J3.4 CALIBRATION OF PROBABILISTIC QUANTITATIVE PRECIPITATION FORECASTS FROM THE RSM ENSEMBLE FORECASTS OVER HYDROLOGIC REGIONS

Huiling Yuan [1]*, Steven L. Mullen [2], Xiaogang Gao [1], and Soroosh Sorooshian [1]

1. INTRODUCTION

Cool season precipitation plays an important role in freshwater supply over the Southwest United States, which is marked by heterogeneous topographic and hydrologic scenarios. More accurate precipitation prediction is highly desirable for both the public and hydrological model users. Numerous studies indicate that ensemble forecasting provides more skillful weather forecasts than a single deterministic forecast run. The National Centers Environmental Prediction (NCEP) Regional Spectral Model (RSM, Juang and Kanamitsu 1994) ensemble system was performed to forecast daily precipitation during winter 2002-2003 over the southwest US (Yuan et al. 2004). Probabilistic quantitative precipitation forecasts (PQPF) from 11 ensemble members are good at discriminating precipitation events in terms of high relative characteristic curve (Wilks 1995) areas. The forecast skill presents large spatial variation, with the highest skills over the California region. However, significant wet biases in the RSM forecasts result in some low statistical scores and unskillful forecast indices, in particular, over the Colorado Basin River Forecast Center (CBRFC) and the Great Basin Region. It is indispensable to calibrate such biases to increase the accuracy of PQPF and quantify the real atmospheric uncertainties in weather forecasts. An artificial neural network is applied to conduct this post-processing over four US Geological Survey (USGS) hydrologic Unit Regions: the Upper Colorado Region, the Lower Colorado Region, the Great Basin Region, and the California Region (see Fig. 1, Yuan et al. 2004).

2. DATA

Currently, the NCEP operates short-range ensemble forecasts using the NCEP RSM over the continental US at the spacing grid of 32 km. In this study, the NCEP RSM produced accumulated precipitation amounts for one control run and five pairs of perturbed ensemble members at an equivalent grid of 12 km. Daily precipitation is available for a total of 151 days from Nov 1 2002 to Mar 31 2003. Only forecasts from 0000 UTC are analyzed in this paper. The verifying data of 24-h accumulated precipitation comes from the NCEP stage IV precipitation analyses on a 4 km national grid.

[1]* Department of Civil and Environmental Engineering, University of California, Irvine, CA, 92697, email: huilingy@uci.edu
[2] Department of Atmospheric Sciences, University of Arizona, Tucson, AZ, 85721

Averaged monthly precipitation and Brier skill scores (BSS, Wilks 1995) show that PQPF in February and March performed worse than the previous three months (Yuan et al. 2004). Because only one cool season forecasts were implemented at the finer resolution, the historical data is short. The first 90 days are selected as the training data, while the rest of the 61 days are as the forecasted and verification periods.

3. METHOD

A 3-layered feed forward neural network (Hsu et al. 1995) is chosen to process the precipitation calibration. In this neural network, a linear least square simplex (LLSSIM) algorithm, is used to search the optimal non-linear relationship between input and output datasets. LLSSIM is an effective way to obtain global and near global optimization. Three layers include an input layer, a hidden layer, and an output layer.

Since PQPF are the focus of this paper, the calibration of precipitation probability at a certain threshold is emphasized. The total number of input nodes is 18, including 7 probabilities and daily precipitation of 11 ensemble members. For each category of 15 thresholds (0.25, 1, 5, 10, 15, 20, 25, 30, 40, 50, 60, 70, 80, 90, and 100 mm), precipitation probability is calculated using the number of members greater than the threshold of 11 ensemble members at each model grid. The closest 7 probabilities are selected. For example, at 15-mm threshold, probabilities for 1, 5, 10, 15, 20, 25, and 30 mm are used as input datasets. Daily precipitation of 11 members is sorted and normalized by the range [0 1]. The output data is the probability between 0 and 1. The verifying data is dichotomous observed probability. Observed event (greater than the threshold) refers to “1”, otherwise, “0”. The objective function is the root mean square error (RMSE). As the error function evaluation reaches the criteria or maximum training times, the training process ends and gains a series of weighting coefficients. The hidden node starts from 2 and can be increased according to convergence criteria and evaluation time. Usually, RMSE changes very small after 5000 times of iterations.

Over each hydrologic region, training datasets for the first 90 days are trained at a selected threshold. The output datasets for the last 61 days are verified compared to the observed probability. To reduce large dry points and increase the efficiency of training, the datasets having probability less than 0.1 have not been trained and calibrated. The NCEP stage IV precipitation is firstly averaged on a 12-km model grid.

