Transaction A: Civil Engineering
Vol. 16, No. 2, pp. 165–172

SCIENTIA
IRANICA

Long Lead Rainfall Prediction Using Statistical Downscaling and Artificial Neural Network Modeling

M. Karamouz[1,*], M. Fallahi[2], S. Nazif[1] and M. Rahimi Farahani[2]

Abstract. *Long lead rainfall prediction is important in the management and operation of water resources and many models have been developed for this purpose. Each of the developed models has its special strengths and weaknesses that must be considered in real time applications. In this paper, field and General Circulation Models (GCM) data are used with the Statistical Downscaling Model (SDSM) and the Artificial Neural Network (ANN) model for long lead rainfall prediction. These models have been used for the prediction of rainfall for 5 months (from December to April) in a study area in the south eastern part of Iran. The SDSM model considers climate change scenarios using the selected climate parameters in rainfall prediction, but the ANN models are driven by observed data and do not consider physical relations between variables. The results show that SDSM outperforms the ANN model.*

Keywords: *Statistical Downscaling Model (SDSM); Artificial Neural Network (ANN); Precipitation; GCM.*

INTRODUCTION

In recent years, many efforts have been devoted to investigating the effects of large scale climate signals and climate change on rainfall variability in different parts of the world. Statistical methods in the form of multiple nonlinear regression methods are used for predicting rainfall. These models can consider more than one predictor for rainfall prediction.

Even though General Circulation Models (GCMs) can be run at high resolutions, still the results from such models need to be downscaled for individual sites or localities for impact studies. Downscaling enables the construction of climate change scenarios for individual sites at daily time-scales using grid resolution GCM outputs. Wilby et al. [1] investigated the SDSM application for spatial and temporal downscaling of the long lead predictions of meteorological variables as well as rain and temperature. Harpham and Wilby [2] predicted the precipitation for different parts of England using different models, including SDSM, a radial neural network and a multilayer neural network. The results of their research show that all of these methods are capable of predicting precipitation; however, in different regions, their capabilities are different. Massah [3] has downscaled long-lead predictions of rain and temperature in the Zayanderud River basin in Iran using Kriging and inversed distance weighting methods.

As for the studies of variability and climate signals, Nazemosadat [4] showed that variations in Sea Surface Temperature (SST) in the Persian Gulf have significant effects on precipitation variability in the southwestern part of Iran. Nazemosadat and Cordery [5] also studied the effects of the EL Nino-Southern Index (ENSO) on precipitation variability recorded on 41 rain gauges in different parts of Iran. The results of their study have shown that there is a low correlation between the Southern Oscillation Index (SOI) and precipitation in Iran. Karamouz et al. [6] have studied the effects of the SST, SLP, ΔSLP, ENSO and SOI on precipitation variability in the western parts of Iran and showed that there is considerable correlation between SST, SLP and ΔSLP of key areas

1. *School of Civil Engineering, University of Tehran, Tehran, P.O. Box 11163-4563, Iran.*
2. *School of Civil Engineering, Amirkabir University of Technology, Tehran, P.O. Box 4413-15875, Iran.*
*. *Corresponding author. E-mail: Karamouz@ut.ac.ir*

Received 17 March 2007; received in revised form 5 November 2007; accepted 2 March 2008

in the Atlantic Ocean, the Mediterranean Sea and some other sea surface areas and precipitation in the western part of Iran.

In this study, ANN models have been used as well. ANNs have been proved an effective alternative to other methods of modeling water resource variables [7-11]. A successful application of ANN to rainfall forecasting has been done by French et al. [12] who applied a neural network to forecast one-hour-ahead, two-dimensional rainfall fields on a regular grid. Toth et al. [13] investigated the capability of ANN in short-term rainfall forecasting using historical rainfall data as the only input information. Another approach dealing with a temporal pattern is to introduce a cyclic (feedback) connection described by direct loops in the network. Once feedback connections are included, a neural network is often called a Recurrent Neural Network (RNN). Anmala et al. [14] reported that recurrent networks may perform better than standard feed forward networks in predicting monthly runoff. Also the models established by Gowariker et al. [15], Tapliyal [16] and Sahai et al. [17] are categorized as Walker's studies. Among these, the model of Sahai et al. [17], which links global SST with Indian monsoon seasonal data, appears to include a successful approach.

