Using Machine Learning to Predict Tornadoes in Sub-seasonal to Seasonal Time Frame

Scott A. Burgholzer1

1 Lewis University, Romeoville IL 60446, USA

scottaburgholzer@lewisu.edu

1. Introduction
   1. Sub-seasonal to Seasonal Forecasting

Sub-seasonal to seasonal forecasting (S2S) generally is trying to predict weather two weeks to a full year in advance. One big difference between weather forecasts and S2S forecasts is that weather forecasts rely heavily on initial atmospheric conditions to use as initial input, while S2S uses that plus the effects of other slowly evolving boundary conditions. These conditions can include sea surface temperature, sea ice, and soil moisture [1].

Another major difference is weather forecasts give specific values to meteorological conditions, while S2S uses probabilistic values to state if meteorological conditions will be above, below or near normal [1]. S2S is a field within meteorology that is usually overlooked over preferred methods of short term and long-range forecasts.

1. Existing Machine Learning Research in Meteorology
   1. Using Machine Learning in S2S Forecasting

When seasonal forecasting began, simple statistical techniques were being used to predict the forecast. In today’s world, however, there are complex forecast models that are used for S2S predictions and are being favored over the statistical methods previously used [2]. Upon initial research, only two journal entries have been found to use machine learning and S2S. In addition, in talking to experts in the field, they have noted that S2S is a relatively untouched sub-field and even more so in using machine learning to assist in S2S forecasts.

Cohen et al. specifically looked at using clustering methods and multilinear regression and compared the results to canonical correlation analysis (CAA) and the North American Multi-Model Ensemble (NMME). In two different studies they proved that these machine learning methods had vast improvement over CAA and NMME [2].

The first study looked at using hierarchical clustering to create winter precipitation hindcasts in Europe. The results showed that the hierarchical clustering skill was much higher than CAA and NMME. The second study looked at winter surface temperature hindcasts. In addition to the hierarchical clustering they used the Ward method. In this case, they used different permutations of meteorological inputs into the models for a total of 127 models. Out of those 127, they discovered that 16 had a skill score greater than 0.1, 12 had a score greater than 0.15 and 12 had a score greater than 0.2. They state that the NMME skill score is 0.02 which is lower than the scores they found [2].

Hwang et al. used machine learning to participate in a contest put on by the U.S. Bureau of Reclamation and the National Oceanic and Atmospheric Administration. Here they used an ensemble of two types of regression models. They showed that the models on their own and combined did much better than the Climate Forecasting System (CFSv2) [3].

The first type of nonlinear regression model Hwang et al. used was a local linear regression model with multitask feature selection (MultiLLR) and their second was a weighted local autoregression with multitask *k*-nearest neighbor features (AutoKNN). Their MultiLLR model starts adding candidate regressors from all the sources of their dataset. The model then uses a multitask backward stepwise criterion to remove all irrelevant predictors. Their AutoKNN collects features only related to the target variable, in this case temperature or precipitation. It also combines that data with a skill-specific form of nearest-neighbor algorithms [3].

Homstrom et al. used linear regression to predict high and low temperatures. They also used a variation of functional regression that predicts the weather based on searching for similar conditions in the past. They showed that professional weather forecasting continually performed better than the ML models. They also noted that the differences between their models and the professional forecast were larger early on and diminished as the forecast went further out in time. They indicated that based on these results, that ML might be better further out in time than professional forecasts [4]. The results from this study in particular show the need to continue researching forecasting S2S timeframes with machine learning.

* 1. Using Machine Learning in Other Meteorological Applications

While S2S and machine learning is a growing subfield, machine learning has been applied in many other meteorological applications. One such application is to identify Midlatitude Mesoscale Convective Systems (MCS) in Radar Mosaics. This study was performed by Alex Haberlie and Walker Ashley at Northern Illinois University. They specifically used three ML algorithms; random forest, gradient boosting and “XGBoost”. The purpose of using ML in this study was to reduce false positive identifications of MCS. They discovered three main false positive classes: unorganized clusters or lines of convective cells (UCC), tropical systems, and synoptic precipitation systems. They also added in ground clutter as a fourth class [5].

