

Precipitation Forecasting Using a Neural Network

TONY HALL*

NOAA/NWS West Gulf River Forecast Center, Fort Worth, Texas

HAROLD E. BROOKS AND CHARLES A. DOSWELL III

NOAA/ERL/National Severe Storms Laboratory, Norman, Oklahoma

(Manuscript received 10 November 1997, in final form 24 November 1998)

ABSTRACT

A neural network, using input from the Eta Model and upper air soundings, has been developed for the probability of precipitation (PoP) and quantitative precipitation forecast (QPF) for the Dallas–Fort Worth, Texas, area. Forecasts from two years were verified against a network of 36 rain gauges. The resulting forecasts were remarkably sharp, with over 70% of the PoP forecasts being less than 5% or greater than 95%. Of the 436 days with forecasts of less than 5% PoP, no rain occurred on 435 days. On the 111 days with forecasts of greater than 95% PoP, rain always occurred. The linear correlation between the forecast and observed precipitation amount was 0.95. Equitable threat scores for threshold precipitation amounts from 0.05 in. (~ 1 mm) to 1 in. (~ 25 mm) are 0.63 or higher, with maximum values over 0.86. Combining the PoP and QPF products indicates that for very high PoPs, the correlation between the QPF and observations is higher than for lower PoPs. In addition, 61 of the 70 observed rains of at least 0.5 in. (12.7 mm) are associated with PoPs greater than 85%. As a result, the system indicates a potential for more accurate precipitation forecasting.

1. Introduction

Forecasts of precipitation are important in a variety of contexts. The probability of precipitation (PoP) is important for many decision makers who are sensitive to the occurrence of precipitation (e.g., Roebber and Bosart 1996). An accurate quantitative precipitation forecast (QPF) can identify the potential for heavy precipitation and possible associated flash flooding, as well as providing information for hydrologic interests. As part of the modernization of the National Weather Service (NWS), more emphasis is being placed on the local generation of QPFs and their subsequent use in hydrological models at River Forecast Centers.

Numerical weather prediction models provide direct QPF guidance for precipitation forecasts. As long as the predictions contain biases and systematic errors, however, postprocessing of the output can improve the raw output. Many statistical methods can be used to do this postprocessing. Traditionally, these have included the

“perfect prog” technique (Klein et al. 1959) and Model Output Statistics (MOS; Glahn and Lowry 1972). Both of these methods use multiple regression techniques to take model output and convert into forecasts of sensible weather. They allow for a single model solution to show uncertainty in the forecast weather. As Murphy (1993) has shown, such an expression of uncertainty can be valuable to forecast users. Recently, hydrometeorologists at the West Gulf River Forecast Center have used another processing technique, neural networks (Müller and Reinhardt 1991), to develop a precipitation forecasting tool for the Dallas–Fort Worth (DFW), Texas, metropolitan area. The neural network scheme uses gridded output from the National Centers for Environmental Prediction’s (NCEP) Eta Model, and upper air soundings from Fort Worth. The forecasts are for the PoP and 24-h precipitation amount from 1200 to 1200 UTC. Upper air soundings are from the beginning of the period and the Eta Model forecast data are from the 1200 UTC initialization. Verifying observations are determined by taking the arithmetic mean of 36 rain gauges in the DFW area (Fig. 1)¹. A precipitation event requires that this

* Current affiliation: NOAA/NWS Forecast Office Jackson, Jackson, Kentucky.

Corresponding author address: Harold E. Brooks, NOAA/NSSL, 1313 Halley Circle, Norman, OK 73069.
E-mail: Harold.Brooks@noaa.gov

¹ We have chosen to use the arithmetic mean for simplicity. It makes dealing with missing data on any day particularly easy and, given the density of the gauges over the area, we feel that more complex techniques may not provide significant additional information.

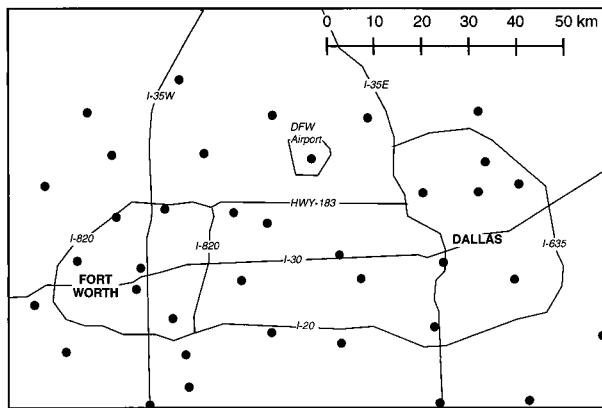


FIG. 1. Map of Dallas–Fort Worth area with rain gauge locations indicated by black dots. Major highways and DFW airport are shown.

areal mean be greater than or equal to 0.01 in. (0.25 mm). In this paper, we describe the development of the neural network forecasts, based initially on data from 1994 and 1995 with subsequent retraining after each month, and report on the verification of forecasts that subsequently were made from 1 March 1996 to 28 February 1998, a period covering 730 days. There are forecasts and data available from every day in the period.

