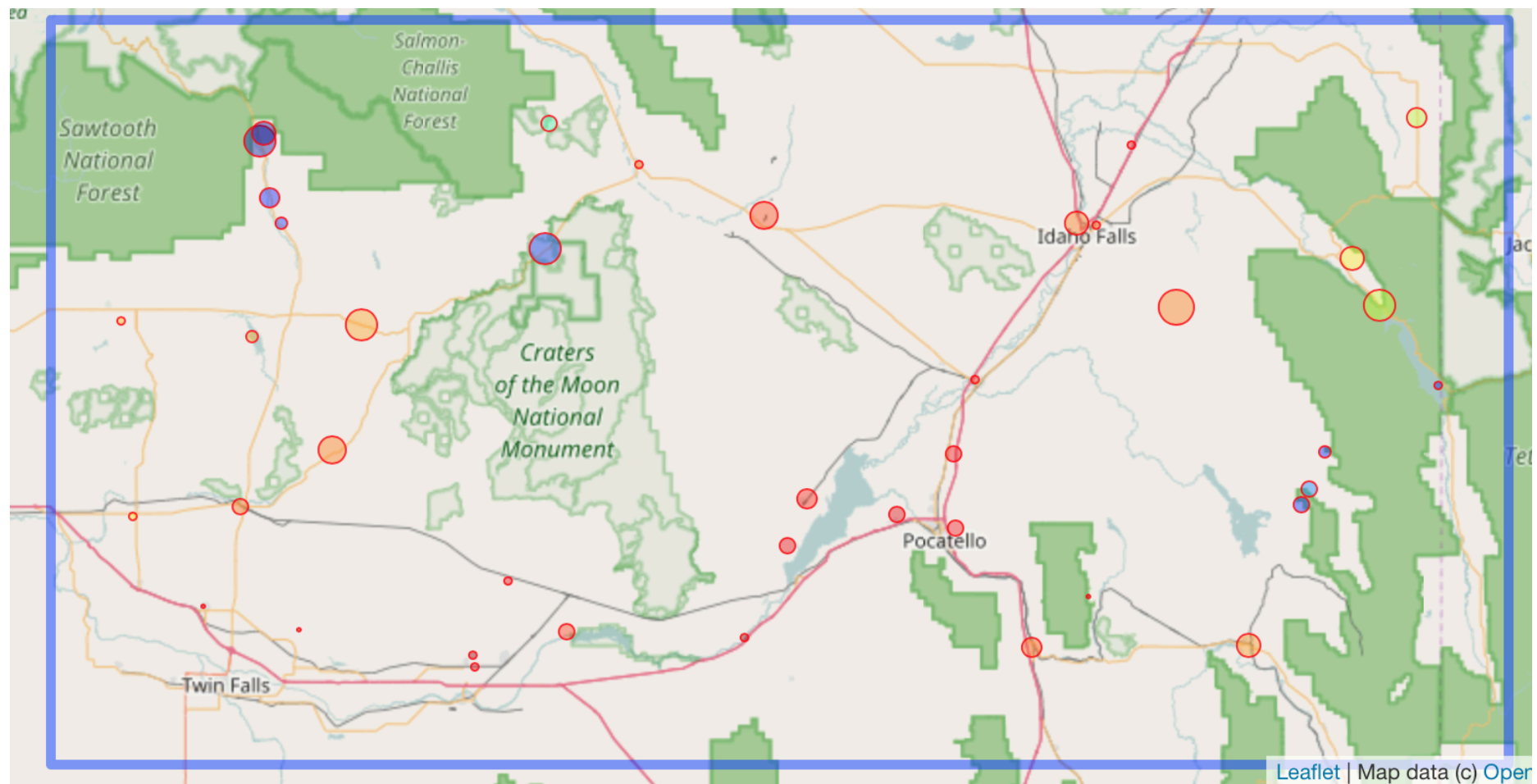


South Idaho Climate Analysis

This report provides climate analysis of temperature, snow and precipitation fields in the south Idaho region. Data used for the analysis is captured from 89 measurement stations located in and around craters of the moon, sawtooth national forest and Idaho falls. Data captured is for last 100 years for some of these stations.



This data specific to this region is downloaded from FTP site.

Analysis is focused based on six measurements which are listed as core elements from the data source description.

- TMAX : Maximum temperature
- TMIN : Minimum temperature
- TOBS : Temperature at the time of observation
- PRCP : Daily Precipitation
- SNOW : Snowfall
- SNWD : Snow depth

Report includes information related to characterization of these measurements listed. Another primary focus of the report is on Principle component analysis (PCA) of these measurements.

1) Importance of snow depth and temperature analysis of Southern Idaho

South Idaho region has a characteristic of very few inches of rain every year. This was noticed in precipitation which is almost same throughout the year with minimal variation. Hence household in this region considered to be living in desert with not much drinking water. Primary source of drinking water is snow fall during winter.

From literature, typically snow water equivalent (SWE) is a common snow pack measurement that is used to understand amount of water contained with in the snowpack. Every year, to predict availability of water for next year is a great need and analysis of snow depths along with temperature is required. A good prediction requires good statistical analysis which in this report is explored through PCA. Inaddition to prediction of drinking water, analysis should help to detect climate changes that should help initiate necessary corrective measures.