The objective of this study is to compare two models, namely Statistical Down-Scaling Model (SDSM) and Artificial Neural Network (ANN) for rainfall prediction. Different measures of error have been used for comparing model performances. The National Center for Environmental Prediction (NCEP) reanalyzed data sets and HadCM3 (Second Hadley Centre Coupled Ocean-Atmosphere AOGCM) is used in the SDSM model and large scale climate signals such as SLP and SST in the ANN model for rainfall prediction. The remainder of this paper is organized as follows.

A brief description of the study area and data is provided in the next section, followed by a description of two methods for rainfall prediction. Then, the results are presented and compared. Finally, a summary and conclusion are given.

STUDY AREA

The study area is the Kajoo River basin in the south Baloochestan region in the south-eastern part of Iran (Figure 1), which has an area of 3659 km^2. The available meteorological station inside the sub-basin, Ghasreghand, is used for downscaling and prediction exercises. The daily rainfall data of this station from 1977 to 2000 has been used as a predictand. Observed data of the effective signals in the same period as the rainfall are received from the NCEP (1977-2000) web site, named predictors.

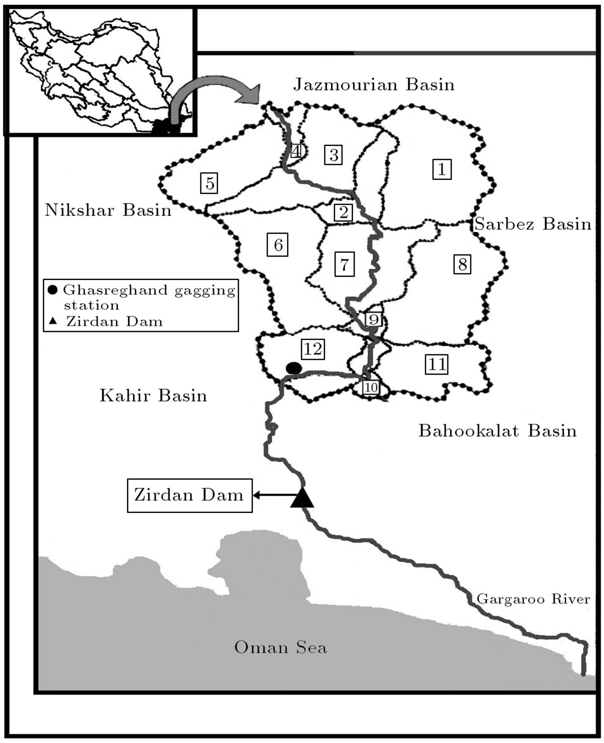

Figure 1. Kajoo River in the South Baloochestan watershed would indicate that the predictor-predictand correlation could be chance related.

LONG LEAD RAINFALL PREDICTION METHODS

Statistical Down-Scaling Model (SDSM)

The first downscaling model introduced here is a multiple regression based method, which is referred to as the Statistical Down-Scaling Model (SDSM). The analytical base of this model is formulated by Wilby et al. [1]. The developed formulation consists of two steps. In the first step, it is determined, whether or not rainfall occurs on each day:

$$\omega_i = \alpha_0 + \sum_{j=1}^{i} \alpha_j \hat{u}_i^{(j)}, \tag{1}$$

where ω_i is the candidate in a random process generator that indicates a state of rainfall occurring or not; accruing on day i, $\hat{u}_i$ is the normalized predictor on day i; and α_j is the estimated regression coefficient. Precipitation in day i occurs if $\omega_i \leq r_i$, where r_i is a stochastic output from a linear random- number generator.

The value of rainfall on each rainy day is estimated in the second step, using the z-score as follows:

$$Z_i = \beta_0 + \sum_{j=1}^{n} \beta_j \hat{u}_i^{(i)} + \varepsilon, \tag{2}$$

$$Z_i = \phi^{-1}[F(y_i)], \tag{3}$$

where Z_i is the z score calculated from the estimated regression coefficients, β_j, and the normally distributed stochastic error term, ε. The rainfall value is then calculated from the cumulative distribution function, ϕ, of the empirical distribution function, $F(y_i)$, of the daily rainfall occurrence and amounts. The predictors are standardized by subtracting the climatological mean and dividing it with standard deviations over the calibration period. During downscaling with the SDSM, a multiple linear regression model is developed between a few selected large-scale predictors and the local scale predictands such as rainfall.