2. The neural network

Neural networks provide a methodology for extracting patterns from noisy data. They have been applied to a wide variety of problems, including cloud classification (Bankert 1994) and tornado warnings (Marzban and Stumpf 1996) in a meteorological context. The advantages and disadvantages of neural networks in comparison to other statistical techniques for pattern extraction are discussed in Marzban and Stumpf (1996). More detail about the construction of neural networks can be found in Marzban and Stumpf (1996) and Müller and Reinhardt (1991) and references therein.

The standard procedure for use of a neural network involves “training” the network with a large sample of representative data. The network has some number of input and output “nodes” representing the predictor and predictand variables, respectively (Fig. 2). In between, there is a number of hidden nodes arranged in layers. The number of hidden nodes and layers is usually determined empirically to optimize performance for the particular situation. Each connection between nodes on a particular layer and the layer above it can be represented by a weight, ω_{ij} , that indicates the importance of that connection between the i th and j th nodes. The training phase of the neural network is designed to optimize the weights so that the mean-squared error of the output is minimized. For each node at a particular layer, the input node values from the previous layer are multiplied by the weight of the connections between the nodes and then all of the different connections are summed to pro-

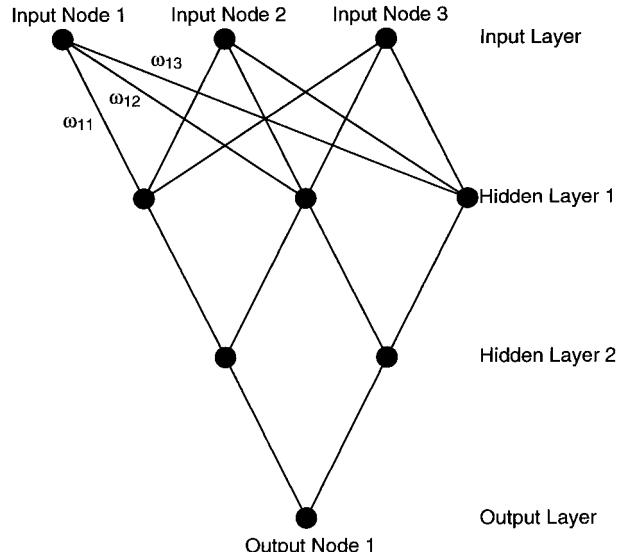


FIG. 2. A schematic of a neural network with three input nodes and one output node, with five hidden nodes on two hidden layers. Three of the 17 weights (links) are also shown. [After Marzban and Stumpf (1996).]

duce the value at that node. This process is repeated for all nodes and then for each layer. The network then can be used to make predictions based on new input values. In our application, we have created two networks: a QPF network for amount of precipitation and a PoP network for probability or confidence in the forecast.²

The QPF network was developed to predict 24-h areal average rainfall, not a point maximum rainfall. We compute 24-h (1200 to 1200 UTC) mean daily precipitation based on precipitation reports from 36 locations around the Dallas–Fort Worth Metroplex, covering approximately 5000 km². For the PoP network, “nonrain” days were identified as zeroes and “rain” days as ones. The resulting output of the PoP network was a number between 0 and 1, which could be multiplied by 100 to give a percentage. The networks were trained initially on data from all days from 1994 and 1995, regardless of the amount of precipitation that occurred.

Input data used to train the networks consist of 19 meteorological variables (Table 1), plus the observed rainfall. Input variables that can be derived from the Fort Worth, Texas, upper air soundings are taken from that source. Other values come from either the initial analysis or forecast fields from the Eta Model. Model data were entered from gridded model output read off of PC-GRIDDS graphics (Petersen 1992) and text values from NWS products (Eta “FOUS”) produced by

² The network was constructed using an “off-the-shelf” commercial neural network software package, Brainmaker, manufactured by California Scientific Software. Note that use and mention of a commercial product does not constitute endorsement by the United States government.

TABLE 1. Atmospheric variables used in neural network. Variables labeled "12-h forecast" are taken from the Eta Model. Other variables come from Fort Worth (FWD) sounding data.

Variable	Source
Precipitable water	Observed
Precipitable water 24-h change	Observed
K index	Observed
K index 24-h change	Observed
850-mb theta-e	Observed
850-mb theta-e advection	12-h forecast
850-mb wind speed and direction	Observed
700-mb theta-e	Observed
700-mb vertical velocity	12-h forecast
700-mb temp	Observed
700-mb temp advection	12-h forecast
500-mb wind speed and direction	Observed
500-mb vorticity	12-h forecast
850–300-mb thickness	12-h forecast
850–300-mb differential divergence	12-h forecast
1000–850-mb moisture divergence	12-h forecast
850-mb mixing ratio	12-h forecast
250-mb divergence	12-h forecast
700–500-mb lapse rate	Observed

NCEP. Various "test" networks were developed to determine the number and combination of meteorological variables that would provide the best possible "picture" of the available moisture, lift, and instability. Initial developmental networks were small and focused primarily on the key ingredients usually present for heavy to excessive rainfall (Junker 1992; Borneman and Kadin 1994). Additional variables were included later, based on experience from decision tree approaches to QPF (e.g., Johnson and Moser 1992). We eventually ended up with the 19 meteorological variables presently being used.

