Debiasing Technique for Numerical Weather Prediction using Artificial Neural Network

Kang, Boosik

Assistant prof., Dept. of Civil and Envir. Engrg, Dankook University, Seoul, Korea

Ko, Ick Hwan

Director, Water Resources and Environmental Research Center, K-water, Daejon, Korea

Abstract: Biases embedded in numerical weather precipitation forecasts by the RDAPS model was determined, quantified and corrected. The ultimate objective is to eventually enhance the reliability of reservoir operation by Korean Water Resources Corporation (KOWACO), which is based on precipitation-driven forecasts of stream flow. Statistical post-processing, so called MOS (Model Output Statistics) was applied to RDAPS to improve their performance.

The Artificial Neural Nwetwork (ANN) model was applied for 4 cases of 'Probability of Precipitation (PoP) for wet and dry season' and 'Quantitative Precipitation Forecasts (QPF) for wet and dry season'. The reduction on the large systematic bias was especially remarkable. The performance of both networks may be improved by retraining, probably every month. In addition, it is expected that performance of the networks will improve once atmospheric profile data are incorporated in the analysis. The key to the optimal performance of ANN is to have a large data set relevant to the predictand variable. The more complex the process to be modeled by the ANN, the larger the data set needs to be.

1 INTRODUCTION

Predictions from large scale atmospheric weather models have biases because of uncertainty in the initial conditions, imperfect knowledge of the relevant processes leading to parameterizations of process behavior, inaccuracies of the numerical method of solution of the model equations (*e.g.*, spectral model or grid-based model), computational scale, etc. These biases give rise to incorrect or inaccurate precipitation forecasts. Therefore, the main objective of this project is to determine, quantify and correct the biases embedded in numerical weather precipitation forecasts by the RDAPS model. The ultimate objective is to eventually enhance the reliability of reservoir operation by Korean Water Resources Corporation (KOWACO), which is based on precipitation-driven forecasts of stream flow. Resolution of numerical weather prediction (NWP) models, such as RDAPS, is constrained by computing resources. Downscaling of NWP model output to the desired local scale can be done by relating it to the observations at specific locations of interest (Kang and Ramírez, 2002; Kuligowski and Barros, 1998).

2 DATA

Data provided by Korean Water Resources Corporation covers the period between July of 2001 and June of 2002. The data set is composed of forecasted and observed variables (Table 1). Forecasts were computed by RDAPS numerical weather prediction model on a 30 x 30 Km grid. RDAPS produced 48-hour forecasts twice a day, at 00 UTC and at 12 UTC. These forecasts are officially provided to natural disaster agencies, water resource management agencies, and universities in South Korea by the Korea Meteorological Administration (KMA). Predicted values are available 8 times a day as shown in Figure 1. Surface meteorological observations were recorded hourly at 452 unmanned stations with Automatic Weather System (AWS). Figure 2 illustrates the distribution of the weather stations in South Korea together with the RDAPS output grid location.

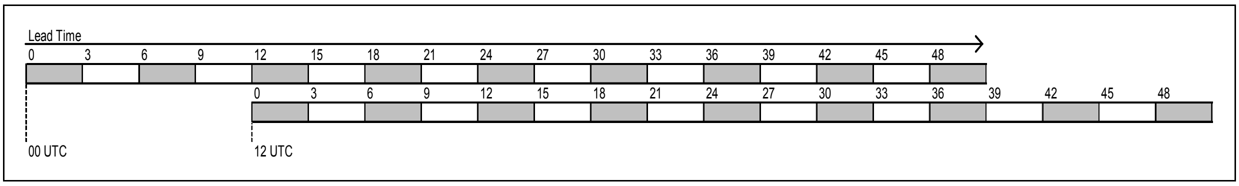

Figure 1. RDAPS NWP model is updated every 12 hours and forecasts variables in lead time increments of 3 hours.

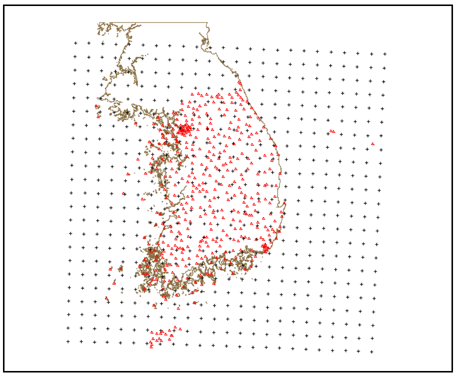

Figure 2. Map of the Korean Peninsula with RDAPS grid (+) and Observation stations (Δ).

