GEOFIZIKA VOL. 25 No. 1 2008

Original scientific paper
UDC 551.501.777

Improving the quality of INSAT derived quantitative precipitation estimates using an neural network method

Sankar Nath, A. K. Mitra and S. K. Roy Bhowmik

Indian Meteorological Department, New Delhi, India

Received 15 May 2007, in final form 17 April 2008

In this paper an Artificial Neural Network (NN) approach has been applied to improve the quality of the INSAT derived sub-division quantitative precipitation estimates (IMD-QPE) over the Indian region for the summer monsoon season. Data for the years 2001, 2003 and 2004 have been used as the training sample. The method is tested with independent sample data for the year 2005. For the subdivisions over the domains of high orographic and monsoon low pressure system, where very rainfall occasionally occurs, different network architectures are applied to minimize the IMD-QPE errors. An inter-comparison between NNQPE (NN model output IMD-QPE), IMD-QPE and actual rainfall indicates that the pattern of NNQPE is closer to the observed rainfall distribution. The weekly mean absolute error of IMD-QPE with respect to observed rainfall, which ranges between 10–99 mm, becomes 4–70 mm in case of NNQPE. The performance statistics shows that the proposed NN model is able to produce better IMD-QPE with higher skill score and correlation co-efficient with respect to observation in most of the sub-divisions. The method is found to be promising for operational application.

Keywords: artificial neural network, INSAT derived QPE, weekly subdivision rainfall, orographic, skill score

1. Introduction

Satellite observations of infrared (IR) and microwave radiances have been used successfully to retrieve precipitation information over many parts of the globe. Satellite derived Quantitative Precipitation Estimates (QPE) promise to provide very useful input for the initialization and validation of Numerical Weather Prediction (NWP) models (e.g. Krishnamurti et al., 1995; Roy Bhowmik and Prasad, 2001). But the accuracy of the product is limited by the relatively indirect relationship between the radiances measured from the satellites and the rate of precipitation at the ground. It has been well established by various studies (Janowiak, 1992; Ebert and Marshall, 1995; Roy Bhowmik and Sud, 2003) that the accuracy of the satellite-derived QPE is limited due to many technical and scientific issues. There are various key factors like topography, prevailing synoptic situation and its interactions with mesoscale systems, stratiform precipitation and calibration issues which create uncertainties in deriving QPE from satellite data sets.

The study of Roy Bhowmik and Sud (2003) showed that the very heavy rainfall events over most part of Indian monsoon region is significantly under-estimated due to the fact that rainfall rate constant (72 mm/day) as introduced by Arkin et al. (1989) is unrealistically low in the context of intense mesoscale convective rainfall in association with monsoon depression or cyclone.

One way to improve the accuracy of rainfall estimates for the Indian monsoon would be to use a statistical technique to post-process the satellite estimates based on their error characteristics. The roots of this approach are found in the work of Glahn and Lowry (1972) who used regression techniques to derive corrections for numerical weather prediction fields to make them more similar to the observed values. However, the relationships between model forecasts and observations are typically nonlinear and thus are only partially captured by linear approaches such as regression. During recent years, the technique of Artificial Neural Networks (NN) has drawn considerable attention for handling these kinds of problem. Neural Networks have been widely 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; Koizumi, 1999), and even long-range monsoon precipitation (Wu et al., 2001). It has a strong potential for pattern reorganization and signal processing problems and also has the ability to predict future values of a time series from past values.

The tool, which is one of the most effective methods for pattern reorganization and signal processing problems, has been applied in this study to improve Indian Geostationary Satellite (INSAT) derived Quantitative Precipitation Estimates (IMD-QPE). The strategy for applying the NN technique involves two phases. The first phase, known as training period, utilizes IMD-QPE and observed rainfall fields to derive statistics and the second phase utilizes the IMD-QPE and aforementioned statistics to obtain the modified IMD-QPE. Details of the NN technique are available in the literature (Shi, 2001; Bishop, 1995).

2. Data and methodology

IMD-QPE from the Indian Geostationary Satellite (INSAT) are derived at the grid resolution of $1° \times 1°$ lat./long. following the algorithm as described by Arkin et. al. (1989). The approach is identical with one, which was used to obtain such estimates for GOES (Arkin and Meisner, 1987). According to this algorithm the rainfall estimates is given by:

$$R = 3 \times f \times h \tag{1}$$

Where R is aerial rainfall estimates in mm, f is the fraction of IR pixels within the square grid with a temperature less than 235 K and total number of pixels within the square grid and h is time in hour.