The networks were designed with three important features. The first was year-round applicability. Networks were developed separately for both the "warm" (April–October) and "cool" (November–March) seasons. The second feature was to expedite the entire process by running both the QPF and PoP networks simultaneously, and generate the output on one computer at the same time. Last, the networks are interactive; as such, one can change any of the variables desired and "rerun" the networks to accommodate any anticipated meteorological changes. This "interaction" is proving to be a great learning technique for forecasters by allowing them to carry out "what if" exercises. For example, if forecasters believe that the model is not handling moisture return well, they can vary the input value and see what effect that has on the forecast. If the timing of frontal passage is in question, values from another location on the other side of the front can be tested.

Numerous sensitivity analyses have been completed on the networks. The network is retrained and a sensitivity analysis is done every month to see if, and how, the importance of various parameters changes. The analysis ranks the variables by how well the individual parameter value correlates with the output forecast. The

TABLE 2. Most significant variables in neural network forecast for (a) warm season and (b) cool season.

(a) Warm season
Precipitable water
850–300-mb differential divergence
Precipitable water → 24-h change
K index → 24 h change
250-mb divergence
850-mb theta-e
700-mb vertical velocity
K index
1000–850-mb moisture divergence
850–300-mb thickness
(b) Cool season
Precipitable water
700-mb vertical velocity
Precipitable water → 24-h change
250-mb divergence
850–300-mb differential divergence
500-mb vorticity
K index → 24 h change
K index
850-mb theta-e
1000–850-mb moisture divergence

most important variables, based on an average of 12 months of analysis, are somewhat different for the warm and cool season networks, but precipitable water is the most important in each case (Table 2). The forecasts presented here are those done in real time, so that information from the previous month's retraining is included. These analyses change slightly after each network retraining, but the performance of the network was not significantly different in the two March–September periods after the initial training period of 2 yr. This indicates that little additional skill has been derived after the initial training period. It is possible that the retraining process may have preserved skill by continually "nudging" the network toward reality.

The training cycle is stopped by the operator. For the initial and retraining processes, the number of iterations used by the software was about 500 000. Ten percent of the data were set aside from each training cycle and used to test the network performance. The sensitivity analysis consists of correlating the value of the individual parameter with the output forecast of the network and seeing if, and how, the importance of the different parameters changes from month to month. Changes were small during the post-initial training phase and the listing of parameters in Table 2 is an average placement over a period of 12 months.

3. Results

a. Probability of precipitation (PoP)

An attributes diagram (Hsu and Murphy 1986) summarizes the performance of the probability of precipitation (PoP) forecasts graphically (Fig. 3). For convenience in presentation, we have combined the forecasts

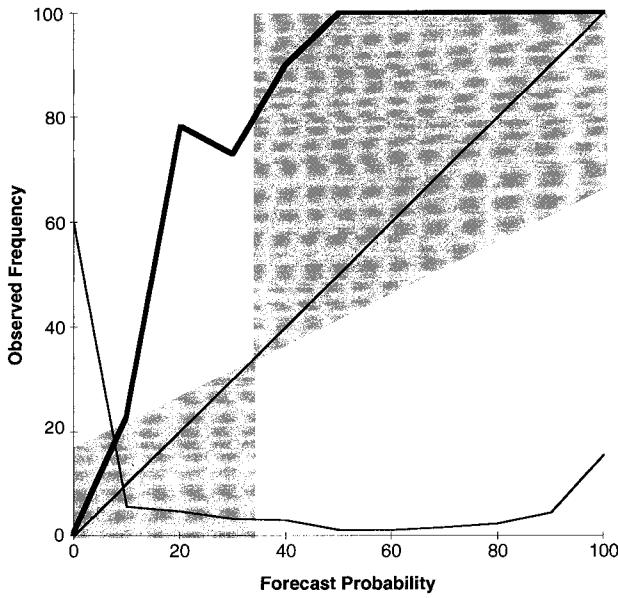


FIG. 3. Attributes diagram for PoP forecasts. Line at 45° is perfect reliability. Heavy line is observed frequency of precipitation given forecast probability and light line is frequency of use of forecast value. Shaded region is area where forecasts contribute positively to skill compared to Brier score of climatological forecast.

into 11 categories, rounded to the nearest 10%. It rained only once out of 436 forecasts in the 0% category and it rained on all 109 forecasts in the 100% category. The PoP was underforecast for all categories from 10% through 90%. In fact, it rained on all forecasts with a PoP greater than or equal to 38.5%. However, it is important to note the frequencies at which forecasts were issued. The two most common forecast categories were 0% (59.7% of forecasts) and 100% (15.2% of forecasts). Only 11.1% of the forecasts were in the 30%–80% categories. As a result, the forecasts had the desirable properties of sharpness and resolution (Murphy 1993). From this sample, it would appear that simple recalibration of the forecasts could be done to increase the forecast PoP to 100% when the system produces values above approximately 40%. The mean forecast PoP was 26.0% and the frequency of observed rain in the sample was 34.2%.

