El Nino and Atlantic Hurricanes

Matt Donahue

For my final ENSO math project, I set out to explore the relationship between El Nino (Pacific Ocean surface temperature) and the Atlantic hurricane season behavior. There has been some related work done on this topic; namely, it is established that there is a relationship between how many Atlantic/Pacific hurricanes there are and El Nino.

	Atlantic		Eastern Pacific	
	Average	El Niño Avg.	Average	El Niño Avg.
Named storms	9.4	7.1	16,7	17.6
Hurricanes	5.8	4.0	9.8	10.0
Intense Hurricanes	2.5	1.5	4.8	5.5

(1. Interaction with El Niño: How Hurricane Frequency May Be Affected)

This figure was taken from a project conducted by University of Illinois called, 'The Weather World 2010 Project.' The conclusion is that El Nino conditions cause more hurricanes in the Eastern Pacific and fewer hurricanes in the Atlantic. This result is consistent with what we know about how hurricanes form.

Waves of low pressure travel from Western Africa over the Atlantic. Here, the dry air starts to pick up moisture, which rises and loses buoyancy as the pressure drops, and its upward flow is accelerated. This process feeds itself as the system moves across the Atlantic, eventually either making landfall in the US or Mexico, or veering northward, where it encounters cool water and stops being able to feed itself.

For this process to create a hurricane, the air column needs to be stable enough that the warm, moist, rising air is able to support the air above it. If there is significant wind difference between the surface level and higher in the atmosphere, the moist air stops being able to support itself and the storm is unable to continue to grow.

In El Nino years, there are more severe weather systems that travel across the southern United States. These systems travel out into the Atlantic, where they can encounter a hurricane or tropical storm, and cause vertical wind shear (difference between surface and high-altitude wind), subduing the storm.

With these findings in mind, I decided to see if I could recreate these findings using publicly available hurricane data and El Nino observations. I also wanted to know if I could find a more detailed relationship between El Nino and Atlantic hurricanes. I had a few general questions to use as a starting point.

1. Does Nino 3.4 affect the severity of hurricanes? It would seem that if strong El Nino years hinder the formation of hurricanes, maybe that also suggests that during La Nina, the storms are capable of becoming more severe. How can we quantify this? In general, to measure a storm's severity, it will be useful to look at the maximum wind speed of a hurricane and the minimum air pressure in the center of a hurricane. These two measurements are very highly correlated, because the pressure differential between the eye of the hurricane and the outside is what directly causes the windy circulation of the hurricane.

It would also be prudent to look at how much damage a storm did monetarily. However, that data is pretty sparse and the landfall location of a storm probably determines that more so than wind speed (Two of the most damaging hurricanes of all time, Katrina and Sandy, were category 3 and 1, respectively, when they made landfall).

(4. The Ten Most Damaging Hurricanes in U.S. History.)

The other metric we might use to measure this is what proportion of named storms became full-blown hurricanes (pun intended)? Some storms are significant enough to be named Tropical Depressions or Tropical Storms but fail to attain hurricane status.

2. Does Nino 3.4 affect the hurricane paths? More specifically, does it affect whether a hurricane makes landfall or not? This question was inspired by this year's hurricane season. Based on this class, it seems like this year has been leaning La Nina. We have had an astounding number of named storms in the Atlantic, but strikingly, very few landfalls. So maybe La Nina increases the number but decreases the landfall?

We can test this a number of ways. One obvious path is purely by looking at the number of storms that made landfall in a given year and checking if that correlates with Nino 3.4. Additionally, we can look at what percentage of storms made landfall. That might be more useful, since my initial hypothesis is that La Nina causes more hurricanes but less frequent landfall. This would mean that counting number of landfalls alone might not be as revealing.

3. What month of Nino 3.4 has the most predictive power for the hurricane season? Is it most important to know what the Nino value is during hurricane season? At the beginning of hurricane season? Last year? I want to test correlations with every month, just in case one happens to be a lot better at predicting information about hurricanes.

4. Can we reproduce the results found in the University of Illinois project with regards to Atlantic hurricanes? This should be a fairly simple process; we just need to separate years/ months into Nino, Neutral, and maybe Nina, then see how many named storms and how many hurricanes there are in each.

