

Ascertaining Time Series Predictability in Process Control – Case Study on Rainfall Prediction

Chandrasekaran Sivapragasam¹, Poomalai Saravanan^{1*}, Saminathan Balamurali², and Nitin Muttil³

¹Center for Water Technology, Kalasalingam Academy of Research and Education, Srivilliputtur-626126, Tamil nadu, India

²Department of Computer Application, Kalasalingam Academy of Research and Education, Srivilliputtur-626126, Tamil nadu, India

³College of Engineering and Science, Victoria University, P.O. Box 14428, Melbourne, Victoria 8001, Australia

Abstract. Rainfall prediction is a challenging task due to its dependency on many natural phenomenon. Some authors used Hurst exponent as a predictability indicator to ensure predictability of the time series before prediction. In this paper, a detailed analysis has been done to ascertain whether a definite relation exists between a strong Hurst exponent and predictability. The one-lead monthly rainfall prediction has been done for 19 rain gauge station of the Yarra river basin in Victoria, Australia using Artificial Neural Network. The prediction error in terms of normalized Root Mean Squared Error has been compared with Hurst exponent. The study establishes the truth of the hypothesis for only 6 stations out of 19 stations, and thus recommends further investigation to prove the hypothesis. This concept is relevant for any time series which need to be used for real time process control.

1 Introduction

Due to rapid growth of population, urbanization and industrialization, the demand for water has increased tremendously. The sustainable planning of water resources of a country needs prediction of hydrological and meteorological parameters among which the spatial and temporal variability of rainfall is indispensable due to its importance in issuing flood warning and to plan for crop scheduling. When considering process automation and control particularly under the changing Industry Revolution 4.0 scenario, the forecasting of a given variable from its time series plays very important role. Before forecasting, it is preferred to ensure that the time series is actually predictable. For example, Qian and Rasheed [1] have studied the predictability of financial time series data. In this study, rainfall prediction is considered.

*Corresponding author: p.saravanan.me.1986@gmail.com

When compared to the traditional statistical methods, the performance of Artificial Neural Network (ANN) is reported to be the best in rainfall prediction [2, 3]. Hence ANN is being more commonly adopted for rainfall predictions [4, 5, 6, 7, 8]. The ANN is also used in many other hydrological applications such as modeling of runoff, improving spatial interpolation of rainfall, bias correction of simulated climate data, modeling of flash flood prediction etc. [2, 9, 10, 11, 12]

Since prediction of rainfall is a very arduous task due its chaotic nature and its dependency on many other climatological parameters, the researchers' adopted many methodologies as recourse of rainfall prediction using ANN. For instance, Havaluddin *et al.*, [13] had attempted to improve the prediction by changing the hidden layer and epochs. Some authors hybridized ANN with other techniques to improve predictions. Ramana *et al.* [14] integrated wavelet technique with ANN. Mohd-Safar *et al.* [6] hybridized ANN with Fuzzy Logic to predict short term rainfall in tropical climate. Yet others used mutual information techniques for the identification of best input variable among various available input parameters to predict rainfall [15]. Pre-processing of inputs has also been commonly used to improve rainfall prediction using ANN [16].

Despite all such efforts, developing a robust rainfall prediction model continue to remain a challenge. Some researchers tried to ascertain first if a rainfall series is predictable or not before attempting prediction by using the knowledge of the memory effect of the time series [5]. If a time series have correlation among themselves either as a short memory or long memory then the predictability will be more. Hurst exponent is a measure of long term memory having value between 0 and 1. If the Hurst exponent value is very near to 0.5, then the time series is random. On the other hand, if the Hurst exponent value is close to either 0 or 1, the current trend will persist or show anti-persistent respectively in future as well. It is believed hypothetically that a persistent time series will always be predictable because of its low randomness. Some authors used Hurst exponent as a measure of predictability, but the actual predictability is not considered in the scope of their reported works [17, 18, 19, 20]. For instance, Rangarajan and Sant [19] proposed a predictability index which is derived from Hurst exponent. Rehman [20] explored dependence of rainfall predictability on pressure and temperature using predictability indices. Similarly, Rehman and Siddiqi [21] indicated the dependency of temperature and pressure with predictability indices of precipitation and wind speed.

