Weather Analysis of New Hampshire-Maine Region

CSE 255 HW5 Report

May 2017

1 Introduction

We analyze the historical weather patterns in the bordering area of New Hampshire and Maine state covering (or centered around) Manchester, Concord city in the New Hampshire state and Portland city in the Maine state. This is showed in Figure 1.We use the weather data from NOAA [1] specifically downloaded from [2].

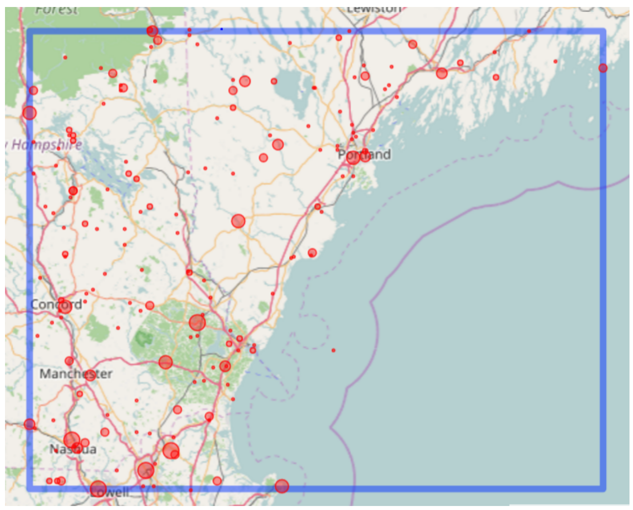


Figure 1: Measurements from stations' location on the map. Size of the circle is proportional to the number of years of data present.

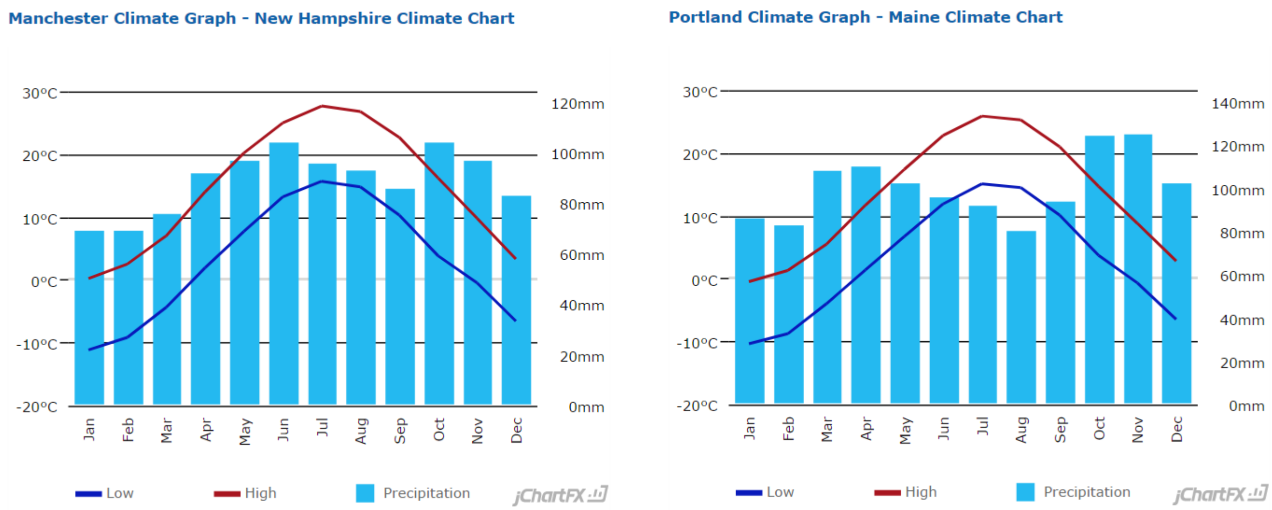
The NOAA dataset has daily measurements of Temperature, Precipitation and Snowfall. We would focus on 6 measurements which are as follows

- **TMIN:** Daily minimum temperature in ° Celsius
- **TMAX:** Daily maximum temperature in ° Celsius
- **TOBS:** Daily average temperature of the day in ° Celsius
- **PRCP:** Daily precipitation in mm.
- **SNOW:** Daily snowfall in mm.
- **SNWD:** Depth of accumulated snow in mm measured daily.

NOAA dataset had TMIN, TMAX, TOBS and PRCP in tenths of the actual value. The values were scaled by (1/10) to get the actual value.

2 Sanity Check

We extracted general statistics like mean and standard deviation and compared it with the US Climate Data [3]. The figure 2a and 2b shows the climate chart obtained from US Climate Data site for Manchester and Portland respectively.



(a) Manchester, New Hampshire Climate

(b) Portland, Maine Climate

Figure 2: Climate data for Manchester and Portland

We extract general statistics from our data which is plotted in Figure 3 and Figure 4.

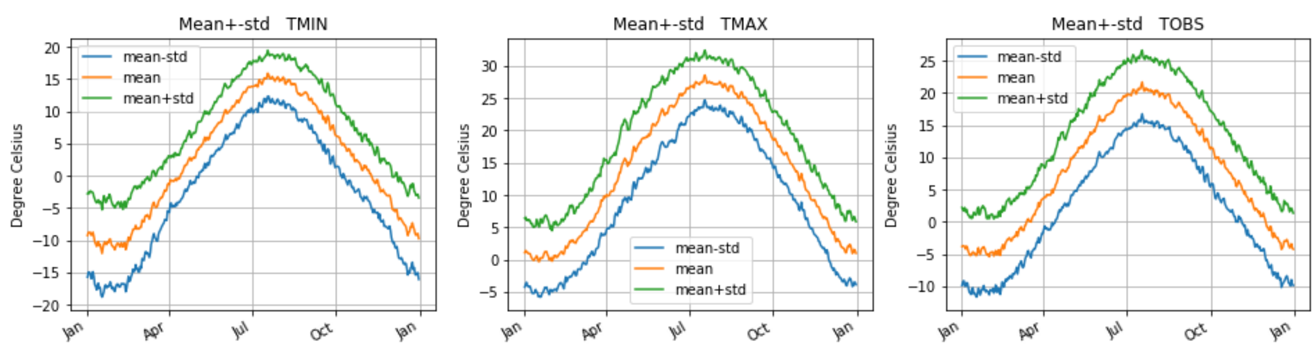


Figure 3: Mean +- std for TMIN, TMAX, TOBS

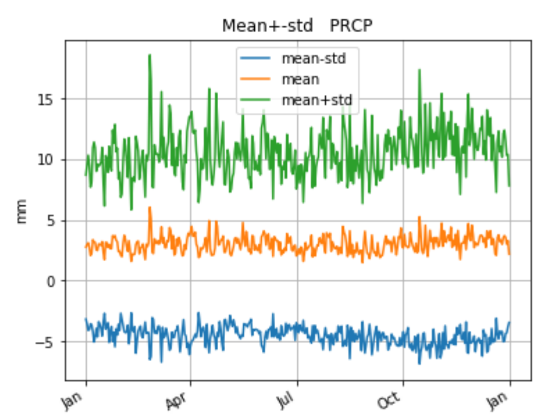


Figure 4: Mean +- std for PRCP

We observe that our minimum and maximum temperature as shown in Figure 3 agree with the US Climate Data (Figure 2). Min temperate in both the plots is around 15° C and maximum is around 25° C.

Average precipitation in US Climate Data (Figure 2) around 100 mm monthly. Mean daily precipitation in our data as shown in the Figure 4 is around 3 mm which translates to 90mm per month. Hence our basic statistics matches with the external source.

2.1 Yearly plots

Figure 5 and 6 shows the yearly plot for station USC00278612. It is hard to make any conclusion for TMIN, TMAX, TOBS, PRCP and SNOW. In Figure 5, we can observe that SNWD is very much correlated with the previous year. SNWD in year-2 is lagging year-1 but the overall trend remains the same.