The constant rain-rate of 3 mm/hour provides the best correlation of the regression relation (Richard and Arkin, 1981).

The subdivision weekly IMD-QPE prepared by the Satellite Division of the India Meteorological Department (IMD) are used in this study. In order to get the corresponding data sets of ground truth, the subdivision weekly rainfall data obtained from Hydrology Division have been used as observed data set.

In general, a neural network is a computer model composed of individual processing elements called neurons. The neurons are connected by links that have weights associated with them. A neural network consists of multiple layers of neurons interconnected with neurons in other layers. These layers are referred to as the input layer, hidden layer(s) and output layer. The inputs and the interconnection weights are processed by a weighted summation function to produce a sum that is passed to a transfer function. The output of the transfer function is the output of the neurons. A neural network is trained with input and output pattern examples. It then constructs a nonlinear numerical model of a physical process in terms of network parameters, Shi (2001).

The NN technique proposed here is based on the three-layer feed forward back propagation (Rumelhart et. al, 1986; Bishop, 1995; Haykin, 1999) using the basic Levenberg-Marquardt training algorithm as illustrated in Figure 1(a,b). This is a commonly used method for the minimization of mean square error criteria (Maqsood et. al., 2002; Marquardt, 1963). The layers 1, 2, and 3 represent the input layers, hidden layers and the output layer respectively. In Figure 1(a), for the subdivisions which are not affected by very heavy rainfall due to orographic or monsoon synoptic system, the neuron of the input layer is represented by corresponding weekly IMD-QPE i.e. W(QPE). The hidden layer is determined during network architecture design and adjusted to achieve best network performance. For this work, 10 hidden neurons are found to produce the best result. Finally, the neuron in output layer is the corresponding improved weekly rainfall W(RF). In Figure 1(b), for the subdivisions which are affected by heavy to very heavy rainfall due to orographic or monsoon synoptic system, the neuron of the inputs layer is represented by three consecutive weeks IMD-QPE i.e., W(QPE), W-1(QPE) and W-2(QPE). The hidden layers are determined during network architecture design and adjusted to achieve the best network performance. In this case also, 10 hidden-layer neurons are found to produce the best performance. The neuron in the output layer is the corresponding improved weekly rainfall W(RF).

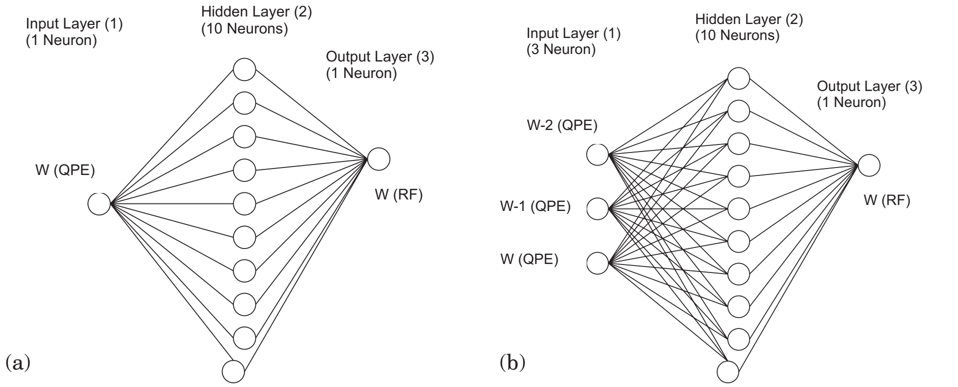

Figure 1. Neural Network architecture for (a) subdivisions where monsoon system is not dominant (b) subdivisions where monsoon system and orography are dominant. Weekly IMD-QPE is noted as W(QPE), weekly rainfall as W(RF), previous week's weekly IMD-QPE as W-1(QPE) and previous to previous week's weekly IMD-QPE as W-2(QPE).

