6B.2 POSTPROCESSING OF PRECIPITATION FORECASTS FOR NEW CONFIGURED NCEP SHORT-RANGE ENSEMBLE FORECASTING (SREF) SYSTEM

Huiling Yuan [1, *], Jun Du [2], John McGinley [1], Paul Schultz [1]
, Binbin zhou [2], Chungu Lu [1, 3], Zoltan Toth [2], and Geoff Dimego [2]
1. NOAA/Earth System Research Laboratory (ESRL)/Global Systems Division (GSD)
2. EMC, NOAA/NCEP and SAIC
3. Cooperative Institute for Research in the Atmosphere (CIRA), Colorado State University

1. INTRODUCTION

Since December 2005, the NCEP (National Centers for Environmental Prediction) short-range ensemble forecasting (SREF) system has been updated by constructing new ensemble components (Du et al. 2006). This system contains perturbed initial/boundary conditions, multiple physics, and multimodel. It covers the North American continent and the adjacent maritime zones. It is important to evaluate the performance of precipitation forecasts of the SREF, especially probabilistic quantitative precipitation forecasts (PQPF). Results of precipitation forecasting from this suite of new configured operational SREF system were verified compared to NCEP Stage IV precipitation analyses over the continental United States (CONUS). The reliability curves of PQPF presented light wet biases. It is of interest to conduct calibration for the SREF PQPF and examine how much the postprocessing can benefit an ensemble system with diversified models and physics.

One tool used to conduct probabilistic postprocessing is a feedforward artificial neural network (Hsu et al. 1995), which was successfully applied to correct PQPF biases in a high-resolution (12 km) Regional Spectral Model (RSM, Juang and Kanamitsu 1994) ensemble system (Yuan et al. 2005, 2007a). Another tool is a linear regression model, which was applied to calibrate PQPF for heavy precipitation events over the American River Basin during the Hydrometeorolgical Testbed (HMT) program at NOAA/ESRL/GSD (Yuan et al. 2007b). Both tools were implemented to calibrate PQPF from the SREF system over the CONUS. According to hydrological and geographic characteristics, bias correction of PQPF was performed for three major regions – western, central, and eastern U.S. (Fig. 1.). The two seasons are defined as the warm season (April-September) and cool season (October-March), respectively. In this study, preliminary results are present for the two periods: (1) April-September 2006, (2) October-December 2006.

Cross-validation was used to compute verification scores and attributes diagrams. The following results are composited from all months during a season in a region.

* Corresponding author address: Huiling Yuan, NOAA/ESRL, R/GSD7, 325 Broadway, Boulder, CO, 80305-3328. email: huiling.yuan@noaa.gov

2. DATA

Current NCEP SREF system (Du et al. 20006) includes 21 members from ETA, RSM models, and Weather Research and Forecast (WRF) models with two dynamic cores: the Advanced Research WRF (ARW) and Nonhydrostatic Mesoscale Model (NMM). The system has various physical schemes to increase the diversity of ensemble construction. Each model uses the breeding perturbations to create ensemble members. The model output is over 212 AWIPS grid domain (~ 40 km, 185 x 129 grids), covering the CONUS. All model runs are initialized at 0900 and 2100 UTC with 3-h interval output. The observation data come from the NCEP 4-km Stage IV precipitation analyses, which combine gauge and radar measurements over the CONUS with a quality control procedure. The 24-h precipitation accumulations from the SREF system with lead times of 27, 51, and 75 hours (i.e., 1, 2, 3 days) are compared to the Stage IV data, which are aggregated to the 212 AWIPS 40-km grids.

3. METHOD

A three-layered feedforward neural network (Hsu et al. 1995) includes the input, hidden (neural), and output layers. The neural network is non-linear with a sigmoid function in the neural layer. The optimization scheme in this neural network uses a linear least square

simplex (LLSSIM) algorithm to obtain best possible weights for fitting the input and output datasets under selected criteria and convergence constrains. Through the LLSSIM algorithm, the global and near global optimization can be achieved.