DATA

Diving into the data I used for this project, we first have El Nino observations. Specifically, I plan to use Nino 3.4 observations, which is a measure of Sea Surface Temperature (SST) in the Mid to East Pacific. I will use the measurements of Nino 3.4 for each month over 40 years or so. The data sets start in 1980 and include El Nino 3.4 values for each month.

http://iridl.ldeo.columbia.edu/home/.tippett/.scratch/.APMA4990/.n34/#info

To separate out this data into what I need, I first extract the data from the format in which it is stored and immediately subtract its average value from it. Most of the measurements are in degrees Celsius and inhabit the range of 24-28 degrees Celsius, so taking 26-ish degrees off of those values will center our measurements around 0. This doesn't make a difference for the actual correlation values, but it's generally easier to analyze data by looking at the anomaly than it is to analyze data by looking at the absolute value.

The other important thing to note is that when I took off the average, I took off the 1980-2010 average, rather than the 1980-2020 average. Again, this practice shouldn't affect the correlation values, but it is a commonly accepted convention in climate science.

The second data set I used was that of Atlantic hurricanes. The data set called HURDAT2 from NOAA's National Hurricane Center, to be specific. The format of this data is a little bit difficult to understand at first, but here is what a few lines look like:

```
AL252005,                WILMA,     48,
20051015, 1800,  , TD, 17.6N,  78.5W,  25, 1004,    0,    0,    0,    0,    0,    0,    0,    0,    0,    0,    0,    0,
20051016, 0000,  , TD, 17.6N,  78.8W,  25, 1004,    0,    0,    0,    0,    0,    0,    0,    0,    0,    0,    0,    0,
20051016, 0600,  , TD, 17.5N,  79.0W,  30, 1003,    0,    0,    0,    0,    0,    0,    0,    0,    0,    0,    0,    0,
20051016, 1200,  , TD, 17.5N,  79.2W,  30, 1003,    0,    0,    0,    0,    0,    0,    0,    0,    0,    0,    0,    0,
20051016, 1800,  , TD, 17.5N,  79.4W,  30, 1002,    0,    0,    0,    0,    0,    0,    0,    0,    0,    0,    0,    0,
20051017, 0000,  , TD, 17.4N,  79.6W,  30, 1001,    0,    0,    0,    0,    0,    0,    0,    0,    0,    0,    0,    0,
20051017, 0600,  , TS, 16.9N,  79.6W,  35, 1000,    0,    0,   40,    0,    0,    0,    0,    0,    0,    0,    0,    0,
20051017, 1200,  , TS, 16.3N,  79.7W,  40,  999,    0,   60,   40,    0,    0,    0,    0,    0,    0,    0,    0,    0,
20051017, 1800,  , TS, 16.0N,  79.8W,  45,  997,   30,   60,   60,   30,    0,    0,    0,    0,    0,    0,    0,    0,
20051018, 0000,  , TS, 15.8N,  79.9W,  55,  988,   60,   60,   60,   60,   20,   20,   20,   20,    0,    0,    0,    0,
20051018, 0600,  , TS, 15.7N,  79.9W,  60,  982,   60,   60,   50,   60,   20,   20,   20,   20,    0,    0,    0,    0,
20051018, 1200,  , HU, 16.2N,  80.3W,  65,  979,  105,   75,   50,  105,   30,   20,   20,   30,   15,   15,    0,   15,
20051018, 1800,  , HU, 16.6N,  81.1W,  75,  975,  120,   75,   60,  120,   50,   30,   30,   50,   15,   15,   15,   15,
```

3. Landsea, Chris, and Jack Beven. "The Revised Atlantic Hurricane Database (HURDAT2)." www.nhc.noaa.gov/data/hurdat/hurdat2-1851-2019-052520.txt. National Hurricane Center, 2019.

