Artificial neural network model of forecasting relative humidity in different humid and arid areas of China

Zhenfang He[1, 2*]

[1]*Cold and Arid Regions Environmental and Engineering Research Institute, Chinese Academy of Sciences, Lanzhou, China, 730000*

[2]*University of Chinese Academy of Sciences, Beijing, China, 100049*

Received 6 October 2013, www.tsi.lv

Abstract

The objective of the present study is to build different models forecasting the daily mean relative humidity (MRH) values in China with the help of the meteorological parameters. A back-propagation artificial neural network (BPANN) models was employed to identify the relationship between meteorological factors and the relative humidity in China. Weather data 1-day lag was the input layer variables, including (1) the highest atmospheric pressure, (2) the lowest atmospheric pressure, (3) the average atmospheric pressure, (4) the average temperature, (5) the highest temperature, (6) the lowest temperature, (7) precipitation, (8) the average wind speed, (9) the maximum wind speed (the average wind speed over 10 minutes), (10) the utmost wind speed, (11) hours of sunlight, (12) the relative humidity. Experimental results: in the validation period for 1-day lead, the comparison of the prediction performance efficiency of the BPANN models indicated that the BPANN models with trainbr algorithm was superior to the remaining two ones (trainlm and traingdx) in forecasting the relative humidity time series in term of correlation coefficient (R). During the training and testing periods for 1-day lead, the best performance for the given problem was arid area, followed by semi-arid area, semi-humid area, and humid area respectively. The possible cause for the results was that the impact of these factors on the relative humidity in arid area was the largest, followed by semi-arid area, semi-humid area, and humid area, respectively. From the prediction results of MRHextrema, humid area was the first; semi-arid area was the second; semi-humid area was the third; and arid area was the fourth. From the prediction results of MRHextrema, trainbr algorithm was the best in arid area, semi-humid area, and humid area; but trainlm was the best in semi-arid area. So trainbr algorithm was further employed to predict MRH for 2, 3 or 4-day lead at Urumqi City. From the training and testing effects, 1-day lead was the best, followed by 2, 3 or 4-day lead respectively. In the prediction results of MRHextrema, the best was 2-day lead; the second was 3-day lead; the third was 1-day lead; and the fourth was 4-day lead. The BPANN model results will assist researchers determining meteorological parameters to forecast MRH.

Keywords: Meteorological Parameters, Humid and Arid Areas, Artificial Neural Network Model, Relative Humidity, Training Algorithms

1 Introduction

Relative humidity as a major meteorological component of the hydrologic cycle plays an important role in climate change studies in climatic regions. The influence of relative humidity in controlled environments (e.g. industrial processes in agro-food processing, cold storage of foods such as vegetables, fruits and meat, or controls in greenhouses) is vital [1]. The black-box modelling method is one of data developing techniques in which the knowledge is abstracted in term of models. When data is not sufficient, empirical models are a good alternative method, and can provide useful results without a costly calibration time [2]. It is crucial to contain all significant variables that influence relative humidity. Too much irrelevant data would increase the network training and restraint the network from learning adequately. Unlike physical models, black-box models can be made adaptive, by transforming their parameters as a function of the actual performance that the models show [3]. Many researchers have applied ANN to predict relative humidity with the data related to weather conditions and climate being collected for certain periods [4]. They used three-layer feed forward neural networks to predict indoor temperature and relative humidity, and temperature and relative humidity predictions got very accurate and satisfactory results [5]. They developed an artificial neural network model to predict the thermal behaviour of an open office in a modern building, and used external and internal climate data recorded over three months to build and validate models for predicting relative humidity, moreover, the results reveal that the model provides reasonably good predictions[6]. They aimed to determine simultaneously relative humidity in air, by employing an optical sensor based on a nafion-crystal violet film and ANN to perform multivariate calibration. In addition, he proposed a kinetic approach in order to improve the ANN performance [7]. They adopted artificial neural network approach to estimate the monthly mean relative humidity (MRH) values with the help of the topographical and meteorological parameters, and used latitude, longitude, altitude, precipitation and months of the year in the input layer of the ANN network, while the MRH in output layer of the network, at last, the

* ***Corresponding author*** e-mail: hzfwhy@126.com

result shows that the obtained values were in the acceptable error limits. To examine prediction suitability of artificial neural networks in terms of relative humidity in different areas, measurement data were taken from four different meteorological stations in China. In the present study, a methodology is presented to predict relative humidity simultaneously in China by using BPANN models. Thus, the methodology is of great practical importance. The applicability of the methodology is demonstrated by using three ANN training algorithms namely trainlm, trainbr, traingdx for forecasting relative humidity.

