2.2 EXAMINATION OF SPATIAL RELATIONSHIPS USING MACHINE LEARNING TECHNIQUES

L. Clemente-Harding[1*], G. Cervone[2], L. Delle Monache[3], and S.E. Haupt[3]
[1]Engineer Research and Development Center, Geospatial Research Laboratory
[2]Department of Geography and Institute for CyberScience,
The Pennsylvania State University
[3]National Center for Atmospheric Research, Research Applications Laboratory

1. INTRODUCTION

This oral presentation investigates the impact of predictor weighting in the Analog Ensemble (AnEn) technique and begins to examine the spatial relationships present through the optimal predictor weighting results and an effort to pre-generate optimal weights using machine learning. Specifically, the AnEn is used for short-term (0-48 hour) probabilistic forecasting of wind speed over the continental United States (CONUS). An analog is a historical case that closely represents a current case. Probabilistic predictions are generated using analogs between a current deterministic forecast and set of corresponding historical forecasts and observations. The original AnEn implementation proposed by Delle Monache et al. 2013, weights each of the predictors equally in the defined similarity metric. This paper extends that original work by investigating the role of different weighting schemes in the computation of the similarity metric for 10-m wind speed and 2-m temperature. Junk et al. 2015 investigated optimal weighting strategies for the AnEn at five specific wind farms for wind power forecasting. This work extends Junk et al. 2015 contribution to 10-m wind speed and 2-m temperature predictions for 669 surface stations across a variety of terrain types, vegetation categories, and population densities.

Delle Monache et al. 2013 have shown that the AnEn can generate well-calibrated probabilistic forecasts. Given a current deterministic forecast, a set of corresponding historical forecasts and their verifying observations, a similarity metric is used to select the most similar past forecasts to the current deterministic forecast. Next, observations corresponding to the best matching historical forecasts are used to generate a probabilistic prediction. Central to the AnEn method is the definition of the metric which computes the similarity between a multivariate current deterministic forecast and a set of historical forecasts.

* *Corresponding author address:* Laura Clemente-Harding, Geospatial Research Laboratory, 7701 Telegraph Road, Alexandria, VA 22315; *email: laura.e.harding@usace.army.mil*

2. DATA

Data includes meteorological observations and corresponding historical Numerical Weather Prediction (NWP) model forecasts as defined in Delle Monache et al. (2013). The observed and historical forecast datasets, spanning a 15 month period, were divided into two sets: Set A contains 12 months of data (days 1-365) and Set B contains three months (days 366-458).

2.1 *Observations*

Available observations span a 15-month period from 1 May 2010 to 31 July 2011. Over this 457-day period, the observational dataset was collected at 669 routine aviation weather reporting stations (METAR, surface) located across the CONUS. The dataset contains 10-m Above Ground Level (AGL) wind speed and 2-m AGL temperature measurements at three-hour intervals.

2.2 *Historical Numerical Forecasts*

Historical numerical forecasts were generated by a regional version of the Environment Canada (EC) Global Environmental Multi-Scale (GEMS) model. GEMS is a variable resolution model capable of forecasting and simulating the atmosphere across global, meso-β and meso-γ scales (Cote et al. 1998).

3. METHODOLOGY

3.1 *Analog Ensemble Method*

The method builds an ensemble of analogs from deterministic NWP output (Delle Monache et al. 2013). Analogs are sought independently at each METAR station and for each lead-time. All historical forecasts are initialized at 00 UTC. The best-matching historical forecasts for the current prediction are selected as the analogs. The best match is determined by the similarity metric described in Delle Monache et al. (2013) as follows,

$$\|F_t, A_{t'}\| = \sum_{i=1}^{N_v} \frac{w_i}{\sigma_{f_i}} \sqrt{\sum_{j=-\tilde{t}}^{\tilde{t}} \left(F_{i,t+j} - A_{i,t'+j}\right)^2}$$

where F_t is the forecast for which analogs are being sought at the given time t and A_t is a historical forecast at time t' before F_t. The process is repeated independently for each METAR location. N_v and w_i are the number of physical variables and their weights, respectively. In this study, N_v is equal to four (temperature, pressure, wind speed, wind direction) and optimal w_i values are determined. σ_{f_i} is the standard deviation of the time series of past forecasts of a given variable at the same location, $\tilde{t}$ is an integer equal to half the width of the time window over which the metric is computed, and $F_{i,t+j}$ and $A_{i,t'+j}$ are the values of the current and past forecasts time window for a given variable.

