The Effects of El Niño on Precipitation Patterns in the United States

Proposal: *Writing a plan: specifying the background, question and experimental design for the project.*

Background: El Nino is a large-scale ocean-atmosphere climate interaction characterized by warm, nutrient poor water, migrating from the western equatorial regions of the pacific to the coast of south America. The migration of the warm water is a direct result of the weakening of the trade winds that in normal years push the water into the Western Pacific and stimulate upwelling of cold, nutrient rich water off the South American Coast. The warmer water influences global climate patterns. A typical El Nino year is associated with an increase in rainfall across the southern band of the United States as well as parts of South America. Increased likelihood of drought in the Western Pacific (particularly Australia), often with accompanying brush fires, also tends to be correlated with El Nino.

El Nino is a quasi-periodic system with a peak spectral density around 5 years and with a range of between 2 and 7 years.

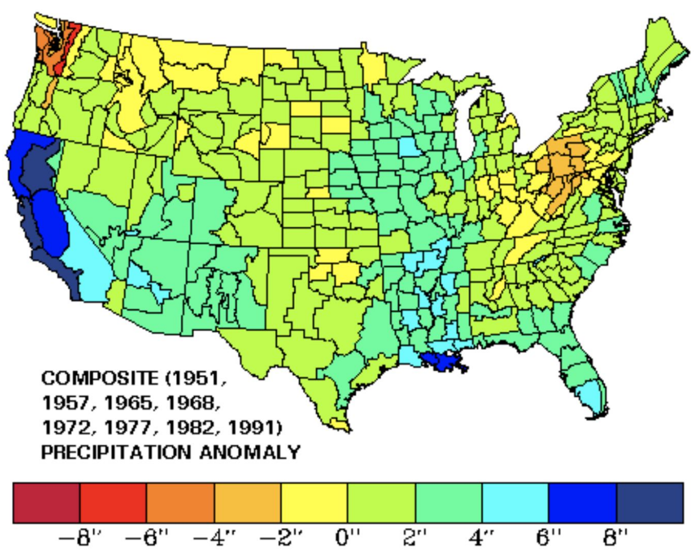


Fig 1: The precipitation anomaly of El Nino compared to non El Nino years. The anomaly is defined as the difference between the long run average and the average over the El Nino years specified. The southern band across the U.S, particularly California, see much increased rainfall while Washington State sees drier conditions. This is due to the bending of the jet stream towards the south, changing the pattern of winter storms.
http://tornado.sfsu.edu/geosciences/elnino/elnino.html

Goal/Question: What is the overall effect of El Nino on the weather of the USA? More specifically, can we detect significant changes in rainfall and snowpack in confirmed El Nino years? Are these changes overall positive or negative (more or less rain/snow)? Do the changes include variations in the pattern of rain/snowfall—e.g. do El Nino years correspond with an earlier/later snow season? Heavier but more infrequent storms?
The end product will be two items:

1. A map of areas/stations in the US that are affected by El Nino showing the spatial and geographic diversity of its impact.
2. Analysis to find and quantify the variables that change in a predictable and in a statistically significant manner during an El Nino year for those stations that are affected. Does El Nino correspond to changes in rainfall, temperature, snowpack? Is it just the amount that is changed or are seasons shifted earlier or later? Can we quantify the variability due to El Nino in the rainfall for a given year in California?

Hypothesis:
From Fig 1 we see that there is a well documented pattern of increased rainfall in CA and the southern band of the U.S as well as decreased rainfall in western Washington State. We would expect to be able to reproduce these results from the weather station data.

From our exploratory data analysis (see section *Preliminary Analysis*) we find that there is some correlation of El Nino conditions with higher rainfall and La Nina conditions (opposite of El Nino) with lower rainfall in TX and CA. This is an encouraging sign that the weather data provided matches our hypothesis that there is a correlation between rainfall and El Nino.

Other variables include snowfall and snow depth which are proxies as well for an increase in winter storms in the southern part of the U.S.

