1.4 IMPROVING SNOWFALL ACCUMULATION PREDICTIONS BY POST-PROCESSING ENSEMBLE FORECASTS WITH AN ARTIFICIAL NEURAL NETWORK

Tyler McCandless*, Sue Ellen Haupt, George Young
[1]The Pennsylvania State University

1. INTRODUCTION

Meteorologists face the difficult task of forecasting complex winter storm systems that can affect millions of people. Forecasting snowstorms is a multifaceted problem with many challenges. Kocin and Uccellini (2005) state that "one can conclude that the accuracy of weather forecasting, even for a rare event such as a major Northeast snowstorm, can be attributed to the introduction of numerical models into the forecast process beginning in the 1950s, the continued improvements made to the numerical models and global data, and the overall professional development of forecasters whose training and education are based heavily on understanding the strengths and weaknesses of the models." In addition to improvements made to the numerical models, improvements in weather forecasting have come from the development of statistical post-processing methods of weather forecasting. Glahn and Lowry (1972) implemented Model Output Statistics (MOS) post-processing that is an objective weather forecasting technique. MOS consists of determining a statistical relationship between a predictand and variables forecast by a numerical model at various projection times. The development of linear regression equations that relate model predicted variables to weather observation enabled the prediction of weather variables not directly forecast by the model, like snow accumulation.

Another development in weather forecasting has been the advent of meteorological ensembles which quantify forecast uncertainty by representing possible realizations of future states of the atmosphere. Advanced statistical post-processing techniques have recently been developed and implemented in order to improve the calibration and the accuracy of NWP ensembles. Several studies have examined different methods of post-processing ensemble forecasts in order to improve weather prediction (Raftery et al. 2005, Greybush et al. 2008, Glahn et al. 2009). Although many of these advanced statistical post-processing methods have been shown to improve general forecasting, only recently have there been attempts to use post-processing to improve snowfall accumulation predictions. Cosgrove and Sfanos (2004) have applied the MOS post-processing technique to forecast the conditional probability of snow and the snowfall amount exceeding a certain threshold, given that snowfall occurs, using the Global Forecast System (GFS) model. Several studies have attempted to improve snowfall forecasting by more accurately predicting the snow density (Roebber et al. 2003, Baxter et al. 2005, Ware et al. 2006, Roebber et al 2007). There have been attempts at improving weather predictions from the use of Numerical Weather Prediction (NWP) ensemble forecast systems, the use of advanced statistical post-processing, and developing an accurate snowfall density climatology. However, no studies have attempted to apply statistical post-processing to current NWP ensemble prediction systems. The goal of our study is to use advanced statistical guidance methods to post-process forecasts from the Global Ensemble Forecast System (GEFS) in order to improve both the accuracy and ensemble calibration 24 hour snowfall accumulation predictions.

In section 2, we discuss the datasets: the Global Ensemble Forecast System and the cooperative observing network. In section 3, we discuss the statistical guidance methods used. In section 4, we summarize and discuss the results. In section 5, we provide conclusions and ideas for future research.

* Corresponding author address: Applied Research Laboratory, P.O. Box 30, The Pennsylvania State University, State College, PA 16804-0030
Email: tcm5026@psu.edu

2. STUDY DATA

a. *Verification Data*

In order to test the validity of any forecast and allow for consistent, reliable statistical post-processing, an accurate observing system is necessary (Allen 2001). The National Climatic Data Center's (NCDC's) Cooperative Summary of the Day reports, or co-op, are used as our snowfall observing network. The co-op reports are taken daily by volunteers at their home or work and then sent to NCDC who collects and processes the data. The variables reported once per 24-hr period are precipitation amount, snowfall accumulation, maximum temperature, and minimum temperature. Co-op stations are established, closed, supervised, and inspected by NWS personnel with annual visits to ensure observer proficiency, adherence to instrument and exposure standards, and network integrity (NWS 2000). This dataset has many challenges, such as data formatting, quality control, varying reporting times, and station changes. The data has been quality controlled by the National Climatic Data Center (NCDC) with 36 different checks for consistency and quality (NCDC 2000) and also quality controlled by the National Weather Service Meteorological Development Laboratory (MDL). The quality control by MDL has eliminated observations with subjective information; such as, "accumulated amount since last measurement" or "subjectively derived value."

