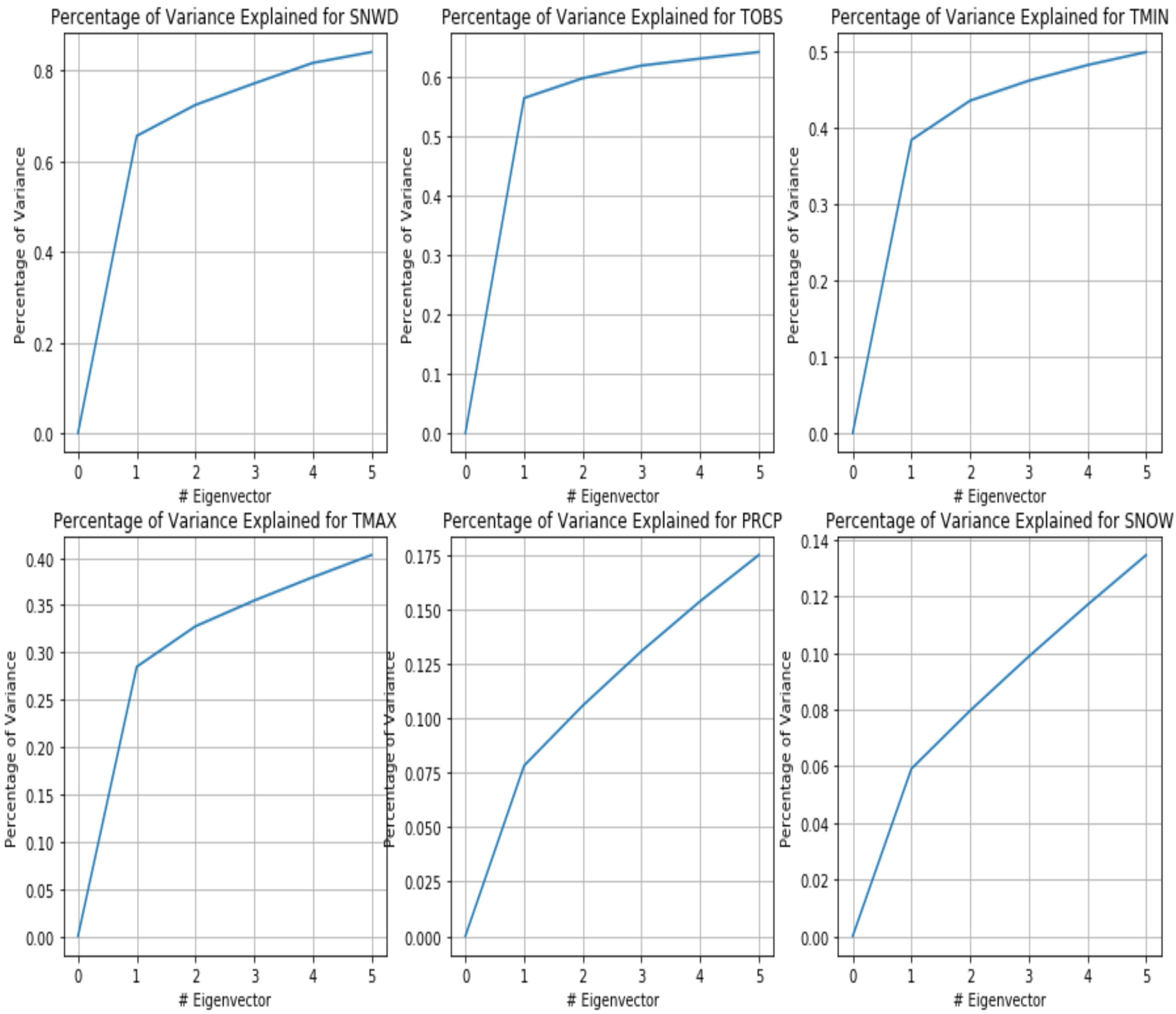
We also see that the mean daily precipitation (in Inches) also almost agrees with January and February months of most precipitation and July, August least



PCA Analysis

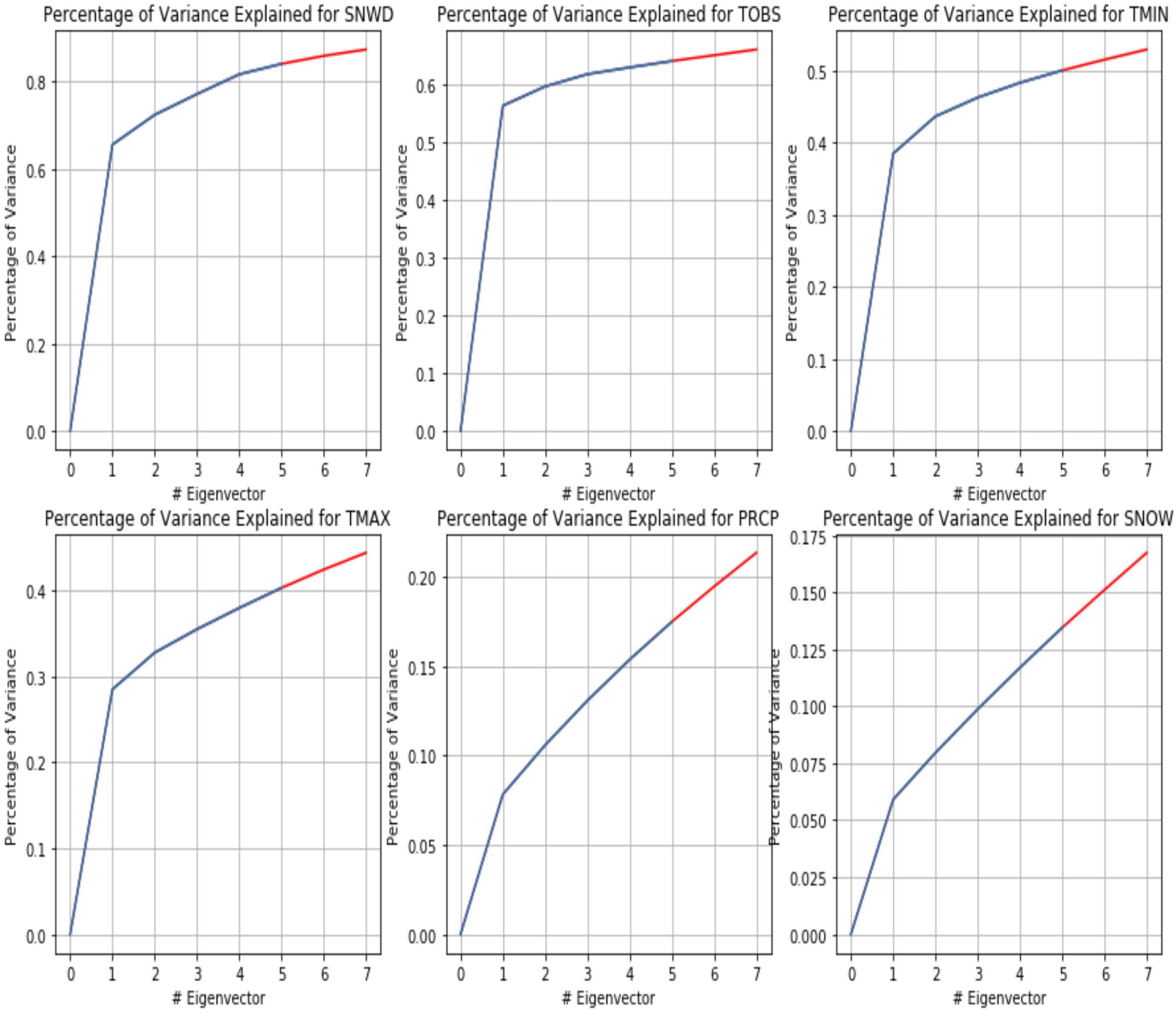
For each of the six measurement, we compute the percentate of the variance explained as a function of the number of eigen-vectors used. For the six measurements analysis based on top 5 eigen vectors, in the decreasing order of variance explained

Measurement	Percentage variance explained
SNWD	82
TOBS	63
TMIN	50
TMAX	41
PRCP	17.5
SNOW	13



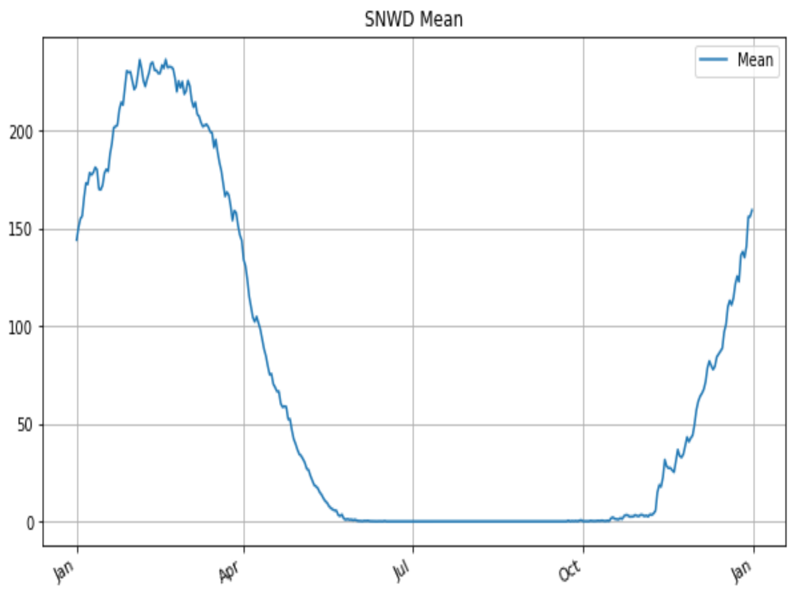
Now let's try to increase the eigen vectors to seven and see if the variance explained gets better

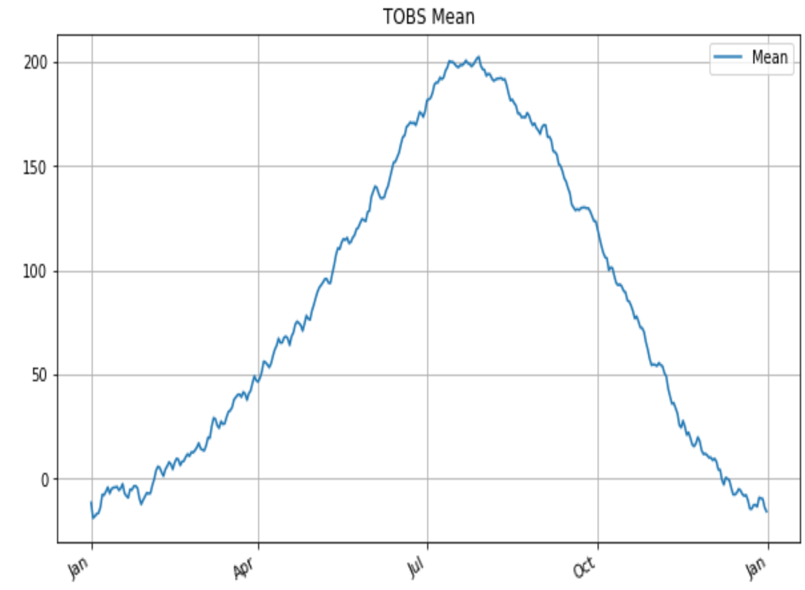
We see that variance explained it almost same for SNWD, TOBS but adds more to the variance for TMIN,TMAX, PRCP and SNOW (delta is in red))

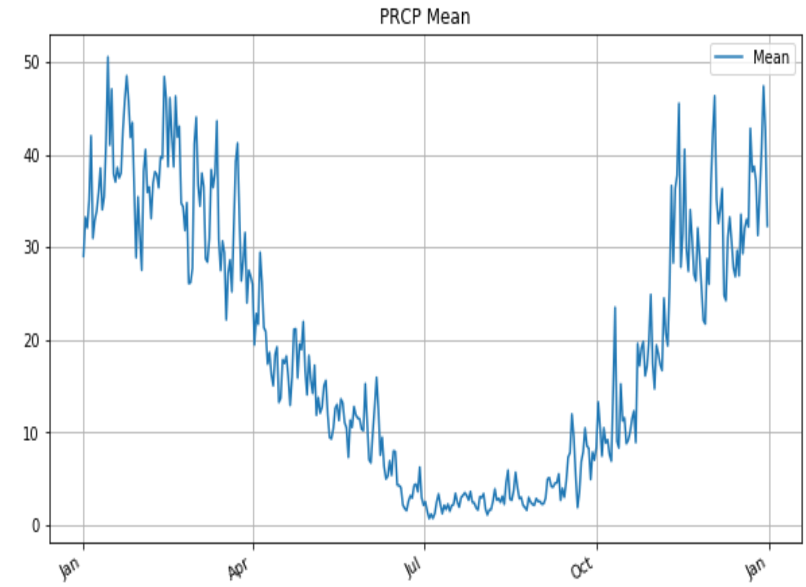


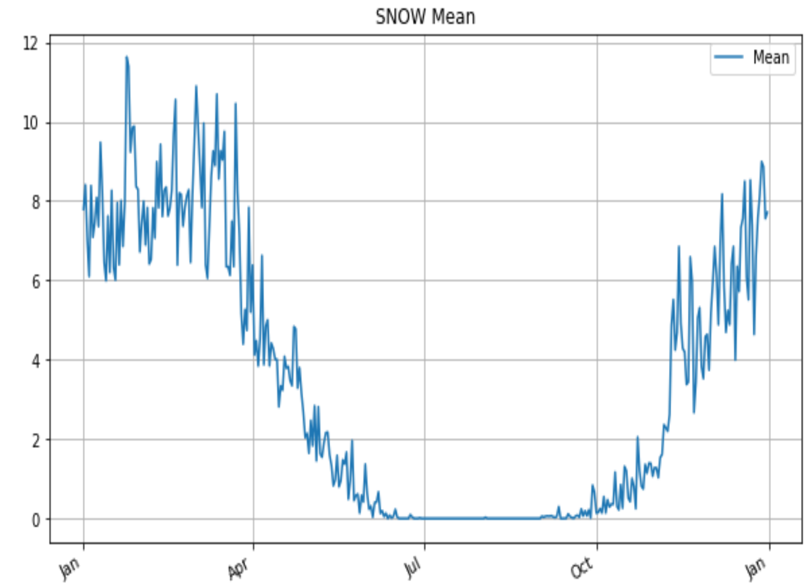
Analysis of Mean

for measurements with eigen vectors with most and least variance i.e. SNWD, TOBS, PRCP and SNOW







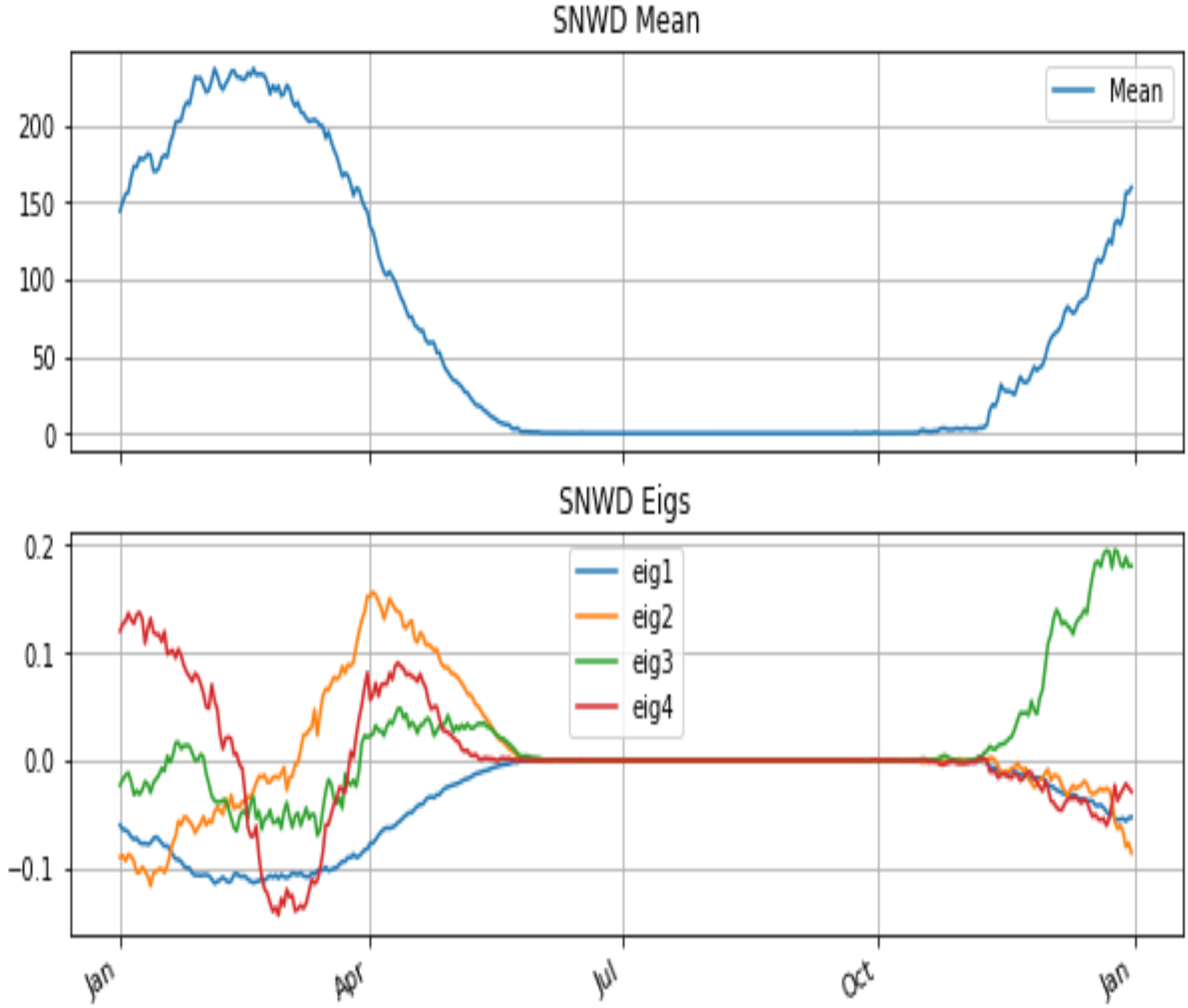


Looking at the four figures together the area is hottest in the months of july to september with no precipitation/snow. The coldest months are mid january to february end with most precipitation/snow.