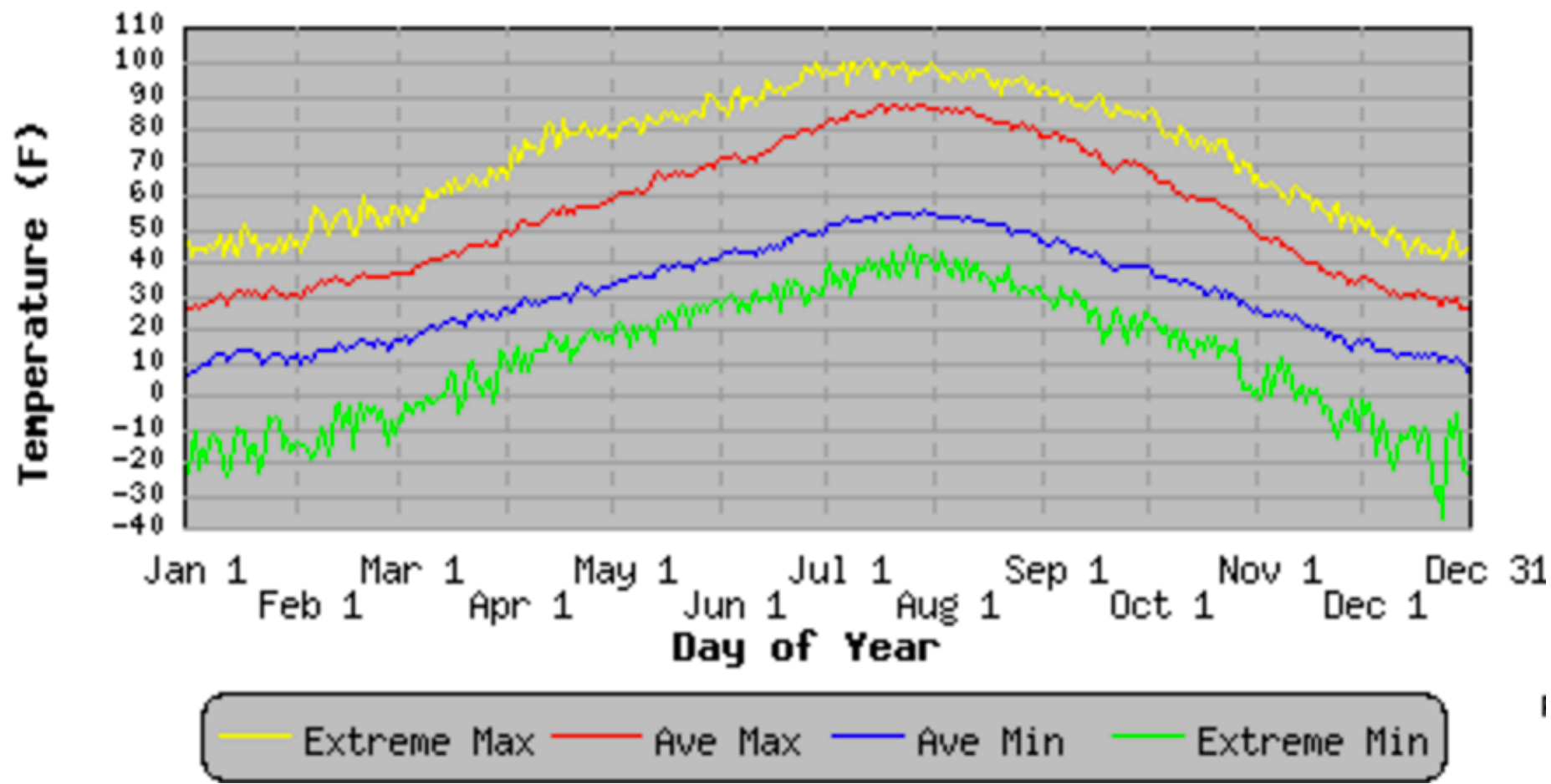
2) Data validation

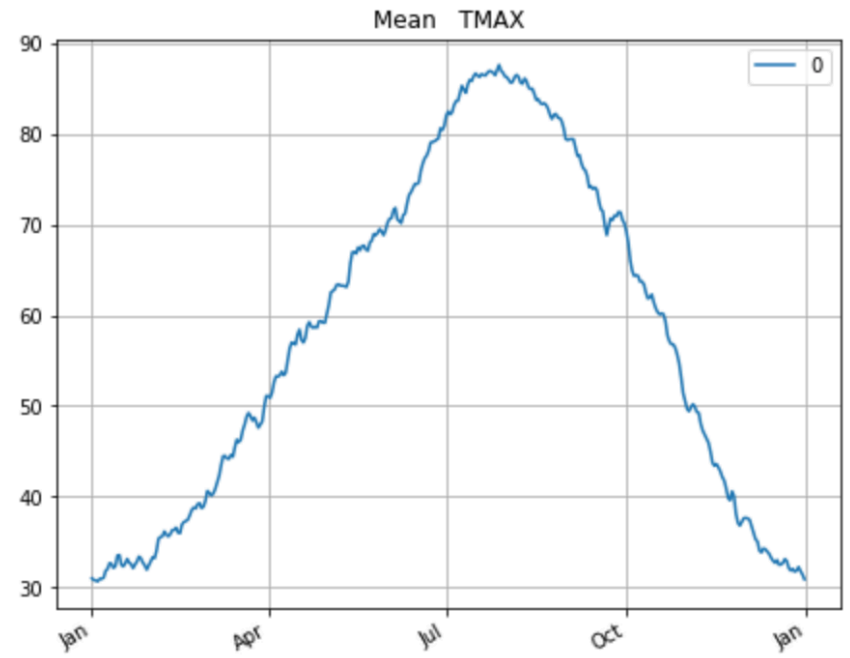
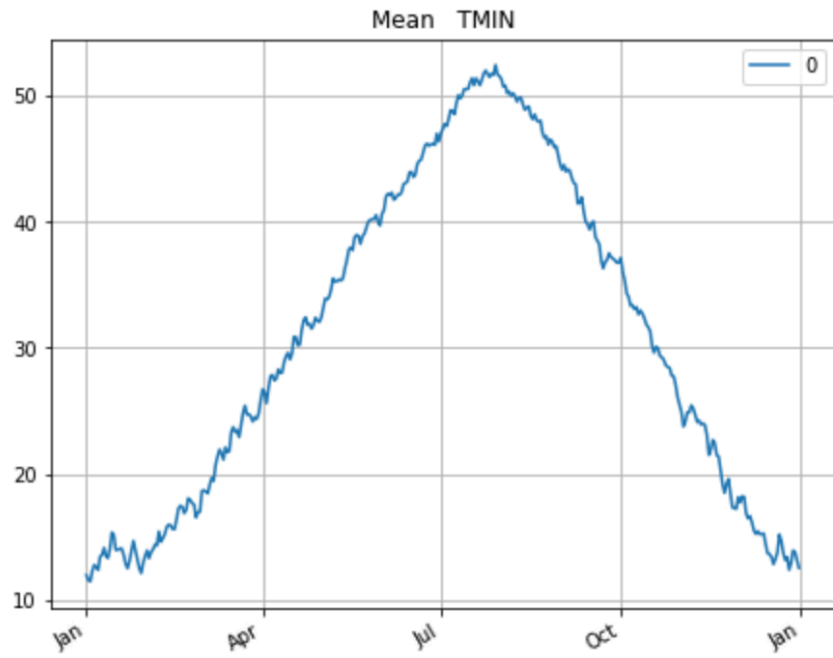
First step in data analysis is to make sure that data being used for analysis is correct. In this section, we provide comparison with general statistics from <http://www.wrcc.dri.edu/cgi-bin/cliMAIN.pl?id2260> (<http://www.wrcc.dri.edu/cgi-bin/cliMAIN.pl?id2260>) to validate that data is good for further analysis.

In next few subsections, data from external source listed is compared with data being used. Comparing all the listed plots in respective subsections, data being used for analysis agrees with the external source. Hence data is validated.

2 (a) Comparison of Temperature (TMIN, TMAX) data

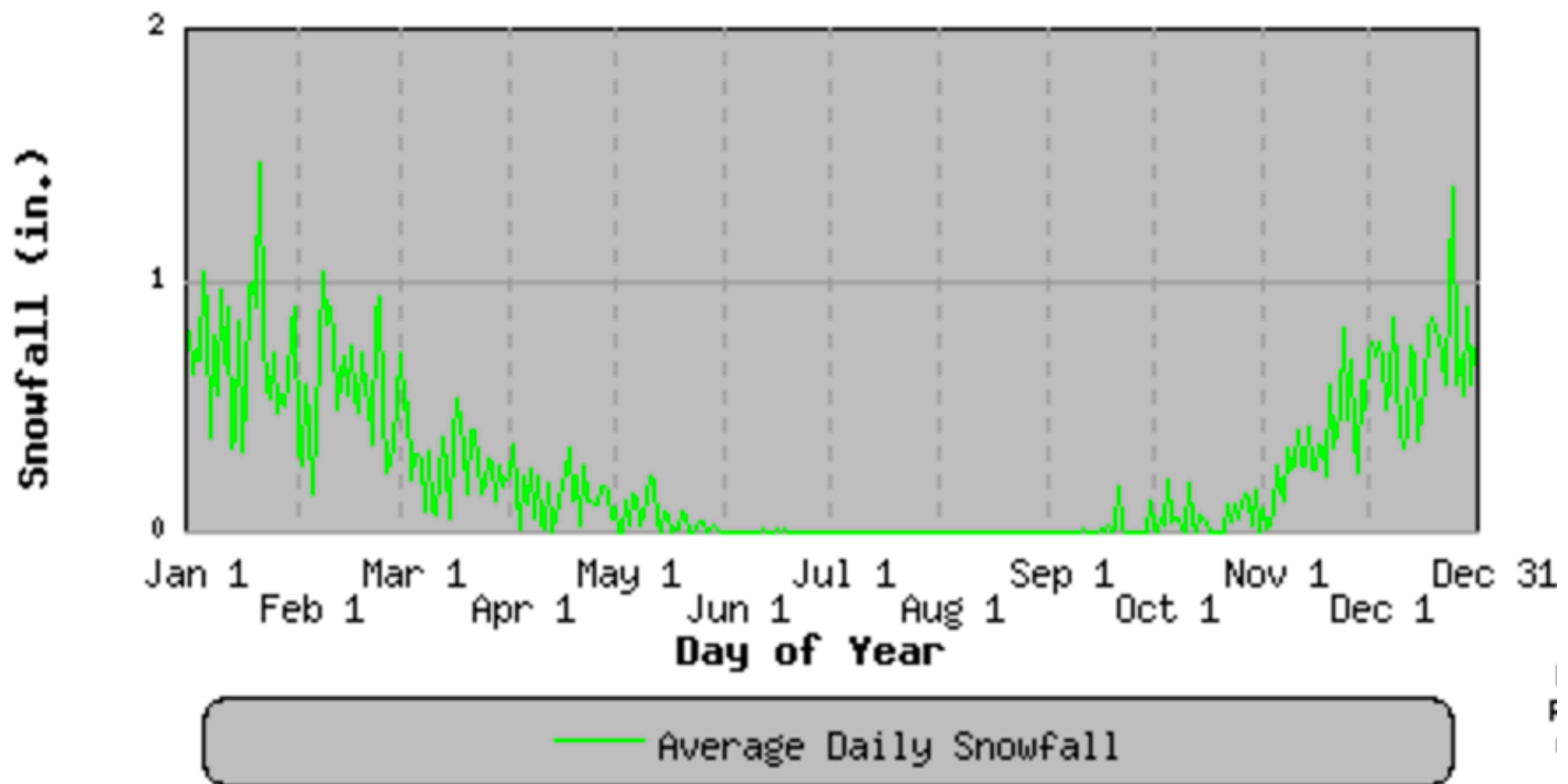
In this section, we compare TMIN, TMAX measurements from the data being used for analysis and external source. Comparing these data, our data used for analysis agrees with the external source.