The structure and operation of SDSM can be divided into five distinct tasks, including:

1. Preliminary screening of potential downscaling predictor variables;
2. Assembly and calibration of SDSM;
3. Synthesis of ensembles of current weather data using observed predictor variables;
4. Generation of ensembles of future weather data using GCM-derived predictor variables;
5. Diagnostic testing/analysis of observed data and climate change scenarios (Figure 2).

In SDSM, the parameters of the regression model are estimated using the efficient dual simplex algorithm. Large-scale relevant predictors (among those presented in Table 1) are selected using correlation, partial correlation and P value analysis. Also, considering physical sensitivity between selected predictors and predictands for the site in question, the correlation statistics and P values indicate the strength of association between the predictor and predictand. Higher correlation values imply a higher degree of association and lower P values indicate a better chance for association between variables. It is important to know that the P values of less than 0.05 as a threshold do not necessarily mean that the result can be statistically significant. On the other hand, the high P value would indicate that the predictor-predictand correlation could be chance related.

As rainfall values cannot take negative amounts, they are modeled as a conditional process in which local rainfall amounts are correlated with the occurrence of wet-days which in turn, are correlated with regional-scale atmospheric predictors. A wet day is defined as a day with rainfall more than 0.3 mm. During calibration of the model, the mean of downscaled daily rainfall is adjusted by a bias correction and a variance inflation factor to force the model to replicate the observed

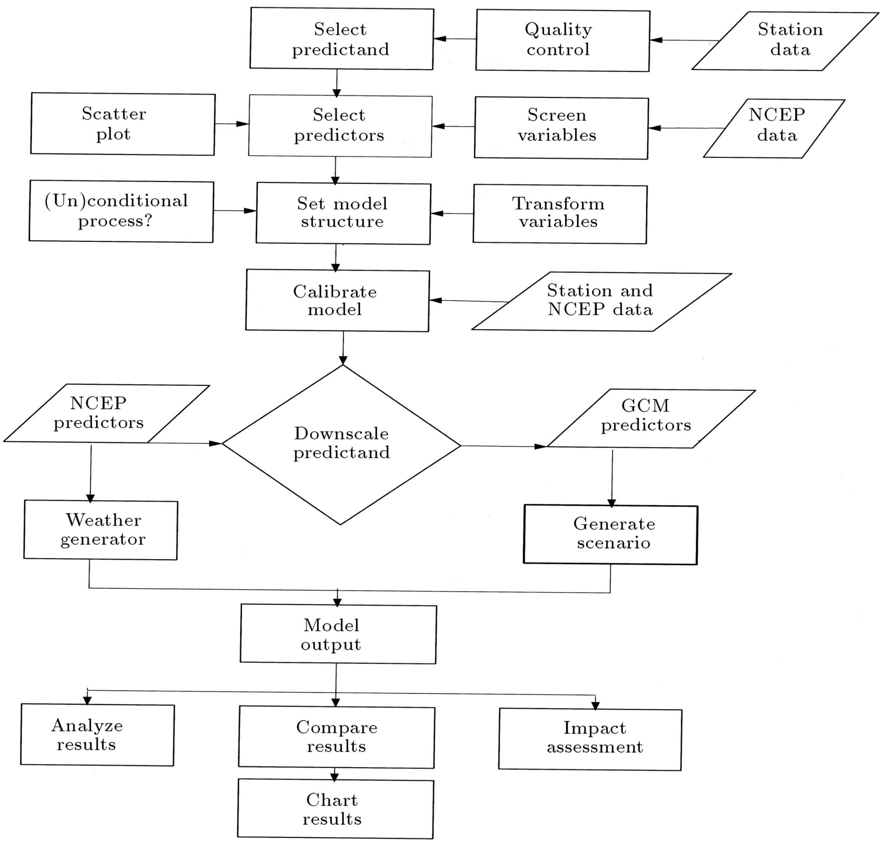

Figure 2. SDSM climate scenario generation process [1].

Table 1. Large-scale predictors for the meteorological prediction with SDSM.

