CHANGES IN SNOWFALL AND SNOW DEPTH IN THE MIDWESTERN UNITED STATES

ABSTRACT

Snowfall and snow depth measurements have decreased in the Northern Plains and in the Southern Plains of the Midwestern United States from 1970 to 2010. We perform a statistical analysis which strongly supports decreasing snow measurements in both regions, in disagreement with previously published claims. Additionally, we perform a principal component analysis which highlights three features that describe 70% of the variance in snow depth measurements in the Northern Plains and in Southern Plains during this time period, and provide a qualitative interpretation of the three most significant features.

1 Research Question

It has been claimed previously [1] that the Northern Plains has experienced a significant increase in snowfall while the Southern Plains has experienced a significant decrease in snowfall during the years 1970 - 2010 . Our research question serves to investigate meteorologist Bob Henson's claim:

Has the Northern Plains experienced a significant increase in snowfall while the Southern Plains has experienced a significant decrease in snowfall during the years 1970-2010?

2 Background

Climate change establishes the notion that temperatures on earth are steadily and ubiquitously rising. Many sources cite the status quo effects of climate change: snow-seasons are shortening, frozen ground snow is melting at increasingly alarming rates, and that average daily temperatures are rising over the years. While many of the effects of increased temperatures are sensible, others are less intuitive. For example, one study [2] found that streamflow in semi-arid basins had decreased as a result of vegetation reduction, whereas one might suspect that with less vegetation, water would not only be more free to move but would be less absorbed by vegetation. The possibility of counterintuitive effects of global warming serves as a motivation to closely examine any claims made by weather reports.

Global warming in the United States is well studied in regions of geographical interest, such as national parks, and coastal and mountainous regions. Less analysis exists of the country's central regions. We explore several claims made by Henson's online article about climate change in the Northern Plains (North Dakota, South Dakota, Minnesota, Iowa, Nebraska) and Southern Plains (Texas, Oklahoma, Kansas) [1]. The first claim that we investigate is that there is an increase in snow in the Northern Plains over the time period 1970 to 2010, while there is a decrease in snow in the Southern plains over the same time period. We also explore and address the article's claim that due to warmer temperatures, snow seasons are shortening across both plains [1].

While we already know that the northern parts of the United States experience more snowfall than the Southern parts of the United States [3], Henson's article claims that the amount of snowfall that northern regions receive has generally increased since the 1970's, while it has decreased in Southern regions since the 1970's [1]. This result is interesting, since our intuition tells us that warmer temperatures should take an equal toll on any region in the country. Henson does note, however, that "One possible explanation is that a warmer climate is allowing these normally cold areas to experience more snowy days with temperatures close to 32 degrees rather than well below freezing. Higher temperatures allow for more water vapor, which in turn supports heavier snow if conditions are otherwise favorable" [1]. Henson's article also claims that snow-seasons in the United States are shortening, and we would like to examine if that is in direct response to a warming climate [1].

3 Summary of Experimental Design

1. We extracted snowfall and snow depth data for all stations in states in both the Northern and Southern Plains which have collected snowfall and snow depth data for every year from 1970 to 2010.
2. We computed a moving average for annual snowfall with various window sizes across the time period for each region to visualize long term trends in the snow data, and perform further exploratory analysis by visualizing spatial and temporal trends in snowfall and snow depth data.
3. We used PCA on snow depth (SNWD) data to determine which components are responsible for the greatest variation in SNWD data from 1970 to 2010 across both regions, and provide an interpretation of those components.
4. We performed a Wilcoxon signed-rank test to statistically compare snowfall and snow depth in the Northern Plains to snowfall and snow depth in the Southern Plains between several time periods, namely 1970-1980 and 2000-2010.

4 Snowfall Data Exploration

Below, we plot the Northern Plains and Southern Plains moving window average mean and standard deviation for snowfall with various window sizes between 5 and 20 years to demonstrate how the differing window sizes smooth the graph. A moving average is widely used as a smoothing function to filter out noise from random short-term fluctuations. Therefore, a large window size is chosen to reduce noise in the dataset while observing long term trends between the 1970's and 2010's. These data visualizations demonstrate how snowfall has changed each year, between the years 1970 and 2010. In particular, we notice a dip which appears significant in both the northern plains and the southern plains snowfall data. It appears around 1986 and continues until around 1991, at which point the snowfall begins to increase again in both regions.

