4.2 THE APPLICATION OF CLIMATE DATA SETS IN CALIBRATING ENSEMBLE GUIDANCE FOR THE PREDICTION OF HAZARDOUS WEATHER

David R. Bright*
NOAA/NCEP/Storm Prediction Center, Norman, OK

Matthew S. Wandishin
University of Arizona and NOAA/National Severe Storms Laboratory, Norman, OK

Steven J. Weiss, Russell S. Schneider, and Joseph T. Schaefer
NOAA/NCEP/Storm Prediction Center, Norman, OK

1. INTRODUCTION

Ensemble forecasts are becoming commonplace in operational meteorology. Yet despite their well documented utility, ensemble systems still suffer from a host of problems resulting from model bias, improper sampling of initial conditions, effects of lateral boundaries, imprecise or misunderstood physical processes, and intrinsic model error. As a result, most ensemble forecasts are under-dispersive and do not encompass a complete range of plausible future atmospheric states (Wandishin et al. 2001).

The post-processing of ensemble output may enhance the statistical attributes of an ensemble system. The goal of ensemble post-processing is to use statistical techniques that improve the scalar attributes of forecast performance such as bias, reliability, resolution, and sharpness (Wilks 1995). Many techniques of varying complexity exist for ensemble post-processing, with the preferred approach dependent upon the forecast problem, operational requirements, and the availability of historical data. Operational requirements generally infer computational efficiency and the capability of efficiently adapting to upgrades in the real-time operational system. Regardless of the post-processing approach used (e.g., regression, artificial intelligence, Bayesian, frequentist) a long-term historical archive of past weather is essential.

2. APPLICATIONS OF CLIMATE DATA

Historical climate data is often used in ensemble forecasting, with an extended climate record becoming particularly important when interested in predicting rare yet high-impact events. Besides direct ensemble post-processing or calibration, some diagnostic techniques use a portion of the climate record to measure predictability or to gauge the impact potential of an event. For example, Grumm and Hart (2001) use the long-term climate record to construct ensemble forecast anomalies as an aid in the prediction of winter storms. A basic normalized variance (i.e., the predicted ensemble variance normalized by the expected climatological variance) can infer ensemble forecast uncertainty since, in a long-term statistical sense, normalized variance approaches two as predictability wanes (e.g., Thompson 1961, pg 148). More recent measures of predictability derived from the ensemble forecast include "forecast confidence" (Durante et al. 2006) and a "relative measure of predictability" (RMOP; Toth et al. 2001). The common thread among these approaches is the inclusion of climate statistics within the ensemble forecast process.

The Storm Prediction Center (SPC) is responsible for specialized weather forecasts of high-impact mesoscale phenomena such as thunderstorms, large hail, damaging wind, tornadoes, excessive precipitation, heavy snow, freezing rain, and dangerous fire weather environments. Although these hazards are often localized in their spatial and temporal extent, they are capable of considerable societal disruption affecting public safety and commerce. The first step toward developing customized ensemble guidance for decision support at the SPC is to understand the physical and dynamical aspects of the phenomenon. Obviously today's operational ensembles do not explicitly resolve most of the aforementioned weather hazards and consequently provide only an

**Corresponding author address*: David Bright, NOAA/SPC, 120 David L Boren Blvd, Norman, OK 73072; E-mail: david.bright@noaa.gov

intimation of their possible occurrence. Understanding the larger scale environments in which these phenomena are normally embedded allows for "downscaling" the ensemble through post processing. Forced mesoscale phenomena such as orographic precipitation or lake effect snow can result in high-resolution probabilistic and non-probabilistic downscaled forecasts. A good example of this is through a reforecast ensemble over the western United States (Hamill et al. 2006). Free-forced mesoscale phenomena not directly linked to a fixed boundary is a more challenging downscaling problem and perhaps best suited toward probabilistic forecasting. Bright and Wandishin (2006) show that with a fundamental consideration of the large-scale environmental parameters most commonly associated with severe convective weather, a skillful probabilistic forecast of severe convection can be produced using the NCEP SREF despite grid lengths of only 32 to 45 km.

Regardless of the specific approach or the phenomena of interest, the existence of a high-quality historical record is critically important to the success of any ensemble post-processing technique. A lengthy historical record is particularly important if the event is rare; often these rare events result in the largest societal impact (e.g., tornadoes). For example, in the case of strong or violent tornadoes ($\geq$ F2 intensity), a historical record on the decadal scale is probably necessary for ensemble post-processing. Unfortunately, problems in the severe record database are well documented and include biased observations, spatial inhomogeneities, and unsteady temporal trends (e.g., Verbout et al. 2006; Weiss et al. 2002).