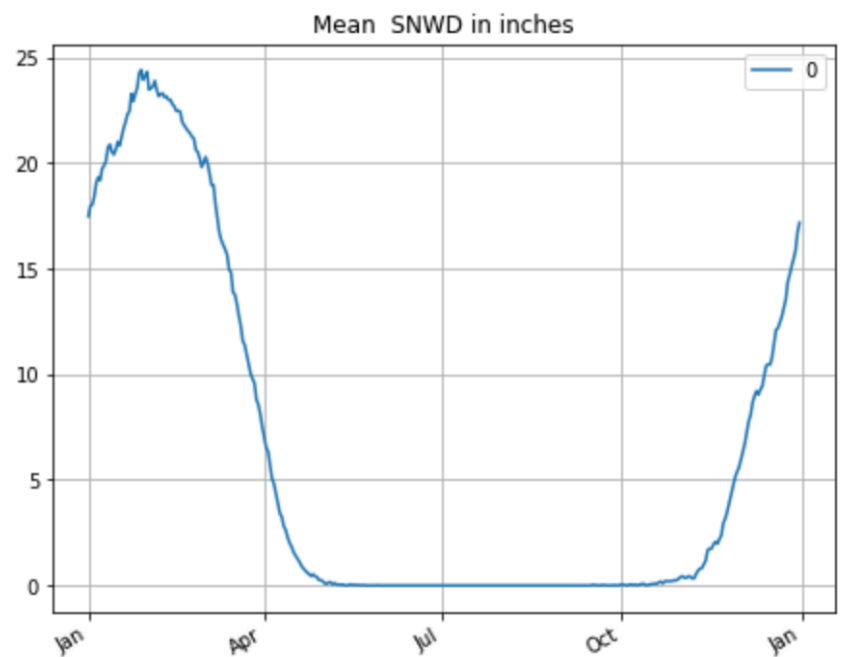
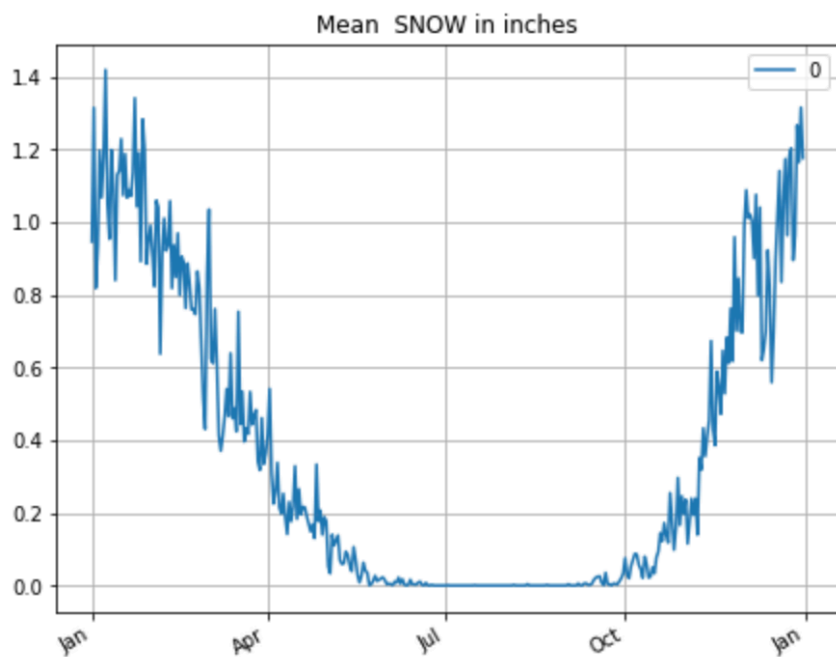
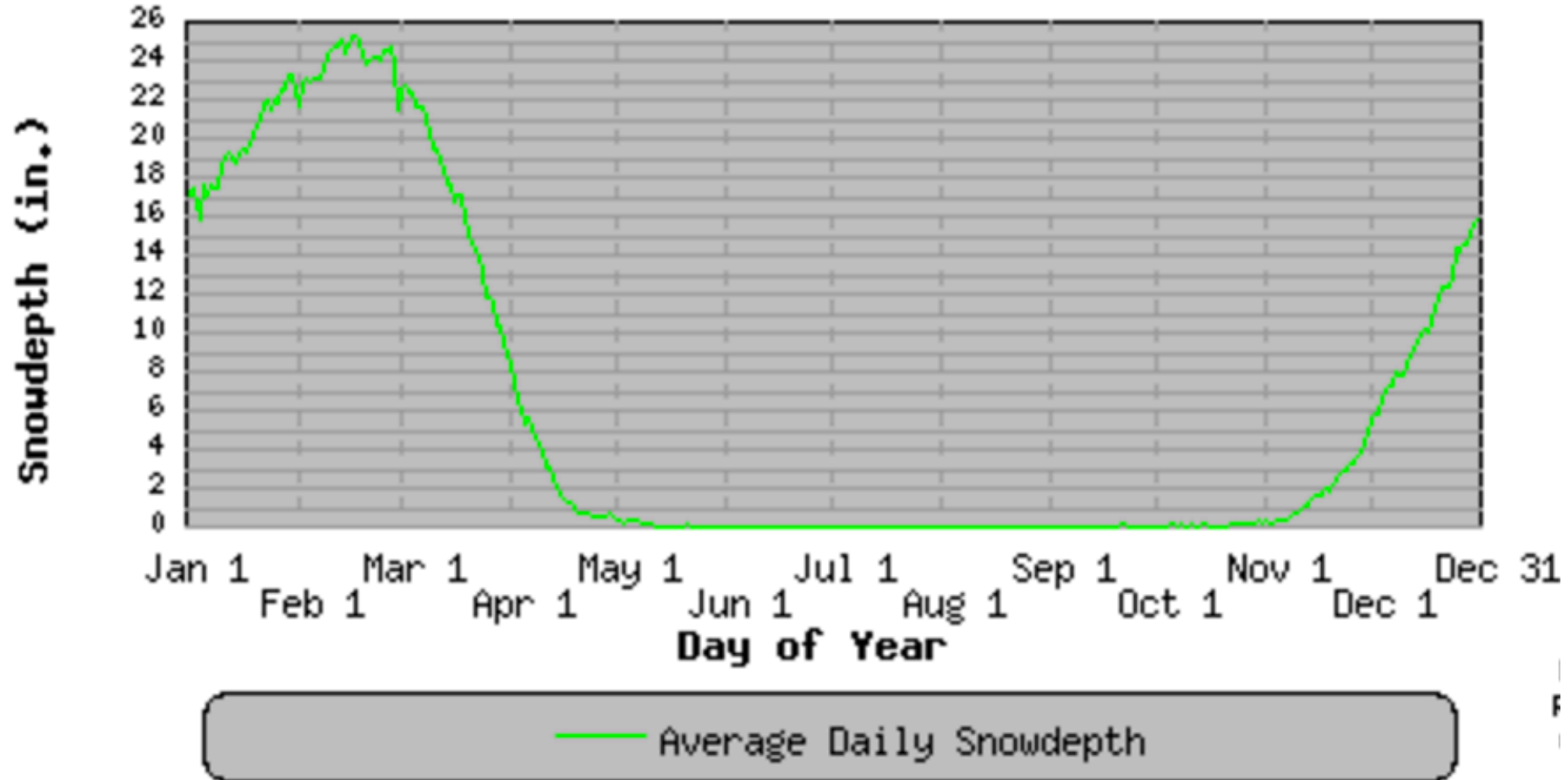