Predictor Variable	Description
Temp	Mean temperature
Mslp	Mean sea level pressure
p_u	Zonal velocity component near surface
p5_u	Zonal velocity component at 500 hPa height
p8_u	Zonal velocity component at 850 hPa height
p_v	Meridional velocity component near surface
p8_v	Meridional velocity component at 850 hPa height
p_z	Vorticity
p_zh	Divergence near surface
p5zh	Divergence at 500 hPa height
p8zh	Divergence at 850 hPa height
p500	500 hPa geopotential height
p850	850 hPa geopotential height
s500	Specific humidity at 500 hPa height
r850	Relative humidity at 850 hPa height
Shum	Near surface specific humidity
Rhum	Near Surface relative humidity

data. The amounts of these variables are determined by trial and error. Bias correction compensates for any tendency to overestimate/underestimate the mean of downscaled variables. Variance fluctuation changes the variance of downscaled daily weather variables by adding or reducing the amount of 'white noise' applied to the regression model estimates of the local process, in order to better accord with observations. Use of this stochastic (random) component also enables the SDSM regression model to produce multiple ensembles of downscaled weather variables.

Rainfall predictions must be tested. There are several indicators that can be used for this purpose. Here, two indicators, namely Root Mean Square Error (RMSE) and Mean Absolute Error (MAE) have been used. These indicators are calculated as follows:

$$\mathrm{RMSE} = \sqrt{\frac{\sum\limits_{m=1}^{n} (X_p - X_o)^2}{n}}, \tag{4}$$

$$\mathrm{MAE} = \frac{\sum\limits_{m=1}^{n} |X_p - X_o|}{n}, \tag{5}$$

where X_P is simulated data, X_o is observed data and n is the number of data.

Artificial Neural Network (ANN)

The second model for forecasting rainfall in this study is the Artificial Neural Network (ANN). This model is a non-linear regression type in which a relationship is developed between a few selected large-scale atmospheric predictors and basin scale meteorological predictands. An ANN that emulates a biological neural network structure distributes the computation to small and simple processing units called artificial neurons or nodes. Neurons with similar characteristics are arranged into a layer. A layer can be seen as a group of neurons, which are connected to other layers or the external environment, which have no interconnections.

Artificial Neural Networks can be trained using a Back Propagation Algorithm (BPA) during supervised learning. The back propagation algorithm provides a way to calculate the gradient of the error function efficiently, using the chain rule of differentiation [18]. In this algorithm, network weights are moved along the negative gradient of the performance function.

Multilayer Perceptron Network

Multilayer Perceptron (MLP), also called a feed forward network, implements mappings from the input pattern space to the output space. MLP is a static network in which information is transmitted through the connections between its neurons, forward only, and the network has no feedback or memory, which covers its initial and past states. MLP can be trained with the standard back propagation algorithm. Figure 3 shows a typical three layer MLP (without recurrent connections) which, mathematically, can be expressed as:

$$a_j^1(t) = F\left(\sum_{i=1}^{R} w_{j,i}^1 p_i(t) + b_j^1\right), \qquad 1 \le j \le S_1, \tag{6}$$

$$a_k^2(t) = G\left(\sum_{j=1}^{S_1} w_{k,j}^2 a_j^1(t) + b_k^2\right), \qquad 1 \le k \le S_2, \tag{7}$$

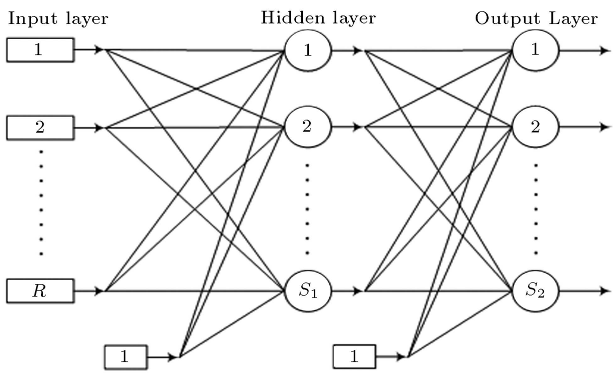

Figure 3. MLP model structure.

where t denotes a discrete time, R is the number of input signals and S_1 and S_2; the numbers of hidden and output neurons, respectively; $w^1_{i,j}$ and $w^2_{k,j}$ are the weight matrices of hidden and output layers; b^1_j and b^2_k are the bias vectors of hidden and output layers; $p_i(t)$ is the input matrix; and a^1 and a^2 are the output vectors of the hidden and output layers. F and G are the activation functions of the hidden and output layers, respectively.

