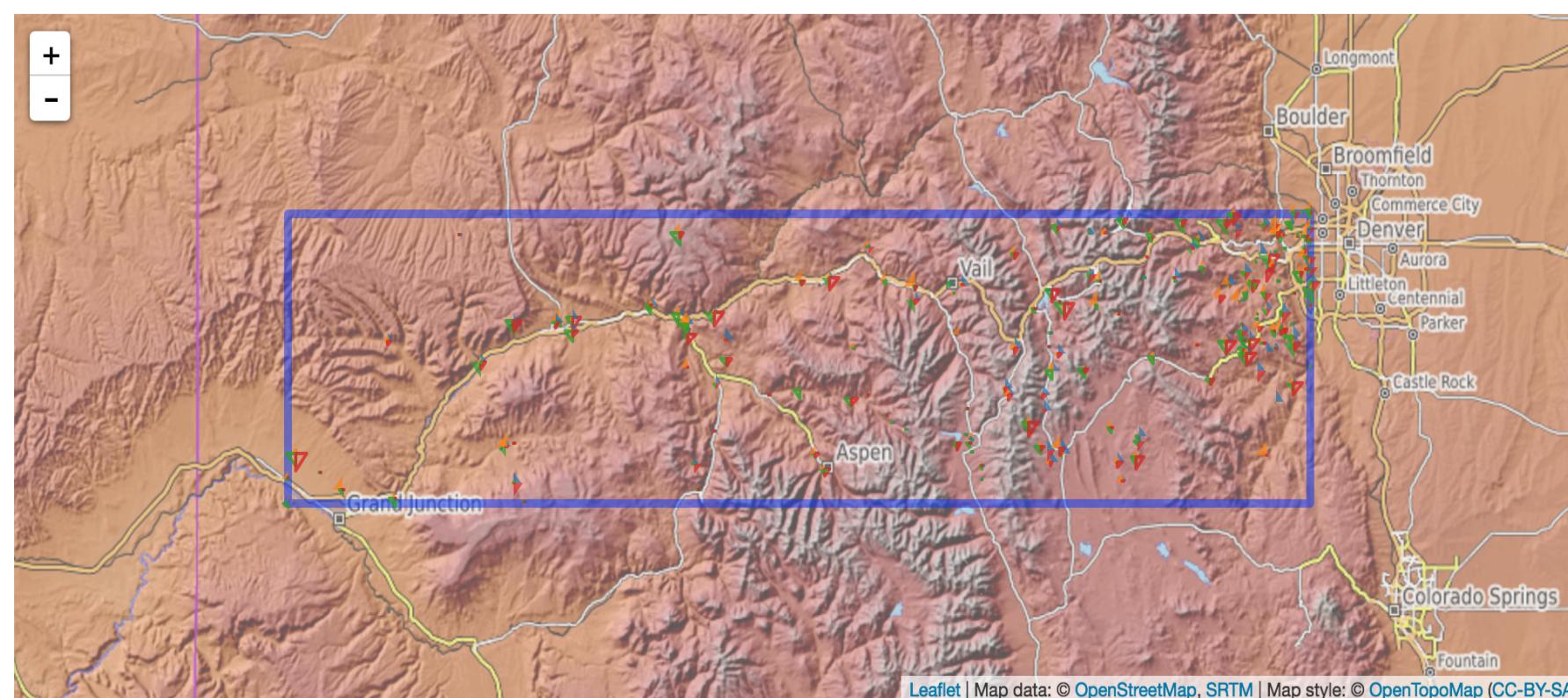


Colorado Weather Analysis



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

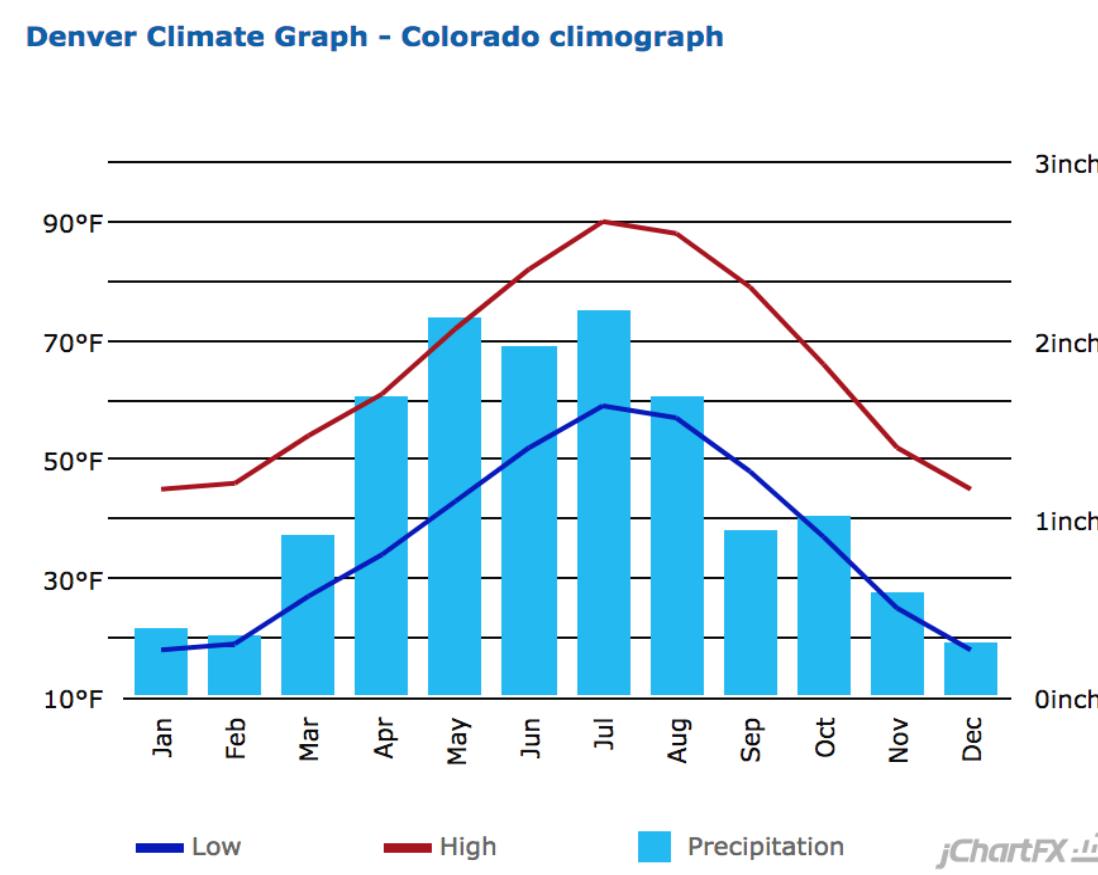
The data we will use here comes from [NOAA](https://www.ncdc.noaa.gov/) (<https://www.ncdc.noaa.gov/>). Specifically, it was downloaded from This [FTP site](ftp://ftp.ncdc.noaa.gov/pub/data/ghcn/daily/) (<ftp://ftp.ncdc.noaa.gov/pub/data/ghcn/daily/>).

