ANALYZING EL NINO SOUTHERN OSCILLATION PREDICTIONS FROM LONG-SHORT-TERM-MEMORY MODELS

BY

ANDREW K HUANG

THESIS

Submitted in partial fulfillment of the requirements
for the degree of Master of Science in Atmospheric Sciences
in the Graduate College of the
University of Illinois at Urbana-Champaign, 2018

Urbana, Illinois

Adviser:

Professor Ryan L. Sriver

ABSTRACT

El Nino Southern Oscillation (ENSO) can have global impacts across the world. Because of its prevalence, scientists run models to forecast its next move. Here, long-short-term-memory models (LSTM) were compared to linear regression models (LR) as first steps to explore the potential benefits of simple deep neural networks for predicting ENSO. Each model's prediction capabilities were tested with sea surface temperatures (SST), warm water volumes, and zonal winds as predictors, individually and in combinations, utilizing both monthly and daily resolution data, across a total of 11 leads. By utilizing these three variables, we examine different forms of climate variability within the coupled system (SST), the subsurface ocean (warm water volume), and the atmosphere (zonal winds), and we quantify the relative importance of each of these processes for ENSO predictability through two statistical modeling approaches: LSTM and LR. Results show that when using monthly data as predictors, predictions from LSTM were similar to predictions from LR. However, with daily data, LSTM exhibited some advantage over LR in terms of the correlation coefficient, especially with daily resolution SST as a predictor and at longer leads. This can be appealing because once the computationally expensive training of LSTM is complete, the predictions employing the trained model can be relatively cheap to perform thereafter.

ACKNOWLEDGEMENTS

I would like to thank my adviser, Ryan Sriver, and my group member, Ben Vega-Westhoff, for their support and guidance on this project. NOAA High Resolution SST data provided by the NOAA/OAR/ESRL PSD, Boulder, Colorado, USA, from their Web site at https://www.esrl.noaa.gov/psd/. NCEP Reanalysis data provided by the NOAA/OAR/ESRL PSD, Boulder, Colorado, USA, from their Web site at https://www.esrl.noaa.gov/psd/. Warm water volume data provided by TAO Project Office/NOAA/PMEL/Seattle from their Web site at http://www.pmel.noaa.gov/tao/. In Python, visuals were made with HoloViews and matplotlib, LSTM models were built with Keras, LR models were built with scikit-learn, and data management and wrangling were done with xarray, pandas, and numpy.

TABLE OF CONTENTS

1. INTRODUCTION

A rock in a stream can affect the flow of the stream. Similarly, a phenomenon known as the El Nino Southern Oscillation can influence the atmospheric flow with its warmer (cooler) than normal sea surface temperatures (SST) in the eastern central Pacific during its El Nino (La Nina) phase. This could lead to anomalous temperature and precipitation patterns across the US and the world (*Ropelewski and Halpert* 1987, 1989; *Mason and Goddard* 2001; *Glantz* 2001), and influence severe and hazardous weather activity, such as tropical cyclones and tornadoes (*Landsea and Pielke.* 1998; *Allen, Tippett, and Sobel* 2015). Because of ENSO's impacts across the world, scientists routinely run both dynamical and statistical models to forecast upcoming ENSO phases from the next month to the next year.

Statistical models, simple and complex, have been formulated in the past to help predict the evolving ENSO. Two were the canonical correlation analysis and constructed analog models. The canonical correlation analysis model is a multi-dimensional analysis between a set of predictand and a set of predictors that maximizes the relationship between those two sets, where a set can be a gridded time series, so that the evolution is captured and can be projected outward for a prediction (*Barnston and Ropelewski,* 1992). The constructed analog model retrieves several analogs, comparable months of the predictor in terms of least error, and applies weights to match the corresponding predictand at those times of the analogs to make a prediction (*Van den Dool,* 1994). These two statistical models were found to be as competitive as the older dynamical models in forecasting ENSO, with all correlation coefficients averaging to about 0.6 at six lead months in the historical period, 1956-1993 (*Barnston et al.* 1994).

However, a more recent study found that dynamical models generally outperform statistical models in forecasting ENSO (*Barnston et al.* 2011). At a lead time of six months in the historical period, 2002-2011, the best dynamical model boasted an average correlation value of 0.7 while the best statistical model at that time only had a correlation value at 0.5, lower than the earlier study, attributed to an interdecadal shift in variability and the mean state of the tropical Pacific Ocean (*Hu et al.* 2012). There have also been hybrid models—statistical techniques used to post process dynamical models with the goal of further improving dynamical models' performance. One such example is the Bayesian updating technique utilized on the North American Multi-Model Ensemble (*Zhang and Villarini*, 2017).

In present times, because of advances in data, hardware, and techniques leading to the advent of deep learning, there are new opportunities for statistical models to redeem its edge in forecasting. Deep learning is a form of machine learning that uses layered, hierarchical representations of data, or neural networks, to learn. Neural networks are inspired by biological brains—the input neurons are activated through stimuli, which may then lead to reactions, often non-linear, from other neurons, based on assigned thresholds. The thresholds, or weights, between each neuron were designated by how effective the connection was in leading to the desired output during its training period (*Schmidhuber*, 2015).

Training a neural network can be as computationally expensive as running a dynamical model, but upon completion of its training, the neural network can be reused inexpensively to conduct predictions. Neural networks in the past have been used to forecast SST (Wu et al., 2006) but here,

the capabilities of long-short-term-memory models (LSTM) (*Hochreiter and Schmidhuber* 1997) were investigated from a variety of aspects. LSTM is an advanced recurrent neural network (RNN), and its specialty is that it can retain short segments of memory for long term, unlike a vanilla RNN which is only able to remember from the previous time step.

The following is an example to emphasize the differences between LSTM and RNN. If RNN was used as a language processor to write a children's book, where sentences in the book are like "Joe saw Bob" and "Bob saw Ace," or a noun following a verb and a verb following a noun, RNN is capable of automatically recognizing these rules through training. However, a couple issues may arise; because RNNs are only able to remember one time step, or word, prior to the current word, it may end up outputting sentences like "Bob saw Bob," or "Joe saw Bob saw Ace." LSTM, in contrast, is capable of remembering not just the previous word, but all prior words, so if it remembers that "Bob" was used as the first word, it will not use "Bob" again, or if it remembers that a noun was already provided, it will end with a period instead of continuing with another verb. This same logic can be applied to ENSO predictions. If the observations show an elongated La Nina phase, then LSTM may realize this and predict an El Nino next.

