A Multilinear Approach to Forecasting the El Niño Southern Oscillation

Anoop Singh†
July 1st, 2018

I. Personal Experience

When I received the letter informing me of my selection to be published in $E = mc^2$, the first thought that came to my mind was "why in the world would I receive a letter from the University of Chicago months after the college admissions process was complete". I, like many seniors in high school, had reached a point where my thoughts were constantly focused on "starting college". Now, you're probably wondering why I would begin this section with this seemingly irrelevant account. You see, the essence of $E = mc^2$ is to "inspire high school students with some of the many possibilities for using mathematics to explore science". While I opened the letter, I realized that even though I was very occupied with the college application process, research was the one thing that truly allowed me to work out of wonder, and not for some greater prize. Over the past few years, I have noticed that many of my research peers are driven by the desire to win accolades to boost their college applications. However, it is clear to me that the true spirit of scientific research is to use what you love and understand to explore problems that you don't have answers to, and this can only be driven by sheer, childlike, curiosity.

---

†Contact: anoop_singh@brown.edu

My first exposure to scientific research took place in 9^{th} grade. I vividly remember shyly walking into Mr. Richard Kurtz' freshman science research class and taking a seat, only to be shocked by the teacher's high expectations. It was only my first day, but Mr. Kurtz acquainted me with different inventions high school students were creating to solve problems across the globe! It was hard to believe that at some point that year I would have to invent and create a machine to solve a problem impacting someone in my local community. My group and I started slowly, but with help from each other, and our teacher, we were able to successfully complete the task. It all began when we met Mr. Glenn Campbell, a local attorney suffering from Quadriplegia. As soon as we met Mr. Campbell, we noticed that a prominent problem in his life was managing a cluttered desk, and therefore, we decided to build a voice-activated desk to move various items of his that were positioned on metal bars (laptop, papers, cell-phone, water bottle, etc.). For example, if a client were to walk into his office, Mr. Campbell could say "all left" and all the items on his desk would move to the left, allowing him to better-interact with his client. At first, I felt that the invention was less than I had hoped for. However, when I saw Mr. Campbell use the desk for the first time, and experience a sliver of the independence he had yearned for, I felt a satisfaction that no other kind of work could give me. It was this feeling that drove me to continue down this path in research.

In my sophomore year of high school, I decided that I would personally benefit from exploring what the different areas of science had to offer. Partnering up with two close friends of mine (one being a fellow 2018 Regeneron STS Semifinalist), I chose to venture into Earth science, studying a "Cold-blob in the North Atlantic". But I soon learned that Earth science research is "painfully excruciating" (which is something I now look back on with a sense of hilarity). No matter how exciting the project was, I always felt that the grunt work was too time-consuming for anyone to have the will to do. Only a year into my research, I was looking to try out something I was more passionate about. The first field that popped into my mind was mathematics, but as a 15-year-old sophomore I was not allowed to work in a lab. A few days had passed while I searched for possible areas of interest, when (at a local science fair) I met a Columbia University researcher by the name of Daniel Bader. Out of genuine interest for my project, Mr. Bader decided to take a look at our board after judging other presentations at the fair. I then realized that instead of switching fields, I should learn the ins and outs of Earth science from a professional. The following week, I nervously composed an email asking Mr. Bader to mentor me on the subject.

Working alongside my research partner, under the tutelage of Mr. Bader, turned out to be one of the greatest experiences I've had. I soon began to love the work I used to loathe doing, and most of this is attributed to my introduction to the MATLAB computing language. Throughout the different research projects that I've worked on over the years, my passion for mathematics has always been evident. However, it was my junior year project that truly allowed me to apply my interest in mathematical problem solving to scientific questions. Not only was it fun finding solutions to different problems, I now had the mathematical and computational tools to create the "most efficient" approach to discovering the solution, an interest that was a driving force in my senior year STS project.

Midway through junior year, I found myself wondering if I had discovered a research area that I truly loved. My passion for computer science and mathematics applied perfectly to Earth science problems dealing with models and statistical analysis.