Figure 2 displays geographical locations of the subdivisions which occasionally receive heavy to very heavy rainfall due to synoptic scale monsoon circulations or orograhic during summer monsoon season. This also presents the subdivision mean weekly QPE for the monsoon season (1 June to 30 September) of 2005. These subdivisions are: Coastal Karnataka, Konkan and Goa. In

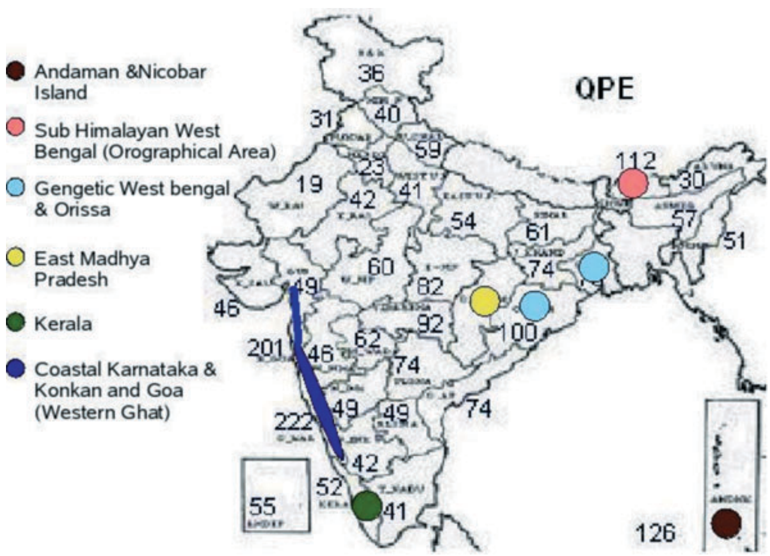

Figure 2. The geographical locations of some important subdivisions and the weekly mean subdivision IMD-QPE for the period 01 June to 30 September 2005.

Table 1. An intercomparison of mean rainfall for the test sample with two different network architectures for the subdivisions where monsoon system and orography are dominant.

Subdivision	IMD-QPE	IMD Rainfall	Single input NN	Multiple inputs (three consecutive weeks) NN
Konkan and Goa	201	182	198	193
Costal Karnataka	222	178	215	145
Sub-Himalaya of West Bengal and Sikkim	112	99	107	93
Gangetic West Bengal	78	67	75	66
Orissa	100	57	87	56

the east coast of India, Gangetic West Bengal, Orissa and neighborhood in the east central India and Sub-Himalaya of West Bengal and Sikkim along the foot hills of the Himalayas. For these subdivisions a number of inputs were compared in constructing the network architectures and it is found that using a previous three consecutive weekly IMD-QPE as inputs gives the optimum network performance for the type of data we used in this study. Table 1 shows an inter-comparison of mean weekly rainfall for these subdivisions based on two different NN architectures against observed weekly rainfall and IMD-QPE. The inter-comparison reveals that the results of the new NN architecture (input layer with three consecutive weeks subdivision rainfall) are closer to the corresponding weekly subdivision rainfall. The inter-comparison also reveals that weekly IMD-QPE for these subdivisions are always higher compare to the observed weekly subdivision rainfall. Over estimation of weekly rainfall for these subdivisions is due to the fact that IMD QPE, though fails to capture mesoscale heavy to very heavy rainfall features, it produces unrealistic rainfall belt over a larger area around the monsoon system, resulting higher subdivision rainfall (Roy Bhowmik et al., 2007). Krishnamurti and Bhalme (1976) showed that monsoon system of Indian summer monsoon follows quasi-biweekly oscillation. This justifies the new NN architecture (input layer with three consecutive weeks subdivision rainfall) designed for the sub-divisions where monsoon systems are dominant.

During the network design, within the training set we used some data for validation which also dependent to prevent overfitting the network so that it will generalize well on independent data. A number of transfer functions were compared in constructing the network architectures and it was found that using a tan-sigmoid transfer function to propagate to the hidden layers and a linear transfer function to propagate to the output layer in a three-layer backpropagation architecture gives the optimum network performance for the type of data we used in this study.

The weekly IMD-QPE from satellite observations of the southwest monsoon seasons of 2001, 2003 and 2004 of a subdivision is taken as signal to the input layer neurons. Training was done with the help of corresponding weekly rainfall data of that particular subdivision and then the results were tested with another set of sample data for the year 2005.

The process of updating of weights is iterated until the error between the derived and actual output becomes less than a predefined small value 10^{-6}. For better performance of network trial and error strategy is used to determine this predefined small value against the training data set.