Temporal Neural Networks

Static networks only process input patterns that are spatial in nature, i.e. input patterns that can be arranged along one or more spatial axes such as a vector or an array. In many tasks, the input pattern comprises one or more temporal signals, as in speech recognition, time series prediction and signal filtering [18]. The dynamic behavior of the system or the recurring (feedback) connections, as recurrent neural networks, should be considered.

RAINFALL PREDICTION RESULTS

SDSM

A combination of predictors has been selected for long lead rainfall prediction because of their maximum correlation with daily rainfall. Correlation coefficients between selected predictors and daily rainfall have been presented in (Table 2). This table also reports the partial correlation and P value between the predictors and rainfall that help identify the amount of explanatory power for each predictor.

As the distribution of daily rainfall is skewed, a fourth root transformation is applied to the original series to convert them to a normal distribution, and then they are used in regression analysis. The model is structured as a monthly model for rainfall downscaling, in which case twelve regression equations are derived for twelve months. The model is calibrated and validated using 15 years (1976-1990) and 9 years (1991-1999) of daily data of predictors and predictands, respectively, and one ensemble of downscaled daily rainfall has been generated.

The reason for choosing 24 years of one ensemble data for calibration and validation purposes is that a longer data set is capable of representing the true climatic condition for the site in question, including less frequent climate events. For this purpose, one of the predicted scenarios for future climate variation such as HadCM3 (Second Hadley Centre Coupled Ocean-Atmosphere GCM) is used as a model input signal and the rainfall is predicted. Figure 4 shows the observed rainfall in 5 months (December, January, February, March, April) and the predicted rainfall by using developed weather generator variables (HadCM3) for the study area. The MAE and RMSE for the predicted rainfalls are 5 and 6 mm, respectively, which are acceptable considering the magnitude of the observed rainfall.

ANN Models

Climate signals that seem to be significant on Iran's rainfall are shown in Figure 5. The correlation between the SLP of each area with different lags between 1 to 12 months and the total 5 month precipitation of the study area have been computed. The results of the best correlation are presented in Table 3.

Results show that the following indices can be considered as indicators of winter precipitation and runoff in the study area:

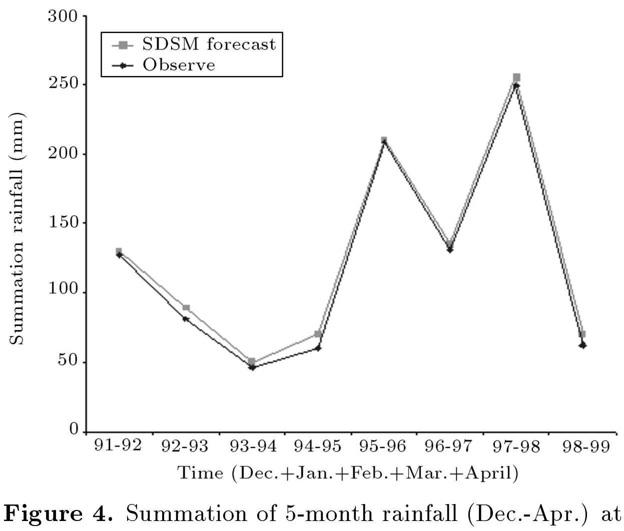

Figure 4. Summation of 5-month rainfall (Dec.-Apr.) at the study area for downscaled data from GCM (HadCM3) predictors and the observed data (1991-1999).

Table 2. The results of selected predictors for the study area.

Predictor	Correlation Coefficient Between Weather-Variables and Rainfall	P_v (P Value)	P_r (Partial)
Relative humidity at 850 hPa height	0.42	0.002	0.48
Near surface specific humidity	0.45	0.000	0.5
Near surface relative humidity	0.40	0.000	0.44