Also, all cases that reported snowfall but not precipitation were eliminated from the dataset.

The next step was determining which observations to use because the co-op sites report once per day. Thus, the hour at which the observation takes place varies between stations. In order to maintain consistency and to provide a valid test of our methods, we have kept all observations between 11z and 17z in order to compare forecasts valid for 12z. Not only are most of the observations recorded between 11z and 17z, but this correlates with a lead time of 12 hours from the Global Ensemble Forecast System. Using this time period parallels the approach used by Cosgrove and Sfanos (2004). Only the observations with a snowfall measurement of a trace or more are retained in the dataset.

b. Ensemble Forecast Data

The National Center for Environmental Prediction (NCEP) Global Ensemble Forecast System (GEFS) is an ensemble forecast system using the Global Spectral Model. Due to several changes in the model configuration and number of members, the longest cold season consistent dataset available was October 1, 2006 to March 31, 2007. During this time period, the GEFS consisted of 15 total ensemble members: one high resolution control run and seven paired perturbations from the NCEP breeding method with Ensemble Transform (ET). The initialization time of the forecasts is 00z each day and each forecast is archived at 95.25 km resolution. The GEFS direct model output consists of forecasts for every six hours from 0-364 hrs. There are 31 forecast elements from the GEFS, which are listed in Table 1. There are four wind speeds forecast for both U- and V-wind components, five temperature levels forecast, four categorical precipitation categories: rain, freezing rain, ice pellets, and snow, four geopotential height fields, six-hour maximum temperature, minimum temperature, and accumulated precipitation. There are also four relative humidity levels forecast and two pressure fields.

In order to compare the GEFS forecasts on a 95.25 km grid to individual co-op reporting sites, a nearest neighbor weighted method was used to interpolate to the co-op locations. First, the closest three grid locations to the co-op sites were selected. The nearest neighbor process then converts the grid point locations and co-op reporting sites from spherical to Cartesian coordinates. Next, the respective distances between the three nearest grid points and the co-op reporting sites are computed. To calculate a forecast for the co-op location, a distance-weighted average of the three nearest neighbor grid points is used. The distance-weighted average forecast for all 31 elements is used in the training data.

The dataset consists of GEFS predictions for all 31 weather variables at forecast valid times of 12-18 hrs, 18-24 hrs, 24-30 hrs, and 30-36 hrs. These 124 variables are combined with each station's latitude, longitude, and elevation as predictors into the statistical guidance model. The statistical guidance methods use these variables to predict the total 24 hr snowfall accumulation. This methodology is displayed in Figure 1. The 11Z-17Z co-op observation, which is approximately 7am for observations in the Eastern Time Zone, correlates with a 12-36 hr prediction from the GEFS.

Table 1. GEFS archived elements.

Element	**Element Description**	**Levels**
Press	Pressure [Pa]	Surface
PRMSL	Pressure reduced to MSL [Pa]	Mean Sea Level
RH	Relative humidity [%]	2M,925mb,850mb, 700mb,500mb
TMP	Temperature [K]	2M,1000mb,850mb, 700mb,500mb
TMAX	Maximum temperature in 6	2-M
TMIN	Minimum temperature in 6	2-M
U GRD	U-comp of wind [m/s]	10-M, 850mb, 700mb, 500mb
V GRD	V-comp of wind [m/s]	10-M, 850mb, 700mb, 500mb
HGT	Geopotential height [gpm]	1000mb, 850mb, 700mb, 500mb
FRZR	Categorical Freezing Rain [1=yes;0=no]	2-M
ICEP	Categorical Ice Pellets [1=yes;0=no]	2-M
SNOW	Categorical Snow [1=yes;0=no]	2-M
RAIN	Categorical Rain [1=yes;0=no]	2-M
PRCP	6h accumulation of Total precipitation [kg/m^2]	2-M

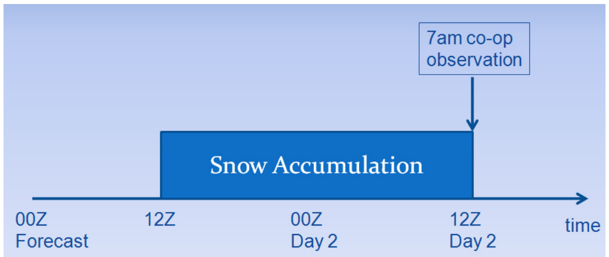

Figure 1. Process layout for the 12-36hour snow accumulation forecast.