The work reported by Khalili *et al.* [5] is an attempt towards demonstrating the relationship between Hurst exponent and predictability of a monthly rainfall time series of 50 years. They studied rainfall series in Mashhad station and found Hurst exponent to be 0.96. They reported that the rainfall prediction is good as predictability is strong with a high Hurst value. Since, the relationship between Hurst exponent and predictability of rainfall time series has not received sufficient attention, it is necessary to investigate in more detail the validity of this hypothesis. This study investigates the Hurst exponent of 19 stations in the Yarra River catchment in Victoria, Australia. ANN is used for one lead prediction.

2 Study area and data set

The Yarra River catchment comprises of three segments which are upper, lower and middle. The lower segment is an urbanized one, which in danger of flash flood, if any extreme rainfall occurs in middle segment. Also the lower segment is highly depends on middle segment catchment for its water requirement [22]. Though area is less, it is one of the highest productive regions in Victoria. Hence, rainfall-related studies are of importance for this catchment [23].

The observed monthly rainfall time series from January 1981 to December 2012 of the middle segment of Yarra River Catchment are used in this study. To study the predictability,

nineteen rain gauge stations are used and its locations are shown in Figure 1. The name of the nineteen stations corresponding to the station numbers can be found in Table 1.

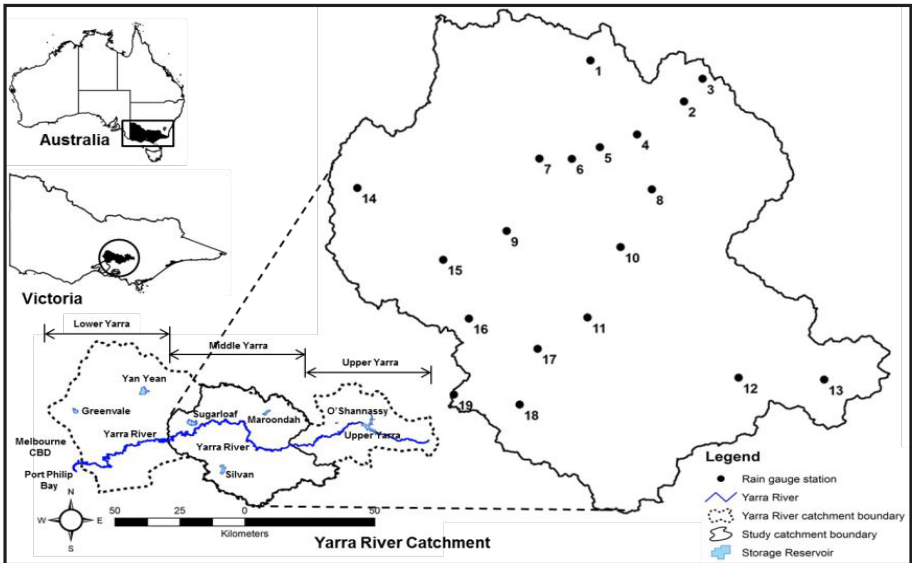


Fig. 1. Location of the study area and the nineteen rain gauge stations

3 Methodologies

Many methods are available to find the Hurst exponent, of which the rescaled range (R/S) analysis method is commonly adopted by many researchers [18, 19, 20].To predict the rainfall, ANN is used in this study. Both R/S analysis method and ANN technique are explained in the upcoming sub-section.

3.1 R/S Analysis

For prediction of Nile River flooding, a British Water Engineer, Harold E. Hurst developed the R/S analysis method [20].