Will be analyzing SNWD, PRCP and TOBS in detail

Analysis on SNWD

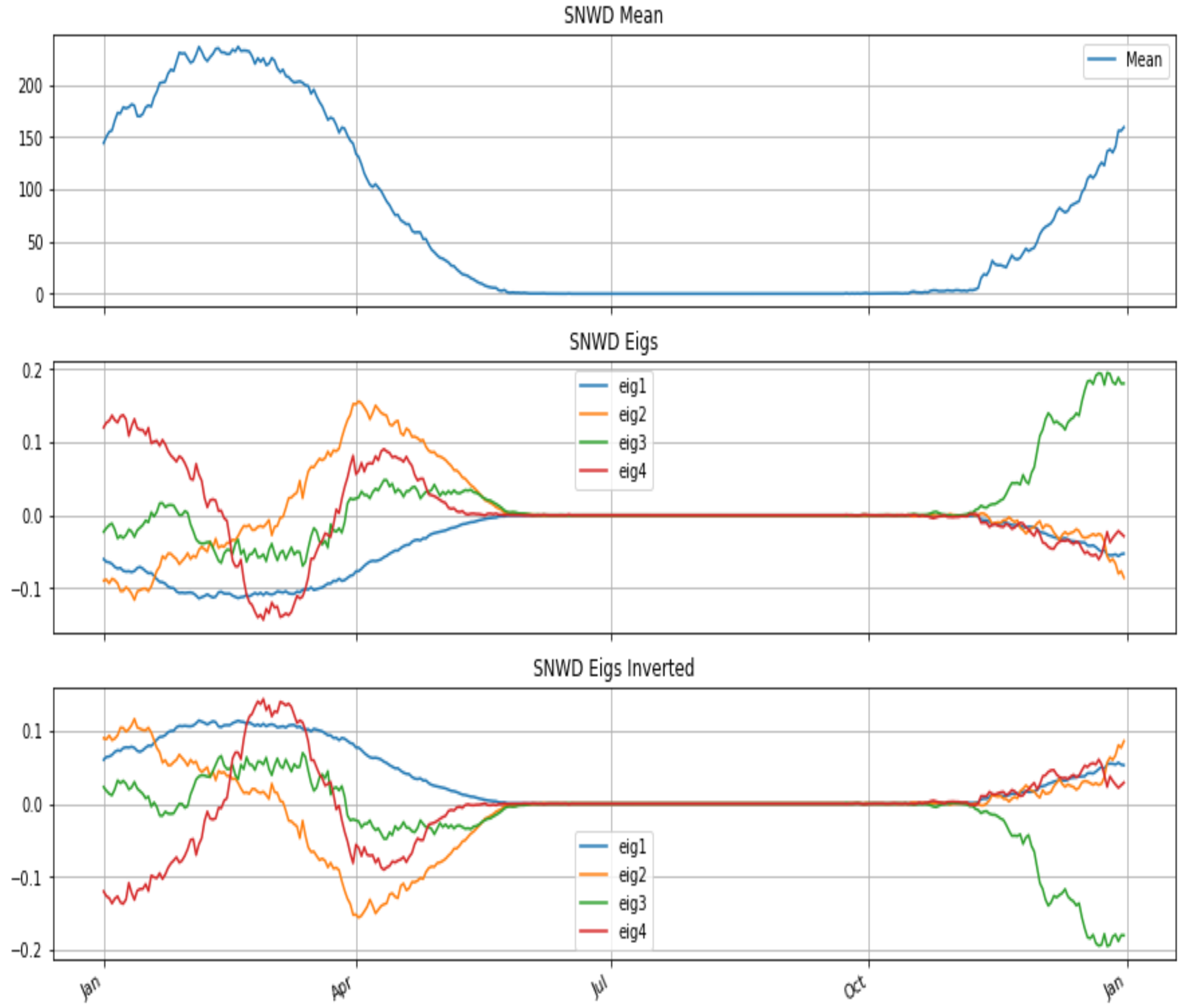
Mean and Eigen vectors Analysis



Next we interpret the eigen-functions. The first eigen-function (eig1) has a shape very similar to the mean function but in opposite direction below 0. The distribution over time stays same.

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).

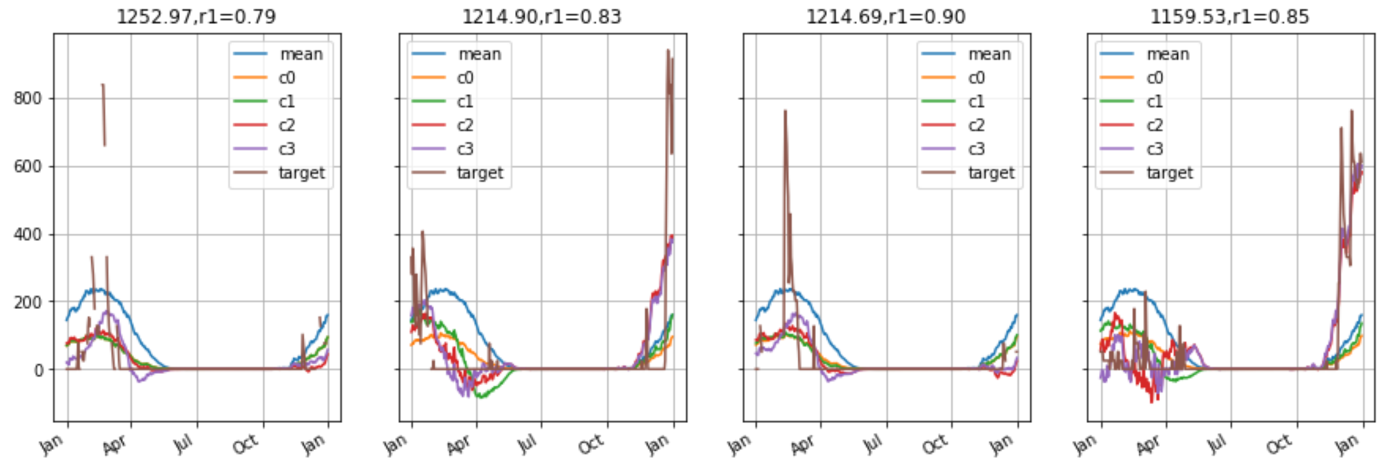
But if we inverse the eigen vectors construction by multiplying by -1 as shown below, they look graphically better; it corroborates our statement that the distribution remains same.



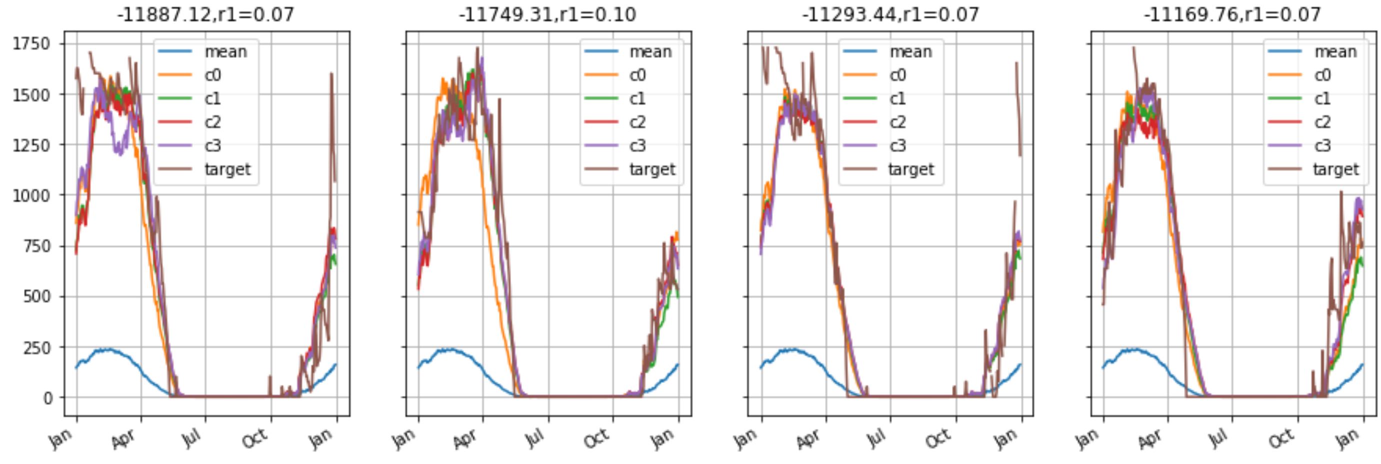
Examples of reconstruction

Coeff1

Coeff1: most positive



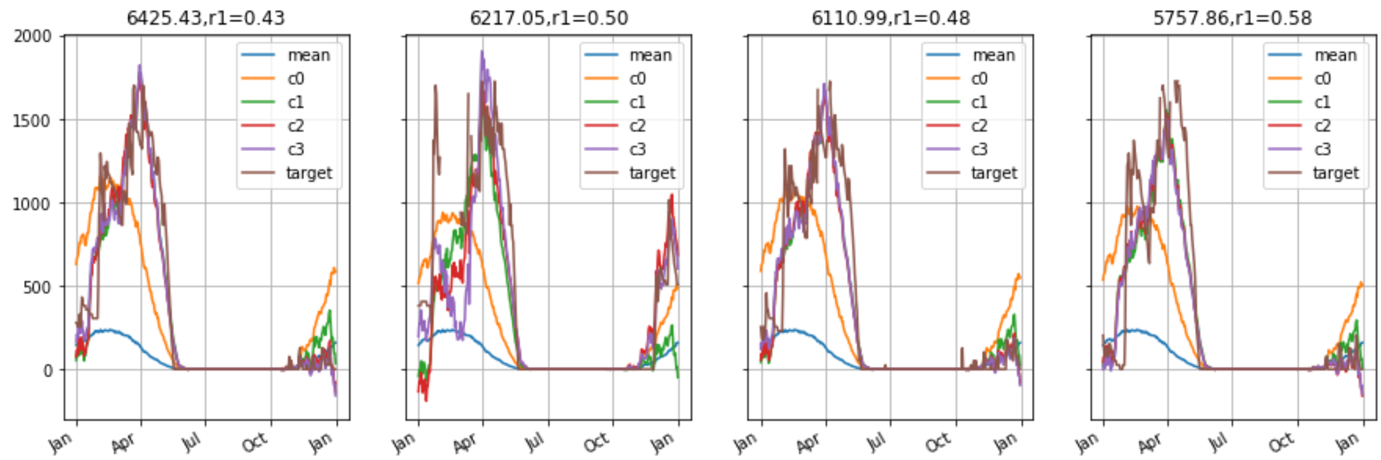
Coeff1: most negative



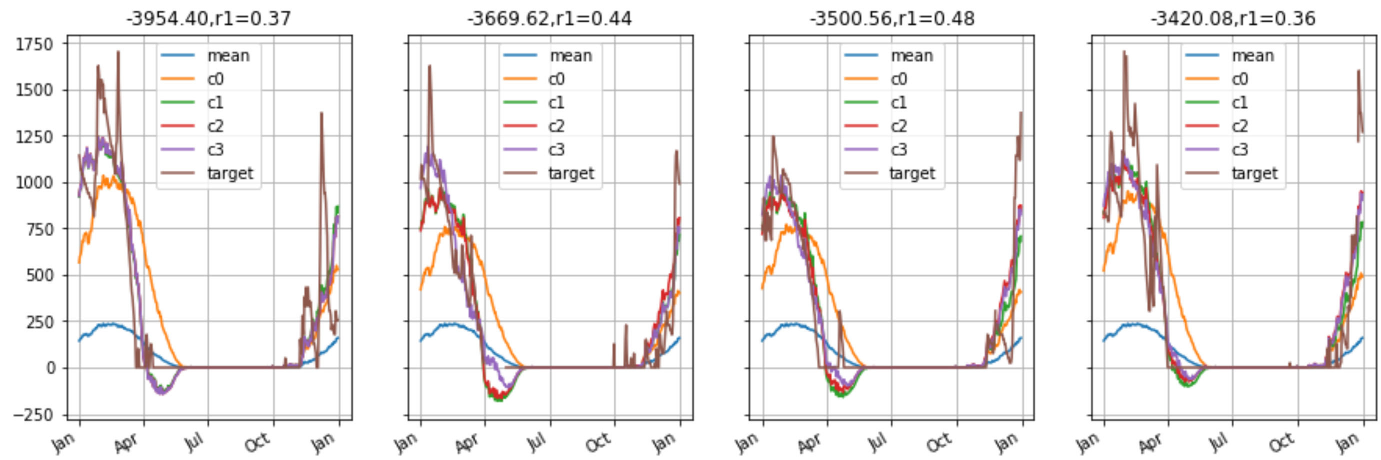
Since the first eigen vector as shown before is in absolute negative direction of mean, the overall effect for top 4 lowered down and for bottom 4 is increased significantly. Even if we are seeing values positive and following the distribution of mean, that indicates more than average snow.

Coeff2

Coeff2: most positive



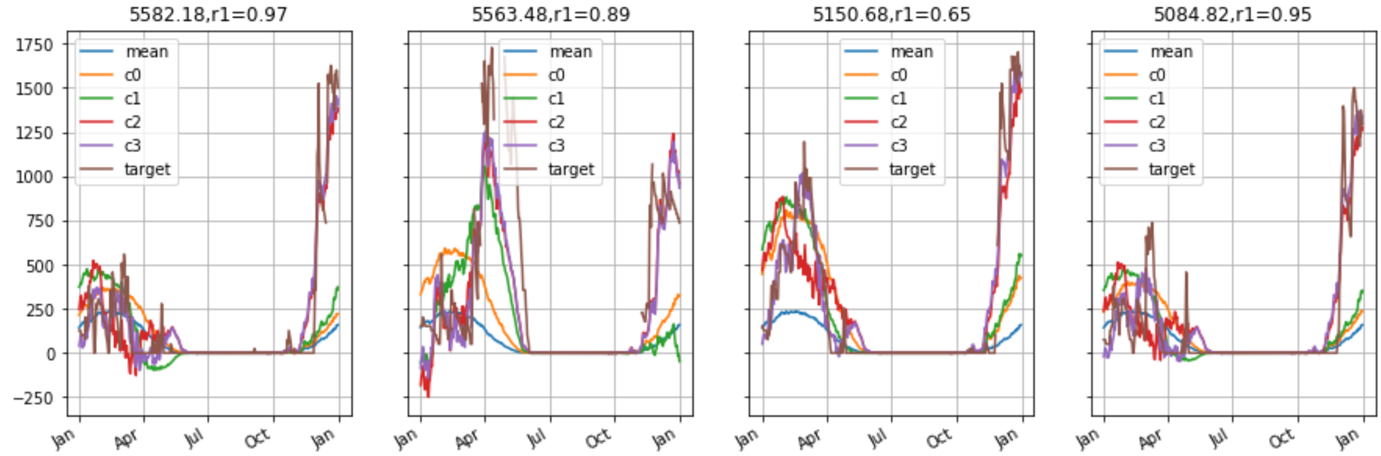
Coeff2: most negative



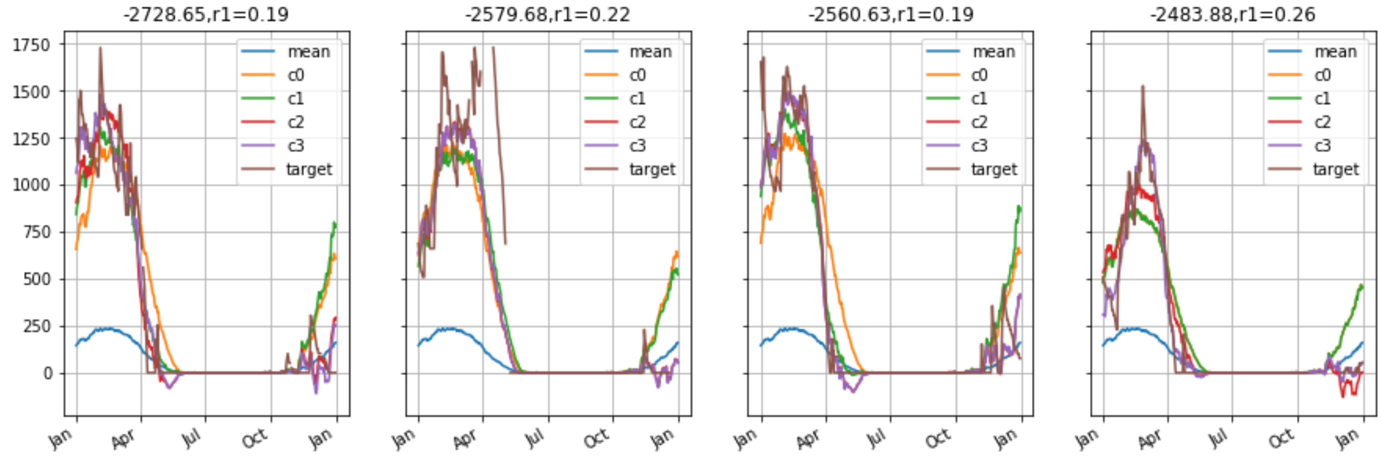
Large positive values of coeff2 correspond to a late snow season (most of the snowfall is after end of April. Negative values for coeff2 correspond to an early snow season (most of the snow is before mid-march.

Coeff3

Coeff3: most positive



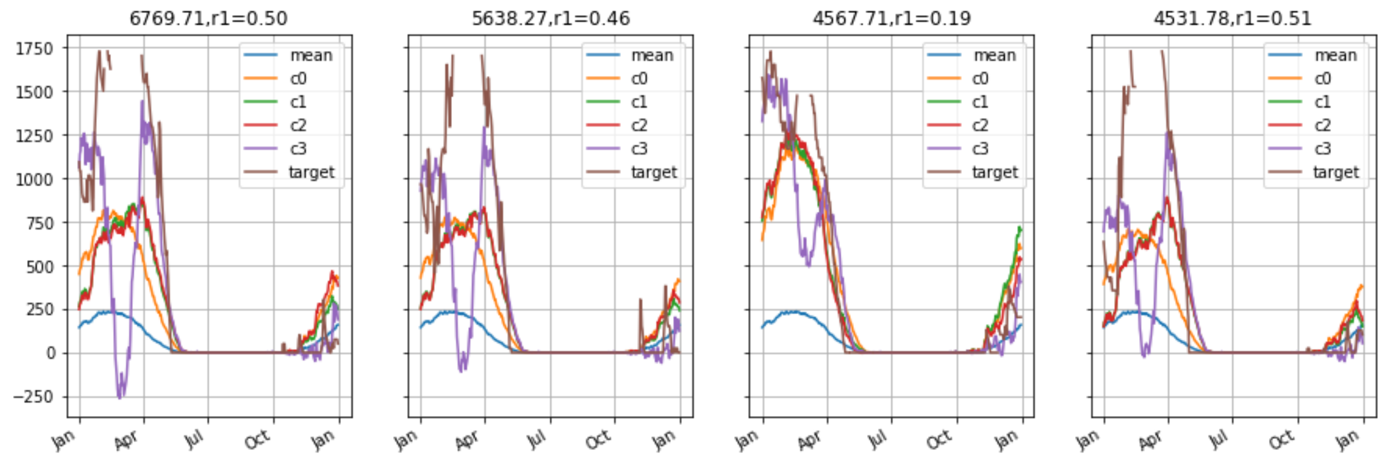
Coeff3: most negative



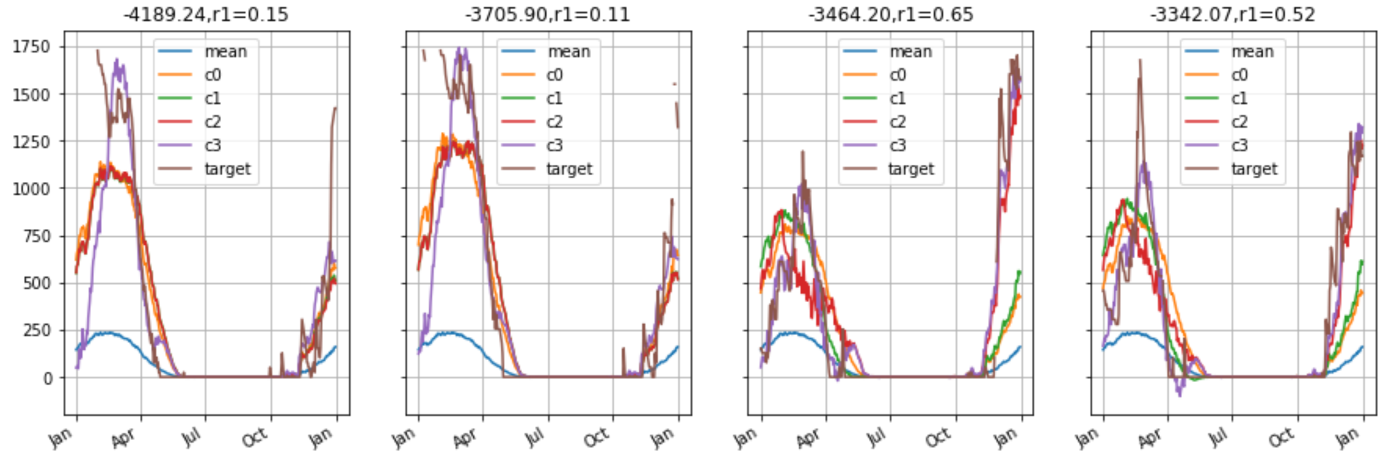
Large positive values of coeff3 correspond to a snow season with one spikes: one in the end of feb, the other at the end of april. Negative values of coeff3 correspond to a season with a single peak at the beginning of Feb.