3. STATISTICAL GUIDANCE METHODS

In order to forecast 24 hr snow accumulation, a method must be devised that translates the predicted weather variables from the GEFS as well as the latitude, longitude, and elevation from the co-op sites into a 24 hr snow accumulation forecast. These are statistical guidance or post-processing methods because they are used after the model has output its predictions. Glahn and Lowry (1972) developed Model Output Statistics (MOS), which determines a statistical relationship between a predictand and the variables forecast by a numerical model at some projection times. MOS uses a linear regression equation to relate the predicted variables to the predictand. MOS is still the statistical technique used in operational forecasting at the NWS to predict variables not explicitly forecast in the model. Thus, we use linear regression as our baseline method for comparison. We attempt to improve upon MOS by capturing non-linear relationships among the predictors using an Artificial Neural Network as depicted in Figure 2. This simplified diagram shows four predictors fed into one hidden layer consisting of five nodes. These five nodes are connected to the output layer, or prediction, which is the 24 hr snow accumulation. The ANN used in this study is a feed-forward neural network trained by a backpropagation algorithm, also known as a multi-layer perceptron. This ANN has a learning rate of 0.3 and a momentum of 0.2. The ANN goes through 50 training cycles to find the optimal model. The activation function is the standard sigmoid function, with the predictor values are scaled to range from -1 and +1. The type of the output node is linear for numerical regression tasks such as snowfall accumulation forecasting.

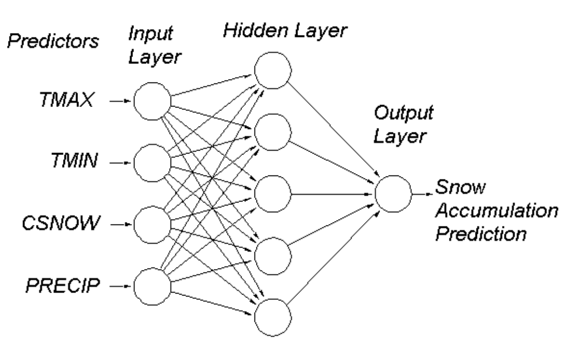

Figure 2. Diagram of an Artificial Neural Network.

The linear regression and ANN statistical guidance methods were trained on the high resolution ensemble member using a ten-fold cross validation. After the Root Mean Square Error (RMSE) of the 24 hr snowfall accumulation prediction is minimized in the cross validation, the statistical guidance models are saved. The linear regression and ANN models are next applied to each ensemble member individually to predict the 24 hr snow accumulation.

4. RESULTS

To test the deterministic forecast accuracy of the statistical guidance methods, the ensemble mean consensus forecast is calculated. This consensus forecast is calculated by averaging the 15 individual ensemble members. The Mean Absolute Error (MAE) of the consensus forecast is the mean absolute difference between the consensus orecast and the observation averaged over all 80219 observations. The MAE calculated for the linear regression method is 0.0328 meters, or 1.29 inches. The MAE for the ANN method is 0.0297 meters, or 1.17 inches. The ANN not only produces a lower MAE for the consensus forecast, but also predicts higher snowfall accumulations than the linear regression, as shown in Figures 3 and 4.

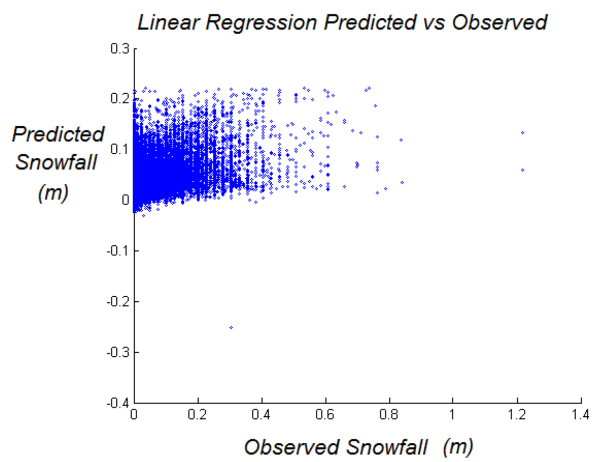

Figure 3. Linear regression predicted snow accumulation versus the observed snow accumulation.