2 (b) Comparison of SNOW fall and Snow depth data data

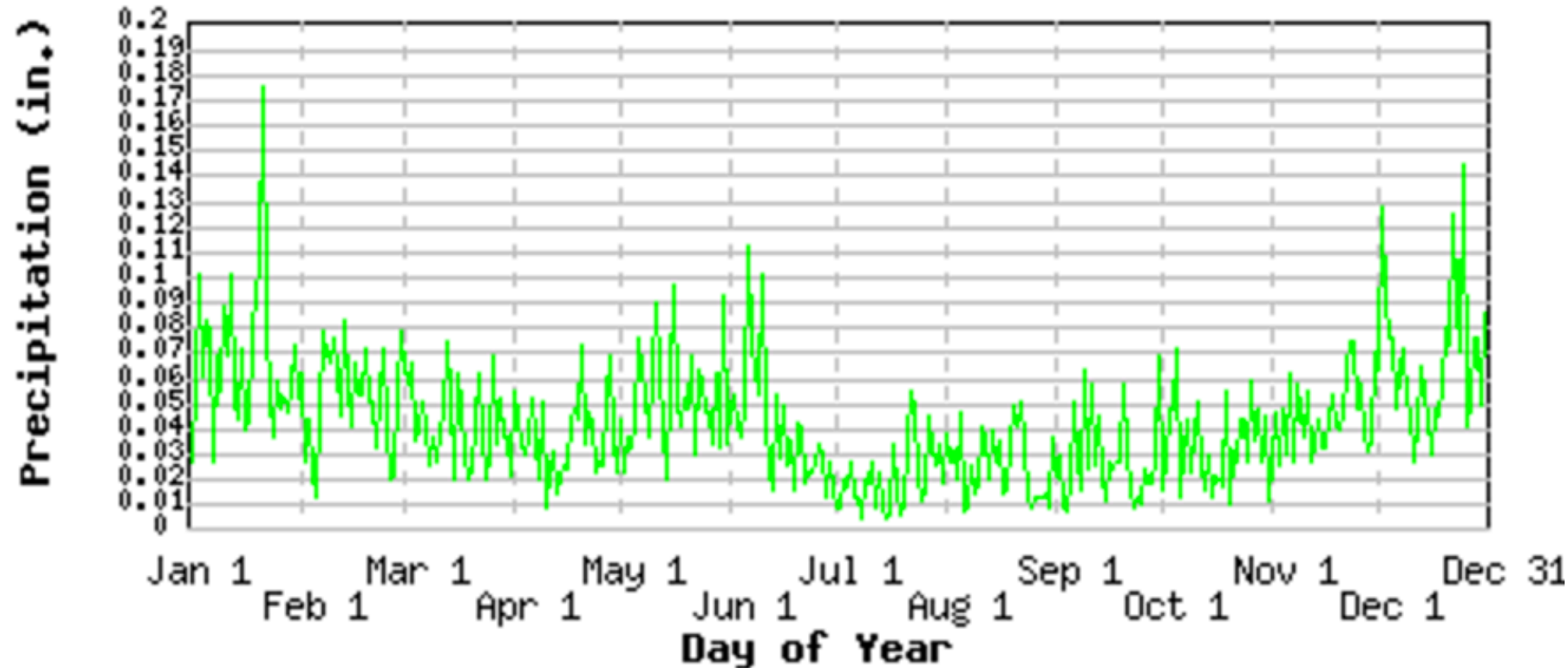
In this section, we compare SNOW fall and SNOW depth measurements from the data being used for analysis with external source. Peak of average snow depth is observed during months of February and March.





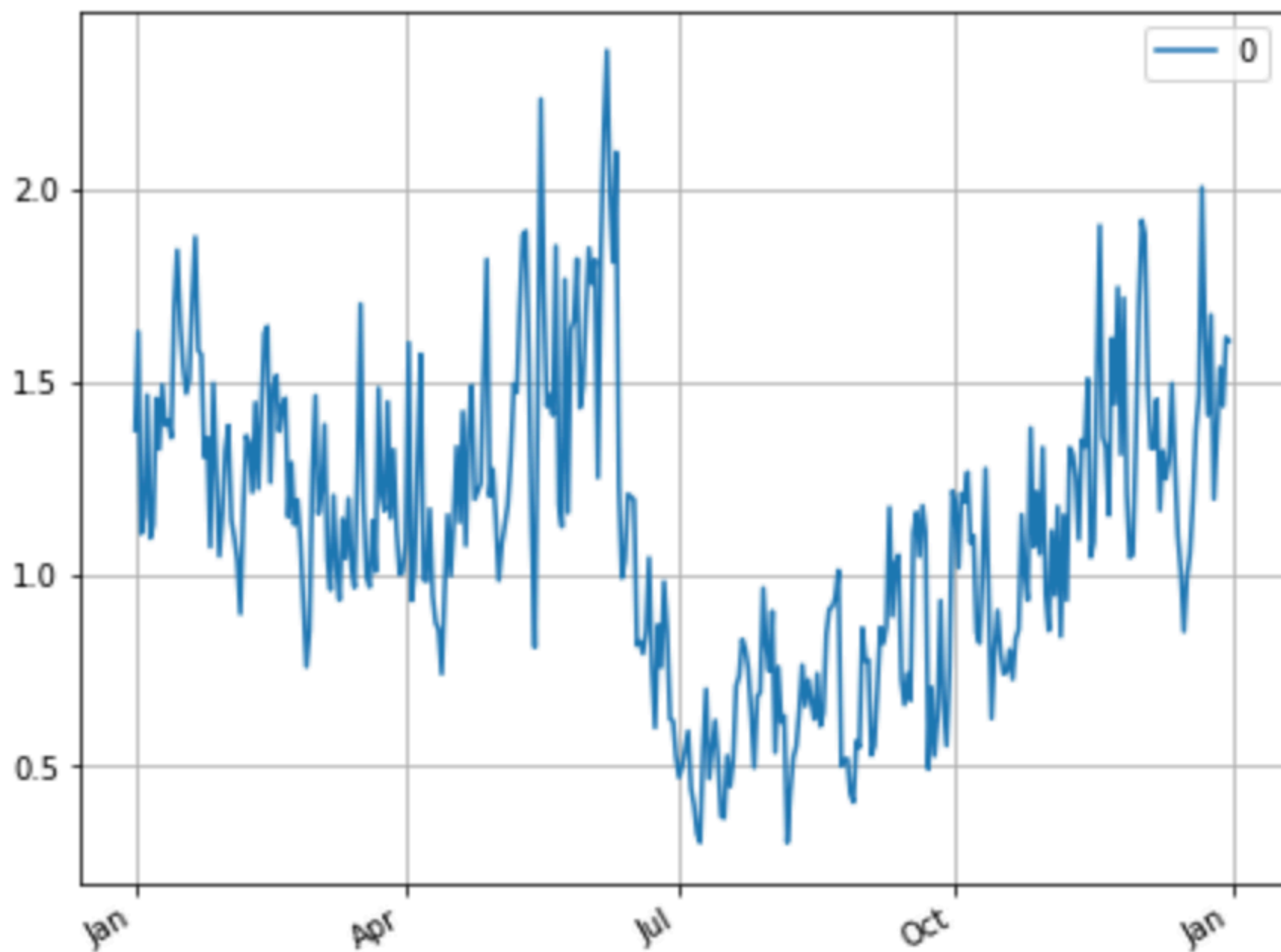
2 (c) Comparison of Precipitation data

In this section, we compare precipitation data.



— Average Daily Precipitation

Mean PRCP in inches



3) Statistical significance test

From the previous plots, we have noticed that precipitation remains constant throughout the year with not much variation. There is not statistical analysis that could be performed on that measurement. But snow depth has a good trend with the season. Hypothesis before further statistical analysis had is that snow depth has a strong correlation with TMin. Variations of the snow depths is in accordance with variation in TMIN measurement.

In this section, we wanted to study statistical significance tests based on correlation between different measurements. For this study, focused analysis based on measurements from one single station. Station with good enough measurements where data from all the measurements is available is consider. Another factor considered is to find a station that could be true representative of this south idaho region. Station and its location on map can be identified based on the lat/Long information. In this case, used a reference below to identify its location.

