Calibration of Probabilistic Quantitative Precipitation Forecasts from the NCEP RSM ensemble system

*[1]Yuan, Huiling, [2]Mullen, S.L., [1]Gao, X., [1]Sorooshian, S.
[1]Center for Hydrometeorology and Remote Sensing, Civil and Environmental Engineering, UC Irvine
[2]Atmospheric Sciences, University of Arizona *email: huilingy@uci.edu

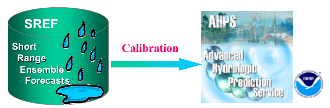
Introduction

Short-range ensemble forecasting of precipitation was evaluated over the Southwest US during winter 2002-2003.

The 12-km RSM ensemble system has a large wet bias.

To use probabilistic quantitative precipitation forecasts (PQPFs) to hydrological models, calibration is necessary.

An artificial neural network is used to calibrate the RSM PQPFs over four USGS Hydrologic Unit Regions.



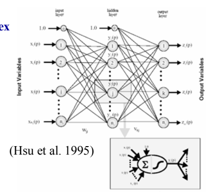
Experiment Design

A 3-layer feedforward neural network

Linear least square simplex (LLSSIM) algorithm:

Search for an optimal non-linear relationship between input and output datasets

Obtain global and near global optimization

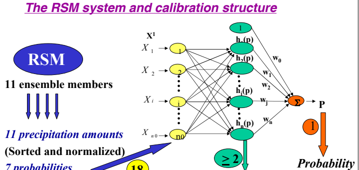


(Hsu et al. 1995)

Dataset

- **Model forecasts:**
  11 ensemble members by the NCEP Regional Spectral Model (RSM, Juang and Kanamitsu 1994) system
  Nov. 2002 – Mar. 2003 (151 days) at 0000 UTC
  Training dataset: **the first 90 days**
  Forecast period: **the last 61 days**
- **Verification:**
  Observed event: occur 1, non-occur: 0 (for a given threshold)
  NCEP Stage IV 24-h precipitation analyses (4 km)
  12 km RSM forecasts bilinearly interpolated to Stage IV grids
- **Input: 18 (11 precipitation amounts and 7 probabilities)**
  Output: 1 probability

The RSM system and calibration structure



Prob. = (# member > threshold)/11

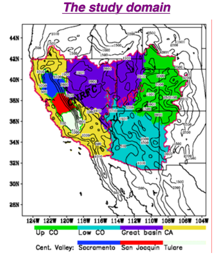
Probabilities at 15 thresholds (e.g. **20 mm**)
(0.25, 1, **5, 10, 15, 20, 25, 30, 40**, 50, 60, 70, 80, 90, and 100 mm)

Input probability at 7 selected thresholds

Application Procedure

- ***Iteration:***
  Up to 5000 times with little change in coefficients
  For efficiency, train and calibrate data with probability > 0.1
- ***Objective function:***
  Root mean square error (RMSE) of Brier Score (BS)
- ***Verification measures:***
  Comparison between the calibrated PQPFs and RSM PQPFs during the forecast period (61 days)
  Brier Skill Score (BSS)
  BS = Reliability term – Resolution term + Uncertainty term
  Attributes Diagrams

The study domain



2 River Forecast Centers:
CNRFC
CBRFC

4 USGS hydrologic regions:
California
Great Basin
Upper Colorado
Lower Colorado

Central Valley in California:
Sacramento
San Joaquin
Tulare Basin

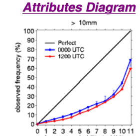
Forecast Bias

On Stage IV 4 km grids, the 24-h RSM PQPFs are verified over the Southwest US for 151 days.

Overestimation according to reliability curves in the attributes diagram
(similar trend for other thresholds)

The 1200 UTC forecasts show a stronger wet bias than the 0000 UTC ones.

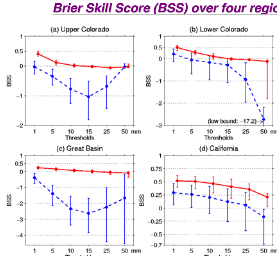
Attributes Diagram



(Error bars indicate 90% confidence levels of the 0000 UTC forecasts)

Improved Brier Skill

Brier Skill Score (BSS) over four regions



RSM
NET

Error bars:

90% confidence levels (CIs)

BSS greatly improves between 1-50 mm thresholds. Considering error bounds, uncertainty of BSS reduces after calibration and skillful forecasts remarkably increase (BSS>0).

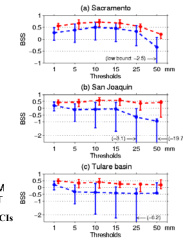
Brier Skill Score (BSS) over the Central Valley

Some variability over different smaller watersheds

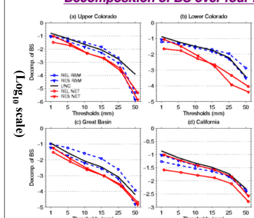
Considering confidence levels, using a localized training dataset gives better BBS than using dataset from the whole California basin, except for 50-mm threshold, owing to few rain events over the smaller catchments.

RSM
NET

Error bars: 90% CIs



Decomposition of BS over four regions



$$BSS = \frac{Resolution - Reliability}{Uncertainty}$$

When *Resolution > Reliability, BSS > 0*

"Uncertainty" depends on sample climatology **only**, the larger values the higher sample climatology frequencies *p* (while *p* < 0.5).