3 BIAS CORRECTION FOR RDAPS PRECIPITATION OUTPUT

An ANN comprises interconnected simple processing units that work in parallel, much as do the neuron networks of the brain, and which can extract patterns from noisy input data (Silverman and Dracup, 2000). The neural networks developed for this study have a *feedforward* structure: signals flow forward from input neuron through any hidden units, eventually reaching the output neurons. Neurons are arranged in layers. The input layer simply introduces the values of the input variables. The hidden and output layer neurons are each connected to all of the units in the preceding layer.

One of the most popular network architecture is perhaps the *multilayer perceptron*. The equation for the so-called feedforward, supervised, multi-layered-perceptron can be written as:

$$y = g\left(\sum_{i=1}^{H} \omega_i f\left(\sum_{j=1}^{J} \omega_{ij} x_j - \theta_j\right) - \omega\right) \quad (8)$$

Where:

- H is the number of hidden units
- f and g are the activation functions
- J is the number of inputs
- ω_{lj} are the weights
- θ_j are the thresholds
- x_j is the input data (independent variables)
- and y is the output

Selection of input variables is a critical part of neural network design. Each input variable adds a dimension to the data space. During training, a neural network attempts to fit a response surface to this data. The number of points needed to do this grows very rapidly with the dimensionality (curse of dimensionality). Principal components analysis is a popular approach to dimensionality reduction. However, for this study only a few variables were available as inputs and therefore dimensionality was not a concern.

To correct the bias in mean aerial precipitation, two neural networks need to be developed[1] for each RDAPS grid box, and for each season (rain, dry)[2] due to the seasonal characteristic of rainfall in South Korea[3]. The first network will predict the probability of precipitation (PoP). This PoP network will be applied to the RDAPS output to forecast wet or dry 3-hour periods. Secondly, a quantitative precipitation forecast (QPF) network will be applied to the RDAPS output corresponding to rainy periods predicted by the probability of precipitation NN (PoP>Threshold). The resulting output will forecast precipitation amount.

The PoP (QPF) neural network selects its independent variables (*i.e.*, its inputs) from a pool of girded values produced by RDAPS, plus the precipitation occurrence (amount) observed in the 6-hour and in the 12-hour periods previous to the RDAPS run. The dependent variable, or neural network's output, is determined by the occurrence (amount) of interpolated precipitation corresponding to the RDAPS' grid box for which the NN is being developed.

4 QUANTITATIVE PRECIPITATION FORECAST NEURAL NETWORK FOR RAINY SEASON

A Multi-Layer Perceptron NN with a performance of 0.80 was selected to predict amount of precipitation. Specifically, the network had 3 layers, with 13 units in the first layer, 13 in the second, and 1 unit in the last layer (Figure 3).

[1] Neural networks for this project were developed using the following software: StatSoft, Inc. (2005). STATISTICA version 7.1. Use and mention of a commercial product does not constitute endorsement by Colorado State University.

[2] Rainy season was composed by data from the months of June, July and August. The dry season data sets contained information from the rest of the months.

[3] To handle precipitation seasonality with only one neural network, all precipitation values must be standardized with climatological means and standard deviations, which are not available for this study.

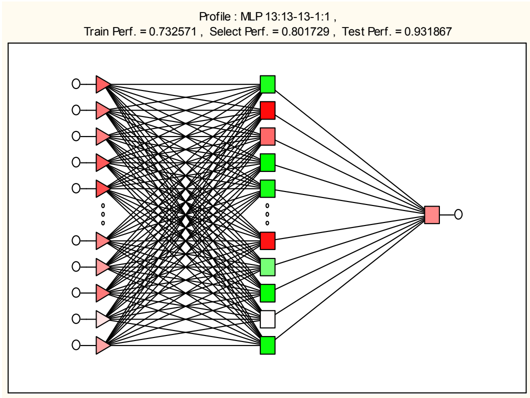

Figure 3. Three-layer perceptron QPF artificial neural network. Units are colored according to their activation levels.

Input data used to train the quantitative precipitation forecast (QPF) neural network are all available RDAPS output variables, plus the output of the PoP NN. Since the observed amount of precipitation was not normally distributed, the logarithm of the observations was used as the desired network output[4].

Sensitivity analysis was conducted on six QPF networks with similar performance. The single most important variable was the geopotential height of 1000 hPa. Mean sea level pressure, surface wind speed, and geopotential height of 750 hPa were also selected as significant variables. Figure 4 shows results of the sensitivity analysis conducted for the QPF NN that performed best.

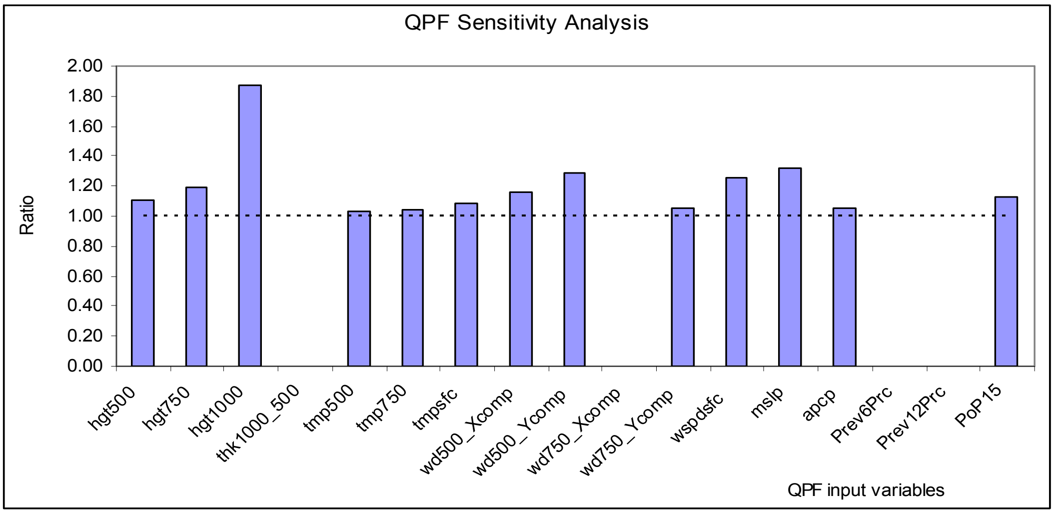

Figure 4. Sensitivity analysis for QPF NN (rainy season).

The output produced by the QPF neural network reduced the systematic biases in precipitation. The distribution of predictions and observations illustrates this improvement (Figure 5). The least squares linear regression indicates over forecast for precipitation amounts less than 14 mm/3hs, and under forecast for higher amounts. The linear correlation coefficient between the observations and the predictions is $r^2 = 0.215$.

[4] It has been noted that as the mean squared error function is used to optimize the ANN free parameters, the data has to be normally distributed (Fortin et al., 1997). This, however, has not been confirmed by experimental trials.

In addition, values for precipitation were accumulated for every 48-hour period, corresponding to each RDAPS forecasting time, where precipitation was observed. Figure 6 shows the improvement in quantitative forecast attained by the QPF neural network. It is worth to note that while total precipitation observed during the rainy season was 1261 mm, RDAPS predicted only 40.8 mm of rainfall, and total precipitation forecasted by QPF NN was 808 mm.

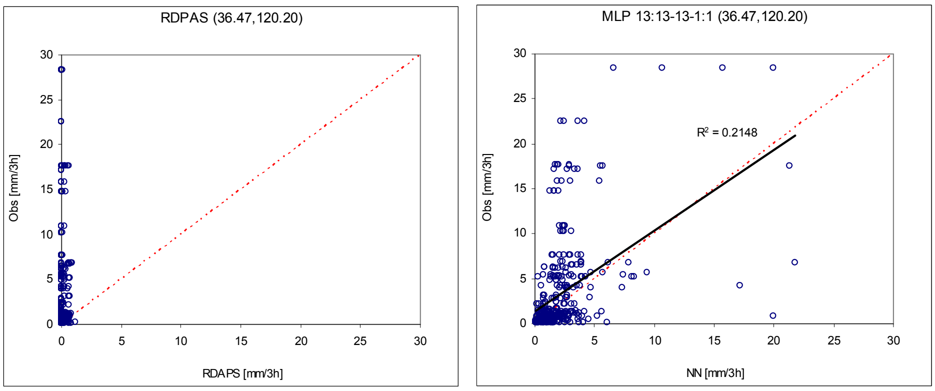

Figure 5. Scatterplots of predictions and observations for precipitation accumulated in 3-hour periods (rainy season): *raw* RDAPS output on the left, and NN QPF forecast on the right. Red, dashed line is the observation = forecast line. The black, solid line is the least squares fit regression line.

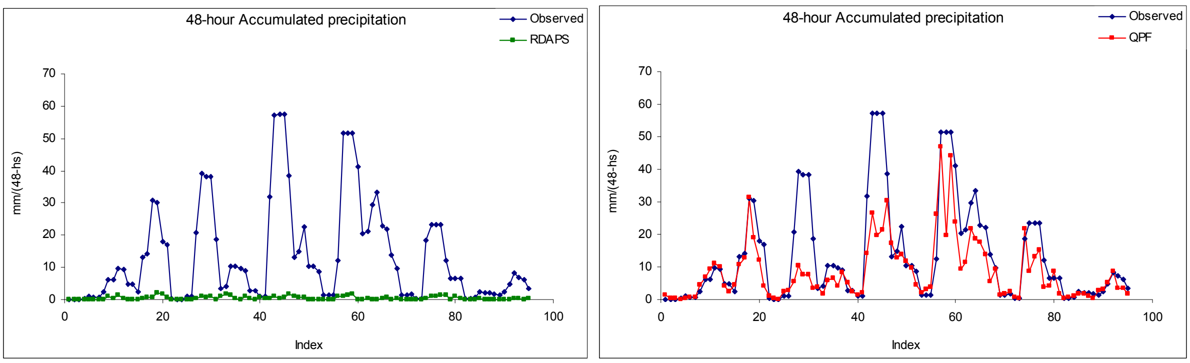

Figure 6. Rainy season precipitation accumulated for each 48-hour RDAPS run. Plot on the left shows comparison between observations and RDAPS predictions; graph on the right compares observations against the forecast of QPF neural network.

	RDAPS	QPF NN
Accumulated Bias [mm]	-1220	-453
Mean Bias [mm]	-3.1	-2.3
St. Deviation of Bias	5.13	4.56
RMS	5.99	4.69

Table 1. Rainy season comparison of RDAPS and QPF NN statistics.

5 CONCLUSION

Both neural networks, probability of precipitation, and quantitative precipitation forecast, produced significant improvement on the RDAPS NWP model rainfall output. The reduction on the large systematic bias was especially remarkable. The performance of both networks may be improved by retraining, probably every month. In addition, it is expected that performance of the networks will improve once atmospheric profile data are incorporated in the analysis.

Artificial neural network is clearly a powerful technique that can be applied to improve the output of NWP models. However, the key to the optimal performance of ANN is to have a large data set relevant to the predictand variable. The more complex the process to be modeled by the ANN, the larger the data set needs to be.

The size of the data set used in this study seemed to be adequate for the proposed objective. However, most available variables used for neural networks training did not prove to be of importance. Clearly, further experimentation is necessary in order to properly utilize the potential of neural networks. This additional work should be conducted utilizing the following additional atmospheric variables:

- Total column precipitable water
- Relative humidity (values at, for example, 100, 750, and 500 hPa)
- Vertical velocity
- Measures of atmospheric instability to vertical displacements (*e.g.*, CAPE)

REFERENCES

Kang, B., and Ramírez, J. A. (2002). "Stochastic Space-Time Downscaling for GCM Precipitation." *2002 American Geophysical Meeting (AGU) Spring Meeting*, Washington, DC.

Kuligowski R. J. and Barros A. P. (1998). "Localized Precipitation Forecasts from a Numerical Weather Prediction Model Using Artificial Neural Networks", *Weather and Forecasting*, 13, pp. 1194-1204

Silverman D. and Dracup J. A., 2000: "Long Range Precipitation Prediction in California; A Look Inside the 'Black Box' of a Trained Network", in *Artificial Neural Networks in Hydrology*, Edited by: R.S. Govindaraju and A.R. Rao, Kluwer Academic Publisher.