The metric describes the quality of the analog chosen and is based upon the similarity of the current forecast window to the past forecast time windows available in the dataset. Analogs are ranked from most to least similar and can come from any past date within the training period. Next, the corresponding observations for each of the *n* best analogs are selected. Together, the corresponding observations generate the *n* members of the ensemble prediction for the current forecast lead-time. Delle Monache et al. (2013) show that the AnEn has several attractive features including the use of higher resolution forecasts and no need for initial conditions, model perturbation strategies, or post processing requirements. The AnEn is able to capture flow-dependent error characteristics and shows superior skill in predicting rare events when compared to state-of-the-art post processing methods (Delle Monache et al. 2011, Delle Monache et al. 2013).

3.2 *Null Case: Equally weighted predictor variables*

All four predictor variables are equally weighted in order to determine baseline improvement. This is referred to as the null case. The improvement of the equally weighted predictor variables is thoroughly discussed in Delle Monache et al. (2013) and readers are directed to this article for a full description of the AnEn null case.

3.3 *Optimal Weight Identification*

In the optimization case, weights are determined for each predictor and for each station using a brute force optimization algorithm. The brute force algorithm used is similar to the algorithm described in Junk et al. (2015) and is used to identify the optimal predictor weights by computing all possible combinations between 0 and 1.0, with a step factor of 0.1 for each of the 669 stations. A total of 257 permutations for each station are required for this analysis because only combinations of weights that add to 1 are used. The root mean square error (RMSE) between predicted (using the AnEn mean) and observed wind speed is calculated for each station, each day, and each of the possible 257 weighting combinations. Finally, the computed RMSEs for each of the days are averaged into a single value. The final result of the optimization is the set of weights for each station that achieve the lowest RMSE score among the weight combinations performed.

3.4 *Experimental Cases*

The following three cases, described in Figure 1, were used to test and validate the method:

- **Train:** The optimal weighting is identified using the training dataset (days 1 - 365) using the leave-one-out method. A total of 364 runs are required, each using a 364 day rotation for training and the remaining day for testing.
- **Test:** Weights are applied to the testing dataset (days 366 - 458). This represents an operational case and the improvement that could be obtained using optimized weights.
- **Theoretical:** The optimal weighting is identified using the testing dataset (days 366 - 458) and the leave-one-out method as previously described in the training phase. The results provide the theoretical maximum possible improvement for the testing dataset.

	Set A (days 1-365)	Set B (days 366-458)
Train	Train on Set A using leave-one-out method. Test weights on the one day left out.	
Theoretical		Train on Set B using the leave-one-out method. Test weights on the one day left out.
Test 'Operational'	Train on Set A. Test weights on Set B.	

Figure 1: Description and depiction of partitioning of data into set A and set B as well as the training, testing and theoretical cases used for both the null and optimized experiments.

4. RESULTS

Results from applying the AnEn method to surface METAR stations are presented for four cases: equal weighting (null case) for 10-m wind speed and 2-m temperature and optimized weighting for both 10-m wind speed and 2-m temperature.

4.1 *Null case*

In the null case each predictor is weighted equally and results in an average RMSE of 1.43 for 10-m wind speed predictions and 2.38 for 2-m temperature predictions when the AnEn technique is applied to set B using the search space of set A. Further descriptions of this improvement are encompassed in Delle Monache et al. (2013).