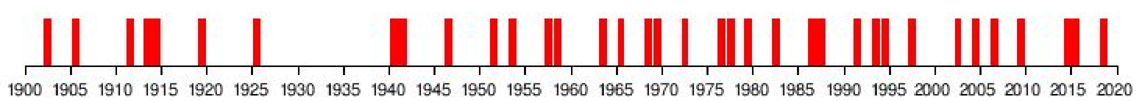


Fig 2: The red bars indicate verified El Nino occurrences over the last 100 years. The strongest El Ninos of recent years have been 1982-83 and 1997-98.

Experiments

- Create a map of the USA giving difference in average rainfall for El Nino years given by Fig 2. We should be able to reproduce Fig 1.
- Analysis of stations in CA for El Nino years and non El Nino years to identify whether there is a statistically significant difference between the two. The first order experiment will be to look at yearly averages of rainfall, snowfall, temperature, and wind. We can also look at cumulative snowfall throughout the year. Since the impact of El Nino is often stronger during the winter season, we may see a stronger signal in looking at the total snowfall across the two-year period of 1982-83 and 1997-98, as opposed to the years individually.

- Second order experiments would be to look at higher order moments to see if there is increased volatility that could be caused by El Nino without skewing the average.
- PCA analysis. PCA takes in data matrix X, an n x p dimensional matrix, to give p eigenvectors of the p x p dimensional covariance matrix X.T X (assuming 0 mean).
  - Average over the stations in a region such as california and then fold the data over by year. This would give us X with n = 365 and p being the number of years in the dataset. Thus X.T X would give us a symmetric p x p matrix whose eigenvectors would correspond to linear combinations of different years e.g., c1 1940+c2 1941+ c3 1942. This could give us an idea of how much of the yearly variation is due to El Nino if we could find a vector corresponding to just the El Nino years and calculate percentage variance explained.
  - For a single variable such as precipitation we have n = days of the year (365), p = # of stations. Thus we can find linear combinations of stations giving us explainable variances. The first eigenvector should correspond to the mean. I'm not sure what to expect but we could possibly see an El Nino effect here. This will be exploratory.

Preliminary Analysis:

In this section we present the evidence we have for the promise of different variables to explain variations caused by El Nino. The idea of this exploratory analysis is to try to discount variables that may not merit further investigation.

1. **Precipitation** is the most well known of El Nino's effects on the U.S. To probe the correlation we extracted the Oceanic Nino Index (ONI). ONI is one of the primary indices used to monitor the El Nino-Southern Oscillation (ENSO). The measurement of ONI is as follows. First, collect the average sea surface temperature in the Nino 3.4 region for each month. Next, average it with values from the previous and following months, which is then compared to a 30-year average. The observed difference from the average temperature in that region is the ONI value for that 3-month "season."

   In Figure 3 we show ONI and yearly averaged precipitation (PRCP) plotted against one another in CA (Upper) and TX (Lower). Although we find only a weak correlation (<0.35) it is strong enough to merit further analysis. Also, we suspect that there is a highly non-linear relationship between PRCP and ONI and thus we may want to use mutual information or some other kind of non-linear analysis technique going forward. The conclusion is that precipitation shows promise for future analysis and we could even disprove the commonly held notion that there is a statistically significant relationship between ONI and El Nino.