Since INSAT was defunct during 2002, the year 2002 has been skipped in this study. This process was applied to all the meteorological subdivisions of India.

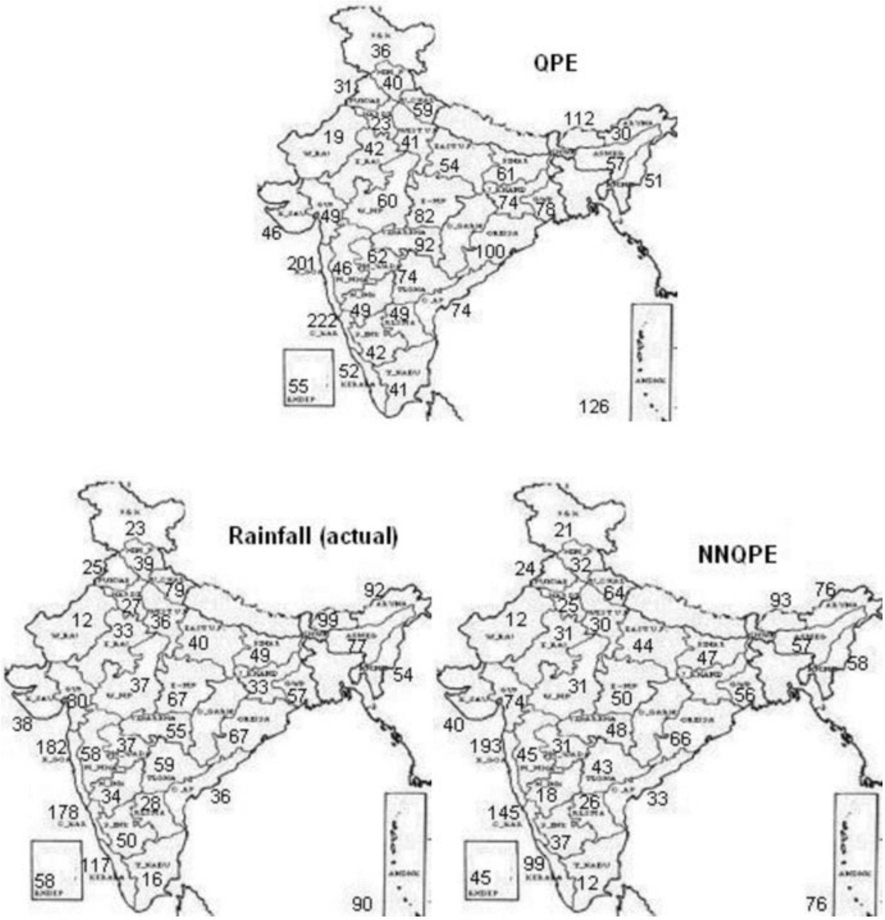

Figure 3. An inter-comparison of mean weekly rainfall on the basis of IMD-QPE, rain-gauge rainfall and the improve QPE by neural network based on the independent data sample for the period from 01 June to 30 September 2005.

3. Results and discussion

3.1. Performance statistics

The mean patterns of subdivisional weekly rainfall of the summer monsoon (1 June to 30 September) 2005 based on IMD-QPE, observations and neural network for an independent test sample are shown in Figure 3. The result shows that the mean pattern of IMD-QPE does not agree with the observations for most of the subdivisions both for the training (not shown) as well as for the test data period. The maximum over estimation occurs over Andaman and Nicobar Islands, Kerala, Coastal Karnataka and East Madhya Pradesh. But with NNQPE the patterns becomes closer to the observations. The geographical locations of these subdivisions are as shown in Figure 2. The weekly mean range of IMD-QPE, actual rainfall observations and NNQPE for the training set (not shown) are respectively 28–230 mm, 13–159 mm and 15–184 mm. The corresponding ranges for the test data set are 19–222 mm, 12–178 mm and 12–193 mm respectively.

Figure 4 represents the mean absolute errors of IMD-QPE and NNQPE with respect to the rainfall observation. The inter-comparison reveals that a IMD-QPE error decreases considerably with the NNQPE model. The weekly mean absolute error of IMD-QPE which ranges between 10–99 mm becomes 4–70 mm in case of NNQPE.