In March, I reached out to professors performing work in this area and was surprised to hear back from Professor Sultan Hameed, of Stony Brook University, who needed someone interested in MATLAB reliant analysis to take on a few projects that would generally be completed by his graduate students. As soon as summer began I started to work on my first project, an analysis of the impact of different pressure systems on the precipitation in California (with a goal to create a better mathematical model for Californian precipitation). The work was constant, but I found myself enjoying every minute of it. I had completed the project extremely quickly, but unfortunately, I failed in my endeavor. However, in this moment of frustration, I truly learned why curiosity is the key to being a strong researcher. Instead of being deterred, I picked up a project my mentor had been wanting to focus on for a while. When I heard that he would like to form models predicting the state of the El Niño Southern Oscillation, I was a bit nervous, but I also knew that it would be a test of my drive and my desire to truly contribute to the solution of a scientific problem. I threw myself into the work, making major procedural changes a just few hours after learning about the project. The long summer days became even longer as I began logging seven or eight hours creating scripts every day. Yet, each day was more exciting than the previous one. While I made sure I hung out with friends when I was offline, I would also look forward to the next breakthrough to come in the project. Professor Hameed and I would e-mail each other 10 times a day, anywhere from 6:00 A.M. to 4:00 A.M., discussing our personal analyses of the work. We continued working together because we are both driven by a common interest in the topics we study, something which must be present for a great partnership. Our work on El Niño is currently in preparation for submission to journals and we have already begun a few more projects in the meantime.

When senior year began, I noticed that the skills I learned in the summer were widely applicable to a variety of topics. I breezed through problems in computer science because I was so used to dealing with real-life problems requiring the same skills. In math, I found that partial differentials and topics in multivariable calculus and linear algebra could clearly be used to solve physical problems in Earth science. Even in a course I studied at Columbia University (under the Science Honors Program) I found that prominent ideas in physics, like the Navier Stokes equations, can tell us so much about Earth as we know it. The reason I began to notice all of these connections to my research was because of the passion I developed for my work. I truly began to treat my research as a part of my identity, because I chose a topic that embodied skills I loved to practice. Ultimately, I believe that when researching fields involving math and/or science, it's not only critical to be motivated by an almost overwhelming curiosity of the subject matter, but equally important to have a passion for the problem-solving the work entails.

II. Research

II.I. Layperson's Summary

Climate change impacts all people living on the Earth. The El Niño Southern Oscillation (ENSO) is a system which influences the climate around the globe. For this reason, it would be helpful to create a procedure for predicting ENSO each year, allowing the population to understand and prepare for a potential climate in their area, months in advance. This study developed a procedure to create predictions of ENSO every year. This procedure is simple, using basic statistics and computer science to create forecasts more accurate than those currently existing. Additionally, the study helped specify the relationship between the pressure systems surrounding the Pacific and ENSO, assisting in creating stronger predictions and allowing us to better understand the phenomenon.

II.II. Background

In recent times many extreme climatic events (ECE) have wreaked havoc on impacted regions. Most recently are Hurricanes Harvey and Irma, two Atlantic hurricanes that severely impacted the United States. ECEs are often influenced by the El Niño Southern Oscillation (ENSO), and incorporating ENSO activity aided in forecasting the increase of 2017 hurricane activity in the Atlantic (Klotzbach *et al*. 2017). Because of ENSO's influence on ECEs, accurately predicting the future of ENSO can not only be helpful for yearly projections of hurricane activity but also for forecasting broad global conditions.

ENSO is a teleconnection defined as a period of irregular winds and surface temperatures over the eastern Pacific Ocean. It is considered to be the most influential climate pattern used in seasonal forecasting [National Oceanic and Atmospheric Administration - NOAA (2014)]. The phenomenon consists of three phases, El Niño (Warm), La Niña (Cold), and the neutral phase. The El Niño phase exhibits warmer than average sea surface temperatures (SST) off the Pacific coast of South America. It is also associated with a high sea level pressure (SLP) in the western Pacific. The La Niña counterpart is associated with anomalously cold SST in the eastern Pacific and low SLP. The neutral phase of ENSO is a transitional phase between El Niño and La Niña. SST and tropical precipitation are near average during the neutral phase, allowing other teleconnections to have more of an impact on climate (Trenberth, 1997).