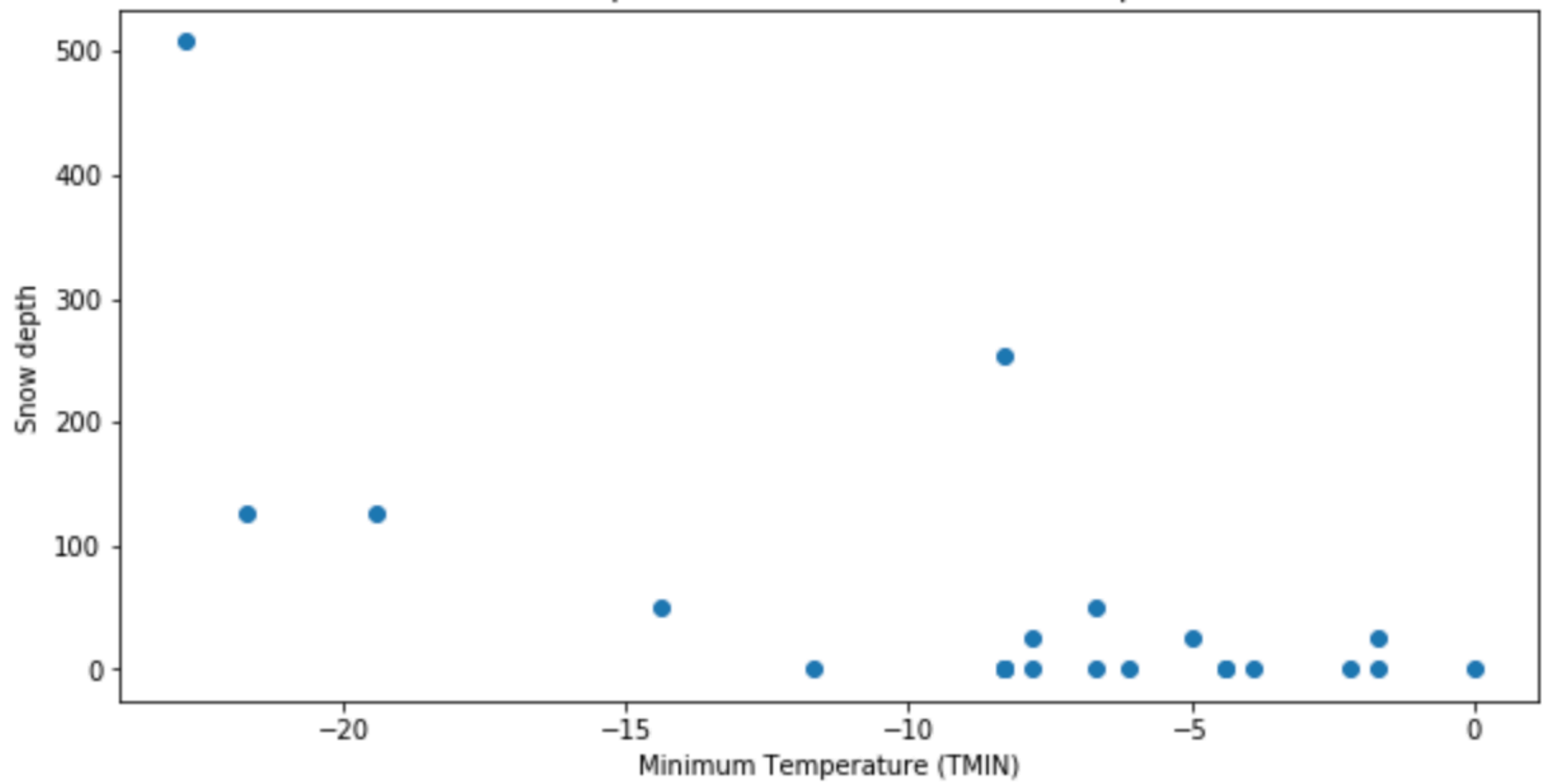
[REF] Mauger, G. S., K. A. Bumbaco, G. J. Hakim, and P. W. Mote, 2013: Optimal design of a climatological network: Beyond practical considerations. *Geosci. Instrum. Method. Data Syst.*, 2, 199–212, doi:10.5194/gi-2-199-2013.

Station with ID USC00100010 is used further for our study in this section and observe if this agrees with hypothesis or rejects. This station is considered as average measurements from this station are similar to the average measurements from the entire cluster of stations. So this could good approximation of the climate behavior of this south idaho region.

3 (a) Correlation between TMIN and SNWD

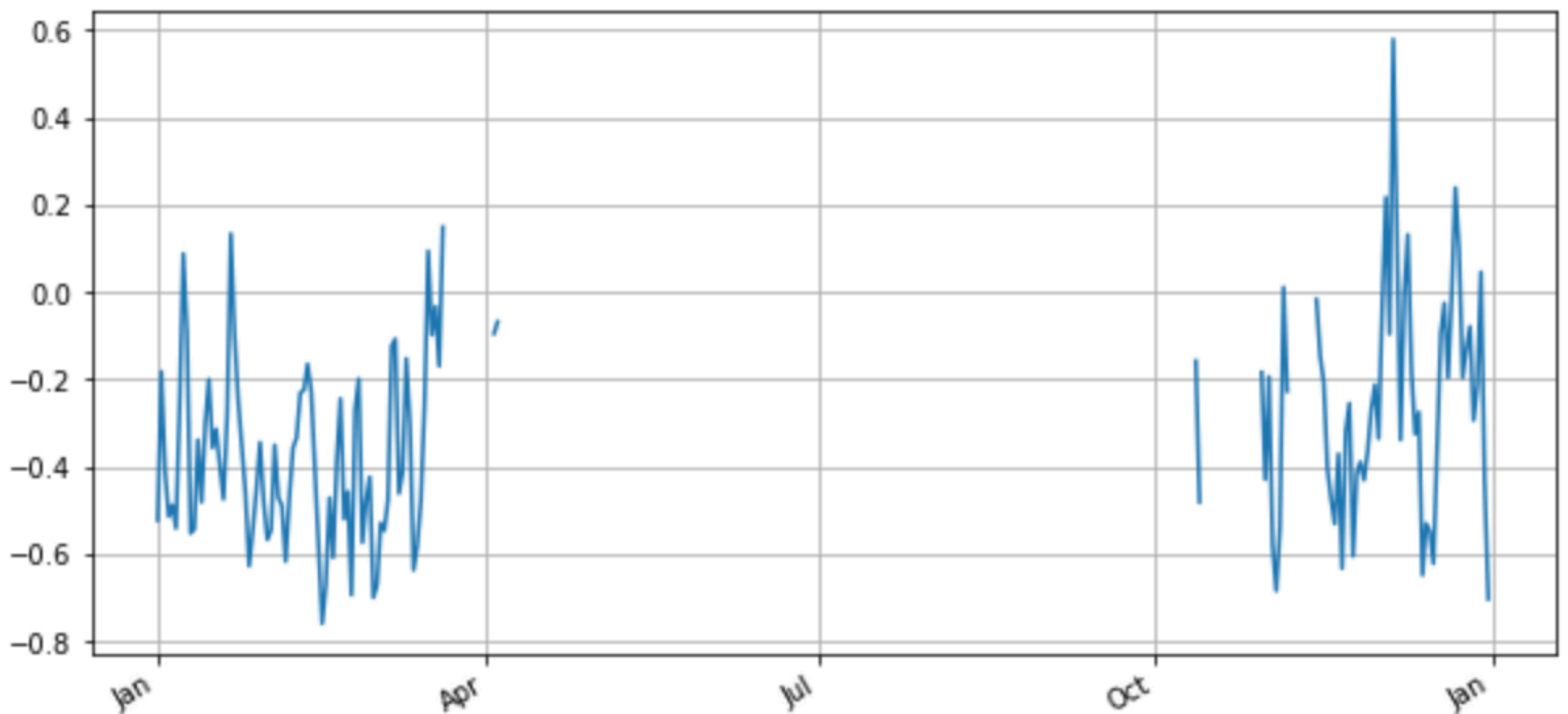
For this station identified, plots listed below are scatter plot and correlation coefficients. Based on these plots below, we have noticed a good correlation between SNWD and Tmin.