The output layer node in the neural network is only one node. The output data, i.e. the calibrated PQPF on every grid pixel, is between 0 and 1. The target observations (expected values) are the dichotomous data, in which the value is 1 when the observation exceeds a selected precipitation threshold, otherwise 0. The objective function is the root mean square error (RMSE) between the calibrated PQPF and target observations. Therefore, the optimization process is equivalent to minimize Brier score, which measures the deviation between the forecasted and observed probabilities. The nodes in the hidden layer are 4 and can be increased by convergence criteria. The input variables (nodes) in the input layer are probabilities (PQPF). On a grid pixel, using the 21 ensemble members from the SREF system, probabilities can be computed for a series of selected exceeding precipitation thresholds: 0.01, 0.1, 0.25, 0.5, 0.75, 1.0, 1.25, 1.5, 2.0, 2.5, 3.0, 3.5, 4.0 inches/24-h (i.e., 0.254, 2.54, 6.35, 12.7, 19.05, 25.4, 31.75, 38.1, 50.8, 63.5, 76.2, 88.9, 101.6 mm/24-h). For a given threshold, the input variables are seven sequential probabilities with the thresholds centered at (the closest) the given threshold in this probability series. For example, the first seven probabilities are used as the input for selected thresholds from 0.01 to 0.5 inches; at 1-inch threshold, probabilities for 0.25, 0.5, 0.75, 1.0, 1.25, 1.5, 2.0 inches/24-h are used as input datasets.

After training of known input and target observations on all grid pixels in a region during a season, the weights in the neural network are obtained. For the validation period, the probabilities for validation data and the weights are used to output calibrated PQPF for a selected threshold. All training and validation are conducted for a single selected threshold. Most probabilities meet the monotonic distribution, i.e., the probabilities at lower thresholds are not less than ones at higher thresholds. A program is conducted to check the monotonic distribution of probabilities and correct the higher probabilities at high thresholds to the probabilities at an adjacent lower threshold.

An alternative technique is the multiple linear regression method. The linear regression equation is simple:

$$P(\mathrm{x},t) = a + \sum_{i=1}^{M} b_i f_i(\mathrm{x},t) \qquad (3.1)$$

where M=7, $f_i(x,t)$ is the seven ordered input probabilities as the input variables in the neural network, $P(x,t)$ is the observed probability, and a is a constant (error residual). The coefficients b_i and a are estimated by minimizing the errors of the target observed probabilities and the linear regression probabilities using N training data samples, *i.e.*, the least square function:

$$\min \sum_{j=1}^{N} [P(\mathrm{x},t) - (a + \sum_{i=1}^{M} b_i f_i(\mathrm{x},t))]^2 \qquad (3.2).$$

Unlike the neural network, the output values of (3.1) can fall outside of the range of [0 1]. For negative values (or > 1), new probabilities are set to 0 (or 1).

In each region – western, central, and east U.S. (Fig. 1), the validation data is rotated for one month of data during a season, and the rest months of data are used as the training dataset. This is one way to increase sample size for obtaining stable weights in both the neural network and linear regression equation. In the training dataset, dry points having probability less than 0.1 are excluded for increasing the efficiency of the training in the neural network. This exclusion does not alter the weights much. All results are composited from monthly validation during the warm season (6 months) and cool season (3 months).

4. RESULTS

For increasing the training sample size, all grid pixels with concomitant observation and PQPF are used to derive the weights in the linear regression or neural network over a region. The cross-validation results using the linear regression calibration are shown in Fig. 2-5. Figure 2 shows the attributes diagrams for the calibrated PQPF over the western region during the warm season for four selected thresholds (0.1, 0.25, 0.5, and 1 inch). The 1:1 diagonal line indicates the perfect forecast as the forecast probabilities equal to the observed occurrence frequencies at each probability category. The forecast probabilities from 0% to 100% with 1/21 interval are grouped to the nearest 10% interval. The original PQPF from the SREF system shows good forecasts with light wet biases as

the reliability curves are below the diagonal line, especially for high probabilities categories. The calibration shows improved PQPF for thresholds 0.1, 0.25, 0.5 inches, but with less confidence at 0.5 inches for a longer lead time of 3 days. For higher thresholds (e.g., 1 inch), while the reliability curves largely vary and lack high probabilities for the lead time of 3 days, the calibration only improves the lower probability categories but lack reliability for higher probability categories. This is due to sample size issue over the relatively dry western region during the warm season, while rare heavy precipitation events occurred during the study period.