The sharpness of the forecasts and the nearly perfect performance for the forecasts of 0% and 100% lead to a skillful forecast system, with respect to climatology. The Brier score (BS; Brier 1950; Wilks 1995) averages the squared differences between forecasts and events:

$$BS = \frac{1}{n} \sum_{i=1}^n (f_i - x_i)^2, \quad (1)$$

where f_i is the i th forecast, x_i is the i th event ($x_i = 0\%$ for no rain and 100% for rain), and n is the number of forecasts. A skill score, SS, can be computed using the performance of some reference forecast system:

$$SS = 1 - \frac{BS}{BS_{ref}}, \quad (2)$$

where BS_{ref} is the Brier score of the reference system (Wilks 1995). Using the sample climatology (34.2%) of the probability of precipitation as the reference system, SS for the neural net forecast is 73.0%.

Unsurprisingly, the cool season PoP network (November–March) is somewhat more accurate than the warm season network relative to climatology (April–October). The SS for the cool season is 79.1% (sample climatology = 32.1%), while it is only 67.4% for the warm season (sample climatology = 35.5%). Although the warm season PoP network has better reliability in the 10%–30% PoP range, as evidenced by the plot of observed frequency of precipitation being closer to the perfect reliability line (Fig. 4), this is offset by the forecasts being noticeably less sharp than the cool season forecasts. Over 83% of the cool season forecasts have PoPs either less than 5% or greater than 95%. Only 68% of the warm season forecasts fall into those categories. The cool season PoP network tends to be particularly confident in no-rain situations. Over 65% (198 of 302) of the total forecasts have PoPs less than 5% and only one of those is associated with rain.

We have also evaluated the Brier score against baselines of persistence. Persistence, defined as “did it rain the day before?,” is a much worse forecast than sample climatology ($SS = -42\%$). It is possible that “was it raining at 1200 UTC?” would be better than standard persistence, but it is unlikely to be significantly better than sample climatology. Using the last 729 days of the period, there are 362 correct persistence “forecasts” of no rain, so that all of those have zero error points. There are 117 persistence forecasts of no rain that are associated with rain events. The simplest assumption is to assume that it was not raining at 1200 UTC in any of those, so that all of those have maximum error points (100% squared). There are also 117 events with no rain that are associated with rain on the previous day. Using the assumption that it was not raining at 1200 UTC on any of those, they all get no error points. There are 133 rain events with rain occurring on the previous day. In order for the “1200 UTC persistence” forecast to be more skillful than the sample climatology, the frequency of rain at 1200 UTC on those days would have to be greater than 35%. Given that the frequency of rain on a day, given that it rained *any time* the day before, is only 53%, it seems unlikely that the frequency of rain at 1200 is as high as 35%. Even if there are no errors in those cases, which seems exceedingly unlikely, it would only lower the Brier skill score improvement to 62%.

b. Quantitative precipitation forecasts

The distribution of forecast and observed precipitation amounts illustrates the accuracy of the forecasts

