HYDROLOGICAL PROCESSES
Hydrol. Process. **22**, 643–650 (2008)
Published online 30 July 2007 in Wiley InterScience
(www.interscience.wiley.com) DOI: 10.1002/hyp.6630

Basin-scale stream-flow forecasting using the information of large-scale atmospheric circulation phenomena

Rajib Maity† and D. Nagesh Kumar*
Department of Civil Engineering, Indian Institute of Science, Bangalore, 560 012, India

Abstract:

It is well recognized that the time series of hydrologic variables, such as rainfall and streamflow are significantly influenced by various large-scale atmospheric circulation patterns. The influence of El Niño-southern oscillation (ENSO) on hydrologic variables, through hydroclimatic teleconnection, is recognized throughout the world. Indian summer monsoon rainfall (ISMR) has been proved to be significantly influenced by ENSO. Recently, it was established that the relationship between ISMR and ENSO is modulated by the influence of atmospheric circulation patterns over the Indian Ocean region. The influences of Indian Ocean dipole (IOD) mode and equatorial Indian Ocean oscillation (EQUINOO) on ISMR have been established in recent research. Thus, for the Indian subcontinent, hydrologic time series are significantly influenced by ENSO along with EQUINOO. Though the influence of these large-scale atmospheric circulations on large-scale rainfall patterns was investigated, their influence on basin-scale stream-flow is yet to be investigated. In this paper, information of ENSO from the tropical Pacific Ocean and EQUINOO from the tropical Indian Ocean is used in terms of their corresponding indices for stream-flow forecasting of the Mahanadi River in the state of Orissa, India. To model the complex non-linear relationship between basin-scale stream-flow and such large-scale atmospheric circulation information, artificial neural network (ANN) methodology has been opted for the present study. Efficient optimization of ANN architecture is obtained by using an evolutionary optimizer based on a genetic algorithm. This study proves that use of such large-scale atmospheric circulation information potentially improves the performance of monthly basin-scale stream-flow prediction which, in turn, helps in better management of water resources. Copyright © 2007 John Wiley & Sons, Ltd.

KEY WORDS El Niño-southern oscillation (ENSO); equatorial Indian Ocean oscillation (EQUINOO); Mahanadi River; artificial neural network (ANN); genetic algorithm based evolutionary optimizer; hydroclimatic teleconnection; stream-flow; forecasting

Received 19 April 2006; Accepted 6 November 2006

INTRODUCTION

Prediction of stream-flow is an important factor for water resources management. It helps in flood control, devising agricultural strategy, reservoir operation, etc. Use of the inherent statistical properties of time series for prediction is practised widely with the intuitive assumption of stationarity. Recently, it is understood that the temporal structure of a hydrologic time series is significantly influenced by large-scale atmospheric circulations (Jain and Lall, 2001). However, it is scientifically and mathematically challenging to use such signals for the prediction of basin-scale hydrologic variables.

Hydroclimatic teleconnection between the large-scale rainfall pattern over India and the large-scale atmospheric circulation patterns from the tropical Pacific Ocean and Indian Ocean is established (Rasmusson and Carpenter, 1983; Parthasarathy *et al.*, 1988; Krishna Kumar *et al.*, 1999; Ashok *et al.*, 2001; Li *et al.*, 2001; Gadgil *et al.*, 2003; Gadgil *et al.*, 2004; Maity and Nagesh Kumar, 2006a). A brief description of large-scale atmospheric circulations, which are established to have a significant link with Indian summer monsoon rainfall (ISMR), is presented along with the physical mechanism.

El Niño and Southern Oscillation

El Niño and southern oscillation is the coupled ocean-atmosphere mode of the tropical Pacific Ocean (Cane, 1992). It is a large scale anomalous warming of sea surface temperature (SST) over the central and eastern Pacific Ocean with associated change in pressure field. In normal years, SST of the western part of the equatorial Pacific Ocean remains warmer than that of the eastern part and pressure at the eastern part of the Pacific Ocean is higher than that of the western part. During anomalous years, SST of the eastern part of the equatorial Pacific Ocean becomes warmer-than-normal and the pressure field is reversed, i.e. the anomalous pressure in the eastern part of the Pacific Ocean becomes lower than that in the western part. A contrary situation may also occur. Anomalous warming (cooling) of SST over the eastern part of the Pacific Ocean is known as El Niño (La Niña) whereas anomalous sea-saw variation of the pressure field, between the eastern and western parts of the Pacific Ocean, is called the Southern Oscillation. Acting together, the oceanic and atmospheric parts are jointly known as El Niño-southern oscillation (ENSO)

