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 advanced. 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, 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. 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].

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 nonlinear regression models. They showed that the models on their own and combined did much better than the Climate Forecasting System (CFSv2) [3].

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 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.



**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).