The attributes diagrams over the central region (Fig. 3) show better calibration results than over the western region. For higher thresholds and a longer lead time, the calibrated PQPF lacks higher probability categories. Again, sample size is one important issue. Also, there are more uncertainties for precipitation forecasts with longer lead times. The best forecasted and calibrated region is the eastern region (Fig. 5). During the warm season, the eastern region possesses more heavy precipitation events. The overcorrection of the calibrated PQPF, which shows dry biases at higher probability categories, occurs for 1-inch threshold with the lead time of 3 days.

The cool season (October-December 2006) has similar calibration results over each region (not shown) using the linear regression method. Improved PQPF is shown in composite reliability curves over the CONUS by combing three regions (Fig. 5) for all thresholds, except for 1-inch with a 3-day lead time. The internal bars indicate the frequencies of different probability categories. After calibration, the conditional biases are greatly reduced and the frequencies are changed for the calibrated PQPF. The reduction of 0% and 100% probability categories in the calibrated PQPF leads to increasing grid pixels with middle-range probability categories.

The calibration using the linear regression and neural network shows slight different results, but with more differences for the cool season's calibration, which has only three months in total and uses two months as the training data. Figure 6 indicates that the calibration during the cool season using the linear regression method is better than using the neural network, in terms of overfitting and unstable reliability curves in the calibration using the neural network at higher thresholds.

Due to a coarse resolution (~ 40 km) and small sample size during a season, the calibration was conducted over a large region, which covers several big watersheds (Fig. 1) and has varied climatology over the grid pixels. This also limits the applications of the neural network to obtain good relationship between the input and output data. The calibration needs longer period of data in the training over a smaller region at current resolution. Another speculation is that the calibration of PQPF favors the results using the linear regression because the observation and the input variables are more linear relationship.

5. SUMMARY

Calibration of PQPF for the newly configured NCEP SREF system was conducted over three major geographic regions based on a linear regression model and a three-layered feedforward neural network. With seven probabilities as input variables and a few months of data as the training, both methods showed similar results in removing conditional biases in PQPF for both cool and warm seasons. The neural network is able to derive the non-linear relationship between the input and output data. However, for shorter data and few samples with rare events, a simple linear regression method showed better results than the neural network, since the artificial neural network had overfitting problem due to sample issues. This study used the offline weights and cross-validation. Both methods can be implemented in a real-time operational postprocessing system. The weights in the linear regression calibration are more easily to be updated and adaptive. On the other hand, the biases in original PQPF could be more linear pattern with the selected input variables. More data from the NCEP SREF with the same configuration are desirable to examine the calibration procedures.

In the future, as more forecast data or retrospective forecasts can be available, multivariate (e.g., humidity, wind speeds, and temperature) calibration is feasible by using more variables associated with PQPF as the input in the training dataset. Using multivariate calibration, the optimal weights in the neural network are expected to improve PQPF and outperform the simple linear regression method.

With the development of the physical models, a fast portable postprocessing procedure is needed with less dependency of the training data.

6. ACKNOWLEDGEMENTS

This research was performed while the first author held a National Research Council Research Associateship Award at NOAA/ESRL. Special thanks to Mr. Jeff McQueen at NCEP for his helps on this project.

7. REFERENCES

Du, J., J. McQueen, G. DiMego, Z. Toth, D. Jovic, B. Zhou, and H. Chuang, 2006: New Dimension of NCEP Short-Range Ensemble Forecasting (SREF) System: Inclusion of WRF Members, Preprint, WMO Expert Team Meeting on Ensemble Prediction System, Exeter, UK, Feb. 6-10, 2006, 5 pp. [Available online at http://wwwt.emc.ncep.noaa.gov/mmb/SREF/reference.html.]

Hsu K, H. V. Gupta, and S. Sorooshian, Artificial neural network modeling of the rainfall-runoff process, Water Resources Research, 31(10), 2517-2530, 1995.

Jolliffe, I. T., and D. B. Stephenson, 2003: *Forecast Verification. A Practitioner's Guide in Atmospheric Science*. Wiley and Sons Ltd.

Juang, H. H., and M. Kanamitsu, 1994: The NMC nested regional spectral model. *Mon. Wea. Rev.*, **122**, 3–26.