* Correspondence to: D. Nagesh Kumar, Department of Civil Engineering, Indian Institute of Science, Bangalore, 560 012, Karnataka, India. E-mail: nagesh@civil.iisc.ernet.in
† Present address: Department of Civil Engineering, Indian Institute of Technology, Bombay, Mumbai - 400 076, India.

phenomenon. It is established that ISMR is significantly related to ENSO through oceanic-atmospheric teleconnection (Rasmusson and Carpenter, 1983; Parthasarathy *et al*., 1988). However, contrary to the long recognized negative correlation between the ISMR and ENSO, some discrepancies have been observed in recent years (Krishna Kumar *et al*., 1999; Li *et al*., 2001; Gadgil *et al*., 2004). These unanticipated experiences suggest that the response of the monsoon to El Niño is not yet assessed adequately (Gadgil, 2003; Gadgil *et al*., 2003) or more pertinently that there are some other causative climate forcing events, which are also influencing the Indian rainfall concurrently.

Indian Ocean dipole (IOD) mode and equatorial Indian Ocean oscillation (EQUINOO)

Indian Ocean Dipole (IOD) mode is a pattern of internal variability with anomalously low sea surface temperatures off Sumatra and high sea surface temperatures in the western Indian Ocean, with accompanying wind and precipitation anomalies (Saji *et al*., 1999). There are two coupled components of IOD—oceanic and atmospheric. The dipole mode index (DMI) is the oceanic component (Saji *et al*., 1999) of the IOD mode, which is defined as the difference in SST anomaly between the tropical western Indian Ocean (50 °E–70 °E, 10 °S–10°N) and the tropical south-eastern Indian Ocean (90 °E–110 °E, 10 °S–0°). However, statistical correlation of the DMI with precipitation over the Asian monsoon regime does not yield a significant relationship. Thus the relationship of the DMI to the ISMR variability is not clear (Saji *et al*., 1999). However, the equatorial Indian Ocean oscillation (EQUINOO), which is the atmospheric component of the IOD mode, is established to have significant influence on the ISMR (Gadgil *et al*., 2004). The convection over the eastern part of the equatorial Indian Ocean (EEIO, 90°–110 °E, 10 °S–0°) is negatively correlated to that over the western part of the equatorial Indian Ocean (WEIO, 50°–70 °E, 10 °S–10 °N). The anomalies in the sea level pressure and the zonal component of the surface wind along the equator are consistent with the convection anomalies. When the convection is enhanced (suppressed) over the WEIO, the anomalous surface pressure gradient is towards the west (east) so that the anomalous surface wind along the equator becomes easterly (westerly). The oscillation between these two states is known as EQUINOO. The equatorial zonal wind index (EQWIN) is considered as an index of EQUINOO which is defined as the negative of the anomaly of the zonal component of surface wind in the equatorial Indian Ocean region (60°–90 °E, 2·5 °S–2·5 °N). Northward propagation of a large scale convective system generated over the Indian Ocean indicates the physical link between ISMR and EQUINOO (Gadgil *et al*., 2004). Recently it was observed that ISMR is influenced by both ENSO and EQUINOO, at seasonal time-scale (Ashok *et al*., 2001; Gadgil *et al*., 2004) as well as at monthly time-scale (Maity and Nagesh Kumar, 2006a).

Basin-scale stream-flow and large-scale atmospheric circulation

Although the strength of the hydroclimatic teleconnection decreases for smaller spatio-temporal scale, significant influence still exists for subdivisional scale too for some particular geographical locations. However, the nature of the relationship varies for different subdivisions and different seasons (Kane, 1998; Maity and Nagesh Kumar, 2006b).

The basin-scale stream-flow, used in this study, is measured at Basantpur, a few kilometres upstream of Hirakud reservoir, one of the most important multipurpose reservoirs, located in the state of Orissa in India (Figure 1). The upstream catchment of Basantpur site is mostly located in the Chattisgarh state in India. Though the spatial variation of ISMR across subdivisions is considerable, ISMR is highly correlated with rainfall over Chattisgarh rainfall (0·603, p-value being 0·00 with null