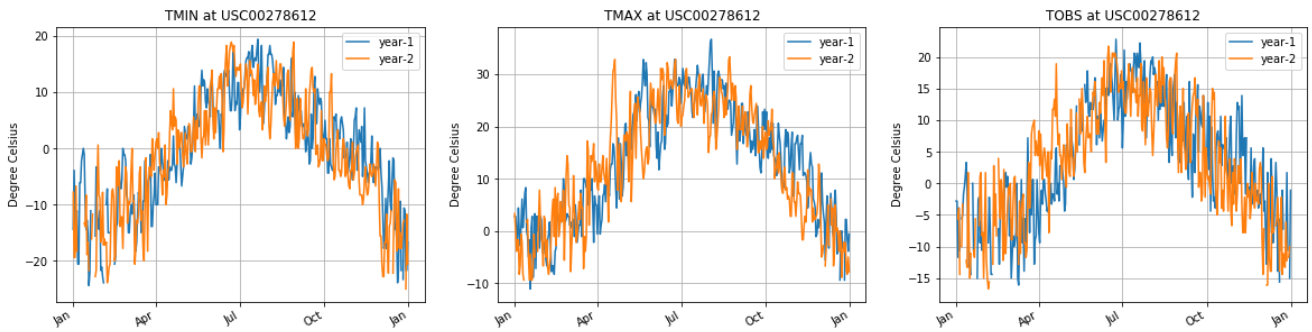


Figure 5: Yearly plots for TMIN, TMAX, TOBS

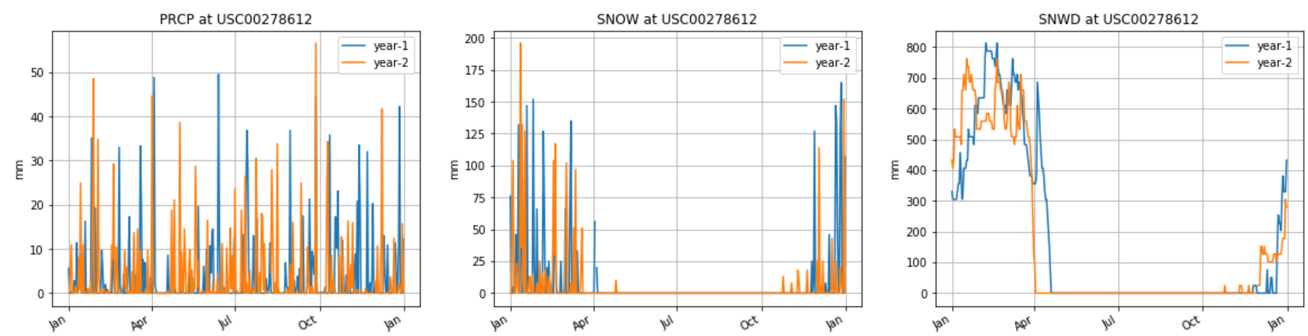


Figure 6: Yearly plots for PRCP, SNOW, SNWD

3 PCA

3.1 Percentage variance explained

Figure 7 and Figure 8 shows the percentage variance explained vs the number of eigenvectors. We observe that top 4 eigenvectors explains around 20% variance for TMIN, 15% for TMAX and 35% for TOBS. Note that TOBS first eigenvector explains around 30% variance.

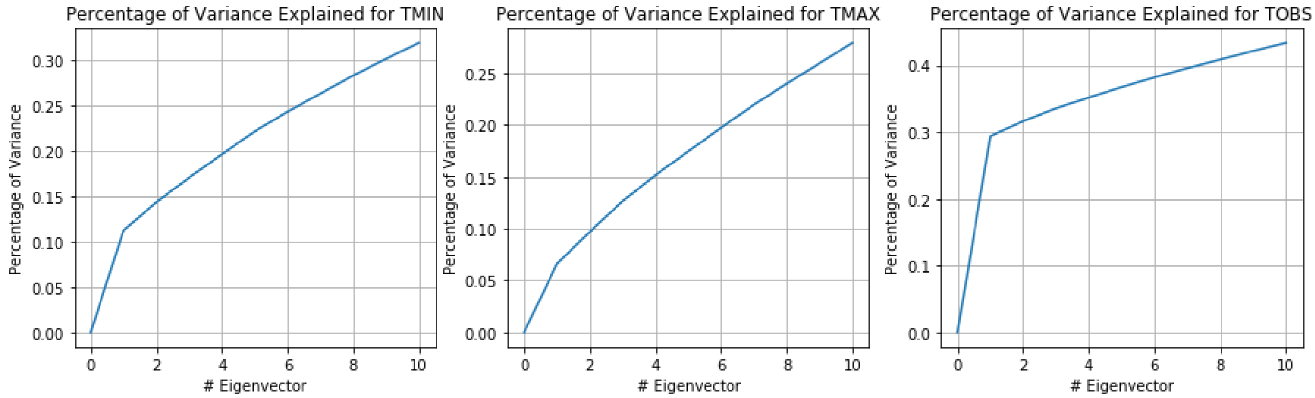


Figure 7: Percentage Variance Explained for TMIN, TMAX, TOBS

Top 4 eigenvectos for PRCP explains around 10% variance and the same is true for SNOW. For SNWD, top 3 eigenvectors explains 80% of the variance and top-4 around 85% variance. Since SNWD is accumulation of snow over the days, it is integral of SNOW and thus is less noisy. Top 4 eigenvector captures most of the variance and we will in the following section that we can make sense out of these eigenvectors.

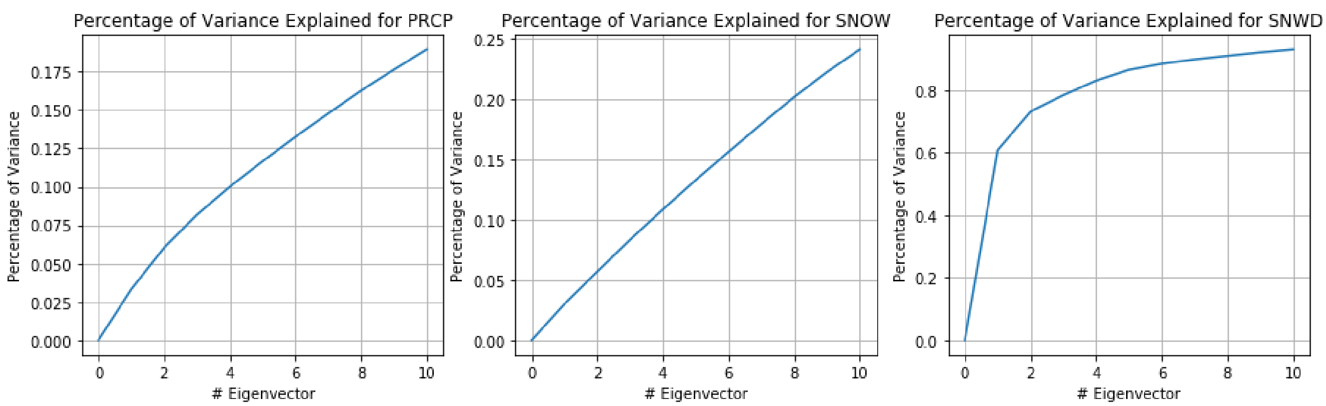


Figure 8: Percentage Variance Explained for PRCP, SNOW, SNWD

3.2 Eigenvectors

3.2.1 TMIN

Figure 9 shows the mean and eigenvectors for TMIN. We can observe that the first eigenvector explains how much a day's TMIN varies above the mean. Other eigenvectors are highly varying to make any observation. These mostly captures the change in TMIN between consecutive days or 2-3 days.

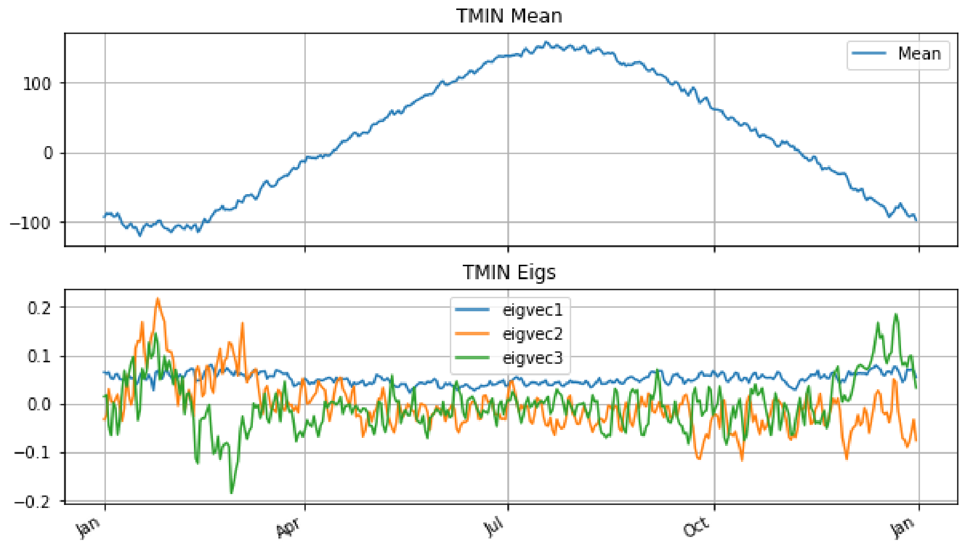


Figure 9: Mean and Eigenvectors for TMIN

3.2.2 TMAX

Figure 10 shows the mean and eigenvectors for TMAX. As in case of TMIN, the first eigenvector explains how much a day's TMAX varies above the mean. Other observations are same as TMIN.