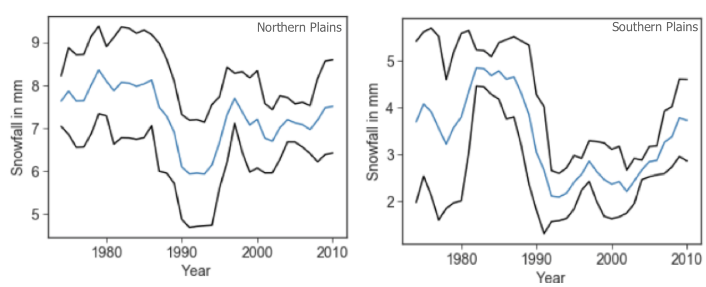

Figure 1: Plotted are mean SNOW measurements (blue) $\pm$ one standard deviation *versus* time using a 5 year rolling average in (left) Northern Plains and (right) Southern Plains.

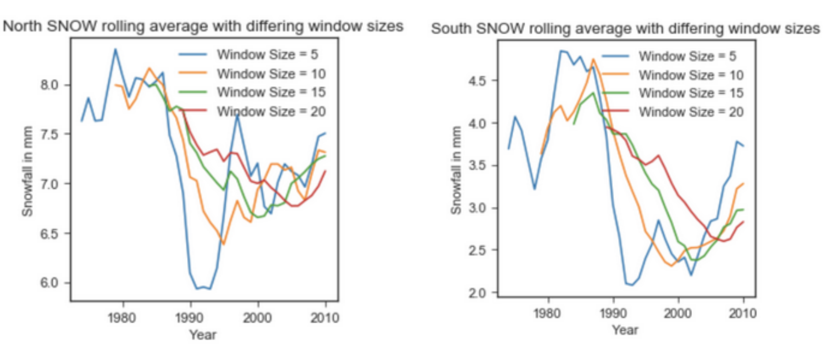

Figure 2: Northern and Southern Plains Snowfall Moving Window Average for Various Window Sizes.

The total amount of change in average winter snowfall from the 1970's to the 2010's (averaged across each state in both plains) is visualized on the map in Figure 3, which summarizes the overall snowfall trends in each state in the Northern and Southern Plains between the 1970's and 2010's. State lines were chosen as artificial boundaries to demonstrate large spatial trends in snowfall data across both regions. We notice immediately that states located more centrally between the plains experience a decrease in winter snowfall between the two decades, whereas states located further north and south experience an increase.

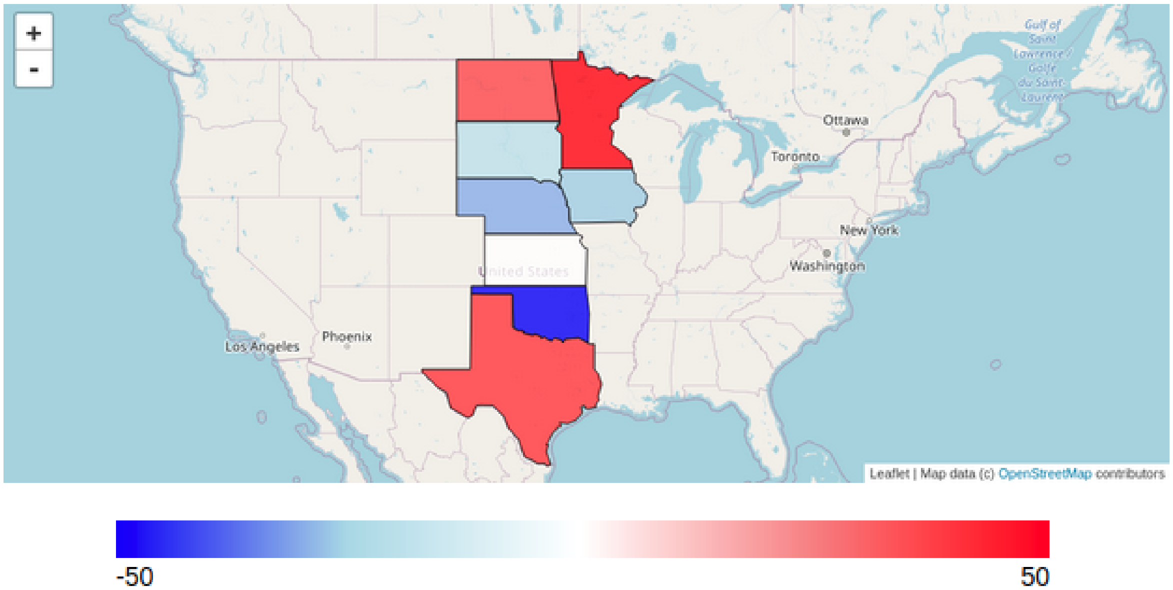

Figure 3: Change in average winter snowfall (mm) from the year 1970 until 2010

5 Principal Component Analysis of Snowfall Depth

Together, the first, second, and third eigenvectors that result from a principal component analysis (PCA) explain approximately 70% of the variance in snow depth measurements in both regions for the time period 1970-2010. We will explore these eigenvalues in greater depth below.

