Romanian Reports in Physics, Vol. 62, No. 2, P. 405–413, 2010

THE JULY 2008 RAINFALL ESTIMATION FROM BARNOVA WSR-98 D RADAR USING ARTIFICIAL NEURAL NETWORK*

GINA TIRON[1], STELUȚA GOSAV[2]

[1]*Faculty of Physics, "Al.I.Cuza'' University of Iasi, 11 Carol I Boulevard, Romania, E-mail: tiron.gina@.yahoo.com*

[2]*Physics Department, Faculty of Science, "Dunărea de Jos" University of Galati, Domnească St.47, Romania, E-mail: stelagosav@yahoo.com*

(Received July 30, 2009)

*Abstract***.** In the last years, the artificial neural networks (ANN) have proved an attractive approach to non-linear regression problems arising in environmental modelling, such as short-term forecasting of atmospheric pollutant concentrations, rainfall run-off modelling and precipitation nowcasting using radar, satellite or meteorological data. The term nowcasting reflects the need of timely and accurate predictions of risk situations related to the development of severe meteorological events. The objective of this work is the very short term prediction of the rainfall field from radar data based on feed forward neural network approach. The radar dates used in this study were measured by the WSR-98D Doppler radar in North-East of Romania. The reflectivity data sets extend over July 2008. The ANN system with reflectivity values as input variables was trained to predict the rain rate on the ground. The output vector consists of one variable namely the rain rate measured by a rain gauge on ground level. The two available rain gauges provided the rain rate in millimetres every one hour. Data-preprocessing or the selection of input variables was performed when necessary. The efficiency of ANN network in the estimation of the rain rate on the ground in comparison with that supplied by the weather radar is evaluated.

Key words: weather radar, artificial neural network, precipitation, reflectivity factor.

1. INTRODUCTION

Rainfall forecasting is one of the most difficult but very important tools in meteorology and hydrology. Rainfall estimation on the ground based on radar measurements is a challenging problem due to space-time variability of rainfall fields. The rainfall on the ground is dependent on the four-dimensional (space-time) structure of the precipitation aloft. Recent research has shown that neural network techniques can be successfully used for ground rainfall estimation from weather radar data [–15]. In addition, in the last years Artificial Neural Networks (ANNs) have been used in various scientific fields such as environmental

* Paper presented at the Annual Scientific Session of Faculty of Physics, University of Bucharest, June 5, 2009, Bucharest-Magurele, Romania.

protection [6, 7], spectroscopy [8], wind meteorology and precipitation estimations [9, 10], etc.

This paper describes an approach meant to estimate ground rainfall based on a feed-forward neural network technique to model the relationship between weather radar data aloft and rain gauge measurements on the ground. The ANN system accepts radar data as an input and is trained to predict the rain rate as measured by the rain gauge. Xiao and Chandrasekar [11] applied a backpropagation neural network for rainfall estimation from radar data. Liu *et all.* [12], Xu and Chandrasekar [1] developed a radial basis function neural network to estimate ground rainfall.

The ANN technique includes two stages, namely, 1) the training and 2) the validation stage. In the training stage, the neural network learns the potential relationship between the rainfall rate and the radar measurements from a training dataset. When a radar measurements set is applied to neural network, the network yields a rainfall rate (RR) estimated as output. This output is compared with the rain gauge measurement, and their difference or the error is propagated back to adjust the parameters of the network. This learning process is continued until the network converges. Once the training process is complete, a relationship between the rainfall rate and the radar measurements is established and the network is ready for operation [12]. In addition, we have achieved a sensitivity analysis of the input weather radar data for the artificial neural network.

2. RAIN GAUGE. WEATHER RADAR

2.1. DATA AVAILABLE

For this study rain-gauge and radar data from the province of Moldova, Romania, were available. The data sets extend over a one year period. The rain gauges work on the tipping bucket principle with a resolution of 0.2 mm. Their locations are showed in Fig. 1, respectively Botosani and Vaslui. Their temporal resolution is 60 minutes.

Reflectivity measurements are gained from the Doppler WSR-98D weather radar station at Barnova, Iasi (Fig. 1). Radar is a high-resolution S – band weather radar situated on an altitude of 396 m above m.s.l. The distance between the radar station and the rain gauge in Botosani is about 106 Km, while the one between the radar station and the rain gauge in Vaslui is about 43 Km.