Coeff4

Coeff4: most positive



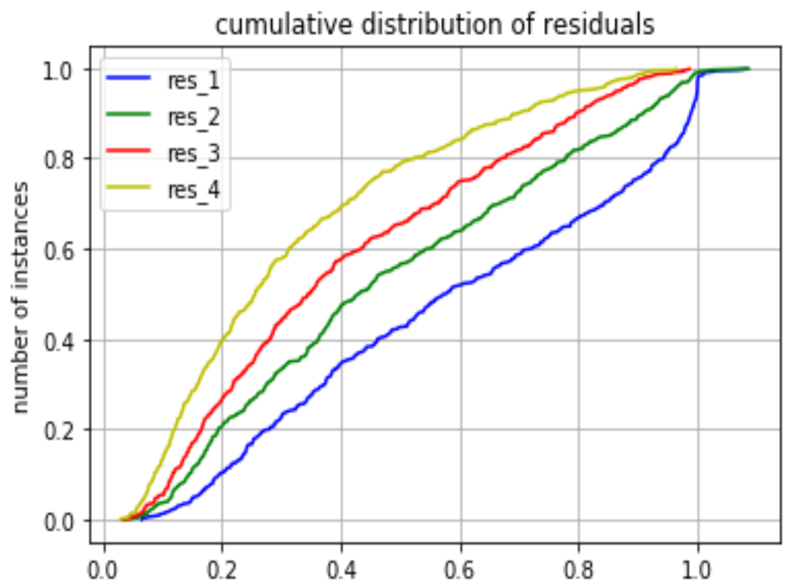
Coeff4: most negative



Positive values show spikes two time when the snow pattern(delta)) changes more than normal i.e it was increasing from Nov to mid Jan then slight dip and back to increasing, then stayed almost constant; then around April end it started decreaing again. Large negative values peak when the snow is highest

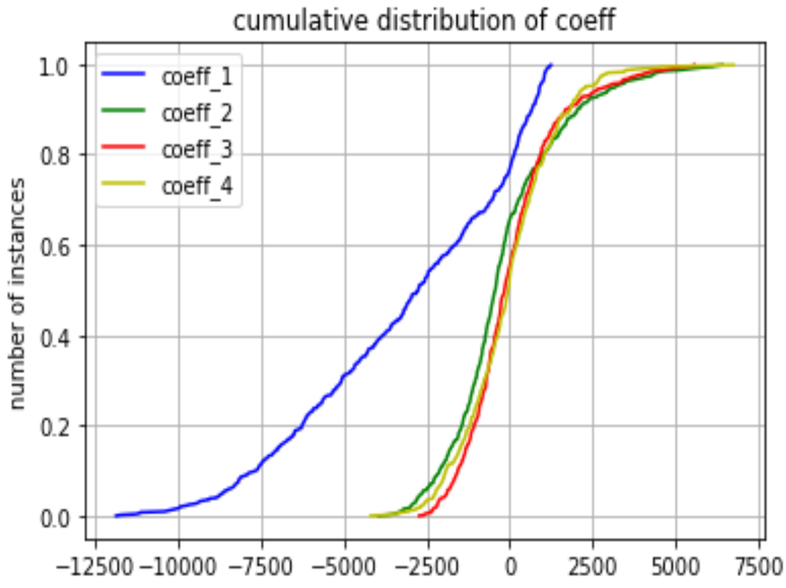
Cummulative Distribution Residuals and Coefficients

Residuals



The residuals for all four coefficient show how well the eigen vectors in the collection are shown

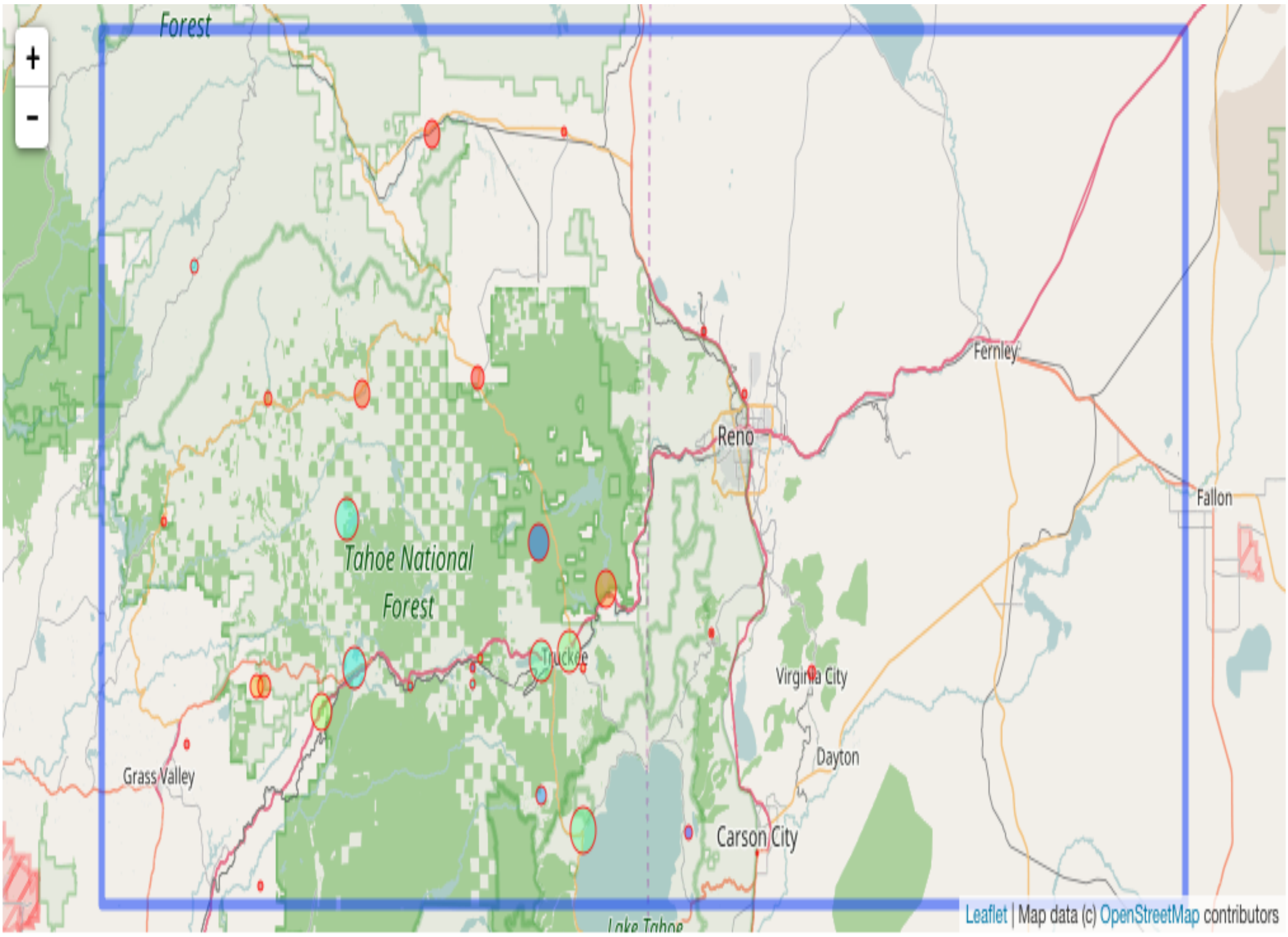
Coeff



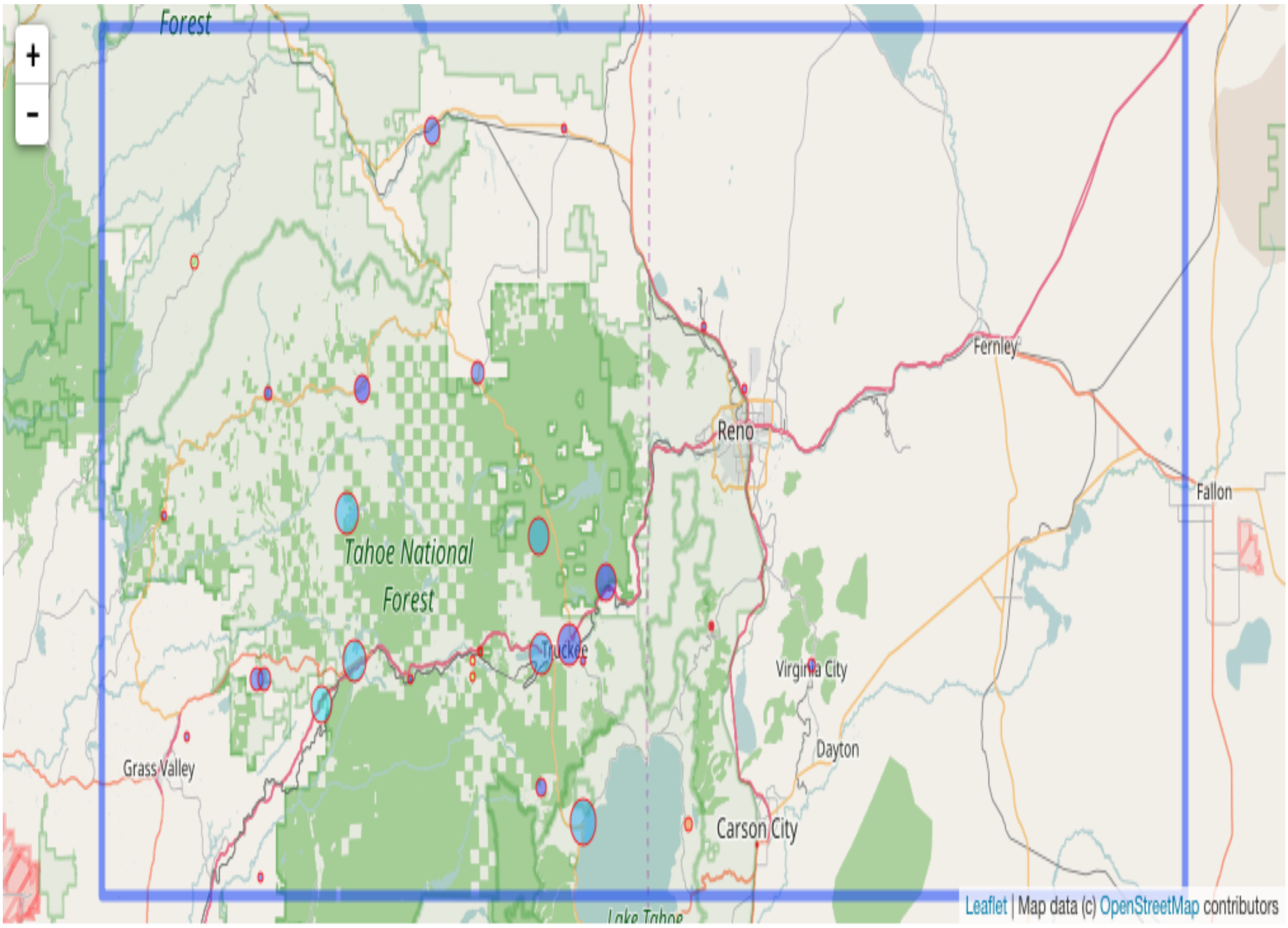
Correlation coeff between stations. 2,3 and 4 shows similar correlation; also depicted in the maps below

Geographical distribution of top 4 coefficients.

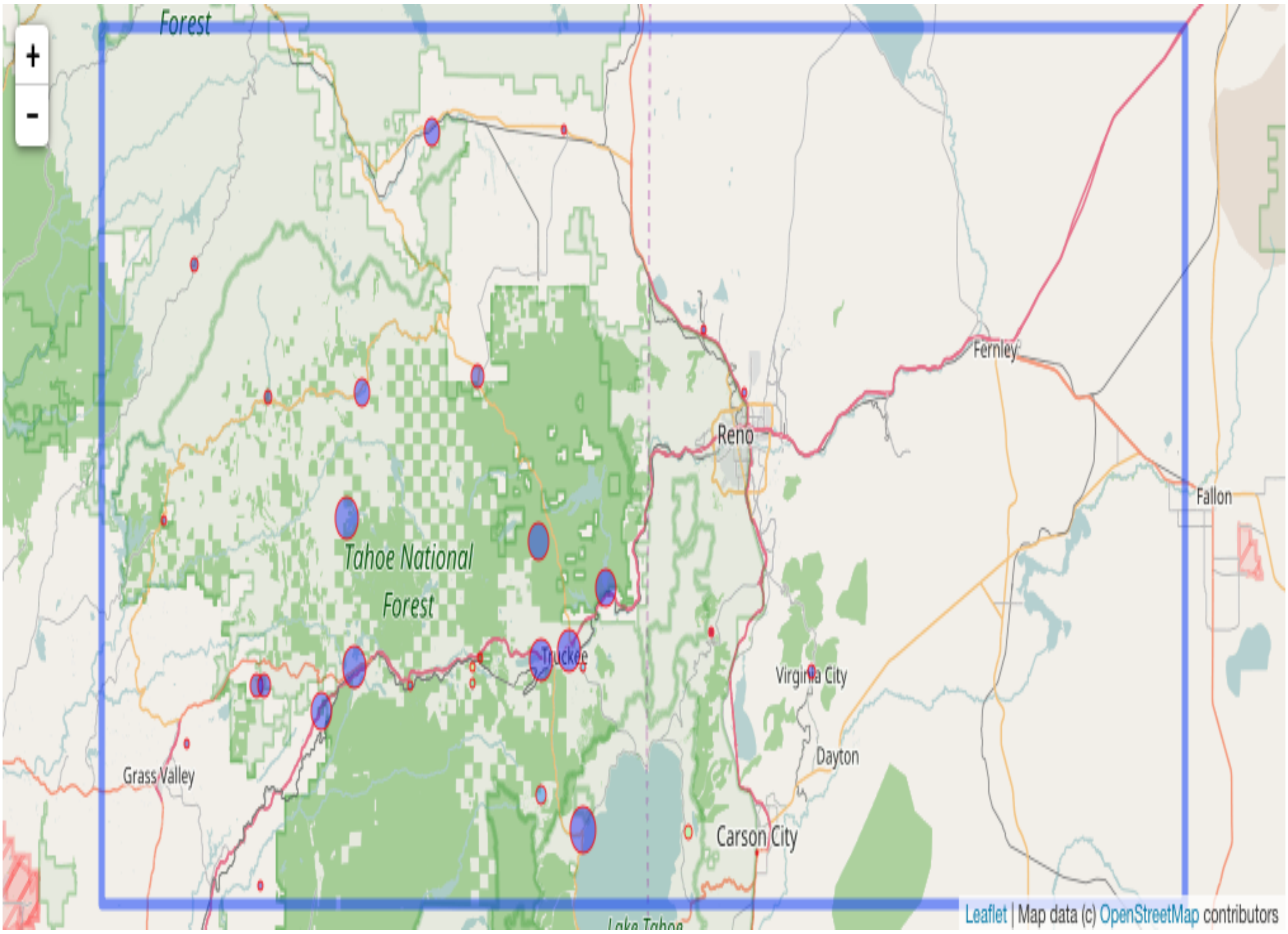
coeff1



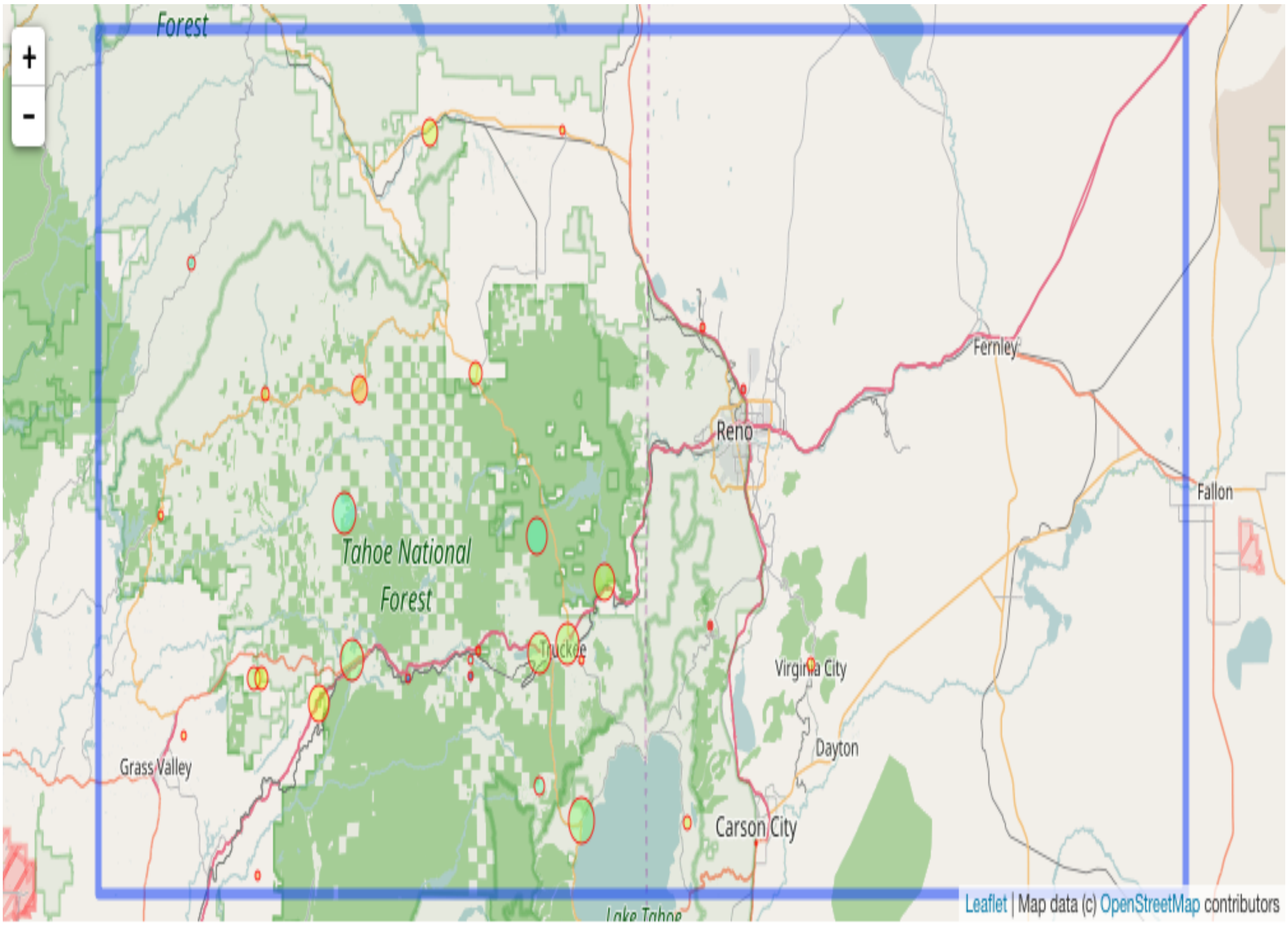
coeff2


+
-
Forest
Fernley
Reno
Fallon
Tahoe National
Forest
Truckee
Virginia City
Grass Valley
Dayton
Carson City
Lake Tahoe
Leaflet | Map data (c) OpenStreetMap contributors

coeff3


Forest
+
-
Tahoe National Forest
Reno
Fernley
Fallon
Truckee
Virginia City
Dayton
Grass Valley
Carson City
Lake Tahoe
Leaflet | Map data (c) OpenStreetMap contributors

coeff4



Most data came from stations located centrally close to each other but the contribution to coeff is more from far away distributed ones. It is interesting that the coeff 3 for most stations is same.