The procedure to find the Hurst Exponent is as follows:

- 1) The data have to be arranged into a number of different time length (size) of data, Then the Rescaled range of each size will be calculated as follows,

(R/S) = R_t / sigma_t for t = 1,2,3,.....n (1)

where:

(R/S) = the rescaled range for the size under consideration

sigma = the standard deviation for the range under consideration

$$R_t = [\max(y_1, y_2, \dots, y_t) - \min(y_1, y_2, \dots, y_t)] \text{ for } t = 1, 2, \dots, n \quad (2)$$

Where:

R_t = the widest spread in each range

$$y_t = \sum_{i=1}^t (x_i - \mu) \text{ for } t = 1, 2, 3, \dots, n \quad (3)$$

y_t = the new time series adjusted for deviations from the mean

x_i = any of the value in the range

μ = the mean for the range

- 2) Calculate the logarithmic values for the size of each region and for each region's rescaled range
- 3) Plot the logarithm of the size (x-axis) of each series versus the logarithm of the rescaled range (y-axis). Slope of the line is the 'Hurst Exponent'

3.2 Artificial Neural Network

The architecture of ANN is motivated by the structure of the human brain and nerve cells. This technique is used identifying the statistical pattern present in the time series and applies it to unknown data to predict. A network of countless simple elements called neurons with a small amount of local memory is considered. The neurons are connected through connections which carry numeric data encoded by various means. Each neuron operates only when it receives data through the connections. The architecture is formed by the learning algorithm which is responsible for the extraction of the regularities present in the data through the finding of a suitable synapses set during the process of observation of the examples. Accordingly, ANNs solve problems by self-learning.

The feed-forward architecture is used in this study. One input layer, one output layer and one or more hidden layers are available in the architecture. The information passes from the input layer to the output layer through hidden layer. Each layer is fabricated by several neurons, and the layers are interconnected by a set of weights. Neurons operate the input and transform it to produce an analog output. More details about ANN have been discussed in [9, 24, 25, 26].

The performance of the forecast is evaluated by the normalized root-mean squared-error (NRMSE) goodness-of-fit measure, represented by the following equation:

$$NRMSE = \frac{RMSE}{MAX - MIN} \quad (4)$$

Where, MAX= highest value of the validation set

MIN = Lowest value of the validation set

and, Root-mean squared-error RMSE is calculated as follows,

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n [(X_m)_i - (X_s)_i]^2} \quad (5)$$

Where X is the variable that is being forecasted; the subscripts m and s represent the measured and the simulated values, respectively, and n is the total number of training records.

4 Results and Discussions

The averages, standard deviations and Hurst exponents for the Rainfall time series of all the 19 rain gauge stations are tabulated in Table 1. For all the 19 stations, 384 values of monthly data are used for one-lead monthly prediction. To develop the ANN model, out of the 383 data values, the first 150 are used for training, the next 150 are used for testing and the remaining 83 are used for validation. The optimal architecture is found to be 13 hidden neurons using logistic activation function. The calculated NRMSE for all the 19 stations are also given in Table 1.

It is seen that for all the stations, Hurst exponent ranges from 0.66 to 0.88 and NRMSE ranges from 0.12 to 0.19. As per the hypothesis, a lower value of NRMSE is expected for a high value of Hurst exponent. But, from Table 1, it is seen that only in six stations this hypothesis is found to be true. Hence this outcome doesn't fully support the conclusions arrived at by Qian and Rasheed [1] for the financial data. This may be due to the difference in the characteristics of the financial data and rainfall data. The structures of the rainfall time series do not have serial correlation as much as in financial time series. From this, it can be concluded that Hurst exponent value also depends on the pattern of the data used.

Further, interesting information can be observed in Table 1 that the stations having identical Hurst exponent can have dissimilar NRMSE also. For example, Hurst exponent for the stations Fernshaw, Gladysdale (Little Feet Farm), Seville and Silvan are same i.e. 0.72, but the NRMSE values are different i.e., 0.15, 0.19, 0.16 and 0.16 respectively. This shows evidently that the higher Hurst exponent time series are predictable but not always.