3. SPC ENSEMBLE CALIBRATION TECHNIQUE

Probabilistic forecasts are the most widely used ensemble product at SPC (Bright et al. 2003; 2004). An ongoing effort at the SPC is to produce calibrated ensemble guidance *specific* to the various program areas. As previously mentioned, any ensemble post-processing technique needs to be computationally efficient and adaptable to ensemble system changes. To meet these demands a frequentist approach designed to correct raw probabilistic forecasts has been adopted. The frequentist approach is simple to implement and maintain; adjusting raw probabilistic forecasts from the ensemble removes the direct relationship to each member. If upgrades and improvements to the ensemble system are gradual then abrupt changes to the raw probabilistic output should not occur. Under this assumption, the post-processing remains functional even as the calibration is adjusting to an upgraded system.

This particular frequentist calibration approach requires two relevant probabilistic predictors from the ensemble. The two raw ensemble probabilistic predictors are listed as Pr[*parameter 1* $\geq$ X] and Pr[*parameter 2* $\geq$ Y. At every grid point the Pr[*parameter 1* $\geq$ X] and Pr[*parameter 2* $\geq$ Y] are each partitioned separately into one of eleven bins based on their forecast probability: 0-5%; 5-15%; … ; 85-95%; 95-100%. Then, the two probabilistic forecasts are considered jointly by placing the joint forecast into one of 121 unique bins: (0%,0%); (0%,10%); (0%,20%); … (100%,100%). This is done at each grid point over the entire calibration domain. Historical reports (e.g., lightning strike data) are gridded onto the same ensemble grid with hits and misses tallied over the appropriate forecast period to estimate the actual occurrence frequency. The length of time required to estimate the matching occurrence frequency grids depends on the phenomena of interest and varies from weeks for common events (e.g., warm season lightning) to months or years for rare events (e.g., severe weather reports). The computational efficiency of the calibration technique allows for the occurrence frequency grids to be updated weekly. The sophistication of the calibration can be adapted to optimize results. For example, temporal weighting is sometimes applied for seasonal considerations and spatial weighting inversely proportional to the distance (1/r) of the grid point from the event allows for regionalization in data rich areas and extrapolation to data sparse areas. Separate adjustment tables are normally produced for each ensemble forecast cycle and for each forecast hour.

4. OPERATIONAL EXAMPLES IN HAZARDOUS WEATHER PREDICTION

4.1 Thunderstorm Guidance

Bright et al. (2005) provides a comprehensive description of the methodology used to produce the SPC ensemble-based calibrated thunderstorm guidance. Verification statistics are also provided. A 3h and 12h calibrated prediction is shown in Figs. 1 and 2, respectively, and their accompanying reliability in Figs. 3 and 4, respectively. This particular forecast is based on the probability the SPC cloud physics thunder parameter exceeds one and the probability total precipitation exceeds 0.25 mm (see Bright et al. 2005 for details). Historical National Lightning Detection Network (NLDN) cloud-to-ground lightning strike data serve as truth in the calibration process. During the summer months when convection is plentiful only about 30 days of NLDN data are required for a useful calibration; the cool season requires a longer training period particularly in thunderstorm sparse regions. The system is currently set to use the previous 366 days for thunderstorm calibration and is updated weekly.

4.2 Severe Thunderstorm Guidance

Bright and Wandishin (2006) describes the SPC ensemble-based severe thunderstorm calibration technique and includes verification statistics. The calibration is largely based on the methodology outlined in section 3 with some modification owing the complexity of predicting severe thunderstorms. These modifications are described in Bright and Wandishin (2006) and not repeated here. An example of a 3h and 24h prediction is shown in Figs. 5 and 6, respectively. The accompanying reliability diagram for the contiguous United States shows the 24h forecast reliability ending at 12 UTC (Fig. 7). The calibration of severe thunderstorm guidance requires an extensive archive of severe weather reports. The calibration period is presently set to use all severe weather events archived in real-time at the SPC over the preceding year, although a longer period of record may be desirable to capture very infrequent events or discriminate probabilistically between hail, wind, or tornadic environments.