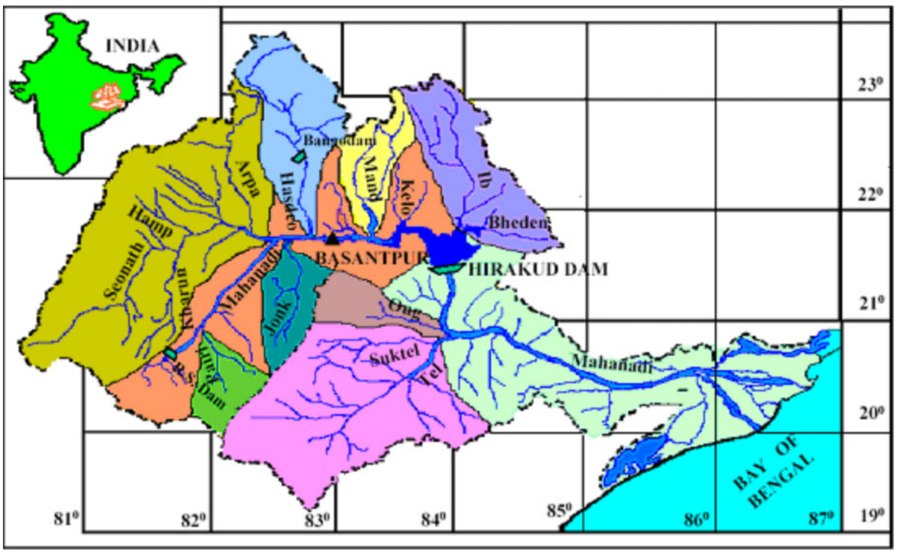

Figure 1. Location map of catchment and sub-basins of Mahanadi River

DOI: 10.1002/hyp

hypothesis as no correlation). Thus a physical relationship between ISMR and ENSO and EQUINOO can be ingenuously assumed to exist, effecting rainfall over Chattisgarh subdivision. The seasonally averaged rainfall is also highly correlated with the seasonally averaged stream-flow. However, at monthly time-scale, the hydroclimatic teleconnection between basin-scale stream-flow and large-scale atmospheric circulation pattern is more complex and cannot be recognized easily. One of the most important reasons is the complex non-linear relationship between rainfall over upstream catchment and the stream-flow. However, the affect of the rainfall pattern on stream-flow is obvious on a monthly scale. Hence, the influence of ENSO and EQUINOO on stream-flow on a monthly scale is supposed to exist as there exists a link between rainfall pattern and ENSO and EQUINOO. However, the influence of these large-scale atmospheric circulation phenomena on the rainfall pattern over Chattisgarh subdivision, on a monthly scale, is not discernible in terms of correlation coefficients. Thus, the initial motivation for this study of the large-scale atmospheric circulation pattern and the basin-scale stream-flow is rather logical than statistical.

The objective of this study is to investigate the influence of large-scale atmospheric circulation information on the basin-scale stream-flow variation and possible improvement of stream-flow prediction by incorporating the information of such large-scale atmospheric circulations. An artificial neural network (ANN) approach is adopted to model the complex relationship between stream-flow and large-scale atmospheric circulations (ENSO and EQUINOO). ANN architecture is decided using a genetic algorithm (GA) based evolutionary optimizer. A description of the algorithm is presented in a later section. Predictability of stream-flow is investigated with the trained networks. Improvement of prediction performance and advantage of considering the large-scale atmospheric circulation information are also investigated.

DATA

Sea surface temperature anomaly (SSTA) from the Niño 3·4 region (120°–170 °W, 5 °S–5 °N) is used as the ENSO index in this study. Monthly SSTA data is obtained from the website of the National Weather Service, Climate Prediction Centre of NOAA (http://www.cpc.noaa.gov/data/indices/) for the period January 1972–December 2003.

Similarly, EQWIN is used as EQUINOO index. Monthly surface wind data for the period January 1972–December 2003 is obtained from the National Centre for Environmental Prediction (http://www.cdc.noaa.gov/Datasets).

Monthly stream-flow data at Basantpur site is obtained from the office of Executive Engineer, Mahanadi Division, Central Water Commission (CWC), Burla, Orissa for the period January 1972–December 2003.

METHODOLOGY

Stream-flows are modelled on a monthly time-scale using the information of large-scale atmospheric circulations of ENSO and EQUINOO in terms of their corresponding indices, which were discussed earlier.

Potential influencing variables

As mentioned earlier, the catchment of Basantpur site is mostly located in the Chattisgarh state in India and ISMR is highly correlated with rainfall over Chattisgarh state. Correlation analysis between large-scale circulation indices (both ENSO and EQUINOO) and monthly variation of summer monsoon rainfall over India shows that for June rainfall, both ENSO and EQUINOO indices from March; for July rainfall, both ENSO and EQUINOO indices from June; for August rainfall, both ENSO and EQUINOO indices from July; and for September rainfall, ENSO index from August and EQUINOO index from July are highest correlated (Maity and Nagesh Kumar, 2006c). However, such correlations between large-scale indices and stream-flows are not so conspicuous due to the complex non-linear relationship between rainfall over upstream catchment and the stream-flow. Still it is worthwhile to note that stream-flow for the month of June is highest correlated with large-scale indices for the month of March as that between rainfall and large-scale indices. Thus, for June stream-flow, both ENSO and EQUINOO indices from March, are used in this study. For other months (July through October), even if the correlation coefficients (provides only the information about linear association) are not conspicuous (due to complex non-linear relationship between rainfall over upstream catchment and the stream-flow), intuition from the correlation analysis between large-scale circulation indices and monthly variation of summer monsoon rainfall over India is used in a broad sense. Thus, for the months July through October, large-scale circulation information is used from the immediate previous month.

