REVIEW ARTICLE

A simple approach for improving spatial interpolation of rainfall using ANN

C. Sivapragasam · V. M. Arun · D. Giridhar

Abstract Interpolation of hydrological variables such as rainfall, ground water level, etc. is necessary to arrive at many engineering decisions. This study suggests an innovative and yet simple idea to effectively interpolate the rainfall at non-sampling locations using artificial neural network (ANN). The method has been demonstrated in the Tamirabarani basin of Tamil Nadu State (India), where rain gauge information for 18 rain gauge stations is available. With the help of land use map of the basin and also the proximity of rain gauge stations to each other in the neighbourhood, the most appropriate input variables for ANN are designed and used for training ANN to estimate rainfall in the unknown stations. It is also interesting to note that with appropriate selection of input variables for training ANN, one of the major lacunae of ANN (to have sufficiently lengthy training records) can be suitably addressed. ANN results are compared with Kriging method. Further, the proposed method is applied to improve the prediction of inflow to Ramanadhi reservoir in Tamirabarani basin, and the results seem to be very encouraging.

1 Introduction

Surface water resources are under increasing stress in view of severe regional imbalances in availability of water and drought conditions in India. Because of increasing demand on water supply from various sectors, timely and analytical information is needed by water managers to assess current and long-term water resource potential in the region. This information is critical to management of water supply and to avoid potential adverse affects on the hydrologic system.

Southern basins in Tamil Nadu are very much drought prone. The sustainable planning of water resources in such region requires hydrological information, the most crucial of which is the estimation of spatial and temporal variability of rainfall (Wilk et al. 2006; Goncalves and Alpuim 2006). This necessarily depends on rain gauge network in a given basin. Often the number of rain gauges in a basin is not adequate to arrive at average or mean precipitations used in any design study due to the large space and time variability of rainfall. Some authors (e.g., Tsintikidis et al. 2002) have attempted to identify possible sites for the placement of additional gauges in order to reduce the precipitation interpolation error using historical hourly data.

Conventionally, however, stochastic interpolation methods such as Kriging are widely recognized and used as standard methods for interpolation of hydrological variables such as groundwater level, precipitation, etc. For example, Vijayakumar and Remadevi (2006) used Kriging for the spatial analysis of groundwater levels and established its superiority over the inverse square distance. Jinfeng (2006) conducted a case study on the impacts of land use/cover change (LUCC) on the temporal and spatial variability of the groundwater level in an arid oasis in the Sangong River Watershed by using geographical information system (GIS), remote sensing (RS) and Kriging. Similarly, Shi et al. (2007) compared IDW, Ordinary Kriging and Simple Kriging interpolation of geostatistics analyst to interpolate rainfall data. Among other important conclusions, they observed that Exponential model of semivariogram has the highest precision than others. Yan-bing (2002) compared five different semivariogram models for monthly rainfall interpolation in their study area.

C. Sivapragasam (✉) · V. M. Arun · D. Giridhar
Department of Civil Engineering, Kalasalingam University,
Krishnankoil, Srivilliputtur, Tamil Nadu 626 190, India
e-mail: sivapragasam25@gmail.com

Although Kriging has been widely used, it is being recognized that traditional Kriging has a major limitation due to the need for a mathematical function for a semivariogram that might fit the surface to be interpolated. Consequently, some researchers have used and demonstrated the advantage of the universal function approximates such as artificial neural network (ANN) as a replacement to fitted authorized semivariogram model within ordinary Kriging (Teegavarapu 2007). Chowdhury et al. (2010) demonstrated that ANNs can also be used to map with noticeably higher accuracy than Kriging the complex and seemingly erratic spatial pattern of groundwater contamination provided that reasonable data preprocessing and exploratory data analysis are performed.