4. RESULTS

Calibrated PQPF shows remarkably improved BSS (Fig. 1) over four hydrologic basins. Except for 25-mm threshold, BSS over the Colorado basin and the Great Basin region become positive. After calibration through the neural network, BSS at 50-mm threshold turns into skillful values over the California region.

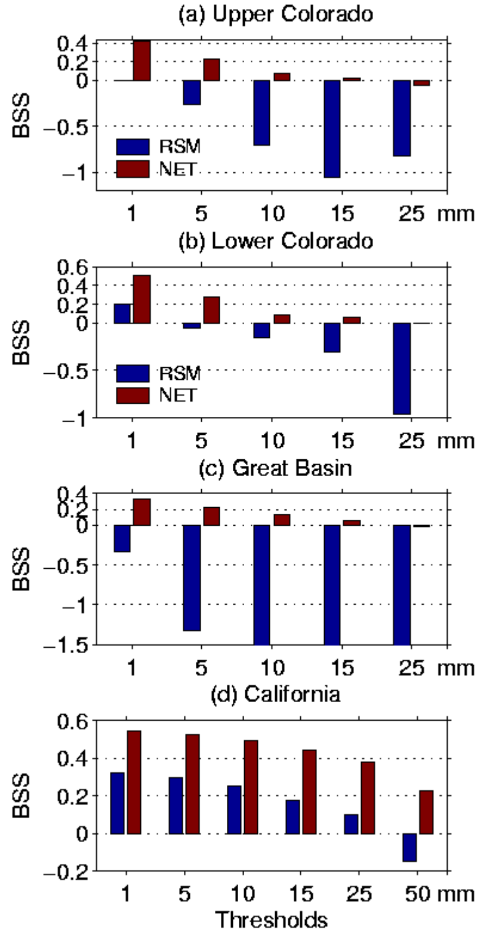

Fig. 1 BSS over four hydrologic regions

BSS can be decomposed into three terms: reliability, resolution, and uncertainty (Wilks 1995, Jolliffe and Stephenson 2003). Whenever the resolution term exceeds the reliability term, BSS is skillful. The original PQPF present the most skillful forecasts over the California region, while most forecasts are generally unskillful over the other three regions. Fig. 2 shows that all reliability terms decrease (better) over four hydrologic regions throughout all thresholds. Over the California region, the resolution term is almost unchanged. Therefore, the improvement of BSS mainly results from the reduction of the reliability term, i.e. the correction of conditional biases. Compared to modeled PQPF, the calibrated resolution term, however, decreases (worse) at higher thresholds over the other three regions and the uncertainty term, depending on the sample climatological frequencies, is much less than it over the California region. The low frequencies of precipitation events lead to fewer sample sizes for training in the neural network. Thus, more historical data is desirable to construct training data and increase the accuracy of calibration for rare cases.