We focused on six measurements:

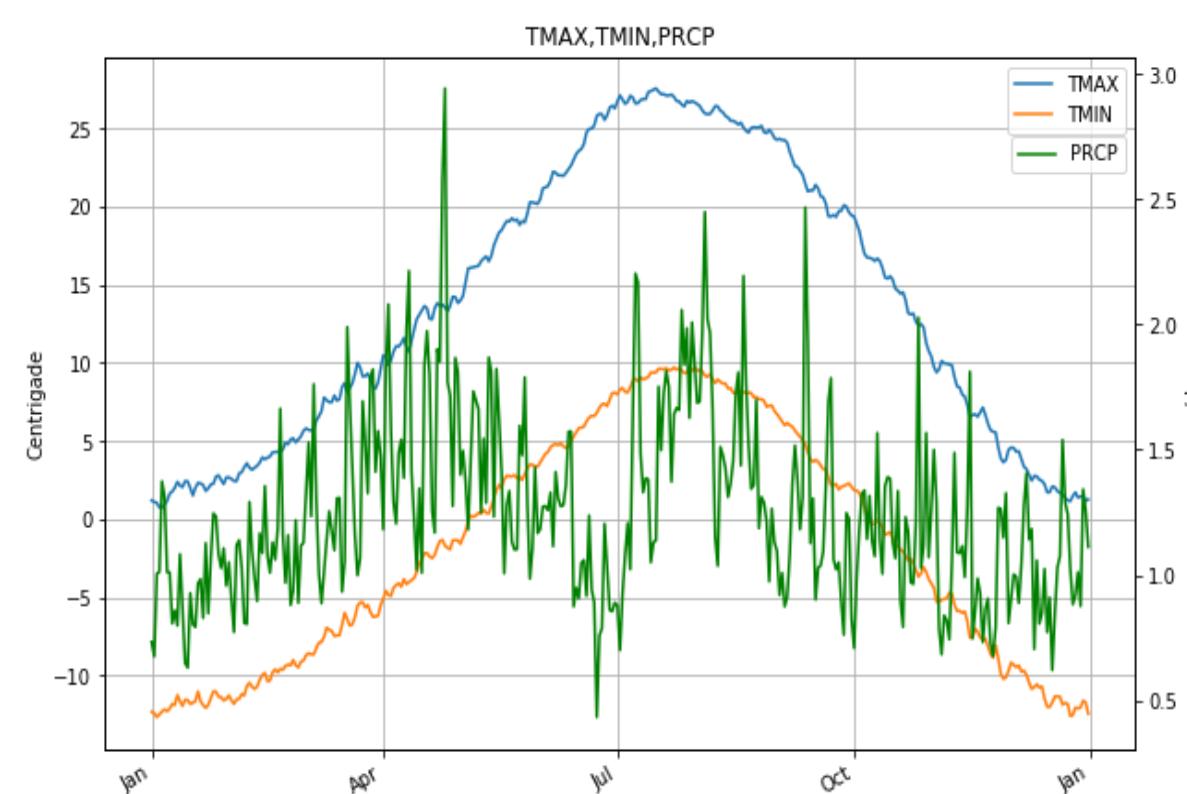
- **TMIN, TMAX:** the daily minimum and maximum temperature.
- **TOBS:** The average temperature for each day.
- **PRCP:** Daily Precipitation (in mm)
- **SNOW:** Daily snowfall (in mm)
- **SNWD:** The depth of accumulated snow.

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/colorado/united-states/3175) (<http://www.usclimatedata.com/climate/colorado/united-states/3175>) 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.



To compare the precipitation we need to translate millimeter/day to inches/month. According to our analysis the average rainfall is 1.25 mm/day which translates to about 1.5 Inches per month. According to US-Climate-Data the average rainfall is closer to 1.2 inch per month. However, there is clear agreement that average precipitation is close to a constant throughout the year

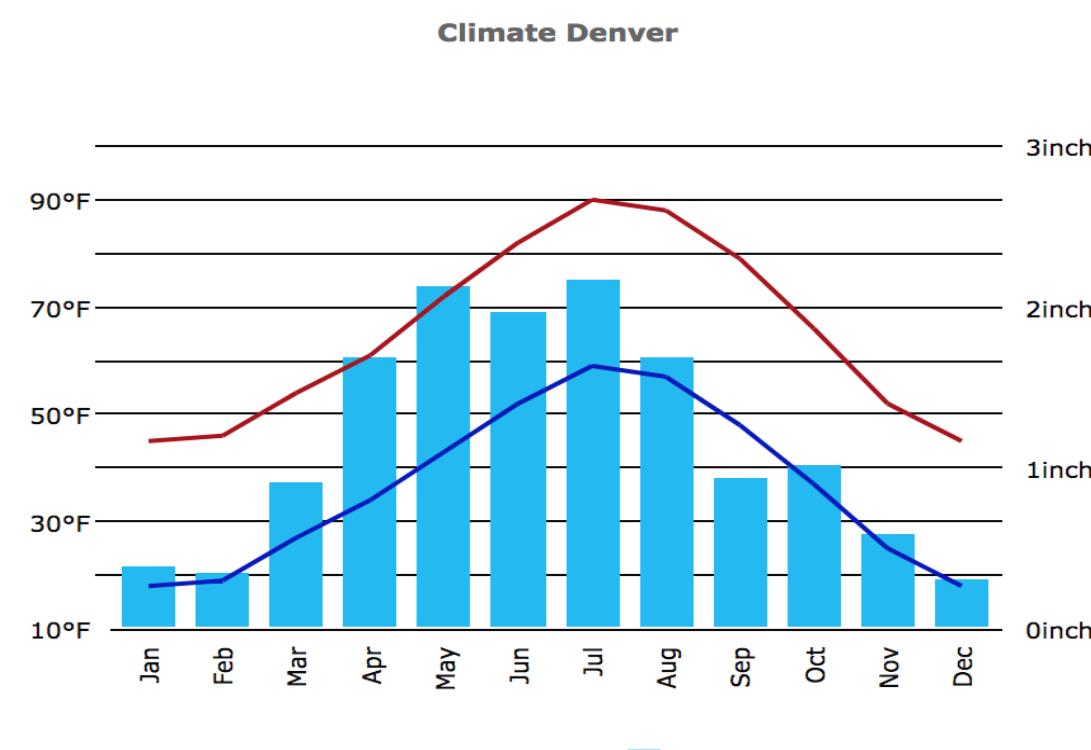
Regional Analysis

If we see the area mentioned in the map, there are 3 different type of regions that we can look at:-