This looks like gibberish, but this is how it generally breaks down:

The top line of each line is a unique identifier for each storm, which is 'ALxxyyyy' for 'Atlantic, number hurricane of the year, year.' Then, the name of the storm that year (they need the unique identifier since names are recycled), followed by how many observations are recorded for that storm. Then, within each hurricane, the lines are sorted as follows:

Date, time (4 hour increments), classification (depression, storm, hurricane, etc.), latitude, longitude, maximum wind speed, minimum air pressure, and then a bunch of data I will not end up using. This data is radius of varying wind strengths, which didn't appear to have enough data to use for anything.

I further organize this data by dividing it up by storm key, so that I can access each storm individually within a pandas array. I also supplement this data by adding SST at each point of each storm. To begin analyzing the data, I had to turn the HURDAT2 data into some testable variables. First, I counted landfall for each storm. In the SST data, a sea surface temperature reading of '0' means that the observation is taken over land. Thus, for each hurricane, I test if the SST at the center of the hurricane ever becomes 0. If so, it gets added to the 'made landfall' category.

For maximum speed and minimum pressure, I simplified the idea to make the project more manageable. For each year, I essentially identify the strongest wind attained and the minimum pressure attained. This may be an oversimplification, and it is possible that I change this methodology in future projects.

To try to compensate for this shortcoming, I separated out the strongest storms in each year by level of severity. I take the 20^{th} and 80^{th} percentile of the maximum speed values, and label each year's maximum speed values based on where it falls on that scale.

The last thing I do is simply count how many storms ever achieve the classification of "hurricane." This is simple, since each storm has a list of what classifications it achieves.

To synthesize this data, I create a new pandas dataframe. This new frame has a column for year, a column for proportion of years which fall above the 80^{th} percentile for storm severity, a column for proportion of named storms which attain hurricane status, a column for highest wind speed of each year, in addition to minimum pressure, a column for both number and proportion of storms which make landfall, and a column for total number of named storms. This seemed like a good number of test statistics to use to explore the relationship between El Nino and Atlantic hurricanes.

METHODS

I used a number of methods to explore the relationship between El Nino and Atlantic Hurricanes. For starters, I ran a simple correlation test. I plotted the correlation between the El Nino 3.4 value of a given month and each individual hurricane statistic. This is shown in figure 1 in the notebook.

For reference, I abbreviated each variable when labeling the graphs to make everything plot-friendly. The correlation graphs are labeled with the following labels:

PropSev – The proportion of years which have maximum storm severity over the 80^{th} percentile

NumLand – The number of named storms each year which made landfall

Prop. Of storms to be Hurr – Proportion of named storms which attained hurricane status

Num of Storms – Number of named storms each year

Prop Landfall – Proportion of named storms which made landfall each year

MinPres – lowest recorded air pressure in a hurricane each year

MaxSpeed – highest recorded wind speed in a hurricane each year

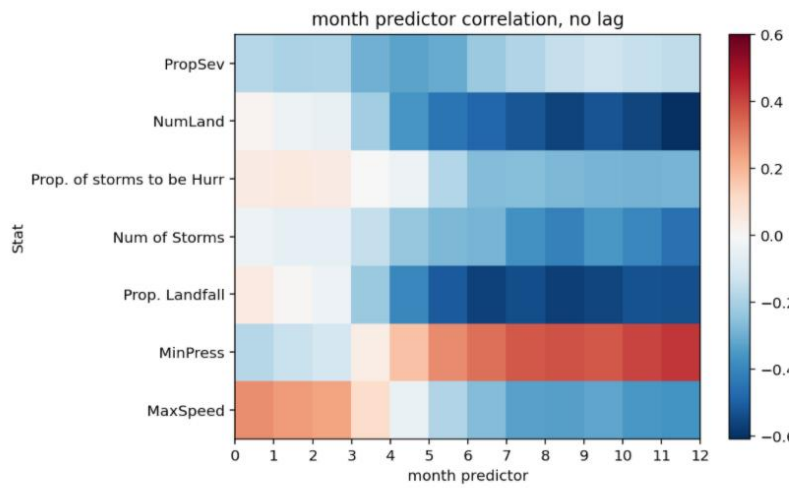


Figure 1: Seems like overall low correlations. Not unexpected. There is a 'predictability barrier' between March and April, which is consistent with El Nino studies. Thus, even though none of these correlations is incredibly strong, they are all near 0 prior to April. It seems that the 'PropSev', 'Prop. of storms to be Hurr', and the 'Num of Storms' variables have very weak correlations entirely. However, it would seem that the strongest storm (min pres and max speed) each year has a nontrivial relationship with El Nino. From what we can see, the maximum intensity and minimum pressure each year is intensified in WEAK El Nino years. IE, strong La Nina years. This is consistent with my hypothesis that La Nina leads to stronger storms, but the correlation is not strong enough to draw any serious conclusions.