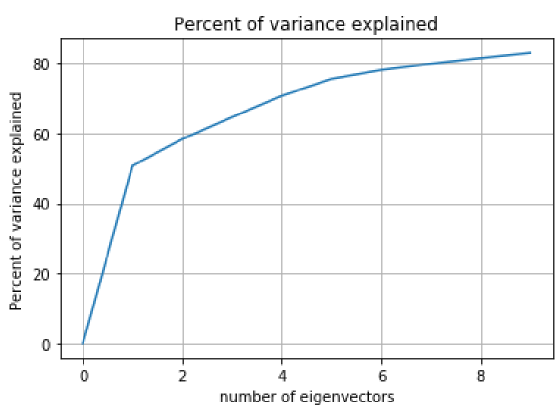

Figure 4: Percentage of variance explained versus number of eigenvectors considered for a principal component analysis of Snowfall Depth in the Midwestern United States.

5.1 Interpretation of Eigenvectors

The first eigenvector, $\vec{u}_1$, (eig1) corresponds closely to the mean daily observation $\vec{\mu}$ and quantifies a heavy or light snow winter. A higher projection of annual snow depth measurements, $\vec{v}$, onto this feature corresponds to heavier snow

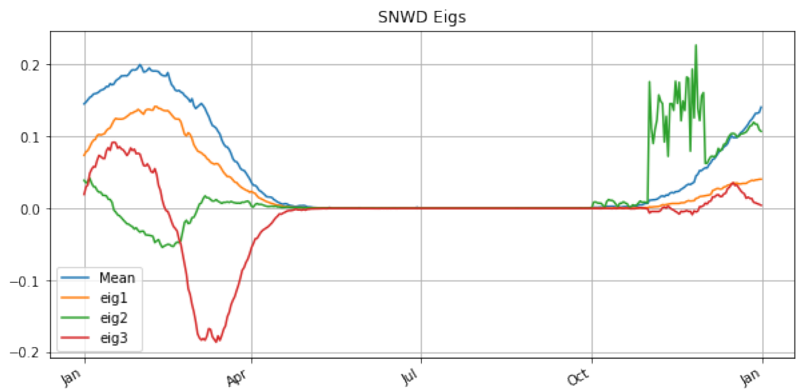

Figure 5: Basis resulting from PCA of snow depth computed for the Northern and Southern Plains together. The above plots represent the first 3 eigenvectors computed using PCA on the SNWD measurement. The blue line represents the mean snow depth, the orange line represents the first eigenvector, which represents the component which contributes to the largest percentage of variance in snow depth data. Similarly, the second and third eigenvectors represent the orthogonal eigenvectors which describe the next largest percentage of variance. The instability apparent in the second eigenvector likely resulted from inconsistent recording of snow depth measurements reported by weather stations in the Midwest.

and vice-versa, which may be computed according to $\vec{v} \cdot \vec{u}_1$. This first eigenvector explains roughly 50% of variance in snow depth data of the Midwestern United States (Fig. 4).

The second eigenvector $\vec{u}_2$ (eig2) is positive starting from October until beginning of January and later becomes negative until March. In order to interpret the meaning of second eigenvector we filtered out the measurements whose second residual is greater than a threshold value taken to be 0.1 mm/year. Residuals are the remainder left after successive approximations. The second residual is defined as follows:

$$\vec{r}_2 \equiv \vec{v} - \vec{\mu} - (\vec{v} \cdot \vec{u}_1)\vec{u}_1 - (\vec{v} \cdot \vec{u}_2)\vec{u}_2$$

After filtering, we sort the data with respect to their second coefficients. The twelve measurements with the highest/lowest coefficients are plotted in the Appendix (Fig. 8) are indicative of an early season as the target snow depth is positive starting from the month of October. We interpret the second eigenvector as indicative of a late/early winter season.

The third eigenvector, $\vec{u}_3$ (eig3) is positive starting from January until beginning of March and later becomes negative until mid of April. Similar to the second eigenvector, we interpret the meaning of third eigenvector by filtering out the measurements whose third residual is greater than a threshold taken to be 0.1 mm/year and then by sorting the data with respect to their third residuals, defined according to:

$$\vec{r}_3 \equiv \vec{r}_2 - (\vec{v} \cdot \vec{u}_3)\vec{u}_3$$

The twelve measurements with the highest/lowest coefficient 3 are plotted in the Appendix, as well 10. We interpret the third eigenvector as how long/short a snow season lasts.

The first three eigenvectors of the snow depth measurements (SNWD) considered here are strong indicators of certain trends in the region. The first eigenvector is synonymous to the mean/average snowfall depth. The second eigenvector indicates whether the snow season is early or late, with a higher second coefficient indicating an earlier season. The third eigenvector indicates whether the snow season is longer or shorter, where a higher third coefficient indicates a longer season. Analysis of SNWD data is be more meaningful when considering a finer regional granularity than the one considered here. This is exemplified in the first coefficient plotted spatially in Fig. 6.