II.III. Previous Models

Mathematical modeling of ENSO has been used in forecasting over the past several decades. Early approaches represented ENSO via a set of unstable equations where the growth of ENSO is nonlinear and unpredictable behavior is caused by chaos (Zebiak and Cane 1987). More recently, a linear approach to modeling the phenomenon has been employed, based on the idea that the random nature of ENSO is based on external variables. This agrees with the idea that ENSO is randomly forced, yet also allows a simple approach for predicting ENSO year to year (Thompson and Battisti 2000). The approach uses external predictors to forecast ENSO up to a year in advance, giving disaster planners more time to prepare for predicted ENSO events (Pegion *et al.* 2017). If a multilinear method is to be used, the main concern is finding the optimal set of systems to incorporate into the models. Pegion *et al.* (2017) suggests that previous sets of systems are not holistic enough, meaning they do not collectively have enough influence on the regions surrounding ENSO. As a result, the study by Pegion *et al.* (2017) proposes the use of five core predictors consisting of the Pacific Meridional Mode, Seasonal Footprinting Mechanism, Trade Wind Charging, Western North Pacific SST, and the Victoria Mode (with the Pacific Meridional and Victoria Modes consisting of multiple attributes). The models produced by using these systems yield high correlations (*r*) with ENSO, however, the models in Pegion *et al.* (2017) are hindcasts and so the results are skewed since the regression model in their study is already fit to the data from the selected time period. Predictions made using <u>hindcasts</u> do not incorporate the increase in predictive error when measuring a different period of time. However, in this study, a <u>forecasting</u> procedure is being investigated which only uses prior data when creating an index for the following year. It would be desirable to have a forecasting procedure with similar prediction strength to the hindcasts of Pegion *et al.* (2017).

II.IV. Predictors and Influence on ENSO

In this study, I use multiple linear regression to produce forecasts for the December-January-February (DJF) season. This forecasting process is further discussed in the methodology (II.V). The set of predictors selected in this study (the Hawaiian High, South Pacific High, South Atlantic High, and Azores High) are known as the Subtropical Highs (STH). Each of these systems has a strong presence during the DJF months, justifying their inclusion (Figure 1). However, the main reason these predictors are included in the study is that they mark the origins of the trade winds, which have been found to influence ENSO state through wind speed in the ENSO region (Anderson *et al*. 2013).

This study uses the Centers of Action (COA) approach to determine the relationship between these moving pressure systems and ENSO. The term COA was coined by Rossby *et al.* (1939) to describe prominent pressure systems such as the four listed above. A COA is described by its latitude, longitude, and pressure, thus helping us understand how all aspects of pressure systems may impact surrounding regions. Using the COA approach, the latitude, longitude, and pressure of a pressure system are separate features of that system, which allows for many different sets of predictors to be considered. For example, the Hawaiian High Latitude may be considered separate from the Hawaiian High Pressure. The many sets of predictors used in this study, along with the placement of the systems surrounding the ENSO region, embodies the idea of a holistic approach and can potentially increase both model accuracy and interpretability.

It should be especially noted that that the predictors chosen in this study were not considered by Pegion *et al.* (2017). While they implemented a set of 7 predictors, leading to 127 possible models, in my study 4 systems (with attributes of latitude, longitude, and pressure, as well as the months averaged to produce the predictor) were implemented. Therefore, under this procedure a set of 216 variables were newly available for each system, allowing me to consider a total of $216^2 - 1 = 46655$ predictors. Assuming an accurate and efficient variable selection procedure exists (as investigated in this study), this will potentially lead to a more feature-rich and accurate modeling framework.