For discussion we considered only Fernshaw and Gladysdale station, due to its identical Hurst exponent with highest deviation in NRMSE. The time series plots of these two stations are shown in Figure 2 and Figure 3 respectively. Both these stations have negative trend which is not statistically significant.

The plot of actual and predicted validation set for Fernshaw and Gladysdale stations are shown in Figure 4 and Figure 5 respectively. Despite, both stations having the same Hurst exponent, the predictability of rainfall for Fernshaw is noticeably better than that for Gladysdale. This leads to the conclusion that there is a need to further investigate the pattern of rainfall and Hurst exponent estimation and it is difficult to conclude clearly about the predictability of rainfall time series with the Hurst exponent alone.

Table 1. Relation between Hurst exponent and NRMSE for all the rain gauges

S.No	Station Name	Average Rainfall (mm)	Standard Deviation of Rainfall (mm)	Hurst exponent	NRMSE
1	Toolangi (Mount St Leonard DPI)	111.14	55.7	0.66	0.17
2	Fernshaw	97.43	52.44	0.72	0.15
3	Black Spur	109.15	58.95	0.67	0.16
4	Maroondah Weir (Melbourne Water)	88.1	44.96	0.71	0.17
5	Healesville (Mount Yule)	74.52	37.9	0.71	0.19
6	Tarrawarra	70.29	37.26	0.74	0.18
7	Tarrawarra Monastery	65.14	33.36	0.66	0.17
8	Coranderrk Badger Weir	114.98	58.36	0.66	0.15
9	Coldstream	68.17	35.99	0.77	0.17
10	Healesville (Valley View Farm)	80.81	40.68	0.69	0.17
11	Seville	76.67	40.12	0.72	0.16
12	Gladysdale (Little Feet Farm)	101.49	51.2	0.72	0.19
13	Powelltown DNRE	114.25	60.63	0.7	0.18
14	Kangaroo Ground	66.24	34.08	0.7	0.14
15	Lilydale	71.64	38.39	0.76	0.15
16	Montrose	97.63	55.79	0.88	0.12
17	Silvan	94.64	48.25	0.72	0.16
18	Monbulk (Spring Road)	105.22	50.9	0.69	0.19
19	Ferny Creek	114.18	59.55	0.78	0.17