Scatter plot between TMIN and SNOW depth



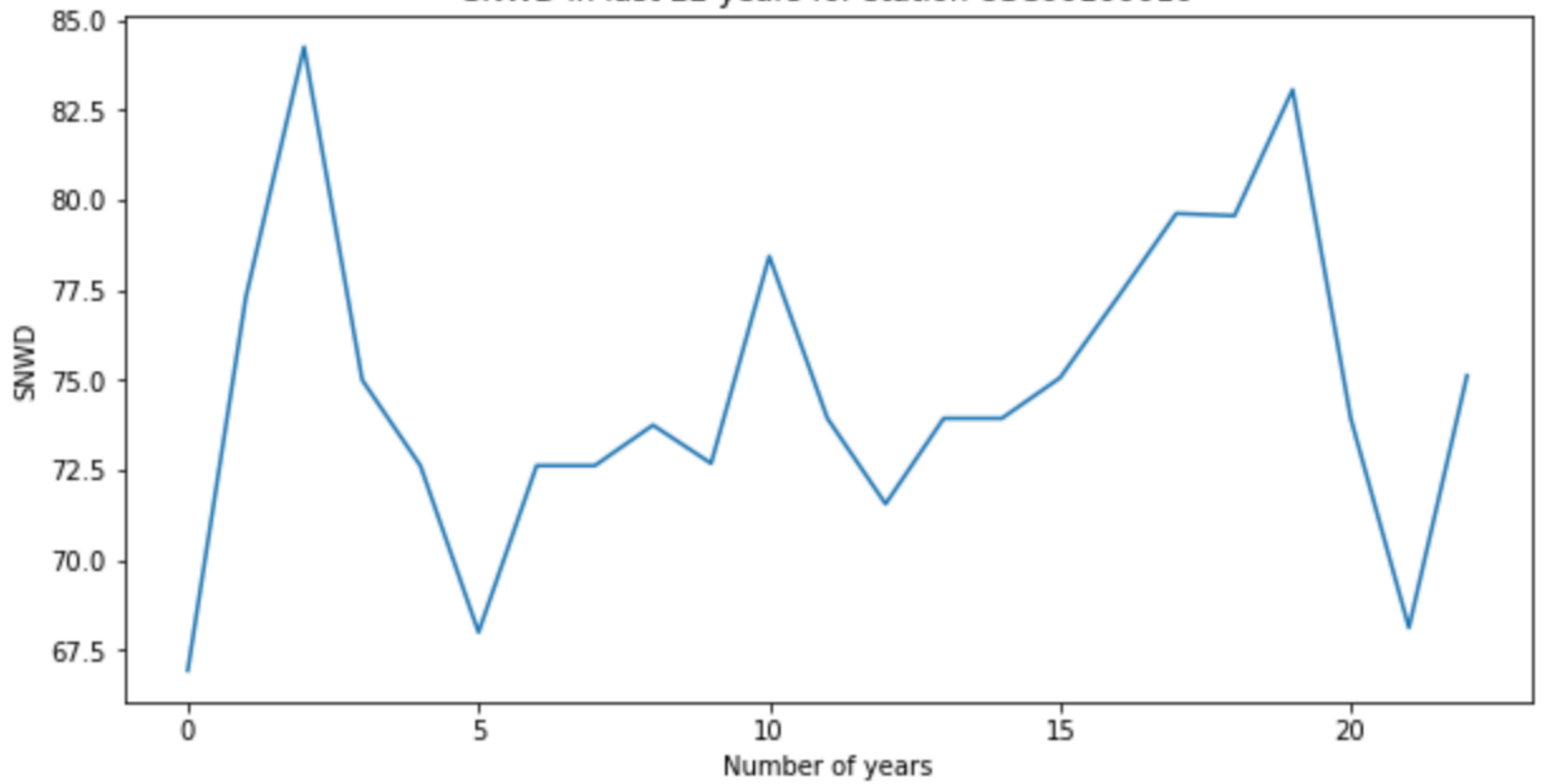
From these observations, hypothesis is proved that they have a strong correlation. Strong negative correlation in spring time where snow is melting fast with increase in minimum temperature and viceversa.

Correlation between TMin and SNWD

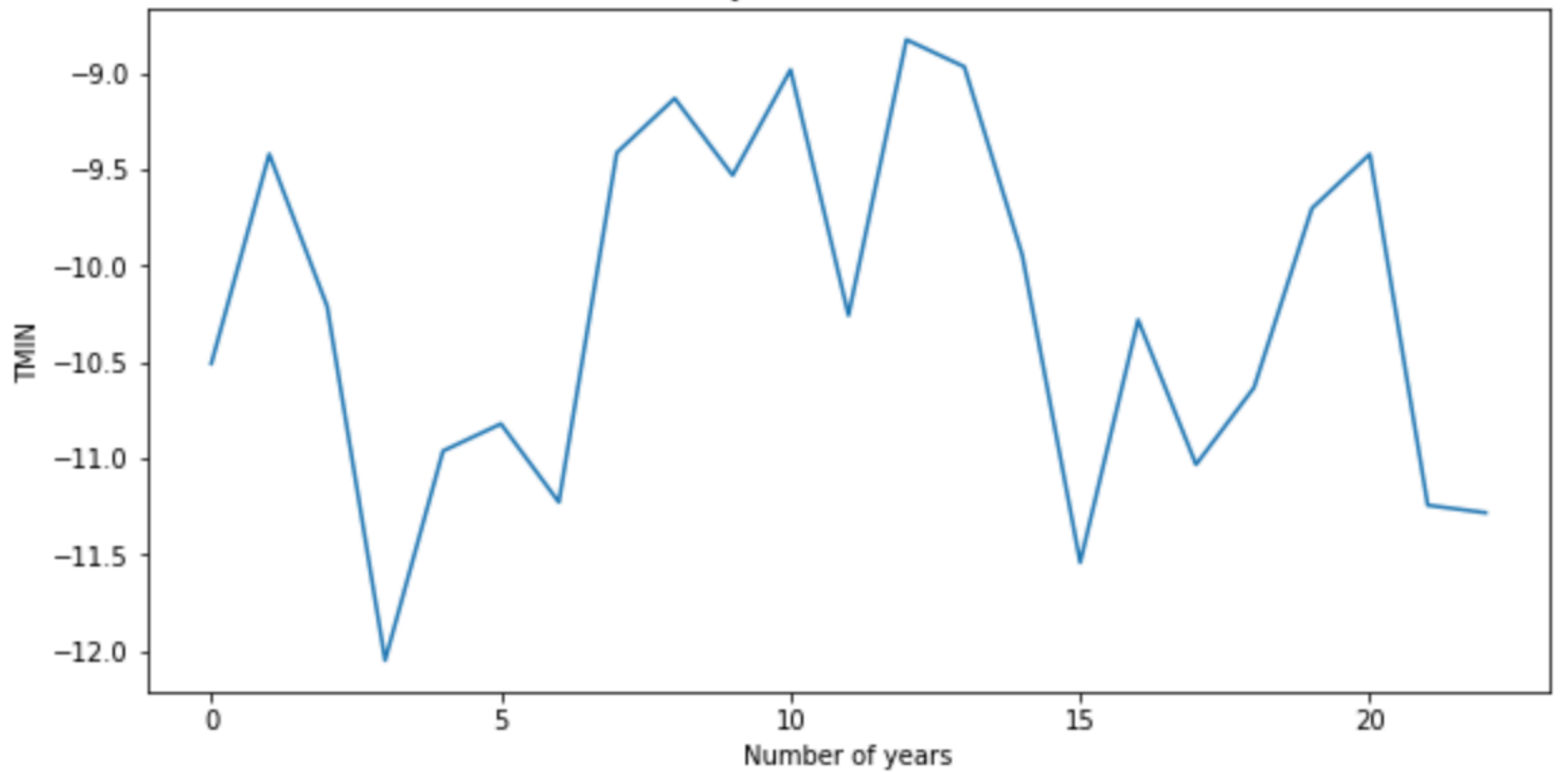


From the analysis and average plots indicated below, there was no increase in TMIN or SNWD in last 22 years.