While the improvement of forecasts was a main aspect of the study, another goal was to use selected features to better understand ENSO. For this reason, I used the pressure system data (see Figure 1) not only in models but also to generate insight into the influence of global climate systems on ENSO. I found that the physical elements of the predictors explained the statistical relationships between ENSO and the STHs (see section II.VI. Results and Discussion).

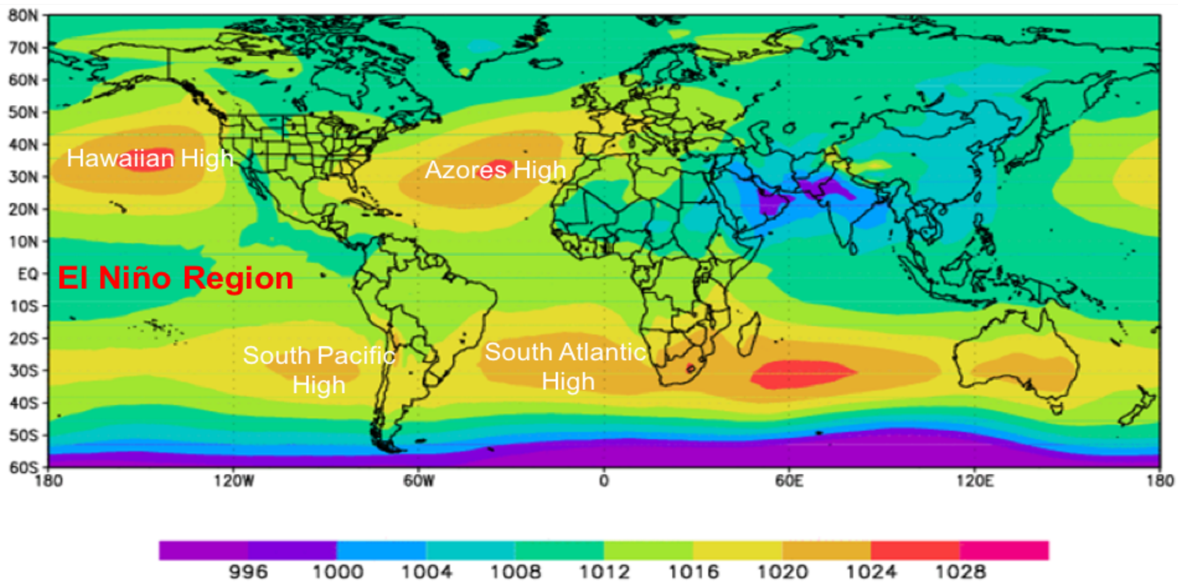

Figure 1: Average pressure of the June-July-August season from 1980-2000. Adapted by author from University Department of Atmospheric Sciences.

II.V. Methodology

There are various index definitions used to determine the strength of ENSO. Because of its observed impact on both SST and SLP, ENSO indices can be measured by either variable. There are four defined indices based on SST in the mid-Pacific, which include ENSO 1+2, ENSO 3, ENSO 3.4, and ENSO 4. These indices are calculated based on temperature anomalies across the given area and are therefore defined for different spatial locations (Figure 2). I also considered the Southern Oscillation Index (SOI) because it relies on SLP, demonstrating the atmospheric aspect of ENSO. All ENSO indices can be obtained from the National Oceanic and Atmospheric Association Physical

Sciences Division (NOAA 2017) while the SOI index can be found through the Australian Bureau of Meteorology (ABM 2017). I constructed models treating only ENSO 3.4 as the dependent variable, to assess consistency and accuracy among multiple modeling frameworks. The other representations were disregarded as ENSO 3.4 has been commonly used in recent times.