4.2 *Optimized Weighting*

Results from the optimized weights are compared to the results of the null (equal weights) case. The improvement is computed as a percentage increase (or decrease in some cases) of the performance with respect to the null case.

4.3 *10-m Wind Speed Predictions*

Figure 2 shows the behavior of the AnEn at one specific sample station out of the 669 stations. Plots (a), (b), (c), and (d) in the figure show the four predictors pressure, temperature, wind direction, and wind speed, respectively. Each figure shows the ensemble of analogs and the deterministic model forecast value over the 0-48 hour forecast period at three-hour intervals. The dotted red line shows the deterministic model forecast value and the boxplot shows the range of the ensemble of analogs chosen for each forecast lead time. The bottom panel compares the chosen analogs (grey boxplots), deterministic forecast model (red dashed line), and the corresponding wind speed observations indicated by blue circles. Note how the analogs more closely capture the observed wind speed values compared to the forecast values. Figure 3 shows improvement in all three cases (training, testing, and theoretical) for each station sorted from the least to most improved. RMSE results show an average improvement of 9.5% for the training set, 6.8% for the maximum theoretical improvement, and 4.9% improvement for the operational (real world) case. This improvement is relative to the null case. A full investigation and verification of the technique for the null case is described in Delle Monache et al. (2013). Three stations achieved greater than 15% improvement. Results indicate that maintaining the inter-relationships present among the variables and, furthermore, highlighting those relationships most important to forecast improvement through predictor weighting is beneficial.

Figure 4 shows the optimal weighting across all stations investigated using the brute force method described. The figure clearly depicts wind speed as the most important predictor receiving an average of 60% weighting across all stations. Wind direction, temperature, and pressure each require lower average weightings for optimal forecast performance. Figure 5 shows stations exhibiting greater than 10% forecast skill improvement reached through optimal predictor weighting.

4.4 *2-m Temperature*

AnEn generated for 2-m temperature show an average improvement of 52% for the training case, 2% for the maximum theoretical improvement, and 1% for the operational (real world) case improvement. Similar to the wind results, the predictor being predicted receives the greatest importance in prediction. Thus, temperature receives the highest weighting across all stations. Wind speed, wind direction, and pressure each require lower or zero weightings for optimal forecast performance. Temperature improvements are much smaller than the wind speed forecasting case. In general, NWP temperature forecasts are better than wind speed forecasts therefore there is less improvement available for temperature. Improvement is smallest in the summer and there is minimal signal in the summer for the best matching to pickup. Furthermore, the NWP model is very good at catching that lack of signal and predicting it well.

4.5 *Impact of optimal weighting on analog selection*

Clear differences in the analog selection occur when predictors are optimally weighted. When w_i is equally weighted for all predictors there is a strong seasonal bias in the analog selection. However, when w_i is optimally weighted, the seasonal analog selection bias is decreased and, sometimes, nonexistent. In some cases, if the seasonal analog selection bias is maintained, then the ordering of the analog selection from most to least similar is reordered. For example, in an ensemble of analogs when a size of *n*=21 is used, the analog selection with optimal weights will either (a) decrease in its seasonal bias or (b) the ordering of the analogs chosen will change whereby an analog with the position of two may move to position seven in the categorization of most to least similar with one being the most similar and 21 being the least similar analog within the ensemble of analogs. This research has so far used the mean of the *n*=21 ensemble of analogs, however, results will suggest how to utilize each of the individual ensemble members.