Here, linear regression models (LR) were compared to LSTM across a total of 11 leads, where SST, warm water volume (WWV) and 925 mb, zonal winds (WND) were used as predictors. Each predictor captures a distinct, key process that helps modulate the evolution of ENSO from either the atmosphere, ocean, or coupled system. During normal conditions, solar heating warms up the surface ocean waters, but with the trade winds blowing from east to west, these warm surface

waters generate a warm pool by piling up in the west. Warm waters then, through the equatorial undercurrents, travel back east below the surface and lead to upwelling in the east.

During El Nino conditions, trade winds weaken and westerly WND bursts occur which lead to Kelvin waves in the ocean, helping to initiate and sustain the El Nino (*Wyrtki* 1975). Because of this, WND was chosen as a predictor to capture the high frequency atmospheric variability. WWV, another predictor, captures the low-frequency variations in ocean heat because it is transported to higher latitudes during the El Niño and must build up again before another El Nino is able to occur (*Wyrtki* 1985). Lastly, SST acts as the interface between the ocean and atmosphere, capturing the coupled ocean-atmosphere variability, and because SST contains memory, where past states have some effect over future states, SST was chosen as a predictor—not to mention that SST anomalies define many ENSO indices, such as the Nino 3.4 index (N34). These predictors' effectiveness was then tested on a monthly and daily resolution, individually and in combinations (SST-WWV, SST-WND, SST-WWV-WND, etc).

2. DATA AND METHODOLOGY

The datasets used were all sliced from 1982 to 2017. N34 and SST predictors, were computed from the NOAA OI SST v2 dataset by taking the daily SST values and subtracting the monthly climatological (1981 – 2010, or the base period) mean, averaged over 5S-5N and 170-120W (*Trenberth*, 1997). WND predictors were computed from the NOAA OI SST v2 dataset by taking the daily WND values and subtracting the base period mean, averaged over 5S-5N and 120-160E (*Wang et al.* 2011). WWV predictors, defined as the volume above the 20 degrees C isotherm over the region 5S-5N and 120E-80W (*Meinen and McPhaden,* 2000), were retrieved from the TAO and its anomalies were taken in a similar fashion by subtracting the base period mean. (Supplemental Figure 1).

ENSO predictions were performed with LR and LSTM. LSTM has three layers: two LSTM layers, the first with 50 nodes and the second with 150 nodes, connected to a dense layer, compiled with a mean square error loss function and the Adam optimizer, and finally trained on eight epochs. These settings were tweaked and tested to not only attain appreciable skill, but also minimize runtime to have the opportunity to run many simulations and test many aspects.

To predict on using monthly data, predictors were averaged monthly and were shifted backwards in time by the number of months equaling the number of leads (Supplemental Figure 2). An example of a one lead prediction is using averaged February predictors to predict the averaged March's N34 value. For four leads, averaged February predictors was used to predict the averaged June's N34 value. To utilize daily data, daily predictors were transposed next to monthly N34

values and shifted backwards in time by the number of months equaling the number of leads. If a month does not contain 31 days, the last daily value was forward filled (Supplemental Figure 3). So, to perform a one lead prediction during February, February 28's N34 value was copied over three times, and the daily values on February 1 to February 28 and pseudo February 29, 30, 31, was then used to predict the averaged month of March. Also, because WWV was only available monthly, its monthly average also was forward filled, or copied over, 31 days for predictions on the daily timescale.

To evaluate skill, the correlation coefficient was computed between the observed and the predicted N34 values. To explore the variability across years, a jackknife resampling technique was employed (*Efron*, 1982). This involves reserving some years in the timeseries as the validation period and the rest as the training (Supplemental Figure 4). For instance, one sample was training the models on 1986 – 2017 to predict 1983 – 1985; another sample was training our model on 1983 – 1985 and 1989 – 2017 to predict 1986 – 1988. Thus, through this method, a total of eleven samples was obtained, each with a sample size of three years.

3. RESULTS

On exploring the three predictors, SST, WWV, and WND, SST was found to have high correlations in the early leads, dropping off rapidly with longer leads, while WWV and WND both started with low correlations in the early leads, peaking at mid leads, and dropping off again at longer leads (Figure 1). When predictions are made with both SST and WWV together or SST and WND together, the drawbacks were offset by the partner variable, leading to comparatively higher correlations across all leads. Also, WWV alone had higher overall correlations than WND alone, and thus, the correlation of SST and WWV together exceeds SST and WND together. There was not much difference when using WWV and WND together. When all three predictors were used, SST, WWV, and WND, the correlations were only slightly improved relative to SST and WWV.

When comparing LSTM and LR using monthly resolution data, little to no difference in correlation was found across the cross validated years for all parameters and leads. However, when using daily resolution data, predictions done with LSTM at longer leads and involving SST resulted in relatively higher correlations compared to LR (Figure 2). This could be because daily data is able to capture the non-linear evolution of SST, which LSTM is able to pick up on, but monthly data does not offer this opportunity. As an example, if SST's evolution was defined as $y = x^3$, and daily data had nine x points evenly spaced from -1 to 1, but monthly data only had three points at x = -1, 0, and 1, the interpolated daily data exhibits a non-linear line while the interpolated monthly data results only in a linear line (Supplemental Figure 5).

However, not all daily resolution data benefits LSTM. Predictions performed using daily, forward-filled WWV data did not indicate any improvements with LSTM over LR, which is not surprising because the monthly resolution was artificially transformed into daily resolution. On the other hand, WND did have actual daily data, but the LSTM predictions were not generally better than the LR predictions.

To test whether the correlation difference between LSTM and LR was statistically significant, hypothesis tests were also performed. These were done through t-tests using the 11 jackknife samples' correlations, comparing the differences between the temporal resolution of predictors, the two types of overarching models, and the lead times. In most cases, the correlations do not differ significantly across samples, having p-values greater than 0.05 suggesting that the null hypothesis cannot be rejected at a 95% confidence level (Figure 5). However, there are a couple exceptions to this: particularly LSTM using the SST predictor at leads greater than seven holding a p-value near 0.