The present study aimed to investigate the efficacy of ANN to interpolate monthly average rainfall in non-sampling locations with the help of rainfall measured in the neighbouring locations. The idea of using only neighbouring rain gauge is derived from one of the commonly used non-parametric interpolation methods, the 'nearest neighbour' method, which is based on minimum Euclidian distance (Brath et al. 2002). Karlsson and Yakowitz (1987) used this method for studying rainfall runoff process. Szolgay et al. (2009) compared the performance of three interpolation methods namely nearest neighbour method, Kriging and inverse distance weighting (IDW) in producing maps of 2-year and 100-year daily precipitation as extreme rainfall indices. In this study, though directly nearest neighbour method is not used, an emphasis is laid in selection of appropriate neighbourhood input variables for training ANN architecture. The efficacy of the proposed approach is demonstrated through an application in improving the inflows to Ramanadhi reservoir in the Tamirabarani basin. The next section briefly introduces Kriging and ANN techniques. This is followed by a description of the study area. Subsequently, the proposed methodology is discussed in detail, results analysed and conclusions drawn.

2 Artificial neural network

The ANN is briefly introduced in this section. Multilayer feed forward network with Back Propagation learning algorithm is one of the most popular neural network architectures, which has been deeply studied and widely used in many fields. Typically a neural network consists of three layers: (1) an input layer; (2) an output layer and (3) an intermediate or hidden layer. The input vectors are $D \in R^n$ and $D = (X_1, X_2,\ldots, X_n)^T$; the outputs of q neurons in the hidden layer, $Z \in R^q$, $Z = (Z_1, Z_2,\ldots, Z_q)^T$, and the outputs of the output layer are $Y \in R^m$, $Y = (Y_1, Y_2,\ldots,$

$Y_m)^T$, assuming that the weight and the threshold between the input layer and the hidden layer are w_{ij} and y_j, respectively, and the weight and the threshold between the hidden layer and output layer are w_{jk} and y_k, respectively; the outputs of each neuron in a hidden layer and output layer are:

$$Z_j = f\left(\sum_{i=1}^{n} w_{ji}X_i - \theta_j\right), \qquad (1)$$

$$Y_k = f\left(\sum_{j=1}^{q} w_{kj}Z - \theta_k\right), \qquad (2)$$

where $f()$ is a transfer function, which is the rule for mapping the neuron's summed input to its output, and by a suitable choice, is a means of introducing a non-linearity into the network design, and θ is the threshold. One of the most commonly used functions is the sigmoid function, and it is monotonically increasing and ranges from 0 to 1. Details on ANN can be found in Karunanithi et al. (1994). In this study, NEUROSHELL software is used for implementing ANN (Neuroshell2).

3 Kriging interpolation technique

Kriging is essentially a method of estimation. The assumption here is that there is a local structure in the sampling domain, and the accuracy of the estimated values depends on how this is accounted for. Moreover, a weighted sum of the observations is assumed to provide the best estimate. Thus, Kriging is a method of determining the optimal weights. The most commonly applied form of Kriging uses a "semivariogram", which is a measure of spatial correlation between pairs of points describing the variance over a distance or lag h. Weights change according to the spatial arrangement of samples. The linear combination of weights is as follows:

$$Y_i = \sum \lambda_i y_i, \qquad (3)$$

where Y_i estimate of Kriging, y_i the variables evaluated in the observation locations and λ_i Kriging weights.
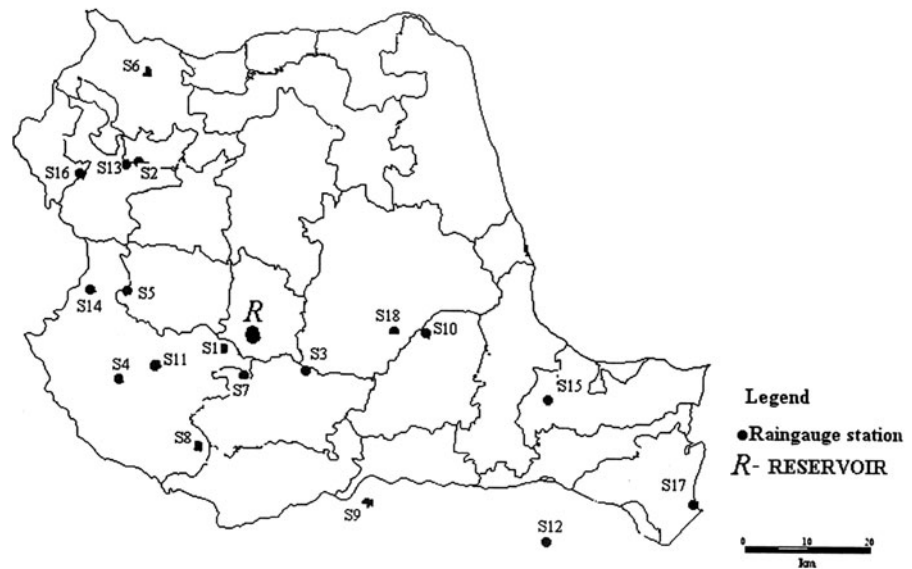
The semivariogram is formulated as follows:

$$\gamma(h) = \frac{1}{2}E\big(y_{(x)} - y_{(x+h)}\big)^2, \qquad (4)$$

where $\gamma(h)$ semivariogram, dependent on distance h; $(x, x + h)$ pair of points with distance vector h; $y(x)$ regionalized variable y at point x; $y(x) - y(x + h)$ difference of the variable at two points separated by h; E mathematical expectation.

Details on Kriging can be found in Kitanidis (1997) and Biau et al. (1999).

Fig. 1 Location of rain gauge stations in the basin

4 Study area

Tamirabarani River has its origin in Western Ghats near Senkottai district of Tamil Nadu State (India) with a catchment area of about 6,000 km^2. There are 18 rain gauge stations in the catchment (refer Fig. 1). Six years of monthly rainfall records are collected from these stations for a time period from January 1979 to December 1984. Out of 18 stations, stations S3 (Cheranmadevi), S6 (Kadayanallur) and S12 (Sattankulam) are used for validation of rainfall interpolation.

The location of Ramanadhi reservoir is also shown in the map. Monthly inflows to Ramanadhi reservoir are available from 1974 to 1986.

5 Interpolation of rainfall with ANN

The following four different models based on ANN for rainfall interpolation are considered in detail.

(a) ANN with only three neighbouring stations (ANN_I): The methodology adopted for interpolation of rainfall is shown in Fig. 2. For the non-sampled location, rain gauge stations in the proximity and conforming to the same land use are selected (Table 1). The justification for use of rain gauges from the same land use can be understood from the fact that rainfall in the same land use (particularly for small- and medium-sized catchments) is expected to have same value. Figure 3 shows the general land use of the basin.
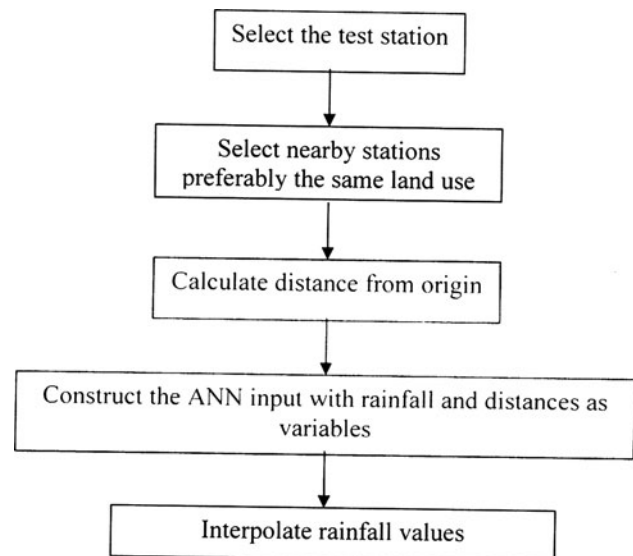

Fig. 2 Flow chart showing methodology for rainfall interpolation

ANN_1:

$$I = f(R_1, R_2, R_3), \tag{5}$$

where R_1 monthly rainfall of first rain gauge in the proximity, R_2 monthly rainfall of second rain gauge in the proximity and R_3 monthly rainfall of third rain gauge in the proximity.
The output consists of rainfall at the non-sampled location.

(b) ANN with three neighbouring stations and their distances from origin as inputs (ANN_II): The easting

Table 1 Proximity stations using land use map

Unknown station	S3			S6			S12		
Proximity stations	S1	S7	S18	S2	S13	S16	S9	S15	S17
Land use details	Cultivable			Hilly region			Dry crops		
Distance (km)	62.3	61.9	87.5	79.6	77.3	73.8	72.4	106.2	125.5

and northing of rain gauge stations are determined from an arbitrary origin, and their distances from the origin are tabulated in Table 2. Since rainfall varies spatially, it is expected that incorporating distance to a particular station from a reference origin (as one of the input variables in ANN) may help in improving the prediction.