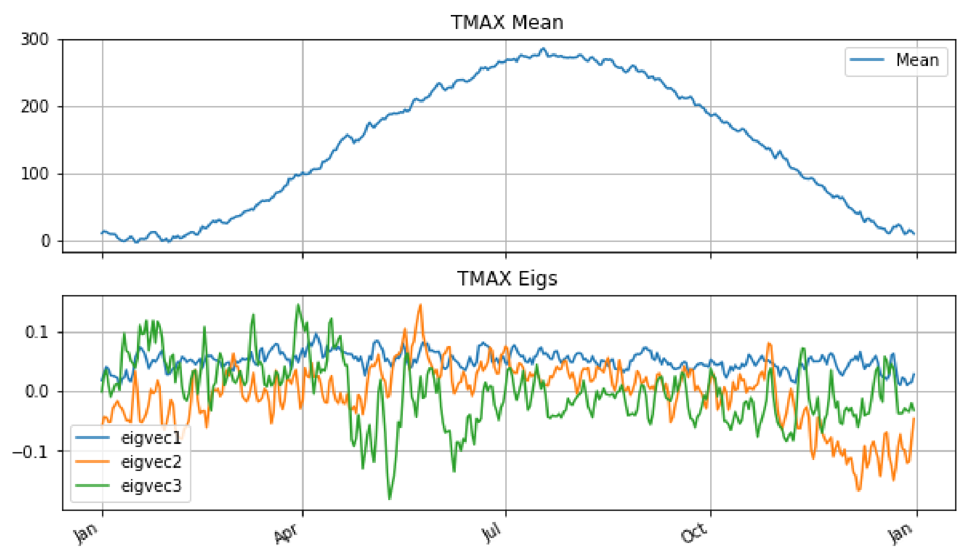


Figure 10: Mean and Eigenvectors for TMAX

3.2.3 TOBS

Figure 11 shows the mean and eigenvectors for TOBS. The first eigenvector explains how much a day's average temperature would be above the mean. Note that eigvec1 is slightly low during winter (Nov to Feb). The second eigenvector captures the summers as it is higher during April to Oct and lower in the rest of the year. It is hard to comment on eigvec-3 but it high during Jan to March signifying relation to the winters.

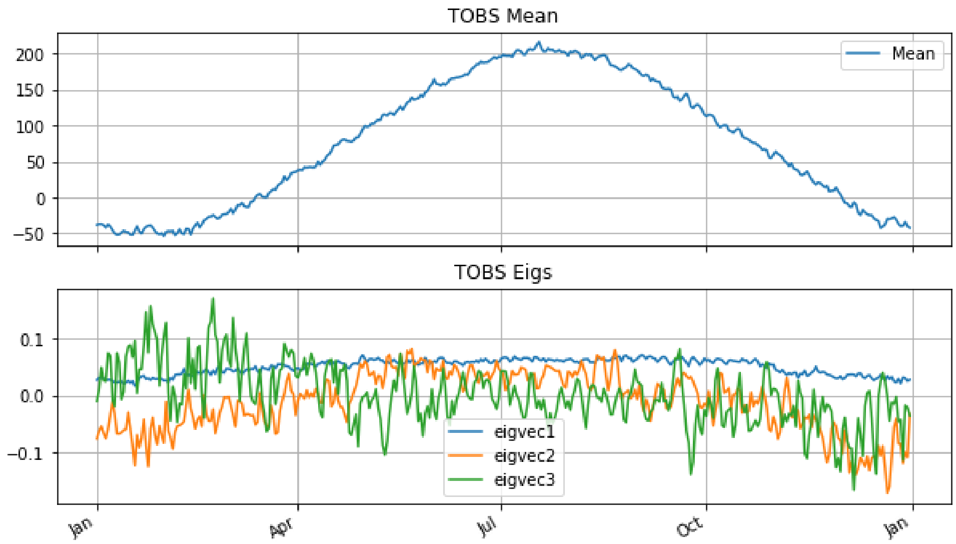


Figure 11: Mean and Eigenvectors for TOBS

3.2.4 PRCP

Figure 12 shows the mean and eigenvectors for PRCP. Mean and eigenvectors of PRCP are highly noisy and it is difficult to may inference from them. As observed from interactive plot, first eigenvector tries to fit the peaks of the PRCP.

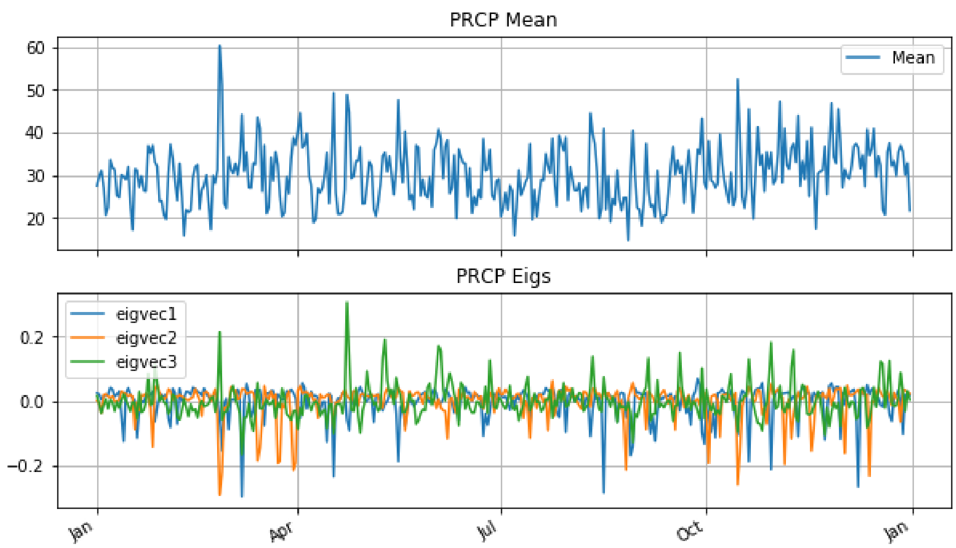


Figure 12: Mean and Eigenvectors for PRCP

3.2.5 SNOW

Figure 13 shows the mean and eigenvectors for SNOW. Obviously there is no snow during summers and hence mean is zero for summers. Eigenvectors are too noisy to make any concrete inference but they try to fit the local treads on the SNOW.

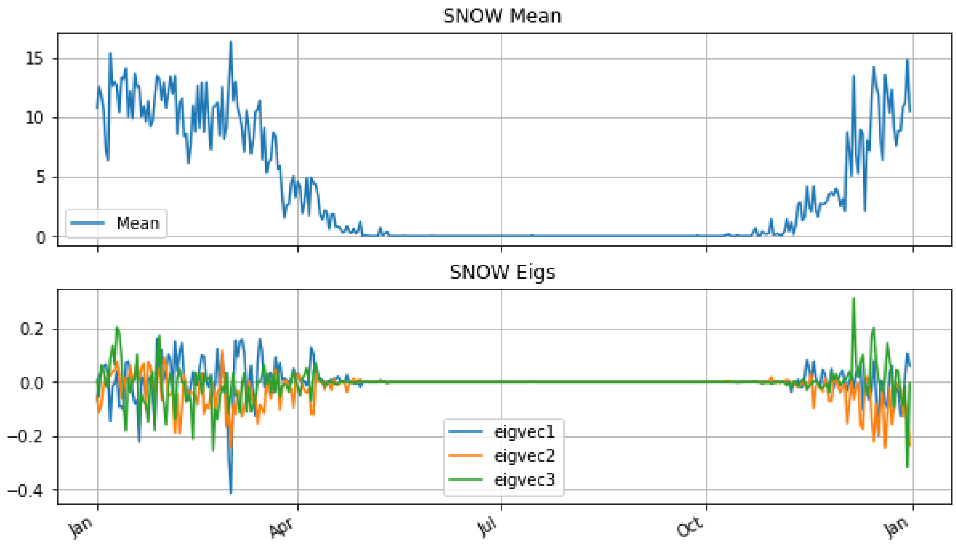


Figure 13: Mean and Eigenvectors for SNOW

3.2.6 SNWD

Figure 14 shows the mean and eigenvectors for SNWD. This measurement is least noisy as it is integral of SNOW over the days. We see that SNWD is maximum around the Feb-March. The first eigenvector explains the SNWD above the mean. Note difference is eig1 is too low between Nov-Dec. This implies that SNWD above the mean is unlikely during this period. The second eigenvector eigvec2 capture the delay in snowfall. It is negative before March and positive during March-April. If contribution of eigvec2 is high, then SNOW is likely delayed. The third eigenvector eigvec3 capture early arrival of SNOW. It is negative during Feb and positive using Dec-Jan. The fourth eigenvector capture very similar trends to eigvec2. Hence would would consider only first 3 eigenvector for our analysis in further sections.