Estimating the effect of the year vs the effect of the station

Computing for coeff_1

total RMS = 1719.45712835

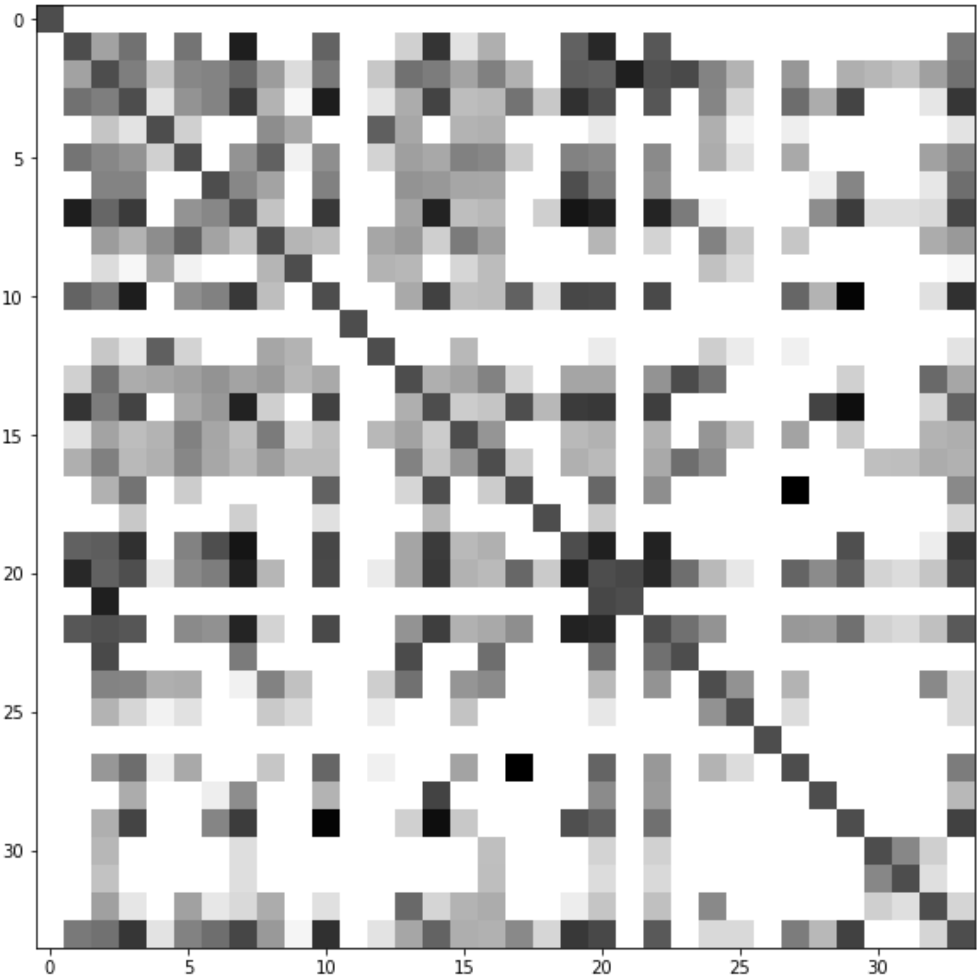
RMS removing mean-by-station= 1609.63676289

RMS removing mean-by-year = 963.98295144

Year has more effect

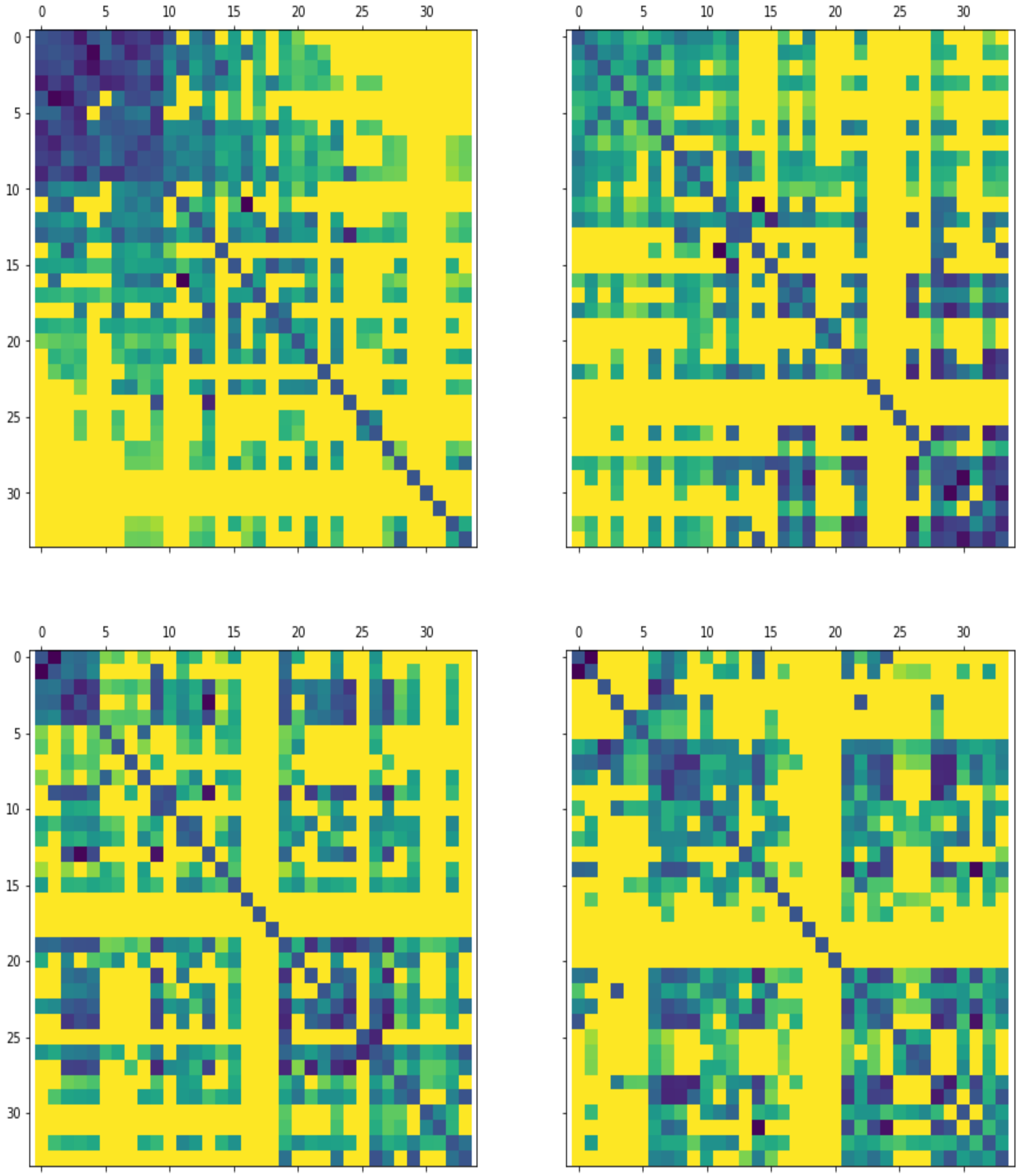
Dependency Matrix to show correlation between stations

i.e. whether or not it snowed on the same day in different stations



The matrix above shows, for each pair of stations, the normalized log probability that the overlap in snow days is random. We see immediately the first 4 stations are highly correlatedwith each other.

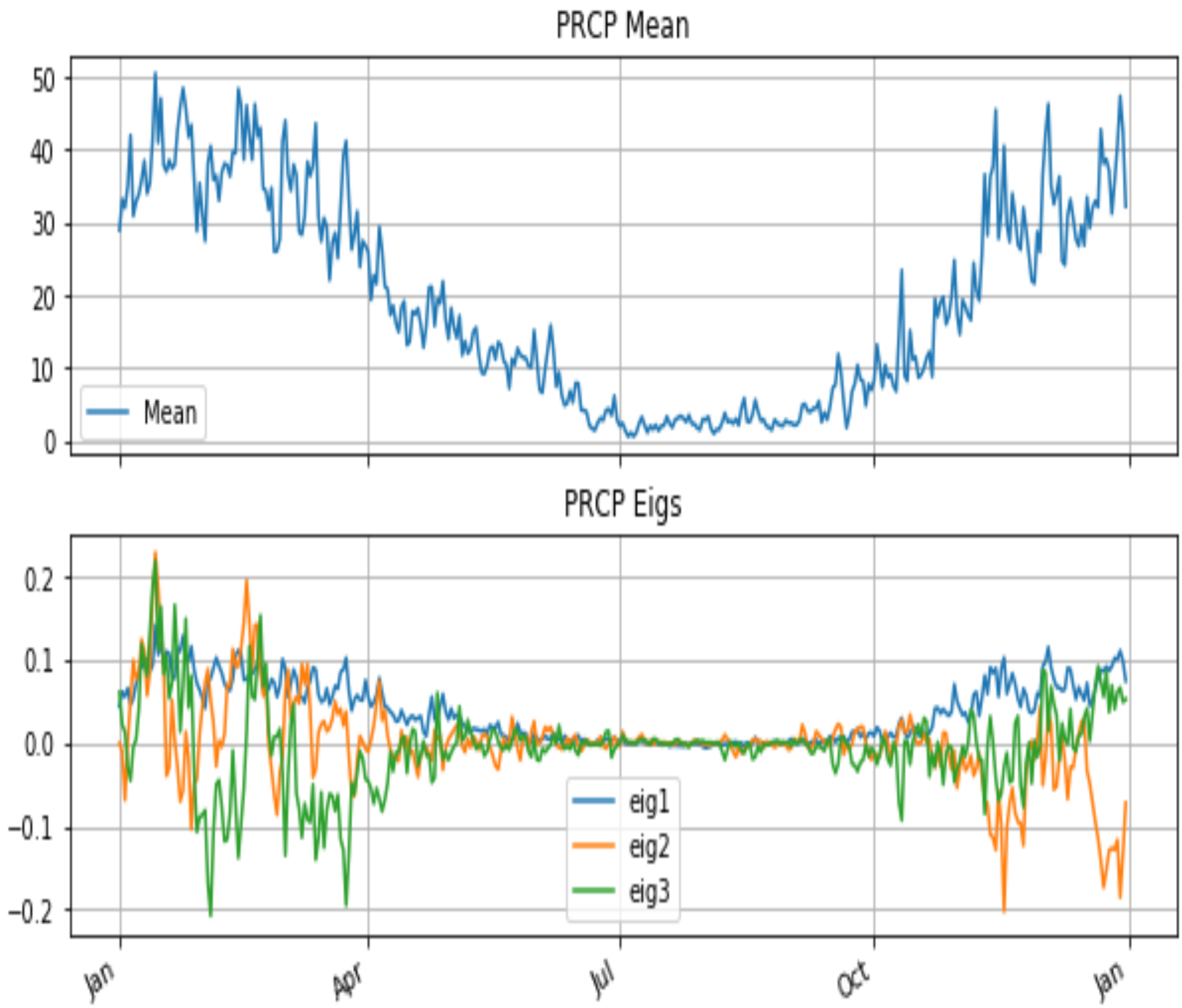
Reordering rows and columns to do more analysis



When we reorder the rows and columns of the matrix using one of the eigenvectors, the grouping of the stations becomes more evident. For example, consider the upper left corner of the first matrix (The upper left one). The stations at positions 0-10 are clearly strongly correlated with each other. On the upper left corner of matrix 2 (The upper right one) stations 0-12 re highly correlated.

Analysis on PRCP

Mean and Eigen vectors Analysis



Using only three eigen vectors for analysis here

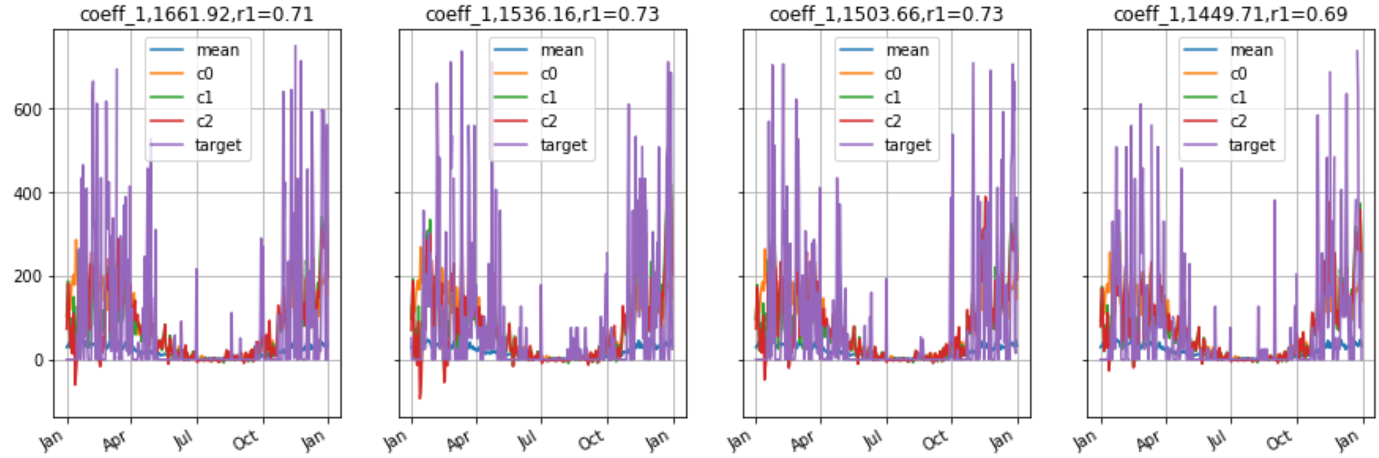
Next we interpret the eigen-functions. The first eigen-function (eig1) has a shape very similar to the mean function. The interpretation of this shape is that eig1 represents the overall amount of precipitation above/below the mean, but without changing the distribution over time.

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 precipitation over the winter months, but they don't change the total (much).

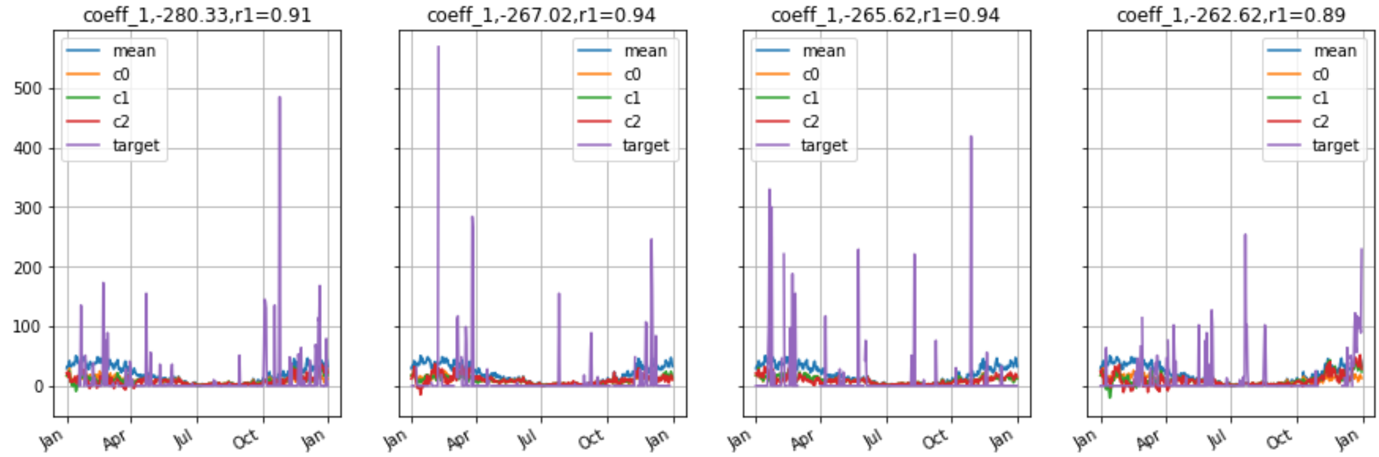
Examples of reconstruction

Coeff1

Coeff1: most positive



Coeff1: most negative

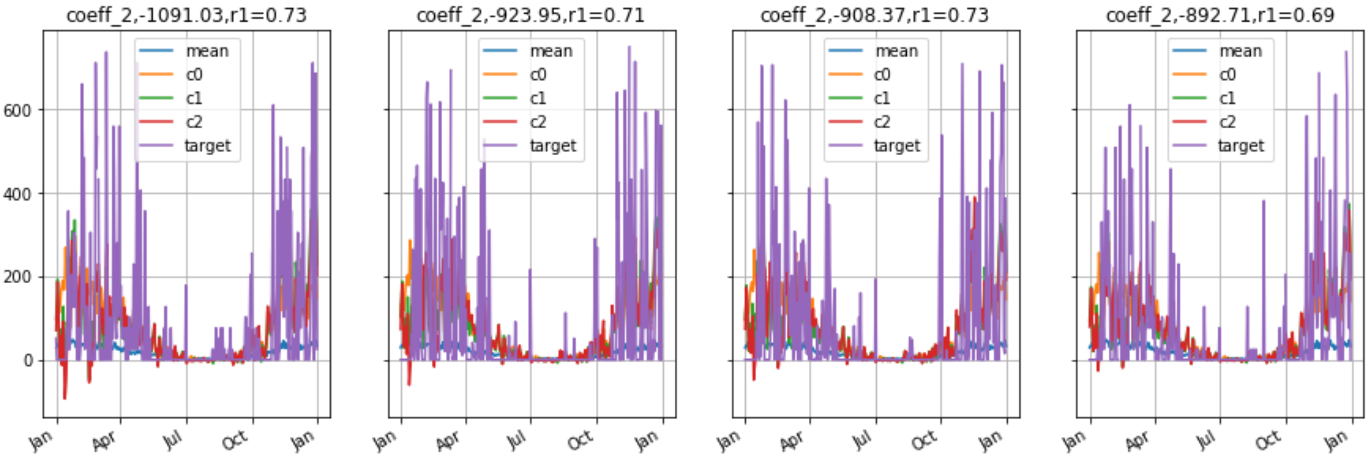


Large positive values of coeff1 correspond to more than average prcp. Low values correspond to less than average prcp.