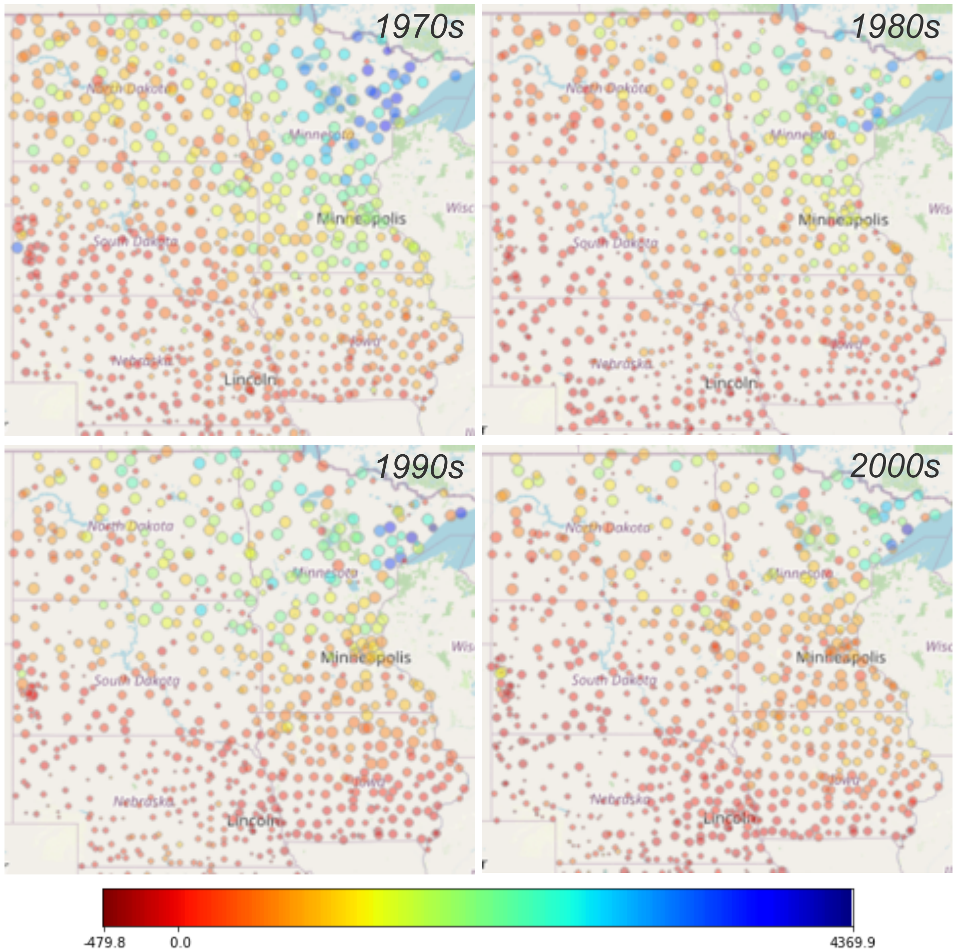

Figure 6: First eigenvector coefficient plotted by station for the Northern Plains, changing over time (top left: 1970s, top right: 1980s, bottom left: 1990s, bottom right: 2000s).

The visualization in Figure 6 depicts the first coefficient of the SNWD PCA for the Northern Plains in each of four decades. In all four of the plots we notice that weather stations experience a stronger coefficient 1 (ie. more snow depth) the further north and east we look. While the trend north is expected, the trend east is more interesting, but we cannot make any definitive conclusions from the data available. The other trend we see in Figure 11 is a clear decrease in snow depth in the most north eastern region after the 1970s, as indicated by "warmer" colors in the region in all but the upper left plot. This trend agrees with our results in Section 5.

6 Results and Discussion

In the following analysis, we computed the time average snowfall and snow depth measurements for each station in the Northern and Southern Plains from 1970 - 2010 using publicly available weather station data [3]. We then show for both regions a statistically significant distributional change in average snowfall and snow depth measurements between the several disjoint time intervals, including 1970-1980 and 2000-2010.