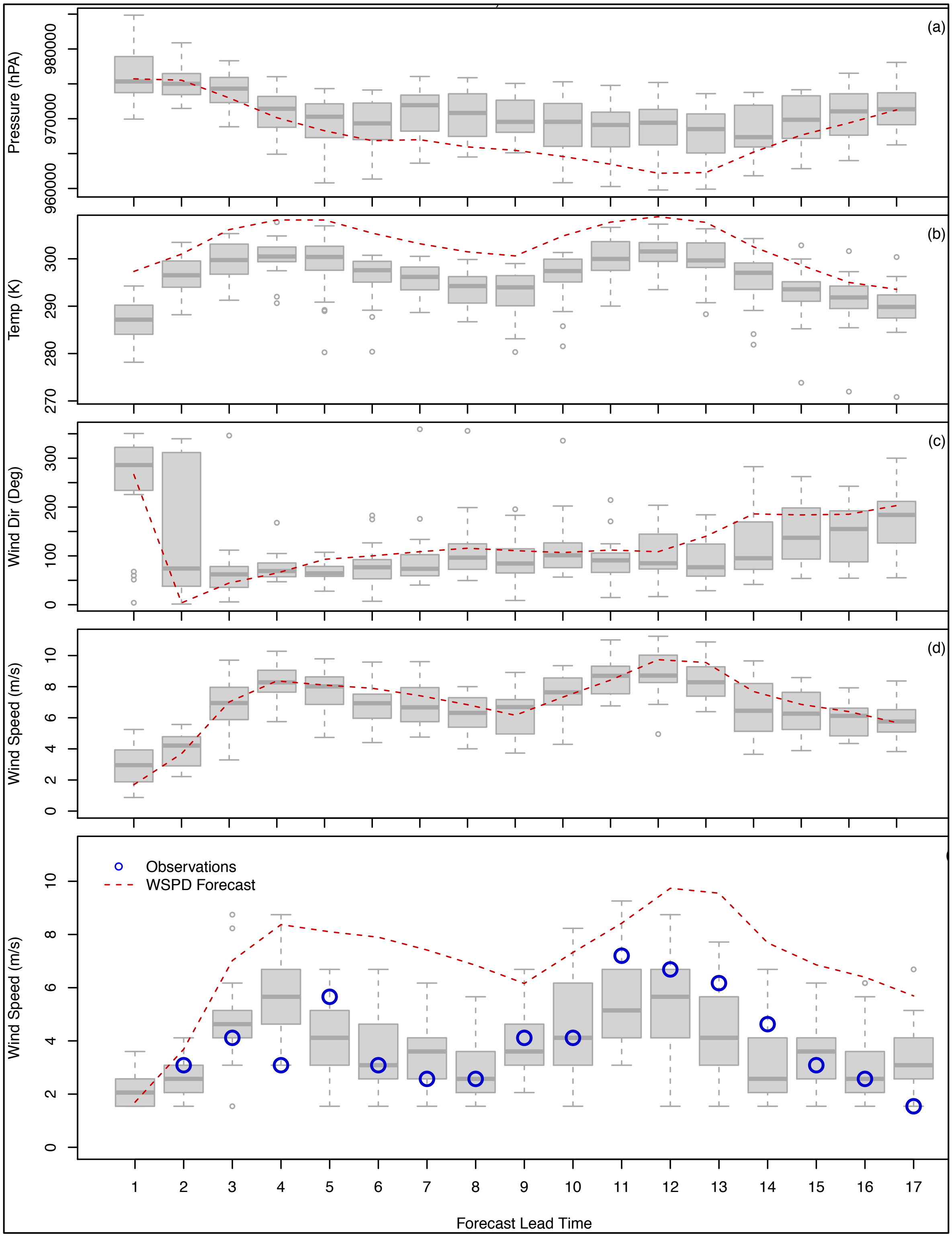

Figure 2: Specific example from one specific station. NWP model forecasts for each of the four parameters are depicted. The dotted red line indicates the deterministic model forecast value. The horizontal axis provides the time interval of the forecast from 0-48 hours at three-hour intervals. The bottom box depicts wind speed forecasts from the deterministic forecast model, corresponding wind speed observations indicated by a blue circle, and boxplots depict the range of analogs chosen.