4.3 Snow accumulation on roads

An experimental probabilistic forecast based on the likelihood of new winter precipitation accumulating on road surfaces is currently under development. Inputs to the prediction are probabilistic forecasts of precipitation type, precipitation amount, and parameters sensitive to model predicted surface temperature and net radiative fluxes at the surface. The specific parameters are at various levels of maturity and consequently are not discussed further herein. The calibration technique is a frequentist approach similar to the methodology used for SPC ensemble-based severe thunderstorm guidance (Bright and Wandishin 2006) except raw probabilistic ensemble forecasts sensitive to the parameter space around expected surface conditions are used as predictors. Truth is provided from the MADIS road state variable available at http://www.madis-fsl.org/; winter data for 2004-2005 were used in the calibration process. An example of a 3h probability of new winter precipitation accumulating on road surfaces is shown in Fig. 8. The system was run in real-time for the entire 2005-2006 winter with the resulting reliability shown as Fig. 9. Presently the system remains under development and will run in real-time for the 2006-2007 winter with *experimental* results available online at https://www.spc.noaa.gov/exper/sref/. The only update for the current winter season is the inclusion of an additional year (2005-2006) of calibration data.

5. SUMMARY

To develop real-time calibrated ensemble guidance for high-impact operational weather support, it is important to maintain accurate and unbiased archives of historical events. As ensembles continue to evolve into operational decision support tools, the availability of historical event archives will become increasingly important. Using historical event data and relevant environmental predictors deduced from the ensemble, straightforward and/or creative calibration techniques can be used to produce reliable probabilistic ensemble guidance.

6. REFERENCES

Bright, D.R. and M.S. Wandishin, 2006: Post processed short range ensemble forecasts of severe convective storms. *Preprints*, 18th Conf. on Probability and Statistics, Atlanta, GA, Amer. Meteor. Soc., CD-ROM (5.5).

Bright, D.R., M.S. Wandishin, R.E. Jewell, and S.J. Weiss, 2005: A physically based parameter for lightning prediction and its calibration in ensemble forecasts. *Preprints*, Conf. on Meteor. Applications of Lightning Data, San Diego, CA, Amer. Meteor. Soc., CD-ROM (4.3).

Bright, D.R., M.S. Wandishin, S.J. Weiss, J.J. Levit, J.S. Kain, and D.J. Stensrud, 2004: Evaluation of short-range ensemble forcasts in predicting severe convective weather during the 2003 SPC/NSSL Spring Program. *Preprints*, 22nd Conf. Severe Local Storms, Hyannis, MA, Amer. Meteor. Soc., P15.5.

Bright, D.R., S.J. Weiss, J. Levit, and D.J. Stensrud, 2003: The utility of short-range ensemble forecasts in the real-time prediction of severe convective weather at the Storm Prediction Center. *Preprints*, 10th Conf. on Mesoscale Processes, Portland, OR, Amer. Meteor. Soc.

Durante, A.V., R.E. Hart, A. Watson, R.H. Grumm, and W. Drag, 2006: The development of forecast confidence measures using NCEP ensembles and their real-time implementation within NWS web-based graphical forecasts. *Preprints*, 18th Conf. on Probability and Statistics, Atlanta, GA, Amer. Meteor. Soc., CD-ROM (5.1).

Grumm, R. and R. Hart, 2001: Standardized anomalies applied to significant cold season weather events: Preliminary findings. *Wea. Forecasting*, **16**, 736-754.

Hamill, T.M., J.S. Whitaker, and S.L. Mullen, 2006: Reforecasts: An important dataset for improving weather predictions. *Bull. Amer. Meteor. Soc.*, **87**, 1-33.

Thompson, P., 1961: *Numerical Weather Analysis and Prediction*. The Macmillan Company, New York.

Toth, Z., Y. Zhu, and T. Marchok, 2001: The use of ensembles to identify forecasts with small and large uncertainty. *Wea. Forecasting*, **16**, 463-477.

Verbout, S.M., H.E. Brooks, L.M. Leslie, and D.M. Schultz, 2006: Evolution of the U.S. tornado database: 1954-2003. *Wea. Forecasting*, **21**, 86-93.

Wandishin, M.S., S.L. Mullen, D.J. Stensrud, and H.E. Brooks, 2001: Evaluation of a short-range multimodel ensemble system. *Mon. Wea. Rev*., **129**, 729-747.

Weiss, S.J., J.A. Hart, and P.R. Janish, 2002: An examination of severe thunderstorm wind report climatology: 1970-1999. *Preprints*, 21st Conf. Severe Local Storms, San Antonio, Amer. Meteor. Soc., 446-449.