For each weather station in the Midwestern United States, daily measurements for snowfall (SNOW) and snow depth (SNWD) were averaged over 1970-1980 and 2000-2010. Comparing these two decades, the median weather station experienced a -2.0 millimeter median decrease in snowfall (interquartile range/IQR: -15.9mm, 5.3mm) and a -9 millimeter median decrease in snow depth (IQR: -22mm, 4mm), which was a significant decrease for both measurements ($p < 0.0001$, $U_{SNOW} \approx 3.1 \cdot 10^5$, $N_{stations,SNOW} = 1279$, $U_{SNWD} \approx 2.2 \cdot 10^5$, $N_{stations,SNWD} = 1260$)) according to a Wilcoxon signed-rank test, which is a nonparametric paired test that does not assume any of the underlying distributions are normally distributed. The test does, however, assume independent sampling, which is not valid if nearby weather stations are spatially correlated. We spatially plotted these average station values, and a clear global trend was not visually apparent. The corresponding plot was omitted for concern of brevity. A list of the stations and values considered here may be found inside the table located at `http://github.com/timtyree/dsc291team4/HW3/Data/snowfall_changes.csv`. Results were consistent according to a paired student's t-test, which makes additional assumptions of normality. Furthermore, comparative tests were repeated with various other choices for other time intervals, and the results were overwhelmingly consistent.

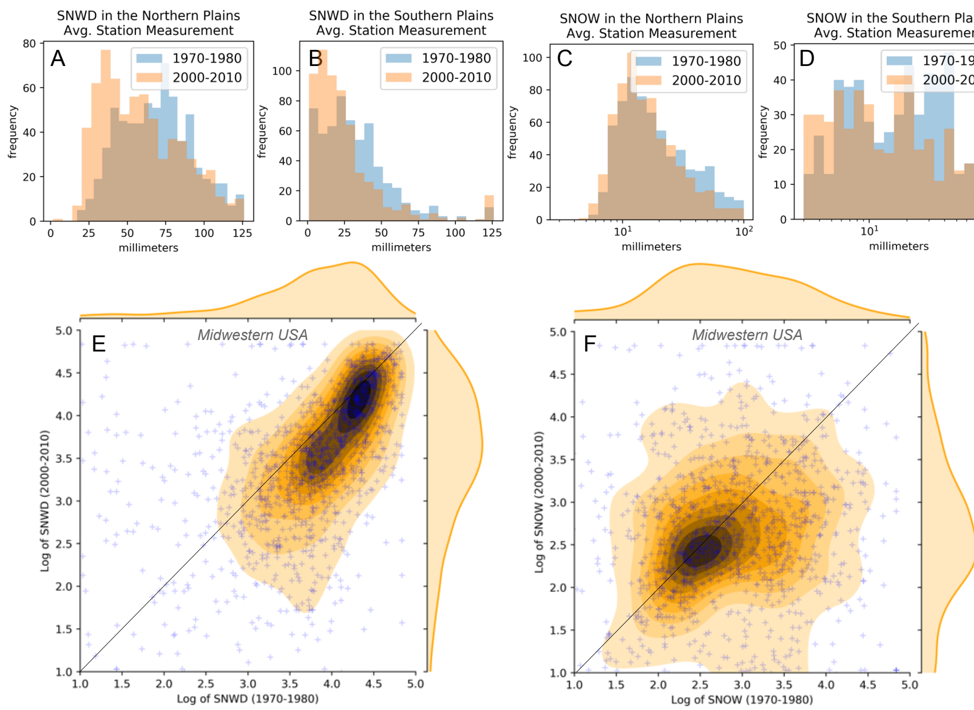

Figure 7: Changes in snow depth and snowfall between 1970-1980 and 2000-2010 in (A, C) the Northern Plains and in (B, D) the Southern Plains. Mean station measurements of (E) snow depth and (F) snowfall for the entire Midwestern United States.

In the Northern Plains, the average weather station measurement of snowfall had a median value of 14 mm (IQR: 10-24 mm) from 2000-2010, which was less than that same value measured from 1970-1980, which took a median value of 17 mm (IQR: 11-30mm), consistent with a significant distributional difference between the two decades ($p < 0.0001$, $U \approx 8.5 \cdot 10^4$, $N_{stations} = 663$). The average weather station measurement of snow depth was even more pronounced, taking a median value of 55mm (IQR: 37-78mm) from 2000-2010, which was less than that same value measured from 1970-1980, which took a median value of 69mm (IQR: 50-85mm), consistent with another significant distributional difference between the two decades ($p < 0.0001$, $U \approx 5.1 \cdot 10^4$, $N_{stations} = 662$).

In the Southern Plains, the average weather station measurement of snowfall had a median value of 8 mm (IQR: 3-22 mm) from 2000-2010, which was less than that same value measured from 1970-1980, which took a median value of 14 mm (IQR: 6-32mm), consistent with a significant distributional difference ($p < 0.0001$, $U \approx 7.1 \cdot 10^4$, $N_{stations} = 616$). The average weather station measurement of snow depth was even more pronounced, taking a median value of 19mm (IQR: 10-31mm) from 2000-2010, which was less than that same value measured from 1970-1980, which took a median value of 28mm (IQR: 15-43mm), consistent with a significant distributional difference ($p < 0.0001$, $U \approx 5.9 \cdot 10^4$, $N_{stations} = 598$).