Inspired by the fact that the correlations of lead SST vs observed SST diverged for positive and negative anomalies, but not much for other variables (Figure 4), one final test was performed by splitting the time series into a training and validation period, and then further separating the monthly SST predictors into two groups: positive and negative anomalies. Within the training time series, two sub models were developed for both LSTM and LR, where the first sub model trained on positive anomalies while the second sub model trained on negative anomalies. The first sub model would be utilized to predict during positive anomaly conditions in the validation time series and the second sub model would be utilized to predict during negative anomaly conditions. Then,

the predictions from both sub models were arranged back into a single chronological time series and the correlation was computed between the N34 observations and the merged predictions.

Most of these split sub models suffered lower correlations compared to the non-split models at early to mid leads, but exhibited higher correlations beginning at lead 8 (Figure 5). This is true on all timescales for both LSTM and LR except one: LSTM using daily data. The non-split LSTM had higher correlations than the split sub models across all leads. This suggests that LSTM possesses the ability to recognize these abstract, underlying phenomena on its own through training and could possibly explain why LSTM predictions involving SST had higher correlations than its LR counterpart at longer leads.

4. DISCUSSION

Overall, this study demonstrated that LSTM had some advantage over LR with predictions involving SST and longer leads, but beyond that, the differences between LSTM and LR were statistically insignificant. However, before dismissing LSTM's forecasting outlook, there is a caveat to note. The goal of this study was not to achieve the highest possible predictability with LSTM; rather it was trying to explore what LSTM could offer so others could build upon it. One of the notable findings of this study was that although SST seems to slowly vary, there seems to be some non-linear component to it that LSTM can pick up on with daily resolution input, or in broader terms, LSTM can capture non-linear signals. This study had also found that LSTM has a statistical significant advantage over LR at long leads, and this may be because of LSTM's memory—through training, it may have recognized the typical length of a given ENSO phase and was able to use this knowledge in its predictions. This capability could hypothetically add value to climate predictions and projections.

And because of this study's exploratory nature, sacrifices had to be made to explore the several parameters, such as sacrificing the spatial aspect of N34, training only on the temporal evolution of N34 to save some processing time. Future studies could provide both the spatial dimension and temporal dimension for training so there can be more non-linear patterns available for LSTM to take advantage of besides the non-linear progression of SST. In addition to that, this work can be considered a proof of concept to isolate effects and interactions between different predictors of ENSO as to gain some scientific insight using deep learning methods contrasted with linear models. Future works can be expanded using more variables with less emphasis on scientific

exploration to potentially improve predictability. Nevertheless, as time progresses, not only will techniques and processors be enhanced, but there will also be more observational data available to train and validate LSTM, so that someday, LSTM can outperform LR in predicting ENSO events without doubt, and perhaps even dynamical models.

REFERENCES

Allen, John T., Michael K. Tippett, and Adam H. Sobel. 2015. “Influence of the El Niño/Southern Oscillation on Tornado and Hail Frequency in the United States.” *Nature Geoscience* 8 (4): 278–83. https://doi.org/10.1038/ngeo2385.

Barnston, Anthony G., Huug M. Van den Dool, David R. Rodenhuis, Chester R. Ropelewski, Vernon E. Kousky, Edward A. O’Lenic, Robert E. Livezey, et al. 1994. “Long-Lead Seasonal Forecasts—Where Do We Stand?” *Bulletin of the American Meteorological Society* 75 (11): 2097–2114. https://doi.org/10.1175/1520-0477(1994)075<2097:LLSFDW>2.0.CO;2.

Barnston, Anthony G., and Chester F. Ropelewski. 1992. “Prediction of ENSO Episodes Using Canonical Correlation Analysis.” *Journal of Climate* 5 (11): 1316–45. https://doi.org/10.1175/1520-0442(1992)005<1316:POEEUC>2.0.CO;2.

Barnston, Anthony G., Michael K. Tippett, Michelle L. L’Heureux, Shuhua Li, and David G. DeWitt. 2011. “Skill of Real-Time Seasonal ENSO Model Predictions during 2002–11: Is Our Capability Increasing?” *Bulletin of the American Meteorological Society* 93 (5): 631–51. https://doi.org/10.1175/BAMS-D-11-00111.1.

Dool, Van den. 1994. “Searching for Analogues, How Long Must We Wait? *Tellus A* Wiley Online Library.” n.d. Accessed March 2, 2018. http://onlinelibrary.wiley.com/doi/10.1034/j.1600-0870.1994.t01-2-00006.x/abstract.

Efron, Bradley. “7. Cross Validation, the Jackknife and the Bootstrap.” 1982. In *The Jackknife, the Bootstrap and Other Resampling Plans*, 49–59. CBMS-NSF Regional Conference Series

in Applied Mathematics. Society for Industrial and Applied Mathematics. https://doi.org/10.1137/1.9781611970319.ch7.

Glantz, M. H., 2001: Currents of Change: Impacts of El Niño and La Niña on Climate and Society. Cambridge University Press, 252 pp.

Hu, Zeng-Zhen, Arun Kumar, Hong-Li Ren, Hui Wang, Michelle L'Heureux, and Fei-Fei Jin. 2012. "Weakened Interannual Variability in the Tropical Pacific Ocean since 2000." Journal of Climate 26 (8): 2601–13. https://doi.org/10.1175/JCLI-D-12-00265.1.

Meinen, Christopher S., and Michael J. McPhaden. 2000a. "Observations of Warm Water Volume Changes in the Equatorial Pacific and Their Relationship to El Niño and La Niña." *Journal of Climate* 13 (20): 3551–59. https://doi.org/10.1175/1520-0442(2000)013<3551:OOWWVC>2.0.CO;2.

Hochreiter, Sepp, and Jürgen Schmidhuber. 1997. "Long Short-Term Memory." Neural Computation 9 (8): 1735–80. https://doi.org/10.1162/neco.1997.9.8.1735.

