Statistical Trend Analyses of Different Time Series Data

SURJEET SINGH
National Institute of Hydrology, Roorkee

INTRODUCTION

A time series is a sequence of observations that are ordered in time or space. There can be different types of time series viz. rainfall, runoff, groundwater level, water quality, temperature, evaporation, etc. All these data series exhibit variation of magnitude with time for any particular location. Since the time can be a common factor during a particular period for some time series, there is always possibility of existence of mutual relationships. Statistically, these mutual relationships can be represented by the correlation among the time series. For example, rainfall generally has strong relationship with runoff, groundwater level, etc. Temperature is generally correlated with evaporation, water quality, etc.

The magnitude of a time series changes from time to time because of dependence of any time series variable on multiple parameters. This is the reason why time series show trend with time. The trend in a time series gives important information whys any particular variable is showing a rising or falling trend. For our hydrological system, these informations are of great concern and are useful for warning purpose. For example, a general declining trend of ground water levels may either be due to decrease in groundwater recharge or increasing groundwater draft. So the trend in any time series plays a vital role to identify the reasons for the cause and avoid any future consequences by taking appropriate actions. Trends in rainfall in response to climate change have been studied by various researchers (Jayawardene et al., 2005; Parta and Kahya, 2006; Obot et al., 2010; Kumar and Jain, 2010; Olofintoye and Adeyemo, 2011; Manikandan and Tamilmani, 2012). These researchers have emphasized that the knowledge of rainfall variations is essential for the proper water management practices. A comprehensive knowledge of the trend and persistence in rainfall of the area is of great importance because of economic implications of rain sensitive operations (Galkate, et al. 1999).

Trend analysis in recent years has gained an important role with the advent of climate change as it offers the first-hand diagnose whether any climatic variable is showing any rising or falling trend. Intergovernmental Panel on Climate Change (IPCC, 2007) has reported that future climatic change is likely to affect agriculture, increase of risk of hunger and water scarcity and may lead to rapid melting of glaciers. Rainfall is one of the main climatic factors that can indicate the climate change (Obot et al., 2010). Kumar and Jain (2010) reported that a higher or lower or changes in its distribution would influence the spatial and temporal distribution of runoff, soil moisture, groundwater reserves and would alter the frequency of droughts and floods. While the observed monsoon rainfall at the all-India level does not show any significant trend, regional monsoon variations have been recorded. A trend of increasing monsoon seasonal rainfall has been found along the west coast, northern Andhra Pradesh, and north-western India (+10% to +12% of the normal over the last 100 years) while a trend of decreasing monsoon seasonal rainfall has been observed over eastern Madhya Pradesh, north-eastern India, and some parts of Gujarat and Kerala (6% to 8% of the normal over the last 100 years) (Kumar and Singh, 2011).

Trend analysis can be equally performed on rainfall, runoff, groundwater levels, water quality data, etc at the time scale depending on the need. With this in view, this paper presents the concept of trend analysis has been presented using parametric and non-parametric approaches.

TECHNIQUES FOR TREND ANALYSIS

There are many techniques for the detection of trends in any data series. The most common and widely used technique is the simple linear regression model. There are two broad approaches identification of trends in the data series:

1. Parametric approach, and
2. Non-parametric approach.

Parametric Approach

Linear regression analysis is a parametric approaches and one of the most commonly used methods to detect a trend in a data series. This model develops a relationship between two variables (dependent and independent) by fitting a linear equation to the observed data. The data is first checked whether or not there is a relationship between the variables of interest. This can be done by done using the scatter plot. If there appears no association between the two variables, linear regression model will not prove a useful model. A numerical measure of this association between the variables is the correlation coefficient, which range between -1 to +1. A correlation coefficient value of ± 1 indicates a perfect fit. A value near zero means that there is a random, nonlinear relationship between the two variables. The linear regression model is generally described by the following equation

$$Y = m.X + C \tag{1}$$

where *Y* is the dependent variable, *X* is the independent variable, *m* is the slope of the line and *C* is the intercept constant. The coefficients (*m* and *C*) of the model are determined using the Least-Squares method, which is the most commonly used method. Statistical tests are used to determine whether the linear trends are significantly different from zero or not at the 5% significance level.