The radar has the following specifications:

- Time interval between measurements: 6 minutes;
- Beamwidth: 1^0;
- Minimum elevation angle: 0.5 gr.;
- Spatial resolution of the volume element: 1 km^3 (1 km × 1 km × 1 km);

- Resolution in measured reflectivity: 14 levels of reflectivity;
- Instrumented range: 230 km.

The measuring unit of the radar is the reflectivity factor *Z*.

The reflectivity factor *Z* is used to estimate the rain rate using relationships of the form:

$$Z = a \cdot R^{b}, \tag{1}$$

where *a* and *b* empirical parameters; *R* – rain rate in mm/h; *Z* – reflectivity factor in $mm^6 \cdot m^{-3}$.

In this paper, we have used the reflectivity measurements at four elevations: 0.5, 1.5, 2.4 and 3.4. For each elevation we have obtained nine adjacent measurements with the rain gauge at the centre of the quadratic grid that is 36 reflectivity measurements. Since the radar data are available every 6 minutes, the reflectivities and rain gauge were transferred into the same temporal resolution (60 minutes) by averaging the radar data. Finally, we have obtained 36 average reflectivity measurements for each precipitation event.

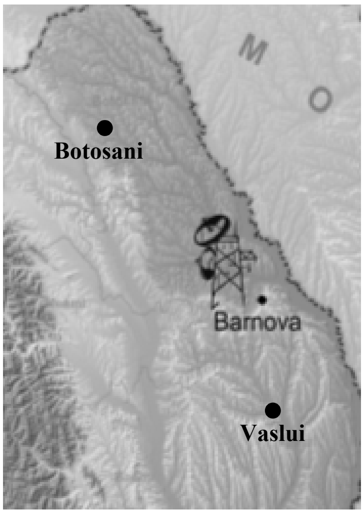

Fig.1. – Study area in the province of Moldova, Romania, showing radar and rain gauge location.

3. EXPERIMENTAL

3.1. DEVELOPMENT OF THE ANN SYSTEM

The structure of the ANN system used in this paper consists of three node layers: an input layer, a hidden layer and an output layer. The nodes in the input layer transfer the input data (average reflectivity measurements) to all the nodes in the hidden layer. These nodes calculate a weighted sum of the inputs that is subsequently subjected to a nonlinear transformation:

$$y_j = f\left(\sum_{i=1}^{m} x_i w_{ji} - b\right), \tag{2}$$

where x_i is the input to the node i in the input layer, m is the number of nodes in the input layer, w_{ji} (weights) are the connection between each node i in the input layer and each node j in the hidden layer, y_j is the output of the node j in the hidden layer, b is the bias for the input layer and f is a sigmoid function:

$$f(z) = \frac{1}{1+\exp(-z)}, \tag{3}$$

where z is the final response (rain rate) of the ANN.

The great power of neural networks stems from the fact that it is possible to "train" them. Training is affected by continually presenting the networks with the "known" inputs and outputs (targets) and modifying the connection weights between the individual nodes and the biases. The output of the network is a weighted sum of the outputs of the hidden layer. In our case, the ANN network was trained with the learning algorithm based on the back-propagation of errors [14].

Learning by error back-propagation (like in all supervised methods) is carried out in cycles, called "epochs". One epoch is a period in which all input-target pairs $\{x_i, t_i\}$ are presented once to the network. The weights are corrected after each input-target pair (x_i, t_i) produces an output vector z_i and the errors from all output components z_s are squared and summed together. After each epoch, the root-mean-squared error (RMSE) is reported:

$$\text{RMSE} = \sqrt{\frac{\sum_{i=1}^{m}\sum_{s=1}^{n}\left(t_{is} - z_{is}\right)^2}{m \cdot n}}, \tag{4}$$

where m and n denote the number of input data and the number of output nodes, respectively.

In conclusion, the RMSE value measures how good the outputs z_i (produced for the entire set of m input vectors x_i) are in comparison with all m, n-dimensional target values t_i. The aim of the supervised training, of course, is to reach a RMSE value as small as possible in the shortest time.

The back-propagation algorithm follows the gradient descent on the error surface. This process is controlled by two parameters: the learning rate α and momentum μ. The learning rate scales the magnitude of each step, down the error surface, taken after each complete calculation in the network (epoch). The momentum acts like a low pass filter, smoothing out progress over some small bumps in the error surface by remembering the previous weight change [14, 15]. Thus, the correction of weights Δw_{ji} is defined as:

$$\Delta w_{ji}^{\ell} = \alpha \delta_j^{\ell} z_i^{l-1} + \mu \Delta w_{ji}^{\ell(\text{previous})}, \tag{5}$$

where δ_j^{ℓ} is the error that occurs on the j-th node of the ℓ-th layer.