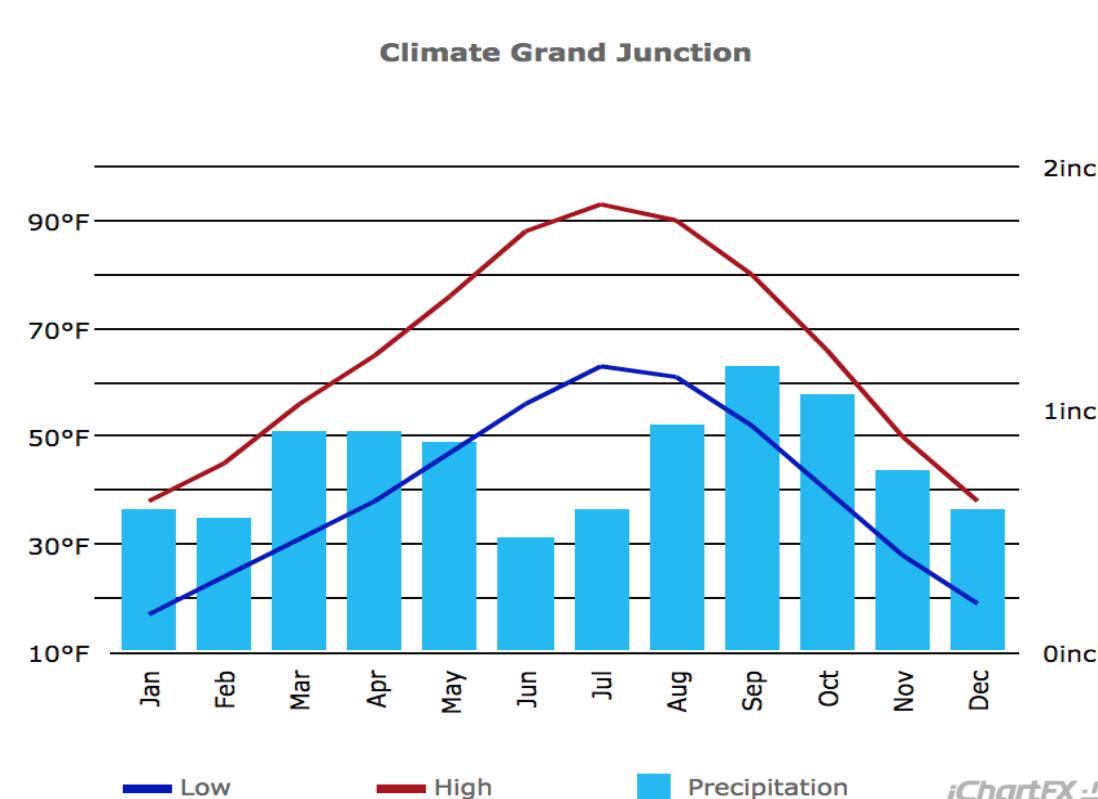
- Mountains (Aspen, Vail, Glenwood Springs)
- West of the Hills (Grand Junction)
- East of the Hills (Denver)

One important thing to note here is that the number of stations in Denver area is high as compared to other areas. So its highly likely that we get more accurate results for Denver.

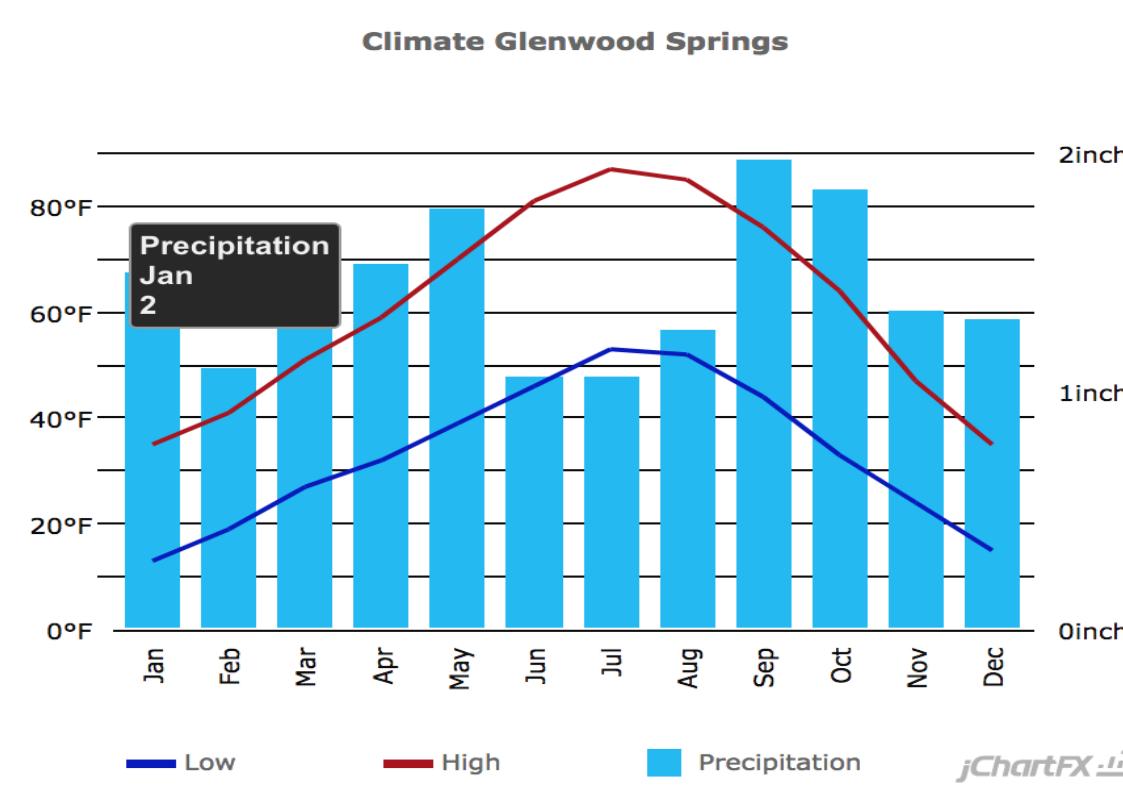
Below are the US Climate Data graphs for the above mentioned regions:-



source: U.S. Climate Data - Climate [Denver](#)



source: U.S. Climate Data - Climate [Grand Junction](#)