Functionally, the model can be represented as:
ANN_II:

$$I = f(R_1, D_1, R_2, D_2, R_3, D_3, D_{ns}),\qquad(6)$$

where R_1 monthly rainfall of first rain gauge in the proximity, D_1 distance to the first rain gauge in the proximity from an assumed origin, R_2 monthly rainfall of second rain gauge in the proximity, D_2 distance to the second rain gauge in the proximity from an assumed origin, R_3 monthly rainfall of third rain gauge in the proximity, D_3 distance to the third rain gauge in the proximity from an assumed origin, D_{ns} distance of non-sampled location in the study.

The output consists of rainfall at the non-sampled location.

(c) ANN with all stations (ANN_III): This model is an extension of ANN_I wherein all the stations in the catchment form input vectors to the ANN model. Therefore, functionally, the model can be represented as:

ANN_III:

$$I = f(R_1, R_2, \dots R_n),\qquad(7)$$

where R_1 monthly rainfall of first rain gauge, R_2 monthly rainfall of second rain gauge, R_n monthly rainfall of nth rain gauge and n number of stations in the catchment.

The output consists of rainfall at the non-sampled location.

(d) ANN with all stations and their distances from origin as inputs (ANN_IV): This model is an extension of ANN_II wherein the all rain gauges stations and their distances from an origin is given as input to the model. Functionally, the model can be represented as:

ANN_IV:

$$I = f(R_1, D_1, R_2, D_2, \dots R_n, D_n, D_{ns}),\qquad(8)$$

where R_1 monthly rainfall of first rain gauge, D_1 distance to the first rain gauge from an assumed origin, R_2 monthly rainfall of second rain gauge, D_2 distance to the second rain gauge from an assumed origin, R_n monthly rainfall of nth rain gauge, D_n distance to the nth rain gauge from an assumed origin,

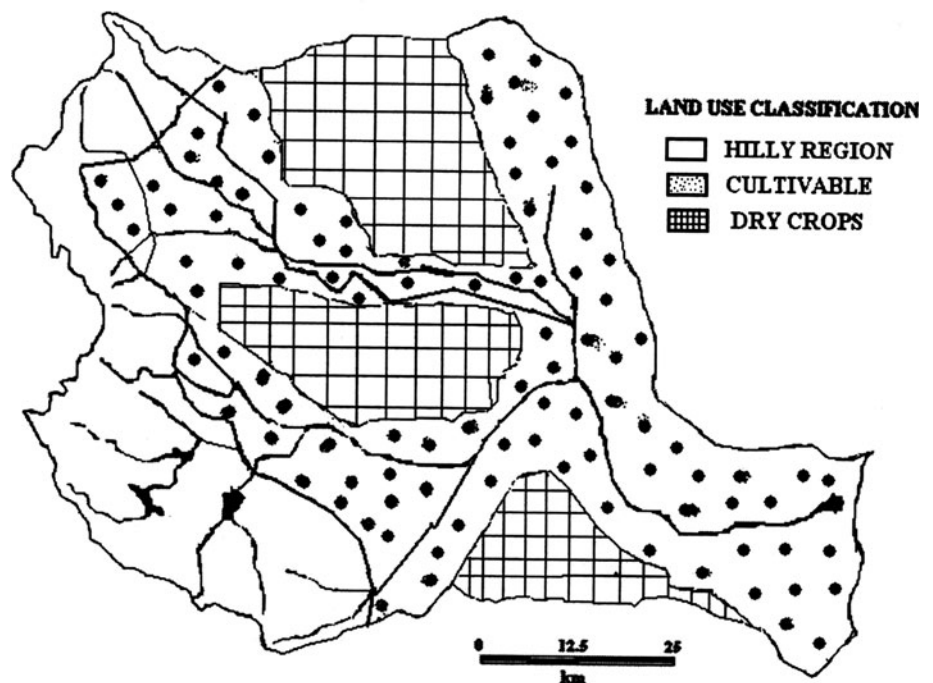**Fig. 3** Land use of Tamirabarani basin