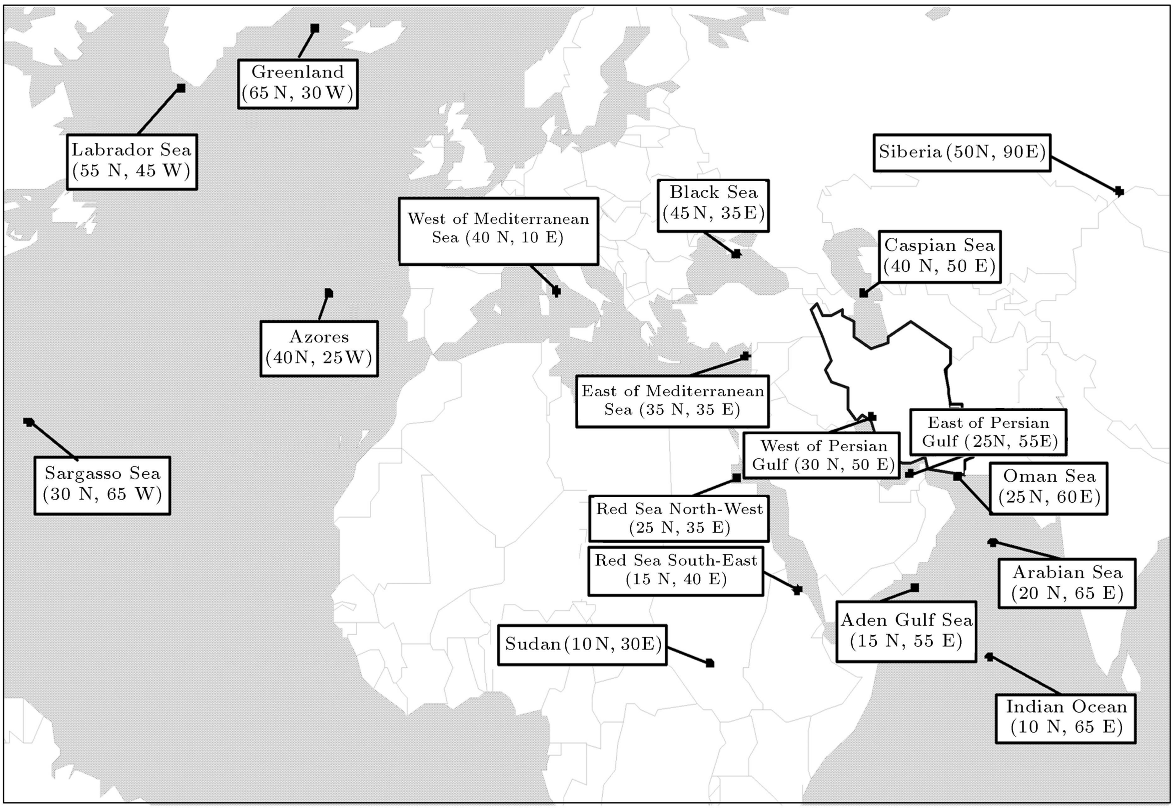

Figure 5. Effective points on Iran climate variations [19].

Table 3. Best correlation for each studied climate signal and (December-April) rainfall.

Study Area	R^2
Δ SLP of Greenland-East of the Mediterranean	0.46
Δ SLP of Greenland-Azores	0.45
Δ SLP of Greenland-West of the Mediterranean	0.49
SLP of Black Sea	0.38

1. SLP difference between Greenland and the east of the Mediterranean Sea;
2. SLP difference between Greenland and Azores;
3. SLP difference between Greenland and the west of the Mediterranean Sea;
4. SLP of the Black Sea.

The time series of the selected SLP differences and SLP as winter rainfall predictors are used to develop an ANN model for winter precipitation predictions in the study area. Twenty years data from 1971 to 1990 are used to calibrate and 9 years (1991-1999) are used to validate the ANN model. Several ANN models with different training steps and neurons and returns have been developed and their performance has been compared. To quantify the ANN performance, RMSE and MAE criteria have been calculated. Finally, a 3-layer ELMAN model with 2 neurons and the least RMSE and MAE has been selected (Table 4). Results of winter rainfall forecasting with selected MLP and ELMAN models have been shown in Figure 6. It can be observed that both of the models have a similar performance and can predict rainfall fairly well.

COMPARING RESULTS

For a better understanding of the models' performance and through comparing their results, an error tolerance is defined. Here, error means the difference between observed and predicted values divided by the observed value as:

$$\text{Error}_i = \frac{|\text{obs}_i - \text{pre}_i|}{\text{obs}_i}, \tag{8}$$

Table 4. Errors of different ANN models used in predicting rainfall as 5-month summation.