Hu, Zeng-Zhen., Arun Kumar, Jieshun Zhu, Bohua Huang, Yu-heng Tseng, and Xiaochun Wang. "On the Shortening of the Lead Time of Ocean Warm Water Volume to ENSO SST Since 2000." 2017. June 27, 2017. https://www.nature.com/articles/s41598-017-04566-z.

Mason, Simon J., and Lisa Goddard. 2001. "Probabilistic Precipitation Anomalies Associated with ENSO." Bulletin of the American Meteorological Society 82 (4): 619–38. https://doi.org/10.1175/1520-0477(2001)082<0619:PPAAWE>2.3.CO;2.

Pielke, Roger A., and Christopher N. Landsea. 1999. "La Niña, El Niño, and Atlantic Hurricane Damages in the United States." Bulletin of the American Meteorological Society 80 (10): 2027–34. https://doi.org/10.1175/1520-0477(1999)080<2027:LNAENO>2.0.CO;2.

Ropelewski, C. F., and M. S. Halpert. 1987. “Global and Regional Scale Precipitation Patterns Associated with the El Niño/Southern Oscillation.” *Monthly Weather Review* 115 (8): 1606–26. https://doi.org/10.1175/1520-0493(1987)115<1606:GARSPP>2.0.CO;2.

———. 1989a. “Precipitation Patterns Associated with the High Index Phase of the Southern Oscillation.” *Journal of Climate* 2 (3): 268–84. https://doi.org/10.1175/1520-0442(1989)002<0268:PPAWTH>2.0.CO;2.

Schmidhuber, Jürgen. 2015. “Deep Learning in Neural Networks: An Overview.” Neural Networks 61 (January): 85–117. https://doi.org/10.1016/j.neunet.2014.09.003.

Tippett, Michael K., Anthony G. Barnston, and Shuhua Li. 2011. “Performance of Recent Multimodel ENSO Forecasts.” *Journal of Applied Meteorology and Climatology* 51 (3): 637–54. https://doi.org/10.1175/JAMC-D-11-093.1.

Trenberth, Kevin E. 1997. “The Definition of El Niño.” *Bulletin of the American Meteorological Society* 78 (12): 2771–78. https://doi.org/10.1175/1520-0477(1997)078<2771:TDOENO>2.0.CO;2.

Wang, Wanqiu, Mingyue Chen, Arun Kumar, and Yan Xue. 2011. “How Important Is Intraseasonal Surface Wind Variability to Real-Time ENSO Prediction?” *Geophysical Research Letters* 38 (July). https://doi.org/10.1029/2011GL047684.

Wyrtki, Klaus. 1975. “El Niño—The Dynamic Response of the Equatorial Pacific Ocean to Atmospheric Forcing.” *Journal of Physical Oceanography* 5 (4): 572–84. https://doi.org/10.1175/1520-0485(1975)005<0572:ENTDRO>2.0.CO;2.

Wyrtki, Klaus. 1985. “Water Displacements in the Pacific and the Genesis of El Nino Cycle.” *Journal of Geophysical Research: Oceans* 90 (C4): 7129-7132. https://doi.org/10.1029/JC090iC04p07129.

Wu, Aiming, William W. Hsieh, and Benyang Tang. 2006. "Neural Network Forecasts of the Tropical Pacific Sea Surface Temperatures." *Neural Networks*, Earth Sciences and Environmental Applications of Computational Intelligence, 19 (2): 145–54. https://doi.org/10.1016/j.neunet.2006.01.004.

Zhang, Wei, Gabriele Villarini, Louise Slater, Gabriel A. Vecchi, and A. Allen Bradley. 2017. "Improved ENSO Forecasting Using Bayesian Updating and the North American Multimodel Ensemble (NMME)." Journal of Climate 30 (22): 9007–25. https://doi.org/10.1175/JCLI-D-17-0073.1.

FIGURES

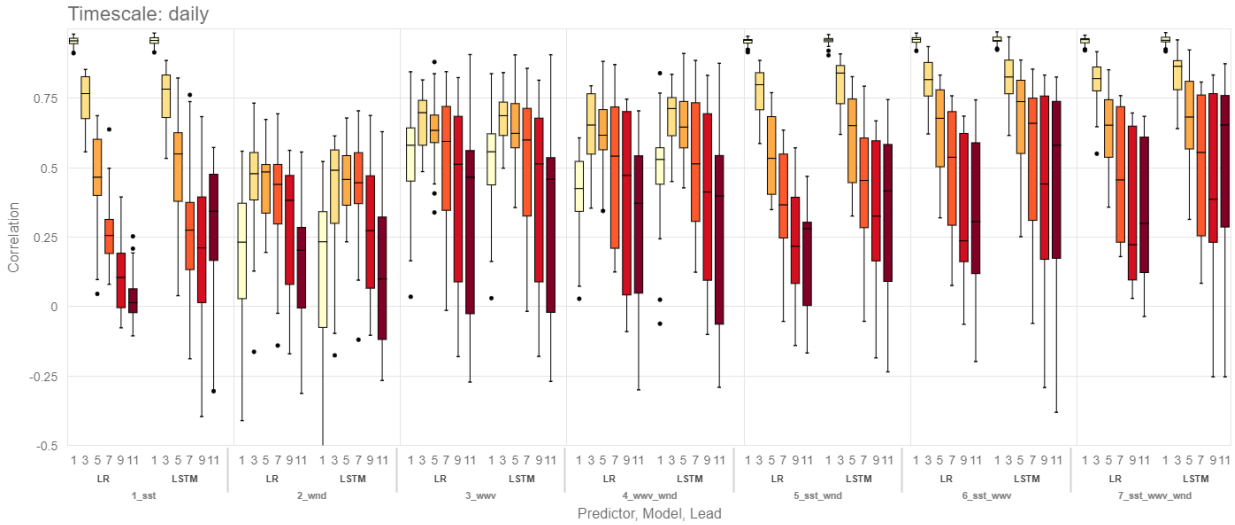

Figure 1 – Box whisker plots of correlations for each lead for each model for each variable. SST has strong correlations at short leads, but rapidly deprecates to zero at longer leads; WWV and WND however, maxes out at around five leads; thus, when SST and WND and/or WWV were combined, higher correlations at longer leads tend to persevere. Comparing LR and LSTM, with monthly data, LSTM has little advantage over LR. However, predictions done with daily timescale data, LSTM generally seems to have higher average correlation across all leads, but at the same time, it shows much more spread as well.