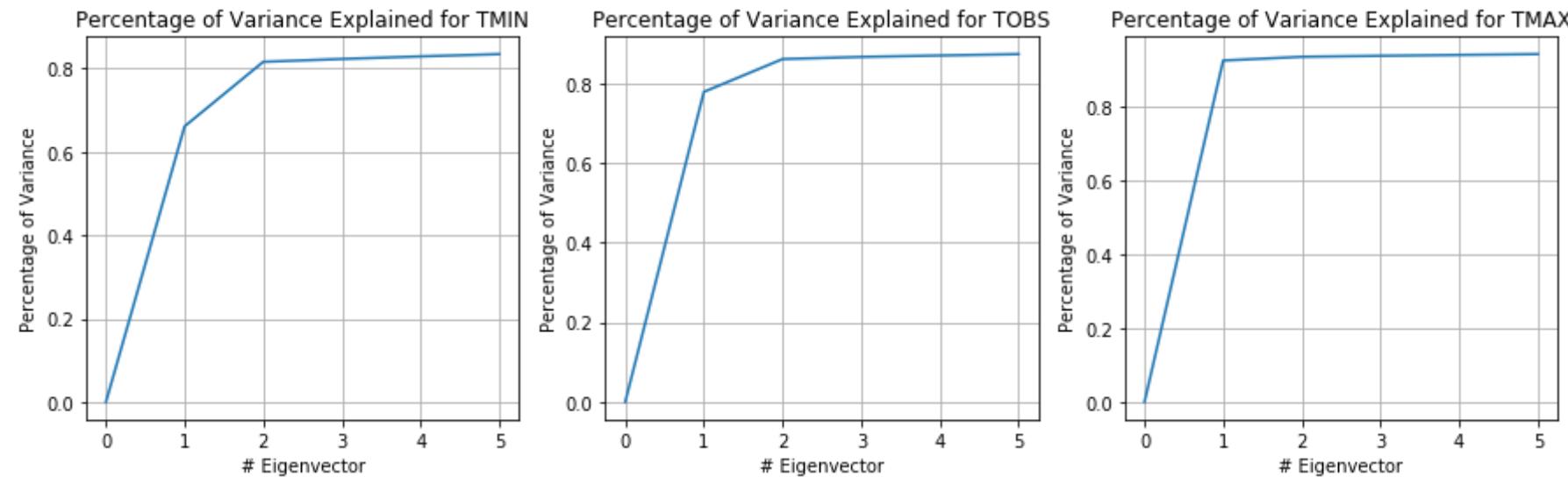
source: U.S. Climate Data - Climate [Glenwood Springs](#)

The average rainfall in Denver and Glenwood springs is more than Grand Junction area.

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.

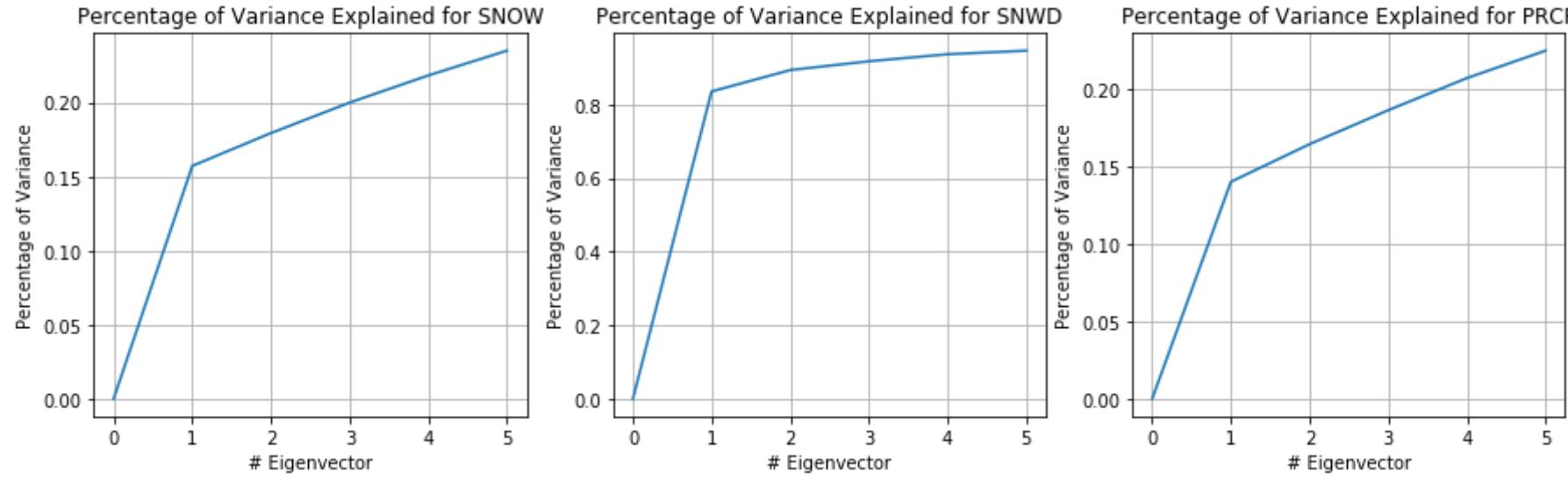
Percentage of variance explained.



We see that the top 2 eigen-vectors explain 80% of variance for TMIN and 85% for TOBS

For TMAX just the 1st eigen-vector explain 85% of variance.

We conclude that of the three, TMAX is best explained by the top 5 eigenvectors. This is especially true for the first eigen-vector which, by itself, explains 85% of the variance.



The top 5 eigenvectors explain 28% of the variance for PRCP and 25% for SNOW. Both are low values. On the other hand the top 5 eigenvectors explain 85% 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.