SNWD in last 22 years for station USC00100010



TMIN in last 22 years for station USC00100010

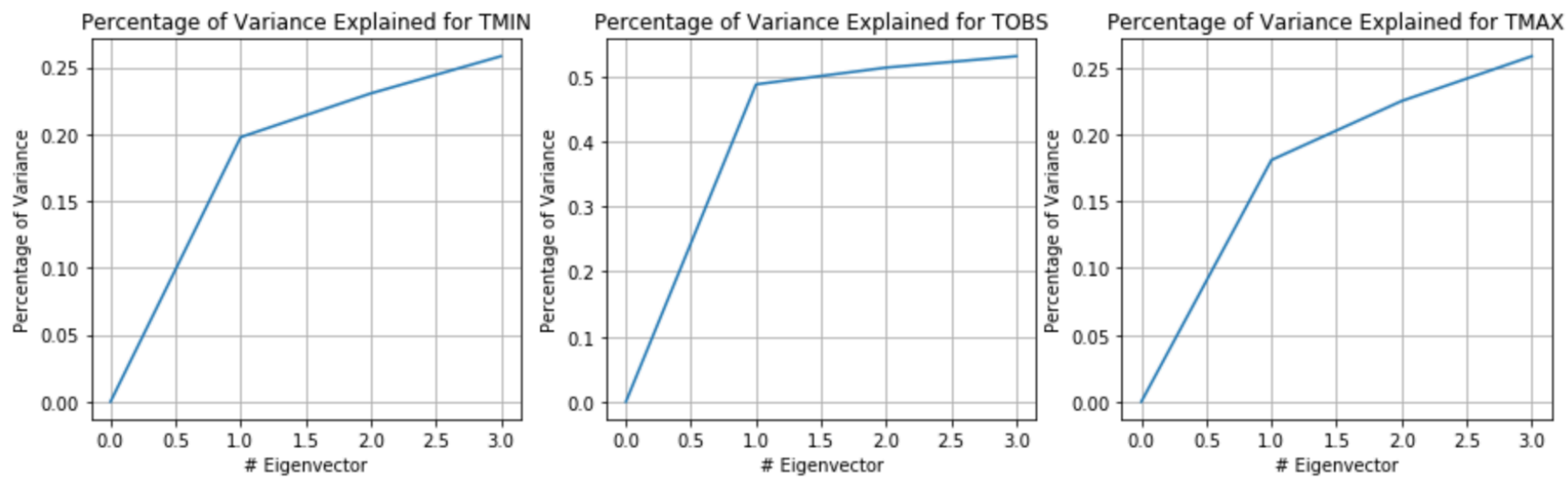


4) PCA analysis

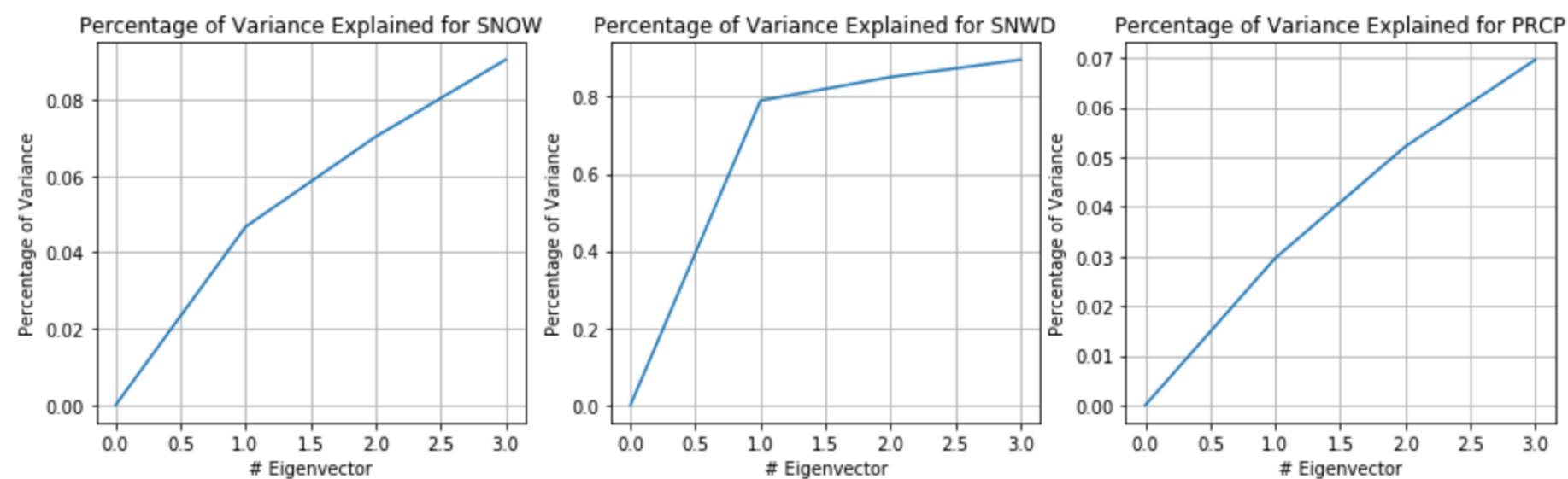
In the previous section, we have focused on statistical significance test. In this section, analysis is focused on PCA of all the 6 measurements. Objective of this section is if these measurements can be better explained with less number of dimensions.

4(a) Percentage of variance vs Principle components

Percentage of variance with respect to top principle components i.e. number of dimensions in new coordinate system is analyzed here. We see that the top 3 pc's explain 25% of variance for TMIN, 55% for TOBS and 25% for TMAX. TOBS is better explained by top 3 dimension where first principle component itself explains 50% of the variance.



For SNOW, top 3 pc's explain, 9% of variance and for PRCP, top 3 pc's explain 7% variance only. But for SNWD, first pc itself explain 80% of variance and 3 pc's explain 90% of variance.



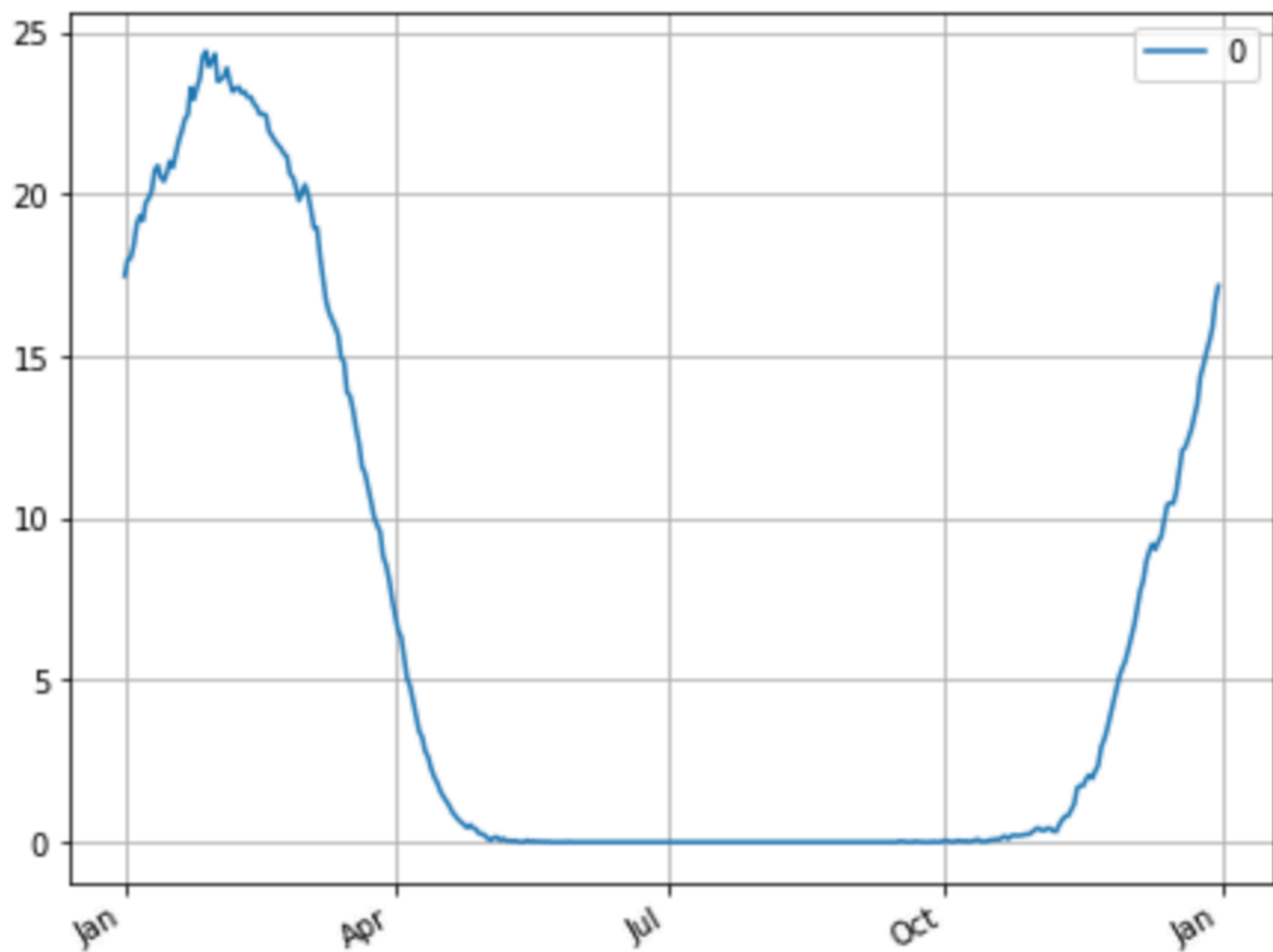
Since SNWD is decaying integral of snow its variation between days is less as well for same date on different years. Hence this is better explained with one dimension. Rest of the report focus on analysis of SNWD measurement in further detail to help explain its reconstruction behavior and residuals.