2 Study background

The four study stations selected for this study are known as Chongqing, Beijing, Lanzhou, Wulumuqi (Urumqi), which are located in China (Figure 1). On the basis of precipitation, China is divided into four areas, viz., humid, semi-humid, semi-arid and arid areas. Precipitation of humid area is greater than 800 mm, precipitation of semi-humid area range from 400 mm to 800mm, precipitation of semi-arid area vary from 200 mm to 400 mm, and precipitation of arid area is less than 200mm. The four study stations are located in the four areas, respectively; that is, study areas include Chongqing, Beijing, Lanzhou, Wulumuqi, which represent humid, semi-humid, semi-arid and arid area, respectively. Eastern China has a typical monsoon climate with high temperature and rainfall mainly occurring in the summer season. Northwestern China has a typical continent climate due to be far away from the ocean.

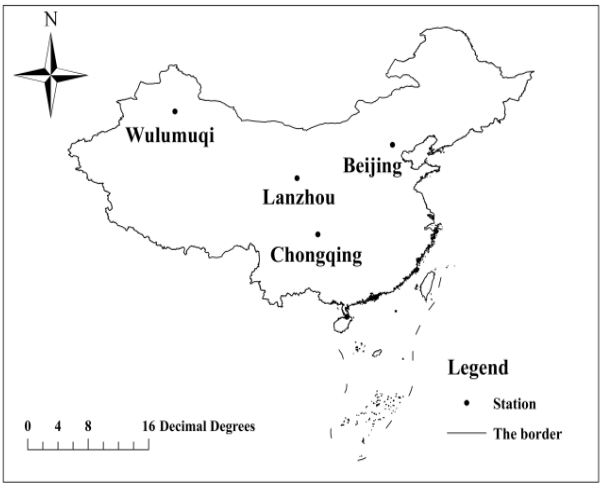

FIGURE 1. Location of the four meteorological observation stations

Table 1 shows the main meteorological parameters of the four stations. The average annual temperatures (AAT) of the four study stations are 14.8 ℃, 11.8℃, 9.3℃ and 6.6℃, respectively. The average annual relative humidities (AARH) of the four study stations are 81%, 52%, 53%, and 54%, respectively. The average annual rainfalls (AAR) of the four study stations are 1138mm, 576.9mm, 316.1mm, and 175.3mm, respectively. The mean daily minimum and maximum relative humidities are 42% and 98% in Chongqing, respectively. The mean daily minimum and maximum relative humidities are 11% and 97% in Beijing, respectively. The mean daily minimum and maximum relative humidities are 17% and 91% in Lanzhou, respectively. The mean daily minimum and maximum relative humidities are 11% and 93% in Wulumuqi, respectively. There is a 2-year (2007-2008) record of the daily mean relative humidity (MRH) in the four study stations, and that is {X (t), t=1, 2… n}. For the ANN models, the data series were divided into a training set (January to December 2007), and a testing set (January to December 2008). The 2007 year time series are used for calibration/ training of the model, and the remaining year data are used for verification or testing purposes.

TABLE 1 The main meteorological parameters of the four study stations in China

Station	Chongqing	Beijing	Lanzhou	Wulumuqi
Latitude (oN)	28.8	39.9	36.0	43.7
Longitude (oE)	108.7	116.2	103.8	87.6
Altitude (m)	665	54	1518	918
AA T (℃)	14.8	11.8	9.3	6.6
AARH (%)	81	52	53	54
AAR (mm)	1138	576.9	316.1	175.3

3 Analysis Method based on ANN Model

An Artificial Neural Network (ANN) is a mathematical model or computational model that is inspired by the structure and functional aspects of biological neural networks. The true power and advantage of neural networks lies in their ability to represent both linear and non-linear relationships and to learn these relationships directly from the data being modelled. But, traditional linear models are simply inadequate when it comes to modelling data that contains non-linear characteristics. ANN is successfully applied to the areas of engineering, mathematics, medicine, meteorology, economy, neurology, psychology, electricity etc. [8, 9]. Although back propagation training has proved to be efficient in lots of applications, it has inherent limitations of gradient based techniques such as slow convergence and the local search nature. In this study, feed forward neural network architecture has been used and three improved ANN training algorithms, viz., gradient descent with momentum and adaptive learning rate back propagation (GDX) algorithm (traingdx), Levenberg-Marquardt (LM) algorithm (trainlm) and Bayesian regularization (BR) algorithm (trainbr) minimize a sum of squared error and to overcome the limitations in the standard BPANN. Three statistical criteria (or statistical indicators) were used in order to evaluate the effectiveness of back propagation artificial neural network (BPANN) models developed in this study. They are root mean square error (RMSE), mean error (ME), correlation coefficient (R), and percentage error of peak (EOP).

The global error function (E) can be calculated by using equation (1) as,

$$E=\frac{1}{2}\sum\left(O_i-P_i\right)^2, \quad (1)$$

where E is the global error function, O_i is the desired output and P_i is the output predicted by the network. The back propagation algorithm uses the gradient descent technique to adjust the weights in which the global error function, E, is minimized by modifying the weights using the following equation (2),

$$\Delta W_{ji}=-\eta\frac{\partial E}{\partial W_{ji}}, \quad (2)$$

where ΔW_{ji} = weight increment from node *i* to node *j*; and η= learning rate, by which the size of the step taken along the error surface is determined. The weights between the hidden layer and the output layer are adjusted first, followed by the weights between the hidden layer and the input layer.