Non-Parametric Approaches

Parametric approaches assume that the data conforms to some probability distribution and makes inferences about the parameters of the distribution. Non-parametric approaches do not rely on data belonging to any distribution. For example, a histogram is a simple non-parametric estimate of a probability distribution. Generally parametric methods make more assumptions than the non-parametric methods. Mann-Kendall and Sen's slope estimator are the main non-parametric techniques used for detection of trends in any data series.

Mann-Kendall Test

The trend analysis and estimation of Sen's slope are done using Kendall (1975) and a Sen (1968) method, respectively for the given data sets. Man-Kendall test is a non-parametric test for finding trends in time series. This test compares the relative magnitudes of data rather than data values themselves (Gilbert, 1987). The benefit of this test is that data need not to conform any particular distribution. In this test, each data value in the time series is compared with all subsequent values. Initially the Mann-Kendall statistics (S) is assumed to be zero, and if a data value in subsequent time periods is higher than a data value in previous time period, S is incremented by 1, and *vice-versa*. The net result of all such increments and decrements gives the final value of S. The Mann-Kendall statistics (S) is given as

$$S = \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} sign(x_j - x_i) \tag{2}$$

where
$$sign(x_j - x_k) = 1, \text{ if } (x_j - x_k) > 0$$
$$= 0, \text{ if } (x_j - x_k) = 0$$
$$= -1 \text{ if } (x_j - x_k) < 0$$

A positive value of S indicates an increasing trend, and a negative value indicates a decreasing trend. However, it is necessary to perform the statistical analysis for the significance of the trend. The test procedure using the normal-approximation test is described by Kendall (1975, p55). This test assumes that there are not many tied values within the dataset. The variance (S) is calculated by the following equation

$$Var\,(S) = \frac{1}{18}\left[n(n-1)(2n+5) - \sum_{p=1}^{g} t_p(t_p-1)(2t_p+5)\right]$$

where n is the number of data points, g is the number of tied groups and t_p is the number of data points in the p^{th} group.

The normal Z-statistics is computed as:

$$Z = \begin{cases} \dfrac{S-1}{\sqrt{Var(S)}}, & if\ S > 0 \\ 0 & , if\ S = 0 \\ \dfrac{S+1}{\sqrt{Var(S)}}, & if\ S < 0 \end{cases}$$

The trend is said to be *decreasing* if Z is negative and the computed Z-statistics is greater than the z-value corresponding to the 5% level of significance. The trend is said to be *increasing* if the Z is positive and the computed Z-statistics is greater than the z-value corresponding to the 5% level of significance. If the computed Z-statistics is less than the z-value corresponding to the 5% level of significance, there is *no trend.*

Sen's Slope Estimator

Simple linear regression is one of the most widely used model to detect the linear trend. However this method requires the assumption of normality of residuals (McBean and Motiee, 2008). Viessman et al. (1989) reported that many hydrological variables exhibit a marked right skewness partly due to the influence of natural phenomena and do not follow a normal distribution. Thus the Sen (1968) slope estimator is found to be a powerful tool to develop the linear relationships. Sen's slope has the advantage over the regression slope in the sense that it is not much affected by gross data errors and outliers. The Sen's slope is estimated as the median of all pair-wise slopes between each pair of points in the dataset (Thiel, 1950; Sen, 1968; Helsel and Hirsch, 2002). Each individual slope (m_{ij}) is estimated using the following equation

$$m_{ij} = \frac{(Y_j - Y_i)}{(j - i)} \tag{3}$$

where $i = 1$ to $n\text{-}1$, $j = 2$ to n, Y_j and Y_i are data values at time j and i ($j > i$), respectively. If there are n values of Y_j in the time series, there will be $N=n(n-1)/2$ slope estimates. The Sen's slope is the median slope of these N values of slopes. The Sen's slope is

$$m = m_{\left[\frac{N+1}{2}\right]}\ , if\ n\ is\ odd$$

$$m = \frac{1}{2}\left(m_{\left[\frac{N}{2}\right]} + m_{\left[\frac{N+2}{2}\right]}\right),\ \ if\ n\ is\ even$$

Positive Sen's slope indicates rising trend while negative Sen's slope indicates falling trend.