The ability to predict greater snowfall totals is beneficial because larger snow accumulations tend to produce the most impact. Figure 3 plots the linear regression predicted snow accumulation versus the observed snow accumulation. From this plot, it is clear that the linear regression does not predict snow accumulations greater than 0.25m and also predicts negative snowfall values. Figure 4 plots the ANN predicted snow accumulation versus the observed snow accumulation. The ANN predicts snowfall accumulations as high as 0.6m and does not predict negative values like linear regression.

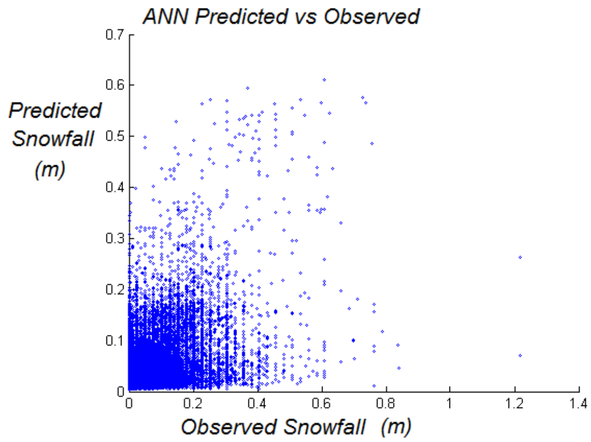

Figure 4. Artificial Neural Network predicted snow accumulation versus the observed snow accumulation.

It is also important to evaluate the ensemble spread, or the forecast uncertainty, from the statistical guidance methods. One method to display ensemble spread is rank histograms. The rank histogram was developed independently by Anderson (1996), Hamill and Colucci (1996, 1997), and Talagrand et al. (1997) to quantify ensemble dispersion. The rank histograms for the linear regression method, Figure 5, and for the ANN method, Figure 6, both display under-dispersive ensembles.

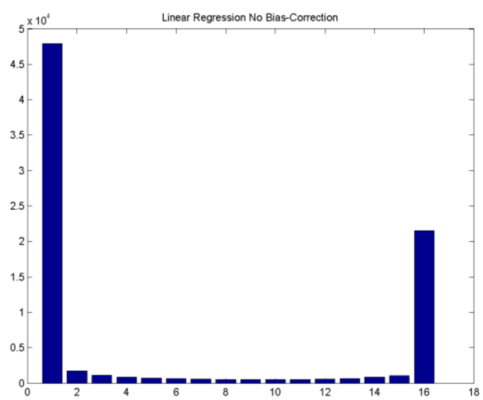

Figure 5. Rank histogram for linear regression method. Bins one and nine are significantly higher than the rest, indicating the ensemble is under-dispersive.

This is evidenced by bins one (leftmost) and 16 (rightmost) having higher tallies than the other bins, meaning that the snow accumulation observation tended to be less than all of the ensemble member forecasts or greater than all of the ensemble member forecasts. These rank histograms show that neither method produces a calibrated ensemble, which would provide accurate forecast uncertainty information.

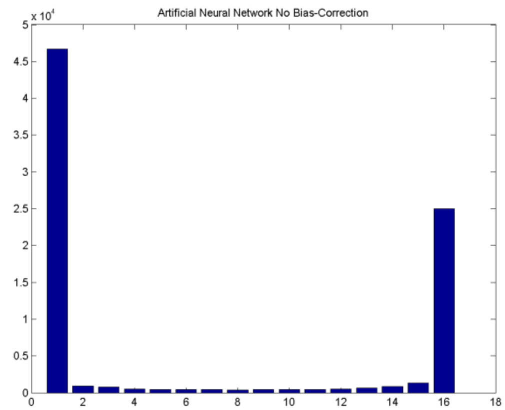

Figure 6. Rank histogram for linear regression method. Bins one and nine are significantly higher than the rest, indicating the ensemble is under-dispersive.