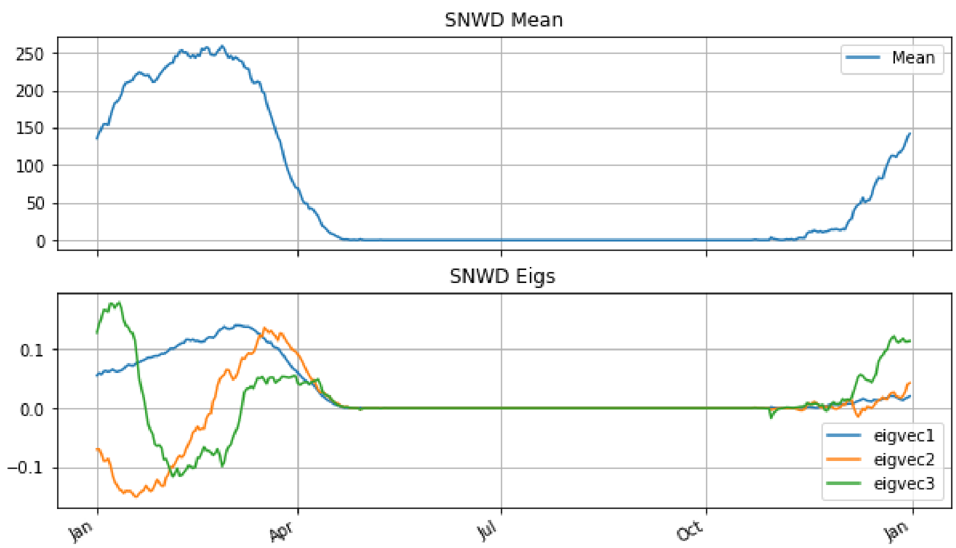


Figure 14: Mean and Eigenvectors for SNWD

4 SNOW Depth Analysis

4.1 Positive coefficients

In this section, we analyze the relation between eigenvalue coefficients and reconstruction when eigenvalue coefficients are highly positive. Note that for all the following figure coeff 1=c0, coeff 2=c1, coeff 3=c2.

Coeff 1

We can observe from the following figure that high positive coeff 1 translates to very high snow depth. Here the contribution of eigvec1 is significantly larger than other eigenvectors.

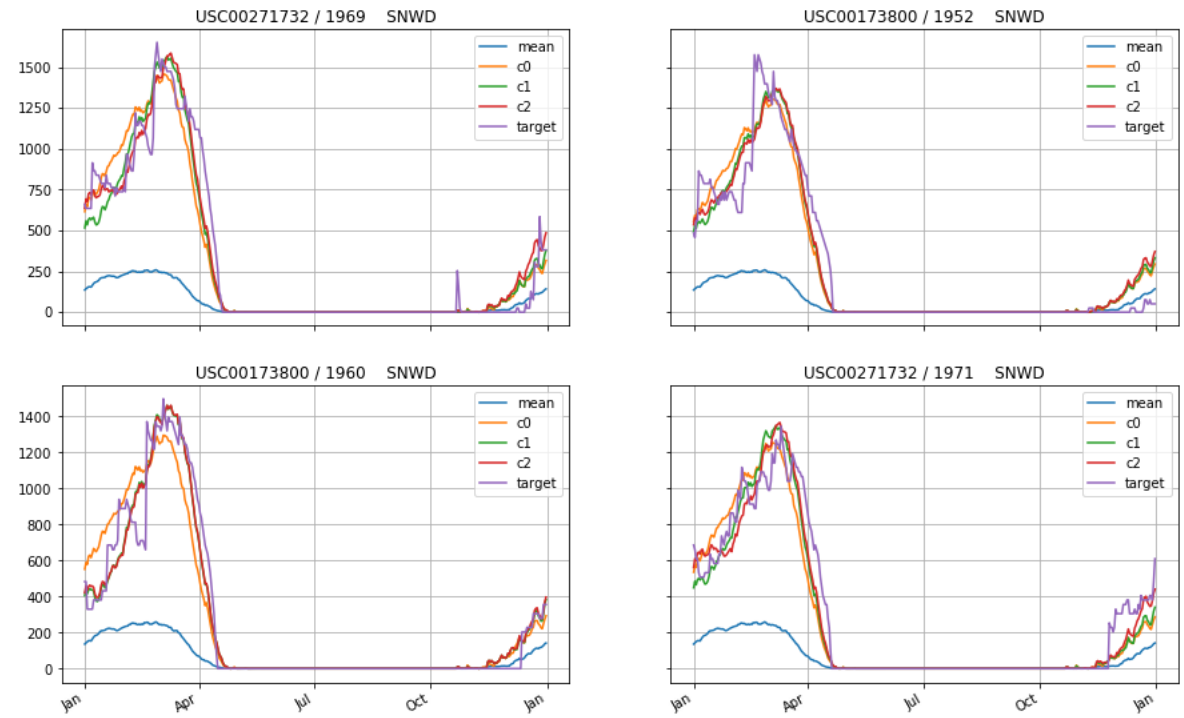


Figure 15: Reconstruction with high positive coeff 1 for SNWD

Coeff 2

We can observe from the following figure that high positive coeff 2 translates to delayed snowfall. The majority of snowfall is during the late winters, around March-April. Here the contribution of eigvec2 is significantly larger than other eigenvectors. Note that eigvec1 just captures amount of snow not the trend. Here the green line is pushing the snowfall towards the April. Note that the difference between red and green line (c2-c1) is very small.

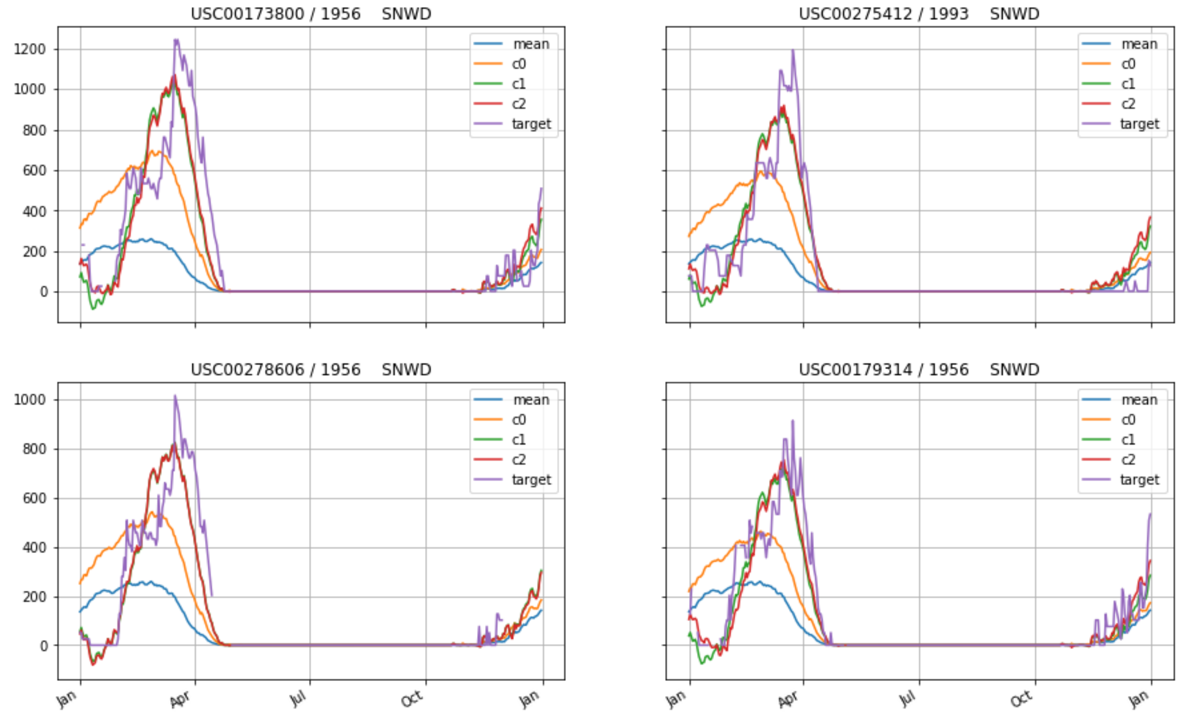


Figure 16: Reconstruction with high positive coeff 2 for SNWD

Coeff 3

We can observe from the following figure that high positive coeff 3 translates to early snowfall. The majority of snowfall is during the early winters, around Jan-Feb. Here the contribution of eigvec2 is significantly larger than other eigenvectors. Here the red line is pushing the snowfall over the Jan.