Coeff2

Coeff2: most positive

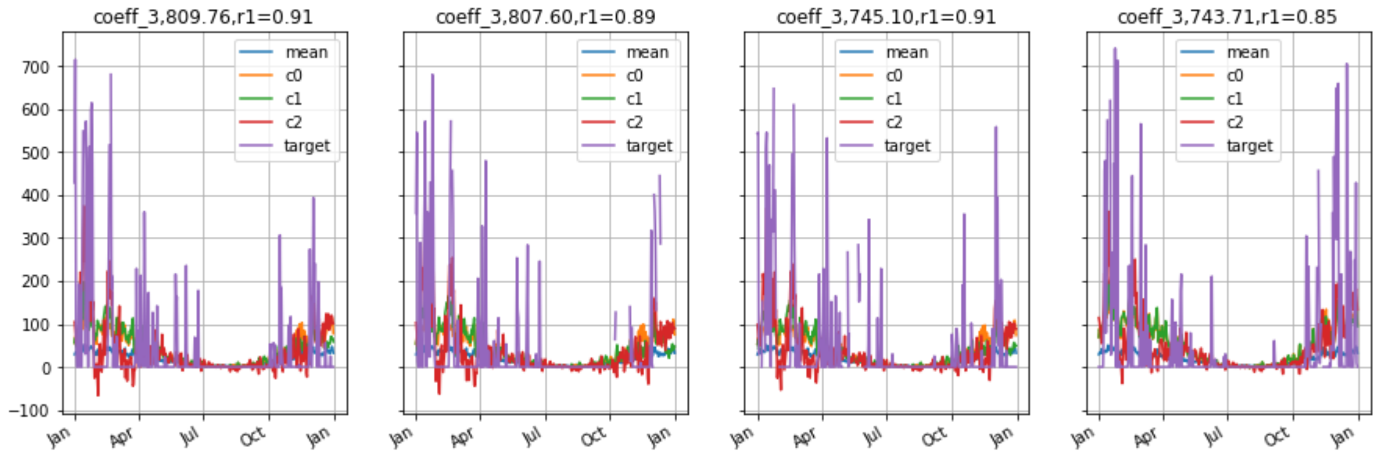
Coeff2: most negative



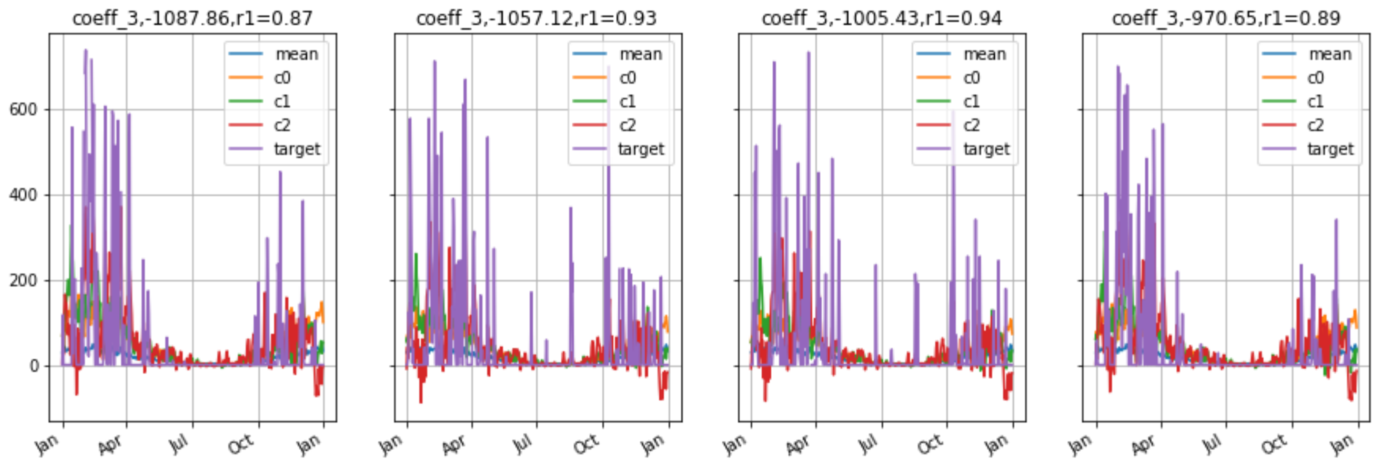
This is very similar to coeff_1

Coeff3

Coeff3: most positive



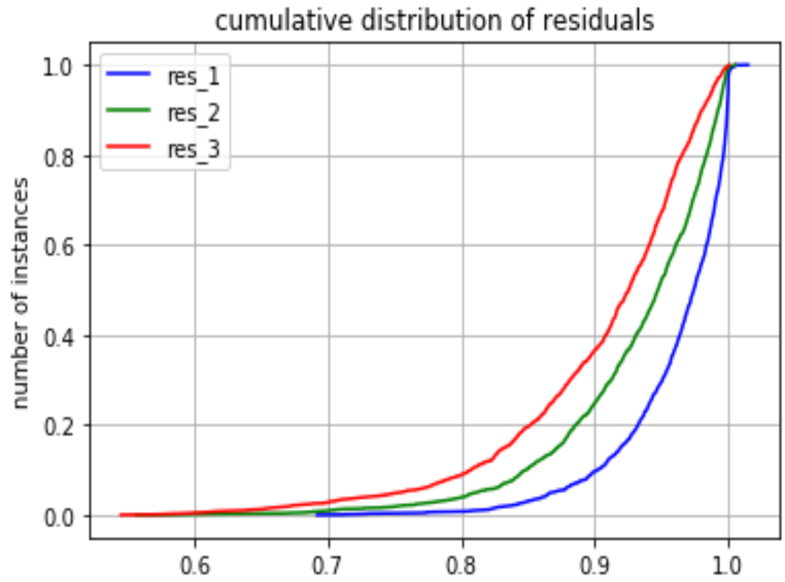
Coeff3: most negative



Extreme Spikes in both positive and negative show where deltas in prcp mean were significant, for instance towards feb and dec.

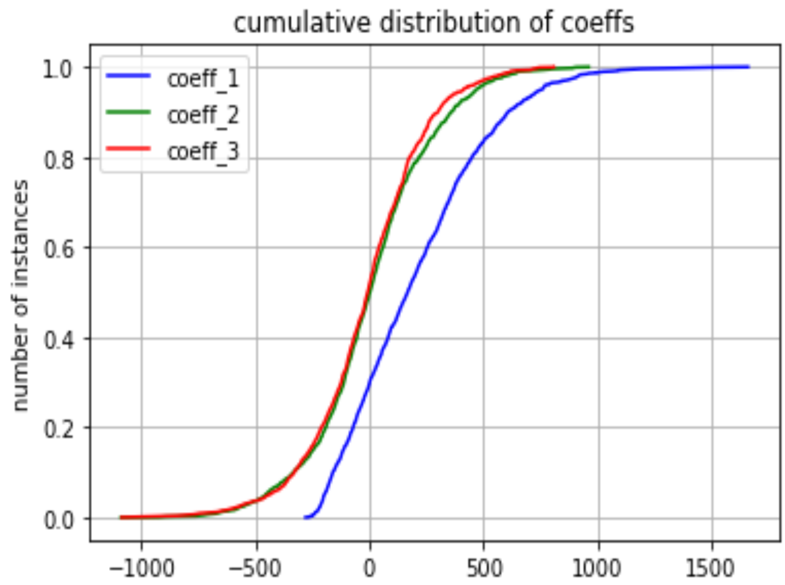
Cummulative Distribution Residuals and Coefficients

Residuals



The residuals for all three coefficient show how well the eigen vectors in the collection are shown

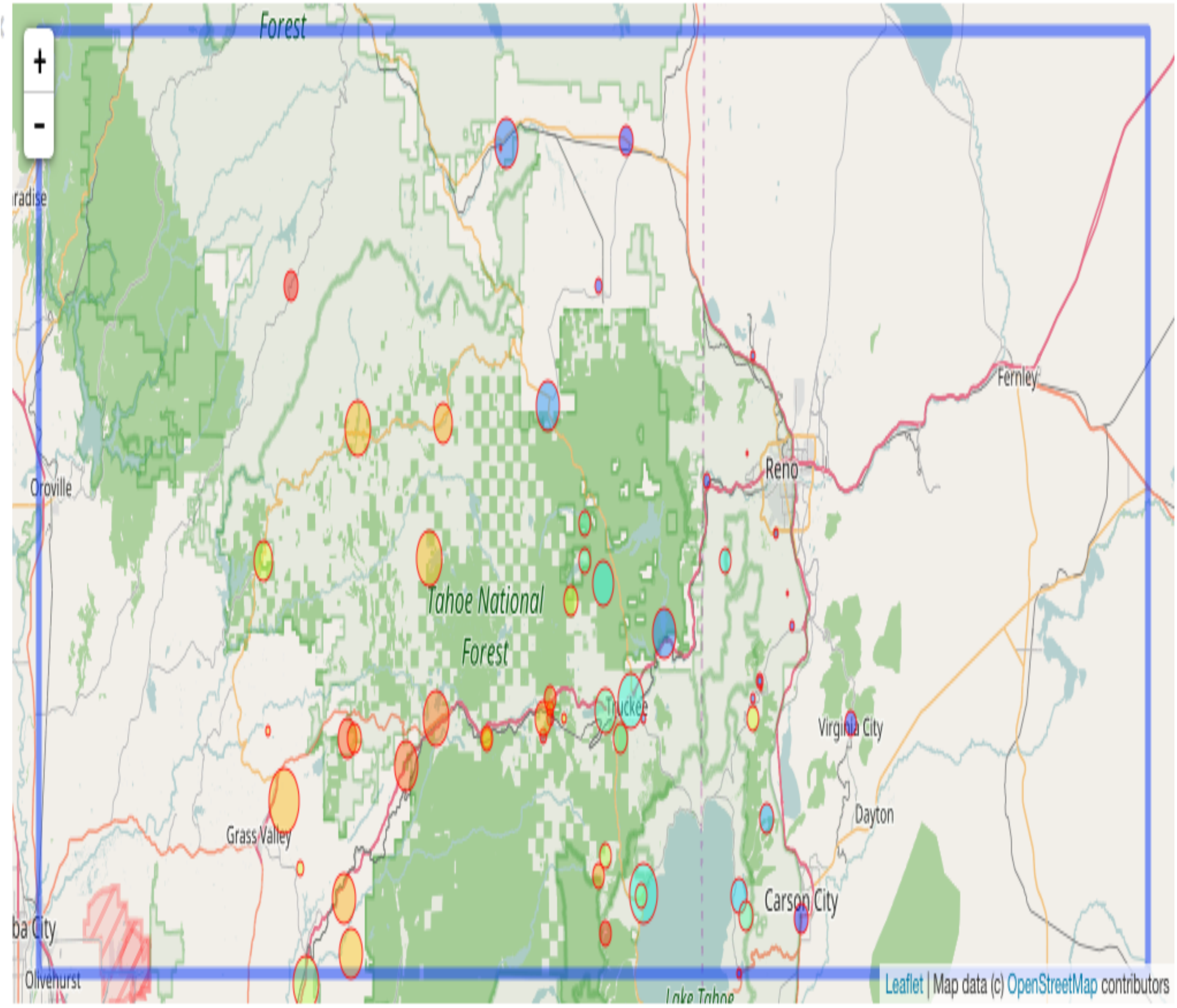
Coeff



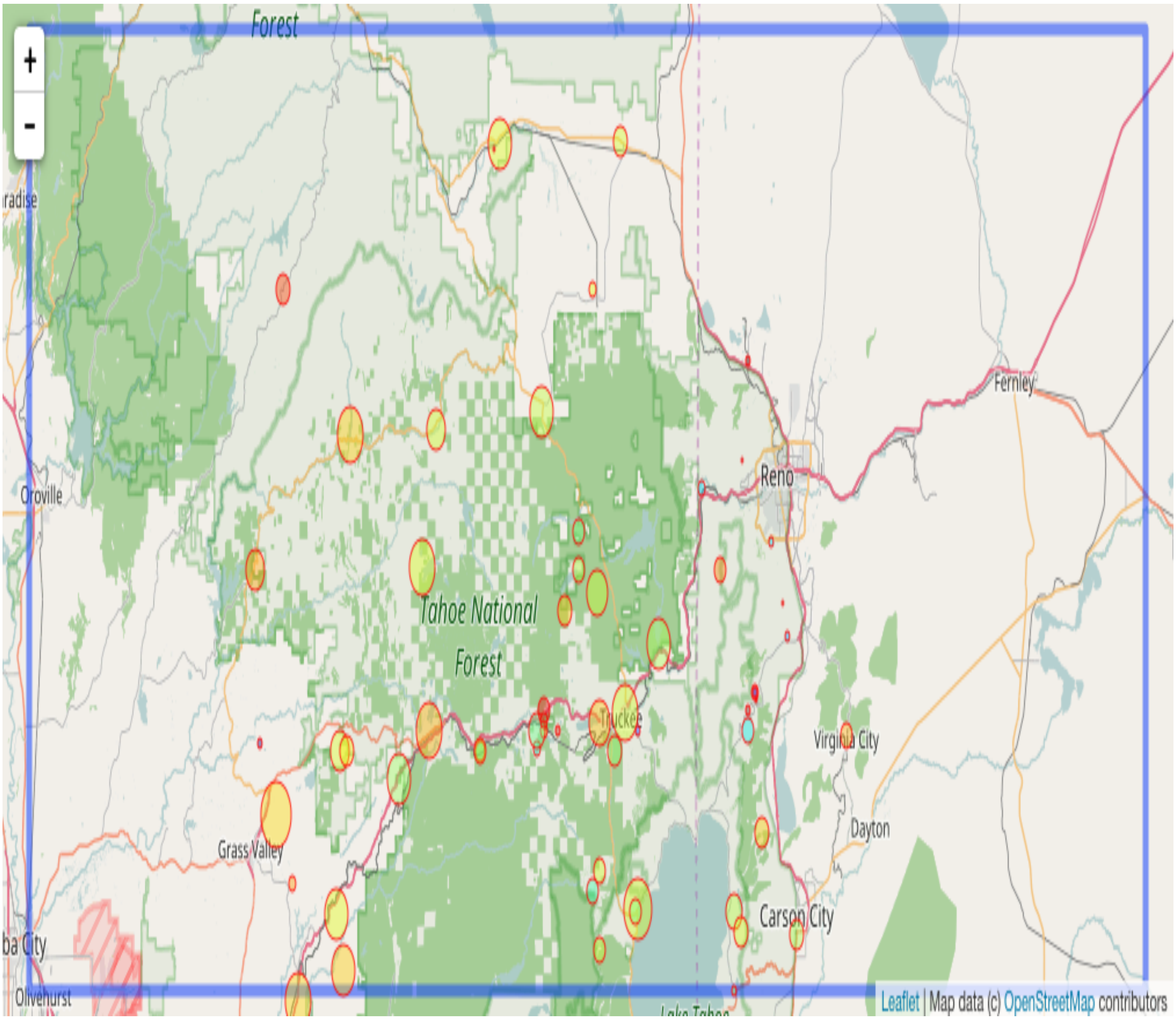
Coeff 2 and 3 show similar correlation pattern

Geographical distribution of top 3 coefficients.