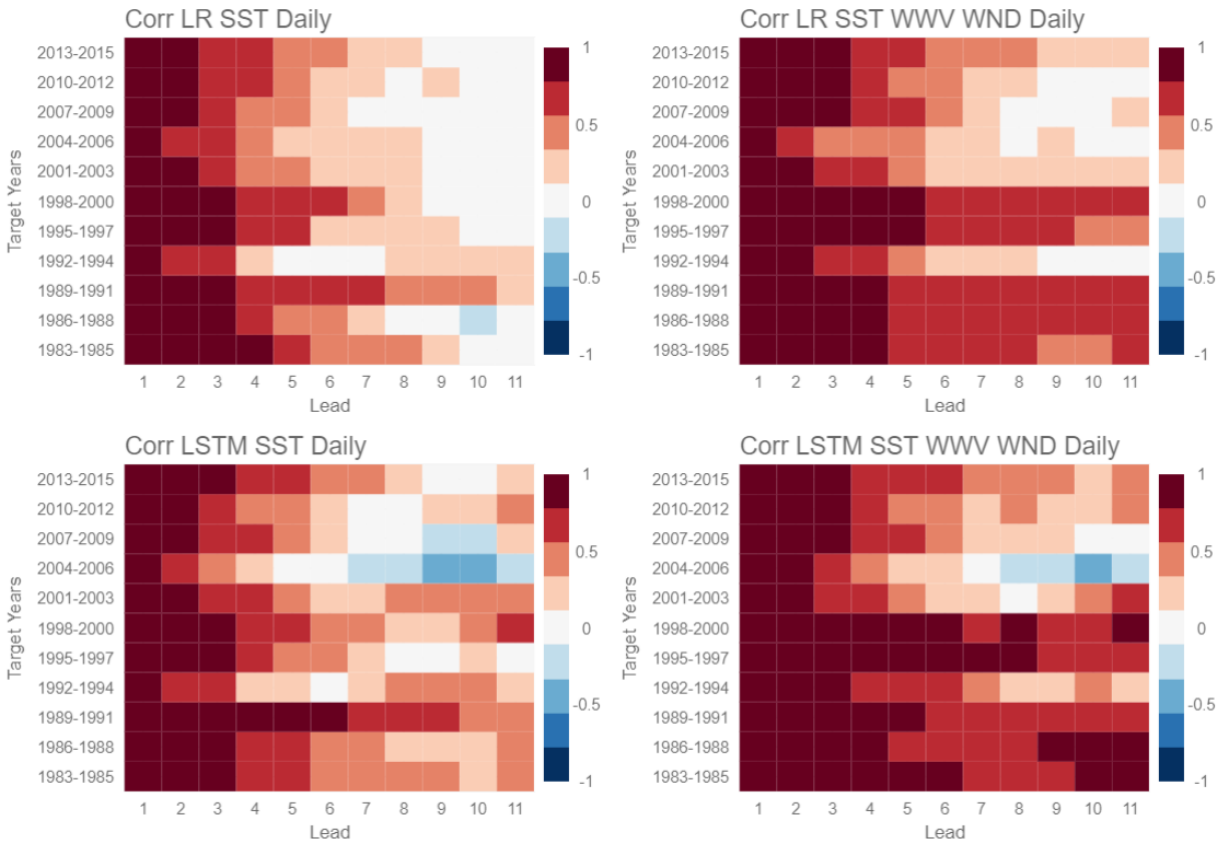

Figure 2 – Correlation heatmaps of LR (top two plots) and LSTM (bottom two plots) using daily SST (left two plots) and a combination of SST, WWV, and WND (right two plots) as predictors. Red indicates positive correlations and blue indicates negative correlations; the darker the color, the higher magnitude. High variability is exhibited across each target year, but there is a noticeable trend towards lower correlations after the 2000s (lighter reds near the top of each plot). LSTM at longer leads (near the right of each plot) generally shows higher correlations (more darker reds visible) when compared to LR, but there is a distinct period from 2004 to 2006 when there seems to be a forecast bust with negative correlations (blue).