4(b) Eigen vector analysis of SNWD measurement

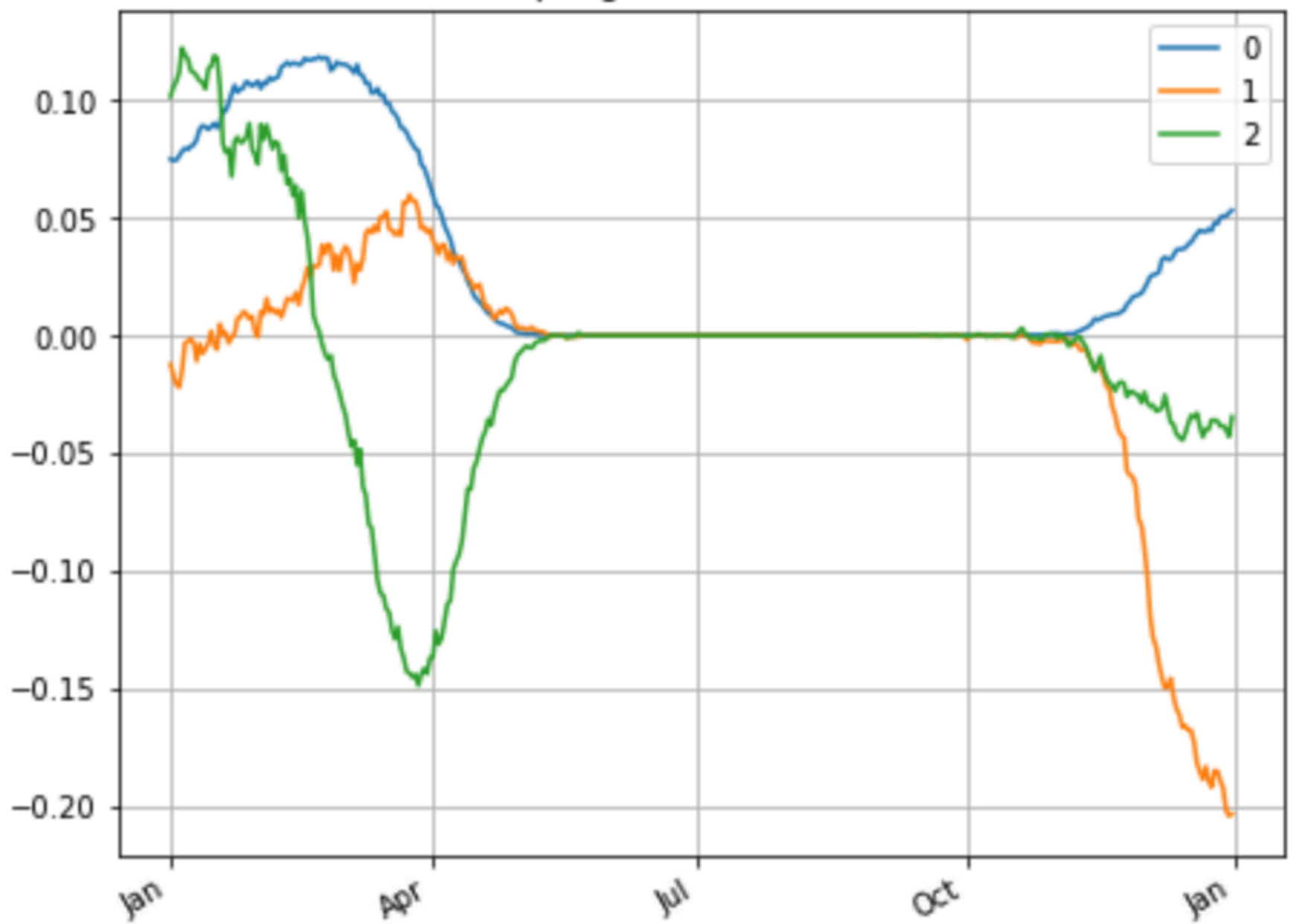
From above, we have noticed that SNWD 80% variance is explained by first eigen vector and 90% by 3 eigen vectors. So we restrict our analysis to these eigen vectors and residual analysis.

Below plot lists top 3 eigen vectors for SNWD measurements. First eigen vector represents mean of SNWD that was listed earlier in the report.

Mean SNWD in inches



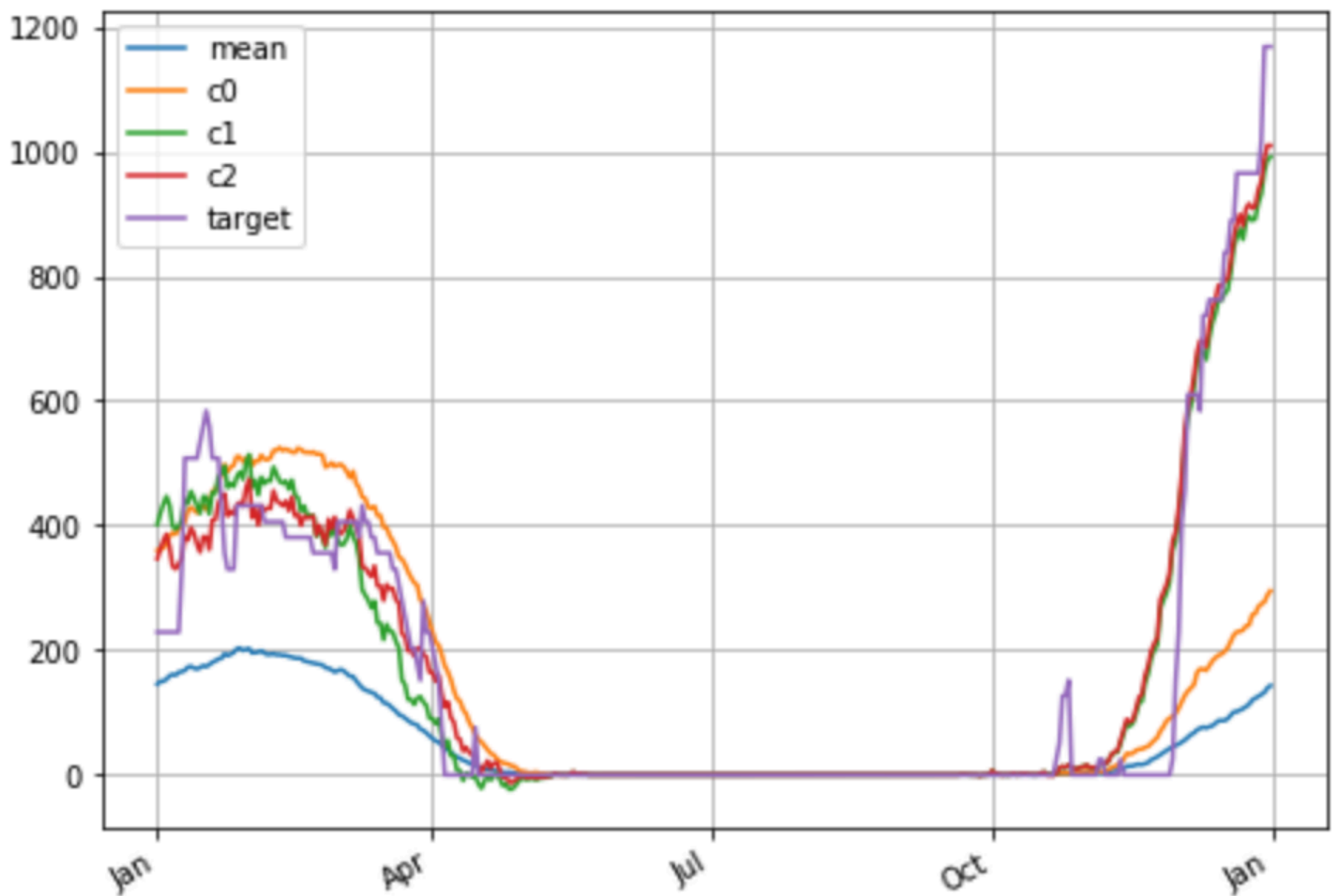
Top Eigenvectors SNWD



4(c) Reconstruction analysis of SNWD measurement

This section we show that reconstructed SNWD measurement with top 3 eigen vectors capture the most part of the measurement as shown below.

Best reconstruction



Conclusions

Based on our statistical significance test using correlation and PCA analysis, we have noticed few observations as listed below:

- There is not much change in precipitation through out the year. Hence larger dimensions are required even in projected coordinate system.
- Snow depth has a strong correlation with minimum temperature
- Not much variation of average TMIN and SNWD in last 3 decades in this region of interest.
- Performing statistical analysis using PCA technique, we observe that SNWD is described by first few principle components. Around 80% of the SNWD is captured with just one eigen vector. With 3 eigen vectors, 90% of the information is captured. Hence performing reconstruction with these top 3 principle components, most part of target signal is reconstructed with very little residual.
- Precipitation on the contrary required large number of dimensions.