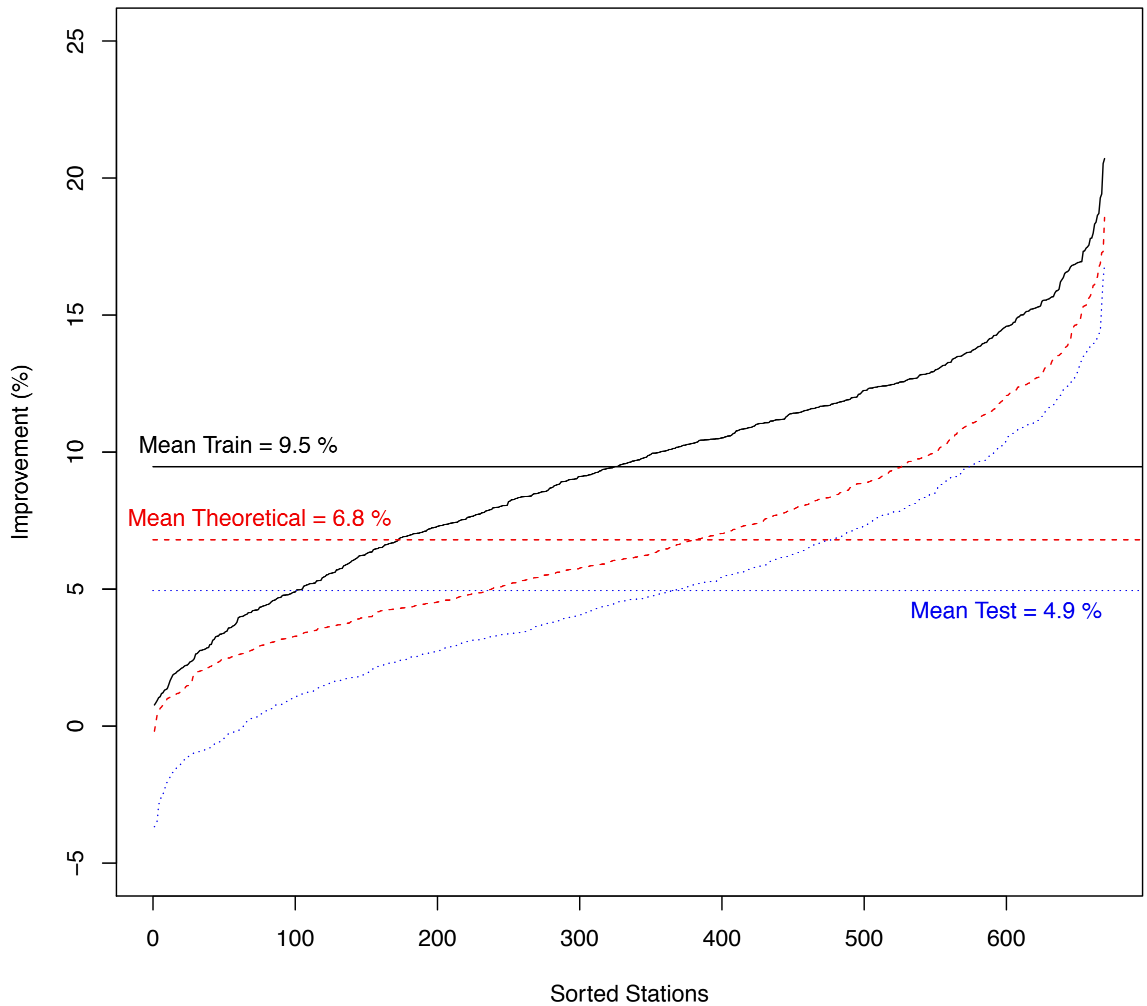

Figure 3: Improvement by station of the AnEn forecasts for wind speed compared to the equally weighted AnEn null case for each of the three cases: training, testing, and optimization. Results are sorted from the least to the most improvement observed at each station and improvement is based on the minimization of RMSE.