Another method of examining the ensemble's uncertainty information is through spread-skill relationships, which is a measure of the correlation between the ensemble spread and the ensemble mean error (Whitaker and Loughe 1998). The ensemble spread is calculated by subtracting the lowest snow accumulation forecast by an ensemble member from the highest snow accumulation forecast by an ensemble member. The ensemble error is the absolute difference between the ensemble mean consensus forecast and the snow accumulation observation. A calibrated ensemble should show a one-to-one correlation, or unity, between the ensemble error and the ensemble spread. For example, if the error of the forecast is 0.1 m, the ensemble spread should be 0.1 m. The spread-skill plots for the linear regression method and the ANN method are shown in Figures 7 and 8 respectively. The slope of the best fit line represents the relationship between the ensemble error and ensemble spread. The slope of the best fit line of the ANN method is closer to one, or unity, than the slope of the best fit line of the linear regression method, 0.552 to 0.496 respectively. The correlation coefficients for the ANN and linear regression methods are 0.344 and 0.217 respectively. The larger correlation coefficients for the ANN method suggest that it produces better uncertainty estimates. These spread-skill relationships show that the ANN method better represents the uncertainty in the snow accumulation forecasts than the linear regression method.

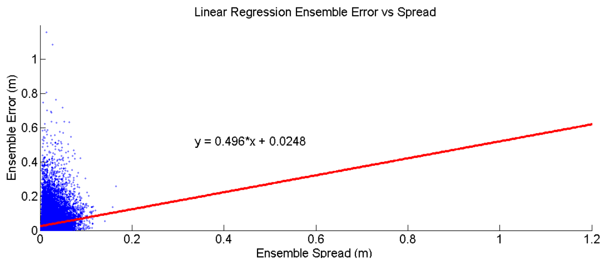

Figure 7. Ensemble error versus ensemble spread for the linear regression method. Blue points represent the 80219 pairs of ensemble spread and ensemble error. The red line is the linear best fit line to the points. The slope of the best fit line, 0.496, represents the relationship.

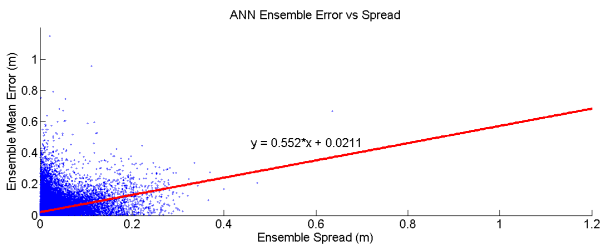

Figure 8. Ensemble error versus ensemble spread for the ANN method. Blue points represent the 80219 pairs of ensemble spread and ensemble error. The red line is the linear best fit line to the points. The slope of the best fit line, 0.552, represents the relationship.

5. CONCLUSIONS AND FUTURE WORK

We have tested two statistical guidance methods for producing 24 hr snow accumulation forecasts from the Global Ensemble Forecast System output. These methods were trained to reduce the error of the control ensemble member and then applied to each ensemble member individually. Averaging these individual ensemble members into a single consensus forecast produces a deterministic snow accumulation forecast. The spread, or the difference between the highest and lowest ensemble member forecasts, represents the uncertainty in the forecast. The linear regression method is used as our baseline since it is the closest method to the National Weather Service standard operational statistical guidance method, Model Output Statistics. The ANN method attempts to improve upon linear regression because it can model non-linear relationships between predictors.

The results indicate that the ANN statistical guidance method of predicting 24 hr snow accumulation provides somewhat more accurate deterministic forecasts than the linear regression method. The ANN method also has a spread-error correlation closer to the ideal 1-1 ratio than the linear regression method, indicating that the ANN method produces better forecast uncertainty estimates. The rank histograms show that both methods produce under-dispersive ensemble; however, Marzban et al (2010) has shown that U-shaped rank histograms may not be due to under-dispersive ensembles but a result of correlations among ensemble members. Rotated Q-Q plots will be included in future work to further test the ensemble reliability. The next step is to improve the calibration of the ensemble in order to provide more accurate forecast uncertainty information. Future work will also test the snowfall predictions on longer lead times, specifically at a lead time of 36 hours that corresponds to winter weather watches issued by the NWS. Future work will also test other statistical guidance methods of forecasting snow accumulation, including random forests and Bayesian approaches.

Acknowledgements: This research is supported by the Educational and Foundational Program of the Applied Research Laboratory of the Pennsylvania State University. The authors wish to thank Dr. Bob Glahn and the rest of the National Weather Service Meteorological Development Laboratory for access to the data as well as helpful guidance and suggestions.

References

Anderson, J. L., 1996: A method for producing and evaluating probabilistic forecasts from Ensemble Model Integrations. *J. Climate*, **9**, 1518–1530.