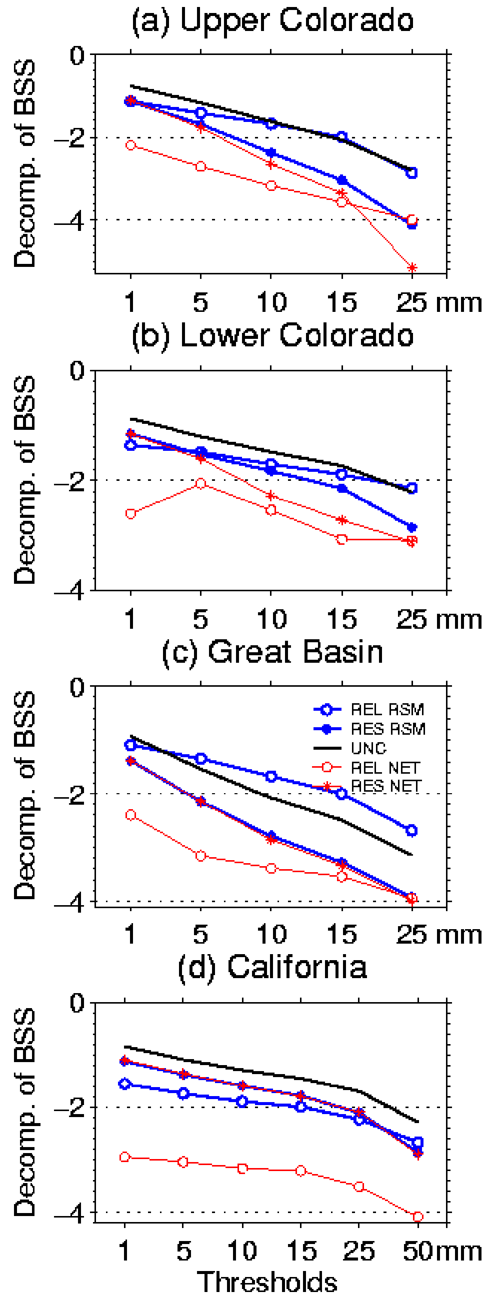

Fig. 2 Decomposition of BSS

The attributes diagrams (Wilks 1995, Jolliffe and Stephenson 2003; Fig. 3.1-3.4) further reflect the correction of model biases. A consistent overestimation (reliability curve bellows the perfect 1:1 diagonal line) in the RSM forecasts appears over four hydrologic regions. Reliability curves on attributes diagrams indicate the decreased reliability term after calibration over four hydrologic regions. However, a severe underestimate is produced over the lower Colorado basin at higher thresholds. In addition, high probabilities have not been produced by the neural network over the Colorado basin and the Great basin. Less sharpness is perhaps caused by lower frequency of precipitation events

over the region. The calibrated PQPF over the California region (Fig. 3.4) reduces overestimation and keeps the

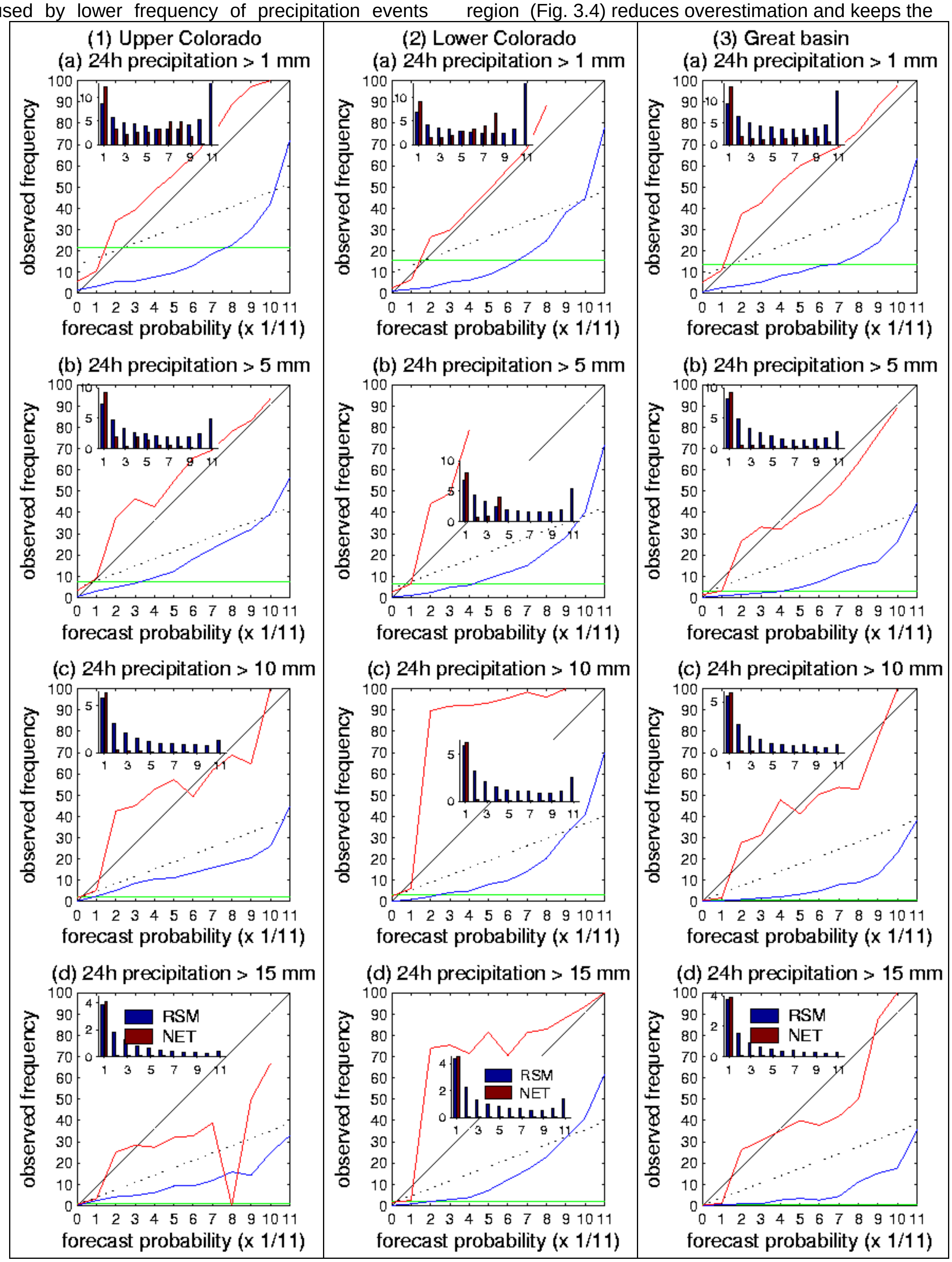

Fig. 3.1-3.3 Attributes Diagrams over the Upper (left column) and the Lower Colorado (middle column), and the Great basin (right column). The internal bars indicate the frequency of different probabilities. The red ones plotted from the calibrated PQPF, while the blue ones from original PQPF. The horizontal green line is climatological frequency, and the slashed blue line is un-skill line. Four thresholds: 1, 5, 10, and 15 mm.