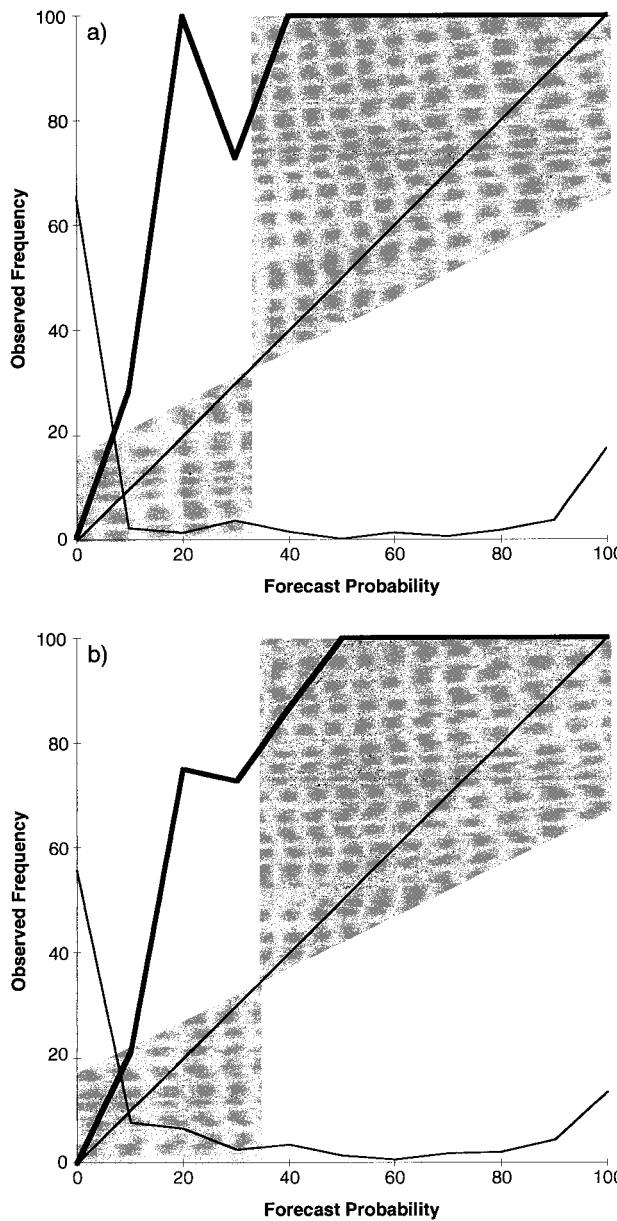


FIG. 4. Same as Fig. 3 except broken down by seasons: (a) cool season (November–March), (b) warm season (April–October).

(Fig. 5). The linear correlation coefficient between the forecast and observed amounts is 0.95. The least squares linear regression fit shows there is a small conditional bias, with precipitation amounts less than 0.13 in. (~ 3 mm) being overforecast and higher amounts being underforecast. Correlation for the cool season QPF is 0.95 and for the warm season it is 0.86, again illustrating the somewhat better performance of the networks in the cool season.

By considering QPF as a forecast of rain greater than a certain threshold, the problem can be broken down into a series of 2×2 contingency tables, each for a different threshold value. Thus, for example, we can

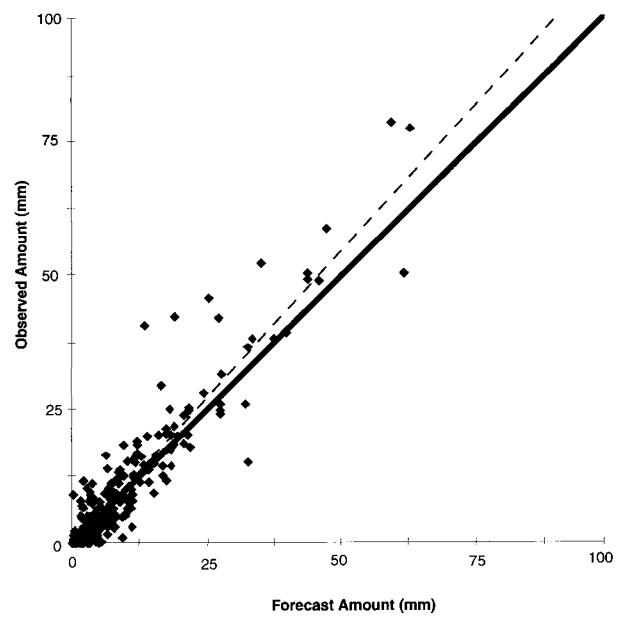


FIG. 5. Forecast and observed amounts for QPF in mm for all cases. Solid line at 45° is perfect reliability. Dashed line is least squares fit to data.

consider the performance of the system for forecasts and observations of rain greater than 0.5 in. (12.7 mm). Then, standard measures of performance of the system for 2×2 contingency tables can be computed for each threshold (see Table 3) (Doswell et al. 1990; Murphy 1996). The probability of detection of rain is above 0.75 for almost all threshold values up to 1 in. (25 mm) (Fig. 6). At the same time, the false alarm rate is quite low, below 0.20 for all thresholds between 0.08 in. (2 mm) and 1 in. (25 mm). As a result, the threat score or critical success index [originally known as Gilbert's ratio of success (Gilbert 1884)³] never goes below 0.63 for any

³ Murphy (1996) discussed the problem of multiple names and repeated "rediscovery" of the basic parameters of the 2×2 contingency table. We cite the original authors here. For more information on the nature of the scores, see Murphy (1996) and Doswell et al. (1990).

TABLE 3. A 2×2 contingency table for dichotomous forecasts of dichotomous events with various scores used in text defined.

Forecast	Events		Sum
	Y	N	
Y	a	b	F1
N	c	d	F0
Sum	X1	X0	N

POD (probability of detection) = $a/X1$. FAR (false alarm rate) = $b/F1$. TS (threat score) = $a/a + b + c$. ETS (equitable threat score) = $a - CH/a + b + c - CH$, where chance (CH) = $F1(X1)/N$. Heidke skill score (HSS) = $2(ad - bc)/X1(F0 + X0)F1$. True skill statistic (TSS) = $ad - bc/X1(X0)$.