Baxter, M.A., C.E. Graves, and J.T. Moore, 2005: A Climatology of Snow-to-Liquid Ratio for the Contiguous United States. *Wea. Forecasting*, **20**, 729–744.

Cosgrove, R. L., and B. Sfanos, 2004: Producing MOS Snowfall Amount Forecasts from Cooperative Observer Report. Preprints AMS 20th Conference on Weather Analysis and Forecasting, January 11-15, 2004, Seattle, WA.

Evans, M., and M.L. Jurewicz, 2009: Correlations between Analyses and Forecasts of Banded Heavy Snow Ingredients and Observed Snowfall. *Wea. Forecasting*, **24**, 337–350.

Glahn, H.R., and D.A. Lowry, 1972: The Use of Model Output Statistics (MOS) in Objective Weather Forecasting. *J. Appl. Meteor.*, **11**, 1203–1211.

Glahn, B., M. Peroutka, J. Wiedenfeld, J. Wagner, G. Zylstra, B. Schuknecht, and B. Jackson, 2009: MOS Uncertainty Estimates in an Ensemble Framework. *Mon. Wea. Rev.*, **137**, 246–268.

Gneiting, T. and Raftery, A. E. (2005). Weather forecasting with ensemble methods. *Science*, **310**, 248-249.

Greybush, S.J., S.E. Haupt, and G.S. Young, 2008: The Regime Dependence of Optimally Weighted Ensemble Model Consensus Forecasts of Surface Temperature. *Wea. Forecasting*, **23**, 1146–1161.

Grimit, E.P., and C.F. Mass, 2002: Initial Results of a Mesoscale Short-Range Ensemble Forecasting System over the Pacific Northwest. *Wea. Forecasting*, **17**, 192–205.

Kocin P. J., and L. W. Uccellini, 2005: *Northeast Snowstorms*. Vols. 1 and 2, *Meteor. Monogr.,* No. 54, Amer. Meteor. Soc., 818 pp.

Marzban, C., R. Wang, S. Sandgathe, F. Kong, S. Leyton, 2010: On the Effect of Correlations on Rank Histograms: Reliability of Temperature and Wind-speed Forecasts from Fine-scale Ensemble Reforecasts.. *Mon. Wea. Rev,* conditionally accepted.

Miller, J.E., 1946: Cyclogenesis in the Atlantic Coastal Region of the United States.. *J. Atmos. Sci.*, **3**, 31–44.

Raftery, A.E., T. Gneiting, F. Balabdaoui, and M. Polakowski, 2005: Using Bayesian Model Averaging to Calibrate Forecast Ensembles. *Mon. Wea. Rev.*, **133**, 1155–1174.

Ralph, F.M., R.M. Rauber, B.F. Jewett, D.E. Kingsmill, P. Pisano, P. Pugner, R.M. Rasmussen, D.W. Reynolds, T.W. Schlatter, R.E. Stewart, S. Tracton, and J.S. Waldstreicher, 2005: Improving Short-Term (0–48 h) Cool-Season Quantitative Precipitation Forecasting: Recommendations from a USWRP Workshop. *Bull. Amer. Meteor. Soc.*, **86**, 1619–1632.

Roebber, P.J., S.L. Bruening, D.M. Schultz, and J.V. Cortinas, 2003: Improving Snowfall Forecasting by Diagnosing Snow Density. *Wea. Forecasting*, **18**, 264–287.

Roebber, P.J., M.R. Butt, S.J. Reinke, and T.J. Grafenauer, 2007: Real-Time Forecasting of Snowfall Using a Neural Network. *Wea. Forecasting*, **22**, 676–684.

Tracton, M.S., and E. Kalnay, 1993: Operational Ensemble Prediction at the National Meteorological Center: Practical Aspects. *Wea. Forecasting*, **8**, 379–398.

Ware, E.C., D.M. Schultz, H.E. Brooks, P.J. Roebber, and S.L. Bruening, 2006: Improving Snowfall Forecasting by Accounting for the Climatological Variability of Snow Density. *Wea. Forecasting*, **21**, 94–103.

Whitaker, J.S., and A.F. Loughe, 1998: The Relationship between Ensemble Spread and Ensemble Mean Skill. *Mon. Wea. Rev.*, **126**, 3292–3302.