Trend Analysis of Streamflow Records of Indian River Basins Accounting Long Term Persistence

Ajay Krishna K S, Arjun Harish, Arjun R S
Undergraduate Student
Department of Civil Engineering
TKM college of engineering
Kollam, India

Adarsh S
Associate Professor
Department of Civil Engineering
TKM college of engineering
Kollam, India

Abstract - **The long term hydro-climatic changes influences streamflow variability and the estimation of long term persistence (LTP) may help to improve the accuracy of hydrological modelling. This study first determined the strength of LTP of streamflow of stations located in five major river basins in India using the classical Rescaled range (R/S) analysis. Subsequently, the standard Mann – Kendall (MK1) is used to analyse the trends in streamflow data. Finally, the Mann-Kendall test with variations accounting for long term persistence (MK2) is employed for trend detection. It is found that standard MK test (MK1) showed statistical significance in the data of 33.8% stations, while MK2 test detected statistical significance in only 21.77 % stations. It is concluded that MK1 test overestimate the inherent trend in the datasets due to the presence of long term persistence in streamflow and the latter method is more reliable estimate in long term water resources planning and management.**

Key words - Persistence, Trend, Streamflow, Mann-Kendall

I. INTRODUCTION

Investigating the gradual changes in hydrological variables like streamflow is important for long-term water resources planning and management. The Mann-Kendall (MK) test [1, 2] is a non-parametric test used in many studies by researchers to identify the gradual changes in time series data such as rainfall, temperature and streamflow [3-5]. In particular, streamflow time series exhibit a large spatial variability and the presence of high and low flow years makes their distribution skewed and do not meet the required assumptions of parametric tests [6].Therefore Mann-Kendall test is very much appropriate for trend analysis of streamflow, as it does not require the data to be normally distributed.

The long term persistence or the Hurst phenomenon was introduced by Harold Edwin Hurst in 1951 [7]. It refers to a long memory in the time series data and often expressed as the “tendency of low values to follow low ones and high values to succeed high ones” [8] and the multiple-scale variability of a time series can explain the Hurst phenomenon. Identifying the Hurst phenomenon on a specific series provides some indication of climatic fluctuations [9]. Long term persistence is determined by the Hurst exponent H, which lies between 0 and 1 and value of H > 0.5 indicates long term persistence in the data [10].There are wide range of methods of finding the LTP by determining the Hurst exponent (H) namely Rescaled range method, whittle estimator, Least square variance method, aggregated variance etc. [10]. The presence of LTP can lead to an underestimation of the serial correlation of the data and an overestimation of the significance of the Mann-Kendall test. Studies on streamflow data have found that the standard Mann Kendall test overestimates the results of trend analysis in the presence of long term persistence[11-12]. Hamed [13] again formulated a modified version of Mann Kendall test under the scaling hypothesis which opens the way to accommodate the Hurst exponent and thereby long term persistence into Mann Kendall test [13].Changing hydrology may lead to under-designed or over-designed projects, which may not meet long-term needs; thus, the traditional assumption of stationarity for hydraulic designs is not enough. To address these issues, this study focuses on evaluating trend in streamflow while taking into account streamflow persistence, which affects long-term trends. Since the study need long records of data or finding the long term behaviour, the studies on LTP of streamflow of Indian rivers is rather scarce in literature.

In this study, the presence of serial correlation in stream flow data sets of five major Indian River basins is estimated by the Hurst exponent H which quantifies the LTP in a time series. Trend analysis is carriedout firstly by the standard Mann-Kendall test (MK1) followed by a modified Mann Kendall test (MK2) that could also account for the presence of LTP in the data.