The neural networks are often affected by the effect called overtraining or overfitting. An overtrained neural network memorizes the small training set instead of generalizing the data and consequently performs badly on new data *e.g.* on the validation set. In this work, the overtraining (or overfitting) was anticipated by the so-called early stopping. Early stopping was implemented by stopping the training when the error of cross-validation of the training data starts going up, as the neural network may loose its generalization ability at this moment [14].

Concerning the topology of the ANN network, the number of neurons in the input layer is equal with the input variables (36 average reflectivity measurements and normalized), the number of neurons in output layer is equal with the number of precipitation classes (3 output neurons representing the following classes: the class 1 has the rainfall rate comprised between 0–5 mm·h^{-1}, the class 2 has the rainfall rate comprised between 5–10 mm·h^{-1} and the class 4 has the rainfall rate comprised between 15–20 mm·h^{-1}), but the optimum number of neurons for the hidden layer must be determined. Thus, the number of neurons for the hidden layer and the weight of the connections were optimized at 13 hidden neurons and 507 weight connections. Also, the learning rate $\alpha = 0.7$ and the momentum $\mu = 0.8$ were fixed after optimization.

The neural networks have been programmed to stop the training process when the average error of training drops below the target error. The target error was set up at 0.01 for both networks. Convergence was touched after 156 training epochs (Fig. 2).

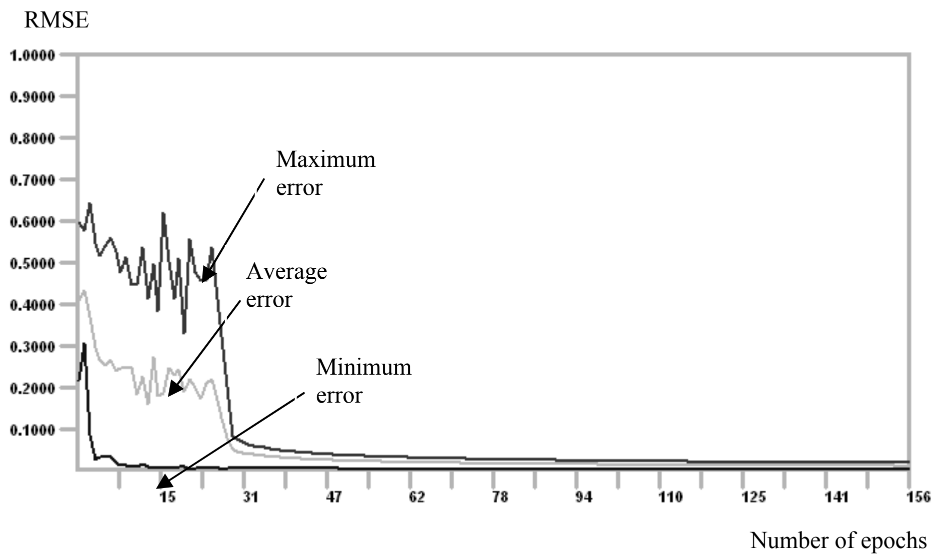

Fig. 2 – The root-mean-squared error (RMSE) of training in function of the number of epochs.

The training set consisted of 12 precipitation events. In our case, we have chosen the same number of events for each class of precipitation: 4 events for class 1, 4 events for class 2 and 4 events for class 4. The remaining 31 precipitation events were included in the validation set.

4. RESULTS AND DISCUSSION

The network was validated using all the events from the database. The method was full cross-validation, as the number of events in the database is relatively small. In order to evaluate the performance of the ANN system, several figures of merit were calculated: the rate of classification (C), of correctly classified events (CC), and of correctly classified events for each class (CC1 for the class 1, CC2 for the class 2 and CC4 for the class 4). We should specify that the C rate represents the percentage of events that were correctly or incorrectly classified. The values of these parameters are presented in Table 1 and offer a complete image of the validation results.