Apart from the large-scale atmospheric circulation, information of stream-flow of previous month(s) is (are) also used as serial autocorrelation is significant. Results of the correlation analysis between stream-flow of successive months are presented in Table I. It can be observed from Table I that, stream-flow in the month of June is not correlated with that of May, because there is hardly any stream-flow in May. Thus for June, previous month stream-flow information is not used. For July, August and October, correlation coefficients are significant only up to one lag and thus, stream-flow information from the previous month is considered. However, for September, correlation coefficient between the stream-flows in September and July is higher than that between September and August. It may not be logical to use the stream-flow information from July without considering the same from the immediate previous month, i.e. August. Hence, stream-flow information of both July and August are considered for the month of September. The complete set of input variables for different months is shown in Table II.

Table I. Correlation coefficient between stream-flow in different months

	June	July	August	September	October
May	0·01	−0·15	−0·41	—	—
June	1·00	**0·69**	0·21	0·04	—
July	—	1·00	**0·41**	**0·32**	0·18
August	—	—	1·00	0·16	0·05
September	—	—	—	1·00	**0·65**
October	—	—	—	—	1·00

Artificial neural network (ANN)

ANN is used in this study to predict stream-flow using the input variables as described earlier. All the input variables and stream-flow values are scaled between 0 and 1 before using in ANN.

ANN can capture the non-linear relationship between two time series, if any, and does not depend on the distributional form of the data set. Details of ANN approach and its application in various fields of engineering can be found elsewhere (Haykin, 1999; ASCE, 2000a,b). One important aspect of ANN methodology is the design of network architecture. In most of the studies, network architecture is decided based on heuristic. Selection of optimum neural network architecture can be obtained by pruning the neural network. Optimal brain damage (Le Cun *et al*., 1990) and optimal brain surgeon (Hassibi *et al*., 1993) can be mentioned in this regard. However, in this study, an evolutionary optimizer based on GA is used to decide the optimum architecture of feed forward network. The back propagation algorithm (Rumelhart *et al*., 1986; Yu and Chen, 1997) is used to train the network. The evolutionary optimizer based on GA approach has been used in some other studies also (Nagesh Kumar *et al*., 2005). The design of the network architecture and its training is performed using the data for the period 1972–1991.

GA based evolutionary optimizer

The evolutionary optimizer uses the principle of GA (Goldberg, 1989) to obtain optimal network architecture. The main steps involved in evolutionary optimizer are as follows:

(a) Parameters initialization: All the parameters like population size (N), number of maximum generations, probability of crossover (P_c) and probability of mutation (P_m) are set to specific values. In this study, N, P_c and P_m are selected as 50, 0·2 and 0·04, respectively.
(b) Generation of initial population: At initial generation, the evolutionary optimizer randomly creates N networks for the initial population, where N is equal to the size of the population.
(c) Training of the network and fitness evaluation: All the networks within the current generation are trained by back propagation algorithm and their fitness values are determined according to the goals to be achieved. These goals are: (i) maximum number of neurons used in the network, (ii) the network's mean square deviation and (iii) maximum square deviation on pattern set reproduction. Thus this is a multiobjective GA, objectives of which are to minimize the above goals so that they satisfy the threshold of each goal. Weighted average (explained as percentage) of the goal parameters is used as the fitness of the individual network. It can be mentioned here that the mean square deviation and maximum square deviation of GA based evolutionary optimizer are subjectively chosen in such a way that the performance of the network is comparable for both training and testing data set. Moreover, in the final generation, among the various networks, the network which satisfies all the goals, is used for testing dataset. The network providing the best performance is selected. Thus, the chances of over-training and under-training are avoided.
(d) Evolution of new generation: A new generation of networks will be created from the present generation according to the following procedure. It must be noted that the following operations are carried out with the 'network architecture', but not with the trained network. Once the population of the new networks is formed, the networks undergo a 'fresh' training as explained in step (c).