The strongest correlation we see is December El Nino with Number of storms that make landfall. While the implication of this relationship isn't useful (in December, we already know how many hurricanes made landfall, so guessing based on statistical analysis does nothing). However, this may be an interesting relationship to explore, since El Nino values themselves have some self-predictability strength. We will test that later.

Figure 2: Given our somewhat strong December Nino – Landfall number relationship, I decided to plot those values to look at the relationship. It looks like a decent negative linear correlation, but there's clearly an outlier.

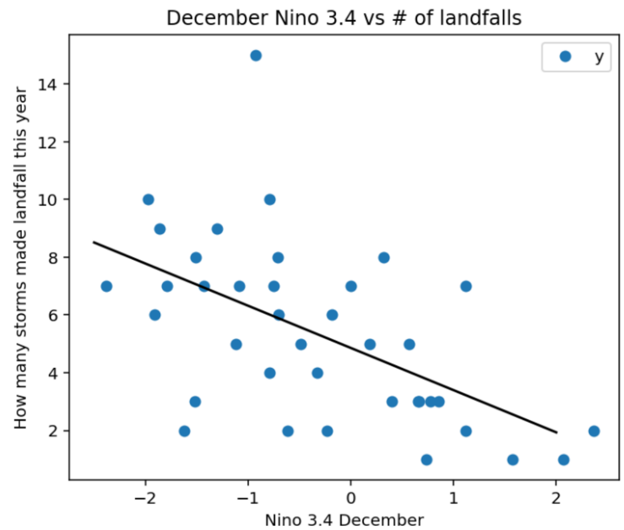
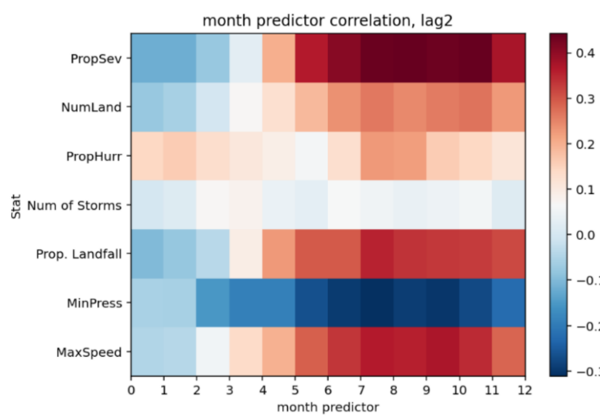


Figure 3: Based on what we know about El Nino's self-predictability, there is almost no use in using El Nino values to forecast things that are one year out, but it is nontrivial to forecast two years into the future. Thus, I check the correlation between each month's El Nino value, and the hurricane test-statistics from two years in the future.

This seems to be a dead end. All of the correlation values are far lower. The only thing of slight interest is that El Nino value during hurricane season seems to have a slight bit of correlation with the proportion of storms that are severe two years later. This weak correlation makes sense given the weak negative correlation between present El Nino and El Nino in two years.



Going back to the hypothesis that El Nino causes fewer and less sever storms, this conclusion would mean that high El Nino values today imply low El Nino values in two years, which then leads to more severe storms in two years.

Figure 4: Given the decent strength of December's El Nino as a predictor of how many hurricanes made landfall, I decided to test the strength of past Decembers predicting current Decembers' El Nino value. It seems that last year's December El Nino value is not a good predictor, but two years ago is decent.