II. METHODOLOGY

A. *Standard Mann Kendall test*

The Mann-Kendall trend test analyzes difference in signs between earlier and later data points**.** The idea is that if a trend is present, the sign values will tend to increase or decrease steadily. Every value is compared to every value preceding it in the time series, which gives a total of n(n – 1) / 2 pairs of data, where “*n*” is the number of observations in the set.Let *x1, x2 . . .* xn is the time series of length n, then the Mann–Kendall test statistic S is given by:

$$s = \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} sgn(x_i - x_j) \qquad (1)$$

$$\text{where } sgn(x) = \begin{cases} 1 \text{ for } x > 0 \\ 0 \text{ for } x = 0 \\ -1 \text{ for } x < 0 \end{cases} \quad (2)$$

$$E(S) = 0 \quad (3)$$

$$V(S) = \frac{n(n-1)(2n+5)}{18} \quad (4)$$

where $E(S)$ is the mean and $V(S)$ is the variance of S. Then the Mann–Kendall z is given by:

$$z = \begin{cases} \frac{S-1}{\sqrt{V(S)}} \text{ for } S > 0 \\ 0 \text{ for } S = 0 \\ \frac{S+1}{\sqrt{V(S)}} \text{ for } s < 0 \end{cases} \quad (5)$$

- Z > 1.96 implies significantly increasing trend at 5% significance level
- Z < - 1.96 implies significantly decreasing trend at 5% significance level

B. Modified Mann Kendall test

Serial correlation in data can be accounted in the Mann Kendall test by either treating the data for its effects or by properly modifying the test. Here the test was modified by introducing a new variance that accounts LTP by accommodating the value of H as suggested by Hamed [13]

Correlation matrix for a given Hurst exponent is

$$C_n(H) = [\rho_{|j-i|}], for\ i = 1{:}n, j = 1{:}n \quad (6)$$

$$\rho_l = \frac{1}{2}\left(|l+1|^{2H} - 2|l|^H + |l-1|^{2H}\right) \quad (7)$$

Where ρ_l is the autocorrelation function for lag l for a given H and the modified variance is given by

$$V(S) = \sum_{i<j}\sum_{k<l}\frac{2}{\pi}sin^{-1}\left(\frac{\rho|j-i| - \rho|i-l| - \rho|j-k| + \rho|i-k|}{\sqrt{((2-2\rho|i-j|)(2-2\rho|k-l|)}}\right) \quad (8)$$

The newly obtained variance is used in equation 5 for getting the results.

C.Methods to find Hurst coefficient

Hurst coefficient is evaluated using dispersonal analysis (DA) and rescaled range(R/S) statistics method

i. Rescaled Range (R/S) Analysis

Different steps involved in the R/S method of estimation of H are

1. For a given time series data $x_1, x_2, x_3 \ldots x_n$compute the mean m

$$m = \frac{1}{n}\sum_{i=1}^{n} x_i \quad (9)$$

2. Create a mean adjusted series,

$$Y_t = X_t - m \text{ for } t = 1,2,3, \ldots\ldots, n \quad (10)$$

3. Cumulative deviate series Z_t is calculated as:

$$Z_t = \sum_{i=1}^{t} Y_i \quad (11)$$

4. Range series R is given by:

$$R_t = max(z1, z2, z3, \ldots zt) - min(z1, z2, z3 \ldots, zt) \quad (12)$$

4. Standard deviation series S is given by,

$$S_t = \frac{1}{t}\sum_{i=1}^{t}\left(x_i - m(t)\right)^2 \quad (13)$$

5. Rescaled range (R/S) is calculated from equation 4 and 5,

$$(R/S)_t = \frac{R_t}{S_t} \quad (14)$$

6. A linear plot between ln (R/S) and ln (n) is made

7. In case of long memory in the series,

$$E\left(\frac{R}{S(n)}\right) \sim Cn^H as\ n \to \infty \quad (15)$$

Where, C is a constant; n is the number of observations and H is the Hurst exponent;
Taking logarithm of the abovefunction, we get

$$ln\left(E\left(\frac{R(n)}{S(n)}\right)\right) = Hln(n) + \ln(C) \quad (16)$$

From equation 8 the value of H can be found as the slope of the line of the linear equation.

ii. Dispersional analysis

Dispersional analysis [14, 15] (also known as Aggregated Variance Method) averages the differenced series over bins of width τ and calculates the variance of the averaged dataset. Consider a time series $X(t)$, t= 1,2... N.

