Machine Learning Enhancement of Storm Scale Ensemble Precipitation Forecasts

David John Gagne II
University of Oklahoma
School of Meteorology
120 David L. Boren Blvd.
Norman, Oklahoma 73072
djgagne@ou.edu

Amy McGovern
University of Oklahoma
School of Computer Science
110 West Boyd St.
Norman, Oklahoma 73019
amcgovern@ou.edu

Ming Xue
Center for the Analysis and
Prediction of Storms
120 David L. Boren Blvd.
Norman, Oklahoma 73072
mxue@ou.edu

1. INTRODUCTION

Precipitation intensity forecasting is among the most challenging for meteorologists because of two main sources of complexity. First, the formation of precipitation requires the conversion of atmospheric moisture into water that falls to the ground and sometimes requires upward vertical motion. Multiple mechanisms can produce, enhance, or interfere with the precipitation processes, and those mechanisms operate on many scales. Numerical models explicitly represent grid resolvable scales but have to parameterize smaller-scale processes with physical and statistical relationships. Denser grids mean fewer parameterizations, but some processes are too small to resolve feasibly. These model resolution and parameterization limitations result in precipitation prediction errors. Second, the limited number of observations to initialize the models and error in such observations results in initial condition imperfections that grow as the model is run.

Numerical model ensembles were developed to quantify the range of possible errors stemming from these sources by perturbing the initial conditions and varying the parameterizations of each model. Sorting the contributions of those error sources requires an archive of previous ensemble forecasts and post-processing to discover patterns and translate them into a calibrated probabilistic forecast. Machine learning techniques provide ways to perform both tasks with varying degrees of skill and understanding. This project compares the ability of multiple machine learning techniques to produce skilled probabilistic precipitation forecasts from a high resolution ensemble of numerical model forecasts and to discover patterns in the ensemble that reveal its strengths and weaknesses as well as where the largest influences in the forecast system lie. The application of machine learning techniques to the problem of ensemble post-processing has so far been very limited in the field of numerical weather prediction (NWP).

KDD-KDMS 2011 San Diego, CA USA

2. DATA AND METHODS

The ensemble prediction system used for this study provides a wealth of information but also major calibration challenges. The University of Oklahoma Center for the Analysis and Prediction of Storms (CAPS, http://www.caps.ou.edu/) has developed a Storm Scale Ensemble Forecast (SSEF) system and run experimental forecasts since 2007 as part of the NOAA Hazardous Weather Testbed (HWT) Spring Experiment [9, 3]. The SSEF consists of a multimodel and multiphysics ensemble of 4 km horizontal grid spacing covering the continental US using the two cores of the Weather Research and Forecasting (WRF) model, the Advanced Research WRF (ARW) and Nonhydrostatic Mesoscale Model (NMM), plus the CAPS Advanced Regional Prediction System (ARPS) [7, 8, 10]. That grid spacing is the largest that allows storm formation. This study uses a core of 15 ensemble members from the 2010 SSEF. We chose those 15 because they include radar data assimilation and initialize with initial condition perturbations derived from the National Center for Environmental Prediction operational Short Range Ensemble Forecast (NCEP SREF) system. The ensemble was run from May 3 to June 18, 2010. The main variable of interest is the 3 hour accumulated precipitation total at different forecast ranges.

Verification data come from the the National Severe Storms Laboratory National Mosaic and Multi-Sensor QPE (NMQ) project [6]. The data consist of a radar-derived precipitation estimate over the SSEF domain, with the original data at 1-km resolution. Radar provides high resolution for rainfall measurement and is the only system that can provide the detail to match the resolution of SSEF. Radar-derived estimates have their limitations as well. The sensitivity of rainfall estimates to the precipitation type, holes in radar coverage, and interpolation errors introduced during the mosaic process are potential sources of error. The NMQ project incorporates measures to quality-control the data and to match storms with the best precipitation type and drop size distribution, so it is the best CONUS-wide precipitation estimate available.

Its short operational timespan provides two major challenges. Heavy-precipitation events have the largest impact on lives and property. Model error also tends to be the greatest during those events. The short time period covered by the model only captures a few heavy precipitation events at each location and is biased to the May-June period. Precipitation totals vary widely over short spatial distances. Addressing these sample size issues requires ways to increase sample sizes without compromising the integrity of the dataset. Building a model based on data sampled from the full domain would capture more heavy-precipitation events. Stratified random sampling provides more balanced numbers of light- and zero-precipitation events to contrast with the heavy events. The downside of the full-domain sampling is the model will be less sensitive to local-scale effects. A smaller training domain would better account for local effects but with the tradeoff of having fewer heavy-precipitation events. Different scales of training domains are compared within this project to find the best balance.

The training data for the machine learning techniques include information from output variables that have an influence on precipitation and aggregations of them that reflect variations in the parameterizations and models. Variables included are accumulated precipitation, or the total rainfall over time; precipitable water, or the total water vapor in a column of air; max positive w, or the upward vertical velocity; max negative w, or the highest downward vertical velocity; simulated maximum and composite radar reflectivity, which is what a radar would see if it were observing the model atmosphere; and echo tops, which is the radar echo height. Aggregations of these variables are done by calculating the variables' maximum, mean, and standard deviations for the full ensemble, each model type, microphysics parameterization, planetary boundary layer parameterization, and land surface model. If one of those parameterizations or models significantly varies the predictions, then the machine learning should include their influence. In addition to grid point values, statistics over a radius of influence are also included to detect spatial and temporal errors.