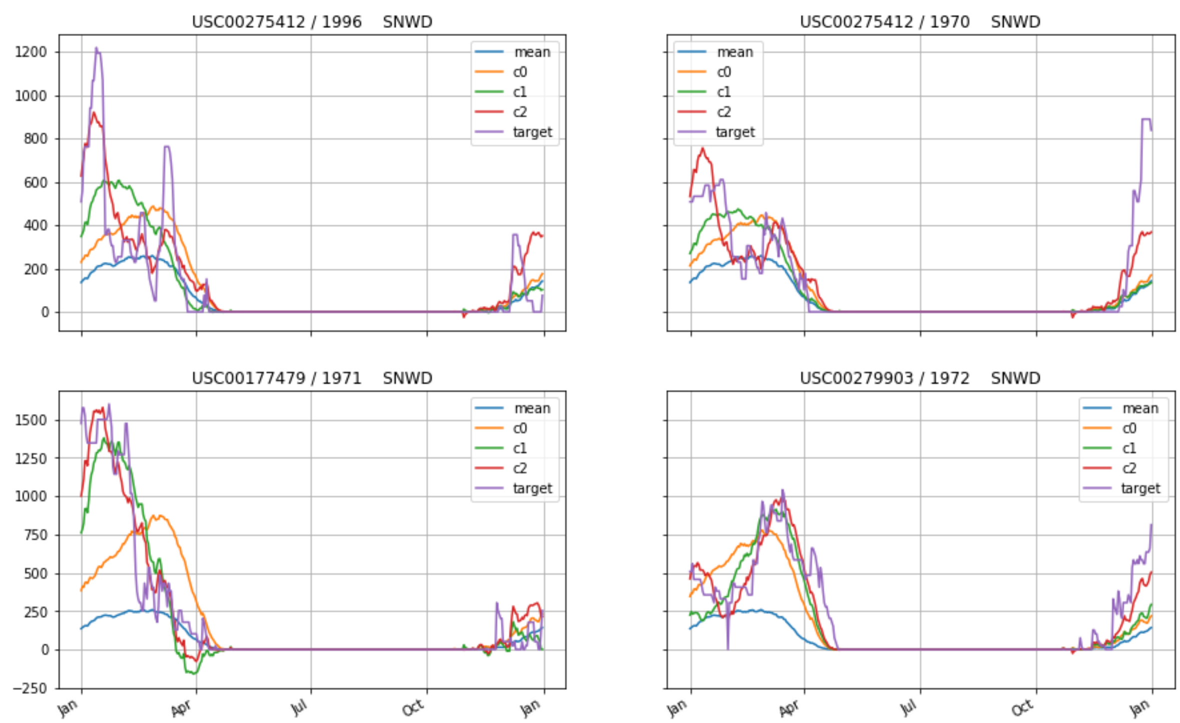


Figure 17: Reconstruction with high positive coeff 3 for SNWD

4.2 Negative coefficients

In this section, we analyze the relation between eigenvalue coefficients and reconstruction when eigenvalue coefficients are highly negative. Note that for all the following figure coeff 1=c0, coeff 2=c1, coeff 3=c2.

Coeff 1

We can observe from the following figure that high negative coeff 1 translates to low snow depth. If we compare Figure 18 with the following figure 15, the difference in the y-axis is significantly large.

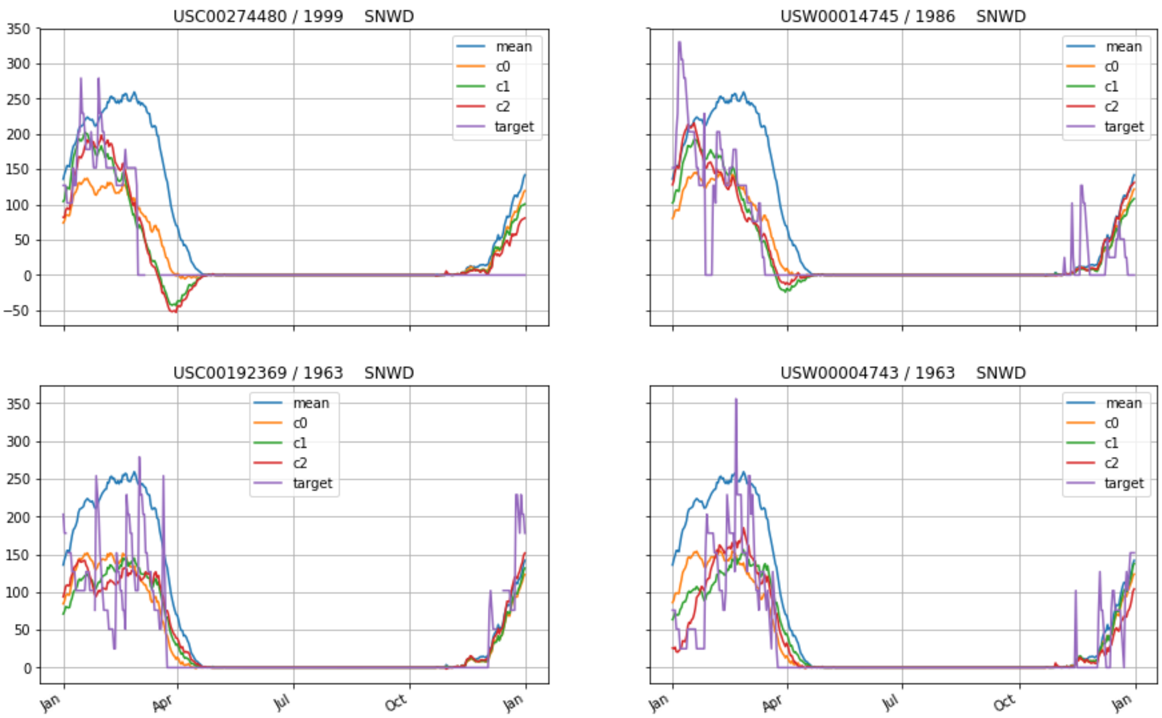


Figure 18: Reconstruction with high negative coeff 1 for SNWD

Coeff 2

We can observe from the following figure that high negative coeff 2 translates to early snowfall, opposite to high positive coeff 2 (Figure 16). This case would mean the snowfall would be early.

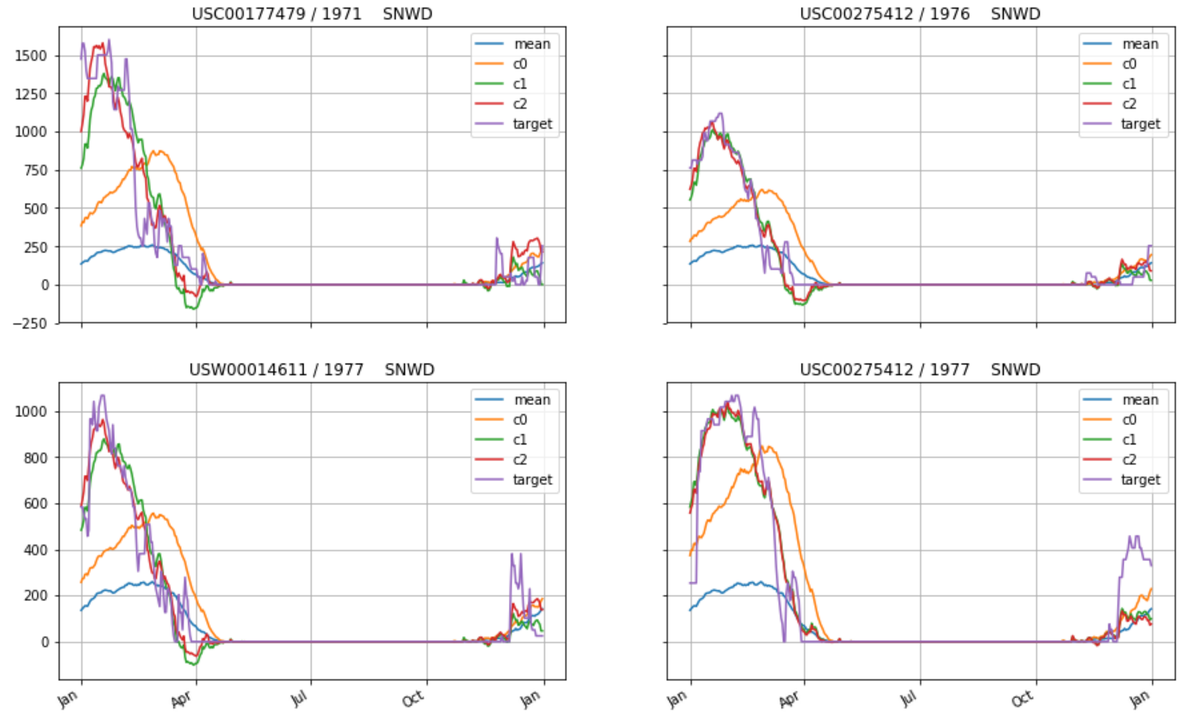


Figure 19: Reconstruction with high negative coeff 2 for SNWD

Coeff 3

We can observe from the following figure that high negative coeff 3 translates to delayed snowfall, opposite to high positive coeff 3 (Figure ??). This case would mean the snowfall would be delayed.