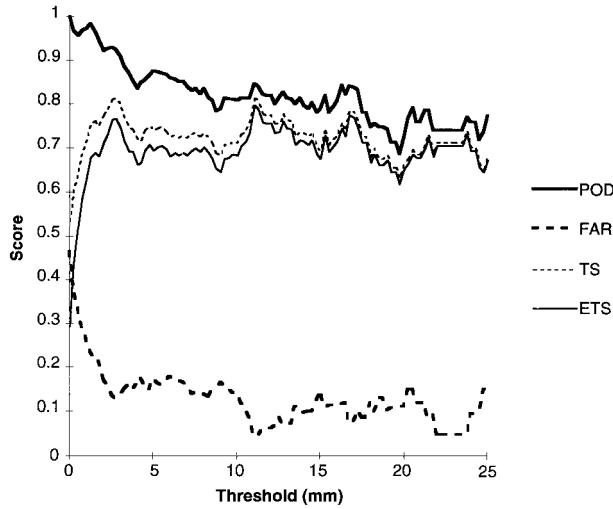


FIG. 6. Summary measures of QPF as functions of precipitation threshold (forecasts and observations greater than amount on abscissa). Plotted functions are probability of detection (POD), false alarm ratio (FAR), threat score (TS), and equitable threat score (ETS).

threshold between 0.03 in. (1 mm) and 1 in. (25 mm). The equitable threat score (ETS; Rogers et al. 1996), which is designed to penalize overforecasting events in comparison to the threat score, follows the threat score and is only slightly lower at the low threshold values [minimum value 0.63 between 0.05 in. (\sim 1 mm) and 1 in. (\sim 25 mm)], where some overforecasting exists. These values are quite high, in comparison to operational numerical weather prediction systems, where the ETS rarely exceeds 0.4⁴ (e.g., Rogers et al. 1996).

Skill measures, such as true skill statistic [TSS, also known as Kuiper's performance index, originally Peirce's "*i*" (Peirce 1884)] and the Heidke skill score [originally Doolittle's association ratio (Doolittle 1888)] are correspondingly high (Fig. 7). They calculate the performance of a forecasting system relative to how well a system based purely on chance would do. The TSS is greater than 0.68 for all values from 0.05 in. (\sim 1 mm) to 1 in. (25 mm) and Heidke skill score is above 0.77 for that range. These values indicate that the neural net is performing far better than chance for a wide range of values of precipitation.

It was noted earlier that rain occurred for all forecasts with a PoP greater than or equal to 38.5%. Recalibration could improve the reliability of the PoP forecasts, but the raw output PoP has value in the QPF problem; that is, the correlation between forecast and observed precipitation amounts increases with the PoP. For the 58

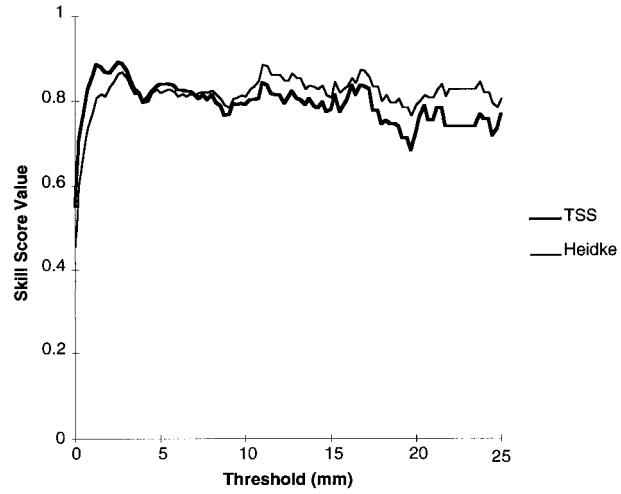


FIG. 7. As in Fig. 6 except for summary skill measures. Plotted functions are true skill statistic (TSS) and Heidke skill score (Heidke).

cases where rain occurred with forecast PoPs less than 38%, the correlation was 0.64. The correlation increases to 0.76 for the 52 cases of PoPs between 38.5% and 85%, and reaches 0.94 for the 140 cases with PoPs greater than 85%. Thus, increasing PoPs indicate greater confidence in the forecast amount of precipitation. This result is consistent with that found by Wilks (1990), using conditional probabilities for precipitation amount, given an MOS forecast PoP. It is possible that the consistency of the two networks is partially a result of the same input variables being used for each network, so that they respond to similar kinds of input.

Separating out those three classes also illustrates another feature of the system, the tendency to increase PoPs with higher forecast precipitation amount (Fig. 8). Forecast or observed precipitation exceeded 0.4 in. (10 mm) in only one case for the low PoP class and exceeded 0.9" (23 mm) in only one case for the medium PoP class. In the high PoP class, 12 of the 102 cases had both forecast and observed precipitation greater than 1 in. (25 mm). The network still produces high PoPs for some light rain events. Twenty-one of the 140 high PoP cases had QPF values of 0.1 in. (2.5 mm) or less. Corresponding with the tendency for heavy rainfall events to have higher PoPs, the average precipitation increases with increasing PoP (Table 4). The mean value of precipitation for the high PoP case is only slightly less than the maximum value in the low PoP case. As a result, forecasters using this system can calibrate their confidence in both ordinary and extreme rainfall events.

4. Closing remarks

The neural network has produced a remarkably good forecast of both the probability and amount of precipitation for the Dallas–Fort Worth area. The sharpness and accuracy of the high confidence PoPs (POP < 5%

⁴ Direct comparison of the ETS between the two methods should be done with caution, since the time periods of the two forecasts (operational NWP and the neural network) are different and, as such, it is possible that some of the difference is related to the time and space scales.