ANN Model	**Variable**	**MAE (mm)**	**RMSE (mm)**
MLP	Winter rainfall	37.19	32.13
ELMAN	Winter rainfall	32.87	25.71

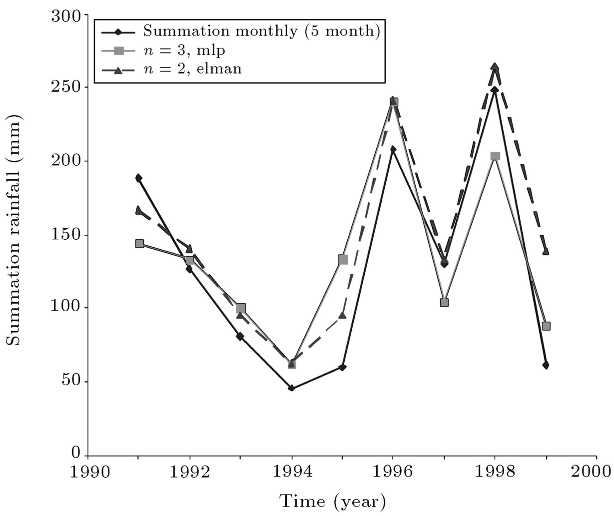

Figure 6. Comparative forecasted and observed rainfall.

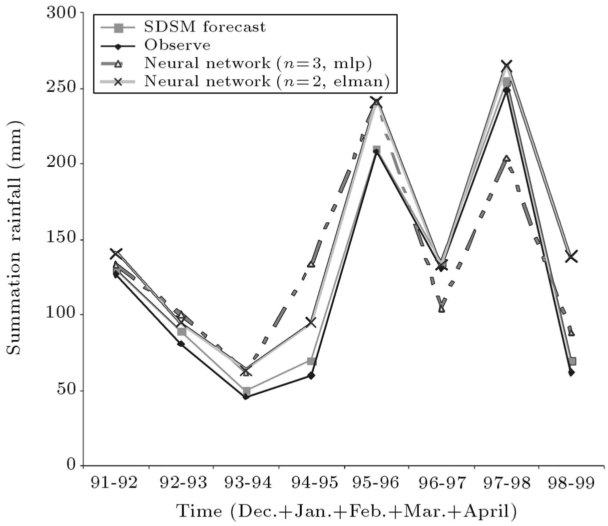

Figure 7. Comparing results of long lead rainfall prediction models including SDSM and ANN Models.

where obs$_i$ and pre$_i$ stand for observed and predicted rainfall, respectively. The percentages of predicted rainfall found in defined error tolerance ranges are calculated and summarized in Table 5 for the selected models.

As seen in Table 5, the performance of the SDSM model is much better than ANN models due to the consideration of local signals in SDSM model, which are more effective than large scale global signals. ANN models overestimated the amount of rainfall, as shown in Figure 7.

CONCLUSION

Long lead rainfall predictions can play an important role in water resources management and operation. It has been investigated that large scale signals can be used for long lead rainfall predictions. Many methods have been developed for utilizing climate signals for the prediction of rainfall for different time scales. One of the commonly used models in this field is the ANN model. In recent years, downscaling models have been developed to predict the amount of rainfall in local scales and in smaller time steps. One of the models used to downscale rainfall prediction is SDSM, which has a statistical base. In this paper, these two models are applied for long lead rainfall prediction in southwestern Iran. In ANN models, SLP and ΔSLP, and in the SDSM model with relative humidity at 850 hPa height, near surface specific humidity and near surface relative humidity have been selected as predictors for long-lead rainfall prediction. These climate signals have been selected with the highest R-Squared values.

Table 5. Error tolerances for selected long lead rainfall prediction models (percentage of occurrence).

Models	**Percentage of Predictions with Error (%) Less than**		
	< 10	**< 20**	**< 30**
ELMAN	22	67	67
MLP	11	45	67
SDSM	75	100	100

Comparing the results, we can conclude that the ELMAN model is better than the MLP model. In this model, MAE is 33 mm and RMSE is 26 mm; 67% of forecast errors are within 30% of the observed values. In the SDSM model, these values are 5mm, 6 mm and 100%, respectively. 75% of forecasts using the SDSM model are within 10% of the observed values. It is concluded that the SDSM performance is better than the ANN model, even though, in comparison, it is a more data intensive model than ANN.