Yuan, H., S. L. Mullen, X. Gao, and S. Sorooshian, 2005: calibration of probabilistic quantitative precipitation forecasts from the RSM ensemble forecasts over hydrologic regions. Preprints, *85th AMS Annual Meeting*, San Diego, CA, Amer. Meteor. Soc. CD-ROM, J3.4.

Yuan, H., X. Gao, S. L. Mullen, S. Sorooshian, J. Du, and H. H Juang, 2007a: Calibration of probabilistic quantitative precipitation forecasts with an artificial neural network. *Wea. Forecasting* (in press).

Yuan, H., J. A. McGinley, P. J. Schultz, C. J. Anderson, and C. Lu, 2007b: Evaluation and calibration of short-range PQPFs from time-phased and multimodel ensembles during the HMT-West-2006 campaign. *J. Hydrometeor.*, in review.

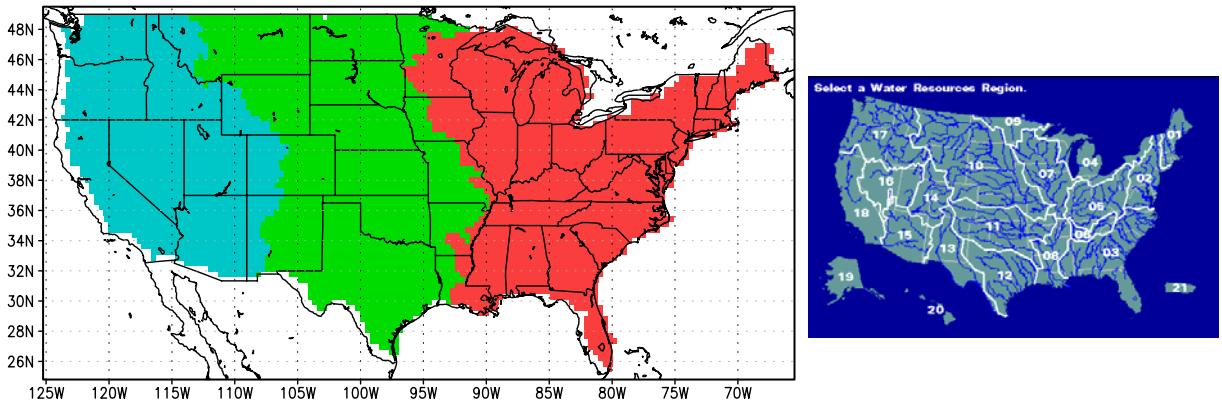

Fig. 1. The western, central, and eastern regions (left) in the continental United States, by combining adjacent watersheds (right) defined by the US Geological Survey (USGS).