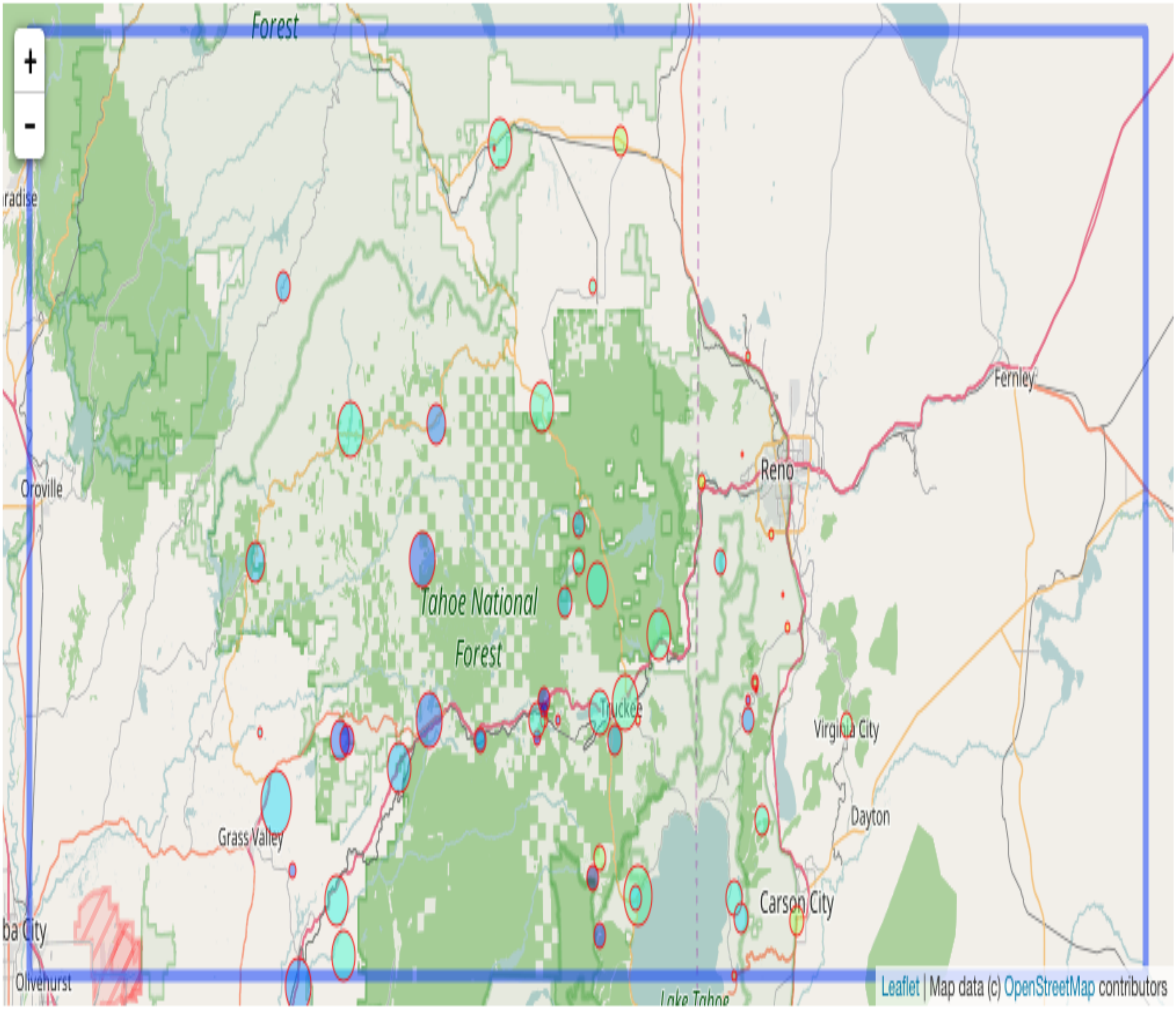
coeff1


Forest
+
-
radise
Oroville
Tahoe National
Forest
Reno
Fernley
Truckee
Virginia City
Dayton
Grass Valley
Carson City
ba City
Olivehurst
Lake Tahoe
Leaflet | Map data (c) OpenStreetMap contributors

coeff2


+
-
Forest
Oroville
Tahoe National
Forest
Fernley
Reno
Truckee
Virginia City
Grass Valley
Dayton
Carson City
Olivehurst
Leaflet | Map data (c) OpenStreetMap contributors

coeff3



Most data recorded from the central stations. The coeff contribution for coeff_1 is significantly different from most stations; shows more variation, but coeff_2 it gets similar and coeff3 is almost same for stations which are closer together

Estimating the effect of the year vs the effect of the station

Computing for coeff_1

total RMS = 273.557600407

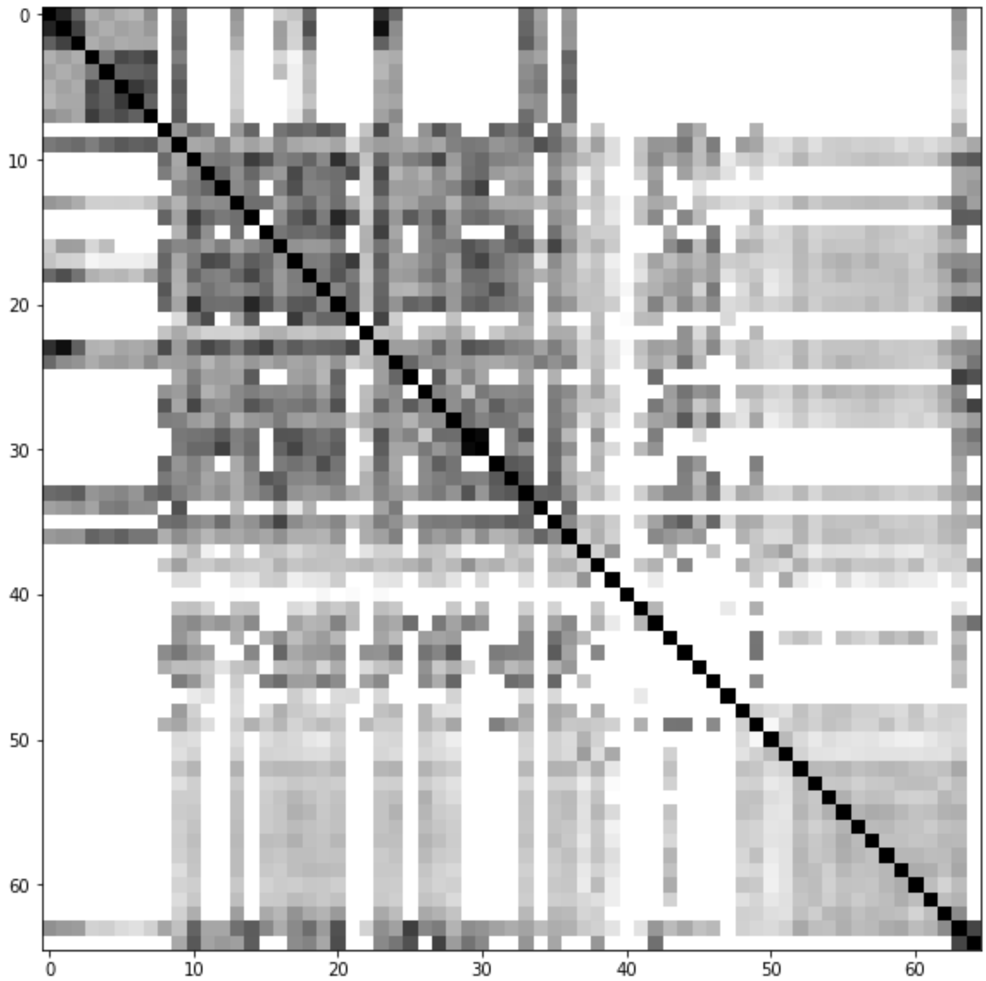
RMS removing mean-by-station= 271.592048642

RMS removing mean-by-year = 111.199352873

Year has more effect

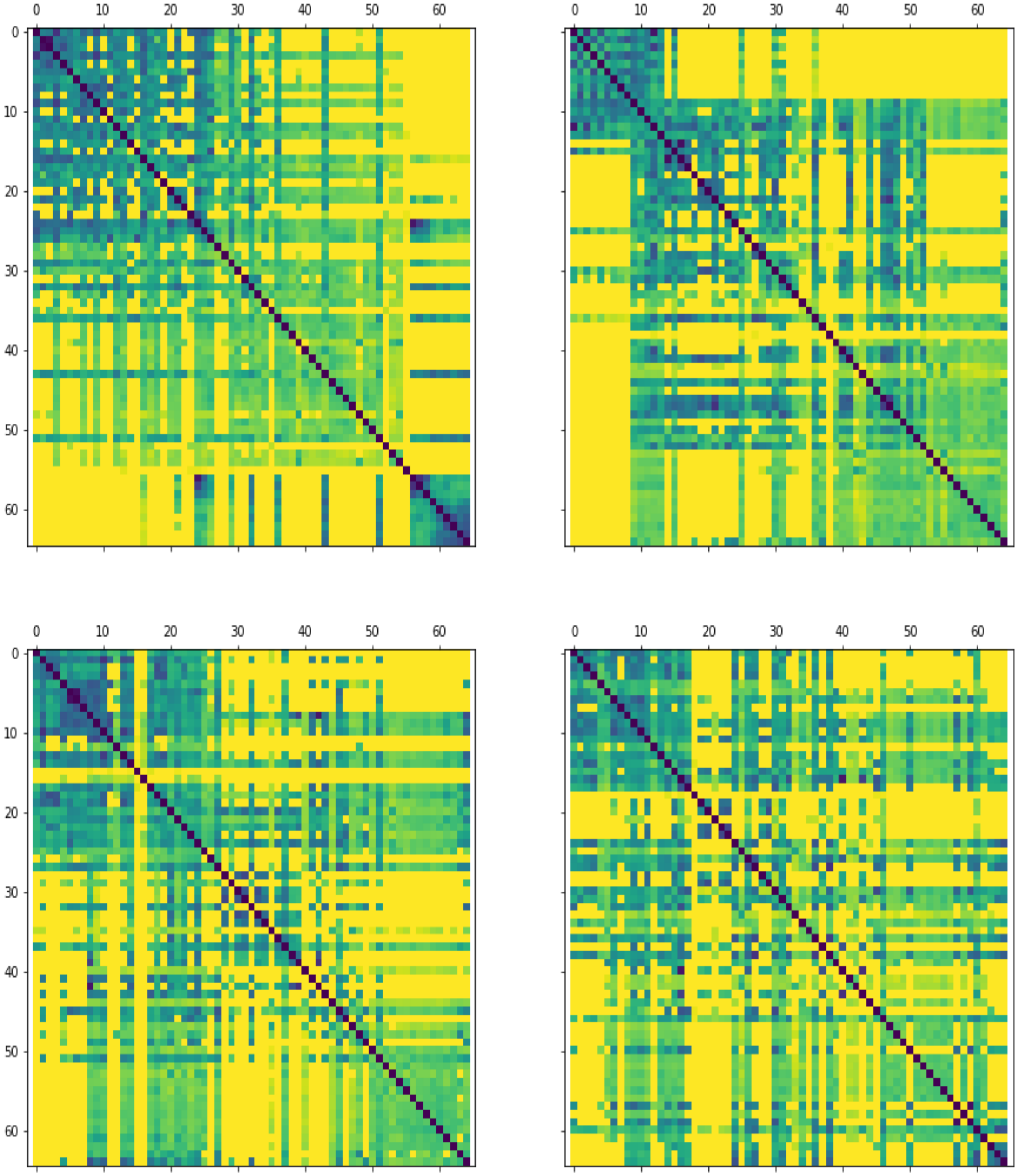
Dependency Matrix to show correlation between stations

i.e. whether or not it snowed on the same day in different stations



The matrix above shows, for each pair of stations, the normalized log probability that the overlap in prcp days is random. We see immediately the first 8 stations are highly correlatedwith each other.

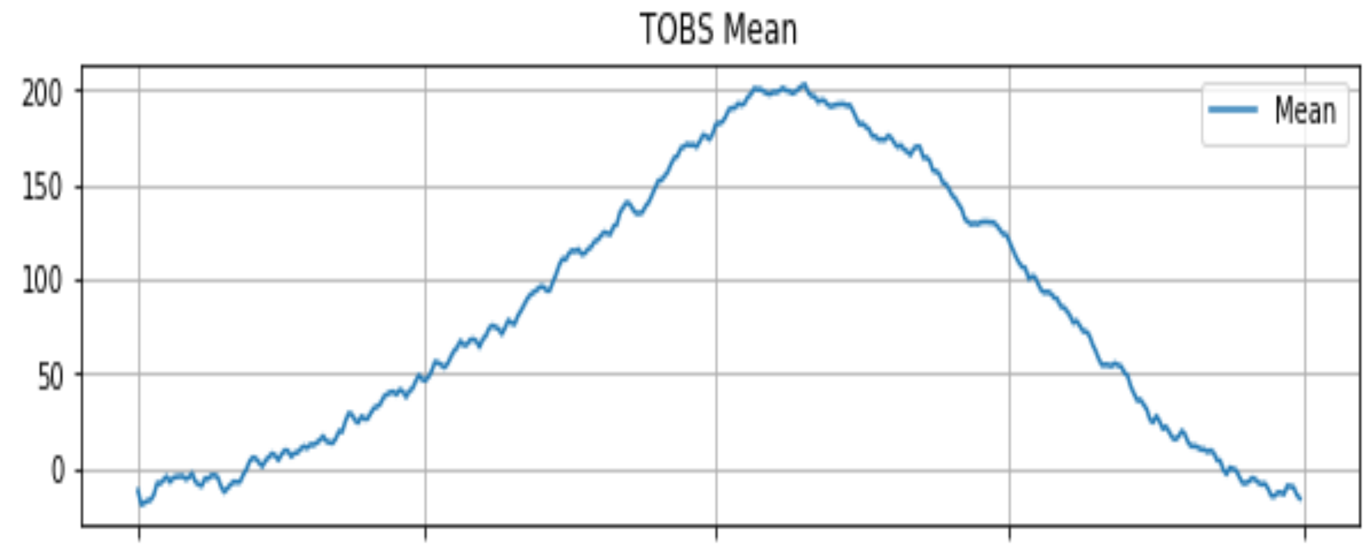
Reordering rows and columns to do more analysis

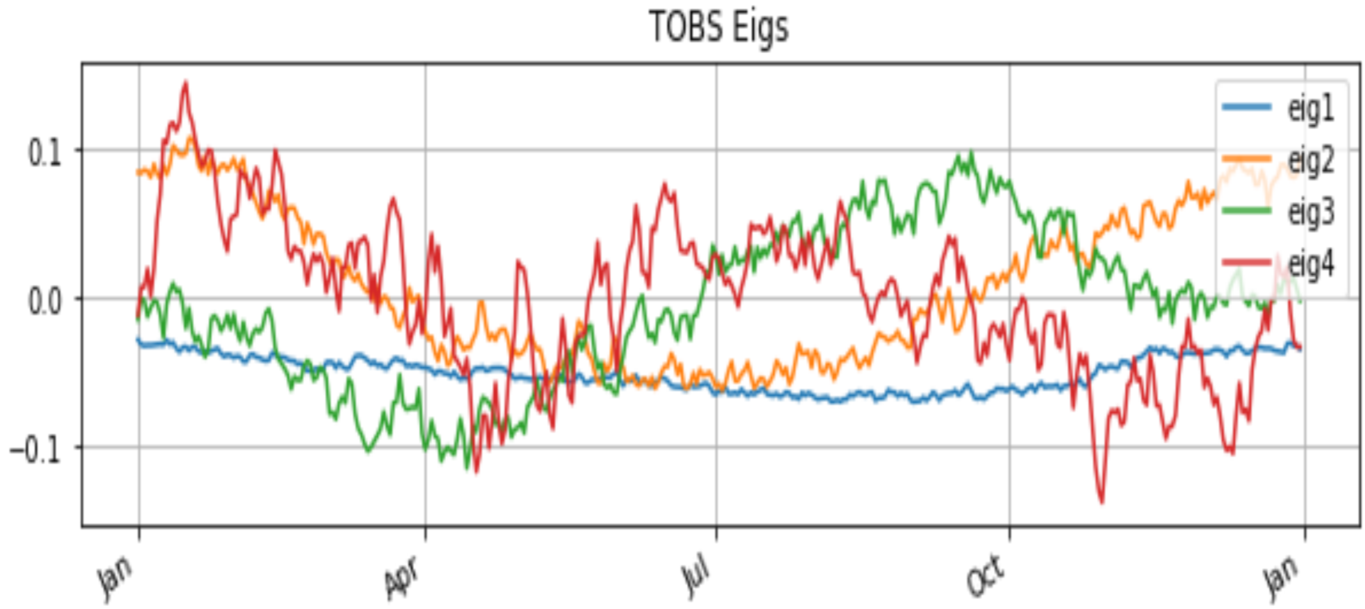


When we reorder the rows and columns of the matrix using one of the eigenvectors, the grouping of the stations becomes more evident. For example, consider the upper left corner of the first matrix (The upper right one). The stations at positions 0-25 are clearly strongly correlated with each other with some variations. Similar correlation can be seen in matrix 2 0-15.

Analysis on TOBS

Mean and Eigen vectors Analysis





Next we interpret the eigen-functions. The first eigen-function (eig1) is close to 0 and in opposite direction. The distribution looks very flat almost constant with some visible variation between July and october, the hotter months.

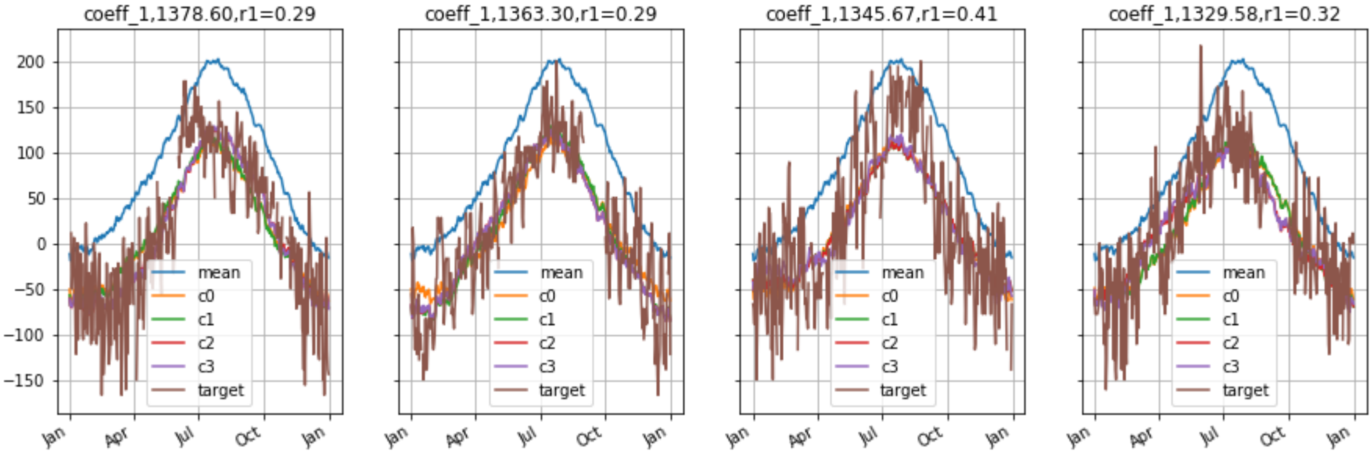
eig2 is like mean but close to 0 in reverse direction.The max reverse peak is towards July, hottest months.

eig3 and eig4 It oscilates between positive and negative values. In other words, they correspond to changing the distribution of the temperature observed , but they doesn't change the total (much).