The transfer function denoted by *f*(*x*), defines the output of a neuron in terms of the induced local field x. linear function is represented by equation (3),

$$f(x)=x. \quad (3)$$

The most commonly used activation function within the neurons is the logistic sigmoid function, which takes the form shown in equation (4),

$$f(x)=\frac{1}{1+e^{-x}}. \quad (4)$$

The bipolarity S function, which has the advantage of a positive or negative output, can also be used, represented by equation (5). And, four statistical indicators were used to evaluate the effectiveness of the ANN models developed in this study. They are the correlation coefficient (R), mean error (ME), root mean square error (RMSE) and peak error percentage (EOP (%)), given by the following equations (6), (7) and (8).

$$f(x)=\frac{e^x-e^{-x}}{e^x+e^{-x}}, \quad (5)$$

$$RMSE=\sqrt{\frac{\sum(O_i-P_i)^2}{N}}, \quad (6)$$

$$ME=\frac{1}{N}\sum(O_i-P_i), \quad (7)$$

$$R=\frac{\frac{1}{N}\sum(O_i-\bar{O})(P_i-\bar{P})}{\sqrt{\frac{1}{N}(P_i-\bar{P})^2}\sqrt{\frac{1}{N}\sum(O_i-\bar{O})^2}}, \quad (8)$$

where O_i= observed value for i^th data, P_i = predicted value for i^th data, = mean of observed value, = mean of predicted value, and n = number of observations. The best fit between observed and predicted values under ideal conditions would yield RMSE = 0, ME = 0, R = 1.

The percentage error of peak MRH, EOP (%) is defined as follows:

$$EO_p(\%)=\frac{P_P-O_P}{O_P}\times 100\%, \quad (9)$$

where P_P denotes the peak data of the predicted MRH, O_P is the peak data of the observed MRH and EO_P is the relative error of the maximum difference in the highest peak MRH.

Levenberg–Marquardt method is a modification of the Newton algorithm for finding an optimal solution to a minimization question. It is used to approach second order training speed and accuracy without having to compute the Hessian matrix. It uses an approximate to the Hessianmatrix in the following Newton-like weight update.

$$W_{i+1}=W_i-[J^TJ+uI]^{-1}J^Te, \quad (10)$$

where, *W* is weights of the neural network. *J* is Jacobian matrix of the performance criteria to be minimized. *u* is a scalar that controls the learning process, and e is residual error vector. When the scalar u is zero, this is just Newton's method to use the approximate Hessian matrix. When *u* is large, the equation becomes gradient descent with small step size. Newton's method is faster and more accurate near anerrorminimum, so the objective is to shift towards Newton's method as quickly as possible.

4 Relationships between MRH and meteorological factors

It is necessary to stress here that the problem of finding the most appropriate structure for a statistical daily MRH prediction model is perhaps one of the major problem for the modellers. Moreover, the link between meteorological factors and daily MRH is non-linear; the selected variables depend on the particular target of the prediction model. Finally, it must be stressed that observed data are affected by various kinds of noise. Generally, some degree of a priori knowledge is used to specify the initial set of candidate inputs [10, 11]. Here, the autoregressive analysis results showed that the autoregressive order was 1, with correlation coefficient R under a 95% level of significance; and this meant that modelling the time series by using regressive models it should consider 1 past sample at least. The samples of the correlation functions computed, considering daily time series were reported in Figure 2 under a 95% level of significance.