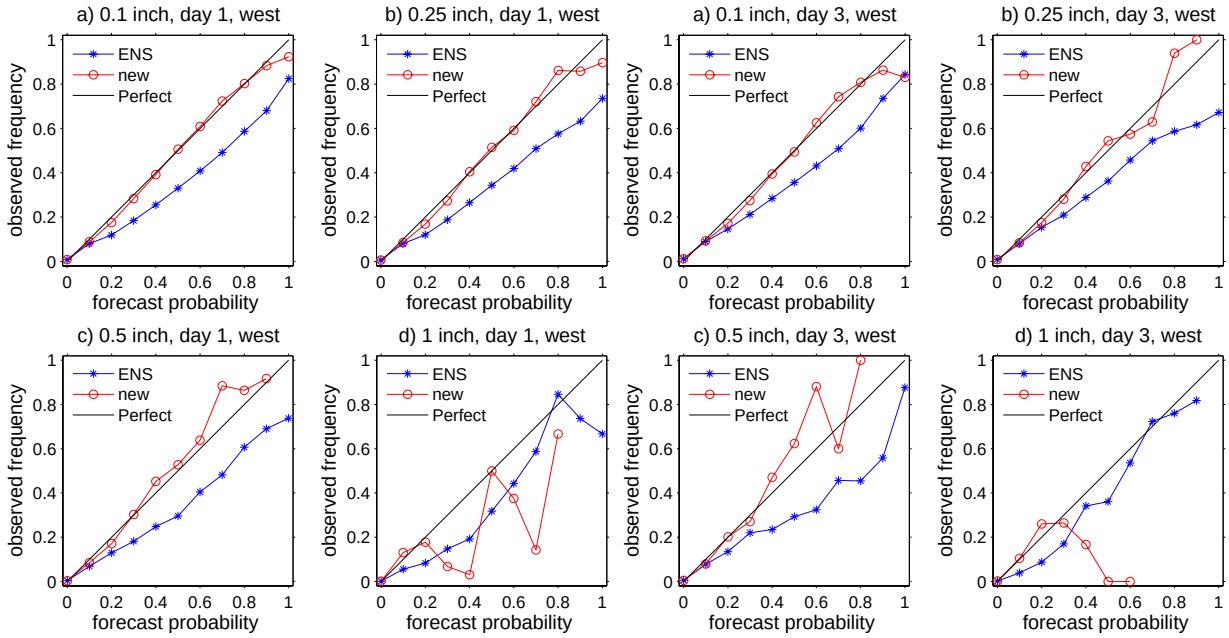

Fig. 2a. (left four frames) The 24-h PQPF calibration over the western region for the warm season (April-September 2006) with the lead time of 1 day. The blue lines are raw PQPF, and the red ones are after calibration. The diagonal line indicates the perfect forecast.

Fig. 2b. (right four frames), same as Fig. 2a but for the lead time of 3 days.

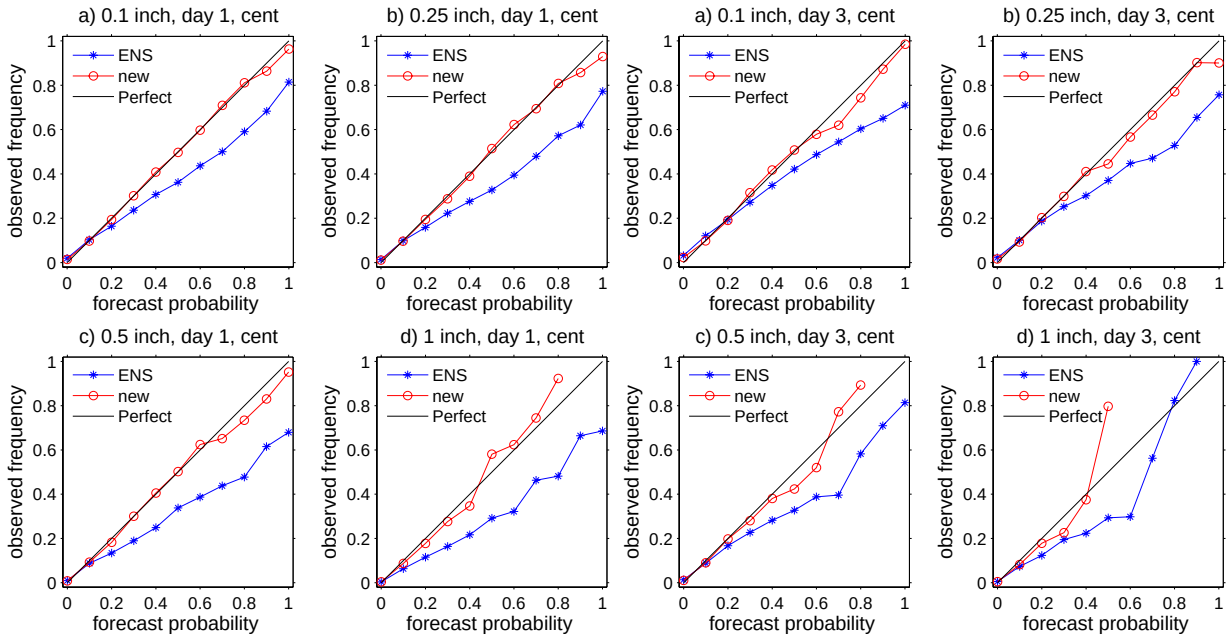

Fig. 3a. (left four frames) The 24-h PQPF calibration over the central region for the warm season (April-September 2006) with the lead time of 1 day. The blue lines are raw PQPF, and the red ones are after calibration.

Fig. 3b. (right four frames), same as Fig. 3a but for the lead time of 3 days.