Data Availability

For finding the trend in various kinds of time series, the data of Sagar district (Madhya Pradesh) is used. The following data series are used:

1. Rainfall of Sagar station (1972-2003).
2. Estimated Runoff at the Garhakota gauging site (1972-2003).
3. Groundwater level of a well in Sonar sub-basin (January, 2000 to November, 2005).
4. Average water quality data of Sagar lake (May, 2006 to May, 2008).

Mann-Kendall's Trend Analysis

The summary of Mann-Kendall analysis of different time series is presented in Table 1. The negative S-statistics indicates a falling trend and a positive S-statistics indicates a rising trend in the data series. It can be seen from the table that rainfall and runoff time series show a falling trend and falling trend is significant at the 5% significance level. The groundwater levels are showing a rising trend and the trend is significant at the given significance level. In case of water quality data, different variables are indicating mixed response. Temperature, DO, total alkalinity, total hardness and total phosphate do not show any trend with time. pH shows a significant falling trend while other parameters nitrate, chloride and total iron show a significant rising trend.

Magnitude of above trends is calculated using Sen's slope estimator method. The estimated Sen's slope and their confidence limits are presented in Table 2. A positive slope gives rising trend and *vice-versa*. It can be seen from the Table 4 that in some cases trend is falling and in some cases trend is rising, as already detected using the Mann-Kendall analysis. The slopes, that show significant trend, can also be used to extrapolate the time series for future. The slopes estimated based on the regression analysis are also presented in Table 2. An analysis was also carried-out to check the test of significance for these regression slopes under various cases at 5% significance level and the significant slopes was indicated with star mark (*).

Table 1. Summary of Mann-Kendall trend analysis of various time series

S.No.	Time Series	S-Statistics	z-statistics$_{comp}$	Trend	Trend at 5% Significance Level
1	Annual Rainfall	-56	-0.892	Falling	No
2	Annual Runoff	-36	-0.568	Falling	No
3	Groundwater Level	85	2.548	Rising	Yes
4	Water Quality Data				
I	Temperature	5	0.219	No Trend	No
II	pH	-53	-2.855	Falling	Yes
III	DO	-10	-0.493	No Trend	No
IV	Total Alkalinity	3	0.109	No Trend	No
V	Total Hardness	-1	0.000	No Trend	No
VI	Nitrate	61	3.285	Rising	Yes
VII	Chloride	37	1.971	Rising	Yes
VIII	Total Phosphate	7	0.329	No Trend	No
IX	Total Iron	63	3.394	Rising	Yes

Table 2. Summary of Sen's slope estimator and regression slope of different time series

S.No	Time Series	Sen's Slope	Trend	Confidence Limits for slope at 5% Significance Level	Regression Slope
1	Annual Rainfall	-3.896	Falling	Lower Limit = -14.691 Upper Limit = 6.349	-4.365
2	Annual Runoff	-1.529	Falling	Lower Limit = -11.259 Upper Limit = 8.325	-2.969
3	Groundwater Level	0.224	Rising	Lower Limit = 0.038 Upper Limit = 0.422	0.269*
4	Water Quality Data				
i	Temperature	0.053	Rising	Lower Limit = -0.437 Upper Limit = 0.639	0.039