   (i) Two 'parent' networks will be chosen out of the old generation. The selection algorithm will choose networks with a high fitness by a higher probability.
   (ii) Two 'children' networks will be created from the two 'parent' networks. Using the cross over probability P_c, the two 'parent' networks will be crossed over, i.e. they will swap a portion of the network with each other.
   (iii) The 'children' will be mutated with a mutation probability P_m. Here, mutation means insertion or deletion of a layer and/or insertion or deletion of a neuron into a layer.
   (iv) A few elitist members of the population in current generation are carried to the next generation. The selection continues until the new generation also has N members. After completion the new generation will be evaluated.

(e) Checking for termination criteria: Evolution is stopped, if the target goals are achieved at least for one network in the population or maximum number of generations is reached. Otherwise, generation counter is increased by one and steps (c) and (d) are repeated.

RESULTS AND DISCUSSIONS

The procedure of the evolutionary optimizer is applied to each of the individual monsoon months. For each case, the best network in the final generation, which fulfils the target goals, is shown in Table II. Correlation coefficients between observed and modelled stream-flows for the

DOI: 10.1002/hyp

Table II. Set of input variables and best networks for different months of the monsoon period

Month of stream-flow	Input variables[a]	Best network	CC_TP[b]
June	March EN March EQ	2-4-5-3-1	0·68
July	June SF June EN June EQ	3-3-5-1	0·88
August	July SF July EN July EQ	3-6-5-1	0·97
September	July SF August SF August EN August EQ	4-3-2-1	0·87
October	September SF September EN September EQ	3-5-1	0·87

[a] SF, Stream-flow; EN, ENSO index; EQ, EQUINOO index.
[b] CC_TP, correlation coefficient during training period (1972–1991).

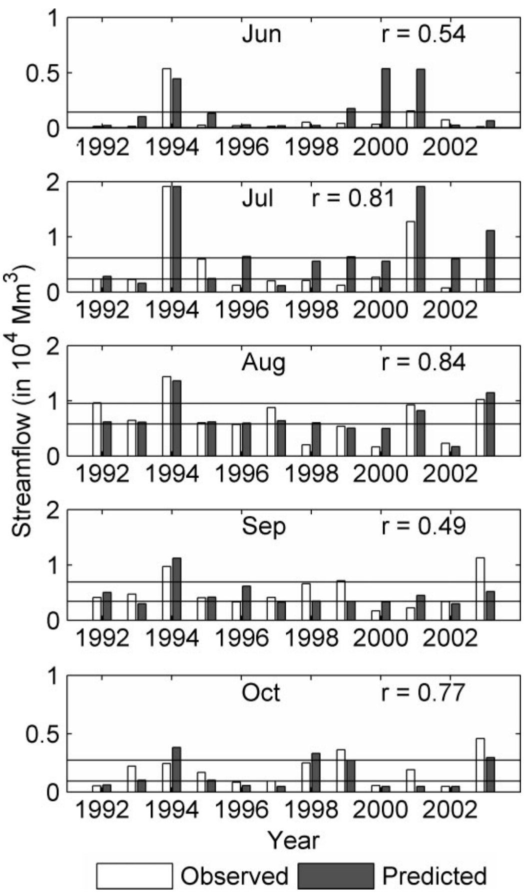

Figure 2. Observed and predicted stream-flows during the validation period, 1992–2003

training period (1972–1991) are also shown in Table II. Performances of trained networks are tested using data for the period, 1992–2003. A comparison between observed and predicted stream-flows is shown in Figure 2.

It is observed that the prediction performance is reasonably good for all months. However, for June, it is not as good as that for other months. The reason for this poor performance can be attributed to the high amount of variance associated with stream-flow during this month. Coefficient of variation, γ, ($\gamma = \sigma/\mu$, where σ is the standard deviation and μ is the mean) for this month, is observed to be 1·91 whereas for the months July through October values of γ are 0·90, 0·49, 0·68 and 0·97, respectively. This indicates that variability associated with June stream-flow is much higher than that with stream-flow during other months. However, it is also true that the stream-flow during June is much lower than that in the other months. Thus, this poorer prediction performance is not very problematic from a water resources management point of view.

For July, very high stream-flow (1994) as well as very low stream-flow (1992, 1993 and 1997) are successfully captured. However, stream-flows are over predicted during 2001, 2002 and 2003. Most successful prediction is observed for the month of August. More or less accurate stream-flows are predicted almost for all the years. It is worth noting that maximum amount of stream-flow occurs during the months of July and August. Model performance during these high stream-flow months can be appreciated. Stream-flows are predicted with reasonable accuracy for the months of September and October too, except for September 2003.

Prediction performance for all monsoon months is investigated in ordinal scale also. An ordinal scale conveys the information of rank order. Sometimes, particularly in case of stream-flow, the analysis in ordinal scale is also useful to decision-makers. However, it is true that as the observed and predicted stream-flows are compared numerically, ordinal scale analysis is not that useful. Still the analysis is presented to show the potential of the model at ordinal scale.