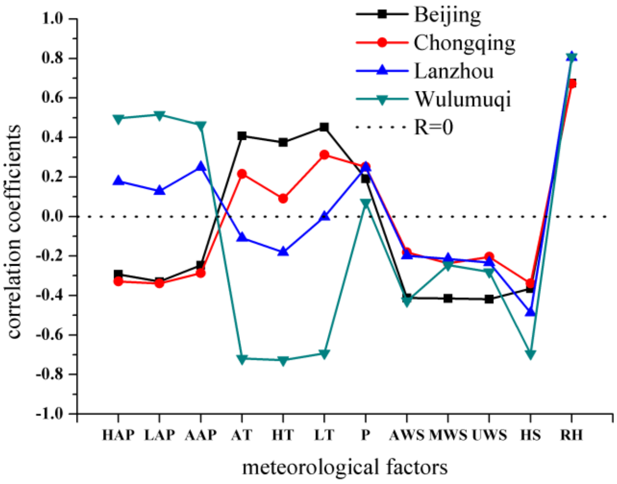

FIGURE 2 The relationship between MRH and meteorological factors

From the correlation coefficient (R) values between MRH and meteorological factors, such as AT, HS, RH 1-day lag, Wulumuqi was the best, followed by Lanzhou, Beijing, respectively, and Chongqing was the worst; that is, from the impact of these factors on the humidity, arid area was the largest, followed by semi-arid, semi-humid area, and humid area was the least. Therefore, 12 antecedent values of the meteorological factors were selected as input for modelling daily MRH. The 12 meteorological factors 1-day lag include (1) the highest atmospheric pressure (HAP), (2) the lowest atmospheric pressure (LAP), (3) the average atmospheric pressure (AAP), (4) the average temperature (AT), (5) the highest temperature (HT), (6) the lowest temperature (LT), (7) precipitation (P), (8) the average wind speed (AWS), (9) the maximum wind speed (the average wind speed over 10 minutes) (MWS), (10) the utmost wind speed (UWS), (11) hours of sunlight (HS), (12) the relative humidity (RH). In other words, the input layer consisted of 12 input nodes/variables and included a 1-day time-lag (X (t)), considering X (t) was the value of a given variable at the present time step for daily MRH. The output of the network was a prediction of daily MRH at time step t+1.

5 Optimization of the hidden layer nodes

In order to apply the ANN models, several network structures were tested to find the most appropriate topology. Numerous studies have shown theoretically that three-layered BP networks can precisely describe any nonlinear mapping relation [12, 13]. Therefore, three-layered BP networks were used in this paper. Sigmoid (tansig) and linear functions were used as activation functions in the neurons of the hidden layer and output neuron, respectively. The training was done for a maximum of 100 iterations. To avoid the over fitting problem, which generally appears with the application of ANN, trains were used. The selection of the network was performed considering a minimum value of RMSE for the train data set. 12 nodes and one node were defined as the input and output layers, respectively, according to above paragraph. The number of hidden layer nodes was calculated by the trial-and-error method. 1 node was initially chosen; 6 nodes were finally selected after debugging. Figure 3 shows RMSE index evolution as a function of the number of hidden nodes for the different variances assessed in Wulumuqi station. In this study, multiple lead-time predictions of 1, 2, 3, 4 time steps were also performed for each input structure. In the case of the ANN, trainlm, traingdx and trainbr were used as learning rules. In all calculations, the best performing ANNs were feed forward back propagation with trainbr algorithm.

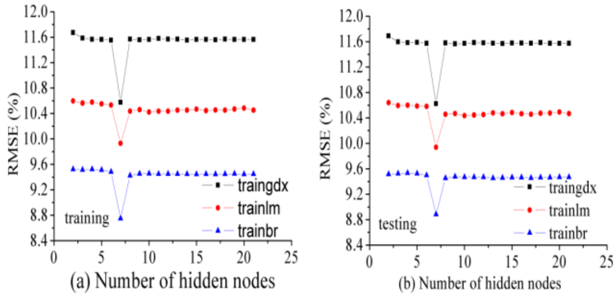

FIGURE 3 Experimental results for optimization of the hidden layer neurons (nodes) for BPANN models

6 Prediction of the daily mean relative humidity (MRH) at 1day lead time

The values of statistical indicators for the three training algorithms for the four stations were shown in Table 2 during training and testing periods. TABLE 2 showed the performance of all the three training algorithms was good during both training and testing periods; all the three training algorithms were able to forecast relative humidity 1 day ahead with a reasonable accuracy in all the four stations. For the trainbr training algorithm during testing period, the correlation coefficient (R) values ranged from 0.7634 to 0.9022, mean error (ME) from -0.3119% to 0.0872%, and root mean square error (RMSE) values from 7.8161% to 34.5310%. For the traingdx training algorithm during testing period, the correlation coefficient (R) values ranged from 0.6859 to 0.8569, mean error (ME) from -0.5927% to 0.3189%, and root mean square error (RMSE) values from 9.2149% to 36.7818%, whereas these figures for the trainlm algorithm were 0.7356 to 0.8769, -0.9683% to 0.3189%, 8.7155% to 31.0500%, respectively.

Table 2 summarized the results of the testing for every network configuration. The best overall performance for the given problem was achieved by the back propagation artificial neural network trained with the trainbr algorithm and the second best by the BPANN trained with the trainlm algorithm. The rest of the network performed relatively well but tended to overestimate the observed dataset. Also all the networks performed very well for 1 day ahead predictions. Since the given problem was aiming at predicting the daily mean relative humidity (MRH), overestimating models were not of particular interest. Thus, the most promising technique seemed to be one using the feed forward neural network trained with the trainbr algorithm. The physical meaning of this result was that the structure of this model allowed its weights to adjust to values that depict the trends of the natural system we were simulating.