Wilks, D.S., 1995: *Statistical methods in Atmospheric Sciences*. Academic Press, New York, 467 pp.

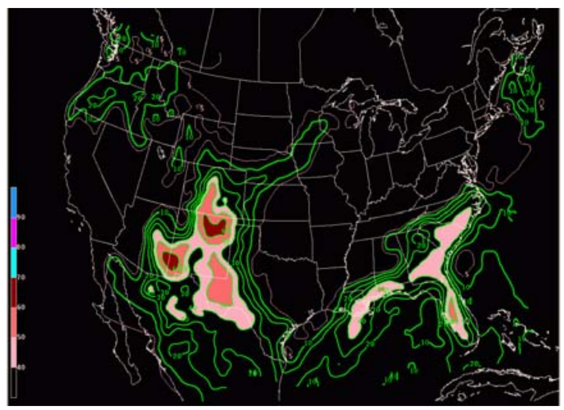

Figure 1. Probabilistic calibrated thunderstorm guidance from the NCEP short-range ensemble forecast ending at 00 UTC 1 September 2004. Forecast shown is a 15h forecast valid between 21 UTC and 00 UTC.

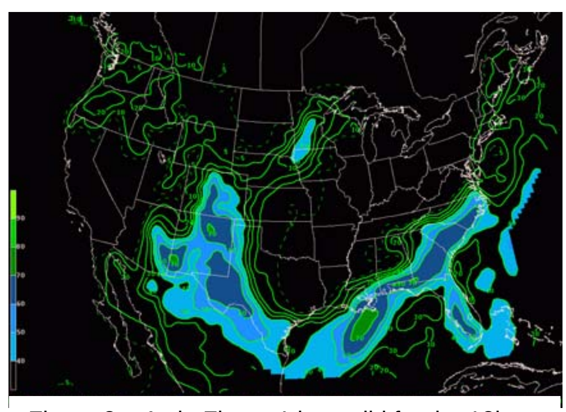

Figure 2. As in Figure 1 but valid for the 12h period from 12 UTC 31 August to 00 UTC 1 September 2004.

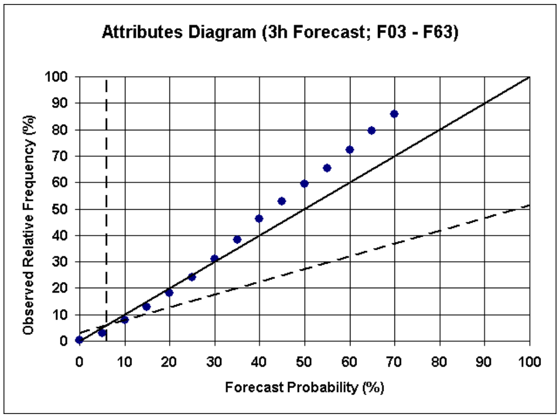

Figure 3. Reliability diagram of calibrated 3h thunderstorm guidance for all forecast hours (F03 through F63) for the period 15 April 2005 to 15 October 2005. Solid diagonal line indicates perfect reliability and dashed lines indicate lower limit of skillful forecast.

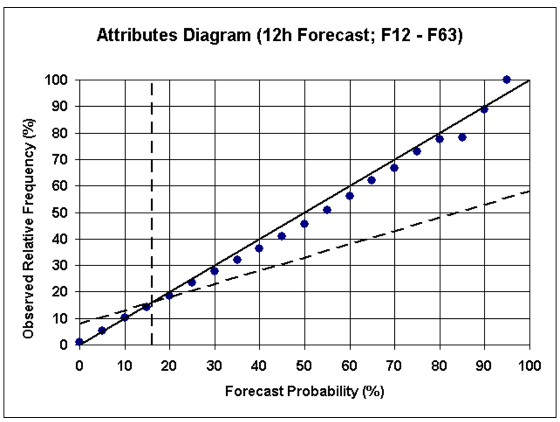

Figure 4. As in Figure 3 except for all 12h forecasts (valid between F12 and F63).

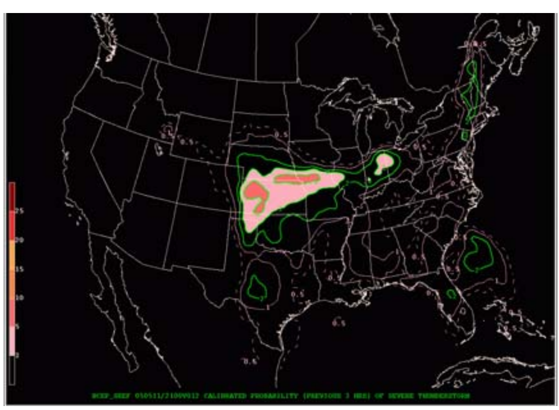

Figure 5. Probabilistic calibrated severe thunderstorm guidance from the NCEP short-range ensemble forecast ending at 21 UTC 11 May 2005. Forecast shown is a 3h forecast valid between 18 UTC and 21 UTC.