1. Set the bin size τ =1
2. Calculate the standard deviation of the N data points and record the point $(\tau, \tau.\sigma_\tau)$
3. Average neighbouring data points and store in the original dataset

$$X(t) \leftarrow \frac{1}{2}[X(2t-1) + X(2t)] \quad (17)$$

and rescale N and τ appropriately

$$N \leftarrow \frac{N}{2} \text{ and } \tau \leftarrow 2\tau \quad (18)$$

4. As long as more than four data points remain (N>4), return to step 2
5. Perform a linear regression on the log-log graph $\log(\tau.\sigma_\tau) = H\log(\tau) + C$ the calculated slope is the best estimate of H

III. STUDY AREA AND DATA

This study is conducted on the streamflow data of 124 gauging stations located atfive major river basins in India namely Krishna, Cauvery, Mahanadi, Godavari and west flowing rivers of Tadri to Kanyakumari (WFR). The 5 basins can be identified from the map of major river basins in India shown in Figure 1.

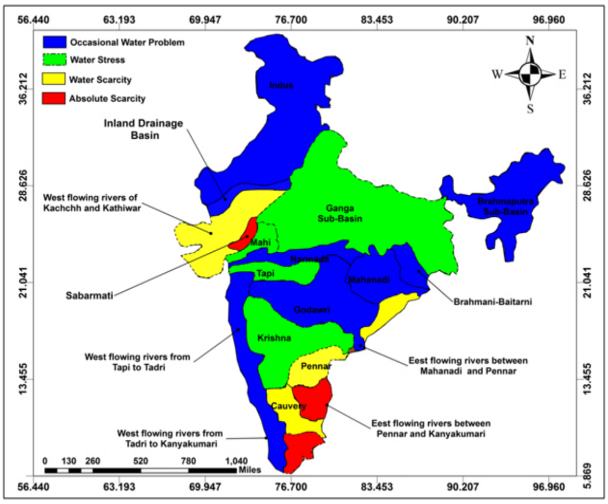

Fig.3.1.Map of river basins in India [16].

The streamflow data wereobtained from the records of Central Water Commission (CWC) (http://indiawris.gov.in/wris/). Streamflow data for all the stations were checked for discontinuity in the time series data and such stations with missing data were avoided. The length of data available for various gauging stations varies and is listed in Table 1. Streamflow records are available from 1966 but not for all stations. Most number of stations used in the analysis have a data length of more than 10 years and most number of stations (33.06%) have data length between 20 to 30 years. Continuous data for alarge period of above 40 years is available only for 18% of total stations studied. Numbers of stations that have data in various range of years are shown in Table 1.

TABLE1. Statistics of discharge data of all gauging stations

	10 to 20 yrs	20 to 30 yrs	30 to 40 yrs	40 to 50 yrs	Total
Mahanadi	3	9	5	2	20
WFR	4	11	6	1	23
Krishna	8	9	14	4	35
Cauvery	4	2	5	4	15
Godavari	7	10	2	12	31
Total	26	41	32	23	124

IV RESULTS AND DISCSUSSIONS

A. *Persistence analysis*

Strength of Long-term persistence was determined by evaluating the Hurst exponent H. Rescaled range analysis and dispersonal analysis were used to evaluate H for the daily streamflow data records of all the 124 stations.Obtained values of H implies that long term persistence is a property for all the stations with an average value of H as 0.6831 and 0.6945 using Dispersonal analysis as well as R/S analysis respectively.The average value of H for different basins are listed in Table 2.