TABLE 2 Comparison of trainbr, trainlm and traingdx algorithms

Algorithm	**Station**	**area**	**RMSE (%)**		**R**		**ME (%)**		**EOP (%)**	
			training	*testing*	*training*	*testing*	*training*	*testing*	*training*	*testing*
trainbr	Chongqing	humid	31.0583	34.5310	0.7639	0.7634	0.2486	-0.3119	-8.4280	-2.5180
	Beijing	semi-humid	12.5529	11.7988	0.7989	0.8167	-0.2219	0.0872	-11.747	-7.4643
	Lanzhou	semi-arid	7.7302	7.8161	0.8814	0.8629	0.2260	-0.1754	-3.3024	-7.3429
	Wulumuqi	arid	8.7508	8.8835	0.8844	0.9022	0.6602	-0.2056	3.4398	-8.7776
trainlm	Chongqing	humid	31.0500	38.0384	0.7638	0.7356	0.3628	-0.3609	-2.9994	-3.2772
	Beijing	semi-humid	13.6846	12.4098	0.7546	0.7947	0.1486	-0.1470	-4.5094	-9.4740
	Lanzhou	semi-arid	8.7155	8.7164	0.8460	0.8254	-0.1174	0.0996	-7.2760	-5.5173
	Wulumuqi	arid	9.9305	9.9392	0.8489	0.8769	0.9733	-0.9683	-10.701	-10.538
Traingdx	Chongqing	humid	36.7818	43.9031	0.7117	0.6859	0.3931	-0.4416	-9.8046	-6.0918
	Beijing	semi-humid	13.6846	13.1006	0.7024	0.7680	-0.1283	0.3189	-20.776	-19.359
	Lanzhou	semi-arid	9.2149	9.7276	0.8263	0.7774	0.0426	-0.1046	-14.699	-15.136
	Wulumuqi	arid	10.5758	10.6243	0.8265	0.8569	1.0433	-0.5927	-11.996	-13.381

The model calibration and validation results for the model design were analysed. In the training stage, at Wulumuqi station, the RMSE values for trainbr, trainlm and traingdx were 8.7508%, 9.9305% and 10.5758%, respectively, the ME values for trainbr, trainlm and traingdx were 0.6602%, 0.9733% and 1.0433%, and the R values for trainbr, trainlmand traingdx were 0.8844, 0.8489 and 0.8265. In the testing stage, RMSE were 8.8835%, 9.9392%, 10.6243%, ME are -0.2056%, -0.9683%, and -0.5927%, and R were 0.9022, 0.8769 and 0.8569. The RMSE values of the trainbr were smaller than trainlm and traingdx in both the training and testing stages, which implied that the calibration and validation capability of the ANN model with trainbr was better than that one with trainlm or traingdx for the given data. However, in testing phase, the magnitude of the ME values at Wulumuqi station was higher than that at Lanzhou station, implying a higher bias of the prediction results at Wulumuqi station. RMSE values also showed that the prediction results for Lanzhou station were better than those for Wulumuqi station. The R values were not much different for the two stations. For Lanzhou station, the performance of the three training algorithms was similar. However, for Wulumuqi station, the performance of the model with trainbr was better than that with trainlm and traingdx, especially for testing phase. However, the correlation coefficient (R) between input variables and daily MRH for Lanzhou station was lower than that for Wulumuqi station. This implied that the time-series data of Lanzhou station could have a higher nonlinearity than that of Lanzhou station and included more noisy data that hinder the model training process. The result of model performances for four stations indicated that the trainbr was more likely to catch the nonlinear relationship between meteorological factors and the relative humidity and to filter out the noise than trainlm or traingdx. FIGURE 4 showed the correlation coefficient (R) values between observed and predicted MRH 1 day ahead by three training algorithms with the observed daily MRH at four stations. From the training effects of all the three training algorithms, Wulumuqi was the best, followed by Lanzhou, Beijing, respectively, and Chongqing was the worst; similarly, from the prediction results, Wulumuqi was the best, followed by Lanzhou, Beijing, respectively, and Chongqing was the worst.

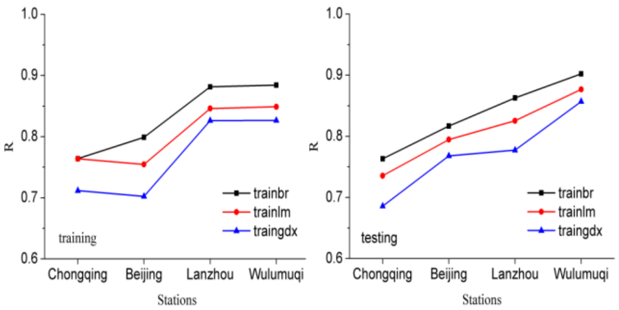

FIGURE 4 Experimental results for the correlation coefficient (R) values between observed and predicted MRH

MRH extrema are affected by random factors, which makes them difficult to predict. The EOP values in TABLE 2 reflected the models' performance in simulating the extremum at the four stations. During the training period, trainbr algorithm was the best at Wulumuqi and Lanzhou station, but trainlm algorithm was the best at Beijing and Chongqing station. During the testing period, trainbr algorithm was the best at Wulumuqi, Beijing, Chongqing station, but trainlm was the best at Lanzhou. From the prediction results of MRH extrema, Chongqing was the best, followed by Lanzhou, Beijing, respectively, and Wulumuqi was the worst. FIGURE 5 showed the comparison of the predicted daily mean relative humidity (MRH) 1 day ahead by three training algorithms with the observed daily MRH at four stations. These figures indicated that there was a very good matching between observed and simulated daily MRH at all the stations. Based on the statistical indicators and the graphical comparison, it could be deduced that all the three algorithms produced more or less same results.