Table 2 Co-ordinates information for 18 stations

Stations	Stations name	X (km)	Y (km)	Distance (km)
S1	Ambasamuthiram	46.5	41.5	62.3
S2	Ayyakudi	34	72	79.6
S3	Cheranmadevi	60	37.5	70.9
S4	Dampcamp	30	36	46.8
S5	Kadananadhi	30.5	50	58.5
S6	Kadayanallur	36	86.7	93.8
S7	Kannadian	50	36.6	61.9
S8	Manimudhar	42.6	24.6	49.2
S9	Nanguneri	71	14.4	72.4
S10	Palayamkottai	81.3	44.7	92.7
S11	Papanasam	36	38	52.3
S12	Sattankulam	100	8.7	100.5
S13	Shenkottai	30.5	70.7	77.3
S14	Sivagiri	25.3	50	56.1
S15	Srivaigundam	101.3	32	106.2
S16	Thenkasi	23.4	70	73.8
S17	Tiruchendur	124.7	14.7	125.5
S18	Tirunelveli	75.3	44.7	87.5

D_{ns} distance of non-sampled location in the study and n number of stations in the catchment.

The output consists of rainfall at the non-sampled location.

The back-propagation algorithm is used for training the ANN architecture consisting of one hidden layer. Out of 18 stations in the catchment, 10 are used for training, 5 for testing and 3 for validation. Now for each of these stations, input vectors are constructed for all the ANN models.

The results are compared and tabulated for the months of January–March for the year 1984 in Table 3.

As can be observed from the table, the RMSE for the month of January by ANN_II is only 9 mm as against 12.7 by ANN_I. On an average for all the three months, ANN_II estimates are better than ANN_I. It is interesting to note that by including the distance to stations as one of the inputs, the ability of ANN to map the complexity in variation of rainfall over the basin is considerably increased despite the fact that a very small sample of data is used for training. This is reflected in the difference in RMSE values in the two cases.

Further, the comparison of ANN_III and ANN_IV clearly shows that it is better to incorporate all the rain gauges stations in the catchment for better estimation of rainfall. However, if the 'distance from origin' variables is dropped (ANN_III), even with inclusion of all stations do not improve the interpolation.

6 Interpolation of rainfall with Kriging

Here, again two different models are considered, namely Kriging with only three immediate neighbourhood stations as used for ANN_I and ANN_II interpolation (Kriging_I) and Kriging with all the stations in the basin as done in conventional methods (Kriging_II). This is illustrated in Table 3, where the results of Kriging analysis are compared with ANN_I and ANN_II. As seen from the table, RMSE for Kriging is significantly more than ANN models. Further, it can be observed that inclusion of all the stations in the Kriging model only marginally improves the result when compared with Kriging_I.

Table 3 ANN results compared with Kriging

	January			February			March		
Stations	S3	S6	S12	S3	S6	S12	S3	S6	S12
Original	172	181	174	57	55	32	20	10	38
Kriging_I	187	129	177	30	23	52	27	53	30
Kriging_II	180	179	181	30	28	29	33	46	32
ANN_I	188	196	176	47	20	48	18	18	63
ANN_II	172	171	161	43	36	43	33	30	41
ANN_III	163	141	142	23	23	32	17	24	20
ANN_IV	167	166	167	42	56	31	24	11	39
RMSE	ANN_I: 12.7	Kriging_I: 33.0		ANN_I: 22.9	Kriging_I: 26.4		ANN_I: 15.2	Kriging_I: 25.6	
	ANN_II: 9.2	Kriging_II: 7.33		ANN_II: 15.1	Kriging_II: 22.44		ANN_II: 14.1	Kriging_II: 22.34	
	ANN_III: 28.97			ANN_III: 27.12			ANN_III: 13.12		
	ANN_IV: 9.1			ANN_IV: 9.04			ANN_IV: 2.39		

All values are in millimetres

RMSE root mean square error

7 Inflow prediction to Ramanadhi reservoir

Here, an attempt is made for improving the prediction of inflows to reservoir by estimating rainfall value in the vicinity of the reservoir using ANN_II (as discussed above). The methodology is illustrated in Fig. 4. For the inflow prediction, first the reservoir for which inflow prediction is required is identified along with a few more sites in the vicinity of the reservoir. In this study, up to five sites in the vicinity are considered, and it is found that best prediction comes for a neighbourhood network of three sites. For these three sites, monthly rainfall is estimated using ANN_II model, which will form a part of input variables for inflow prediction. The input vector for inflow prediction can be mathematically depicted as:

$$I_{i+1} = f(I_{i-1}, I_i, R_{1i}, R_{2i}, R_{3i}), \tag{9}$$

where I_i and I_{i-1} are antecedent inflows. R_{1i}, R_{2i} and R_{3i} are antecedent rainfalls in the neighbouring sites (as estimated using ANN_II).