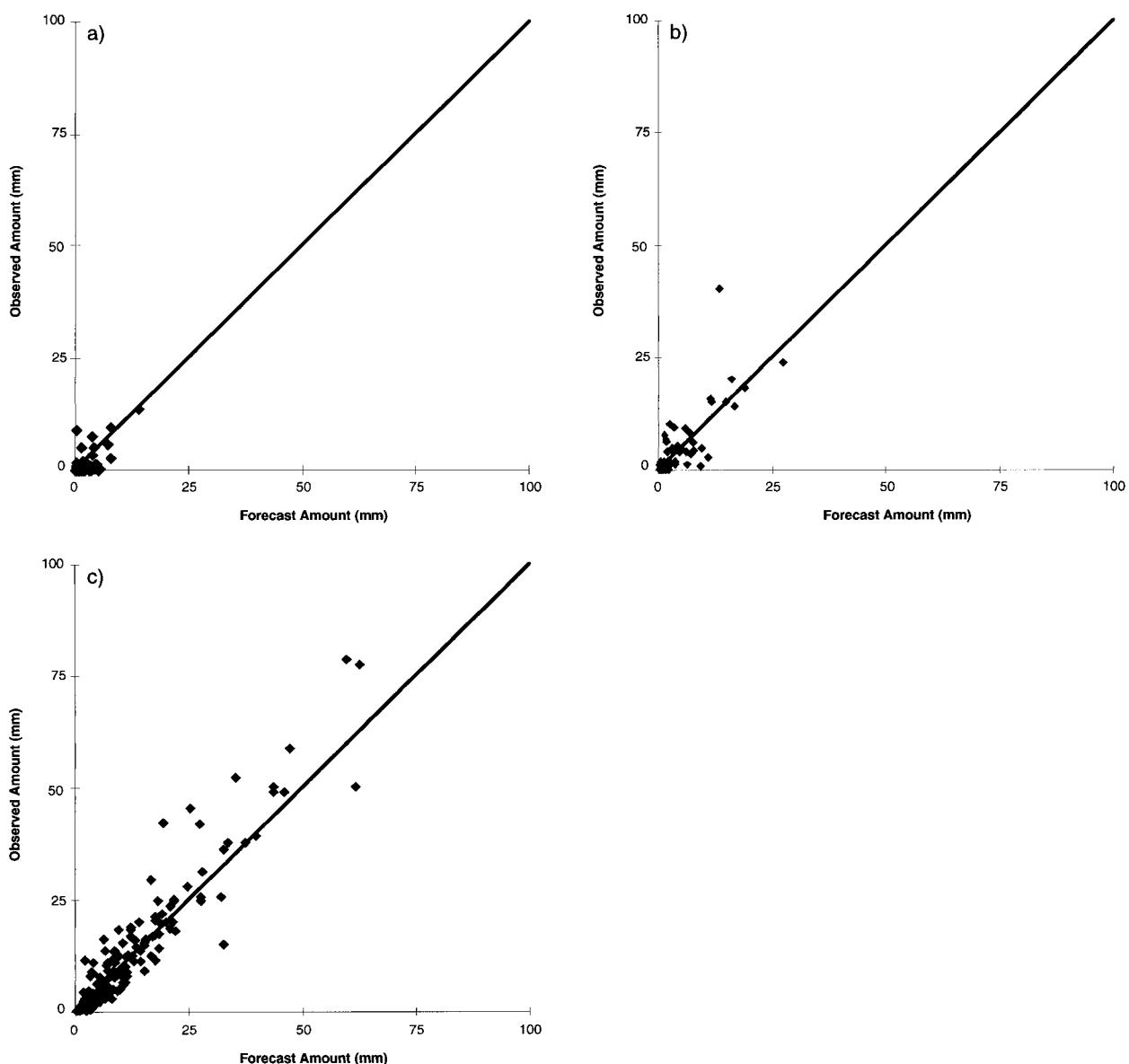


FIG. 8. As in Fig. 5 except forecasts are broken down by PoP: (a) $\text{PoP} < 38\%$, (b) $38\% < \text{PoP} \leq 85\%$, (c) $\text{PoP} > 85\%$.

and $\text{PoP} > 95\%$) are especially notable. Clearly, additional work needs to be done to test the network. In particular, the applicability to other locations is an important issue. However, the relative performance of the network could fall off significantly and still be a valuable forecast tool. Such techniques show promise for improving precipitation forecasts dramatically, partic-

ularly for applications requiring accurate areal precipitation forecasts. It is possible that they may be a useful tool for processing model output for other forecast problems as well.

Acknowledgments. We thank Messrs. Greg Story and Mike Thompson of the West Gulf River Forecast Center

TABLE 4. Summary measures of association and central values for QPF by PoP class for cases in which rain occurred. Number of cases in parentheses after correlation. Values under forecast and observed columns in in. with mm in parentheses.