However, the performance of the trainbr algorithm could be considered superior based on the performance criteria used in this study. On the other hand, the traingdx algorithm could effectively be used for large networks with little less accuracy than the trainlm algorithm and the trainbr algorithm respectively. As a matter of fact, however, any of these three algorithms could be used for the daily mean relative humidity (MRH) prediction in the stations.

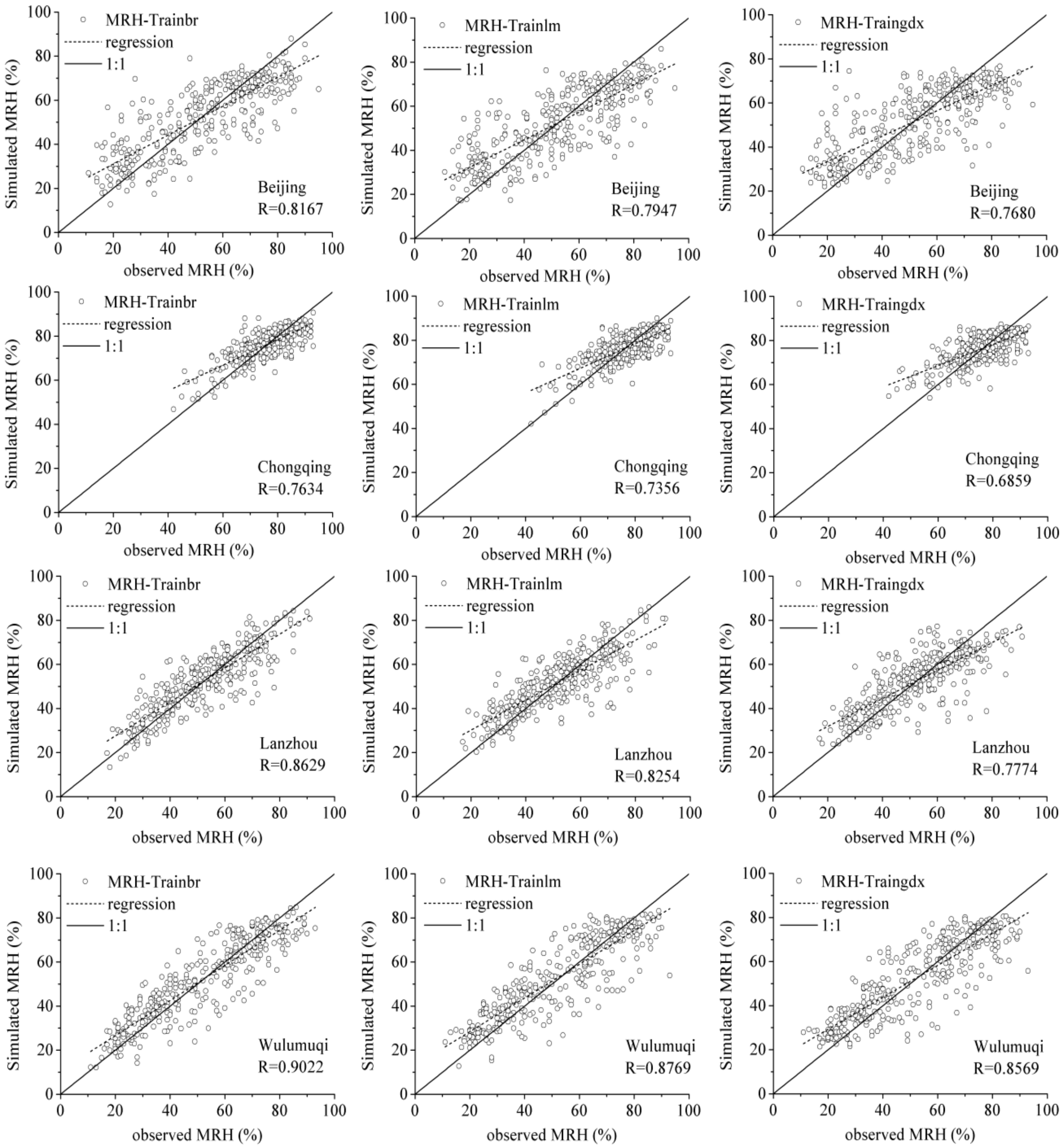

FIGURE 5 Comparison of observed and predicted MRH 1 day ahead at the four stations

It is very difficult to know which training algorithm will be the fastest or accurate for a given problem. Traingdx has adaptive learning rate. It is a faster training than traingd (the gradient descent algorithm), but can only be used in batch mode training. Trainbr (Bayesian regularization) is modification of the Levenberg–Marquardt (LM) training algorithm to produce networks which generalize well. It reduces the difficulty of determining the optimum network architecture. Trainlm (Levenberg–Marquardt) is normally used for training purpose if enough memory is available. However, the LM algorithm appeared to be the fastest training algorithm, because the LM method must solve a linear system of equations in order to obtain the search direction, the computation becomes expensive when the number of input elements and the volume of the training data increase. Therefore, when the volume of the data is large, the standard gradient descent algorithm is used for training.