Figure 5 shows the overall pattern of co-relation coefficient (CC) of IMD-QPE and NNQPE with respect to the observed rainfall. The CC of NNQPE is found to be higher over the most subdivisions compare to that of IMD-QPE.

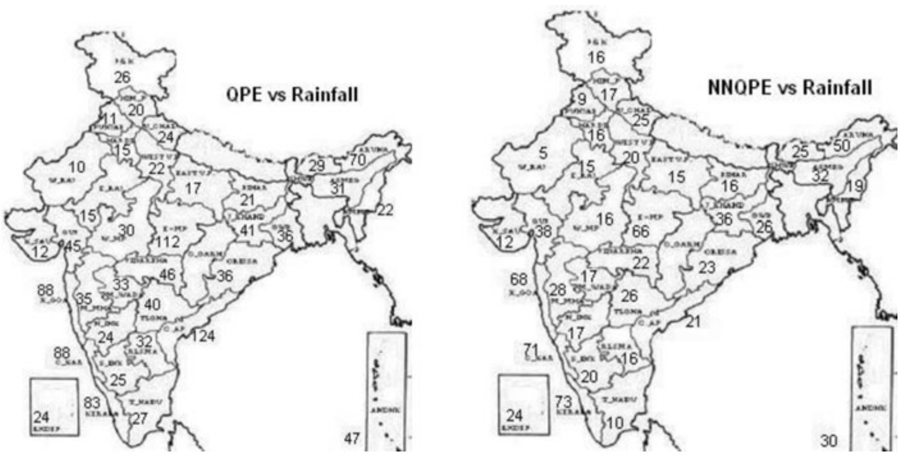

Figure 4. An inter-comparison mean absolute errors between IMD-QPE and neural network QPE based on the independent data sample of 01 June to 30 September 2005.

3.2 Skill score

For the inter-comparison of performance skill, the skill scores are obtained as

$$\text{Skill Score} = \left(1-\frac{MSE^2_{NNQPE}}{MSE^2_{QPE}}\right) \text{‰}\ 100\% \qquad (2)$$

Where MSE_{NNQPE} and MSE_{QPE} stand for MSE (Mean square error) of the neural network model and the satellite precipitation estimates respectively.

The performance skill of the NNQPE model for the testing data of the meteorological subdivisions has been shown in Figure 6. A positive value of the skill score stands for a better performance of the model over satellite precipitation estimates, while a negative value of the skill score indicates that the model does not have skill to match the satellite precipitation estimates. Though some of the subdivisions have smaller positive skills values, the figure clearly indicates that the neural network model has overall positive skill and performs better than the conventional satellite precipitation estimates with training (not shown) as well as independent data sets.

The error analysis in Figure 7 indicates that the NNQPE generally has a lower percentage of large errors (≥ 90 mm) and a higher percentage of smaller errors (≤ 30 mm) than the uncorrected IMD-QPE. This is another indication that the NN post-processing of the IMD-QPE improves its accuracy.

4. Conclusions

In this study a NN approach have been used to improve the quality of INSAT derived quantitative precipitation estimates over Indian region for the southwest monsoon season. The study shows that the neural network model is able to improve the quality of sub divisional weekly IMD-QPE. The weekly

Figure 5. An inter-comparison of CC of observed subdivision weekly rainfall with IMD-QPE and neural network QPE for the independent data sample of 01 June to 30 September 2005.

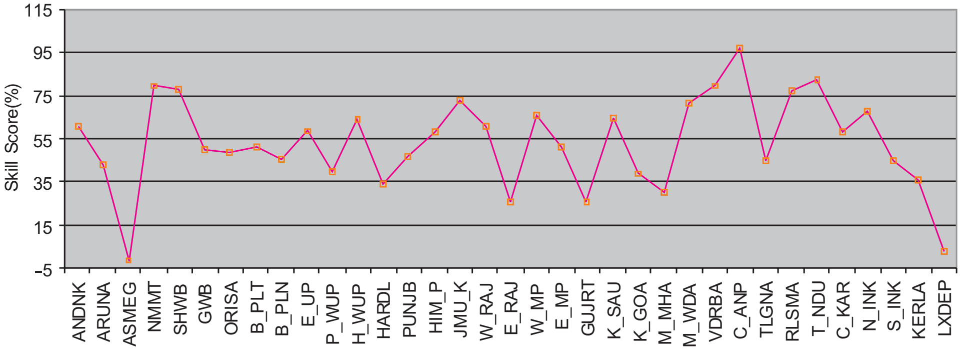

Figure 6. The performance skill of the NNQPE model for the period from 01 June to 30 September 2005.