ACKNOWLEDGMENT

This study is a part of a flood management project entitled "Flood plane zoning downstream of Kajoo nad Kariani River basins and the design of flood warning system" sponsored by the Sistan and Baloochestan water Authority.

REFERENCES

1. Wilby, R.L., Dawson, C.W. and Barrow, E.M. "SDSM–a decision support tool for the assessment of regional climate change impacts", *Environ. Model Software.*, **17**, pp. 147-159 (2002).
2. Harpham, C. and Wilby, R. "Multi-site downscaling of heavy daily precipitation occurrence and amounts", *J. Hydrology*, **312**, pp. 235-255 (2005).
3. Massah, A. "Risk assessment for climatological variety and it's effects on water resources, case study : Zayan-

derud basin", PhD thesis, Tarbiat Modares University (2006).

4. Nazemosadat, M.J. "Persian Gulf sea surface temperature as a drought diagnostic for southern parts of Iran", *Drought News Network*, **10**, pp. 12-14 (1998).
5. Nazemosadat, M.J. and Cordery, I. "On the relationships between ENSO and autumn rainfall in Iran", *International Journal of Climatology*, **1**, pp. 17-62 (2000).
6. Karamouz, M., Razavi, S. and Araghinejad, S. "Long-lead rainfall forecasting, using dynamic neural networks: Case study of western part of Iran", *Proceeding of the First Iran-Korea Joint Workshop on Climates Modeling*, **92**, pp. 100 (2005).
7. Karunanithi, N., Grenney, W.J., Whithley, D. and Bovee, K. "Neural networks for river flow prediction", *J. Comp. Civ. Engrg., ASCE*, **8**(2), pp. 201-220 (1994).
8. Smith, J. and Eli, R.N. "Neural-network models of rainfall-runoff process", *J. Water Resour. Plng. and Mgmt., ASCE.*, **121**(6), pp. 499-508 (1995).
9. Mair, H.R. and Dandy, G.C. "Use of artificial neural networks for prediction of water quality parameters", *Water Resour. Res.*, **32**(4), pp. 1013-1022 (1996).
10. Shamseldin, A.Y. "Application of neural network technique to rainfall-runoff modeling", *J. Hydro.*, **199**, pp. 272-294 (1997).
11. Clair, T. and Ehrman, J.M. "Using neural networks to assess the influence of changing seasonal climate in modifying discharge, dissolved organic carbon, and nitrogen export in eastern Canadian rivers", *Water Resour., Res.*, **35**(4), pp. 1191-1197 (1998).
12. French, M.N., Krajewski, W.F. and Cuykendall, R.R. "Rainfall forecasting in space and time using a neural network", *J. Hydrol.*, **137**, pp. 1-31 (1992).
13. Toth, E., Brath, A. and Montanari, A. "Comparison of short-term rainfall prediction models for real-time flood forecasting", *J. Hydro.*, **239**, pp. 132-147 (2000).
14. Anmala, J., Zhang, B. and Govindaraju, R.S. "Comparison of ANNs and empirical approaches for predicting watershed runoff", *J. Water Resour. Plng and Mgmt., ASCE*, **126**(3), pp. 156-166 (2000).
15. Gowariker, V., Tapliyal, V., Sarker, R.P., Mandal, G.S. and Sikka, D.R. "Parametric and power regression models: New approach to long range forecasting of monsoon in India", *Mausam*, **42**, pp. 115-122 (1989)
16. Tapliyal, V. "Long range prediction of summer monsoon rainfall over India: Evolution and development of new models", *Mausam*, **91**, pp. 339-346 (1990).
17. Sahai, A.K., Grrimm, A.M., Satyan, V. and Pant, G.B. "Long lead prediction of Indian summer monsoon rainfall from global SST evolution", *Climate Dyn.*, **20**, pp. 855-863 (2003).
18. Bose, N.K. and Liang, P., *Neural Network Fundamentals with Graphs, Algorithms, and Application*, McGraw-Hill, New York (1996).
19. Karamouz, M. "Long-lead rainfall forecasting using large scale climate signals", Report to Iranian Meteorological Organization (2005).