Analyzing the values of validation parameters, we have observed that the ANN system has a very good percentage for the rate of classification of precipitation events (C = 95.35%) so that only two events were unclassified. The values for the CC1, CC2 and CC4 rates are quite encouraging (Table 1), as these are initial results, obtained without any optimization of the nature and/or number of events included in the training set. Thus, the biggest value is for the CC4 rate (81.82%), then follows the CC2 rate which has 62.5%, and finally comes the CC1 rate which has the smallest value (CC1 = 54.55%). In our case, the smaller values obtained for the CC1 and CC2 rates do not constitute a disadvantage because the events included in classes 1 and 2 correspond to small rainfall rates *i.e.* RR is from 0 to 5 $mm{\cdot}h^{-1}$for class 1 and RR is from 5 to 10 $mm{\cdot}h^{-1}$for class 2.

Table1

The results of validation

Validation parameter	Value (%)
C	**95.35**
CC	63.34
CC1	54.55
CC2	62.5
CC4	**81.82**

Finally, it is worth emphasizing the fact that the precipitation events with rainfall rate comprised between 15 and 20 $mm{\cdot}h^{-1}$(class 4) were very well classified by the artificial neural network.

4.1. SENSITIVITY ANALYSIS

Sensitivity analysis should be considered an essential pace to all mathematical-based modeling. The main advantage of performing sensitivity analysis is to identify sensitive parameters or processes associated with model output [13]. In ANN modeling, like any mathematical-based model, sensitivity analysis provides feedback as to which input parameters are the most significant.

In order to find the most sensitive input variables, we have achieved a sensitivity analysis (Fig. 3). The absolute sensitivity is a measure of how much the outputs change when the inputs are changed. The change in the outputs is measured as each input is increased from the lowest to the highest to establish the sensitivity to change. The sensitivity analysis is a method for measuring the cause - effect relationship between the input layer and the output layer.

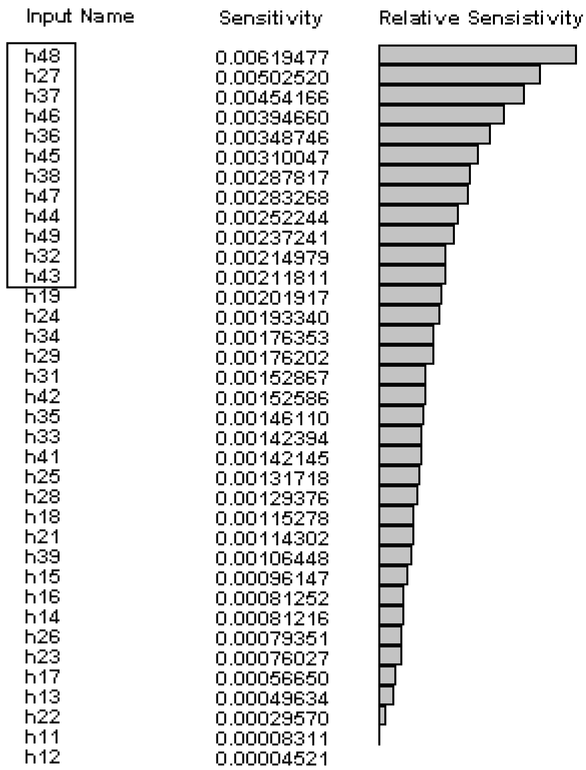

Fig. 3 – The sensitivity analysis of input weather data.

Figure 3 shows us that the majority average reflectivity measurements (input data) corresponding to the elevation 4 (h48, h46, h45, h47, h44, h49 and h43) has the highest values for the relative sensitivity (from 0.0021 to 0.0062). This fact proves that the rate of correctly classified events in the case of class 4 has the highest value (CC4 = 81.82%, Table 1). Class 4 has the rainfall rate comprised between 15–20 $mm \cdot h^{-1}$. From the meteorological point of view, the importance of this result is confirmed by the fact that the reflectivity measurements from the elevation 4 are very important in the rainfall estimation.

Using relation (1), when $a = 300$ and $b = 1.4$, for a rate of precipitation between 15-10 $mm \cdot h^{-1}$ we get a reflectivity factor of about 45 dBZ. Reflectivity values of 45 dBZ represent a sign of the presence of a severe convection [16]. For the clouds with vertical development, the highness to which we find the radar beam corresponds to elevation 4.

5. CONCLUSIONS

The results of the present study indicate that the use of artificial neural network as a rainfall forecasting system is feasible and efficient. The exploratory analysis of the ANN system specialized in the identification of the precipitation classes proved that discrimination between class 1, class 2 and class 4 is possible.