The reader should note that the paired tests considered here make assumption that are not obviously met. We would expect nearby weather stations to be spatially correlated. Therefore, a statistical test that considers spatial correlation is needed to prove that snowfall has decreased beyond a reasonable doubt. The Wilcoxon signed-rank test assumes that each pair of observations observation is all drawn the same population both randomly and independently. It is hard to argue that weather stations are placed randomly and independently. The Wilcoxon signed-rank test also assumes that within-pair comparisons are robust to any measurement uncertainty, which is reasonably met here due to the long period of time averaging.

Overall, the Wilcoxon test is used here because it does not assume that the data is normally distributed. The test does assume, however, that the data is independently distributed, but in reality, the data from the weather stations is not likely to be independent, as stations that are closer to each other will have correlated measurements. Future work will need to be done to determine whether or not that assumption holds in order to provide statistical evidence which would refute the claim made previously [1].

7 Conclusion

The Northern Plains has experienced a decrease in snowfall from 1970-1980 to 2000-2010, refuting previous claims [1]. In fact, both the Northern Plains and the Southern Plains experienced a significant decrease both in snowfall and in snow depth when comparing 1970 - 1980 against 2000 - 2010, when comparing the time intervals 1980 - 1985 against 1990 - 1995, and when comparing a variety of other year intervals. This work comes with the caveat that the weather stations considered here cannot truly be considered independently drawn from the same population, and a deeper consideration of spatial auto-correlation between weather stations is needed. If we consider different weather stations as independent observations, then our work provides statistically significant evidence that overwhelmingly supports a decreasing snowdepth in the Midwestern United States. Additionally, our principal component analysis produces three principal eigenvectors which accounts for 70% of the variance in snow depth data for the time period 1970 - 2010. We interpret these eigenvectors as indicative of the following features, respectively: heavy/light snow, early/late snow season, and short/long snow season length.

References

[1] Henson, Bob. *Snow Season is Getting Shorter Across Most of the U.S., New Analysis Finds.* The Weather Channel. February 2020. https://weather.com/safety/winter/news/2020-02-05-snow-season- shorter-us-climate-central-study. Accessed May 2020.

[2] Guardiola-Claramonte; Peter A. Troch, David D. Breshears, Travis E. Huxman, Matthew B. Switanek, Matej, Durcik, Niel S. Cobb. *Decreased streamflow in semdi-arid basins following drought-induced tree die-off: A counter-intuitive and indirect climate impact on hydrology.* Science Direct. June 2011. https://www.sciencedirect.com/science/article/abs/pii/S0022169411004264

[3] National Climate Data Center. https://www.ncdc.noaa.gov/snow-and-ice/daily-snow/ Accessed May 2020.

[4] NOAA National Centers for Environmental Information May 2020. *nClimDiv Maximum and Minimum Temperatures.* Accessed May 2020.