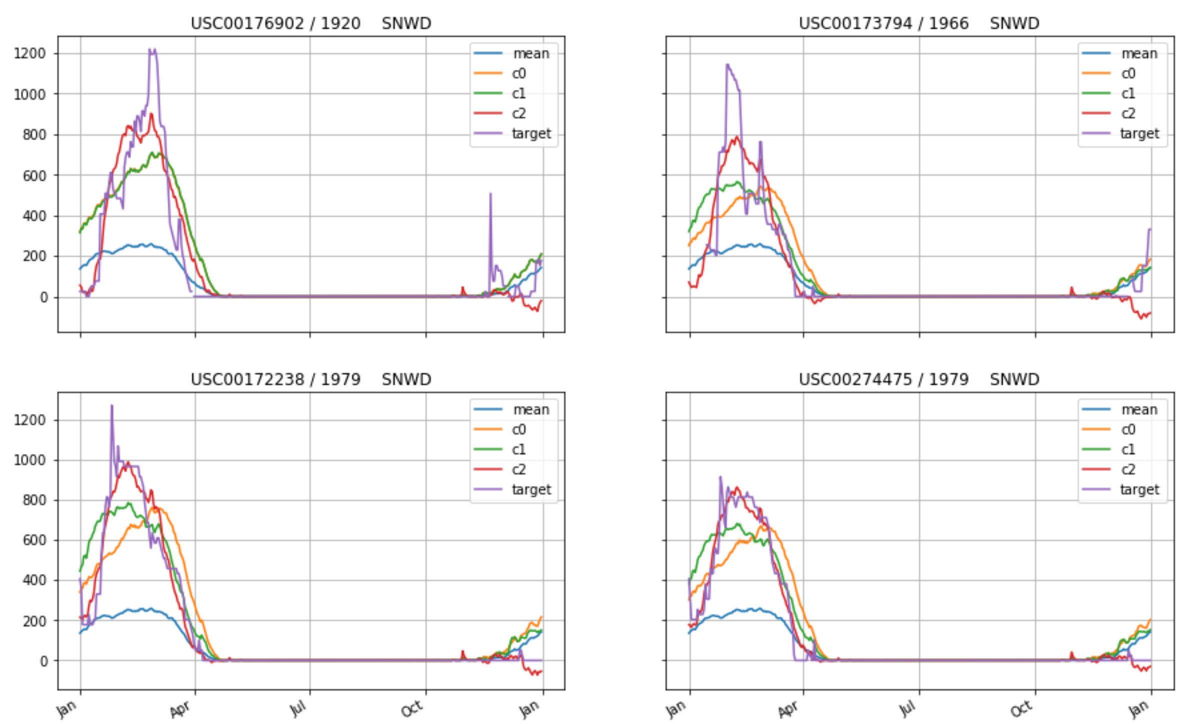


Figure 20: Reconstruction with high negative coeff 3 for SNWD

4.3 Low residual 3

When the 3rd residual is low, then we get a very good fit of our reconstruction using 3 eigenvector with the target, which can be seen in the following figure.

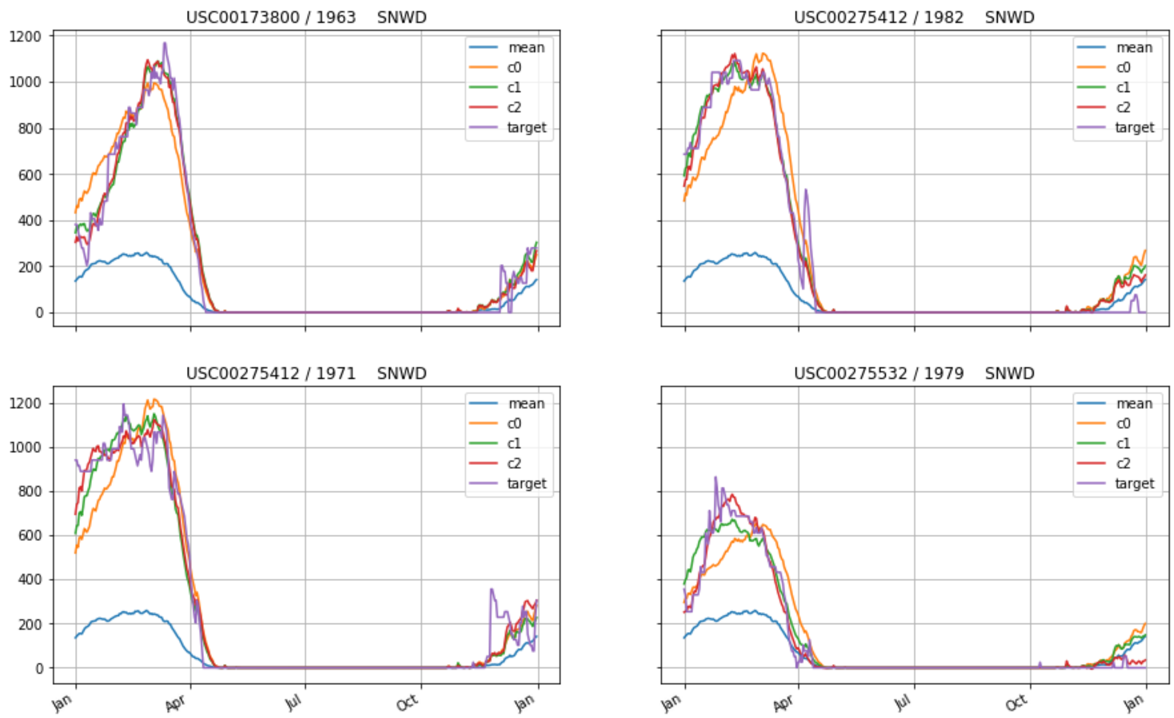


Figure 21: Reconstruction with low residual 3 for SNWD

4.4 High residual 3

When the 3rd residual is high (sometimes greater than 1), then we get a very poor fit of our reconstruction using 3 eigenvector with the target. This can also be attributed to noisy and bad data or very unusual treads in the SNWD. We can observe from the following SNWD is very noisy and has peaks sharp peaks which is unlikely for snow depth until there are geographical factors.

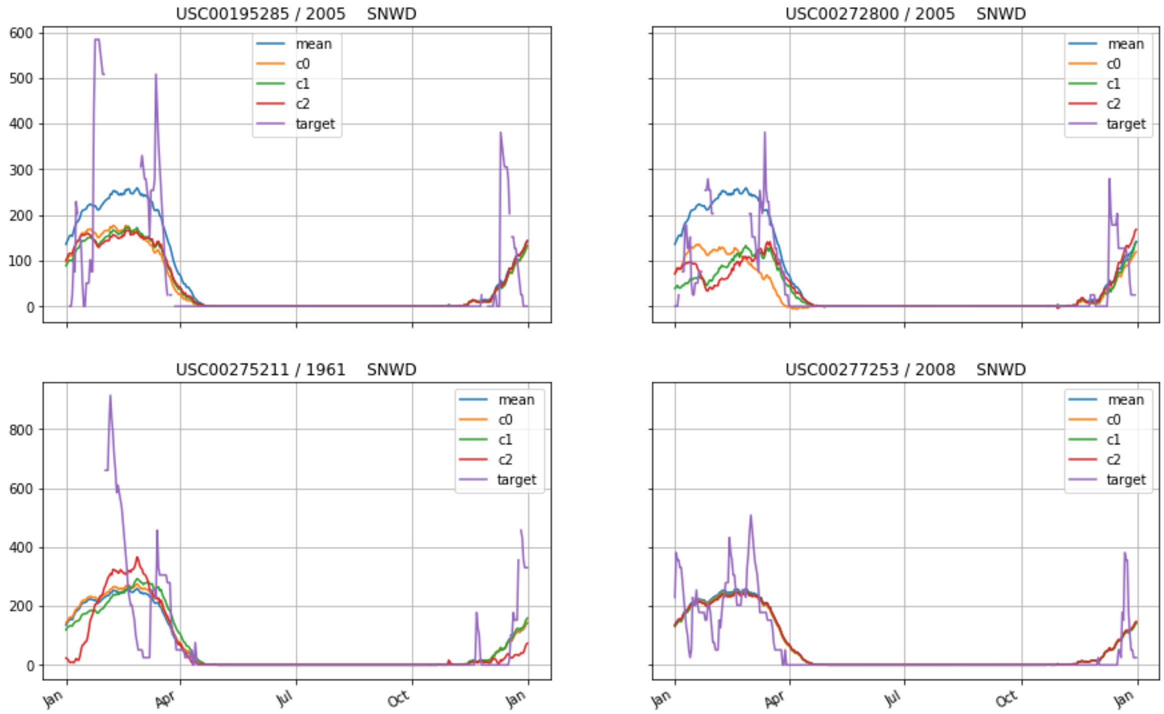


Figure 22: Reconstruction with high residual 3 SNWD

4.5 Temporal vs Spatial

In this section we compare the temporal and spatial dependence of SNWD. We compare RMS after removing mean-by-station (spatial) and mean-by-year (temporal) for 3 coefficients of the SNWD.

Coeff 1

```
total RMS                   =  1857.21540069
RMS removing mean-by-station=  1394.96185094
RMS removing mean-by-year   =  1340.12426857
```

We conclude that for coeff 1 which capture the SNWD above mean has little more temporal dependence than spatial. This is mostly because coeff 1 related to total snowfall.

Coeff 2

```
total RMS                   =  978.569620587
RMS removing mean-by-station=  930.754194373
RMS removing mean-by-year   =  493.268247967
```

We conclude that for coeff 2 high temporal dependence than spatial. Since coeff 2 is related to timing of the snowfall (delayed or not), variations by years explains better than variations over stations.

Coeff 3

```
total RMS                   =  634.164702073
RMS removing mean-by-station=  594.297547809
RMS removing mean-by-year   =  301.323808032
```

We conclude that for coeff 3 high temporal dependence than spatial. Since coeff 3 is related to timing of the snowfall (early or not), variations by years explains better than variations over stations.