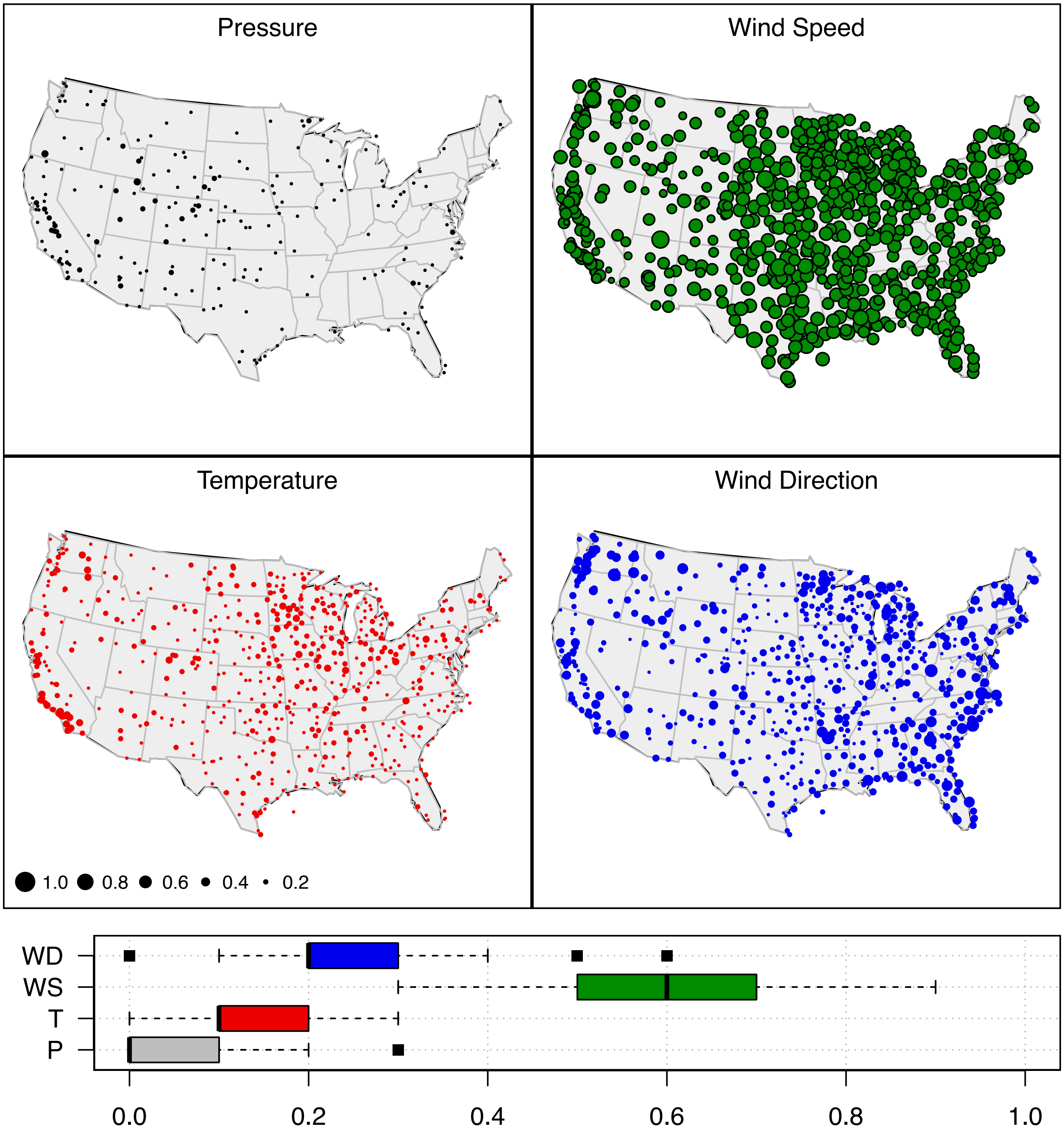

Figure 4: Top: The four panel chart depicts the optimal weight at each station for the four predictor variables: pressure, wind speed, temperature, and wind direction. The optimal weighting value corresponds to the circle size drawn. Only stations with weights greater than zero are plotted. Bottom: Boxplot describing the overall distribution of parameter weighting for all 669 stations. The boxplot identifies the range of weights across all 669 stations, the black line within the box represents the median, and the black squares indicate outliers.