Haberlie and Ashley decided to use the Heidke skill score (HSS) as the metric to determine the best classifier configuration. They were able to get an HSS average score of 0.91 by using a random forest that uses 100 estimators, gradient boosting that uses 500 estimators and a learning rate of 0.01, and an XGBoost that uses 500 estimators and the same learning rate as gradient boosting. They also used probability of detection and probability of false detection to evaluate the class-specific performance. These two values were calculated using the best classifiers created using the testing data set [5].

Haberlie and Ashley’s results showed that MCS were able to be correctly detected 91-93% of the time while the non-MCS labels were predicted as an MCS 5-9% of the time. Each algorithm had a hard time with classifying tropical systems with the highest probability of being correctly labeled at 62%. Looking deeper at all the misclassified tropical results, 15% received an MCS label while 21% were assigned synoptic label. UCC had accuracy of 87-90%. They discovered running an ensemble did not improve over the single best algorithm, but it did improve over the single worst algorithm. These results were for a 5-class classifier. When switched to a binary classifier (MCS or non-MCS) the HSS improved to 0.93 and the accuracy improved to 0.89-0.97. In general, the binary approach did not improve much on the MCS accuracy, but it decreased false positives [5].

Liu et al. used Deep Neural Networks (DNN) to investigate using it for forecasting different meteorological conditions. First, they used a 4-layer DNN to predict temperature and dew point at the very next time point. In this case they were able to predict temperature with very little error. They do note that predicting temperature is not a very challenging task, so they move on to trying to predict Mean Sea Level Pressure (MSLP). They note that in this experiment, there are obvious errors between the actual data vs the DNN predicted value. The next experiment they ran was to predict wind speed forecasting. They made a change to the DNN from the previous two experiments and that was to add another layer and enable sparsity on the layer. They show that the DNN can predict the main trend of changes in wind speed [6].

McGovern et al. provided multiple examples of using ML within meteorology. Their first example is using ML to predict the duration of a specific storm. For this example, they used Gradient boosted regression trees, random forests, and elastic nets. They determined that for both random forests (RF) and gradient boosted regression trees (GBRT) that the best models had 100 trees and a max depth of 5. With the GBRT the best loss function was the Huber loss function. Overall, GBRT had the best results in this experiment [7].

Their next experiment was to predict severe wind (greater than or equal to 50 knots/58 MPH). In this case they used a GBRT ensemble. After the ensemble is trained, they train isotonic regression (IR) with a completely different data set. The IR is trained to correct the bias of the GBRT. Finally, combining the GBRT and IR to form a calibrated model, they test it again on another data set. The results of this experiment show that the area under receiver operation characteristic is >0.9, which they state is usually considered to be excellent. They also state that the reliability curve shows almost perfect results [7].

Ghosh and Krishnamurti used a generalized regression neural network (GRNN) algorithm to improve on hurricane intensity forecasts. In their results, they look first at the 2012 hurricane season specifically. These results showed that there was indeed a major reduction of errors on intensity forecasts. The absolute error between forecast hours of 24 and 120 were between 7 and 9 knots. The 12-hour forecast error was close to 6 knots. The neural network also shows a 10% reduction in error for forecast hours between 48 and 120 hours compared to other models. The 2013 season was looked at next where on the short term forecasts the GRNN gave errors of less than 5 knots while generally at longer term forecasts the GRNN produces the same error as an ensemble mean. Overall, the results showed that the GRNN produced forecasts with very low errors, especially for forecasts past 48 hours in advance [8].

Wimmers et. al. used ML to estimate the strength of a hurricane based on data from weather satellites. The model they created for this is called DeepMicroNet. DeepMicroNet had results that were 16 miles per hour different from the historical data of human forecasters. However, the results of DeepMicroNet did improve when they used only data measured by aircraft, where the difference was less than 11.5 mph [9].

Liu et al. used a Convolutional Neural Network to detect extreme weather by using climate datasets. The CNN they created has 4 layers, 2 of which are convolutional layers and the other two are fully connected layers. The convolutional layers have a max pooling layer immediately after it. The activation function used in both convolutional layers and the first fully connected layer is the Rectified Linear Unit (ReLU). The second fully connected layer uses a logistic activation function. Their results showed that their model had an accuracy between 89% and 99% for classification. They note that there was no overfitting seen [10].

All these studies show that there is indeed use for ML in meteorology, even if there is still more research to be done on them. These studies also give hope to ML in S2S forecasting as the current methods could be applied to and modified for S2S forecasting.