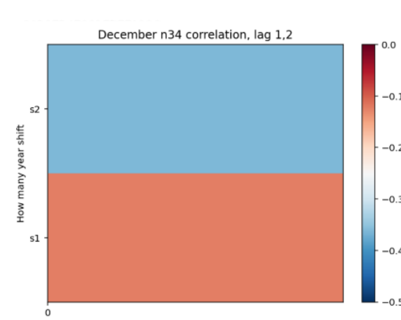


Figure 5: Based on the previous discovery, I was curious to see a plot of December's El Nino values, two years apart. I did a linear regression as well. Based on this, it looks like there's a negative correlation, but not a strong one.

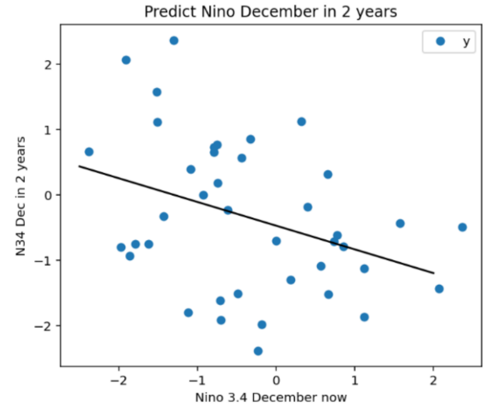
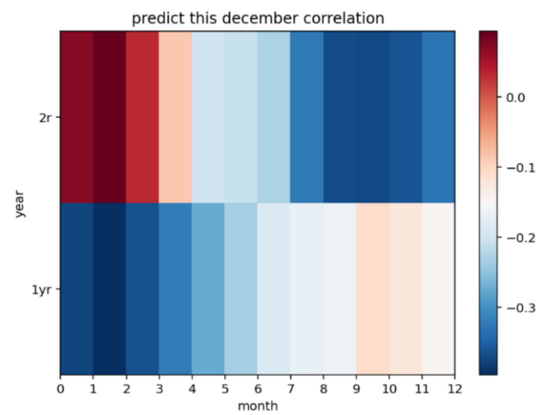
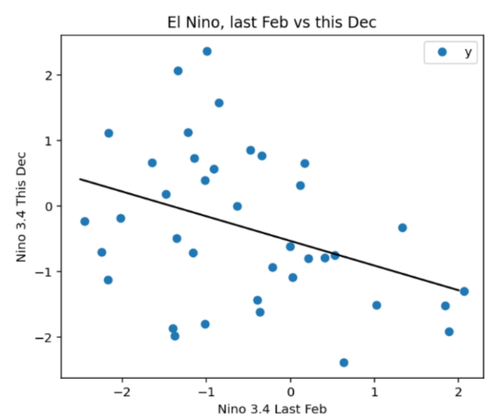


Figure 6: I then decided that given this interesting two-year-lag relationship for current December's El Nino value, it might be interesting to test the relationship between this December and other months. After all, we're currently locked into a 24-month lag. It might be worth testing other lags.



This test leads us to a conclusion that is consistent with El Nino auto-regression that we've done. It seems that this December's El Nino value is best predicted around 22-23 months ago. Empirically, the strongest predictor seems to be 22 months ago - February of last year.

Figure 7: Based on the previous conclusion that February seems to be the best predictor of next December's El Nino value, I plotted the two and ran a linear regression.



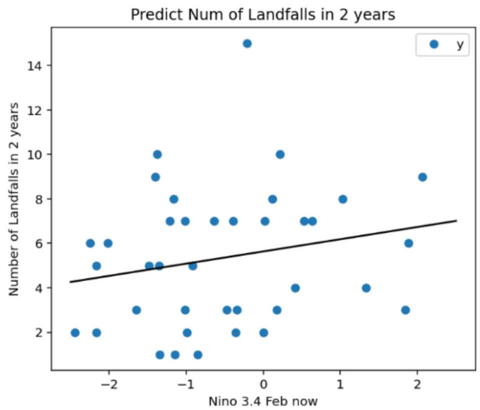


Figure 8: I ran sort of a double-regression. I used February's El Nino value to forecast next December's El Nino value, then used the forecast to predict the associated hurricane number.