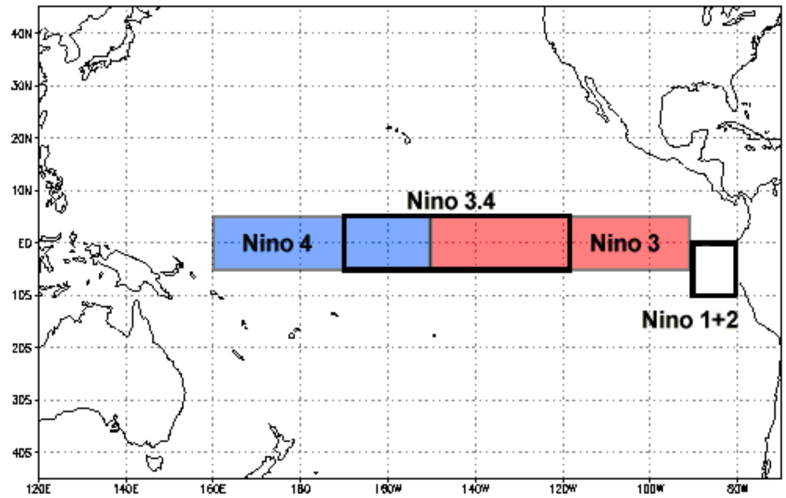

Figure 2: Regions of ENSO. Obtained from the National Oceanic and Atmospheric Association (NOAA) – (www.ncdc.noaa.gov/teleconnections/enso/indicators/sst.php)

In this study, the DJF season for ENSO 3.4 is being forecasted using the STH variables from the previous year. For every year being forecasted, a statistical model is developed based on the past 30 years. For example, when forecasting the 1979 season, a model is developed using data from 1949-1978. The steps for forecasting each year are as follows:

1. The different attributes of the STH systems are averaged over 2, 3, and 4 month periods. One example of this would be the Hawaiian High Pressure averaged over May-June-July-August (this variable is denoted as HHPRS_MJJA).
2. All the variables are assembled in a list and a set of heuristically-determined correlation thresholds is applied, to pick two variables that can be used to create

a model for 1979. There are two requirements for a set of variables to be used in creating a forecast. Firstly, both STH variables for 1948-1977 must have a correlation greater than 0.42 $(p < 0.02)$ with ENSO 3.4 (DJF) from 1949-1978. Secondly, the variables must also have a mutual correlation less than 0.42. After the restrictions are applied, several sets of predictors remain which have variables that are mutually independent but highly correlated with the following ENSO season.

3. A multilinear regression is applied to each set of 2 predictors found in step 2. The regression coefficients found after applying the regression represent a model for the next year's ENSO. Once the 1978 variables are substituted into the equation, the model produces an output predicting the 1979 ENSO season.
4. Step 3 is repeated with the next set of predictors which pass the restrictions in step 2. All predictions for the 1979 ENSO are compiled and averaged to produce an ensemble forecast.
5. Steps 2-4 are repeated for all seasons from 1980-2018. The results are then statistically compared to the existing data for ENSO.

I have created scripts, in the MATLAB computing language, to run the procedure above and produce figures displaying a comparison of predictions and Observed-ENSO. The main results can be found in the following section of this paper (see II.VI. Results and Discussion), but a copy of the scripts can also be found at: https://github.com/ASingh-2000/A-Multilinear-Approach-to-Forecasting-ENSO.

II.VI. Results and Discussion

The procedure described above was tested in 5 different scenarios. The variables involved in the forecast also rely on time (specifically the month), and so, I simulated forecasting ENSO (DJF) during the previous June, July, August, September, and October. For June, it was found that forecasts cannot be created for every year (when testing 1979-2018) because in specific years the variables only including months before July did not pass the restrictions in step 2 (see previous page). However, forecasts for July, August, September, and October were not only created but are extremely strong. For each of these months, the correlation was statistically significant at $p < 0.01$. The following table outlines the values produced by the analysis.

Table 1: Statistical comparisons of the forecasts with the observed ENSO 3.4 from 1979-2018.