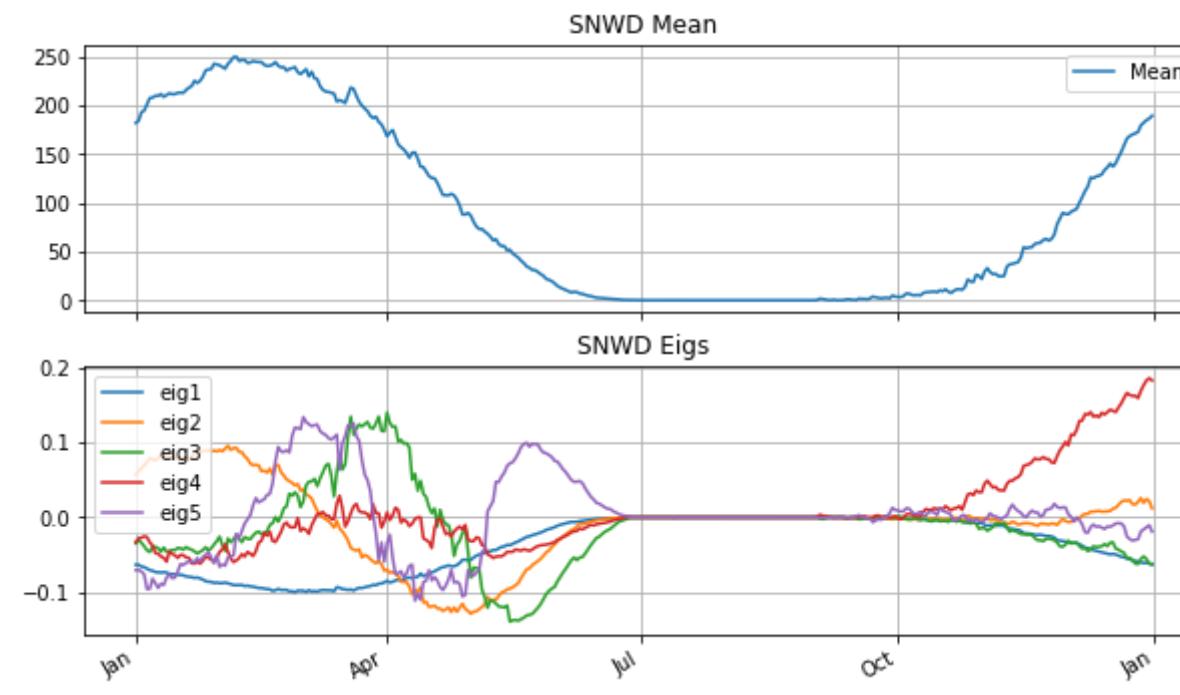
It makes sense that SNWD would be less noisy than SNOW. That is because SNWD is a decaying integral of SNOW and, as such, varies less between days and between the same date on different years.

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 eigen-vector alone explains about 80% of the variance.

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

We observe that the snow season is from October to the end of June, where 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 opposite of the mean function. The only difference is that the eigen-function is not as tall as the mean. 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,eig3,eig4 and eig5 are similar in the following way. They all oscillate 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:

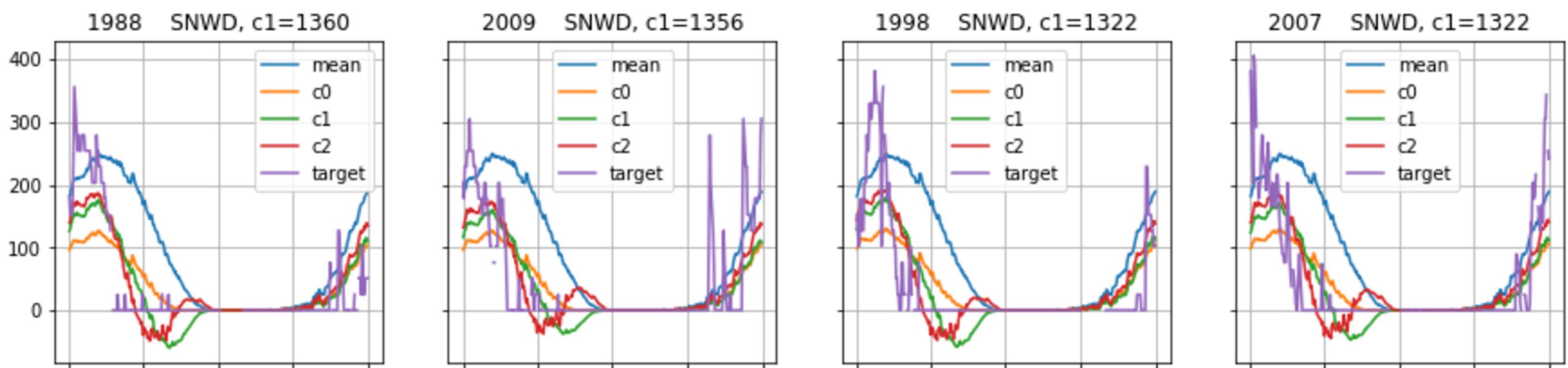
- **eig2:** more snow in jan - apr, less snow in april to july.
- **eig3:** less snow in jan,feb | more snow in mar,apr | less snow in June.
- **eig4:** less snow in jan | almost 0 snow in Mar,Apr | less snow in June.
- **eig5:** less snow in jan,feb | more snow in Mar | less snow in Apr,May | more snow in June.

From October to Dec, eig2 very well describes the mean. Eig1 and eig3 are sane from Oct to Dec. Eig4 and eig5 have close to 0 significance from Oct to Dec.