TABLE 2: Average value of H for different basins

Name of basin	**Number of stations**	**Average Value of H**	
		Dispersonal Analysis	R/S analysis
Cauvery	20	0.7313	0.7443
WFR	23	0.684	0.6906
Krishna	35	0.691	0.7152
Mahanadi	15	0.6669	0.6759
Godavari	31	0.6609	0.662
Total	124	0.6831	0.6945

The average values are comparable suggesting that the values of H are similar using the two different methods. However, for different stations disparities werenoticed in the Hurst exponent obtained from two methods this may be attributed to the length of the time series data and its influence on different methods used. Cauvery basin is showing strong long term persistence with H value of 0.7313 and 0.7343 respectively. As a whole it is observed that average Hurst exponentfrom Rescaled range analysis is greater than that obtain by the Dispersonal analysis for all the 5 basins studied. But it is not so for individual stations. 95% of stations have H value in the range 0.6 to 0.8 in from dispersonal analysis and 90% in the same range using R/S analysis.Number of stations showing the strength of persistence in different intervals with H value varying from 0.5038 to 0.8022 in dispersonal analysis and 0.5014 to 0.9256 in R/s analysis are listed in Table 3.

TABLE3: Number of stations with H value in different range

H	**Dispersonal Analysis**	**RS analysis**
<.5	0	1
.5-.6	4	3
.6-.7	76	68
.7-.8	43	44
.8-.9	1	6
>.90	0	2
Total	124	124

i.Spatial variations of LTP

It is evident from dispersonal analysis that gauging stations in the southern river basins are showing strong LTP than that of the rest. That is 66.66% of stations of Cauvery, 45.71% of stations falling in Krishna basin and 47.82% of stations of WFR show H value in the range of 0.7 to 0.8 whereas it is 10% and 16.12% for Mahanadi and Godavari respectively. In other words, 36 out of 44 stations showing H greater than 0.7 lies in the relatively southern basins. Thus trend is same for results from R/S analysis also. The spatial distribution of H values are shown in Figure 2.