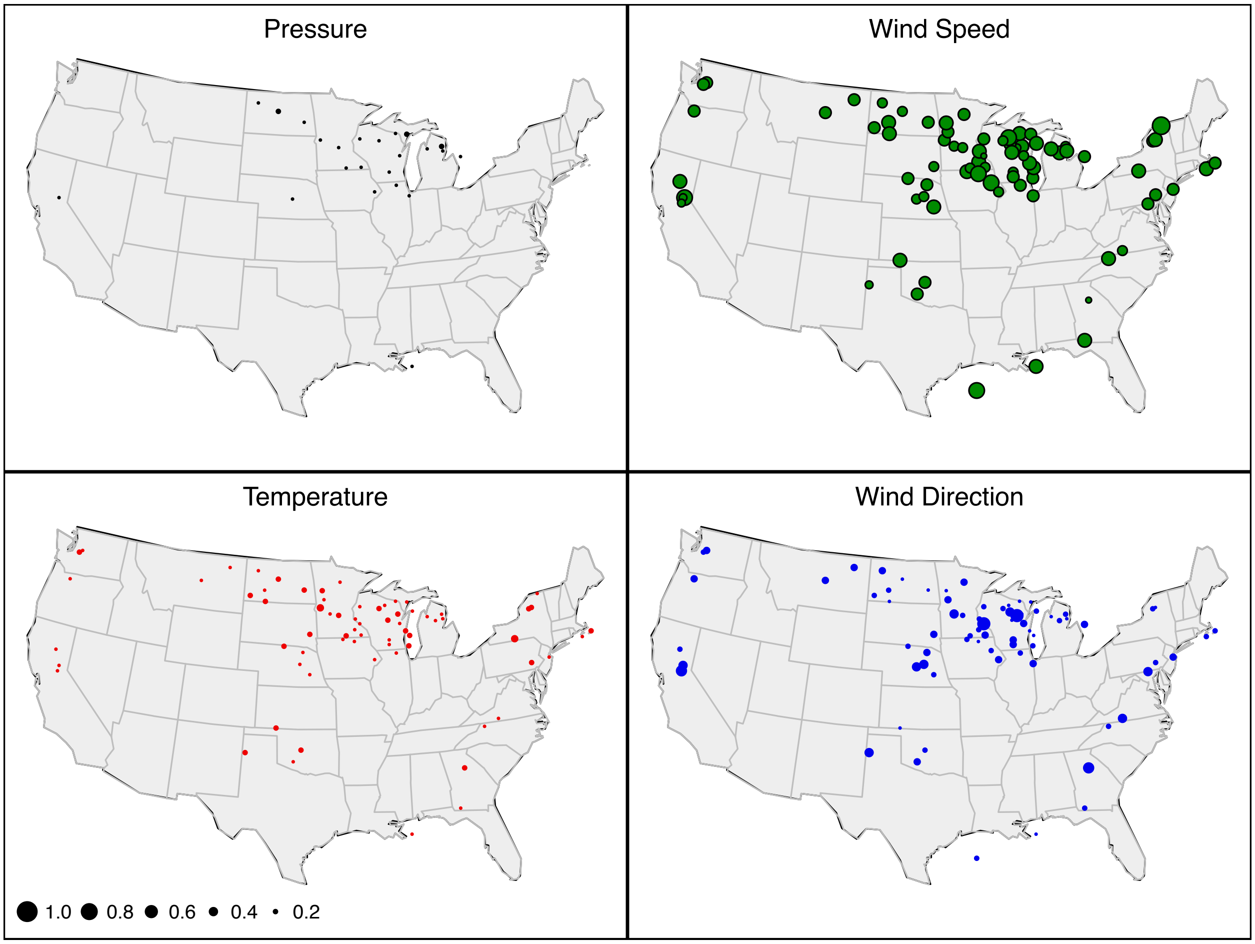

Figure 5: Four panel chart, similar to Figure 4 (top), showing all stations with greater than 10% forecast improvement and the corresponding optimal weights indicated by circle size.