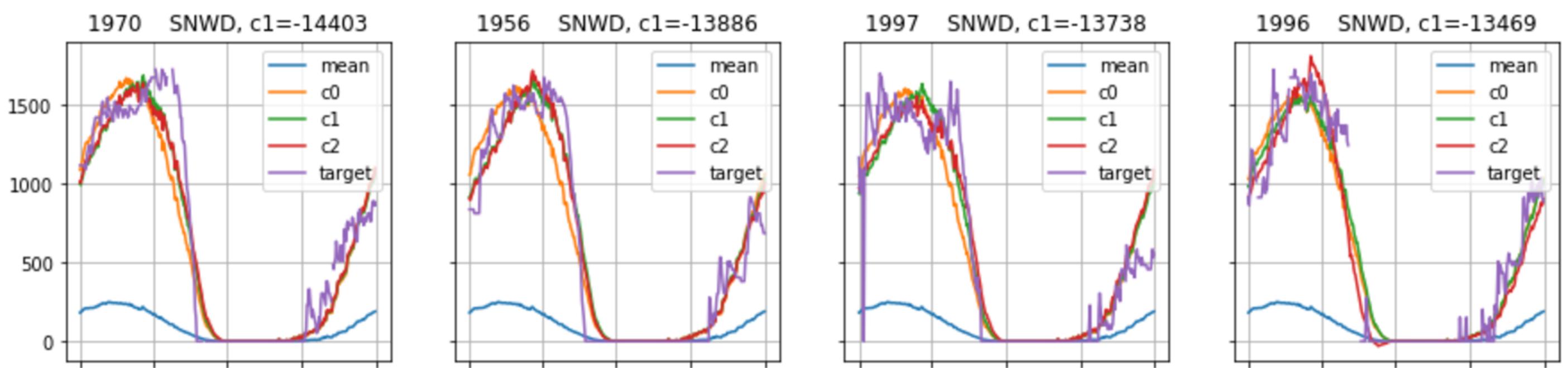
Examples of reconstructions

Coeff1

Coeff1: most positive



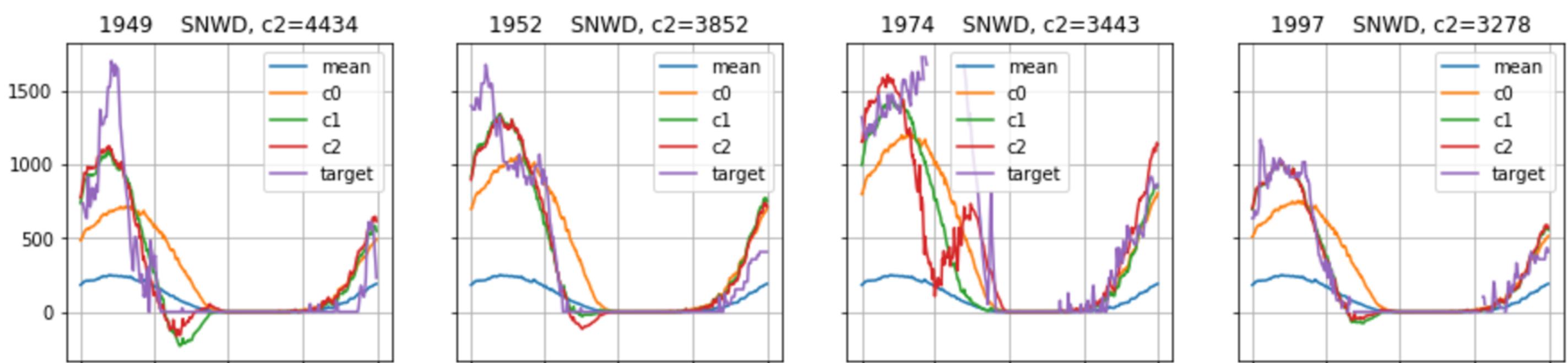
Coeff1: most negative



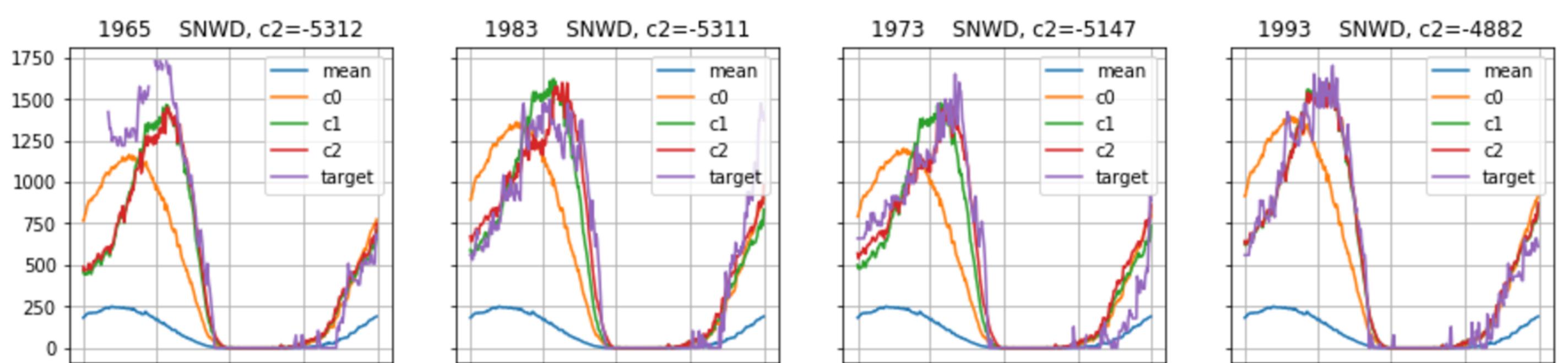
Large positive values of coeff1 correspond to less than average snow. Negative values correspond to more than average snow. Negative value also explain the more than average snow from Oct to Dec.