4.6 Coeff 1 over location

Following is a map of coeff 1 value over the location of stations. The size of the circle is proportional to the counts of the measurement.

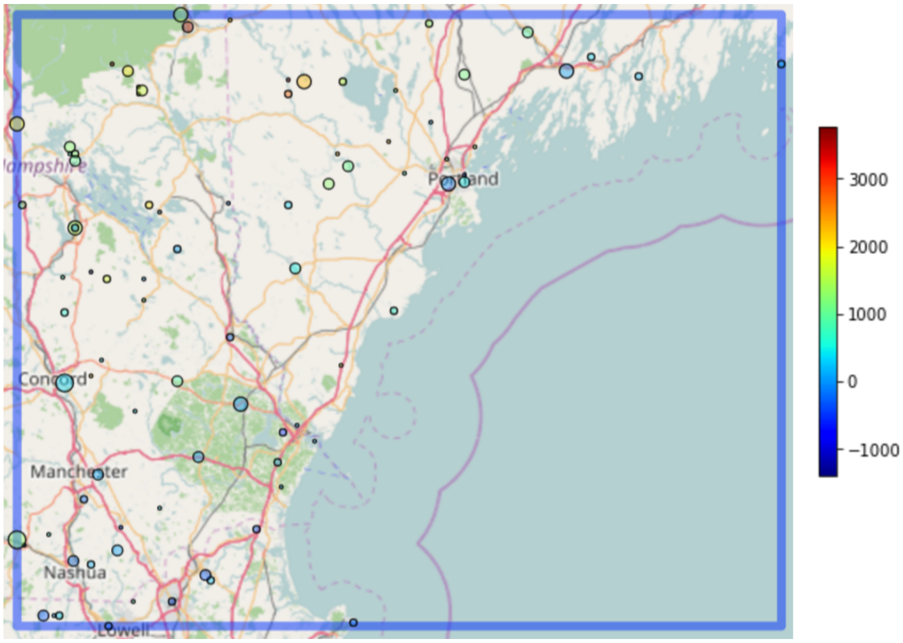


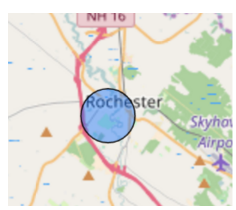
Figure 23: Plot of coeff 1

We observe that the map is well justified with the observation that coeff 1 is related to the total snowfall. Following are some of the stations with their annual snowfall taken from US Climate Data. This matches with the color coding of our map. Note that blueish means less SNWD and reddish means more SNWD.

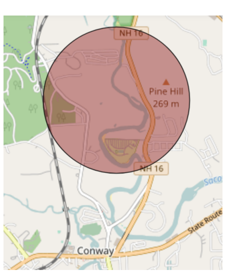
```
Conway, New Hampshire:   Av. annual snowfall: 203 cm
Concord, New Hampshire:  Av. annual snowfall: 155 cm
Hiram, Maine:            Av. annual snowfall: 196 cm
Rochester, New Hampshire:  no snowfall
Manchester, New Hampshire: no snowfall
```

Following are the figures for these stations

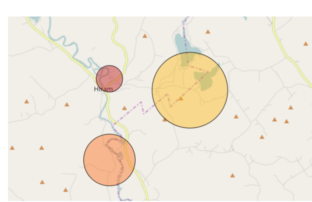
(a) Manchester



(b) Rochester



(a) Conway



(b) Hiram

4.7 Coeffs over location

Following figure shows the plot of 4 eigenvector coefficients for SNWD over the locations.

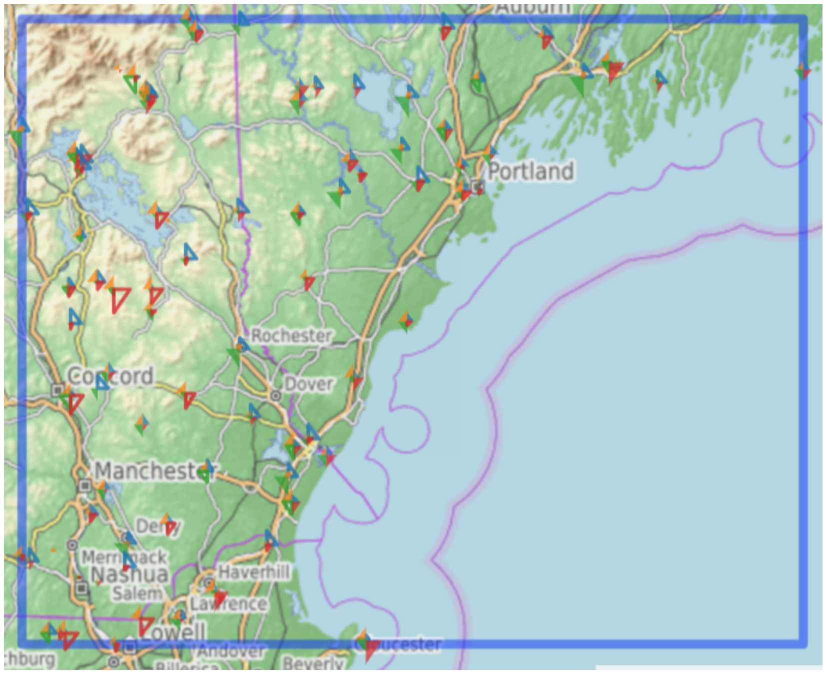


Figure 26: Plot of coeffs

We can over the local trends of the coefficients over the map.

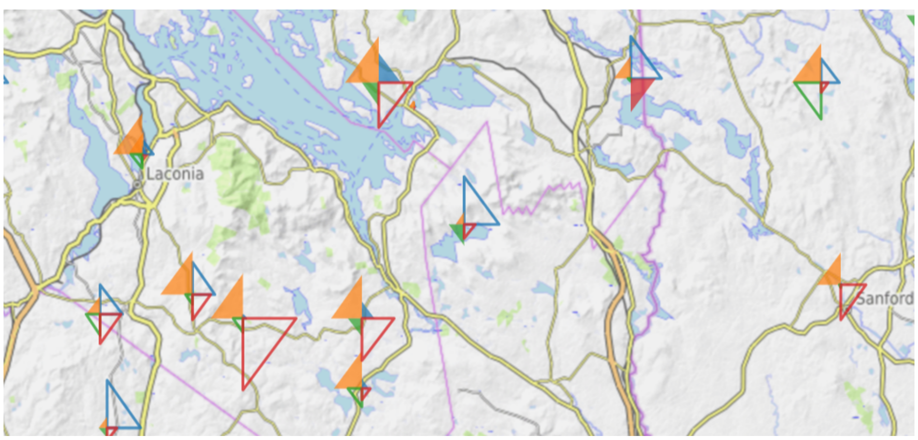


Figure 27: Local trends

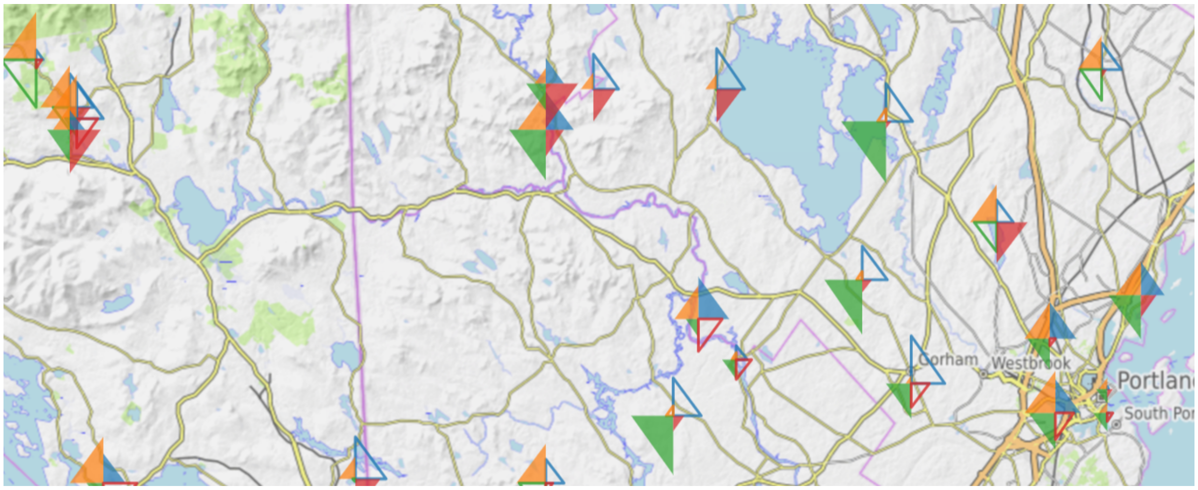


Figure 28: Local trends

5 Statistical Significance

5.1 SNWD

Following figures shows the correlation of stations for top-4 PCA components for SNWD. We can observe block diagram structure in each of these Rependency matrix. Coeff 1 figure (top, left), we can see at the corners that 0-20 stations are highly correlated as well 70-80. There is another block at the diagonal 35-55. Similarly, (top, right), there is stations 50-100 are high correlations (except around 75, there is a yellow strip here). Since the 2nd coeff of SNWD captures the timing of the now, we can say that if snow is delayed for some stations 50-100, then it is likely that it will be delayed for other stations as well. We can further divide this block into multiple blocks to see correlation at fine granularity. We can observe the spatial relationship from these matrices.