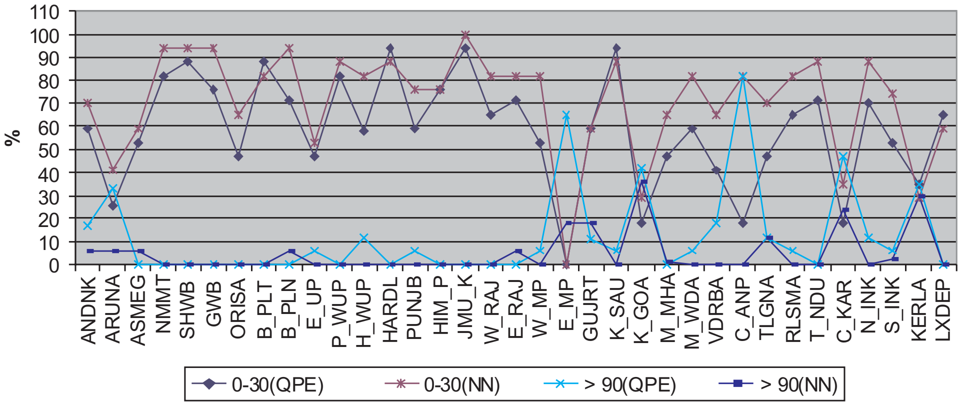

Figure 7. The quantitative error range of IMD-QPE and NN model for the period from 01 June to 30 September 2005.

mean absolute error of IMD-QPE which ranges between 10–99 mm reduces to 4–70 mm in case of an NNQPE. This indicates that the neural network approach reduces the IMD-QPE errors and provides a pattern which is closer to the observed rainfall pattern.

Orographically-enhanched heavy rainfall pattern is difficult to capture by IMD-QPE because of the low threshold being used for cloud top temperature. However, considerable improvement in the quality of IMD-QPE over the subdivisions of very heavy rainfalls (orographic and monsoon low pressure system) are also noticed from the use of three consecutive weeks IMD-QPE as input. The inter-comparison of skill scores confirms the better performance and effectiveness of the proposed NN model. The results have clearly indicated the feasibility of our approach. The approach shows encouraging results to reduce errors of satellite precipitation estimates.

Further improvement can be achieved through the use of high quality daily rainfall analysis and suitable adjustment between cloud top temperature threshold and rain-rate. An ensemble neural network can be developed to perform multi-class classification with location and elevation information to improve the orographic-affected areas.

Acknowledgments – Authors are grateful to the Director General of Meteorology for providing all facilities to carry-out this work. Authors gratefully acknowledge the supports extended by Satellite Meteorology, Hydrology and Telecom Division of the India Meteorological Department, New Delhi. Authors are thankful to the anonymous reviewers for their valuable suggestions and comments to improve the quality of the paper.

References

Arkin, P. A. and Meisner, B. N. (1987): The relationship between large-scale convective rainfall and cloud top temperature over the western hemisphere during 1982–84, *Mon. Weather Rev.*, **115**, 51–74.

Arkin, P. A., Rao, A. V. R. K. and Kelkar R. R. (1989): Large-scale precipitation and out-going long wave radiation from INSAT-1B during 1986 south-west Monsoon season, *J. Climate,* **2**, 619–628.

Bishop, C. M. (1995): *Neural Networks for Pattern Recognition*. Oxford University Press, Oxford, UK.

Ebert, E. and Marshall, J. L. (1995): An evaluation of infrared satellite estimates techniques over Australia, *Aust. Meteorol. Mag*., **44**, 177–190.

Glahn, H. R. and Lowry, D. A. (1972): The use of model output statistics (MOS) in objective weather forecasting, *J. Appl. Meteorol.*, **11**, 1203–1211.

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

Haykin, S. (1999): *Neural Networks. A Comprehensive foundation.* Prentice-Hall, Upper saddle River, New Jersey.

Janowiak, D. E. (1992): Tropical rainfall: A comparison of satellite derived rainfall estimates with QPE precipitation forecast, climatologies and observations, *Mon. Weather Rev.,* **120**, 448–462.