Month	Mean Absolute Error (MAE)	Root Mean Square Error (RMSE)	Correlation with Observed ENSO	2018 Prediction
July	0.68	0.85	0.64	-0.48
August	0.61	0.76	0.74	-0.37
September	0.57	0.71	0.79	-0.42
October	0.54	0.70	0.81	-0.50

From Table 1, not only is it clear that the forecasting procedure proposed in this study produces accurate predictions of ENSO, but the predictions are high even at longer lead times, comparable to hindcasts such as Pegion *et al.* (2017) and other forecasting methods such as that in Zhang *et al.* (2017). Furthermore, all predictions for the 2018 ENSO (DJF) season, which occurred recently, were correct in forecasting what appeared to be a neutral to weak La Niña phase. On the visual side of things, Figure 3 conveys the strong fit between my procedure's ENSO predictions and observed phase.

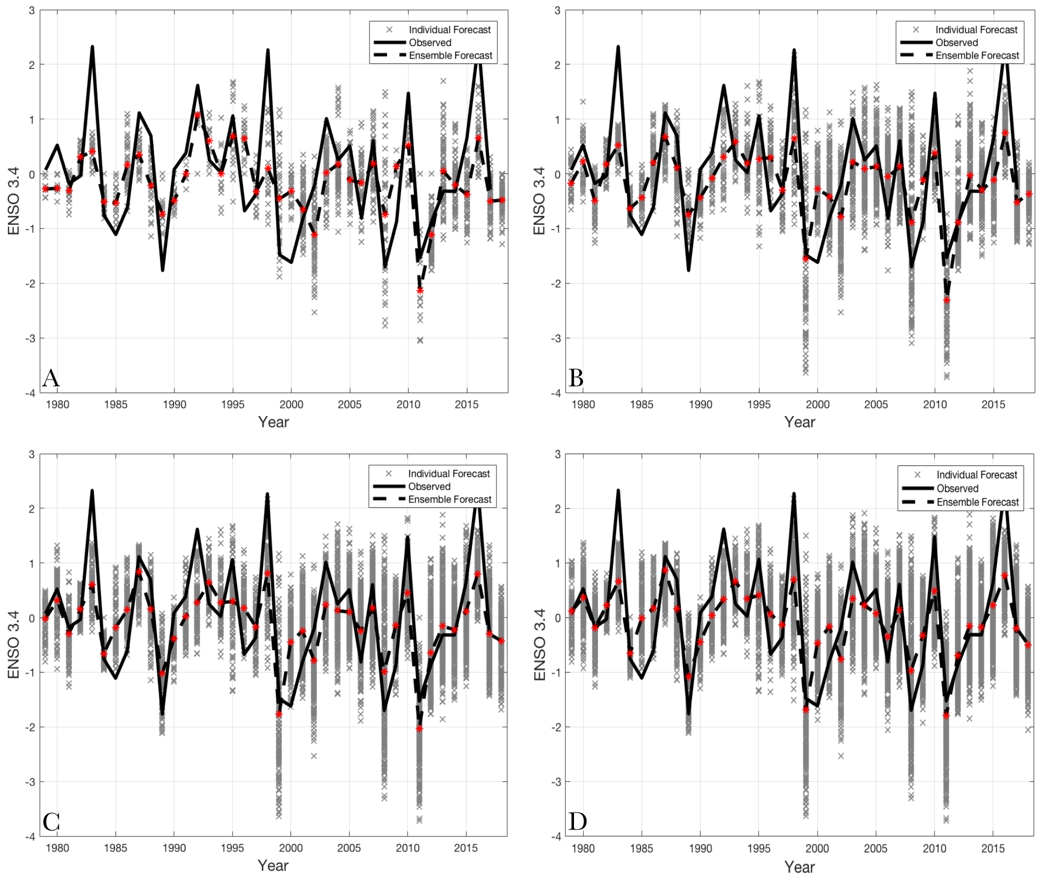

Figure 3: Forecasts of ENSO 3.4, from 1979-2018, using the method proposed by this study. The figures simulate forecasting in (and are blind to data post) July (A), August (B), September (C), and October (D).

Both graphically and statistically, it seems that the procedure outlined in this study is preferable to other proposed forecasting methods. Graphically, the forecasts seem to fit ENSO extremely well. Statistically speaking, the correlations and errors in Table 1 are preferable to other forecasting methods available. Furthermore, the approach outlined in this paper is computationally simple, with script runtimes in MATLAB being less than a minute (you may test this for yourself, as the repository link is attached in section II.V. Methodology).