Analysis of this method

From the start, the weak correlations across the board were not promising. A few of the hurricane metrics have almost zero correlation with El Nino (proportion of storms to be severe/hurricane, max intensity of a given year). It seems like even the only two with nontrivial correlation are weak. Furthermore, the strongest correlation of all is a value that only gives information about the past, making it useless as a forecast. Attempting to work around this flaw by using El Nino's self-correlation did not prove particularly helpful either.

To improve upon these analyses, there are a few things I can try in the future:

1. Use more data. I only used El Nino data going back 40 years, but I'm sure it's available in longer time periods. The hurricane dataset goes back a couple hundred, so a lot more exploration could be done.

2. Look at each storm more specifically. I only looked at the most extreme storm of each year when determining max speed for a year, but maybe it would be more prudent to keep track of every storm's peak intensity and use that instead.

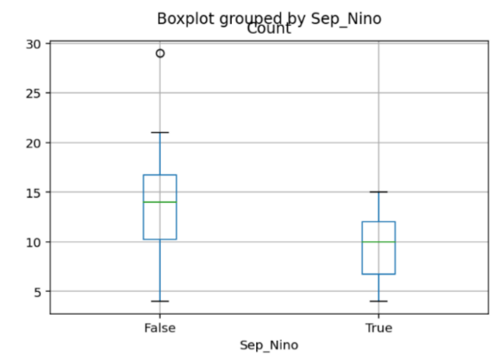
3. Include other data as predictors besides just Nino 3.4 Maybe I could explore the relationship between Nino 3.4 and Atlantic SST, since Atlantic SST certainly has an impact on hurricane season.

I do plan to continue to explore this field in the future with some of these considerations.

Bonus

I tried to reproduce the numbers that came from the University of Illinois project. First, I separated each month into El Nino vs non-El Nino. The conventional method of labeling something as El Nino is to determine whether the 5-month average temperature is more than 0.5 degrees above average. However, for simplification, I simply used the temperature anomaly of each individual month.

Given that the University of Illinois data is from September, that is the month I chose. First, I just wanted to see if there was a notable difference in hurricane count between years labeled as El Nino vs not. **Figure 9** shows the associated boxplots, which indicate to me that there is a substantive difference. This is promising if we are trying to reproduce UI's results.



I separated out Atlantic years into El Nino vs. all, and like the UI study, I counted average storm count, average hurricane count, and average severe hurricane count (>80th percentile). The results:

	Average	El Nino
Named Storms	12.68	9.5
Hurricane	6.3	4.1
severe hurricane	2.2	2.1

	Atlantic	
	Average	El Niño Avg.
Named storms	9.4	7.1
Hurricanes	5.8	4.0
Intense Hurricanes	2.5	1.5

We see that the numbers are similar, but a little different. These differences may be explained by the possibility that the UI study used a different range of years than I did. It is also possible that our thresholds for 'intense' or 'severe' hurricanes are different. However, my result demonstrates the same conclusion: El Nino seems to lead to fewer Atlantic hurricanes.

Conclusion

This study needs to be done with more data. While 40 years gave us a good idea of whether or not relationships exist, it would be helpful to use more years and more variables. That being said, this study accomplished a couple of things. First, El Nino years do seem to bring fewer hurricanes. Second, there exists an unexpected relationship between December El Nino and hurricane frequency. I need to explore further to figure out why this might be the case.

References

1. “Interaction with El Niño: How Hurricane Frequency May Be Affected.” *The Weather World 2010 Project*, University of Illinois, ww2010.atmos.uiuc.edu/(Gh)/guides/mtr/hurr/enso.rxml.

2. http://iridl.ldeo.columbia.edu/home/.tippett/.scratch/.APMA4990/.n34/#info

3. Landsea, Chris, and Jack Beven. “The Revised Atlantic Hurricane Database (HURDAT2).” www.nhc.noaa.gov/data/hurdat/hurdat2-1851-2019-052520.txt. National Hurricane Center, 2019.

4. “The Ten Most Damaging Hurricanes in U.S. History.” *Story.maps.arcgis.com*, ESRI, story.maps.arcgis.com/apps/MapSeries/index.html?appid=50aea84a9853491f994f775cb989ea92.