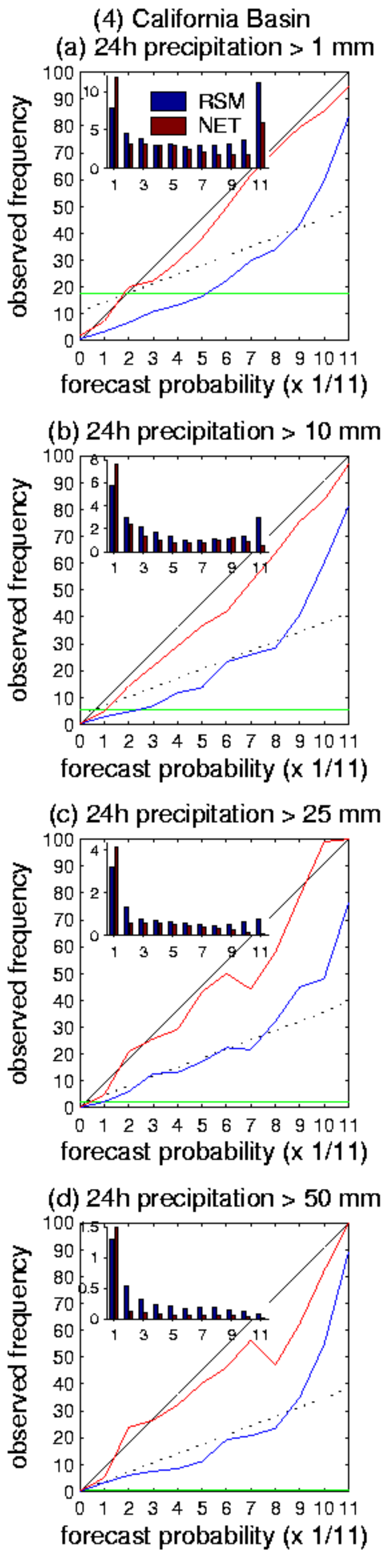

Fig. 3.4 same as Fig. 2.1-2.3, but for California region at four thresholds (1, 10, 25, 50 mm).

sharpness as well. Overall, over each hydrologic region, the reliability curve approaches to the diagonal line, which indicates that the conditional biases are removed to some degrees. Internal bars indicate that post-processing changes the original frequency distribution, and therefore, calibration removes the conditional biases. In general, the frequencies of lower probabilities increase (0% category not shown) and frequencies at high probabilities decrease. Over the first three regions, some frequencies of high probabilities become zero, which is caused by over correction.

5. SUMMARY

Based on a 3-layered feed forward neural network, PQPF from the RSM ensemble system at a grid spacing of 12 km were calibrated over four hydrologic regions of the southwest US. The results indicate that the neural network can correct conditional wet biases in original ensemble forecasts. The post-processing through the neural network assists to increase the forecast skill, such as BSS. Due to limited datasets in a cool season, the training size is not enough to provide information for calibration of rare cases. The resolution term has not been reduced over the California region, while the removal of conditional biases harm the resolution term at higher thresholds over the Colorado basin. More seasons of data are needed to increase the training size.

In this study, only PQPF are used as input datasets. Other weather conditions, such as humidity, wind speeds, need be considered in the future. In addition, the climate zone needs to be classified according to similar climate regions other than using four large hydrologic regions. The elevation and location needs to count into additional factors in calibration. Moreover, in order to apply PQPF to hydrologic models, accurate precipitation at a finer temporal interval is needed. Calibration of 6-hour accumulated PQPF is requisite for current operational hydrologic models to conduct general flood forecasting.

6. ACKNOWLEDGEMENTS

The authors acknowledge the support of NASA EOS-IDS Grant NAG5-3460 and NSF STC program (Agreement EAR-9876800). The authors wish to thank Dr. Jun Du and Dr. Hann-Ming H. Juang at NCEP for assisting the RSM forecasts and thank Dr. Kuo-lin Hsu for providing the neural network code.

7. REFERENCES

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, S. Sorooshian, J. Du, and H. H Juang, 2004: Verification of probabilistic quantitative precipitation forecasts over the southwest United States during winter 2002-2003 by the RSM ensemble system. *Monthly Weather Review* (in press).