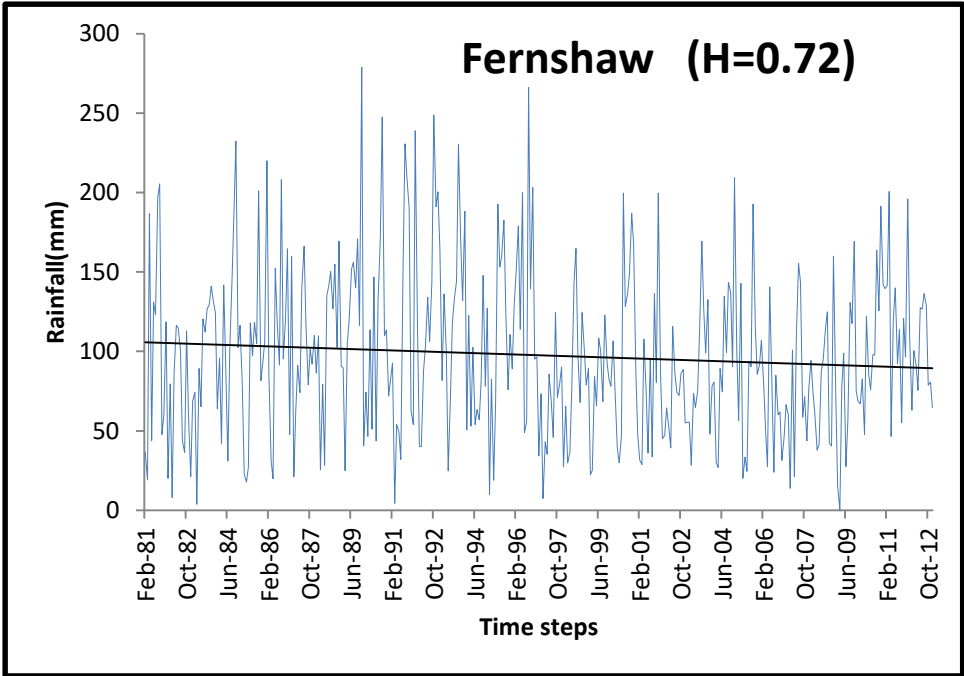


Fig. 2. Time series plot of rainfall at the Fernshaw station

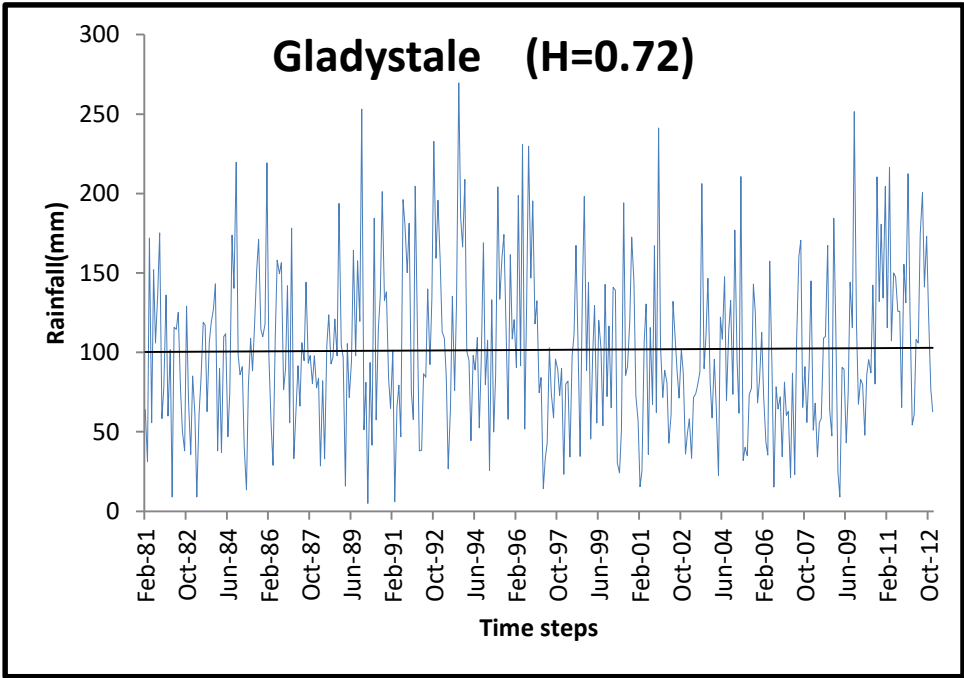


Fig. 3.Time series plot of rainfall at the Gladystale station

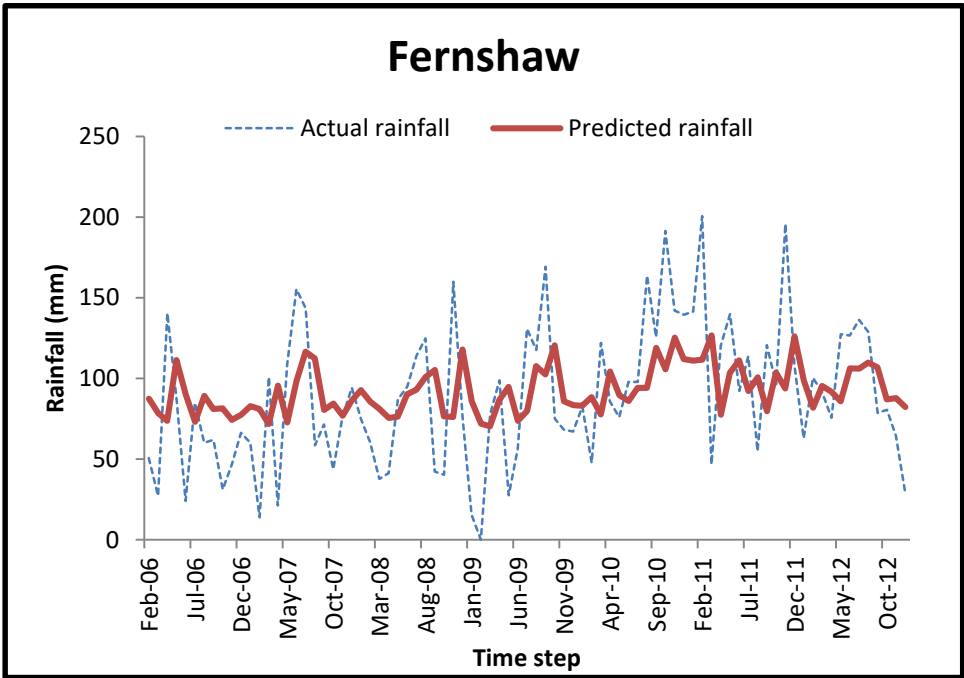


Fig. 4. Comparison of actual vs predicted rainfall of validation set in Fernshaw station

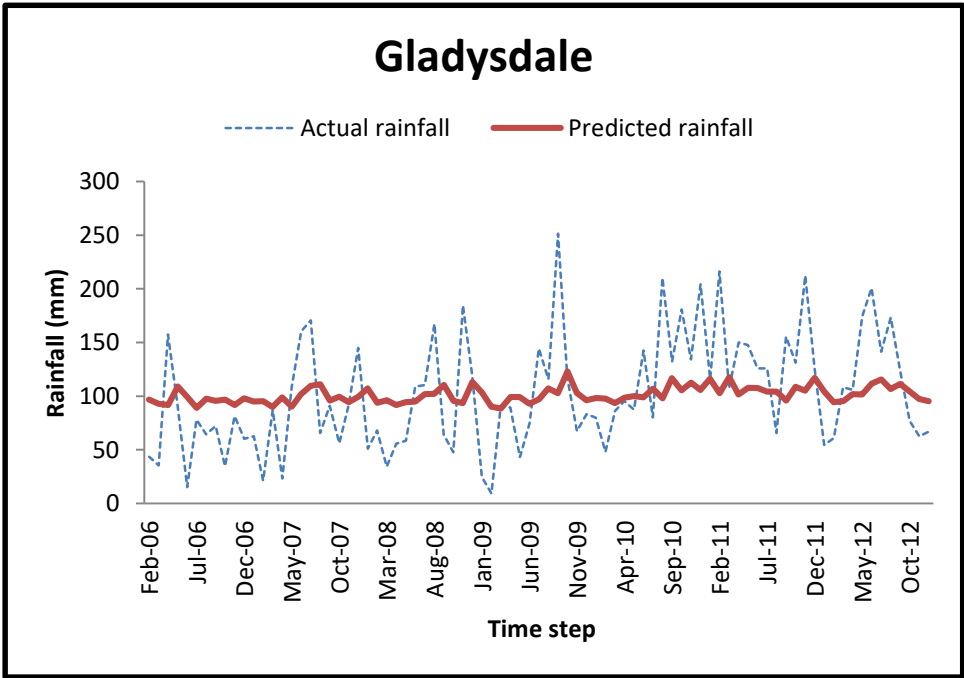


Fig. 5. Comparison of actual vs predicted rainfall of validation set in Gladysdale station

5 CONCLUSIONS

- 1) Out of the nineteen stations considered in this study, only in 6 stations a strong Hurst indicated a stronger predictability.
- 2) The hypothesis that a value of Hurst exponent greater than 0.5 indicates persistence of a time series and hence predictability is not found to hold true always. Thus, this hypothesis needs further investigation.