In ordinal scale stream-flow categories may be termed as 'high', 'normal' and 'low'. Thus, prediction in ordinal scale is meant to indicate whether the ensuing season will have high, normal or low stream-flow. Stream-flow is defined as 'normal', if it lies within the range of $\pm$50% of standard deviation (SD) from mean. Stream-flow, which lies above (below) this range is defined as 'high' ('low') stream-flow. In Figure 2, two lines (mean $\pm$0·5 SD) are shown demarcating the high, normal and low stream-flows for each of the monsoon months. It may be noted that 'mean' and 'standard deviation' are of observed values.

Categorical correspondence between observation and prediction can be visually assessed from these plots. For statistical investigation, a contingency table is prepared showing the correspondence between the observed and predicted stream-flows (Table III). Polychotomous (more than two category) forecast performance can be investigated by Heidke skill score (HSS) (Wilks, 2006). HSS is defined as:

$$HSS = (CF - CF_{rand})/(N - CF_{rand})$$

Table III. Contingency table showing the correspondence between the observed and predicted stream-flows

Stream-flow category Observed	Predicted		
	Low	Normal	High
Low	12	6	3
Normal	3	20	4
High	0	3	9

where CF is the number of correct forecasts, CF_{rand} is the correct forecast by chance of random forecast and N is the total number of forecasts or maximum number of correct forecasts that can be achieved. HSS ranges from $-\infty$ to 1. Higher value of HSS indicates better performance. A value, greater than 0·15, indicates a reasonably good forecast.

For the present case (Table III), $N = 60$, $CF = 41$ and $CF_{rand} = 20$, being a chance of one-third of a case to be correctly predicted. Thus, the value of HSS is 0·525, which indicates that categorical stream-flow information for the ensuing month can be provided by this approach with an applaudable accuracy.

An obvious question may come to mind, whether the commendable results are owing to the consideration of stream-flow information of previous month(s). The analysis is repeated and the performance of stream-flow prediction is checked using only stream-flow information of previous month(s), i.e. information of large-scale atmospheric circulation is discarded. Results are shown in Table IV. Model performance is measured in terms of correlation coefficient (CC), mean absolute error (MAE) and root mean square error (RMSE) between observed and predicted stream-flows for the model validation period (1992–2003). It can be observed that, prediction performances are poorer while using stream-flow information of previous month(s) alone. This observation is conspicuous for the high stream-flow months, i.e. July, August and September. A minor inconsistency for the months June and October is discussed later. In general, improvement of the prediction performance, by considering both the information of stream-flow of previous month(s) and large-scale atmospheric circulations can be easily appreciated. This is because the auto-correlation structure of stream-flow series, varies considerably for different parts of the stream-flow series. In other words, higher-than-normal (lower-than-normal) stream-flow of previous month does not always assure higher-than-normal (lower-than-normal) stream-flow for the present month.

For the month of June, apparently poorer values of MAE and RMSE are obtained while considering both stream-flow and large-scale atmospheric circulation information than those while considering stream-flow information of previous month alone. However, the correlation coefficient is much better while considering both stream-flow and large-scale atmospheric circulation information. Moreover, it may be noted that, if the information of ENSO and EQUINOO are ignored, very low stream-flows are predicted for all the years and the high flow in the year 1994 is not predicted at all. As the magnitude of stream-flows are small for the month of June (already mentioned), resulted MAE and RMSE values are also low.

A small deviation for the month of October is also observed. This is due to the fact that summer monsoon rainfall in India ends in the months of September. Stream-flow for the month of October is mainly sourced by surface storage of the upstream catchment and sub-surface flow. Basically October is the transition period from high stream-flow to low stream-flow. This indicates a correspondence between the September stream-flow and October stream-flow. Eventually, stream-flow in the month of October is well related to the September stream-flow. Thus, a marginally better model performance is achieved for the month of October only while using September stream-flow alone. However, for other months (July, August and September), improvement of the prediction performance, by considering both the information of stream-flow of previous month(s) and large-scale atmospheric circulations can be easily appreciated.