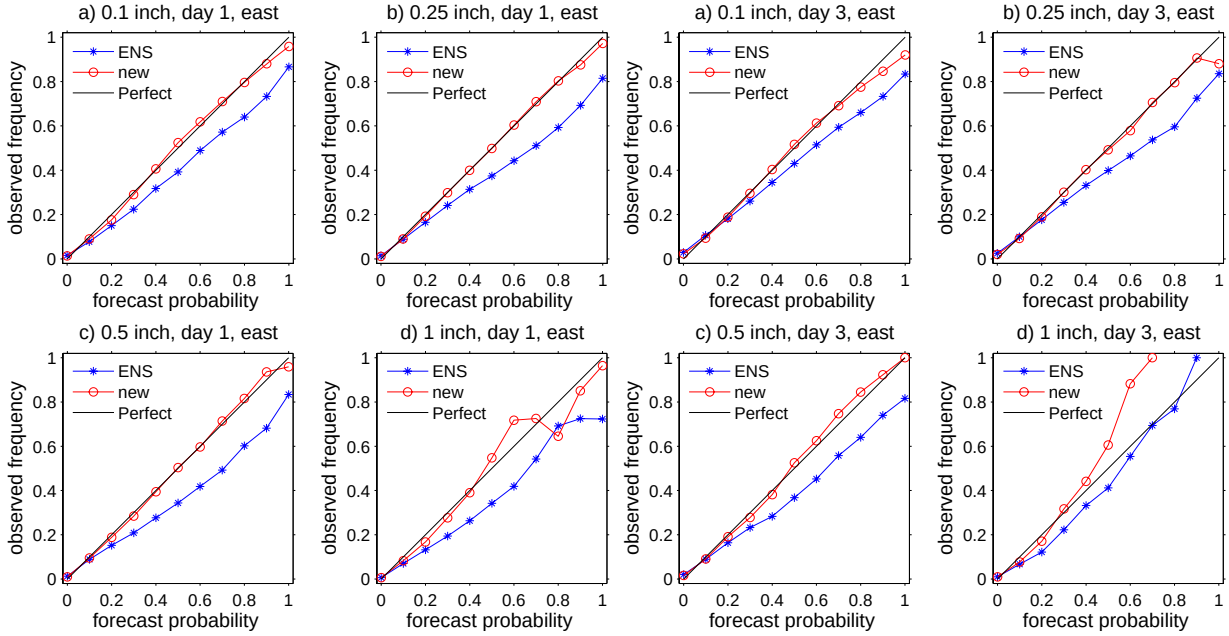

Fig. 4a. (left four frames) The 24-h PQPF calibration over the eastern region for the warm season (April-September 2006) with the lead time of 1 day. The blue lines are raw PQPF, and the red ones are after calibration.

Fig. 4b. (right four frames), same as Fig. 4a but for the lead time of 3 days.