4.6 *Machine learning to determine optimal weights*

In this component of the research we seek to determine the optimal weights without the use of a brute force algorithm and with only the available historical forecasts and observations, and if there is a spatial component to this. First, a machine learning classification effort is used. However, the number of categories of optimal weights are too large for the size of the dataset available. Next, an unsupervised clustering algorithm called *k*-means is used. For this, dataset A is used and broken into seasons. *k*-values from two to 30 are tested and each test is performed 30 times for a total of 840 *k*-means tests. Stations remaining in the same cluster throughout the entire experiment were annotated as anchor points. These anchor points were used to identify other stations within a cluster and then enable investigation into the behavior at the stations in each cluster. While this developed an understanding of the characteristics in each cluster, this did not result in a clear way to pre-determine the optimal weights without using a brute force algorithm and additional machine learning techniques will need to be studied.

5. CONCLUSION

The optimal combination of AnEn predictor weighting was investigated for 10-meter wind speed and 2-m temperature forecasting for each METAR stations across the CONUS. The AnEn has shown to be an effective method for providing bias corrected and calibrated short-term forecasts. In the majority of previous studies, metric weighting was equal for all predictors and optimally weighted results focused on a small subset of solar or wind farms. Research results show that optimal weighting can lead to improved forecast performance of 4.9% on average with some stations seeing greater than 15% improvement for predicting 10-m wind speed with fewer computational resources expended. Results show improvement in 2-m temperature forecasts, albeit smaller than the wind improvements. In particular, the optimal weighting investigations highlight the impact of optimal weighting on analog selection within the search space. When predictors are weighted equally, there is a seasonal bias to the analog selection within the search space whereas with optimal predictor weighting this bias is significantly reduced.

This research uncovers a spatial variation to the optimal weighting strategies whereby stations with similar weights appear to cluster with some potentially geographically related regions depending on terrain, land use land cover, and general synoptic features. Future study will investigate the underlying reasons for the variability in station improvement with optimal weighting. The objective would be to determine the differing characteristics for each site and the contribution to improved forecast skill. Knowledge from this research can lead to rapid predictions of parameters near the land surface using the AnEn. This method and optimal weighting could be utilized in regions where forecast models are known to have significant bias, ensemble forecasting is too expensive or computationally prohibitive, and a record of historical observations exists.

6. REFERENCES

Côté, J., Desmarais, J. G., Gravel, S., Méthot, A., Patoine, A., Roch, M., & Staniforth, A. 1998: The operational CMC–MRB global environmental multiscale (GEM) model. Part II: results. *Monthly Weather Review*, *126*(6), 1397-1418.

Delle Monache, L., Nipen, T., Liu, Y., Roux, G., & Stull, R. 2011: Kalman filter and analog schemes to postprocess numerical weather predictions. *Monthly Weather Review*, *139*(11), 3554-3570.

Delle Monache, L., Eckel, F. A., Rife, D. L., Nagarajan, B., & Searight, K. 2013: Probabilistic weather prediction with an analog ensemble. *Monthly Weather Review*, *141*(10), 3498-3516.

Junk, C., Delle Monache, L., Alessandrini, S., Cervone, G., & von Bremen, L. 2015: Predictor-weighting strategies for probabilistic wind power forecasting with an analog ensemble. *Meteorologische Zeitschrift*, *24*(4), 361-79.