1. Capstone Proposal

This proposal looks to continue the trends of using machine learning within Sub-seasonal to Seasonal forecasts. This proposal specifically looks at using ML to improve S2S tornado prediction.

* 1. Overview

This project will use the National Centers for Environmental Prediction (NCEP) North American Regional Reanalysis (NARR) to predict tornado activity. The NARR has data from January 1, 1979 through October 1, 2019. Variables relevant to severe weather that is within the NARR include air temperature, dew point temperature, geopotential height, potential temperature, sea level pressure, upper air temperature, vertical wind velocity/speed, water vapor and wind shear.

* 1. Methodology

The first step will be to download all the NARR dataset. The Research Data Archive from NCAR/UCAR will be used. This site offers a way to get a subset of data (time frames and variables) in order to reduce the file size needed. Depending on how big this gets, it might be necessary to find a way to save just the needed values from the files and store them in a way to reduce file size.

Due to the limited time of this project, some concessions must be made. The first is an average of grid points within the NARR will have to be calculated. It would be preferred to use all the grid points individually, however this would take more time than allowed. Secondly, to increase the speed of these average calculations, a small sub region of the United States will be used. Fig. 1 shows the approximate area to be used.

While using different machine learning algorithms would also be preferred, it is planned that a neural network will be used for this project. Initially, it will be a binary output, a yes or no to tornado activity. Depending on the success of this, it may become a regression problem to output actual values.

If there is success in this study, given the changing meteorology conditions, it is expected that this model will be continuously trained to use the new data that might not have been seen before by the model or even human forecasters.



**Fig. 1.** Approximation of area to be used in this project

References

1. Zhang, Z., Cayan, D.R., Pierce, D.W.: SUBSEASONAL TO SEASONAL TEMPERATURE PREDICITON SKILL OVER ..., https://www.energy.ca.gov/sites/default/files/2019-07/Projections\_CCCA4-CEC-2018-010.pdf.
2. Cohen, J., Coumou, D., Hwang, J., Mackey, L., Orenstein, P., Totz, S., Tziperman, E.: S2S reboot: An argument for greater inclusion of machine learning in subseasonal to seasonal forecasts. Wiley Interdisciplinary Reviews: Climate Change. 10, (2018).
3. Hwang, J., Orenstein, P., Cohen, J., Pfeiffer, K., Mackey, L.: Improving Subseasonal Forecasting in the Western U.S. with Machine Learning. Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining - KDD 19. 2325–2335 (2019).
4. Holmstrom, M., Liu, D., Vo, C.: Machine Learning Applied to Weather Forecasting. (2016).
5. Haberlie, A.M., Ashley, W.S.: A Method for Identifying Midlatitude Mesoscale Convective Systems in Radar Mosaics. Part I: Segmentation and Classification. Journal of Applied Meteorology and Climatology. 57, 1575–1598 (2018).
6. Liu, J.N.K., Hu, Y., You, J.J., Chan, P.W.: Deep Neural Network Based Feature Representation for Weather Forecasting. In: Deep Neural Network Based Feature Representation for Weather Forecasting. pp. 1–7 (2014).
7. Mcgovern, A., Elmore, K.L., Gagne, D.J., Haupt, S.E., Karstens, C.D., Lagerquist, R., Smith, T., Williams, J.K.: Using Artificial Intelligence to Improve Real-Time Decision-Making for High-Impact Weather. Bulletin of the American Meteorological Society. 98, 2073–2090 (2017).
8. Ghosh, T., Krishnamurti, T.N.: Improvements in Hurricane Intensity Forecasts from a Multimodel Superensemble Utilizing a Generalized Neural Network Technique. Weather and Forecasting. 33, 873–885 (2018).
9. Verbeten, E.: Machine learning and its radical application to severe weather prediction, https://news.wisc.edu/machine-learning-and-its-radical-application-to-severe-weather-prediction/.
10. Liu, Y., Racah, E., Correa, J., Khosrowshahi, A., Lavers, D., Kunkel, K., Wehner, M., Collins, W.: Application of Deep Convolutional Neural Networks for Detecting Extreme Weather in Climate Datasets. In: Application of Deep Convolutional Neural Networks for Detecting Extreme Weather in Climate Datasets. , San Francisco.