References

1. B.Q.K. Rasheed, and B., Qian, In IASTED conference on Financial Engineering and Applications FEA 2004, 203-209(2004)
2. A.S. Tokar, and P.A., Johnson, J. Hydrol. Eng. **4**, 232-239 (1999)
3. S. Chattopadhyay, Acta. Geophys. **55**, 369-382 (2007)
4. A.K. Sahai, M.K. Soman, and V. Satyan, Clim. Dynam. **16**, 291-302 (2000)
5. N. Khalili, S.R. Khodashenas, K. Davary, M.M. Baygi, and F. Karimaldini, Arab. J. Geosci. **9**, 624 (2016)
6. N.Z. Mohd-Safar, D.L. Ndzi, D.A. Sanders, H.M. Noor, and L.M. Kamarudin, March. In SAI Intelligent Systems Conference (2016) Intelli. Sys.**751**, 325-348 (2016)
7. A. Dahamsheh, and H. Aksoy, Meteorol. Appl. **16**, 325-337(2009)
8. H. Aksoy, and A. Dahamsheh, Stoch. Env. Res. Risk. **23**, 917-931(2009)
9. C. Sivapragasam, S. Vanitha, N. Muttill, K. Suganya, S. Suji, M.T. Selvi, R. Selvi, and S.J. Sudha, Neural. Comput. Appl. **24**, 1785-1793 (2014)
10. C. Sivapragasam, V.M. Arun, and D. Giridhar, Meteorol. Atmos. **109**, 1-7 (2010)
11. P. Saravanan, C. Sivapragasam, M. Nitin, S. Balamurali, R.K. Ragul, S.S Prakash, G.S Ganesan, and V.V Murugan, December. In International Conference on Lect Notes Comput Sc. Springer, Cham. **10398**, 362-370 (2016)
12. C. Dinu, R., Drobot, C. Pricop, and T.V., Blidaru, Mathematical Modelling in Civil Engineering, **13**, 10-20 (2017)
13. M., Haviluddin, S Hardwinarto, and M., Aipassa, International Conference on Computer Science and Computational Intelligence Procedia Computer Science **59**, 142-151 (2015)
14. R.V., Ramana, B., Krishna, S.R. Kumar, and N.G., Pandey, Water. Resour. Manag. **27**, 3697-3711 (2013)
15. M.S., Babel, G.B. Badgujar, and V.R., Shinde, Meteorol. Appl., **22**, 610-616 (2015)
16. C.L., Wu, K.W. Chau, and C., Fan, J. Hydrol. **389**, 146-167 (2010)
17. S.R., Koirala, R.W., Gentry, E., Perfect, P.J. Mulholland, and J.S., Schwartz, J. Hydrol. Eng. **16**, 717-724 (2010)
18. G. Rangarajan, and D.A., Sant, Chaos. Soliton. Fract. **19**, 285-291(2004)
19. G. Rangarajan, and D.A., Sant, Geophys. Res. Lett. **24**, 1239-1242(1997)
20. S., Rehman, Chaos. Soliton. Fract. **39**, 499-509 (2009)
21. S. Rehman, and A.H., Siddiqi, Chaos. Soliton. Fract.**40**, 1081-1090 (2009)
22. S., Barua, N., Muttill, A.W.M. Ng, and B.J.C., Perera, Hydrol. Process. **27**, 1727-1738 (2013)
23. Melbourne Water, Port Phillip and Westernport Regional River Health Strategy - Yarra catchment, 38-56, (2013)

24. S. Govindaraju, Rao J Hydrol. Eng. **5**, 115-123(2000)
25. S. Govindaraju, Rao.: **5**,124-137(2000)
26. S Simon. Haykin, Tsinghua University Press, (2001).