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.

Krishnamurti, T. N., Roy Bhowmik, S. K., Oosterhof, D., Rohaly, G. and Surgi, N. (1995): Meso-scale signatures in the tropics generated by physical initialization, *Mon. Weather Rev*., **123,** 2771–2790.

Krishnamurti, T. N. and Bhalme, H. N. (1976): Oscillation of monsoon system, Part I: Observational aspects, *J. Atmos. Sci*., **33,** 1973–1954.

Maqsood, I., Riaz Khan, M. and Abraham, A. (2002): Intelligent weather monitoring systems using connectionist models, *Neural, Parallel & Scientific Computations,* **10** (2), 157–178.

Marquardt, D. W. (1963): An algorithm for least-squares estimation of non-linear parameters, *SIAM J. Appl. math.,* **11** (2), 431–441.

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

Marzban, C. and Stumpf, G. J. (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.

Richard, F. and Arkin, P. (1981): On the relationship between satellite observed cloud cover and precipitation, *Mon. Weather Rev*. **109**, 1081–1093.

Roy Bhowmik, S. K. and Prasad, K. (2001): Some characteristics of limited area model precipitation forecast of Indian monsoon and evaluation of associated flow features, *Meteorol. Atmos. Phys*., **76**, 222–236.

Roy Bhowmik, S. K. and Sud, A. M. (2003): Seasonal and spatial characteristic of QPE errors over Indian region, *Mausam,* **54**, 653–658.

Roy Bhowmik, S. K. and Das, Ananda K. (2007): Rainfall Analysis for Indian region over using the merged rain gauge observations and satellite estimates: evaluation of monsoon rainfall features, *J. Earth Syst. Sci*., **116**,187–198.

Rumelhart, D. E., Hinton, G. E. and Williams, R. J. (1986): Learning internal representations by error propagation. In Rumelhart, D. E., McClelland, J. L. and the PDP Research Group (Eds): *Parallel distributed processing. Explorations in the microstructure of cognition,* Volume 1: Foundations. The MIT Press, Cambridge, MA, 318–362.

Shi, L. (2001): Retrival of temperature profiles from AMSU-A measurement using neural network approach, *J. Atmos. Ocean. Tech*., **18**, 340–347.

Xiaodan, W., Hongxing, C., Flitman, A., Fengying, W. and Guolin, F. (2001): Forecasting monsoon precipitation using artificial neural networks, *Adv. Atmos. Sci.*, **18**, 951–958.

SAŽETAK

Poboljšanje kvalitete INSAT izvedenih procjena količinske oborine korištenjem metode neuronske mreže

Sankar Nath, A. K Mitra i S. K. Roy Bhowmik

U ovoj studiji se koristi umjetna neuronska mreža (NN) za poboljšanje INSAT izvedenih podrazreda procjena količinske oborine (IMD-QPE) nad područjem Indije tijekom sezone ljetnog monsuna. Korišteni su podaci za 2001., 2003. i 2004. godinu kao probni uzorak. Metoda se testira na nezavisnom skupu podataka iz 2005. Za podrazrede nad domenama visokog orografskog tlaka i monsunskog niskog tlaka gdje se opažaju vrlo jake kiše, primijenila se različita mrežna arhitektura radi minimaliziranja IMD-QPE grešaka. Usporedba između NNQPE (izlaz IMD-QPE NN modela), IMD-QPE i stvarne oborine upućuje da je uzorak NNQPE bliži opaženoj distribuciji oborine. Tjedna srednja apsolutna pogreška IMD-QPE u odnosu na opaženu oborinu, koja se nalazi unutar intervala od 10–99 mm, postaje 4–70 mm u slučaju NNQPE. Statistika je pokazala da je predloženi NN model sposoban bolje reproducirati IMD-QPE s boljim pokazateljima uspješnosti i koeficijentima korelacije u odnosu na opažanja u većini podrazreda. Pokazano je da se metoda može uspješno primijeniti u svakodnevnoj praksi.

Ključne riječi: umjetna neuronska mreža, INSAT izveden QPE, tjedna oborina, orografija, pokazatelj uspješnosti

Correspondence author's address: Dr. S. K. Roy Bhowmik, Director (NWP), India Meteorological Department, Lodi Road, New Delhi–110003, India, e-mail: skrb@imdmail.gov.in