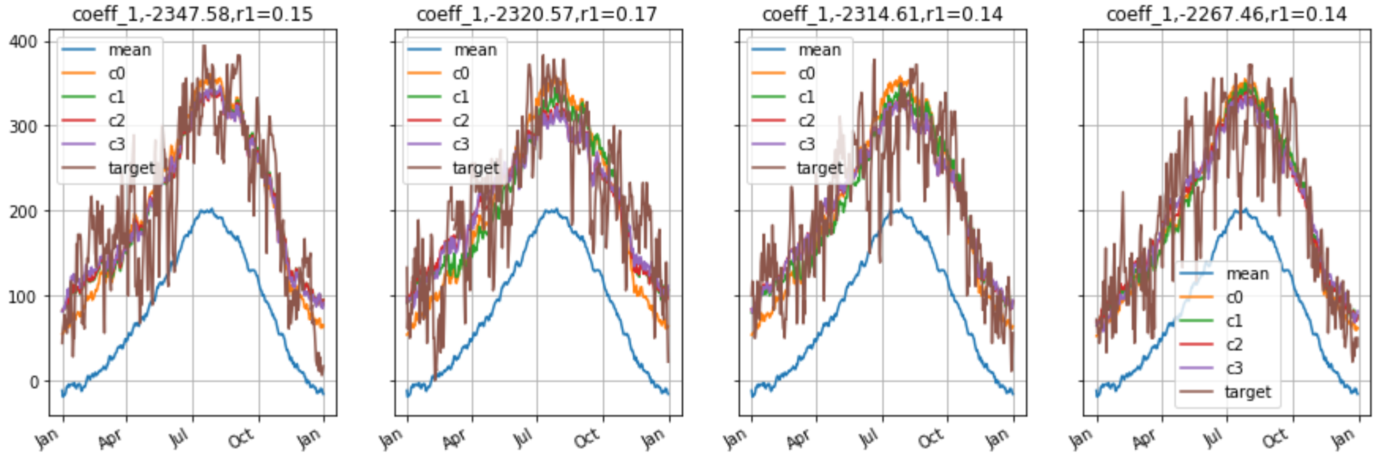
Examples of reconstruction

Coeff1

Coeff1: most positive



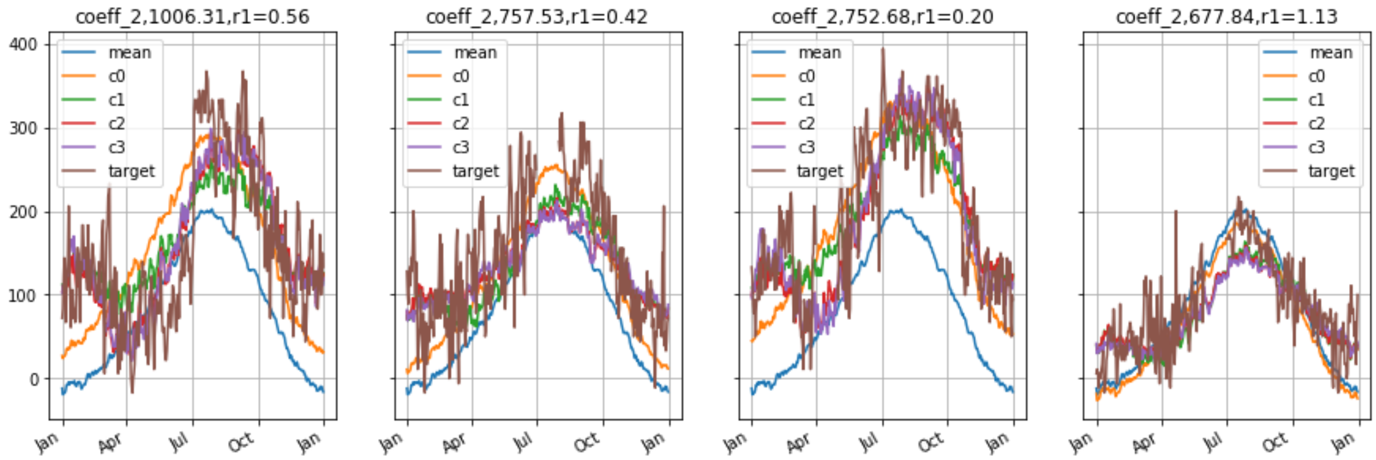
Coeff1: most negative



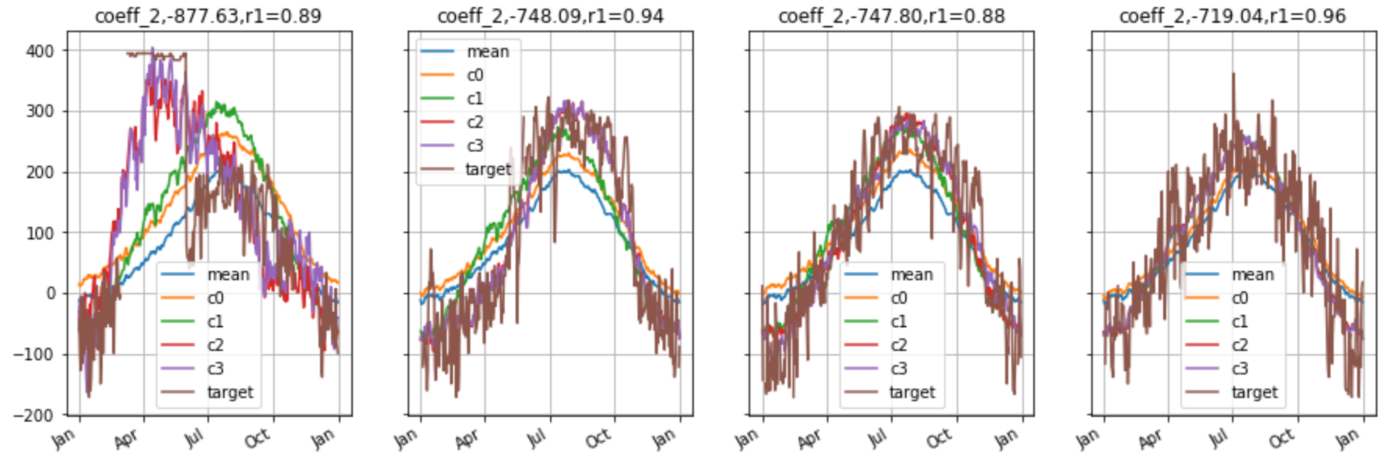
Large positive values of coeff1 correspond to more than average tobs. Low values correspond to less than average tobs.

Coeff2

Coeff2: most positive



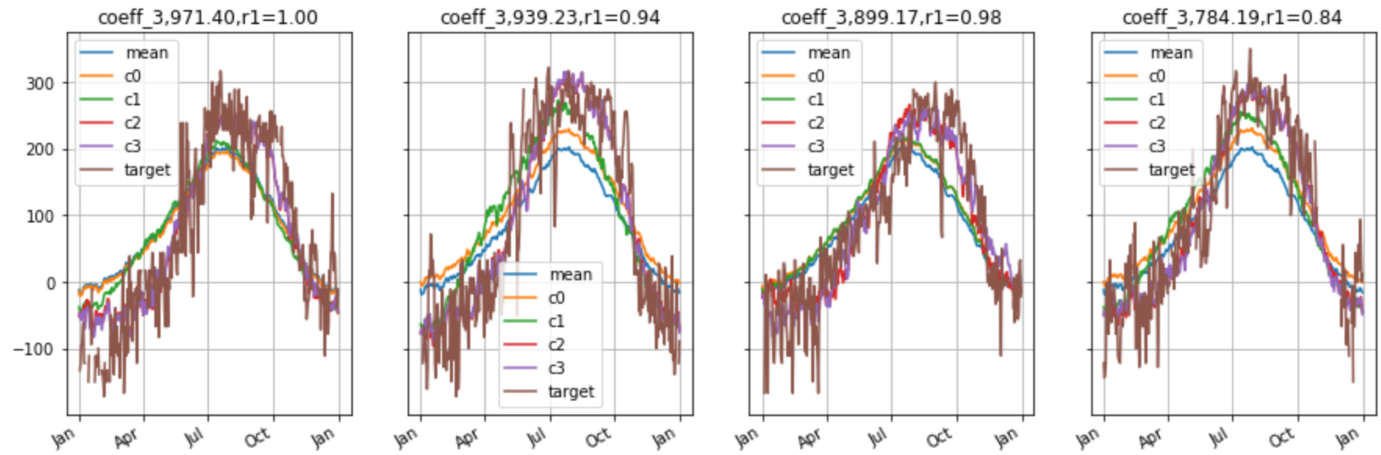
Coeff2: most negative



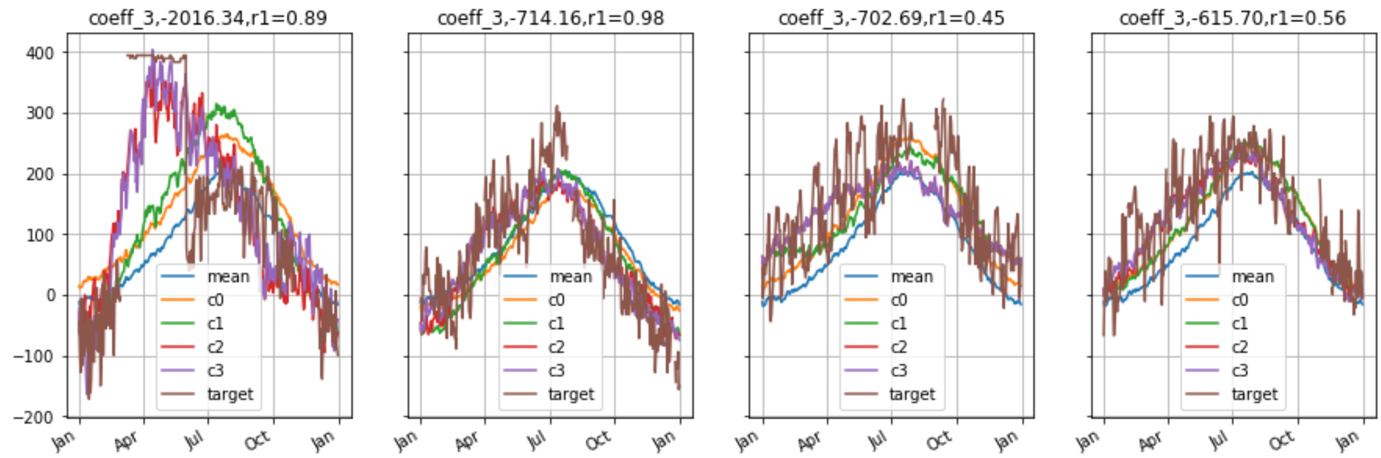
Similar to coeff_1

Coeff3

Coeff3: most positive



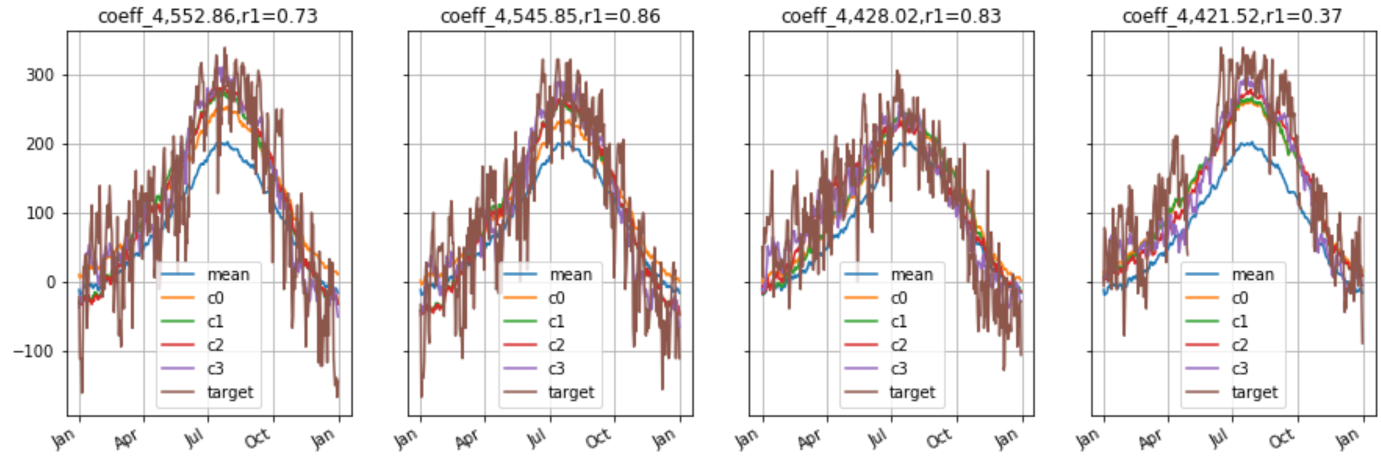
Coeff3: most negative



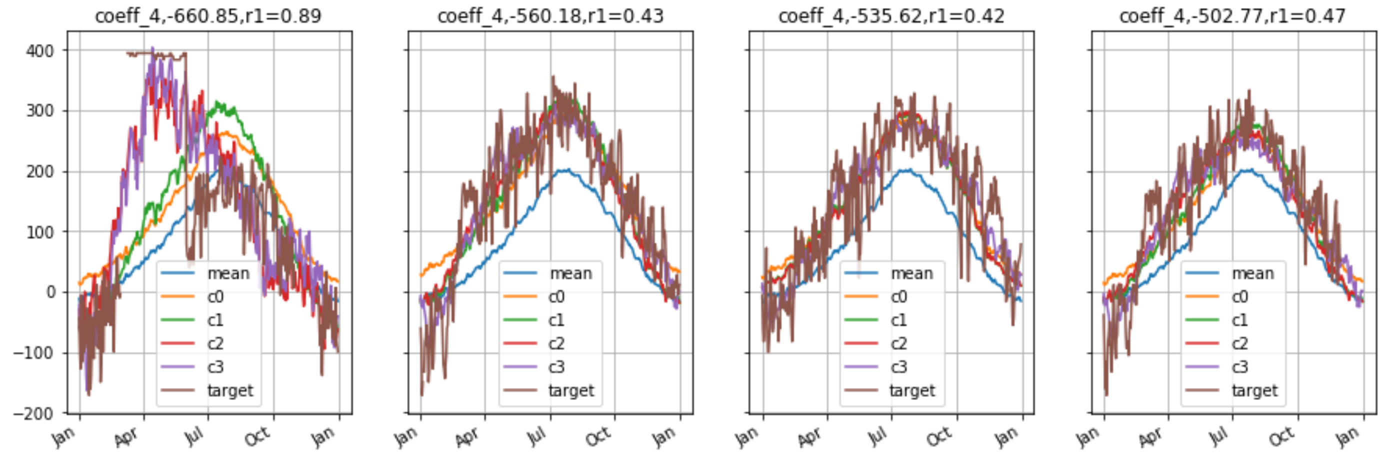
Postive values like coeff_1 and coeff_2. Negative shows a peak between april to july

Coeff4

Coeff4: most positive



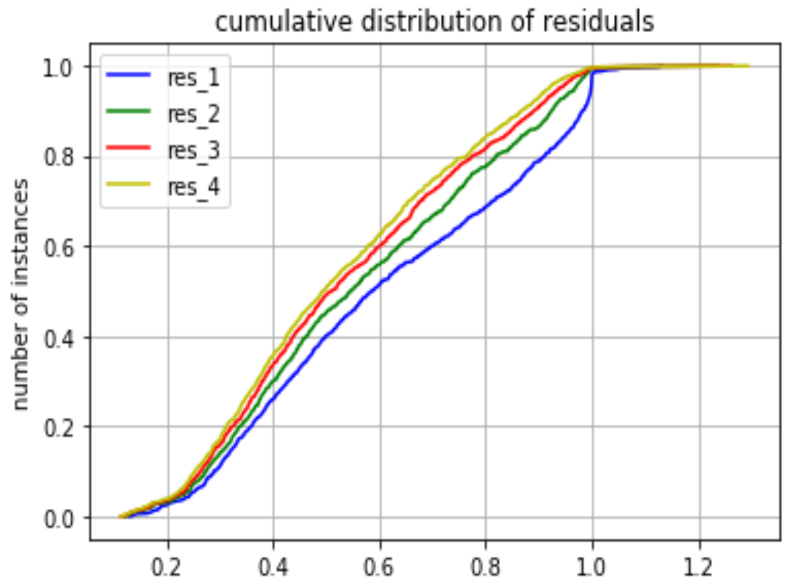
Coeff4: most negative



Similar as coeff_3

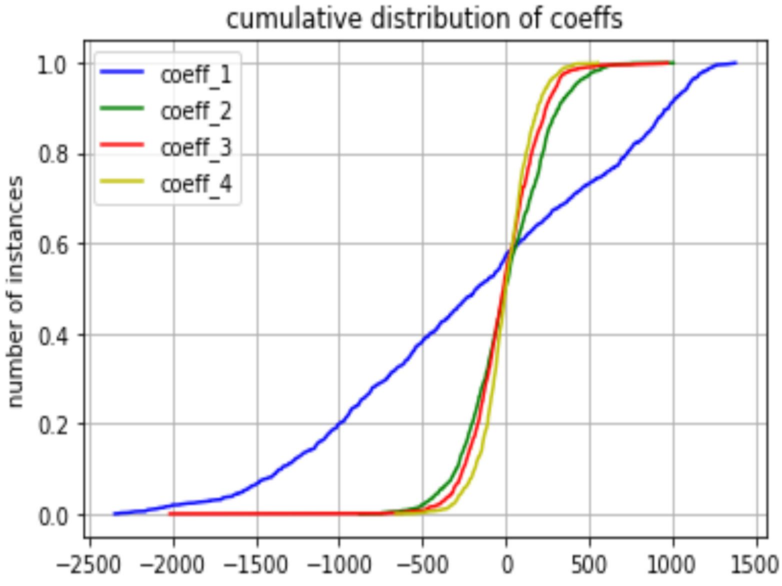
Cummulative Distribution Residuals and Coefficients

Residuals



The residuals for all four coefficient show how well the eigen vectors in the collection are shown