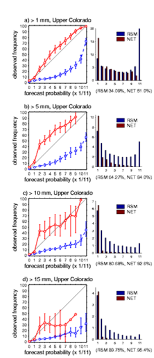
Improvement of BSS is due to the reduction of the reliability term.

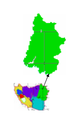
Over the Upper and Lower Colorado regions, and the Great Basin region, the resolution term decreases at higher thresholds. Bias mitigates at the expense of discrimination.

Over the California, bias reduces without harming the resolution term except for 50-mm. This region possesses much higher climatology frequencies at higher thresholds than the other three regions.

Improved Attributes Diagrams

Attributes Diagrams over the Upper Colorado region





For 1, 5, 10, and 15 mm thresholds:

The left graphs show reliability curves. Error bars indicate 90% CIs.

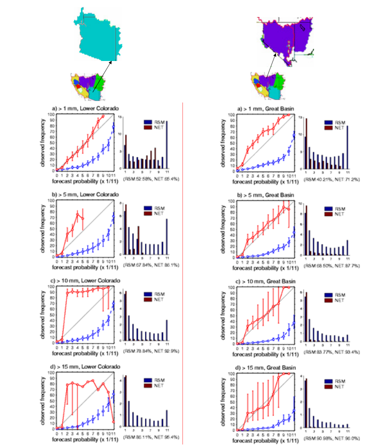
The right histograms show that **Frequency distribution changes** (0% category is shown in the bottom bracket).

Reliability improves; curves lie closer to 1:1 perfect diagonal line.

Lacking high probabilities for higher thresholds and large uncertainties.

High probabilities for higher thresholds show very little error bounds since some resampling tests (randomly 10,000 times) in calculating CIs lack high probabilities.

Attributes Diagrams over the Lower Colorado and the Great Basin

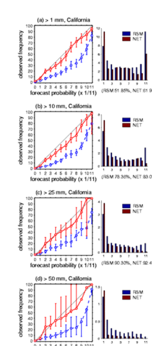


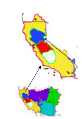
Calibration over the Lower Colorado region causes underestimation at higher thresholds and lacks high probabilities.

The two Colorado regions and the Great Basin region have less heavier rainfall events, however, **overfitting problems** are very severe over the Lower Colorado region for 10-mm and higher thresholds with more unreliable results. **Except for the sample size of rainfall events, the regional variation plays an important role in using neural networks.**

Over the Great Basin region, the calibrated PQPFs exhibit little underestimation for 1-mm threshold and large uncertainties with lacking high probabilities for higher thresholds, which are similar to calibration over the Lower Colorado region.

Attributes Diagrams over the California region



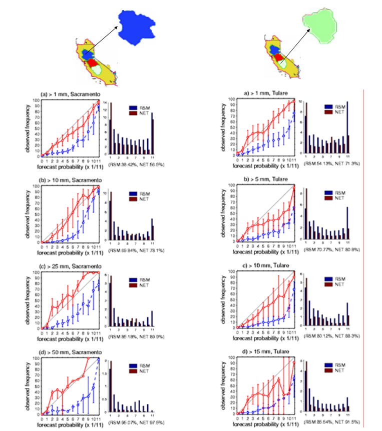


For 1, 10, 25, and 50 mm thresholds:

The best calibrated reliability curves are over the California region of the four hydrologic regions. Reliability curves improve for high thresholds 25- and 50-mm.

Considering error bounds, bias greatly reduces. Reliability curves present smaller uncertainties than ones for the other three regions.

Attributes Diagrams over the Sacramento and the Tulare Basin



Over the Sacramento Basin, reliability curves improve for 1-50 mm thresholds as the California region.

Over the Tulare Basin, large uncertainties exhibit for 10-mm and higher thresholds. The San Joaquin Basin shows the similar trend but smaller error bounds. Both basins are much drier than the Sacramento Basin, especially the Tulare Basin.

Conclusions

A 3-layer feedforward neural network is able to mitigate conditional biases and increase the Brier skill in the RSM daily forecasts.

While the reliability term is significantly reduced, improvement of the resolution term, which is related to the discrimination of events, poses challenges.

Best results are over the California region. Calibration should also consider the climate zone and terrain elevation.

More historical data is desirable for training datasets.

Multivariate calibration of PQPFs and other weather elements related to runoff is likely needed to improve forecasts of general flooding from hydrological models.

References

1. Hsu K, H. V. Gupta, and S. Sorooshian 1995: Artificial neural network modeling of the rainfall-runoff process, *Water Resour. Res.*, **31**, 2517-2530.
2. Jolliffe, I. T., and D. B. Stephenson, 2003: *Forecast Verification. A Practitioner's Guide in Atmospheric Science*. Wiley and Sons Ltd.
3. Juang, H. H., and M. Kanamitsu, 1994: The NMC nested regional spectral model. *Mon. Wea. Rev.*, **122**, 3–26.
4. 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. *Mon. Wea. Rev.*, **133**, 279-294.

Acknowledgements

The study was supported by **NASA EOS-IDS** Grant NAG5-3460 and **NSF STC** program (Agreement EAR-9876800). The ONR supported computer resources. The NCEP has assisted the RSM ensemble system. We thank Dr. Henry Juang, Dr. Jun Du, Dr. Kuo-lin Hsu, Dr. Zoltan Toth, Dr. Yuejian Zhu, and others at NCEP.

UCI