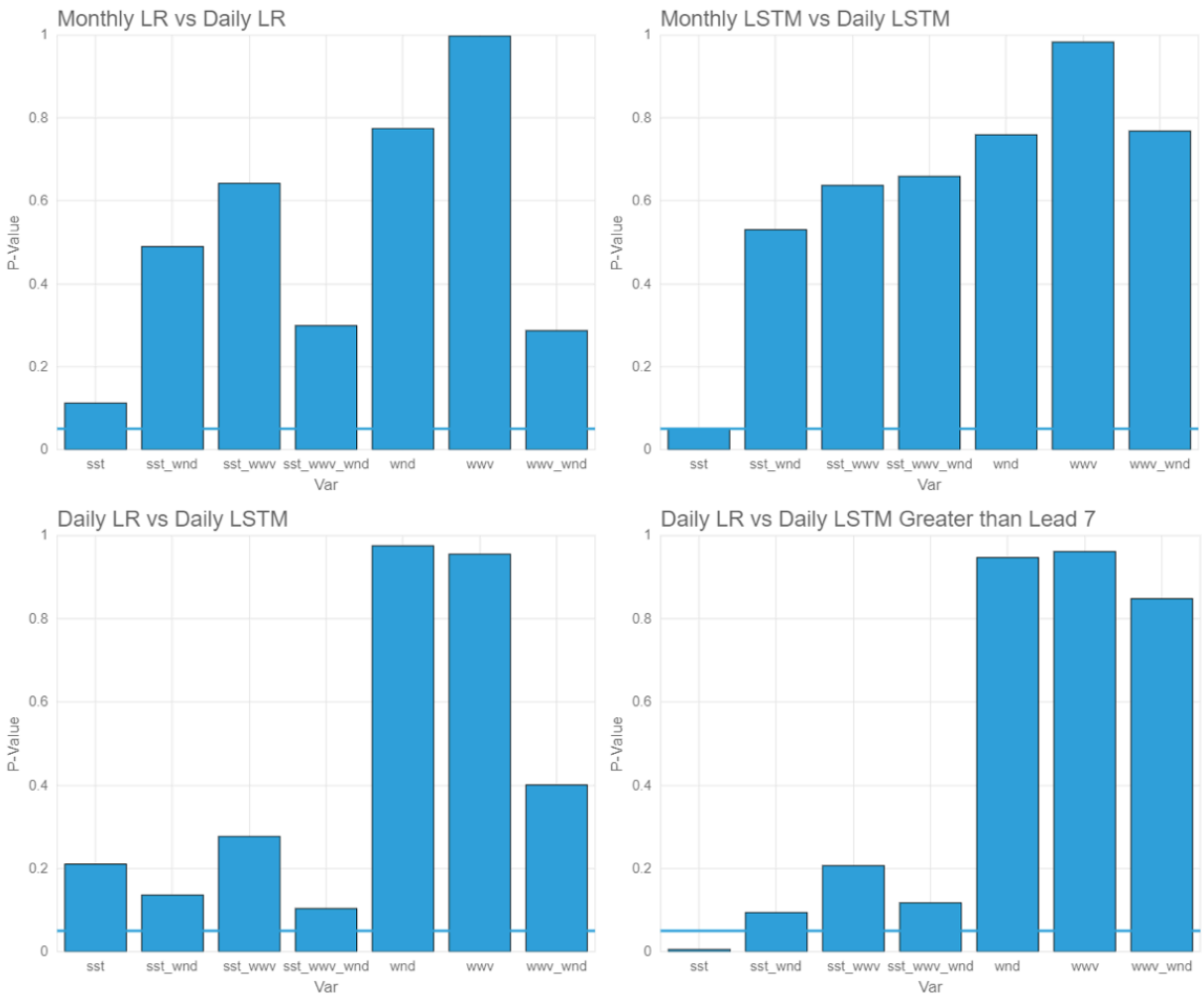

Figure 3 – Bar plots of p-value vs predictors. Top row tests whether there's a statistically significant difference with monthly and daily resolution data in LR and LSTM. Bottom row tests whether there's a statistically significant difference with daily LR and LSTM at all leads and leads greater than 7. In most cases, the p-values are greater than 0.05 indicating that the null hypothesis cannot be rejected at a 95% confidence level, but the notable exception is with the predictor, SST.

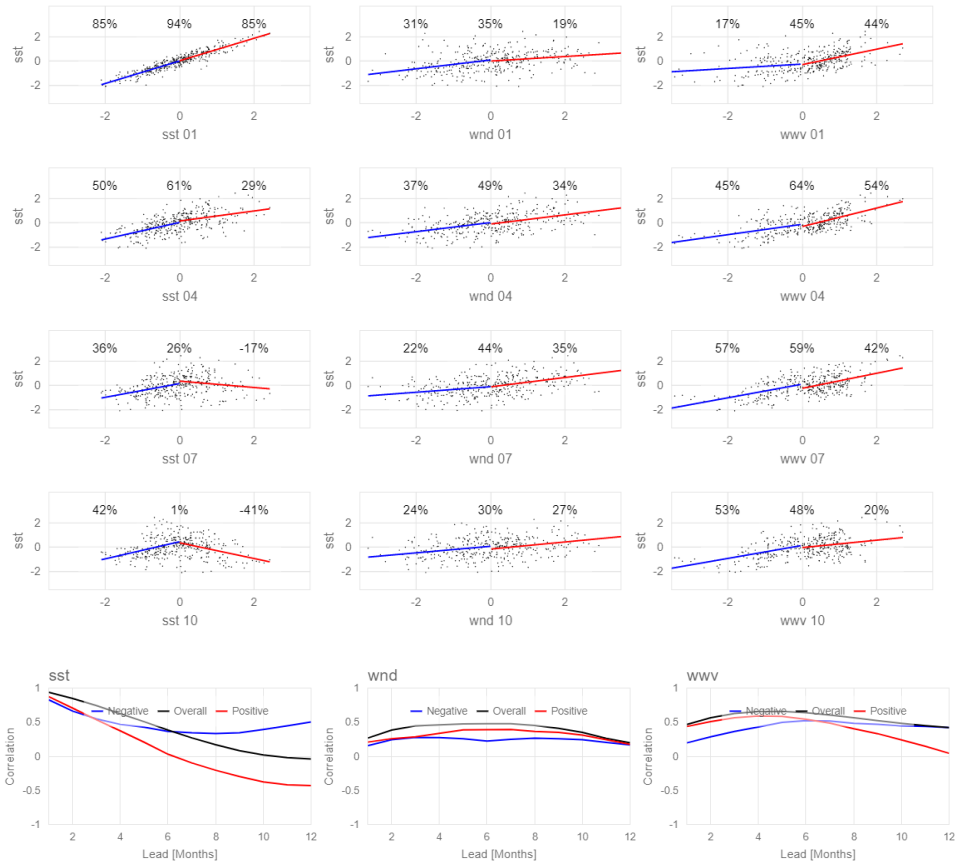

Figure 4 – Scatter plots of seasonally averaged (DJF, MAM, etc.) Nino 3.4 SST vs lead WWV, WND, or SST anomalies. Each given variable has its time series shifted backward by one, four, seven, and ten months. Lines in the first four rows represent least square fits, split by negative and positive anomalies of the given lead variable. The percentages shown indicate the correlation between the negative, overall, and positive anomalies of the given lead variable and Nino 3.4 SST. The last row is the "timeseries" of the just mentioned correlations. SST has its maximum overall correlation of 94% at lead 1 while both WND and WWV has its max correlation of 49% and 64% at lead 4. With the SST "timeseries," negative anomalies tend to preserve a positive correlation: negative SST anomalies were still negative SST anomalies ten months later, or in other words La Nina's persists as La Nina's. El Nino's, however, at ten months, show a tendency to change to La Nina's.