The effect of considering the information of large-scale atmospheric circulation alone is also investigated. A similar observation, i.e. poorer prediction performances are noticed in such cases too (Table IV). This is due to the fact that prevailing conditions/characteristics of watershed are very important factors for stream-flow prediction. While considering stream-flow of previous month(s), some information of such conditions/characteristics is being considered. Moreover, to investigate the relative importance of ENSO and EQUINOO on stream-flow, one of the large-scale indices (ENSO and EQUINOO) is considered at a time and the performance of stream-flow prediction is investigated. It is found that, prediction performances are poorer while considering either one of ENSO and EQUINOO as compared to that while considering both the large-scale indices simultaneously. Logically, this is due to the fact that, variation of ISMR can be better explained and modelled by the combined information of ENSO and EQUINOO as these indices jointly influence the variation of ISMR as mentioned earlier (Gadgil *et al.*, 2004; Maity and Nagesh Kumar, 2006a).

Thus, the overall observation of this study indicates the significant influence of large-scale atmospheric circulation patterns from the tropical Pacific and Indian Ocean on basin-scale stream-flow. Use of such information improves the prediction performance of stream-flow, which will be very useful for better management of water resources.

Finally, the role of GA based evolutionary optimizer is worth mentioning here. It is obvious that performance of ANN significantly depends on its architecture, which is generally decided by heuristics and experience of the researcher. However, with such procedure, it may not be possible to achieve the best possible results for

DOI: 10.1002/hyp

Table IV. Comparison of performances for different cases

Month of stream-flow	Considering only stream-flow information of previous month(s)			Considering only large-scale atmospheric circulation information			Considering both stream-flow and large-scale atmospheric circulation information		
	Inputs[a]	Best network	Model performance[b]	Inputs[a]	Best network	Model performance[b]	Inputs[a]	Best network	Model performance[b]
June	May SF	1-2-2-1	−0·030 833·0 1500·3	March EN March EQ	2-4-5-3-1	0·542 1209·4 1926·4	March EN March EQ	2-4-5-3-1	0·542 1209·4 1926·4
July	June SF	1-2-2-2	0·731 4312·8 5025·6	June EN June EQ	2-3-2-1	0·298 5643·6 6059·4	June SF June EN June EQ	3-3-5-1	0·813 3570·9 4425·2
August	July SF	1-2-3-1-1	0·412 2929·2 3575·7	July EN July EQ	2-2-1	0·511 2721·5 3225·6	July SF July EN July EQ	3-6-5-1	0·836 1486·0 2008·5
September	July SF August SF	2-2-3-1-1	0·321 2331·9 3260·2	August EN August EQ	2-3-1-2-1	0·402 1877·1 2719·7	JulySF August SF August EN August EQ	4-3-2-1	0·493 2101·0 2635·6
October	September SF	1-2-1	0·838 611·6 761·4	September EN September EQ	2-4-3-1	0·372 898·8 1228·1	September SF September EN September EQ	3-5-1	0·765 742·2 926·5

[a] SF, stream-flow; EN, ENSO index; EQ, EQUINOO index.
[b] Model performance in terms of (from top to bottom) correlation coefficient (CC), mean absolute error (MAE) and root mean square error (RMSE) between observed and predicted stream-flows during the model validation period (1992–2003).

many cases, particularly for unknown new problems. The proposed approach, in this study, assures the best possible results for different cases as mentioned earlier. This is due to the application of GA which decides the best network architecture. The approach may be used for other applications of ANN.

CONCLUSIONS

In this study, the possible influence of large-scale atmospheric circulation phenomena on basin-scale monthly stream-flow variation is investigated. Information of two large-scale atmospheric circulation phenomena namely, ENSO from the tropical Pacific Ocean and EQUINOO from the Indian Ocean, are used. ANN approach is used to capture the complex relationship between basin-scale monthly stream-flow and the large-scale atmospheric circulation phenomena. Instead of heuristic basis, a GA based evolutionary optimizer is used to identify the optimum network architecture for a particular pattern set.

It is shown that the basin-scale stream-flow is influenced by large-scale atmospheric circulations phenomena. Information of stream-flow from previous month(s) alone, as used in most of the traditional modelling approaches, is shown to be insufficient. It is successfully established that incorporation of large-scale atmospheric circulation information significantly improves the prediction performance for the monthly scale. Again, prevailing conditions/characteristics of watershed are also important. Thus, consideration of both the information of previous stream-flow and large-scale atmospheric circulations is important for basin-scale stream-flow prediction for the monthly time scale. Adopting this approach, monthly stream-flows are predicted with better accuracy which is a very useful input for better management of water resources for the downstream area.

ACKNOWLEDGEMENTS

This work is partially supported by the Department of Science and Technology, Government of India, through project number ES/48/010/2003 and INCOH, Ministry of Water Resources, Government of India, through project number 23/52/2006-R&D.

REFERENCES

ASCE Task Committee. 2000a. Artificial neural networks in hydrology. I: Preliminary Concepts. *Journal Hydrologic Engineering* **5**(2): 115–123.