Coeff2

Coeff2: most positive



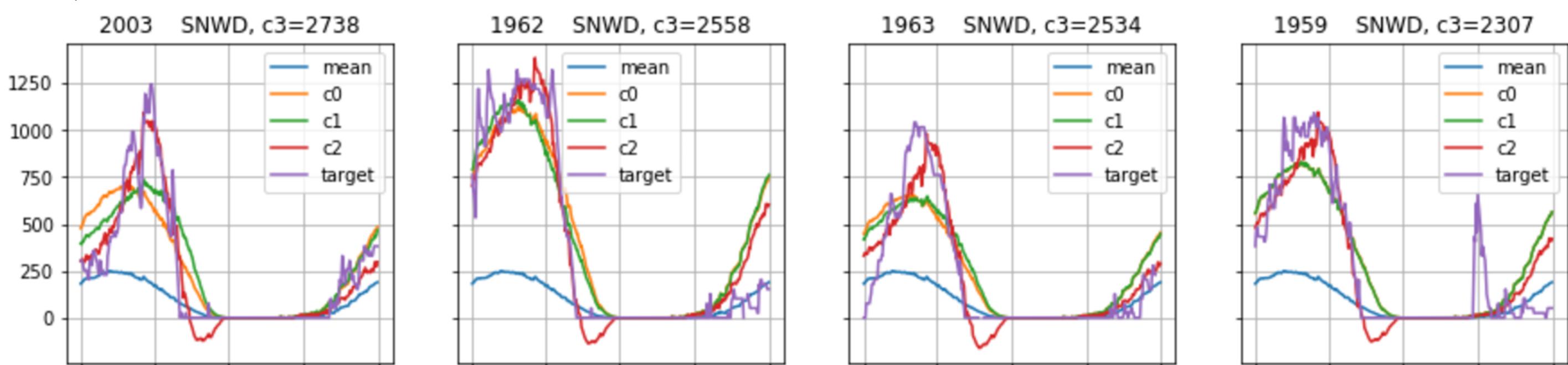
Coeff2: most negative



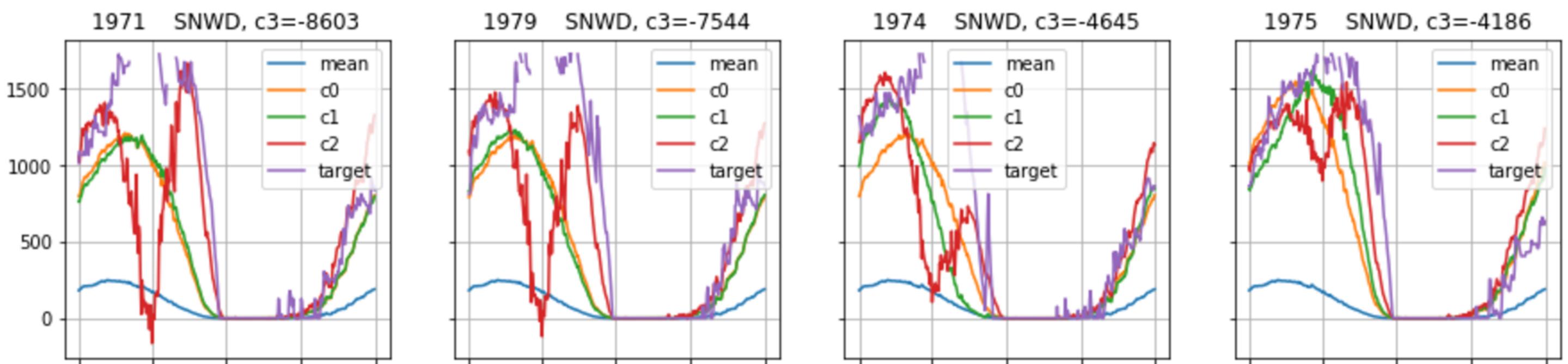
Large positive values of coeff2 correspond to an early snow season (most of the snowfall is before Mar). Negative values for coeff2 correspond to later snow season (most of the snow is after Feb).

Coeff3

Coeff3: most positive



Coeff3: most negative



Large positive values of coeff2 correspond to a snow season during feb to mar. Negative values of coeff3 correspond to the snow season later in the year from Oct to Dec. We can see 2 spikes in: once early in the season and another later during the end of snow season.

The variation in the timing of snow is mostly due to year-to-year variation

In the previous section we see the variation of Coeff1, which corresponds to the total amount of snow, with respect to location. We now estimate the relative importance of location-to-location variation relative to year-by-year variation.

These are measured using the fraction by which the variance is reduced when we subtract from each station/year entry the average-per-year or the average-per-station respectively. Here are the results:

coeff_1

```
total RMS =21881817.8096
RMS removing mean-by-station=3103406.07866, fraction explained=85.8
RMS removing mean-by-year =14144675.0957, fraction explained=35.36
```

coeff_2

```
total RMS =2475544.81348
RMS removing mean-by-station=850826.88867, fraction explained=65.63
RMS removing mean-by-year =1983660.22302, fraction explained=19.8
```

coeff_3

```
total MS =1193249.50671
MS removing mean-by-station=715035.27075, fraction explained=40.1
MS removing mean-by-year =912269.626739, fraction explained=23.5
```

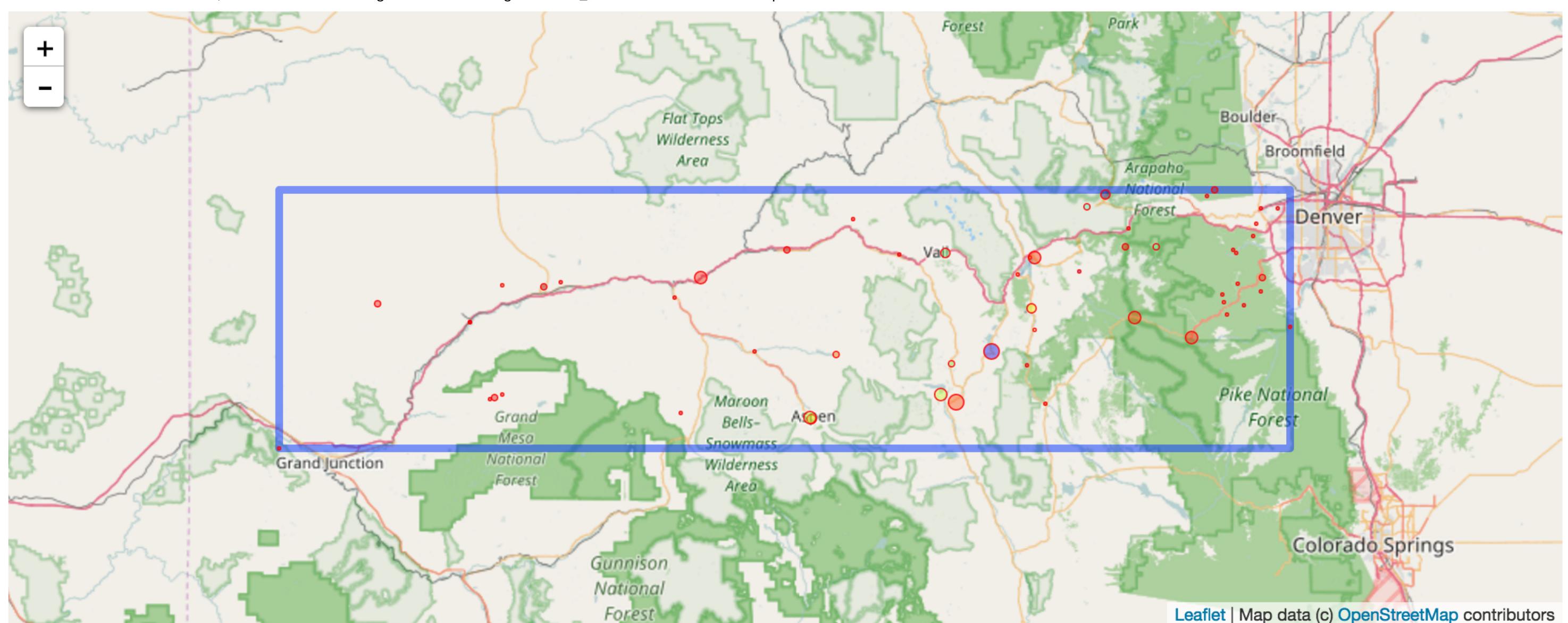
We see that the variation by station explains more than the variation by year. However this effect is weaker consider coeff_1, which has to do with the total snowfall, vs. coeff_2,3 which, as we saw above have to do with the timing of snowfall.