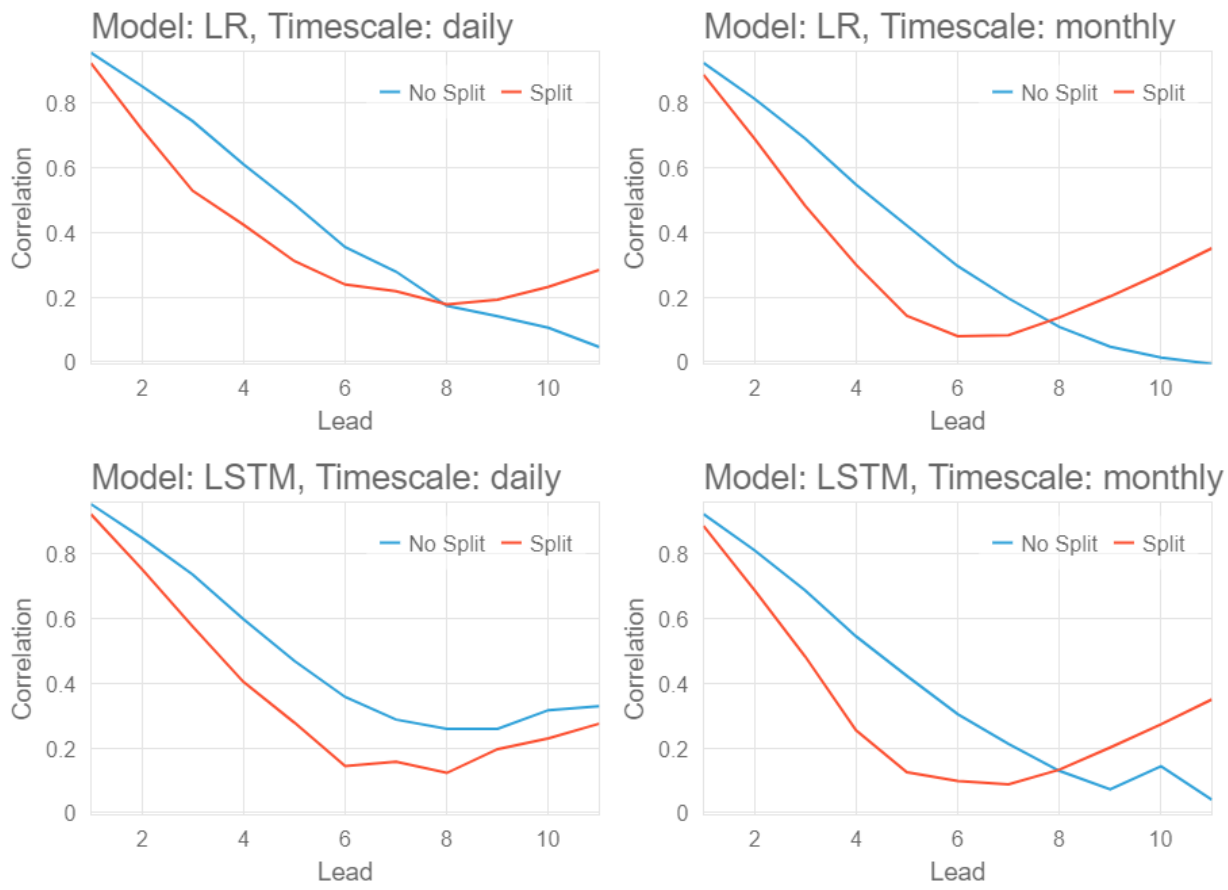

Figure 5 – Line plots of lead vs correlation using just SST as a predictor. Split signifies that each model has two separate sub-models; one sub-model is trained on when the SST anomalies were positive and the other is trained on when the anomalies were negative. Forecasts are then assigned accordingly: when the "present" SST anomalies were positive, utilize the positive model and when the "present" SST anomalies are negative, utilize the negative model. At longer leads, these split models demonstrate higher correlations than the no-split models, but not at shorter leads. Note that all models have its split version outperform the no split version, except LSTM using daily data. This suggests that LSTM can recognize this phenomenon and learn it automatically.

APPENDIX A: SUPPLEMENTAL FIGURES

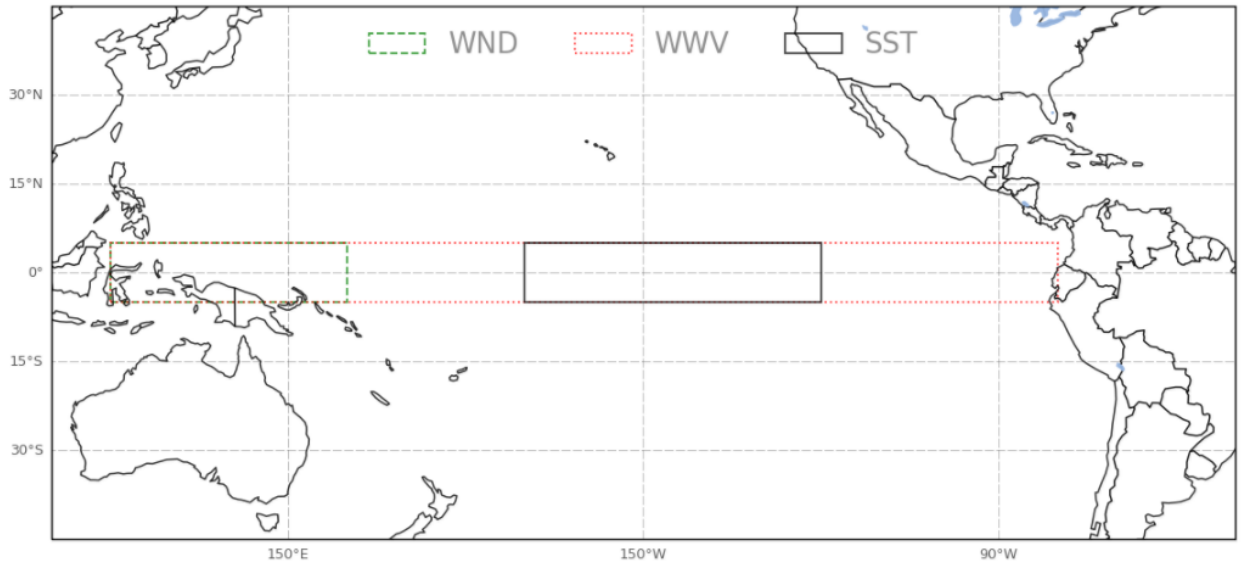

Figure 1 – Regions of predictors. WWV: 5S-5N, 120E-80W; SST: 5S-5N, 170W-120W; WND: 5S–5N, 120E–160E