A Largest and Smallest Principal PCA Coefficients

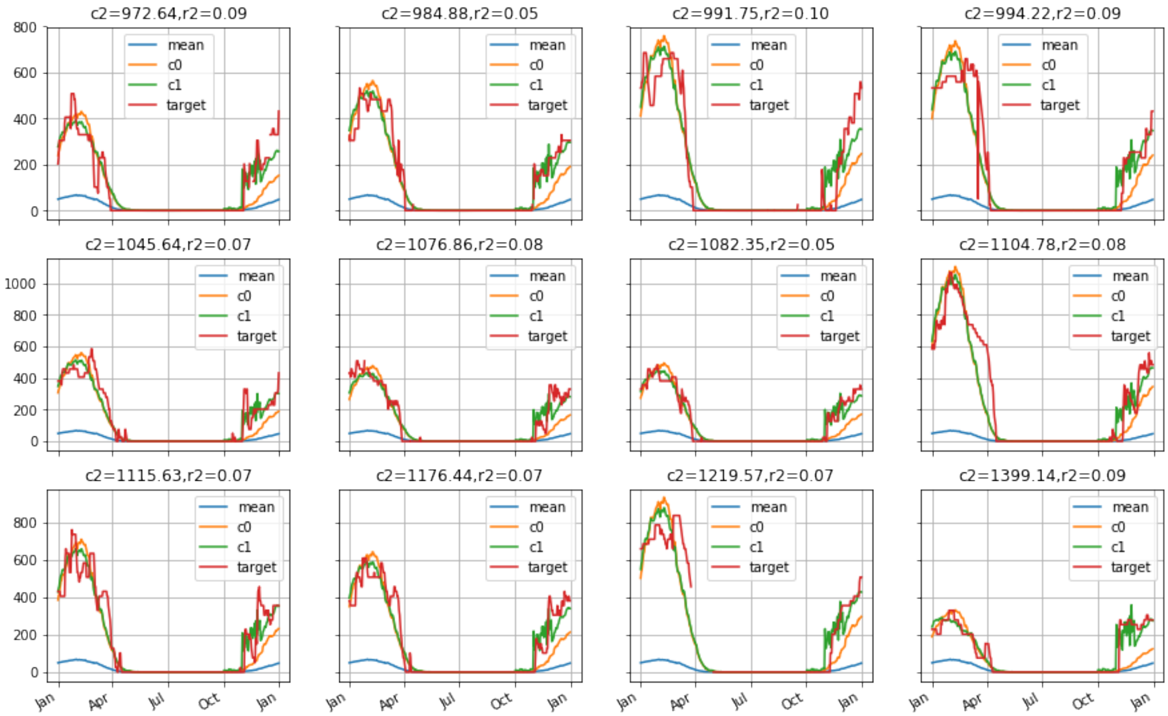

Figure 8: Measurements with largest coefficient #2.

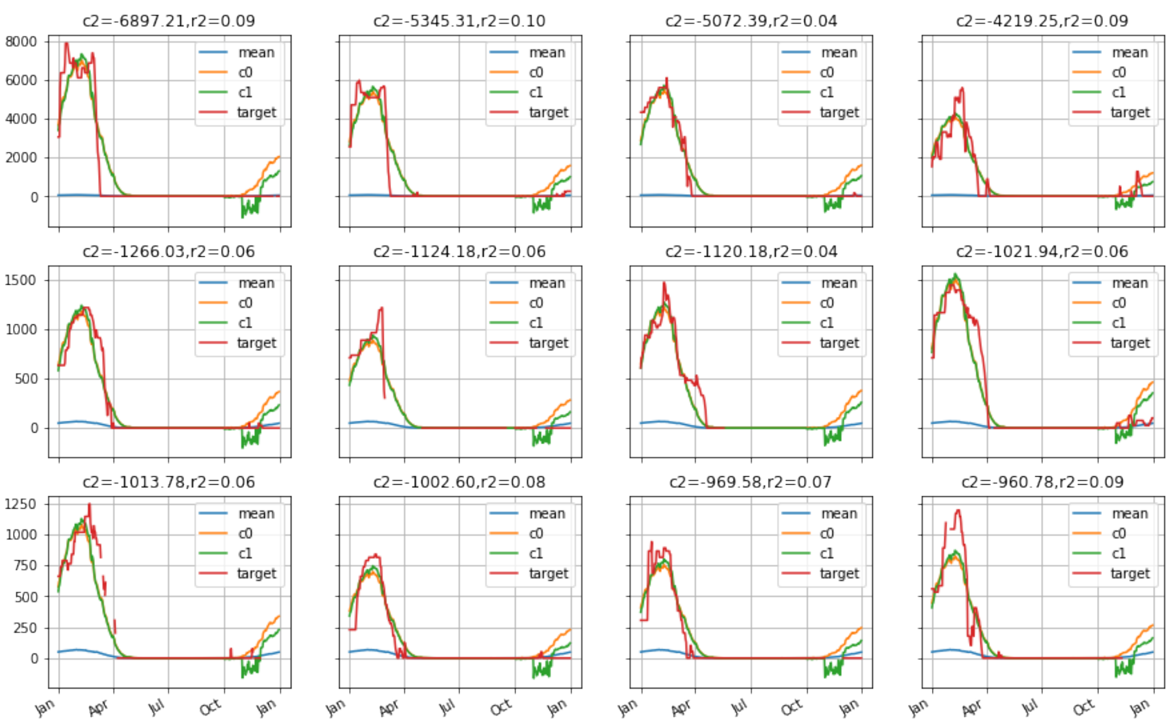

Figure 9: Measurements with smallest coefficient #2.