The ANN is trained, and the results of inflow prediction for the year 1984 are tabulated in Table 4 and Fig. 5. As seen from the table, there is a drastic improvement in the RMSE value with rainfall incorporated as input variable.
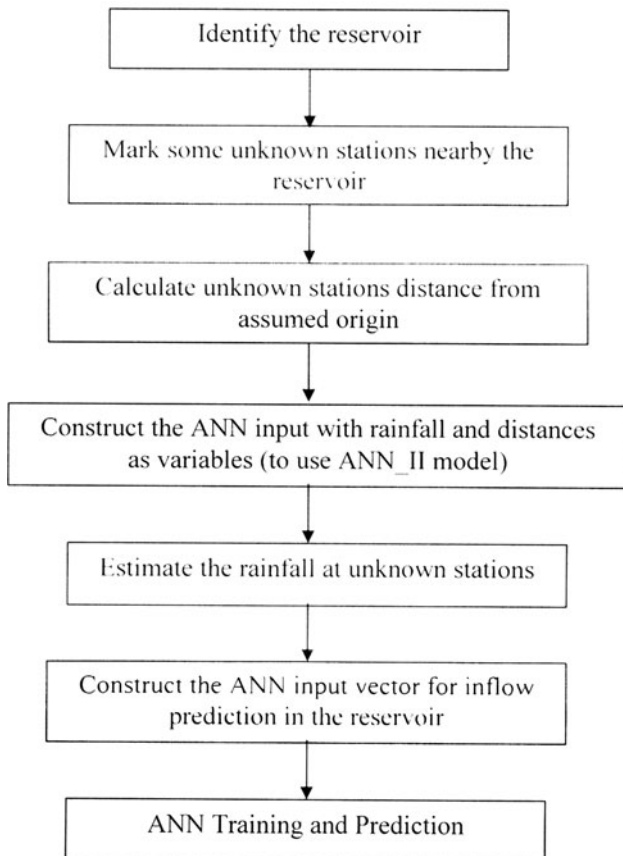

Fig. 4 Flow chart showing methodology for inflow prediction

Table 4 Inflow predictions for Ramanadhi reservoir

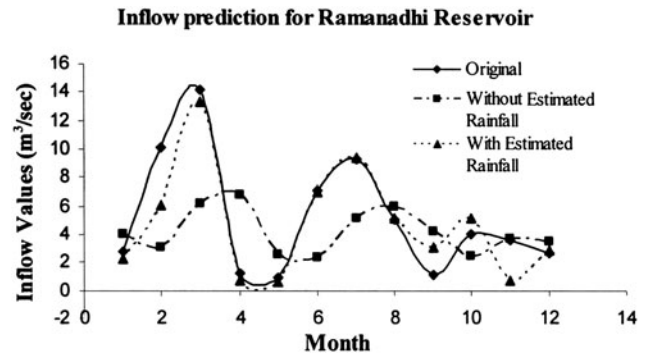| Period 1984	Actual inflow (m^3/s)	Predicted inflow (m^3/s)	
		Without estimated Rainfall	With estimated Rainfall
Jan	2.76	3.98	2.28
Feb	10.14	3.11	6.08
Mar	14.17	6.15	13.36
Apr	1.21	6.79	0.70
May	0.85	2.58	0.56
June	7.06	2.37	6.97
July	9.24	5.09	9.38
Aug	4.99	5.94	5.16
Sep	1.13	4.20	3.12
Oct	4.03	2.50	5.16
Nov	3.62	3.65	0.65
Dec	2.68	3.49	2.89
RMSE (m^3/s)		4.1	1.6

Fig. 5 Inflow predictions for Ramanadhi reservoir

The RMSE value decreases from 4.1 to 1.6 cumec. This is clearly reflected in figure. Particularly, the peak value is very closely predicted. For example, the inflow for the period of February 1984 is predicted as 6.08 cumec as against 3.11 cumec for prediction without using rainfall as input. Similarly, the inflow for the period of March 1994 has been predicted as 13.36 cumec which is very close to the actual value of 14.17 cumec.