The evaluation of the performance of the ANN network was carried out by calculating several figures of merit characterizing the class identity assignment. Thus, for the classification rate of the precipitation events we have obtained a very good value (C = 95.35%). The precipitation events belonging to classes 1 and 2 were satisfactorily classified (CC1 = 54.55% and CC2 = 62.5%) while those belonging to class 4 were much better classified (CC4 = 81.82%).

From the sensitivity analysis of the input variables, it has ensued that the most sensitive reflectivity measurements (input variables) are those from the elevation 4. This result confirms the good percentage of the CC4 rate *i.e.* the fact that the ANN system has a better modeling power in the case of precipitation events with higher rainfall rates.

As the precipitation events with heavy rain accumulations generate damages, we intend to continue the research regarding the rates improvement of correctly classified events especially in the case of these events (*i.e.* CC4, CC5, etc.) using the artificial neural network.

REFERENCES

1. Gang Xu and V. Chandrasekar, *Operational Feasibility of Neural-Network-Based Radar Rainfall Estimation*, IEE Trans. Geosci. Remote Sensing, ***2,*** *1*, 13–17 (2005).
2. Reinhard Teschl, Walter L. Randeu, and Franz Teschl, *Weather Radar Estimates of Rainfall Adjusted to Rain Gauge Measurements using Neural Networks*, Proc. Internat. Joint Conf. On Neural Network, Vancouver, BC, Canada, 5126-5130, July 16–21, 2006.

3. Reinhard Teschl, Walter L. Randeu, and Franz Teschl, *Improving weather radar estimates of rainfall using feed-forward neural networks,* Neural Networks, **20**, 519–527 (2007).
4. M. Nasserri, K. Asghari, M.J. Abedini, *Optimized scenario for rainfall forecasting using genetic algorithm coupled with artificial neural network*, Expert Syst. Applications, **35**, *3*, 1415-1421 (2008).
5. H. Liu, V.Chandrasekar and E. Goegucci, *Detection of rain/no rain condition on the ground Based on Radar Observations,* IEE Trans. Geosci. Remote Sensing, **39**, *3*, 696–699 (2001).
6. Plamen N. Penchev, George N. Andreev, Kurt Varmuza, *Automatic classification of infrared spectra using a set of improved expert-based features,* Anal. Chim. Acta, **388,** 145–159 (1999).
7. R. Linker, I. Shmulevich, A, Kenny, A. Shaviv, *Soil identification and chemometrics for direct determination of nitrate in soils using FTIR-ATR mid-infrared spectroscopy*, Chemosphere, **61**, *5*, 652–658 (2005).
8. Gosav S., R. Dinica**,** M. Praisler, *Choosing between GC–FTIR and GC–MS spectra for an efficient intelligent identification of illicit amphetamines,* J. Mol. Struct., **887**, *1-3*, 269–278 (2008).
9. Tony Hall, Harold E. Brooks and Charles A. Doswell III, *Precipitation Forecasting Using a Neural Network,* Wea. Forecasting, **14***,* 338–345 (1998).
10. S. Vogt, D. Sacher, *A neural network method for wind estimation using wind profiler data*, Meteorologische Zeitschrift, **10,** *6*, 479–487 (2001).
11. Rongrui Xiao and V.Chandrasekar, *Development of* a *Neural-Network-Based Algorithm for Rainfall Estimation from Radar Observations.,* IEE Trans. Geosci. Remote Sensing, ***35,*** *1*, 160–171 (1997).
12. Honping Liu, V. Chandrasekar and Gang Xu, *An adaptive neural Network Scheme for Radar Rainfall Estimation from WSR-88 D Observation,* J Appl. Meteor., **40**, 2038–2050 (2001).
13. Skaggs, T.H., and Barry, D.A., *Sensitivity methods for time-continuous, spatially discrete groundwather contaminant transport models*, Water Resources Research*,* **32**, 2409–2420 (1996).
14. F. Despagne, D.L. Massart, *Tutorial review: neural networks in multivariate calibration*, Analyst, **123**, 157R-178R, (1998).
15. D. Dumitrescu, H.Costin, *Reţele neuronale.Teorie şi aplicaţii,* Edit. Teora, Bucharest, 1996.
16. L.J. Battan, *Radar Observation of the Atmosphere,* University of Chicago Press.