7 Prediction of the daily mean relative humidity (MRH) at higher lead times

As the three ANN training algorithms were determined in this study, the trainbr algorithm performed slightly better than the other two algorithms, and it was further employed to predict the daily mean relative humidity (MRH) at 2, 3 and 4-day lead at the four stations. It was worth referring that the ANN inputs used for this analysis were the same as that used for predicting daily MRH 1 day ahead. The performance of the model for Wulumuqi station in terms of correlation coefficient (R), mean error (ME), and root mean square error (RMSE) statistics along the prediction time horizon during the testing period was shown in TABLE 3. The values of the performance criteria have been acquired for the station. Taking Wulumuqi station as an example, it was obvious from this table that the R value varied from 0.9022 for 1-day lead time prediction to 0.8410 for 4-day lead time prediction, the value of ME varied from -0.2056% for 1-day lead time to -0.6214% for 4-day lead time, and the value of RMSE varied from 8.8835% for 1-day lead time to 11.2188% for 4-day lead time. ME values for 1, 4-day lead time indicated an overestimation for the model, and an overestimation for 2, 3-day lead time. However, the magnitude of the ME values for 4-day lead-time was higher than that for 1-day lead time, implying a higher bias of the prediction results for 4-day lead-time. At Wulumuqi station, from the training effects of trainbr algorithm, 1-day lead time was the best, followed by 2, 4-day lead time respectively, and 3-day lead time was the worst; Similarly, from the prediction results, 1-day lead time was the best, followed by 2,4-day lead time respectively, and 3-day lead time was the worst. Thus, based on the performance criteria, it could be deduced that the predicted the daily mean relative humidity (MRH) for the higher lead times (2 to 4 days) were rational in this study, but the performance of the BPANN model normally reduced with an increase in the lead time. In the simulated results of MRH extrema, during the training period, 4-day lead time was the best, followed by 1,2-day lead time respectively, and 3-day lead time was the worst; during the testing period, 2-day lead time was the best, followed by 3,1-day lead time respectively, and 4-day lead time was the worst.

TABLE 3 Goodness-of-fit statistics using trainbr for different lead time forecasts at Wulumuqi station

lead time (day)	**RMSE (%)**		**R**		**ME (%)**		**EOP (%)**	
	training	*testing*	*training*	*testing*	*training*	*testing*	*training*	*testing*
1	8.7508	8.8835	0.8844	0.9022	0.6602	-0.2056	3.4398	-8.7776
2	13.5752	13.7296	0.8193	0.8496	1.2087	-0.5599	-4.3926	1.1988
3	58.8667	55.3289	0.7831	0.8204	0.8514	-0.9652	-6.6652	-2.5899
4	11.0688	11.2188	0.8038	0.8410	0.4782	-0.6214	3.0198	7.9712

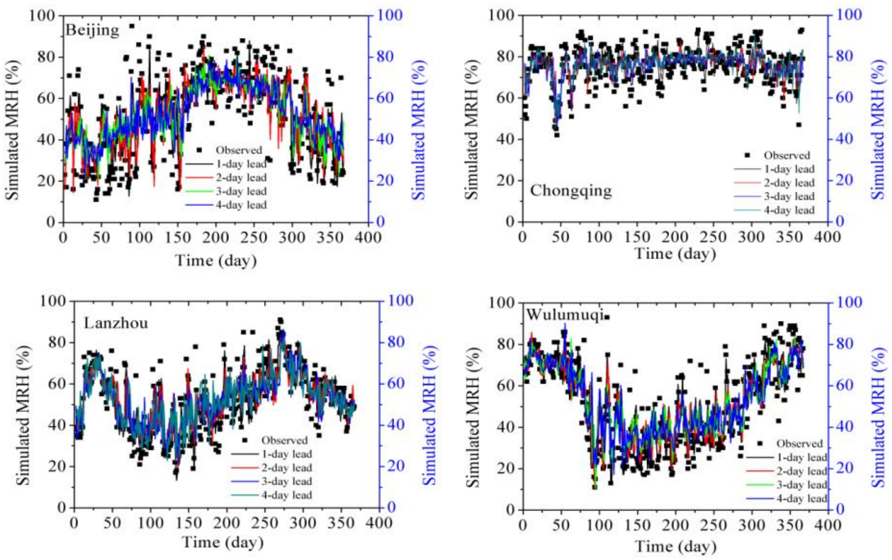

FIGURE 6 Comparison of observed and predicted MRH using trainbr from 1 to 4 days ahead at at the four stations

At the same time, FIGURE 6 also showed a comparison of observed and calculated for the daily mean relative humidity (MRH) predictions from 1 to 4 days ahead at four stations by the best performing network. The model results showed relative good prediction of the trend of the daily mean relative humidity (MRH); however the model prediction accuracy decreased slightly with increasing horizon of prediction. In the mass, it could be deduced that the designed BPANN models could predict the daily mean relative humidity (MRH) over the area reasonably well for 1-, 2-, 3- and 4-day lead times. Thus, the BPANN technique was more appropriate where the knowledge of meteorological parameters was restricted.