Three machine learning methods are used. Stepwise logistic regressions [2] select significant variables and fit a logistic curve to generate a probabilistic forecast. The variable choices are simple to interpret but do not allow for deeper explorations of the variable relationships. Decision trees [4] produce a visual model that generates probabilities and can be analyzed for significant variables and relationships through tree locations. Random forests [1] are an ensemble of random trees that offer better skill than single trees but are not as easily analyzed. Variable importance, which ranks variables by their effects on forest skill, can show predictive correlations better than other statistical ranking methods because it also accounts for variable interactions [5]. By comparing the machine learning models and rankings, meteorologists can discover what model variables and parameterizations provide the best precipitation forecast.

3. CURRENT AND FUTURE WORK

Initial results have shown that machine learning techniques can improve precipitation prediction at various points within the ensemble domain. Stepwise logistic regression and random forest performed with similar skill at all points and thresholds tested. Decision trees had slightly worse skill but also performed well. Variable importance results show high importance for upward vertical velocity and precipitable water, so the random forests are consistently focusing on physically significant processes. Future work will focus on implementing the full-domain and regional models described in Section 2. Cross-correlation analysis also will be performed. If successful, machine-learning forecasts will be used in the 2012 HWT Spring Experiment.

4. ACKNOWLEDGMENTS

Special thanks goes to Fanyou Kong for assistance with the SSEF dataset. Thanks to the anonymous reviewers for their feedback. This research was funded through the National Science Foundation's Graduate Research Fellowship and grant NSF/IIS/0746816 and and NSF AGS-0802888.

5. REFERENCES

[1] L. Breiman. Random forests. *Mach. Learn.*, 45:5 – 32, 2001.

[2] T. M. Hamill, R. Hagedorn, and J. S. Whitaker. Probabilistic forecast calibration using ECMWF and GFS ensemble reforecasts. Part II: Precipitation. *Mon. Wea. Rev.*, 136:2620–2632, 2008.

[3] F. Kong, M. Xue, K. W. Thomas, Y. Wang, K. A. Brewster, X. Wang, J. Gao, S. J. Weiss, A. J. Clark, J. S. Kain, M. C. Coniglio, and J. Du. CAPS multi-model storm-scale ensemble forecast for the NOAA HWT 2010 spring experiment. In *24th Conf. Wea. Forecasting/20th Conf. Num. Wea. Pred.*, page Paper 457, Seattle, WA, 2011. Amer. Meteor. Soc.

[4] J. R. Quinlan. *C4.5 Programs for Machine Learning.* Morgan Kaufmann Publishers, Inc., Santa Mateo, CA, 1993.

[5] C. Strobl, A.-L. Boulesteix, T. Kneib, T. Augustin, and A. Zeileis. Conditional variable importance for random forests. *BMC Bioinformatics*, 9(1):307, 2008.

[6] S. Vasiloff, K. W. Howard, J. Zhang, D. H. Kitzmiller, M. G. Mullusky, W. F. Krajewski, E. A. Brandes, R. M. Rabin, D. S. Berkowitz, H. E. Brooks, J. A. McGinley, R. J. Kuligowski, and B. G. Brown. Improving QPE and very short term QPF: An initiative for a community-wide integrated approach. *Bull. Amer. Meteor. Soc.*, 88:1899–1911, 2007.

[7] M. Xue, K. K. Droegemeier, and V. Wong. The Advanced Regional Prediction System (ARPS) - A multiscale nonhydrostatic atmospheric simulation and prediction model. Part I: Model dynamics and verification. *Meteor. Atmos. Phys.*, 75:161–193, 2000.

[8] M. Xue, K. K. Droegemeier, V. Wong, A. Shapiro, K. Brewster, F. Carr, D. Weber, Y. Liu, and D. Wang. The Advanced Regional Prediction System (ARPS) - A multiscale nonhydrostatic atmospheric simulation and prediction model. Part II: Model physics and applications. *Meteor. Atmos. Phys.*, 76:134–165, 2001.

[9] M. Xue, F. Kong, K. W. Thomas, Y. Wang, K. Brewster, J. Gao, X. Wang, S. J. Weiss, A. J. Clark, J. S. Kain, M. C. Coniglio, J. Du, T. L. Jensen, and Y. H. Kuo. CAPS realtime storm scale ensemble and high resolution forecasts for the NOAA hazardous weather testbed 2010 spring experiment. In *24th Conf. Wea. Forecasting/20th Conf. Num. Wea. Pred.*, page PAPER 9A.2, Seattle, WA, 2011. Amer. Meteor. Soc.

[10] M. Xue, D. Wang, J. Gao, K. Brewster, and K. K. Droegemeier. The Advanced Regional Prediction System (ARPS), storm-scale numerical weather prediction and data assimilation. *Meteor. Atmos. Phys.*, 82:139–170, 2003.