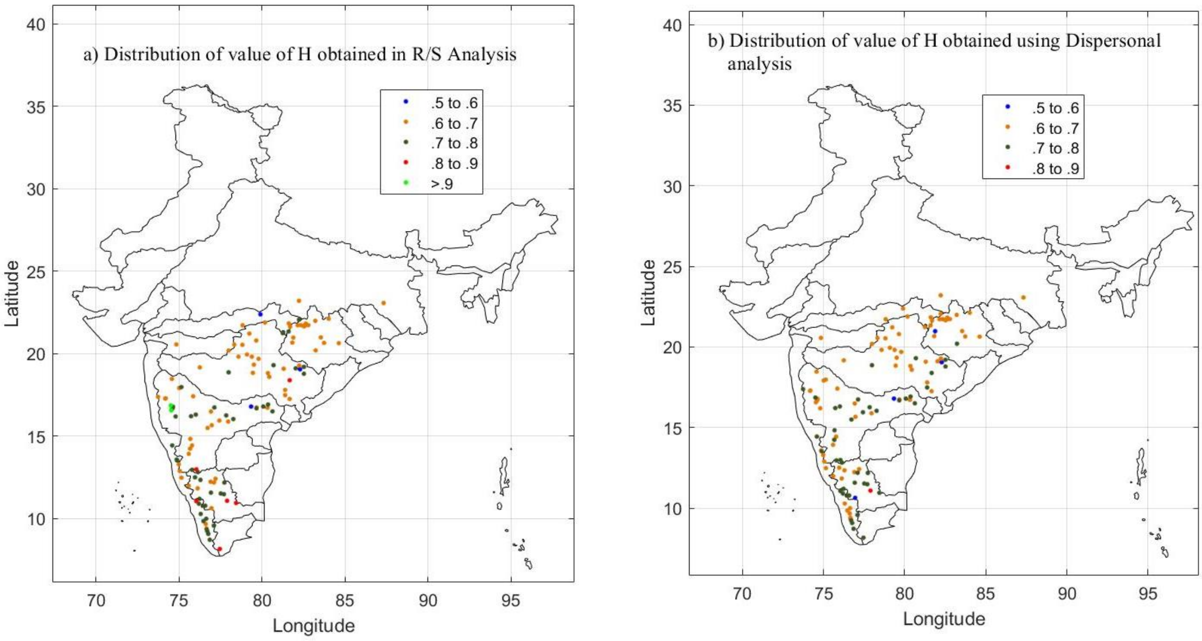

Figure 2: Spatial distribution of the Hurst coefficient
2a) H values obtained using Dispersonal analysis
2b) H values obtained using R/S analysis

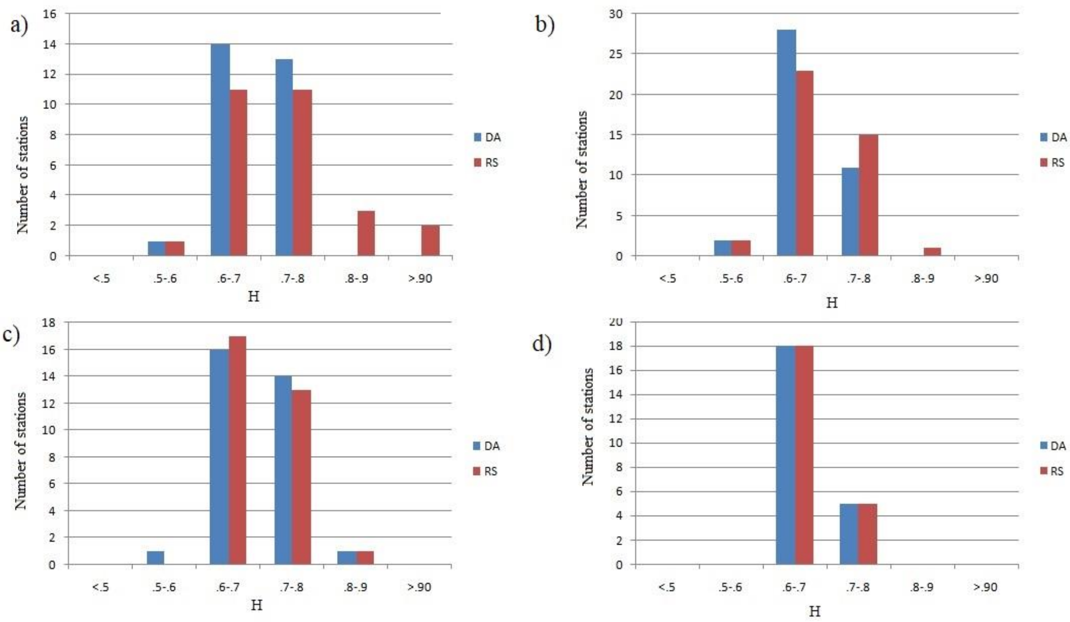

Figure 3: Influence of data length on evaluation of H

a) length < 20 years b)20-30 years c) 30-40 years d) 40-50 years *ii. Influence of data length on LTP*

The gauging stations used in the study have varying data length from less than 10 to a maximum of 50 years.This may be due to various reasons like closing of gauging stations and opening new ones that makes the availablestreamflow records limited in length. Since the same test was done to all the different gauging stations it is interesting to see how the length of available data affects these tests. Figure 3 shows the number of stations that have values of H in different ranges. Figure 3a is the graph for stations with less data length that is less than 20 years. Similarly figures 3b 3c and 3d are for the stations having data length between 20 t0 30, 30 to 40 and 40 to 50 years.