PoP class	Correlation (n)	Forecast mean	Observed mean	Forecast median	Observed median
$\text{PoP} < 38\%$	0.64 (58)	0.10 (2.5)	0.07 (1.8)	0.07 (1.8)	0.03 (0.8)
$38\% < \text{PoP} \leq 85\%$	0.76 (52)	0.22 (5.6)	0.24 (6.0)	0.13 (3.3)	0.17 (4.3)
$85\% < \text{PoP}$	0.94 (140)	0.53 (13.5)	0.58 (14.7)	0.39 (9.8)	0.44 (11.2)

for their work in support of this project. Special thanks go to Mike Foster, science and operations officer at NWSFO Fort Worth for his efforts. Credit for the original concept behind developing these networks is given to Mr. Mike Gillispie (Gillispie 1993). In addition, comments from three anonymous reviewers were extremely helpful.

REFERENCES

- Bankert, R. L., 1994: Cloud classification of AVHRR imagery in maritime regions using a probabilistic neural network. *J. Appl. Meteor.*, **33**, 909–918.
- Borneman, R., and C. Kadin, 1994: Catalogue of heavy rainfall cases of six inches or more over the continental U.S. NOAA Tech. Rep. NESDIS 80, 64 pp. [Available from NOAA/NESDIS, 5200 Auth Rd., Suitland, MD 20746.]
- Brier, G. W., 1950: Verification of forecasts expressed in terms of probabilities. *Mon. Wea. Rev.*, **78**, 1–3.
- Doolittle, M. H., 1888: Association ratios. *Bull. Philos. Soc. Washington*, **7**, 122–127.
- Dowell, C. A., III, R. Davies-Jones, and D. L. Keller, 1990: On summary measures of skill in rare event forecasting based on contingency tables. *Wea. Forecasting*, **5**, 576–585.
- Gilbert, G. K., 1884: Finley's tornado predictions. *Amer. Meteor. J.*, **1**, 166–172.
- Gillispie, M., 1993: The use of neural networks for making quantitative precipitation forecasts. NWS Southern Region Technical Attachment SR/SSD 93-42, 21 pp. [Available from National Weather Service, Southern Region, 819 Taylor St., Room 10A26, Fort Worth, TX 76102.]
- Glahn, H. R., and D. A. Lowry, 1972: The use of Model Output Statistics (MOS) in objective weather forecasting. *J. Appl. Meteor.*, **11**, 1203–1211.
- Hsu, W.-R., and A. H. Murphy, 1986: The attributes diagram: A geometrical framework for assessing the quality of probability forecasts. *Int. J. Forecasting*, **2**, 285–293.
- Johnson, G.A., and J. Moser, 1992: A decision tree for forecasting heavy rains from mid-latitude synoptic patterns in Louisiana generally from late fall through spring. NOAA Tech. Memo. NWS ER-87, 24 pp. [Available from National Weather Service, Eastern Region, 630 Johnson Ave., Bohemia, NY 11716.]
- Junker, N. W., 1992: *Heavy Rain Forecasting Manual*. National Weather Service, 91 pp. [Available from National Weather Service Training Center, 617 Hardesty, Kansas City, MO 64124-3032.]
- Klein, W. H., B. M. Lewis, and I. Enger, 1959: Objective prediction of five-day mean temperature during winter. *J. Meteor.*, **16**, 672–682.
- 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.
- Müller, B., and J. Reinhardt, 1991: *Neural Networks: An Introduction*. Vol. 2, The Physics of Neural Networks Series, Springer-Verlag, 266 pp.
- Murphy, A. H., 1993: What is a “good” forecast? An essay on the nature of goodness in weather forecasting. *Wea. Forecasting*, **8**, 281–293.
- , 1996: The Finley affair: A signal event in forecast verification. *Wea. Forecasting*, **11**, 3–20.
- Peirce, C. S., 1884: The numerical measure of success of predictions. *Science*, **4**, 453–454.
- Petersen, R. A., 1992: A PC-based system for the display of gridded WAFFS data. *Proc. WMO Tech. Conf. on Tropical Aeronautical Meteorology (TECTAM-92)*, Geneva, Switzerland, World Meteorological Organization, 1–6.
- Roebber, P. J., and L. F. Bosart, 1996: The complex relationship between forecast skill and forecast value: A real-world analysis. *Wea. Forecasting*, **11**, 544–559.
- Rogers, E., T. L. Black, D. G. Deaven, G. J. DiMego, Q. Zhao, M. Baldwin, N. W. Junker, and Y. Lin, 1996: Changes to the operational “early” Eta analysis/forecast system at the National Centers for Environmental Prediction. *Wea. Forecasting*, **11**, 391–413.
- Wilks, D. S., 1990: Probabilistic quantitative precipitation forecasts derived from PoPs and conditional precipitation amount climatologies. *Mon. Wea. Rev.*, **118**, 874–882.
- , 1995: *Statistical Methods in the Atmospheric Sciences*. Academic Press, 467 pp.