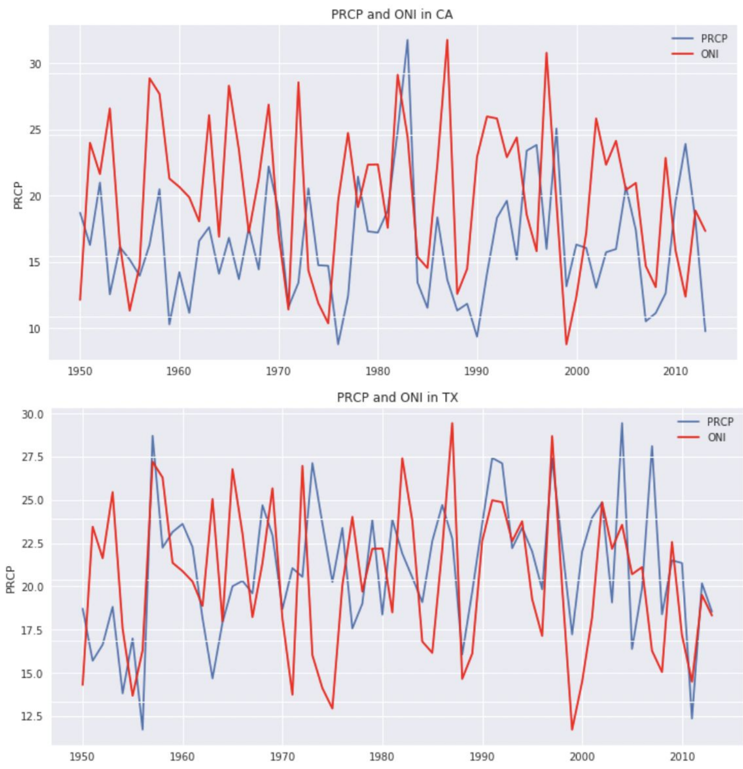
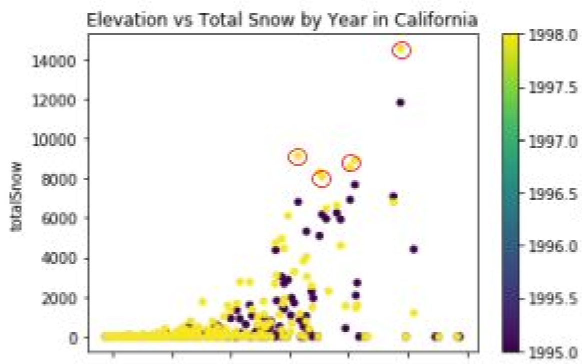


Fig 3. The Oceanic Nino Index (ONI) is one of the primary indices used to monitor El Nino. Plotted is the relationship between PRCP and ONI in CA (upper) and TX (lower). The correlation coefficient between yearly average PRCP and ONI: CA: 0.134, TX: 0.342
Cosine similarity between yearly average PRCP and ONI: CA: 0.033, TX: 0.058
Our conclusion from this analysis is that there is a weak correlation between precipitation and ONI.

2. **Snowfall** is a measurement of the amount of snow that has fallen in a unit of time (usually taken every 6 hours or so). Therefore the cumulative snowfall can tell us how much snow has fallen, compared to snow depth which is a function of both the snowfall and the amount of melting over a given period. Snowfall is complementary to the rainfall data since both measure the amount of actual precipitation. Preliminary analysis, Fig 4a/4b, hint that there are some interesting observations to be gleaned from analyzing snowfall data. For instance, in Fig 4a we see that some stations did indeed experience higher cumulative snowfall in the El Nino year than the non El Nino year. Another observation from Fig 4b is that the snow depth in Feb is much higher for 1998 indicating a possible early season compared to 1995.



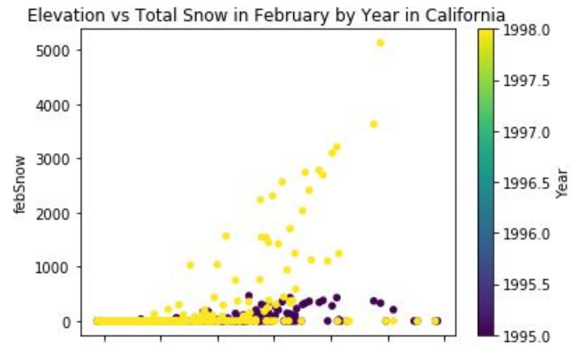


Fig 4a. Plot of the cumulative snowfall (mm) vs elevation across all stations in California in 1998 (one of the strongest El Nino years) and 1995 (not El Nino year). As we look at higher elevations for stations, we see a more noticeable difference in cumulative snowfall in the 1998 vs 1995 (circled in red). This indicates that for future analysis, finding El Nino signals or effects may be easier to detect in higher elevations.
Fig 4b. Plot of cumulative snowfall (mm) vs elevation across all stations in California in the month of February in 1998 and 1995. The single month focus highlights the seasonal increase in snowfall in February. The remaining snowfall during 1998 tended to follow the yearly average.