Analysing SNWD on map

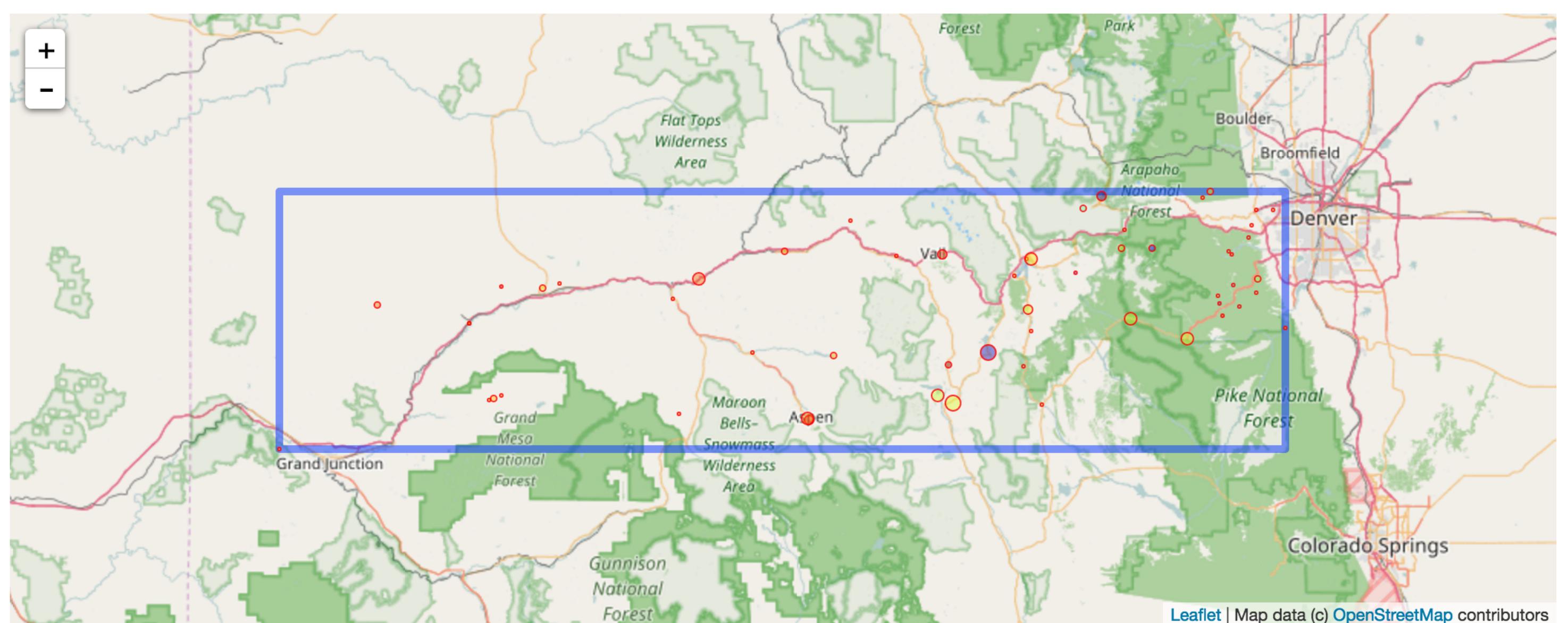
We display the 3 map images below, one for each coeff_1, coeff_2, coeff_3. The circle position represent the location of a station, it radius represents the number readings from that station and color represents the coeff values. Color red is the more negative values where the yellowish color is the positive values.

Coeff 1

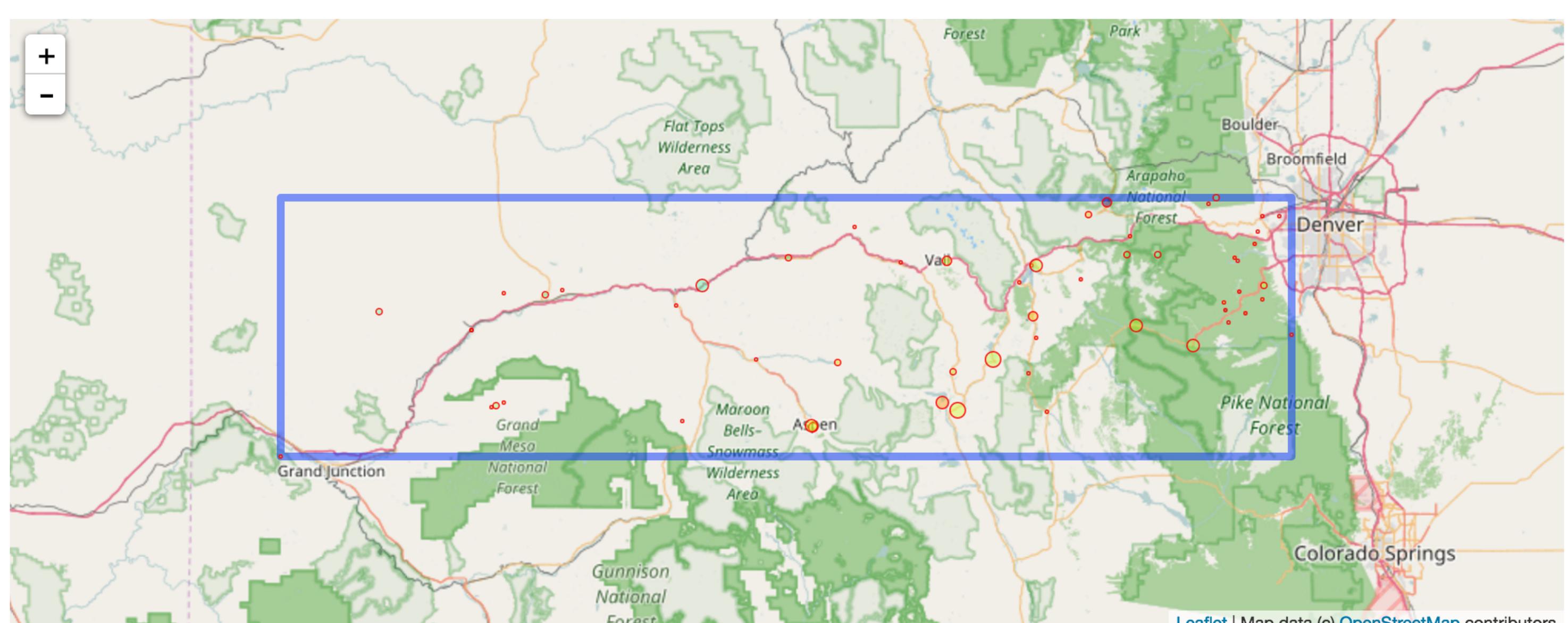
From the above reconstructions, we can see that the negative values are high for coeff_1 which is evident in the map.



Coeff 2



Coeff 3



Explanation of Correlation Matrix for top 4 eigen-vectors for SNWD

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 1st and 2nd matrix. The stations at positions 0-16 are clearly strongly correlated with each other. Even though there are some stations which are more related to each other than to the rest of this block.