8 Conclusion

This paper presented a multi-objective strategy for the optimal design of BPANNs when disposing nonlinear modelling relative humidity time series. The most suitable configuration for this task was proven to be a (12; 6; 1) feed forward network using trainbr (Bayesian regularization) as it showed the most accurate predictions of relative humidity 1-day in advance. From the prediction outcome 1day lead time, arid area was the best, followed by semi-arid area, semi-humid area, respectively, and humid area was the worst. The possible reason was that the impact of these factors on the relative humidity in arid area was the largest, followed by that in semi-arid, semi-humid area, and that in humid area is the least. In predicting the extremum, trainbr algorithm was the best at Wulumuqi station, Beijing station and Chongqing station, but trainlm was the best at Lanzhou station. From the prediction results of MRH extrema, Chongqing was the best, followed by, Lanzhou, Beijing, respectively, and Wulumuqi was the worst.The BPANN model trained with trainbr algorithm was further used to predict relative humidity at 4 stations with higher lead times (2- 3- and 4-daily lead times). From the prediction results for higher lead times at Wulumuqi station, 1-day lead time was the best, followed by 2, 4-day lead time respectively, and 3-day lead time was the worst. It was found that the relative humidity prediction was reasonably good for all the three lead times, but the accuracy of prediction decreased with increasing lead times. In predicting the extremum at Wulumuqi station, 2-day lead time was the best, followed by 3, 1-day lead time respectively, and 4-day lead time was the worst.

References

[1] Martinez-Romero A, Ortega J F, de-Juan J A, Tarjuelo J M, Moreno M A 2010 Development of relative humidity models by using optimized neural network structures *Spanish Journal of Agricultural Research* **8**(S2) 162-71

[2] Krishna B, Satyaji Rao Y R, Vijaya T 2008 Modelling groundwater levels in an urban coastal aquifer using artificial neural networks *Hydrological Processes* **22**(8) 1180-8

[3] Ruano A E, Crispim E M, Conceicao E Z E, Lucio M 2006 Prediction of building's temperature using neural networks models *Energy and Buildings* **38**(6) 682-94

[4] Lu T, Viljanen M 2009 Prediction of indoor temperature and relative humidity using neural network models: model comparison. *Neural Computing & Applications* **18**(4) 345-57

[5] Mustafaraj G, Lowry G, Chen J 2011 Prediction of room temperature and relative humidity by autoregressive linear and nonlinear neural network models for an open office *Energy and Buildings* **43**(6) 1452-60

[6] Raimundo Jr I M, Narayanaswamy R 2001 Simultaneous determination of relative humidity and ammonia in air employing an optical fibre sensor and artificial neural network *Sensors and Actuators B-Chemical* **74**(1) 60-8

[7] Yasar A, Simsek E, Bilgili M, Yucel A, llhan I 2012 Estimation of relative humidity based on artificial neural network approach in the Aegean Region of Turkey *Meteorology and Atmospheric Physics* **115**(1-2) 81-7

[8] Liu S Q, Lu X F, Liu R M, Zeng W H, He M C 2012 The application of the Bp neutral network in the estimation of non-point source pollution *Procedia Environmental Sciences* **13**(2012) 263-73

[9] Fleming Z L, Monks P S, Manning A J 2012 Review: Untangling the influence of air-mass history in interpreting observed atmospheric composition *Atmospheric Research* **104**(2012) 1-39

[10] Thirumalaiah K, Deo M C 2000 Hydrological Forecasting Using Neural Networks *Journal of Hydrologic Engineering* **5**(2) 180-9

[11] Wang G 2013 Research on supply chain demand prediction based on BP neural network algorithm *INMATEH-Agricultural Engineering* **40**(2) 27-34

[12] Jiang S, Ren Z, Xue K, Li C 2008 Application of BPANN for prediction of backward ball spinning of thin-walled tubular part with longitudinal inner ribs *Journal of Materials Processing Technology* **196**(1) 190-6

[13] Wang H 2013 Research on supply chain performance evaluation of fresh agricultural products *INMATEH-Agricultural Engineering* **40**(2) 35-42

Author

Zhenfang He, born in January, 1983, Lanzhou County, Gansu Province, P.R. China

Current position, grades: the PHD of University of Chinese Academy of Sciences, China.
University studies: received her B.Sc. in GIS from Shandong Agricultural University in China. She received his M.Sc. from Northwest University in China.
Scientific interest: Her research interest fields include computer application.
Publications: more than 20 papers published in various journals.
Experience: She has completed five scientific research projects.