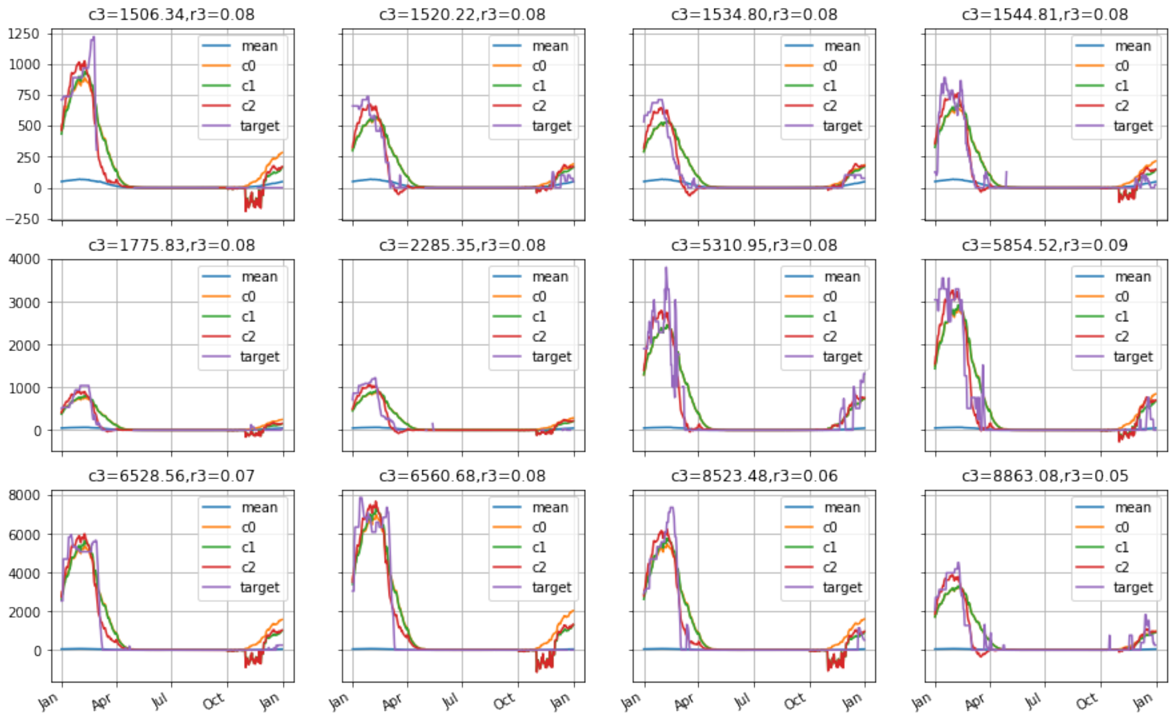

Figure 10: Measurements with largest coefficient #3.

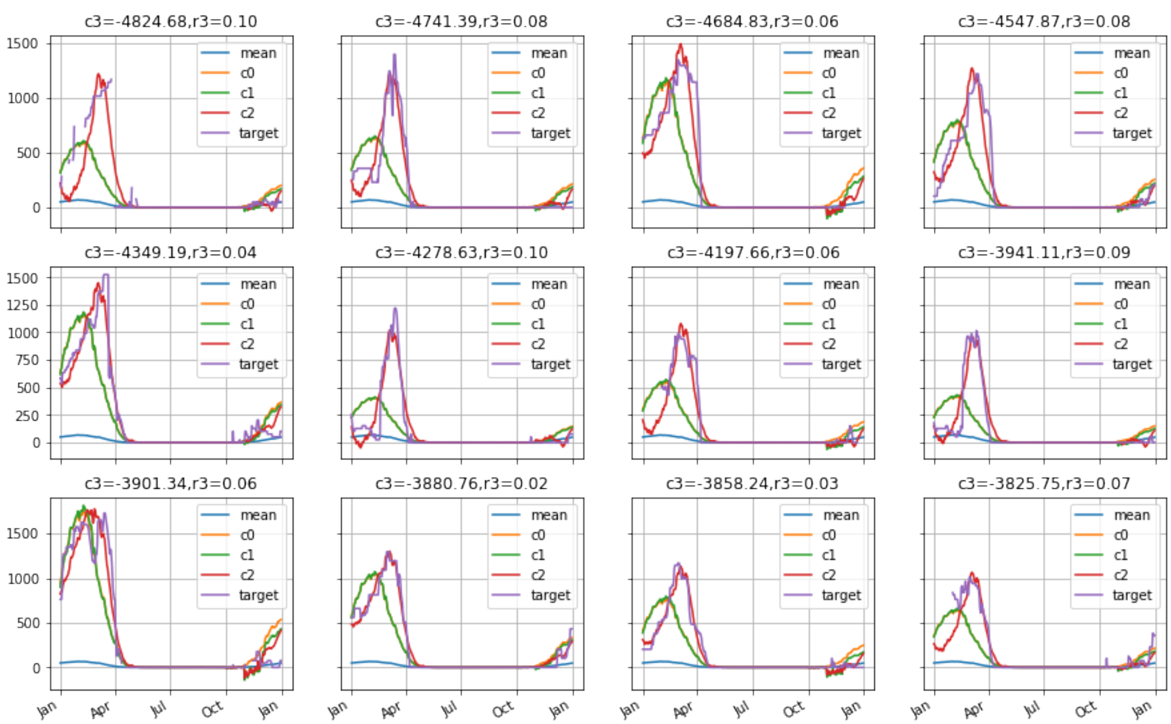

Figure 11: Measurements with smallest coeffient #3.