This improvement in the prediction of inflow to the reservoir is very vital for system engineers/irrigation engineers for optimal planning of reservoir operation.

8 Conclusion

A novel method is proposed for interpolating rainfall values in a region with the help of available rain gauge stations. Further, the proposed method is applied for improving the inflow to a reservoir. The following conclusions can be arrived at from the study:

(a) ANN seems to be a powerful alternative to the traditional Kriging method of spatial interpolation.

(b) The effectiveness of ANN training crucially depends on not only the length of training records but also the quality of input variables.

(c) Results are only marginally improved in both ANN and Kriging models when all the stations in the catchment are used when compared to stations in the proximity.

(d) Inclusion of some kind of 'distance' variable in the input seems to play an important role in mapping the complexity of variability *provided* all the stations are included. Further studies should be carried out in both interpolations of rainfall as well as other hydrological variables such as ground water level, electrical conductivity variation, etc.

(e) Inflows to a reservoir can be improved using antecedent rainfall in the neighbourhood.

References

Biau G, Zorita E, Storch HV, Wackernagel H (1999) Estimation of precipitation by Kriging in the EOF space of the sea level pressure field. J Clim 12(4):1070–1085

Brath A, Montanari A, Toth E (2002) Neural networks and non-parametric methods for improving realtime flood forecasting through conceptual hydrological models. Hydrol Earth Syst Sci 6(4):627–640

Chowdhury M, Alouani A, Hossain F (2010) Comparison of ordinary Kriging and artificial neural network for spatial mapping of arsenic contamination of groundwater. Stoch Environ Res Risk Assess. doi:10.1007/s00477-008-0296-5

Goncalves AM, Alpuim T (2006) Precipitation measurement and the analysis of hydrological resources in a river basin. In: 7th international symposium on spatial accuracy assessment in natural resources and environmental sciences (Accuracy 2006), pp 851–860

Jinfeng Y, Chen X, Geping L, Quanjun G (2006) Temporal and spatial variability response of groundwater level to land use/land cover change in oases of arid areas. Chin Sci Bull 51(Suppl 1):51–59

Karlsson M, Yakowitz S (1987) Nearest neighbor methods for nonparametric rainfall-runoff forecasting. Water Resour Res 23(7):1300–1308

Karunanithi N, Grenney WJ, Whitley D, Bovee K (1994) Neural networks for river flow prediction. J Comput Civil Eng 8(2):201–220

Kitanidis PK (1997) Introduction to geostatistics: applications to hydrogeology. Cambridge University Press, New York

Neuroshell2 (Release 4.0).Ward systems Group Inc., 1993–1998

Shi Y, Li L, Zhang L (2007) Application and comparing of IDW and Kriging interpolation in spatial rainfall information. In: Proceedings of SPIE, Geoinformatics 2007, vol 6753

Szolgay J, Parajka J, Kohnová S, Hlavcová K (2009) Comparison of mapping approaches of design annual maximum daily precipitation. Atmos Res 92:289–307

Teegavarapu RSV (2007) Use of universal function approximation in variance dependent surface interpolation method—an application in hydrology. J Hydrol 332:16–29

Tsintikidis D, Georgakakos KP, Sperfslage JA, Smith DE, Carpenter TM (2002) Precipitation uncertainty and rain gauge network design within Folsom Lake watershed. J Hydrol Eng 7(2):175–184

Vijayakumar V, Remadevi R (2006) Kriging of groundwater levels—a case study. J Spat Hydrol 6(1):81–94

Wilk J, Kniveton D, Andersson L, Layberry R, Todd MC, Hughese D, Ringrose F, Vanderpost C (2006) Estimating rainfall and water balance over the Okavango River Basin for hydrological applications. J Hydrol 331(1–2):18–29

Yan-bing T (2002) Comparison of semivariogram models for Kriging monthly rainfall in eastern China. J Zhejiomj Univ Sci 3(5):584–590