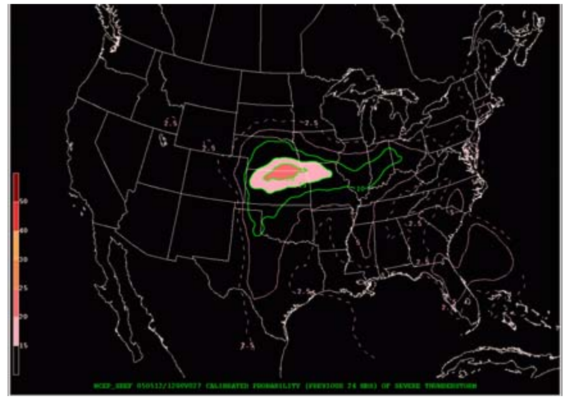

Figure 6. As in Figure 5 but valid for the 24h period from 12 UTC 11 May to 12 UTC 12 May 2005.

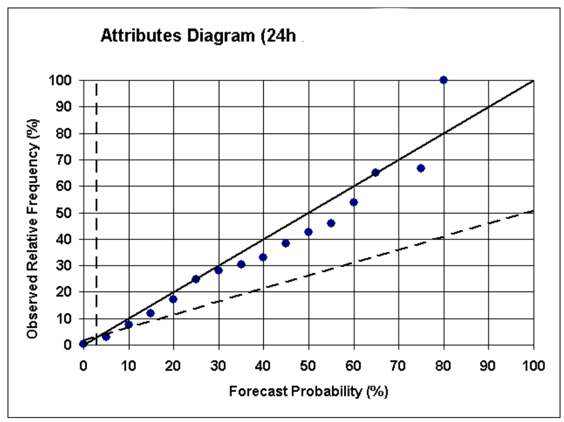

Figure 7. Reliability diagram of calibrated 24h severe thunderstorm guidance ending at forecast hour 39 from the 21 UTC short-range ensemble from 15 April through 15 October 2005. (Solid and dashed lines as in Fig. 3.)

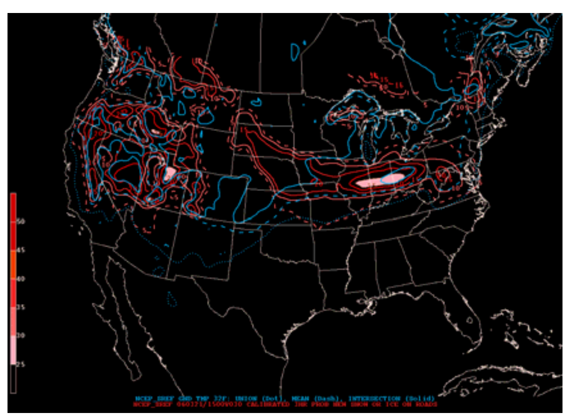

Figure 8. Probabilistic calibrated snow accumulating on road surface guidance from the NCEP short-range ensemble ending at 15 UTC 21 March 2006. Forecast is a 30h forecast valid between 12 UTC and 15 UTC.

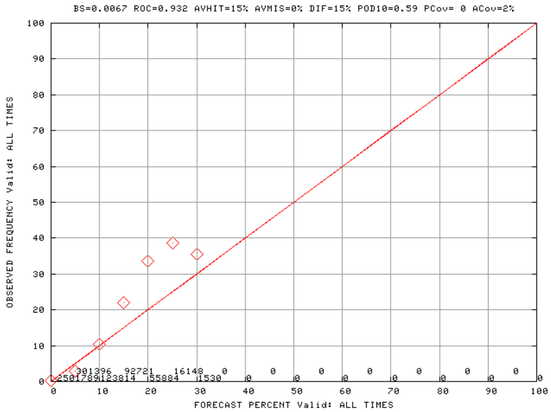

Figure 9. Reliability of calibrated snow accumulating on road surface guidance from the NCEP short-range ensemble forecast. Includes all 3h forecasts (F03 through F63) from 1 October 2005 through 30 April 2006.