From the graphs it can be seen that there is no increase in strength of LTP with an increase in data length. This was not an expected trend. In all the four cases more number of stations are showing H in the range 0.6 to 0.7 followed by 0.7 to 0.8. The higher values of H are also found in shorter data length but it cannot be taken as a trend since only R/S analysis shows it. And for larger data lengths both analysis shows similar results unlike the other. The average H value for stations decreases when more data length was available. The two stations with least data length of 10 years show an average H value of 0.7551. For stations with data length between 10 and 20 years it is 0.700 and the average is 0.6739 for stations with the most available data length. This is same for both the methods while R/S analysis shows higher values in each case.

B.*Trend analysis*

i. Standard Mann Kendall test (MK1)

Standard Mann Kendall test is done for monthly average flows of all the 124 stations at 5% significance level.Streamflows of 59 out of 124 stations shown increasing trends but the records of only 21 stations shown significantly increasing trend (Z value >=1.96). Therest of 82 stations do not show any significant trend. The rest 21 stations shows significantly decreasing trends. That means the number of stations with significant increase as well as significant decrease are equal.

Considering various basins, for WFR streamflows of more than 82% stations do notexhibit any significant trend. Streamflows of most of the stations in Krishna are showing a significantly decreasing trend. In rest of all basins, both significantly increasing and decreasing trends are similar. Stations with significantly increasing and decreasing trend in streamflows are well distributed in all over the basins studied. There is no evident aggregation of streamflow trend at particular regions. The summary of trend analysis (MK1) is in Table 4.

TABLE4: Number of stations showing significant trend in MK1 test

	Number of stations		
	Significantly increasing trend	No significant trend	Significantly decreasing trend
Mahanadi	5	15	0
WFR	3	19	1
Krishna	6	19	10
Cauvery	2	10	3
Godavari	5	19	7
Total	21	82	21

ii. Modified Mann Kendall test (MK2)

An important objective of this study is to find the effect of long term persistence in identification of trend using the Mann-Kendall test. For accounting serial correlation the test is modified by changing the variance and accounting the Hurst exponent whileperforming the MK test. Since all the station exhibits LTP and have the value of H above 0.5 the modified test can be done for all the 124 stations. The result of MK2 is shown in Table5.

TABLE5: Number of stations showing significant trend in MK2 test

	Number of stations		
	Significantly increasing trend	No significant trend	Significantly decreasing trend
Mahanadi	4	16	0
WFR	1	22	0
Krishna	6	24	5
Cauvery	1	13	1
Godavari	3	22	6
Total	15	97	12

It is evident that the stations showing significant trends have reduced from 33.8% (MK1) to 21.77% (MK2). This shows an overestimation in identifying significant trend in the presence of LTP in the time series data.Significantly increasing trends are reduced from 16%(MK1) to 12%(MK2) and significantly decreasing trends have reduced from 16% (MK1) to 9% (MK2). The spatial distribution of Mann-Kendall z value of MK1 and MK2 are shown in Figure 4a and Figure 4b respectively. The Mann Kendall z value also reduces in the modified version of the Mann Kendall test.

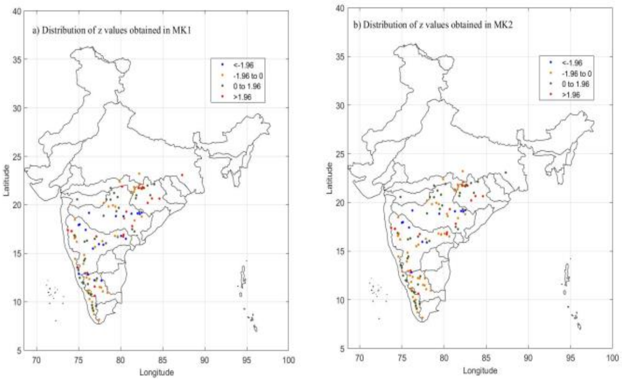

Figure 4 : Spatial distribution of Mann Kendall z values
4a) MK1 z values   4b) MK2 z values