ASCE Task Committee. 2000b. Artificial neural networks in hydrology. II: Hydrologic applications. *Journal of Hydrologic Engineering* **5**(2): 124–137.

Ashok K, Guan Z, Yamagata T. 2001. Impact of Indian Ocean dipole on the relationship between the Indian monsoon rainfall and ENSO. *Geophysics Research Letters* **28**: 4499–4502.

Cane MA. 1992. Tropical Pacific ENSO models: ENSO as a mode of the coupled system. In *Climate System Modeling*, Trenberth KE (ed.). Cambridge University Press: Cambridge; 583–689.

Gadgil S. 2003. The Indian Monsoon and its variability. *Annual Review of Earth Planetary Science* **31**: 429–467.

Gadgil S, Vinayachandran PN, Francis PA. 2003. Droughts of the Indian summer monsoon: role of clouds over the Indian ocean. *Current Science* **85**(12): 1713–1719.

Gadgil S, Vinayachandran PN, Francis PA, Gadgil S. 2004. Extremes of the Indian summer monsoon rainfall, ENSO and equatorial Indian Ocean oscillation. *Geophysics Research Letters* **31**: L12213.

Goldberg DE. 1989. *Genetic Algorithms in Search, Optimization, and Machine Learning*. Addison-Wiley: Reading, NY.

Hassibi B, Stork DG, Wolff GJ. 1993. Optimal brain surgeon and general network pruning. In *IEEE International Conference on Neural Networks 1993*, San Francisco, CA. IEEE Publications: Los Alamitos, CA; vol. 1, 293–299.

Haykin S. 1999. *Neural Networks: A Comprehensive Foundation*. Prentice Hall: Englewood Cliffs, NJ.

Jain S, Lall U. 2001. Floods in a changing climate: does the past represent the future? *Water Resources Research* **37**(12): 3193–3205.

Kane RP. 1998. Extremes of the ENSO phenomenon and Indian summer monsoon rainfall. *International Journal of Climatology* **18**: 775–791.

Krishna Kumar K, Rajagopalan B, Cane MA. 1999. On the weakening relationship between the Indian Monsoon and ENSO. *Science* **284**: 5423.

Le Cun Y, Denker J, Solla S, Howard RE, Jackel LD. 1990. Optimal brain damage. In *Advances in Neural Information Processing Systems*, Touretzky DS (ed.). Morgan Kauffman Publishers: San Mateo, CA; vol. 2, 598–605.

Li T, Zhang YS, Chang CP, Wang B. 2001. On the relationship between Indian Ocean sea surface temperature and Asian summer monsoon. *Geophysics Research Letters* **28**: 2843–2846.

Maity R, Nagesh Kumar D. 2006a. Bayesian dynamic modeling for monthly Indian summer monsoon rainfall using El Niño-Southern Oscillation (ENSO) and Equatorial Indian Ocean Oscillation (EQUINOO). *Journal of Geophysics Research* **111**: D07104.

Maity R, Nagesh Kumar D. 2006b. Hydroclimatic teleconnection between global sea surface temperature and rainfall over India at subdivisional monthly scale. *Hydrological Processes* **21**(14): 1802–1813.

Maity R, Nagesh Kumar D. 2006c. Hydroclimatic association of monthly summer monsoon rainfall over India with large-scale atmospheric circulations from tropical Pacific Ocean and Indian Ocean region. *Atmospheric Science Letters* **7**(4): 101–107.

Nagesh Kumar D, Reddy MJ, Maity R. 2005. Regional rainfall forecasting using large scale climate teleconnections and evolutionary algorithm. In *Proceedings of the 2nd Indian International Conference on Artificial Intelligence (IICAI-05)*, Pune, India; WR 105.

Parthasarathy B, Diaz HF, Eischeid JK. 1988. Prediction of all/India summer monsoon rainfall with regional and large-scale parameters. *Journal of Geophysics Research* **93**(5): 5341–5350.

Rasmusson EM, Carpenter TH. 1983. The relationship between eastern equatorial Pacific sea surface temperature and rainfall over India and Sri Lanka. *Monthly Weather Review* **111**: 517–528.

Rumelhart DE, Hintont GE, Williams RJ. 1986. Learning representations by back-propagating errors. *Nature* **323**: 533–536.

Saji NH, Goswami BN, Vinayachandran PN, Yamagata T. 1999. A dipole mode in the tropical Indian Ocean. *Nature* **401**: 360–363.

Wilks DS. 2006. *Statistical Methods in Atmospheric Sciences*. Academic Press: New York.

Yu XH, Chen GA. 1997. Efficient back propagation learning using optimal learning rate and momentum. *Neural Networks* **10**(3): 517–527.