Given the accuracy of this forecasting procedure, it is important to address the relationship between STHs and ENSO. The strength of the predictors in the study is due to the physical relationship between variations in the Highs and ENSO. The trade winds are surface winds found in the tropics. They originate at the STHs and flow west into the tropical region (NOAA 2004). The El Niño phenomenon is triggered by the trade winds in a process called wind-induced charging, which involves the increase of heat in the Pacific Ocean due to trade wind variations (Anderson *et al.* 2013). In order to demonstrate how attributes of the STHs impact the SST, and in turn ENSO index values, composite graphs were created through the NOAA Physical Sciences Division website. The predictor that had the strongest correlation with the observed ENSO index for each pressure system was identified. The years for which the variable's value was one standard deviation above the mean were used to form a composite SST anomaly of the ENSO region along with a composite anomaly of the Zonal Wind. This was also done for years when the variable's value was one standard deviation below the mean. For example, the following are the composites for the South Pacific High Pressure averaged over April, May, June, and July:

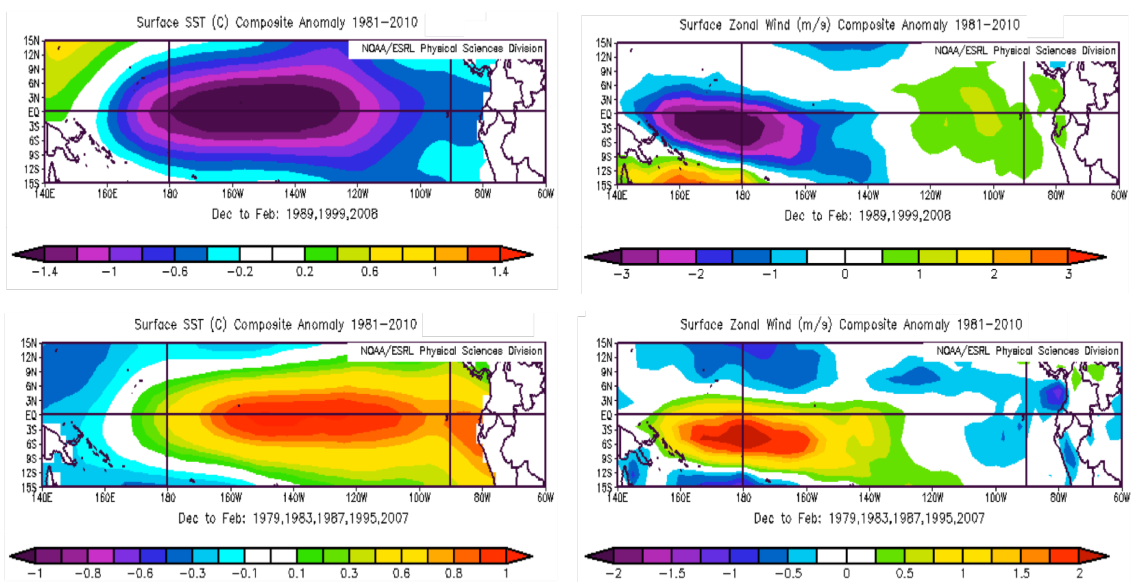

Figure 4: Composite anomalies of the ENSO region. A depiction of SST of the ENSO region when the South Pacific High pressure was abnormally high (Top Left). Zonal Wind when the South Pacific High pressure was abnormally high (Top Right). SST when the South Pacific High pressure was abnormally low (Bottom Left). Zonal Wind when the South Pacific High pressure was abnormally low (Bottom Right).

In Figure 4, it is clear that abnormally high South Pacific High pressure corresponds to negative wind velocity and in turn a lower SST, while an abnormally low South Pacific High pressure corresponds to a positive wind velocity and a positive SST. A relationship exists between all the STHs, the wind above the ENSO region, and the SST in the region (through the trade winds). This physical relationship (trade wind generation at STHs impacting ENSO) is a strong reason why these pressure systems have been successful in this study and should be used in the future for forecasts beginning even earlier in the year. Earlier forecasts are important for our understanding of global climate and disaster planning, and the year-round relationship between these systems and ENSO can be useful in forming them.