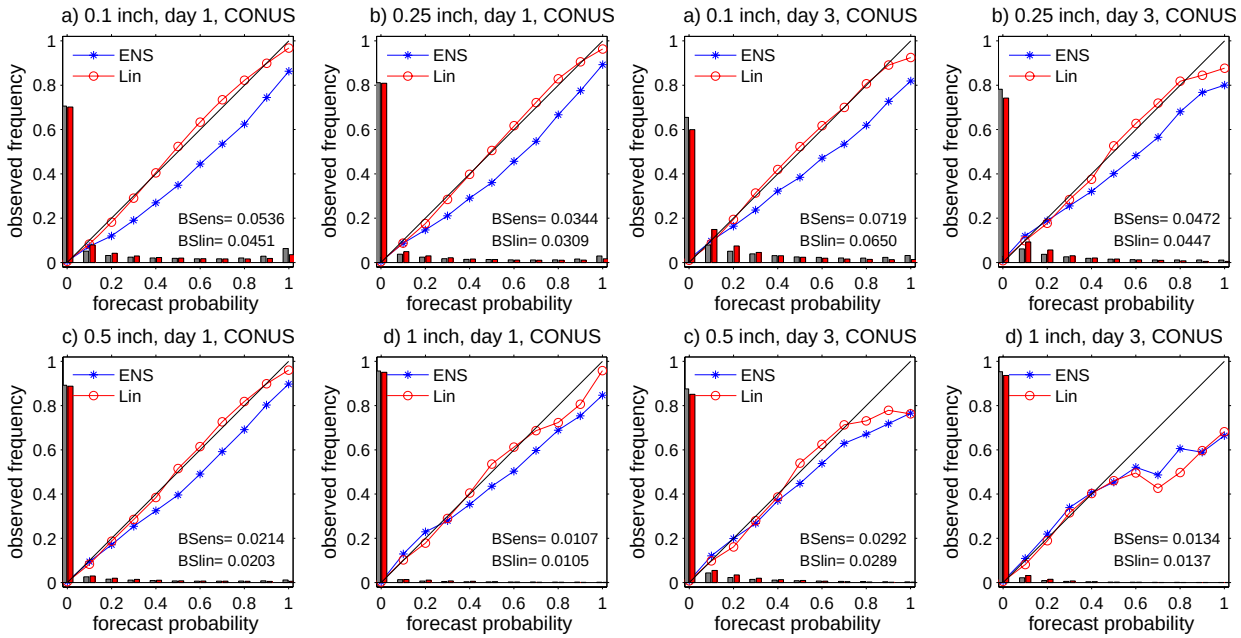

Fig. 5a. (left four frames) The 24-h PQPF calibration over the CONUS for the cool season (October - December 2006) with the lead time of 1 day. The blue lines are raw PQPF, and the red ones are after calibration. The gray bars are frequencies of raw PQPF, and the red bars are for PQPF after calibration.

Fig. 5b. (right four frames), same as Fig. 5a but for the lead time of 3 days.

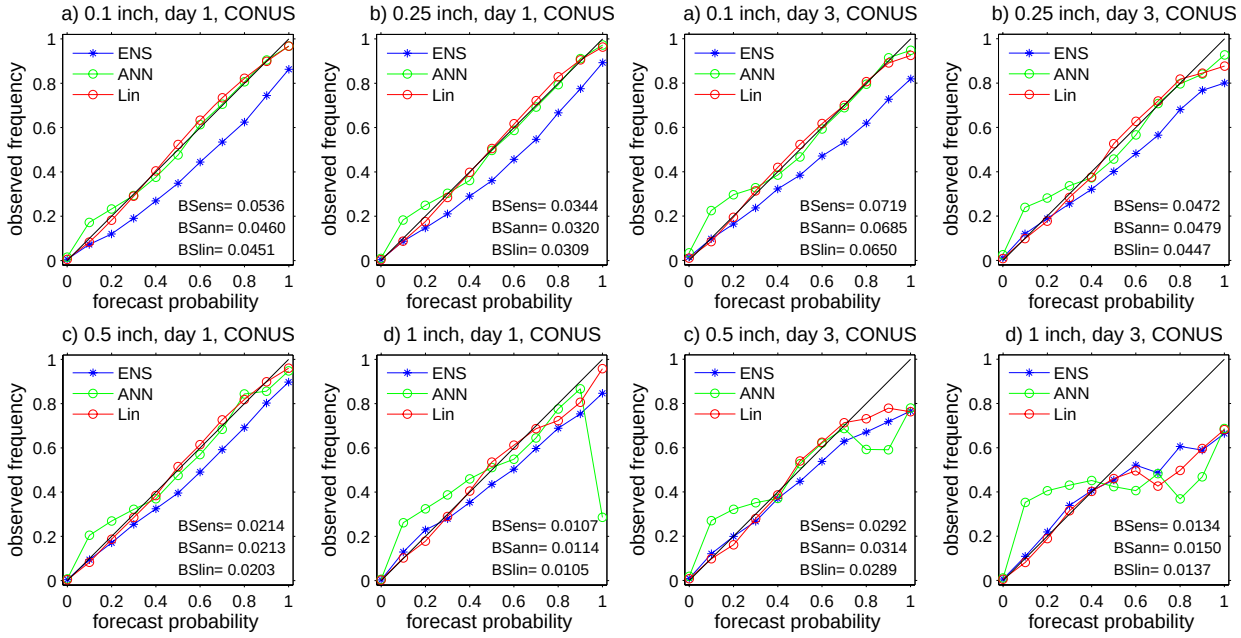

Fig. 6a. (left four frames) The 24-h PQPF calibration over the CONUS for the cool season (October - December 2006) with the lead time of 1 day. The blue lines are raw PQPF, the red ones using linear regression calibration, and the green using the neural network.

Fig. 6b. (right four frames), same as Fig. 6a but for the lead time of 3 days.