V CONCLUSIONS

Long term persistence is one of the important domain of research in hydro climatic studies. The widely used Mann Kendall test for detecting trend in time series data has to be modified to account the serial correlation in data, which is against the assumption of independency of data. The following are the important conclusions made from this study:

- All the streamflow records considered in the study displayed Long term persistence.
- The strength of LTP is more in southern river stations
- Data length used in analyzing LTP is found to be influencing the R/S analysis. It can be seen that greater data length shows a weaker LTP by a lower value of H.
- Streamflow of majority of the stations do not show any significant trend by both methods of MK test
- Number of stations having streamflow with significant trend have reduced from 42 to 27 when modified Mann Kendall test was employed
- The standard Mann Kendall test overestimates the identification of significant trends in the presence of LTP and therefore use of modified Mann Kendall test is a more reliable option for long-term water resources planning in India.

VI REFERENCES

[1] H.B.Mann, "Non-parametric tests against trend", Econometrica, vol.13, pp. 245–259, 1945

[2] M.G. Kendall, Rank Correlation Methods. Charles Griffin: London, UK, 1975

[3] S.K. Jain, V. Kumar, "Trend analysis of rainfall and temperature data for India-A Review", Curr. Sci, vol. 102(1), pp. 37-49, 2012

[4] P. Sonali, D. Nagesh Kumar, "Review of trend detection methods and their application to detect temperature change in India" , J. Hydro, vol. 476, pp. 212-27, 2012

[5] N.S. Abeysingha, M. Singh andV.K. Sehgal, "Analysis of trends in streamflow and its linkages with rainfall and anthropogenic factors in Gomti River basin of North India", Theoret.App. Climatol, vol. 123, pp. 785–799, 2016

[6] K.P. Panda., A. Kumar, A. Ghosh andR.K. Mohanthi, "Streamflow trends in the Mahanadi River basin (India): Linkages to tropical climate variability",J. Hydrol, vol.495, pp.135–149, 2013

[7] H.E. Hurst, "Long term storage capacities of reservoirs", Transactions of the American Society of Civil Engineers, vol. 116, pp. 776–808, 1951

[8]Y. Markonis, Y. Moustakis, C. Nasika, P. Sychova, P. Dimitriadis, M. Hanel et al.,"Global estimation of long-term persistence in annual river runoff", Adv. in wat. Resour., vol. 113, pp.1-12, 2018

[9]D. Koutsoyiannis, "Climate change, the Hurst phenomenon, and hydrological statistics", Hydrol. Sci. J, vol. 48 (1), pp. 3-24, 2003

[10] E. Szolgayova, G. Laaha, G. Blöschl andC. Bucher,"Factors influencing long range dependence in streamflow of European rivers", Hydrol. Process, vol. 28, pp.1573–1586, 2014

[11] S. Sagarika, V KalraXand S. Ahmad, " Evaluating the effect of persistence on long-term trendsand analyzing step changes in streamflows of the continental United States", J. Hydro,. Vol.517, pp. 36-53, 2014

[12] S. Kumar, W. Merwade,J. Kam and K. Thurner"Streamflow trends in Indiana: Effects of long term persistence, precipitation and subsurface drains", J. Hydrol., vol. 374 , pp. 171–183, 2009

[13] K.H. Hamed, "Trend detection in hydrologic data: TheMann–Kendall trend test under the scalinghypothesis", J. Hydrol, vol. 363, pp. 349:350, 2008

[14] J.B. Bassingthwaighte, G.M. Raymond. "Evaluating rescaled range analysis for time series", Annals of Biomedical Engineering, vol. 22, pp. 432–44, 1994

[15] H.J. Blok.. On the nature of the stock market: Simulations and experiments. PhD thesis, University of British Columbia, 2000. http://rikblok.cjb. net/lib/blok00b.html, arXiv:cond-mat/0010211, 2000

[16] U.K. Singh, B. Kumar, "Climate change impact on hydrology and water resources of Indian river basins", Current world environment, vol. 13, pp.32-43, 2018