3. **Snow Depth** is a measure of the depth of the snow that has fallen and stuck to the ground over time. It is measured in the dataset in millimeters. It therefore is a function of both the amount of snowfall and the temperature.

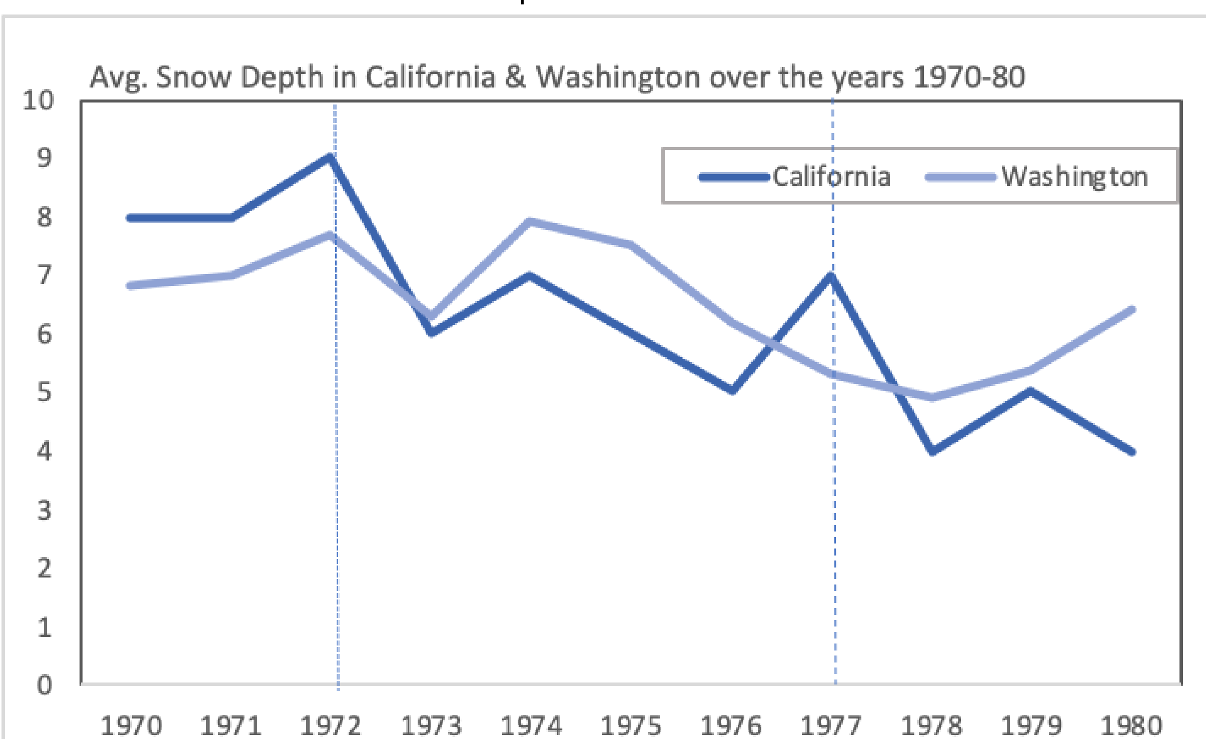


Fig 5. Total average snow depth (inches) over a year for the years 1970-1980 for CA and WA.

In order to assess if there's an increase in snowfall, we conducted exploratory data analysis on snow depth during the decade 1970-1980 across 2 stations each in California and Washington. The two stations had approximately the same elevation.
The graph above plots the average depth of snow across for both CA and WA for this decade. From this graph we can glean 2 insights:

(1) Comparing el nino years to non-el nino years: 1972, and 1977, we can see a slight increase in the depth of the snowpack, when compared to the non-el nino years.