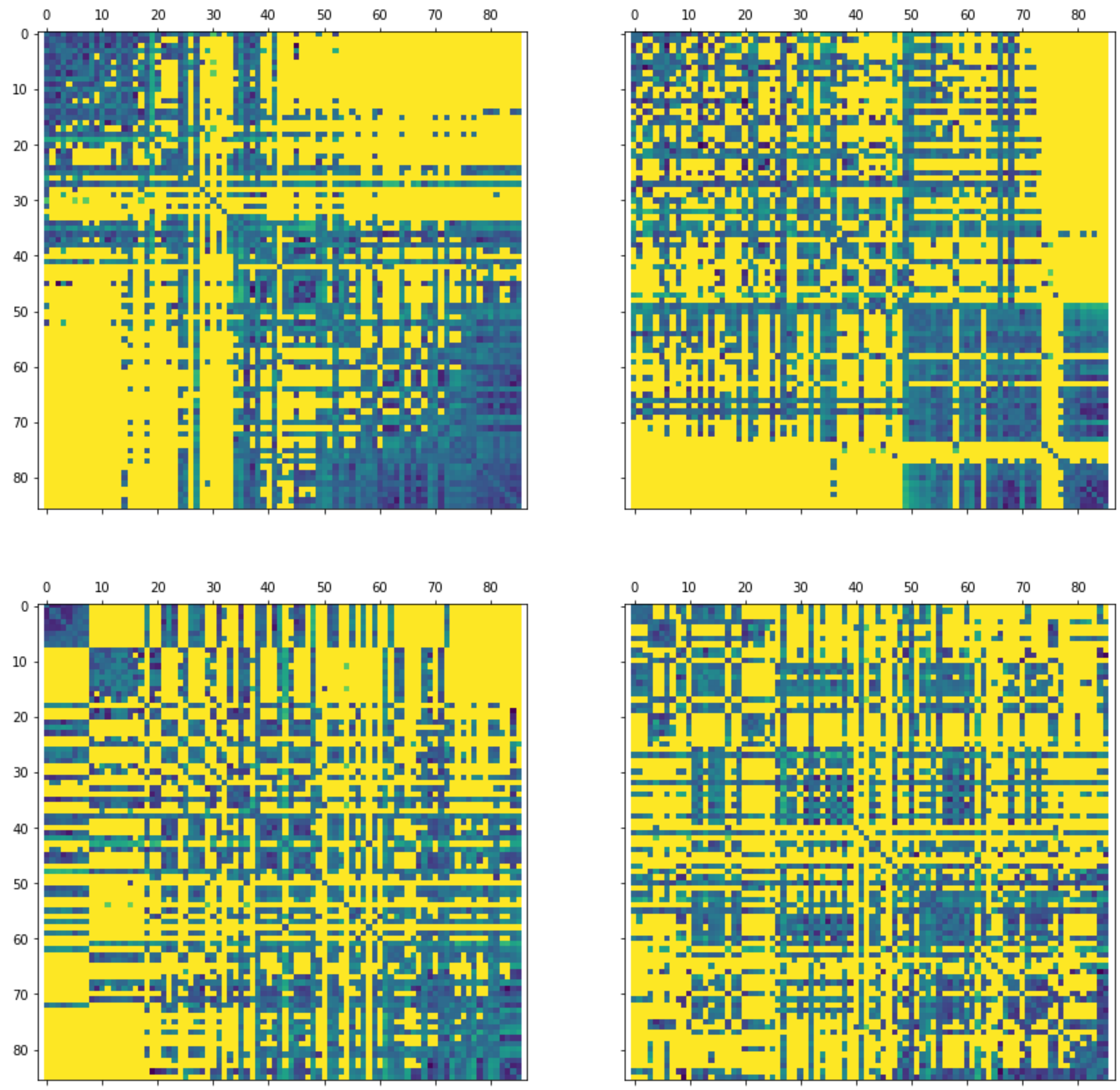


Figure 29: Rependency matrix for SNWD

5.2 PRCP

Following figures shows the correlation of stations for top-4 PCA components for PRCP. We can observe block diagram structure in each of these Rependency matrix. Coeff 1 figure (top, left), we can see at the top-left corner that 0-85 stations are highly correlated. This means that if it rains in any of 0-85 stations, it is likely that it will rain in other 0-85 stations. This captures the spatial relationship between stations for PRCP measurement. We can also observe block diagonal structures for other matrices.

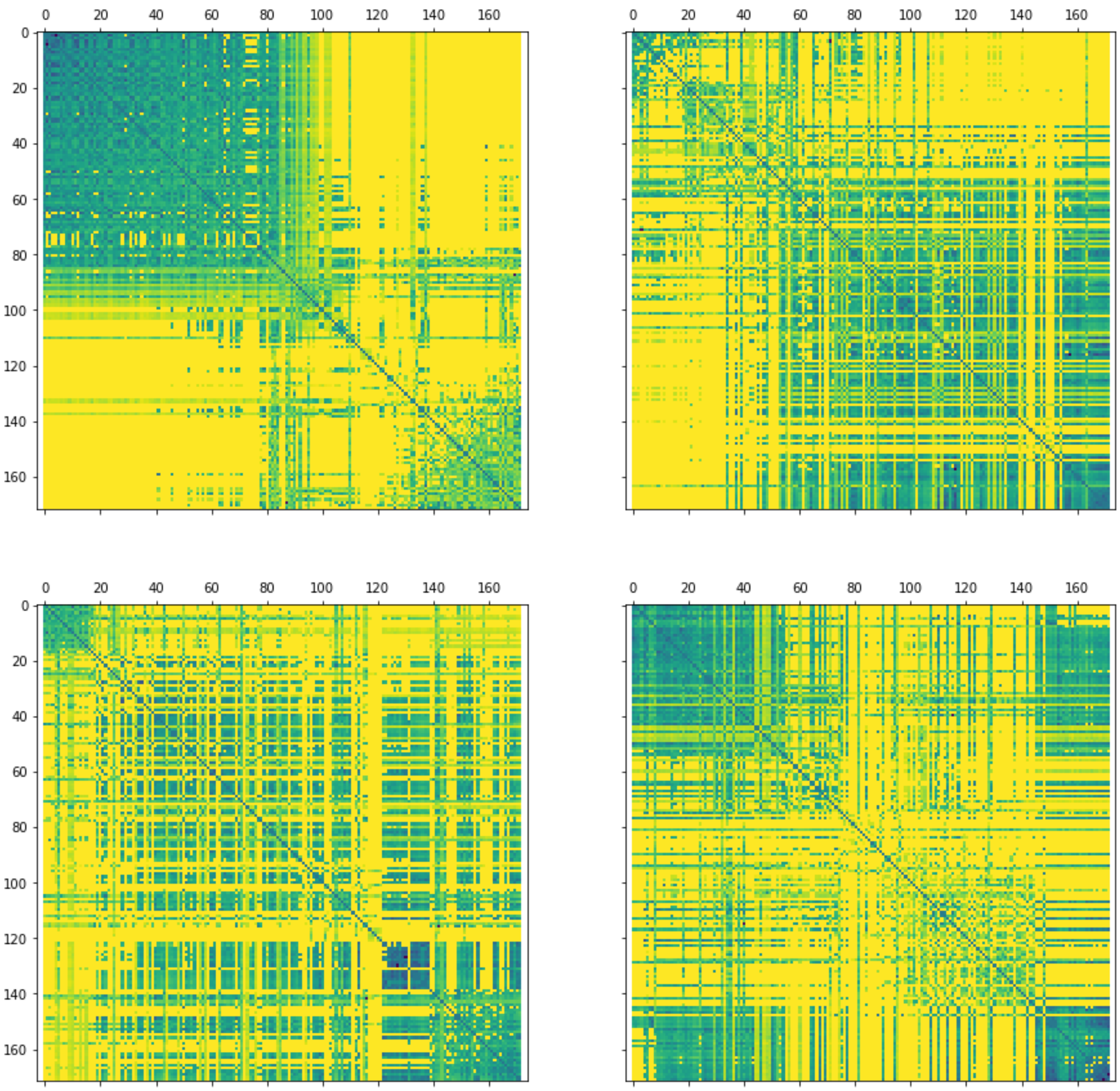


Figure 30: Rependency matrix for PRCP

References

[1] NOAA. https://www.ncdc.noaa.gov/.

[2] NOAA Data. ftp://ftp.ncdc.noaa.gov/pub/data/ghcn/daily/.

[3] US Climate Data. http://www.usclimatedata.com/climate/united-states/us.