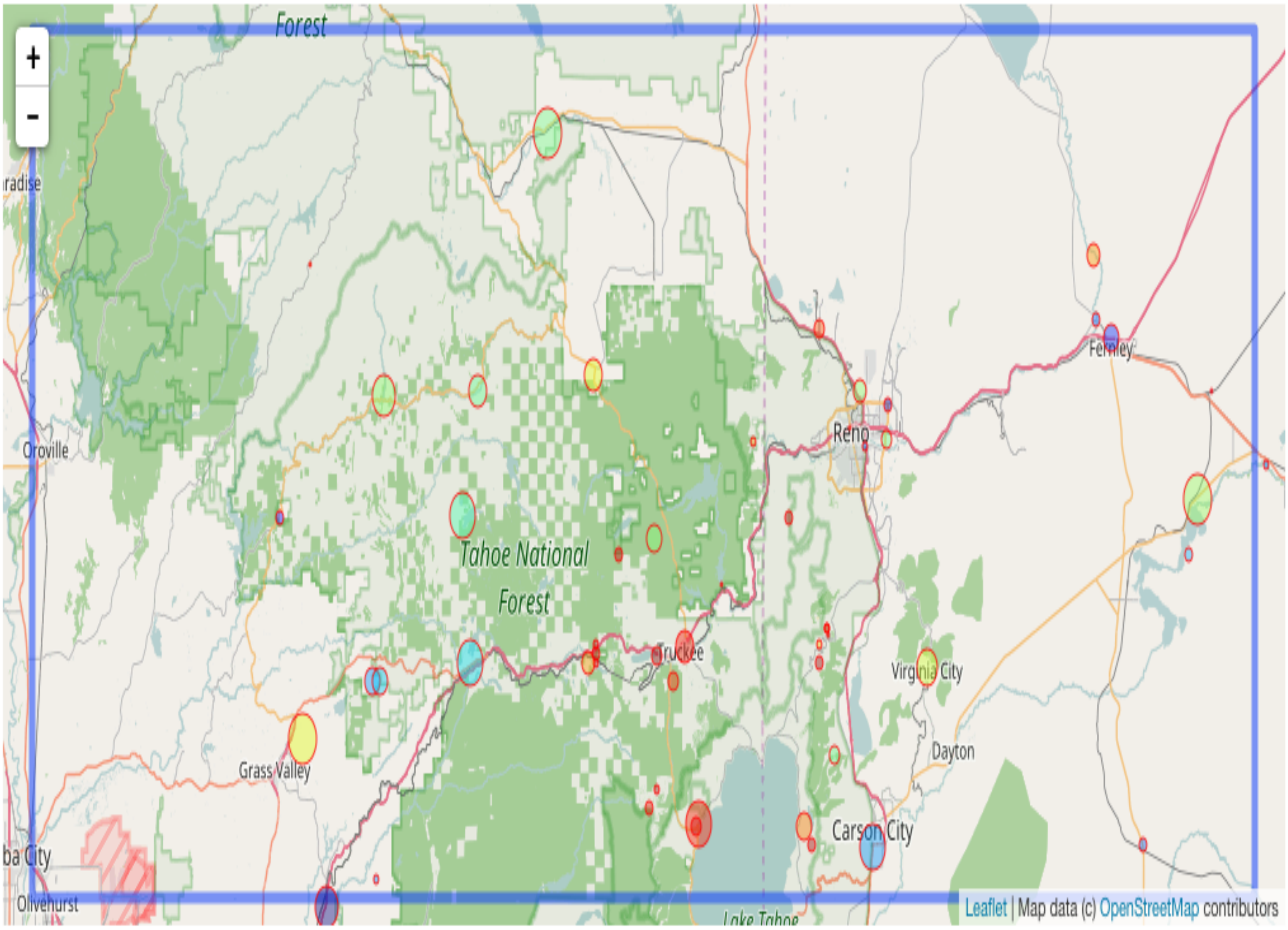
Coeff



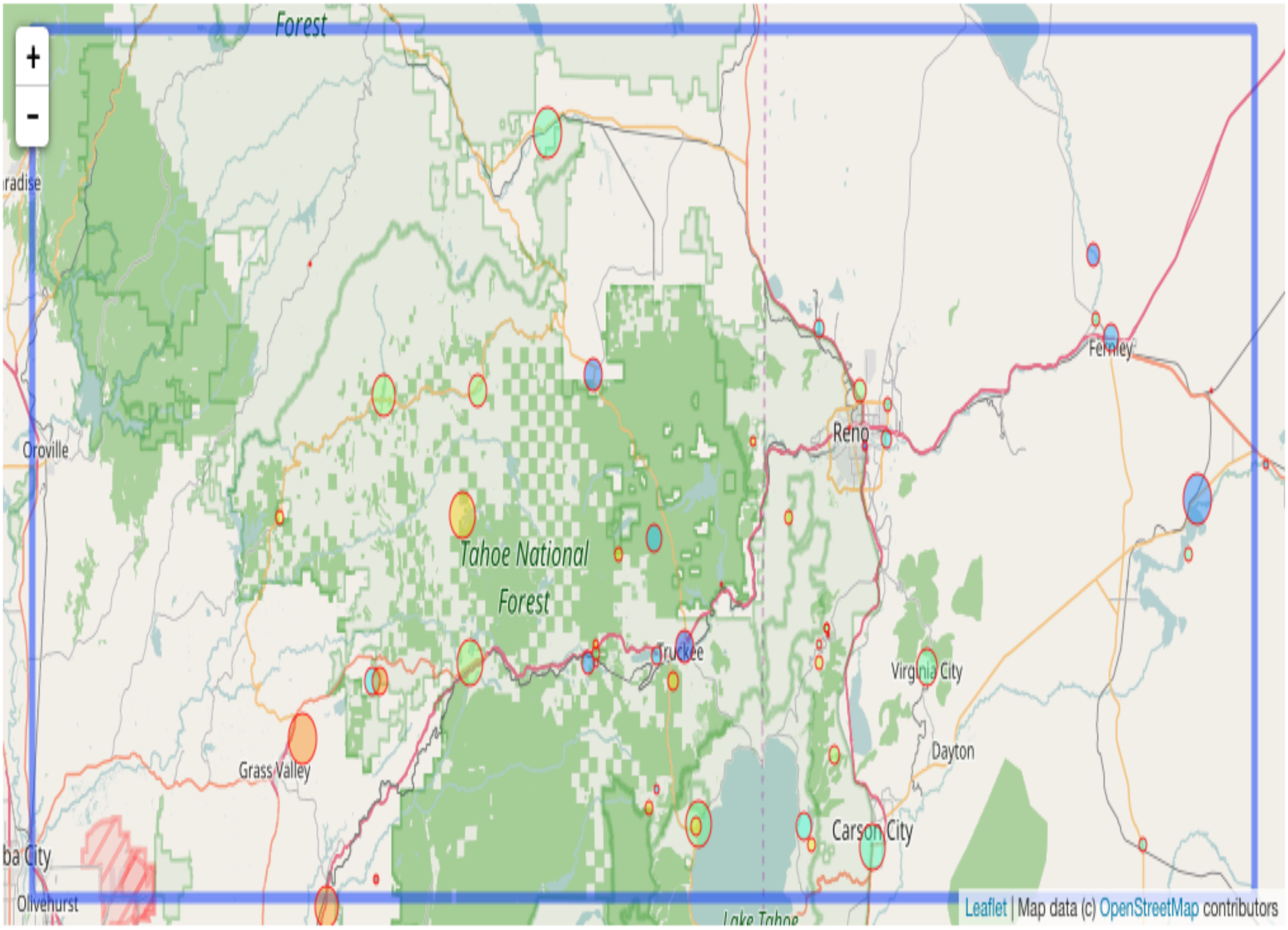
Coeff 1 shows most variation across stations; 2,3 and 4 shows similar correlation

Geographical distribution of top 4 coefficients.

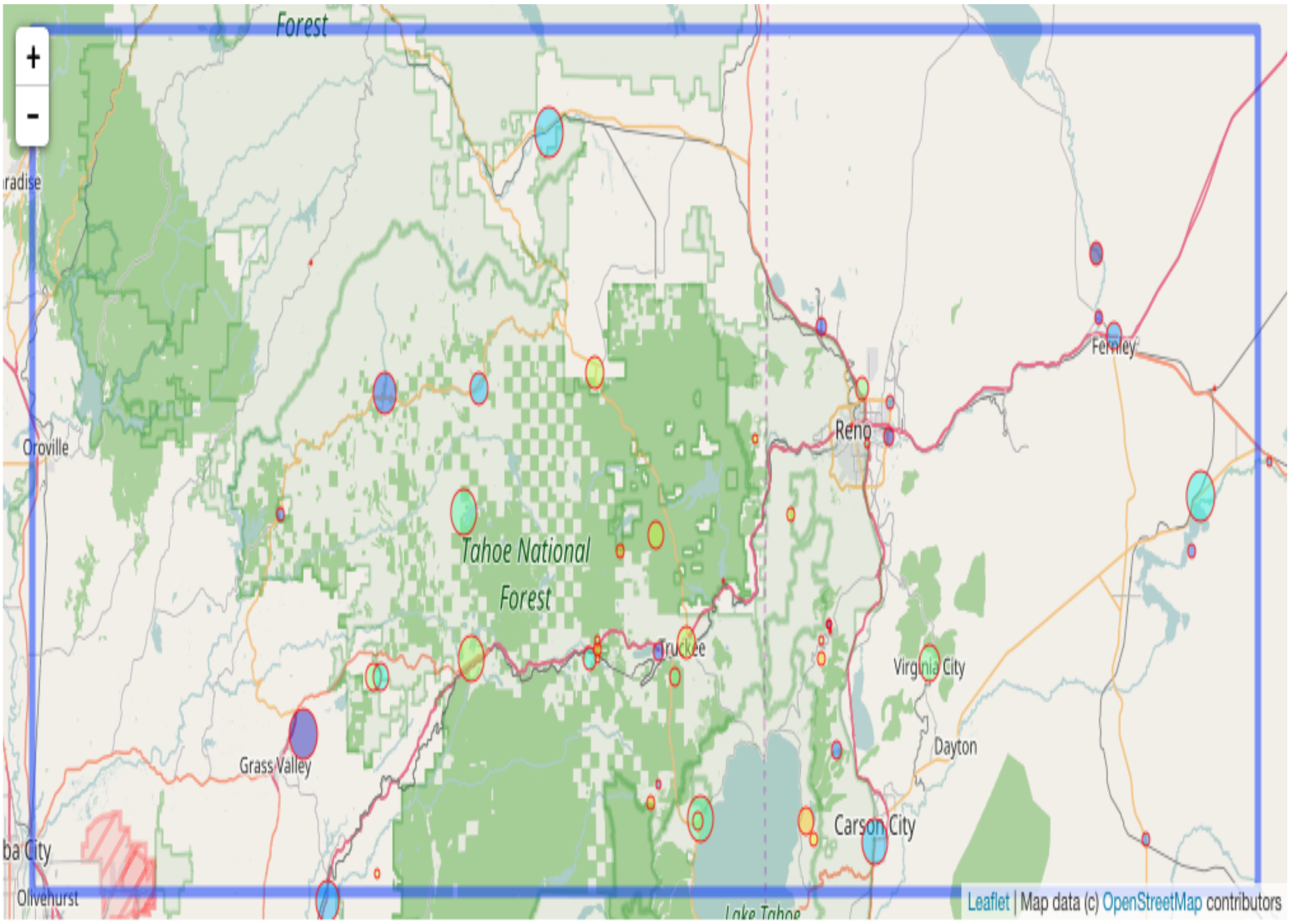
coeff1



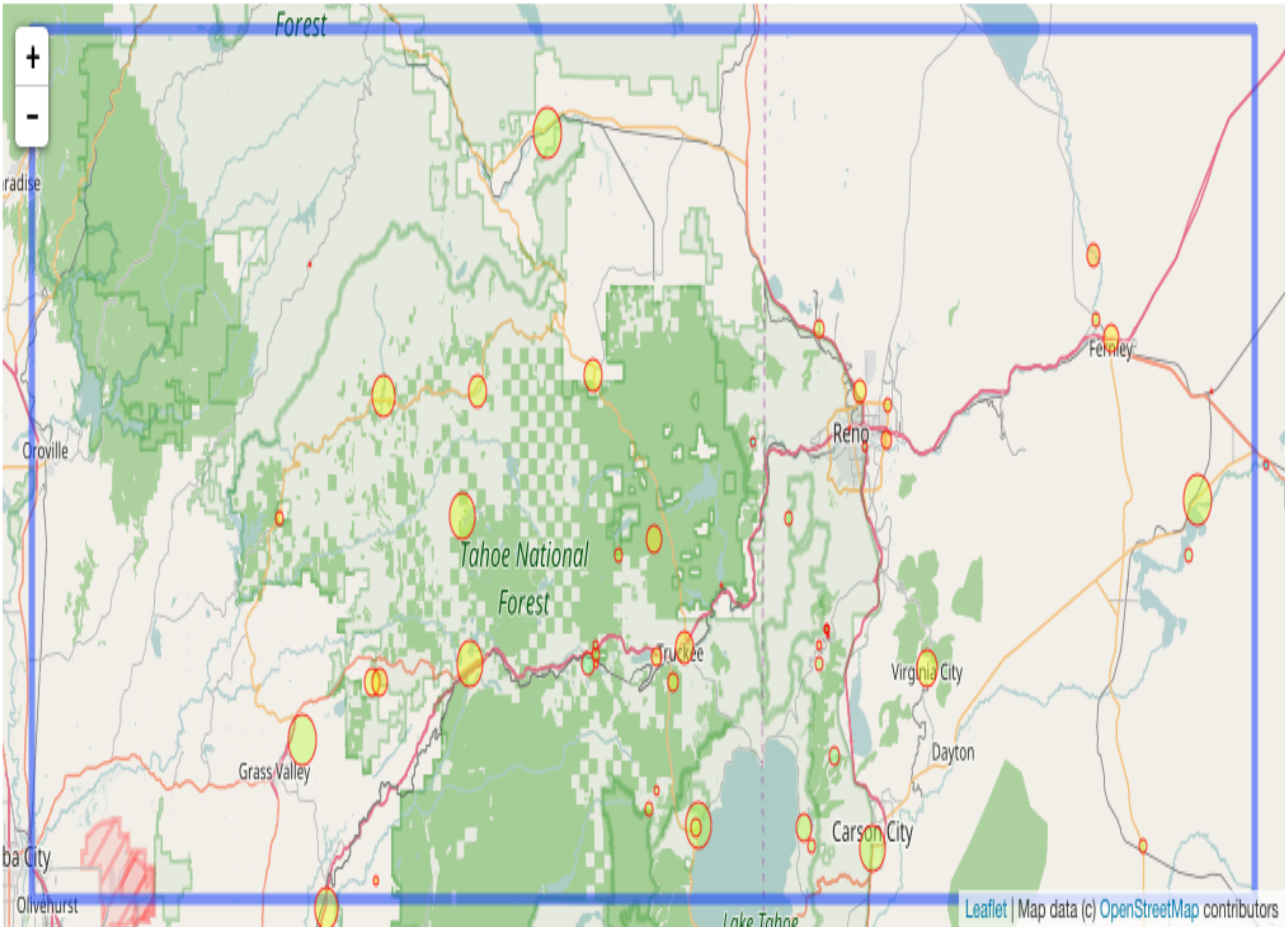
coeff2


Forest
+
-
radise
Oroville
Tahoe National
Forest
Reno
Fernley
Truckee
Virginia City
Dayton
Grass Valley
Carson City
ba City
Olivehurst
Lake Tahoe
Leaflet | Map data (c) OpenStreetMap contributors

coeff3


+
-
Forest
radise
Oroville
Tahoe National Forest
Reno
Fernley
Truckee
Virginia City
Grass Valley
Dayton
Carson City
ba City
Olivehurst
Lake Tahoe
Leaflet | Map data (c) OpenStreetMap contributors

coeff4



Most data came from stations located centrally close to each other but the contribution to coeff is more from far away distributed ones. It is interesting that looking at contribution to coeff from 1 to 4, the variation from stations decreases, i.e. for coeff_1 coeff contribution is very different across stations, most variation but for coeff_4 almost same.

Estimating the effect of the year vs the effect of the station

Computing for coeff_1

total RMS = 249.16627865

RMS removing mean-by-station= 163.56249498

RMS removing mean-by-year = 210.545871956

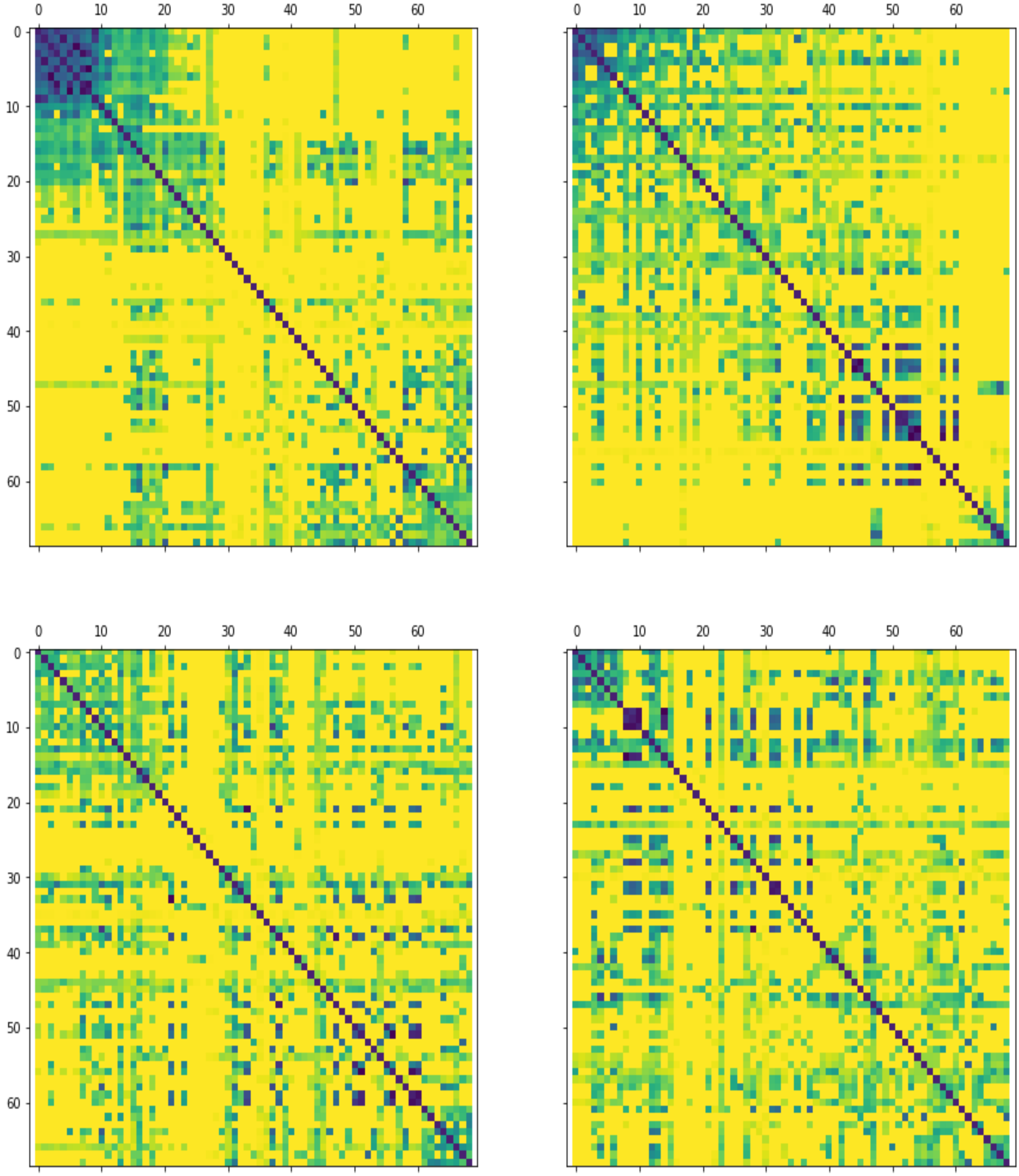
Station has more effect

Dependency Matrix to show correlation between stations

i.e. whether or not it snowed on the same day in different stations

The matrix above shows, for each pair of stations, the normalized log probability that the overlap in tobs days is random. We see that stations are not very correlated, only very few stations are highly correlated from 60 onwards

Reordering rows and columns to do more analysis



When we reorder the rows and columns of the matrix using one of the eigenvectors, the grouping of the stations becomes comparatively more evident. For example, consider the upper left corner of the first matrix (The upper left one). The stations at positions 0-10 are clearly strongly correlated with each other.

In [ ]: