ADVANCES IN ATMOSPHERIC SCIENCES, VOL. 23, NO. 3, 2006, 405–414

A Short-Range Quantitative Precipitation Forecast Algorithm Using Back-Propagation Neural Network Approach

FENG Yerong*[1,2] (冯业荣) and David H. KITZMILLER[3]

[1] *Department of Atmospheric Science, Zhongshan University, Guangzhou* 510275

[2] *Guangdong Provincial Meteorological Observatory, Guangzhou* 510080

[3] *Hydrology Laboratory, Office of Hydrologic Development, National Weather Service, NOAA, USA*

(Received 19 January 2005; revised 25 June 2005)

ABSTRACT

A back-propagation neural network (BPNN) was used to establish relationships between the short-range (0-3-h) rainfall and the predictors ranging from extrapolative forecasts of radar reflectivity, satellite-estimated cloud-top temperature, lightning strike rates, and Nested Grid Model (NGM) outputs. Quantitative precipitation forecasts (QPF) and the probabilities of categorical precipitation were obtained. Results of the BPNN algorithm were compared to the results obtained from the multiple linear regression algorithm for an independent dataset from the 1999 warm season over the continental United States. A sample forecast was made over the southeastern United States. Results showed that the BPNN categorical rainfall forecasts agreed well with Stage III observations in terms of the size and shape of the area of rainfall. The BPNN tended to over-forecast the spatial extent of heavier rainfall amounts, but the positioning of the areas with rainfall $\geqslant$25.4 mm was still generally accurate. It appeared that the BPNN and linear regression approaches produce forecasts of very similar quality, although in some respects BPNN slightly outperformed the regression.

Key words: quantitative precipitation forecast, BP neural network, WSR-88D Doppler radar, lightning strike rate, infrared satellite data, NGM model

doi: 10.1007/s00376-006-0405-7

1. Introduction

In the past decade, many new unconventional observations from Doppler radars, surface automated meteorological observing systems, lightning detection devices, and satellite measurement systems have proved to be an effective way for field weather forecasters and forecast system developers to capture the signals of thunderstorm weather phenomena such as strong hail, damaging wind gusts, tornadoes and flash floods.

The 0–3-h quantitative precipitation forecast (QPF) is thought to be important, because in the rainy season, heavy rainfall accompanying a convective weather system often takes place in a very short period of time during a day and causes great damage to property, disruption to traffic or even loss of life. Aiming at solving the short-range precipitation problem, the U.S. National Oceanic and Atmospheric Administration (NOAA) National Weather Service (NWS) has developed an operational short-range (0–3-h) QPF (Kitzmiller et al., 2001) under the framework of the System for Convective Analysis and Nowcasting (SCAN, Smith et al., 1998). This automated ADvective STATistical system (ADSTAT) utilizes data from remote sensor observations (radar reflectivity, lightning strike rate, and satellite infrared imagery) and from operational numerical weather prediction model forecasts to produce rainfall probabilities and rainfall amount forecasts for areas of interest. The probabilities are for rainfall exceeding approximately 2.5, 12.5, 25.0, and 50.0 mm during the 3-h period immediately following the observation time. The rainfall amount forecasts are for five categories: <2.54, 2.54–12.60, 12.70–25.30, 25.40–50.70, and >50.70 mm. These values represent the highest amount at any place within boxes of a 40-km forecast grid, as estimated from an operational radar or gauge observations.

The ADSTAT probabilities are produced from equations derived through forward-selection linear

*E-mail: Yerong_feng@grmc.com.cn

screening regression within a set of 34 candidate predictors. Some of the predictors are presented to the procedure in both original and transformed versions, the transformation being based on the subjective selection of predictors and sometimes the truncation of their statistical range, depending on those predictors' relationships with the various predictands (Kitzmiller et al., 2001; Charba, 1979). Presently, the rainfall amount forecasts are derived by comparing the probabilities to pre-determined thresholds (25%, 23%, 18%, and 9% for each category, respectively). The thresholds were selected such that they yield forecasts that detect most rain events without undue overprediction.

Recently, neural network (NN) techniques have been applied to many meteorological problems, such as predicting tornadoes (Marzban and Stumpf, 1996), damaging winds (Marzban and Stumpf, 1998), thunderstorms (McCann, 1992), quantitative precipitation (Hall et al., 1999; Kuligowski and Barros, 1998; Koizumi, 1999), and even long-range monsoon precipitation (Wu et al., 2001).

Although linear regression seems to be a commonly-used approach in many forecast algorithms, we chose a back-propagation neural network (BPNN) method (Bishop, 1995) to develop a short-range (0–3-h) QPF algorithm. The BPNN approach considers the nonlinear effect of all candidate predictors as a whole, while the linear regression only selects a limited set of predictors. It is possible to incorporate some nonlinear effects and predictor interactions in a linear regression approach by deriving new predictors from the original set. However, this procedure is time consuming, not necessarily comprehensive, and subjective.

In this experimental study, the BPNN approach used 30 candidate predictors that were utilized in the ADSTAT system. This BPNN approach, which still requires some manual intervention, does not require subjective choices regarding individual predictors, as was done in the ADSTAT system , and it is more economical and guides the developer through the error reduction process. Our aim was to test the potential for simplification of the forecast algorithm derivation procedure and to determine whether the BPNN method could yield better results than linear regression when applied to the 0–3-h QPF problem. Both probabilistic and categorical forecasts from the BPNN approach were compared to those from a multiple linear screening regression approach by evaluating the forecasts' performance on a set of cases not used in developing either system.

2. Neural network architecture

A three-layer BPNN was used in this study. The network architecture includes 30 input nodes, 10 hidden-layer nodes and 2 output nodes. The number of input nodes (predictors) and the number of output nodes (the predictands) are generally determined by the prediction problem itself. The number of nodes in the hidden layer is usually selected based on the number of input and output nodes (Aviolat et al., 1998; Kuligowski and Barros, 1998). We chose to use 10 hidden nodes. Aviolat et al. (1998) pointed out that if there are too few hidden-layer nodes, the NN will not be able to solve the learning problem, while if there are too many, the convergence of the learning process will be slowed and the output could be unreliable when applied to independent data. In the course of our study, some tests on the optimum number of hidden nodes (ranging up to 21) were carried out. The use of more than 10 nodes led to no consistent improvement in the forecasts. The three-layer BPNN consists of two matrices of multiplication weights. If there are N input nodes, M hidden layer nodes, and K output nodes, then the first matrix has $M \times N$ elements, and the second has $K \times M$ elements. The set of values in the hidden layer $\boldsymbol{H}$ is determined from the values in the input layer by:

$$\boldsymbol{H}^{\mathrm{T}} = \boldsymbol{F}(\boldsymbol{W} \cdot \boldsymbol{X}^{\mathrm{T}} + \boldsymbol{Q}^{\mathrm{T}}) \,, \tag{1}$$

where $\boldsymbol{W}$ is an $M \times N$ weight matrix that logically maps N input nodes to M hidden-layer nodes, $\boldsymbol{X}$ is a vector of length N containing input predictor values, and $\boldsymbol{Q}$ is a vector of length M containing activation threshold values. T denotes the transpose of a matrix. $\boldsymbol{F}(\boldsymbol{C})$ is a function vector whose components are sigmoid functions of each component c in vector $\boldsymbol{C}$. The sigmoid function itself is defined by:

$$f(c) = \frac{1}{1 + e^{-c}} \,. \tag{2}$$

Similarly, the vector of output values $\boldsymbol{Y}$ is determined from values in the hidden layer by:

$$\boldsymbol{Y}^{\mathrm{T}} = \boldsymbol{F}(\boldsymbol{V} \cdot \boldsymbol{H}^{\mathrm{T}} + \boldsymbol{O}^{\mathrm{T}}) \,, \tag{3}$$

where $\boldsymbol{V}$ is a $K \times M$ weight matrix that logically maps M hidden-layer nodes to K output nodes, and $\boldsymbol{O}$ is a vector of length K containing activation threshold values.

The training portion of BPNN development involves minimizing the squared error E of the BPNN output for a training sample consisting of n cases, as expressed by

$$E = \sum_{i=1}^{n} \|\boldsymbol{Y}_i - \boldsymbol{P}_i\|^2 \,, \tag{4}$$

where $\| \ \|$ is the Euclidean distance between the BPNN output vector $\boldsymbol{Y}$ (forecasted rainfall category vector) and the observed rainfall category vector $\boldsymbol{P}$.

To minimize the error function E, an iterative gradient descent algorithm (Oh and Lee, 1999) was applied to adjust the weight matrices $\boldsymbol{W}$ and $\boldsymbol{V}$ (also

the activation threshold vectors $\boldsymbol{Q}$ and $\boldsymbol{O}$) toward optimum values, as represented by

$$\boldsymbol{W}_\tau = \boldsymbol{W}_{\tau-1} - \varepsilon_1 \boldsymbol{D}_{\tau-1} \,, \tag{5}$$

$$\boldsymbol{V}_\tau = \boldsymbol{V}_{\tau-1} - \varepsilon_2 \boldsymbol{D}_{\tau-1} \,, \tag{6}$$

where τ denotes a step of iteration. ε_1 and ε_2 represent two small positive numbers. $\boldsymbol{D}$ indicates a matrix whose entries are the partial derivatives of error E with respect to the corresponding entries in weight matrices $\boldsymbol{W}$ or $\boldsymbol{V}$.

3. Training dataset

The training sample was created from data collected during the warm seasons (May-September) of 1996–1998 over the eastern two-thirds of the continental United States (CONUS). Predictor variables were chosen from among the set utilized in the operational ADSTAT algorithms. These included predictors based on initial-time and advected radar reflectivity, lightning, and satellite observations, and others derived from forecasts of the Nested Grid Model (NGM; Hoke et al., 1989), including humidity, stability, moisture divergence, and precipitation. The remote-sensor predictors were all interpolated to a polar stereographic grid with approximately 10-km mesh length over the CONUS; this grid has precisely four times the resolution of the 40-km forecast grid, which basically meets the requirements of operational forecasts of flash floods for strong, warm season, thunderstorm precipitation. The initial time of the NGM runs was 1200 UTC for predictors valid between 1500 and 0259 UTC, and 0000 UTC for predictors valid between 0300 and 1459 UTC. This accounts for the operational lag time in the preparation and dissemination of the model runs. The 6-h precipitation amounts were those whose valid period included the 3-h valid period of the extrapolative forecasts. The nominal spatial resolution of the NGM variables was 80 km.

A total of eight "initial times" of the extrapolative forecasts were considered, starting at 0230 UTC and continuing at 3-hour intervals through 2330 UTC; the corresponding valid periods were 0300–0600 UTC through 0000–0300 UTC.

Forecasts of precipitation based on the remotely-sensed fields (reflectivity, cloud-top temperature, and lightning strike rate) were made by an advection algorithm through which the initial-time fields were advected with the NGM-predicted 700–500 hPa mean wind vector, which proved to be a robust estimator of radar echo motion over the United States during the warm season. The advection procedure was performed at 15-minute intervals and carried out on the 10-km polar stereographic radar analysis grid.

A rainfall amount forecast was made by converting the forecasted reflectivity to rainfall rate via the Weather Surveillance Radar-1988 Doppler (WSR-88D) default $Z - R$ relationship:

$$R = (Z/300)^{0.7142} \,, \tag{7}$$

where R is rainrate in mm h^{-1} and Z is reflectivity in mm^6 m^{-3}. Though this forecast has a large positive bias relative to observed rainfall amounts, it has still proved to be useful as a statistical predictor.

Radar- and lightning-based predictors were derived from either the largest value or the time-averaged value of reflectivity or lightning strike rate within the 40-km forecast grid boxes; satellite-based predictors were based on the minimum forecasted cloud-top temperature by finding the maximum difference between the 700-hPa temperature and the satellite-observed temperature.

One useful radar-based predictor was derived by summing up the number of 10-km grid boxes within a 3×3 region of 40-km grid boxes (120 km$\times$120 km) that were forecasted to have a particular echo level during the three-hour period. Such predictors account for both the presence of intense precipitations and their areal coverage near a given forecast grid box. One of these, called NLVL456, was calculated by summing up the number of boxes with level 4 or higher echoes, the number with level 5 or higher echoes, and the number with level 6 echoes.

Information on large-scale atmospheric conditions (e.g., low level jet, lower layer convergence, and helicity) is important to the evolution of strong convective thunderstorm systems (Xu et al., 2000; Fei and Tan, 2001). Thus the environmental conditions were derived from NGM model forecasts of several variables, interpolated to the forecast grid. These predictors included surface-to-500-hPa mean relative humidity, precipitable water, K index, lifted index, model-generated rainfall, and moisture divergence.

A total of 30 predictor variables were used (refer to Table 1). Prior to being used in the training process, all predictors were normalized to the range 0.01–0.99, rather than the range 0–1, which can lead to singularity problems. Predictors were derived from data at the initial time, during each 1-h interval within the forecast period, and over the entire 3-h forecast period. Additional predictors were derived by finding the largest value within a square region of 120×120 km^2 centered on the 40 km grid of concern. These predictors reflect the influence of convective features expected to pass near the grid box.

The statistical rainfall predictands were derived from WSR-88D Stage III precipitation estimates produced operationally by the National Weather Service River Forecast Centers (Breidenbach et al., 1998). These 1-h analyses are based on a combination of radar estimates and gauge observations. The data represent average rainfall over 4-km grid boxes. Predictands

were derived from the highest 3-h rainfall value observed within contiguous 10×10 grid box subsections. These contiguous subsections comprise the Manually-Digitized Radar (MDR) 40-km grid. The four predictands which stand for the four rainfall categories are binary, being zero when the threshold rainfall amount was not met, and unity when the threshold was met or exceeded. A "case" is thus defined as the predictor and predictand values for one 40-km grid box during one 3-h period on one day. Individual cases were extracted from a subset of the available grid boxes.

It is necessary to deal with missing radar data in the reflectivity imageries, since radar observation coverage gaps always exist. Forecasts derived from other data such as satellite, NGM, and lightning data must be used to fill such gaps. Therefore, an algorithm containing no radar-based predictors provides forecasts for areas where radar data are lacking at the initial time or for areas downstream of the coverage gaps.

First-guess probability fields are derived under the assumption that missing grid boxes contain reflectivity that is 15 dBZ, which is the most likely value. Then, another radar forecast is created by extrapolation, in which missing indicators are tracked in the same manner as radar echoes. Two missing-indicator fields are derived from this forecast. One contains the percentage of each 40-km grid box that is covered by missing indicators at the initial time. The other field contains the percentage of the space-time domain over each 40-km box that is occupied by missing indicators during the forecast period.

The reflectivity is quality-controlled by comparison with satellite data, to identify and remove anomalous propagation and ground clutter. Within any grid box, the analysis contains the highest observed reflectivity, expressed as one of 7 categories (levels):

Category	0	1	2	3	4	5	6
Reflectivity (dBZ)	<15	15–29	30–39	40–44	45–49	50–54	>54

The training dataset had 89556 cases. The relative frequencies of events in each rainfall category were 85.2%, 8.4%, 3.4%, 2.3%, and 0.7%, which decreased from lowest to highest dramatically.

4. Training process

We originally attempted to develop a five-output-node NN to be applied to the five-category categorical precipitation forecast problem. However, the results were unsatisfactory in that almost all forecasts were in the lowest two amount categories.

This was apparently due to the rarity of higher rainfall events in the training sample. From the relative frequencies of occurrence of each rainfall category, we can see that non- or low-precipitation events dominate the sample. The high precipitation events ($\geqslant$25.40 mm) are very rare, comprising only 3% of all cases. Thus the error minimization process selected weights that best differentiated between only the lower-amount categories.

We therefore transformed the five-category forecasting problem into a set of binary decisions by training four sets of two-output-node NNs as follows:

NN1: $<$2.54 and $\geqslant$2.54 mm;
NN2: 2.54–12.60 and $\geqslant$12.70 mm;
NN3: 12.70–25.30 and $\geqslant$25.40 mm;
NN4: 25.40–50.70 and $\geqslant$50.80 mm.

To train neural network NN1, all 89556 cases were divided into two parts, 76311 cases with rainfall $<$0.1 and 13245 cases with rainfall $\geqslant$2.54 mm. The training sample for NN2 was taken by dividing all 13245 cases with observed rainfall $\geqslant$2.54 mm into two groups, namely 7499 cases with rainfall of 2.54–12.60 mm and 5746 with rainfall $\geqslant$12.70 mm. Training samples for NN3 and NN4 were constructed in the same manner, by extracting those cases with rainfall exceeding the lower of the two thresholds and dividing the sample into those that did and did not reach the upper threshold.

Thus each training data sample contains a population in which higher and lower categories are both well represented. As will be shown, this approach worked well when applied to independent data.

We refer to the output of the two NN nodes as P_{low} and P_{high}. If the output of the NN is such that $P_{\text{low}} \geqslant P_{\text{high}}$, then the lower rainfall category is forecasted. If the converse holds, then the higher category is selected. To produce a categorical forecast from these NNs, the output of NN1 is first evaluated. If the output indicates that at least 2.54 mm should be forecasted, then the output of NN2 is evaluated. If NN2 indicates that at least 12.70 mm should be forecasted, then NN3 is evaluated, and so on.

Of course, the NN approach is also applicable to rainfall probability forecasting. For example, the probability of rainfall in excess of 2.54 mm is given by the value of P_{high} in NN1.

The performance of each NN in the training phases was measured by scores derived from a contingency table based on the four possible outcomes. Here, "higher" and "lower" refer to the two rainfall categories:

x: Higher forecasted, higher observed (correct forecast for higher)
y: Lower forecasted, higher observed (missed event)
z: Higher forecasted, lower observed (false alarm)
w: Lower forecasted, lower observed (correct forecast for lower)

Three standard scores and a new index were used to determine the effects of the back-propagation train-

Table 1. The input predictors in the neural network's training processes.

Abbreviation	Description
INITIAL MAX REFL	Maximum reflectivity at initial time
HR1 MAX REFL	Maximum reflectivity at 1st hour extrapolation
HR2 MAX REFL	Maximum reflectivity at 2nd hour extrapolation
HR3 MAX REFL	Maximum reflectivity at 3rd hour extrapolation
HR0-3 MAX REFL	Maximum reflectivity over 3-h extrapolation
HR0-3 NLVL456	Number of 10-km meshes in a 40-km grid with echo levels 4, 5 and 6
HR0-3 NLVL456 3X3	Number of 10-km meshes in a 120 km×120 km area with echo levels $\geqslant 4, \geqslant 5, \geqslant 6$
HR1 MAX RAINFALL	Maximum radar-derived rainfall at 1st hour extrapolation
HR2 MAX RAINFALL	Maximum radar-derived rainfall at 2nd hour extrapolation
HR3 MAX RAINFALL	Maximum radar-derived rainfall at 3rd hour extrapolation
HR0-3 MAX RAINFALL	Maximum radar-derived rainfall within 3 hour extrapolation
PCT MSG RADR INIT	Percentage of missing radar data in a 40-km grid at initial time
CT MISSING RADR	Percentage of occupied missing data over a 40-km grid in 3 h forecast period
INIT MX LTG STRIKE RT	Maximum lightning strike rate at initial time
HR1 MX STRIKE RATE	Maximum lightning strike rate at 1st hour extrapolation
HR2 MX STRIKE RATE	Maximum lightning strike rate at 2nd hour extrapolation
HR3 MX STRIKE RATE	Maximum lightning strike rate at 3rd hour extrapolation
HR0-3 MX STRIKE RATE	Maximum lightning strike rate over 3-h extrapolation
INIT MIN SAT TEMP	Minimum satellite temperature at initial time
INIT MAX T700-TSAT	Maximum temperature difference between 700-hPa and cloud top (initial time)
0-1H MAX T700-TSAT	Maximum temperature difference between 700-hPa and cloud top (1 hour)
0-3H MAX T700-TSAT	Maximum temperature difference between 700-hPa and cloud top (3 hour)
K INDEX	NGM model-predicted K index
PW	NGM model-predicted precipitable water
MEAN RH %	NGM model-predicted surface–500-hPa mean relative humidity
PRECIP EFFICIENCY	NGM generated precipitation efficiency
NGM 6-h RAINFCST	NGM 6-h rainfall forecast
MDIV850	NGM moisture divergence at 850 hPa
850-500 LI	NGM predicted/analyzed lifted index between 850 and 500 hPa
INITIAL MX REFL 3×3	Maximum reflectivity within a 3×3 40-km grid

ing. The standard scores were probability of detection (POD):

$$\mathrm{POD} = \frac{x}{x+y}\,, \tag{8}$$

false alarm ratio (FAR):

$$\mathrm{FAR} = \frac{z}{x+z}\,, \tag{9}$$

and critical success index (CSI):

$$\mathrm{CSI} = \frac{x}{x+y+z} \tag{10}$$

A fourth index, called π, was developed in the course of this work to facilitate monitoring the progress of the BPNN training process. The value of π indicates the prevalence of errors relative to successes in the BPNN output. It is defined as follows.

The value of π approaches 0 for perfect forecasts (no errors) and increases as errors begin to predominate. This quantity proved useful in determining when to terminate training iterations.

$$\pi = 0.125\left(\frac{y+z}{x} + \frac{y+z}{w} + \frac{z}{w+z} + \frac{y}{w+y} + \frac{z}{x+z} + \frac{y}{x+y}\right)\,. \tag{11}$$

In training neural networks, it is common to terminate iterations when it is apparent that further training is leading to only limited gains in forecast skill. The forecast skill as a function of the number of training iterations is shown by the performance scores in Fig. 1. Each iteration corresponds to one adjustment to the weight matrices $\boldsymbol{W}$ and $\boldsymbol{V}$ and to the activation vectors $\boldsymbol{Q}$ and $\boldsymbol{O}$. The ratio $E/E0$ indicates the squared error of the NN output relative to its initial error. This ratio decreases slowly after the first 50 training iterations. The π index decreases sharply in the beginning and reaches a minimum at the 50th iteration, and, after some perturbations, decreases slowly with training time. Higher POD and CSI scores corre-

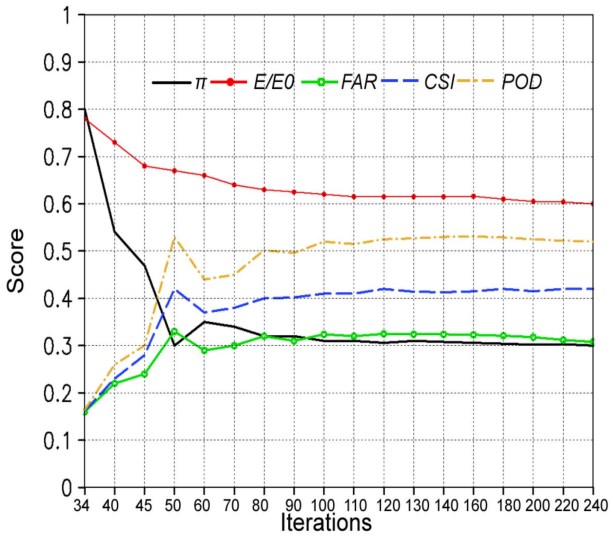

Fig. 1. The performance of NN1 during 240 training cycles.

spond to smaller values of π. We believe that π could therefore serve as a useful termination criterion for the training process.

5. Sample forecasts

On 3 May 2001, a westerly trough at 500 hPa affected the southern part of the country, causing widespread heavy rainfall over the southeastern part of the country. The categorical rainfall forecast and corresponding observations for this convective precipitation event appear in Fig. 2 and Fig. 3. The BPNN categorical rainfall amount forecasts agree well with Stage III observations (Fig. 3) in terms of size and shape of the area with rainfall $\geqslant$2.54 mm. The BPNN tended to over-forecast the spatial extent of heavier rainfall amounts, but the positioning of the areas with rainfall $\geqslant$25.40 mm is still generally accurate. Output from NN1 is shown in Fig. 4. As has been mentioned

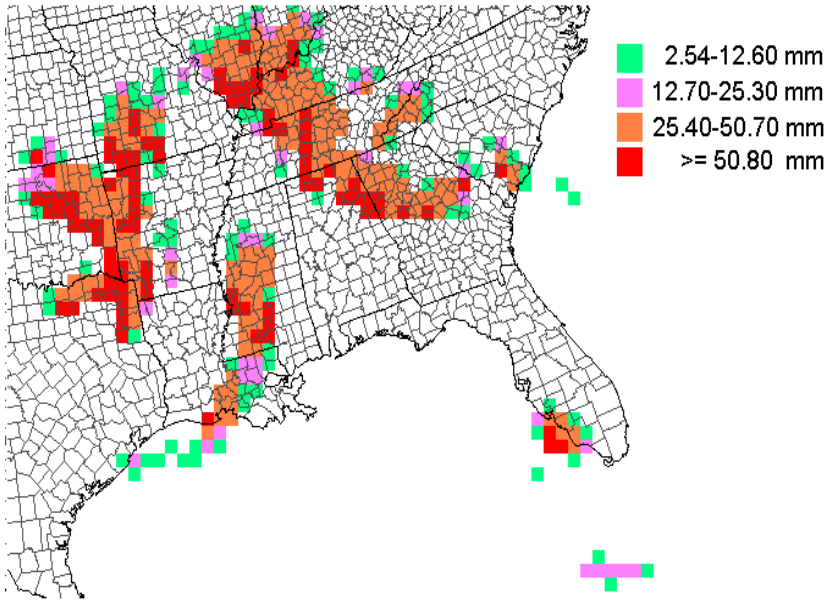

Fig. 2. Rainfall amount forecasts from BPNN output, valid for 2100–0000 UTC 2–3 May 2000.

above, this output is equivalent to the probability of rainfall $\geqslant$2.54 mm.

The size and shape of the area with a probability of 30% agree well with those of the area of observed rain $\geqslant$2.54 mm. Outputs from NN2-NN4 denote the likelihood to produce $\geqslant$12.70 mm, $\geqslant$25.40 mm, and $\geqslant$50.80 mm rainfalls respectively. As is shown in Fig. 5, the colored areas indicate the areas in which rainfall $\geqslant$50.80 mm are possible to occur. Larger values indicate that the areas are more likely to experience heavy rainfalls within 3 hours.

6. Comparative verification

To objectively assess the quality of the trained NNs when applied to independent data, we evaluated them for a set of cases from the 1999 warm season. There were a total of 43884 cases. The relative frequency of events in each rainfall category was (from lowest to highest) 85.3%, 7.9%, 3.6%, 2.3%, and 0.9%, respectively. The verification results were compared with those from the regression procedure. A verification contingency table for the BPNN output for the 1999

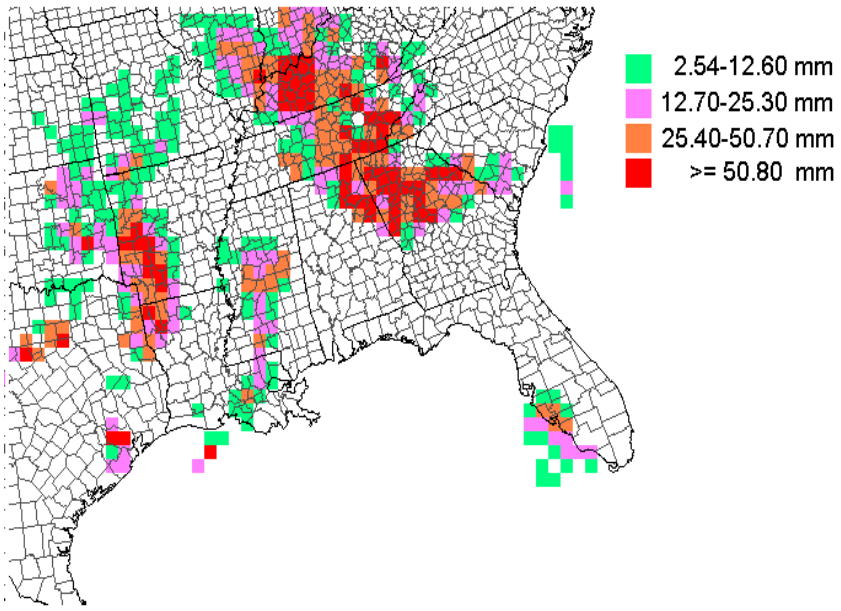

Fig. 3. Stage III radar/gauge rainfall estimates, for the period 2100–0000 UTC 2–3 May 2000.

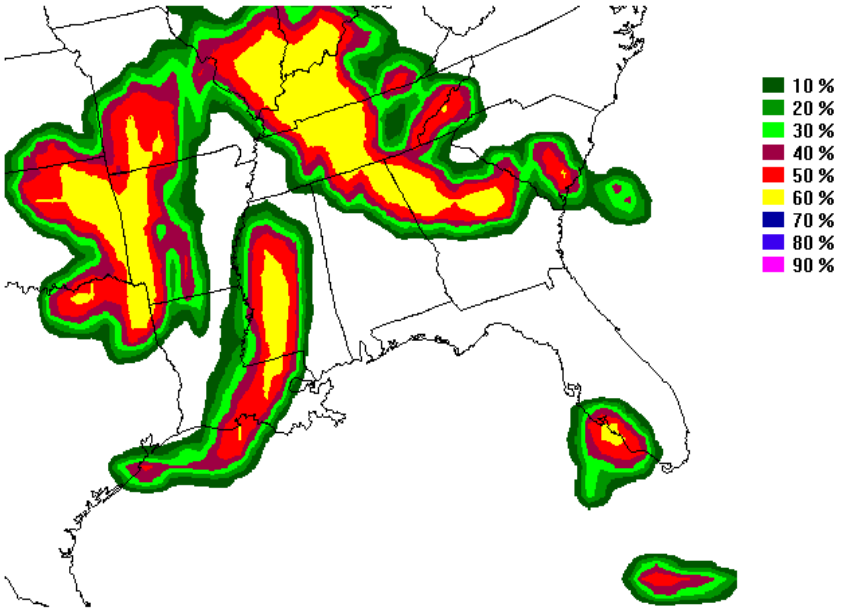

Fig. 4. Probabillty of rainfall $\geqslant$2.54 mm, based on output of NN1, valid for 2100–0000 UTC 2–3 May 2000.

cases appears in Table 2. The forecasts define the columns, and the verifying observations, the rows. Thus correct forecasts are counted in the diagonal from upper left to lower right.

For comparison, results from the linear regression system are shown in Table 3. These forecasts were derived from the set of regression-based probabilities, i.e., equations listed in Table 4, hereafter referred to as Y_1, Y_2, Y_3, and Y_4. Categorical forecasts were derived from the probabilities by comparing each probability to threshold values of 25%, 22%, 18%, and 9%, respectively. The forecasted category corresponded to the highest rain amount for which the threshold probability was exceeded, subject to the condition that the probabilities for all lower amounts also exceeded their thresholds.

One of the simplest measures for assessing differences between the BPNN and regression forecasting

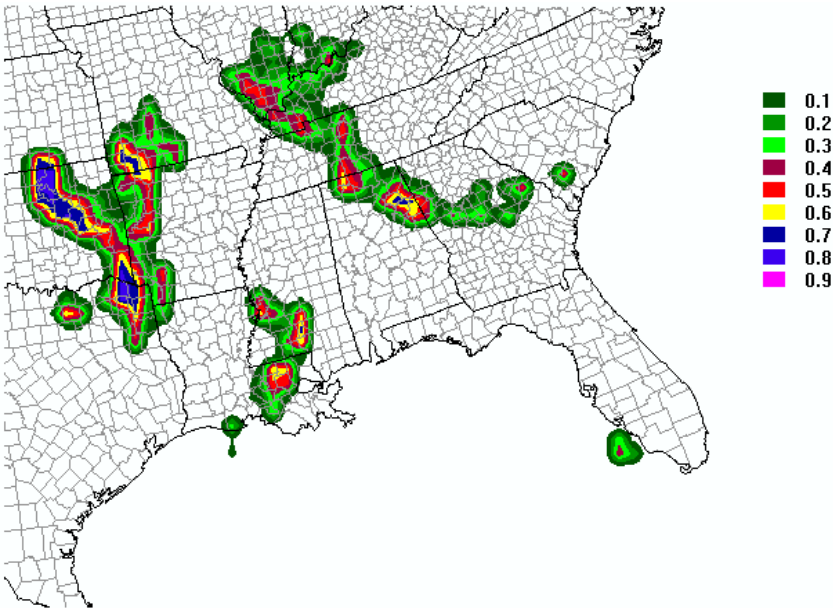

Fig. 5. Output from NN4, for rainfall ⩾50.80 mm, valid for 2100–0000 UTC 2–3 May 2000.

Table 2. Verification of categorical rainfall forecasts based on BPNN, 1999 warm season (mm).

Observation	BPNN Forecasts				
	⩾50.80	25.40–50.70	12.70–25.30	2.54–12.60	<2.54
⩾50.80	39	144	62	71	93
25.40–50.70	76	322	174	227	227
12.70–25.30	59	318	217	479	492
2.54–12.60	66	364	337	1202	1474
<2.54	72	343	257	1893	34876

Table 3. As in Table 2, except forecasts based on linear regression (mm).

Observation	Regression forecasts				
	⩾50.80	25.40–50.70	12.70–25.30	2.54–12.60	<2.54
⩾50.80	46	120	98	92	53
25.40–50.70	90	288	239	268	141
12.70–25.30	78	330	356	489	312
2.54–12.60	74	419	615	1298	1037
<2.54	44	293	738	2968	33398

systems is the percentage of all forecasts that are verified in the correct category. The percentage correct is derived from the total number of cases on the ascending diagonal divided by the total number of cases (83.5% for the BPNN and 81% for the regression approach). However, these percentages are dominated by the large number of minimal-skill cases in which the <2.54 mm category was both forecasted and observed. A potentially more useful indication of the skill is the percentage of forecasts correct within the subset of cases where 2.54-mm rain was forecasted or observed. In these tables, this is the set of cases excluding the box at the extreme lower right; the percentage correct is based on the total number of cases on the diagonal, divided by the total number of cases minus those for which <2.54 mm was both forecasted and observed. This percentage reduces to the standard CSI if the contingency table has only two categories. The values for Tables 2–3 are 20% for the BPNN and 19% for the regression procedure. Finally, the percen-

Table 4. Probabilistic rainfall forecast equations based on linear regression.

Rainfall categories	Probabilistic equations
⩾50.80 mm	Y_4=−0.00889+6.570(HR0-3 MX STRK RATE)+ 8.409(NGM 6-h RAINFCST) +6.311(INITIAL MX REFL)+4.892(HR3 MX STRIKE RATE) −5.754(0-1H MAX T700-TSAT)+4.027(INIT MAX T700-TSAT)
⩾25.40 mm	Y_3=−0.02477+14.493(INITIAL MAX REFL)+12.470(HR0-3 MX STRK RATE) +10.886(NGM 6-h RAINFCST)+7.279(HR3 MX STRIKE RATE) −6.945(0-1H MAX T700-TSAT)+ 5.58(INIT MAX T700-TSAT)
⩾12.70 mm	Y_2=−0.02476+26.043(INITIAL MAX REFL)+12.594(HR0-3 MX STRK RATE) +11.640(NGM 6-h RAINFCST)+ 7.626(HR0-3 MX STRIKE RATE)
⩾2.54 mm	Y_1=−0.03299+12.017(INITIAL MAX REFL)+11.541(HR0-3 MX STRIKE RATE) +14.895(NGM 6-h RAINFCST)+8.450(HR0-3 MX REFL) +4.937(HR0-3 MX STRK RATE)

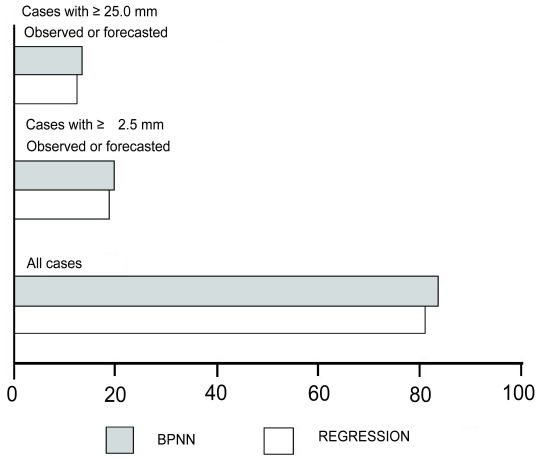

Fig. 6. Percentages of categorical forecasts verified in the correct category, for all cases, for all cases involving >2.54-mm rainfall, and for all cases involving >25.4-mm rainfall.

tage correct among the 25.4-mm cases was derived by similar logic. These scores were again rather low (13.5% for the BPNN, 12.5% for regression). The results of all three comparisons appear in Fig. 6.

Thus it is apparent that there are minimal differences between the two sets of categorical forecasts. Moreover, the forecasts are not highly accurate. However, we noted that for many of the forecasts the verifying observations were within one category of the forecasted one. These cases are those falling on the diagonal or boxes immediately above or below the diagonal in Tables 2–3. For all cases involving rainfall ⩾2.54 mm, 53% of BPNN and 65% of regression forecasts had observed rainfall within one category of the forecasted one. For all cases involving forecasts or observations ⩾25.4 mm, these percentages were 34% for the BPNN and 33% for the regression forecasts.

7. Summary

A BPNN was applied to establish relationships between the 0-3-h rainfall and the predictors ranging from extrapolative forecasts of radar reflectivity, satellite-estimated cloud-top temperature, lightning strike rates, and NGM model outputs. Quantitative precipitation forecasts and the probabilities of categorical precipitation were obtained. Results of the BPNN algorithm were compared to the results obtained from the multiple linear regression algorithm for an independent dataset from the 1999 warm season. It appears that the two approaches produce forecasts of very similar quality, although in some respects BPNN slightly outperformed the regression.

Though generally the BPNN algorithm could capture the basic spatial pattern of the precipitation ⩾2.54 mm, the location and spatial extent of heavier categorical rainfall forecasts still lacked precise. Future improvement should include the evolution features in the convective rain cells. Predictors describing the growth and dissipation processes of the convective echoes should be tactically considered in the algorithm.

In addition, a π index was developed in this effort to facilitate monitoring the progress of the BPNN training process. This π index indicated the prevalence of errors relative to successes in the BPNN output. It proved to be a useful termination criterion for the training process.

A plan to develop an operational 0–3-h QPF forecasting system using similar techniques is under consideration at the Guangdong Provincial Meteorological Observatory of China. This future work will focus on interpreting China's new generation Doppler radar (CINRAD) observations, Guangzhou mesoscale numeric weather prediction model products, surface automatic meteorological observations and China's geostationary meteorological satellite FY2C information.

Acknowledgments. The authors would like to thank Dr. Harry R. Glahn, chief of the Meteorological Development Laboratory (MDL), for his encouragement in

this work and kind guidance in the preparation of this paper. Tribute must also be paid to staff at the MDL for their help in many aspects such as data collection and computer resources, etc. This work was carried out during the first author's tour of duty at MDL, National Weather Service (NWS), USA, while the second author was also employed there. The cooperation between the China Meteorological Administration (CMA) and the U.S. NWS in the atmospheric sciences and technology facilitated the completion of this effort.

REFERENCES

Aviolat, F., T. Cornu, and D. Cattani, 1998: Automatic cloud observations improved by an artificial neural network. *J. Atmos. Oceanic Technol.*, **15**, 114–126.

Bishop, M. A., 1995: *Neural Networks for Pattern Recognition*. Clarendon Press, 482pp.

Breidenbach, J. P., D. J. Seo, and R. A. Fulton, 1998: Stage II and III post processing of NEXRAD precipitation estimates in the modernized Weather Service. Preprints, 14th Interna. Conf. on Interactive Information Processing Systems, Phoenix, Amer. Meteor. Soc., 263–266.

Charba, J. P., 1979: Two to six hour severe local storm probabilities: An operational forecasting system. *Mon. Wea. Rev.*, **107**, 268–282.

Fei Shiqiang, and Tan Zhemin, 2001: On the helicity dynamics of severe convective storms. *Adv. Atmos. Sci.*, **18**, 67–86.

Hall, T., H. E. Brooks, and C. A. Doswell III, 1999: Precipitation forecasting using a neural network. *Wea. Forecasting*, **14**, 338–345.

Hoke, J. E., N. A. Phillips, G. J. DiMego, J. J. Tucillo, and J. G. Sela, 1989: The regional analysis and forecast system of the National Meteorological Center. *Wea. Forecasting*, **4**, 323–334.

Kitzmiller, D. H., S. D. Vibert, and F. G. Samplatsky, 2001: Short-range forecasts of rainfall amount from an extrapolative statistical technique utilizing multiple remote sensor observations. Preprints. Symposium On Precipitation Extremes, Albuquerque, Amer. Meteor. Soc., 266–270.

Koizumi, K., 1999: An objective method to modify numerical model forecasts with newly given weather data using an artificial neural network. *Wea. Forecasting*, **14**, 109–118.

Kuligowski, R. J., and A. P. Barros, 1998: Localized precipitation forecasts from a numerical weather prediction model using artificial neural network. *Wea Forecasting*, **13**, 1194–1204.

Marzban, C., and G. J. Stumpf, 1996: A neural network for tornado prediction based on Doppler radar derived attributes. *J. Appl. Meteor.*, **35**, 617–626.

Marzban, C., and G. J. Stumpf, 1998: A neural network for damaging wind prediction. *Wea. Forecasting*, **13**, 151–163.

McCann, D. W., 1992: A neural network short-term forecast of significant thunderstorms. *Wea. Forecasting*, **7**, 525–534.

Oh, S. H., and S. Y. Lee, 1999: A new error function at hidden layers for fast training of multiplayer perceptrons. *IEEE Transactions on Neural Networks*, **10**, 960–964.

Smith, S. B., J. T. Johnson, R. D. Roberts, S. M. Zubrick, and S. J. Weiss, 1998: The System for Convection Analysis and Nowcasting (SCAN) 1997–1998 field test. Preprints, 19th Conf. on Severe Local Storms, Minneapolis, Amer. Meteor. Soc., 790–793.

Wu Xiaodan, Cao Hongxing, A. Flitman, Wei Fengying, and Feng Guolin, 2001: Forecasting monsoon precipitation using artificial neural networks. *Adv. Atmos. Sci.*, **18**, 951–958.

Xu Hui, Zhang Weiping, Lang Xuxing, Guo Xia, Ge Wenzhong, and Dang Renqing, 2000: The use of dual-Doppler radar data in the study of 1998 Meiyu frontal precipitation in Huaihe River basin. *Adv. Atmos. Sci.*, **17**, 403–412.