To see if this difference was significant, we conducted an independent samples' t-test comparing the means of el nino years and la nina years to find that this difference is only marginally significant at $p=0.076$. This might be because we are comparing the data of 2 years to the other 8 years in this decade. However, this is worthy of further investigation across all years.

(2) Comparing California to Washington: Washington has a smoother curve with less drastic changes. On the other hand, California's curve is sharper.

To see if this difference between California and Washington was significant, we conducted an independent samples' t-test comparing the means of CA to WA to find that this difference was not significant at $p<0.05$.

4. **Average Temperature** could also give us some insight into the effects of El Nino.

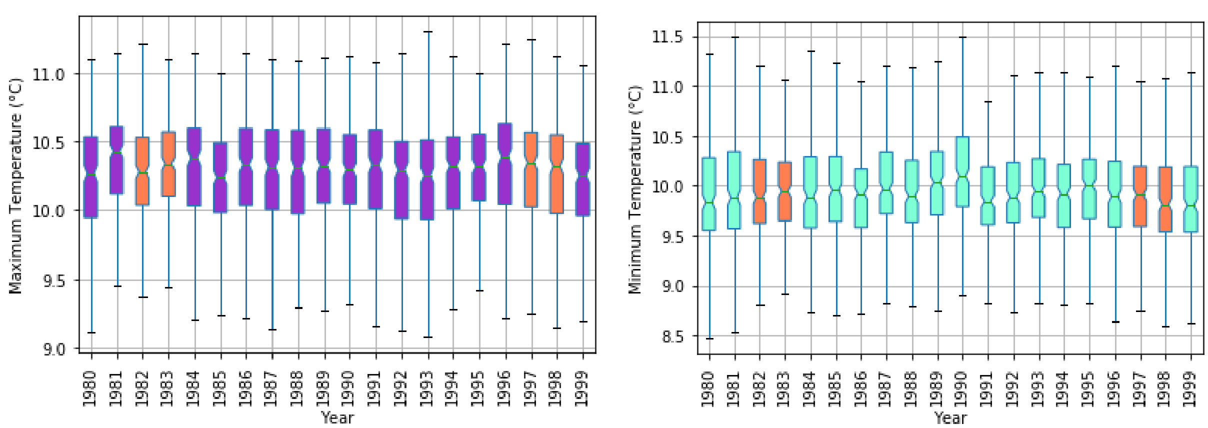


Fig 6. The above two graphs plot how maximum (left) and minimum (right) temperature data changes across 1980-2000 in the CA area. The maximum and minimum temperature data are collected per day throughout the 20 years range by each of 200 different stations in the CA area, which allows the comparison across years valid and unbiased. The box mean and

standard deviation are computed over the 200 station's yearly average maximums and minimums. El Nino years are highlighted in orange.

It is noticeable that the deviation upper bound of TMAX and the deviation lower bound of TMIN of EN years (1982-1983, 1997-1998) is clearly higher than neighboring years, which might suggest an up-toward shift and disturbance in temperature affected by EN. Such kind of changes may be indicative of EN analysis. It is unclear, however, if this is a promising direction of research since the variation seen is fairly minimal.

5. **Other Weather Measures:** We also explored Evaporation of water from the pan (EVAP = tenths of mm), Daily Total Sunshine (TSUN = minutes), and Daily percentage of possible sunshine (PSUN = %). However, the data was sparse and it was hard to identify trends across time. Therefore, the EDA of these measures has not been included in this report.

Conclusion:

In conclusion, we are proposing to look at how El Nino affects the weather, principally precipitation and snowfall, across the US with a particular focus on California. Our exploratory analysis has given us an idea of the scope of the problem and we have seen a few promising first results. The focus going forward will be to tease out the relationships hinted at in the data, whether by more rigorous analysis or better data selection.