	Lead 1	Lead 2	Lead 3	…	Lead 11
2012-Jan	2011-Dec-Avg	2011-Nov-Avg	2011-Oct-Avg	…	2011-Feb-Avg
2012-Feb	*2012-Jan-Avg*	*2011-Dec-Avg*	*2011-Nov-Avg*	…	*2011-Mar-Avg*
2012-Mar	*2012-Feb-Avg*	*2012-Jan-Avg*	*2011-Dec-Avg*	…	*2011-Apr-Avg*
2012-Apr	*2012-Mar-Avg*	*2012-Feb-Avg*	*2012-Jan-Avg*	…	*2011-May-Avg*
2012-May	*2012-Apr-Avg*	*2012-Mar-Avg*	*2012-Feb-Avg*	…	*2011-Jun-Avg*
2012-Jun	*2012-May-Avg*	*2012-Apr-Avg*	*2012-Mar-Avg*	…	*2011-Jul-Avg*
2012-Jul	*2012-Jun-Avg*	*2012-May-Avg*	*2012-Apr-Avg*	…	*2011-Aug-Avg*
2012-Aug	*2012-Jul-Avg*	*2012-Jun-Avg*	*2012-May-Avg*	…	*2011-Sep-Avg*
2012-Sep	*2012-Aug-Avg*	*2012-Jul-Avg*	*2012-Jun-Avg*	…	*2011-Oct-Avg*
2012-Oct	*2012-Sep-Avg*	*2012-Aug-Avg*	*2012-Jul-Avg*	…	*2011-Nov-Avg*
2012-Nov	*2012-Oct-Avg*	*2012-Sep-Avg*	*2012-Aug-Avg*	…	*2011-Dec-Avg*
2012-Dec	*2012-Nov-Avg*	*2012-Oct-Avg*	*2012-Sep-Avg*	…	*2012-Jan-Avg*

Figure 2 – Table of initial conditions (yellow) for 2012 monthly predictions (green). To predict Jan 2012 at lead 1, average of Dec 2011 was used as the initial condition. To predict Jan 2012 at lead 2, average of Nov 2011 was used as the initial condition.

Lead 1

2012-Jan	*2011-Dec-1*	*2011-Dec-2*	*2011-Dec-3*	...	*2011-Dec-30*	*2011-Dec-31*
2012-Feb	*2012-Jan-1*	*2012-Jan-2*	*2012-Jan-3*	...	*2012-Jan-30*	*2012-Jan-31*
2012-Mar	2012-Feb-1	2012-Feb-2	2012-Feb-3	...	2012-Feb-29	2012-Feb-29
2012-Apr	*2012-Mar-1*	*2012-Mar-2*	*2012-Mar-3*	...	*2012-Mar-31*	*2012-Mar-31*
2012-May	*2012-Apr-1*	*2012-Apr-2*	*2012-Apr-3*	...	*2012-Apr-30*	*2012-Apr-30*
...	...	...	...	...	...	...
2012-Dec	*2012-Nov-1*	*2012-Nov-2*	*2012-Nov-3*		*2012-Nov-30*	*2012-Nov-30*

Figure 3 – Table of initial conditions (yellow) for 2012 monthly predictions (green). To predict Mar 2012 at lead 1, Feb 1, 2012 to Feb 29, 2012 was used as the initial condition, with Feb 29 forward filled twice (to create an even data frame).

1983	1984	1985	1986	1987	1988	…	2013	2014	2015	2016	2017
1983	1984	1985	1986	1987	1988	…	2013	2014	2015	2016	2017
…	…	…	…	…	…	…	…	…	…	…	...
1983	1984	1985	1986	1987	1988	…	2013	2014	2015	2016	2017

Figure 4 – Table of training years (yellow) for validation years (green); synopsis of jackknife method.

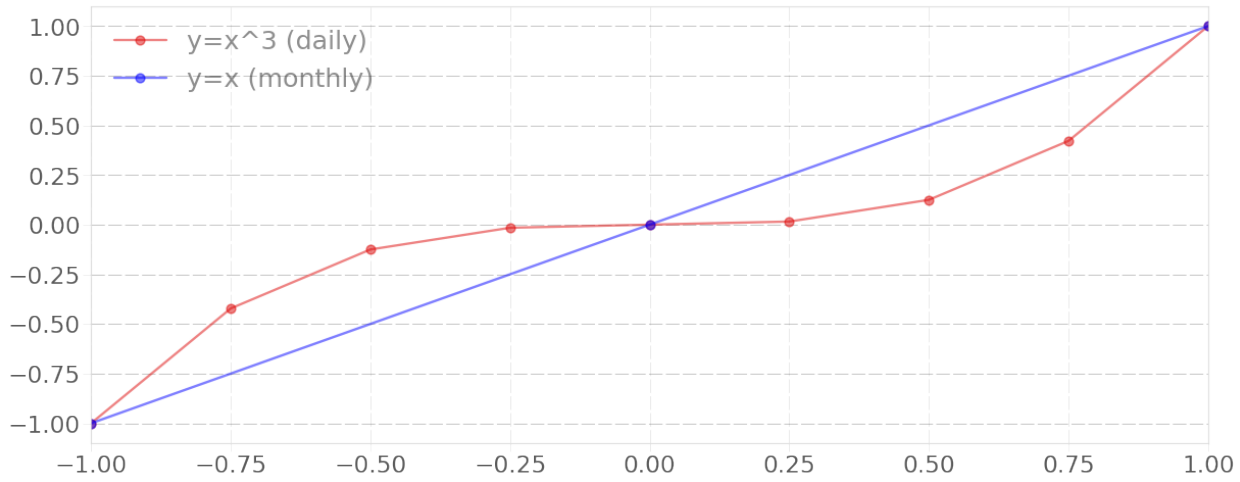

Figure 5 – Red line symbolizes daily data and can exhibit the non-linearities; the blue line symbolizes monthly data and is unable to exhibit the non-linearities.