II.VII. Conclusion

The goal of this study was to form a stronger approach to predicting ENSO than previously available. Accurate procedures, such as that found in this paper, would allow us to not only predict ENSO earlier in the year but also form better predictions, giving the general population a longer time to prepare for ENSO and stronger predictions to base these preparations on. I found that using the Subtropical Highs, along with their attributes, in a multiple linear regression produces more accurate forecasts than models that currently exist. This is likely due to the fact that these STHs contribute to the trade winds, which is known to have an influence on ENSO. The identification of this relationship is another key finding in this paper because it offers an unconventional understanding of ENSO: specific systems like the STHs seem to contribute to ENSO phase, and so ENSO should not be considered a truly randomly forced system.

II.VIII. References

Anderson, B. T., Perez, R. C., & Karspeck, A. (2013). Triggering of El Niño onset through trade wind–induced charging of the equatorial Pacific. *Geophysical Research Letters*, *40*(6), 1212-1216.

Chang, P., Ji, L., Li, H., & Flügel, M. (1996). Chaotic dynamics versus stochastic processes in El Niño-Southern Oscillation in coupled ocean-atmosphere models. *Physica D: Nonlinear Phenomena*, *98*(2-4), 301-320.

Hameed, S., and Reimer, N. (2012): Relationship of Sahel precipitation and atmospheric centers of action. Adv. Meteor., 2012, 953853, doi:10.1155/2012/953853.

Santer, B. D. (1988): Regional validation of general circulation models. Ph.D. thesis, University of East Anglia.

Pegion, K. V. and Selman, C. (2017). Extratropical Precursors of the El Niño–Southern Oscillation, in Climate Extremes: Patterns and Mechanisms (eds S.-Y. S. Wang, J.-H. Yoon, C. C. Funk and R. R. Gillies), John Wiley & Sons, Inc., Hoboken, NJ, USA. doi: 10.1002/9781119068020.ch18

Philander, S. G. H., & Pacanowski, R. C. (1986). A model of the seasonal cycle in the tropical Atlantic Ocean. *Journal of Geophysical Research: Oceans*, *91*(C12), 14192-14206.

Rossby, C.-G., and Coauthors (1939). Relation between variations in the intensity of the zonal circulation of the atmosphere and the displacements of the semi-permanent centers of actions. J. Mar. Res., 2, 38–55, doi:10.1357/002224039806649023.

S. (n.d.). Southern Oscillation Index (SOI). Retrieved August 15, 2017, from https://www.ncdc.noaa.gov/teleconnections/enso/indicators/soi/

Thompson, C. J., & Battisti, D. S. (2000). A linear stochastic dynamical model of ENSO. Part I: Model development. *Journal of Climate*, *13*(15), 2818-2832.

Trenberth, K. E. (1997). The definition of el Niño. *Bulletin of the American Meteorological Society*, *78*(12), 2771-2777.

US Department of Commerce, National Oceanic and Atmospheric Administration. (2004, December 19). Currents. Retrieved September 14, 2017, from https://oceanservice.noaa.gov/education/kits/currents/05currents2.html

Wang, C., and J. Picaut, Understanding ENSO physics - A review, in Earth's Climate: The Ocean-Atmosphere Interaction, Geophysical Monograph Series, Volume 147, edited by C. Wang, S.-P. Xie, and J. A. Carton, pp. 21-48, AGU, Washington, D. C., 2004.

Zhang, W., Villarini, G., Slater, L., Vecchi, G. A., & Bradley, A. A. (2017). Improved ENSO Forecasting Using Bayesian Updating and the North American Multimodel Ensemble (NMME). Journal of Climate, 30(22), 9007-9025. doi:10.1175/jcli-d-17-0073.1