ii	pH	-0.026	Falling	Lower Limit = -0.058 Upper Limit = -0.008	-0.028*
iii	DO	-0.027	Falling	Lower Limit = -0.132 Upper Limit = 0.091	-0.015
iv	Total Alkalinity	0.199	Rising	Lower Limit = -5.047 Upper Limit = 5.606	-0.258
v	Total Hardness	-0.445	Falling	Lower Limit = -4.566 Upper Limit = 7.800	-0.931
vi	Nitrate	1.470	Rising	Lower Limit = 0.710 Upper Limit = 2.586	1.784*
vii	Chloride	2.307	Rising	Lower Limit = -0.054 Upper Limit = 4.138	1.875**
viii	Total Phosphate	0.008	Rising	Lower Limit = -0.018 Upper Limit = 0.045	0.023**
ix	Total Iron	0.173	Rising	Lower Limit = 0.097 Upper Limit = 0.269	0.161*

(*) indicates significant at 5% level, and (**) indicates significant at 10% level.

CONCLUSIONS

This paper describes the identification of trends in any time series data. The trend analysis is carried out using the Mann-Kendall analysis, Sen's Slope estimator and simple regression approach. The analysis is illustrated by performing trend analysis on various time series viz. rainfall, runoff, groundwater levels, and water quality variables. The data used under this paper are taken from Sagar (Madhya Pradesh) from the previous studies. The detection of trends and computation of their magnitude and significance are presented and discussed in the paper. The regression slopes, their magnitude and significance are also presented and discussed.

REFERENCES

Galkate, R.V. and Thomas, T. (1999). Statistical analysis of rainfall in Sagar Division, NIH Report No. CS(AR) 5/98-99.

Gilbert, R.O. (1987). Statistical Methods for Environmental Pollution Monitoring. Van Nostrand Rienhold Company, Inc., New York.

Helsel, D. R., and Hirsch, R. M. (2002). Statistical methods in water resources: U.S. Geological Survey Techniques of Water-Resources Investigations, book 4, chap. A3, 524 p. Available online at http://water.usgs.gov/pubs/twri/twri4a3/.

IPCC (2007). Summary for policymakers In: Climate Change 2007: The Physical Sciences Basis (ed. By S. Solomon, D. Qin, M. Manning, Z. Chen, M. Marquis, K.B. Averyt, M. Tignor and H.L. Miller). Cambridge University Press, UK.

Jayawardene, H.K.W.I., Sonnadar, D.U.J. and Jayewardene, D.R. (2005). Trends of rainfall in Sri Lanka over the last century, Sri Lankan J. Phys., 6, 7-17.

Kendall, M.G. (1975). Rank correlation methods, 4th ed. Charles Griffin, London.

Kumar, C.P. and Singh, S. (2011). Impact of climate change on water resources and National Water Mission of India, National Seminar on Global Warming and Its Impact on Water Resources held on 14th January, 2011 at Kolkata.

Kumar, V. and Jain, S.K. (2010). Rainfall trends in Ganga-Brahmputra-Meghna river basins of India (1951-2004), Hydrology Journal, 33 (Special Issue), 59-66.

Manikandan, M. and Tamilmani, D. (2012). Statistical analysis of spatial pattern of rainfall trends in Parambikualam Aliyar sub-basin, Tamil Nadu, IWRS, 32(1-2): 40-49.

McBean, E and Motiee, H. (2008). Assessment of Impact of Climate Change on Water Resources : A Long Term Analysis of the Great Lakes of North America, Hydrology andEarth System Sciences, Vol. 12, pp 239–255.

Obot, N.I., Chendo, M.A.C., Udo, S.O. and Ewona, I.O. (2010). Evaluation of rainfall trends in Nigeria for 30 years (1978-2007), Int. J. Phys. Sci., 5(14), 2217-2222.

Olofintoye, O. and Adeyemo, J. (2011). The role of global warming in the reservoir storage drop at Kainji dam in Nigeria, Int. J. Phy. Sci., 6(19), 4614-4620.

Parta, T. and Kahya, E. (2006). Trend analysis in Turkish precipitation data, Hydrol. Process, 20, 2011-2026.

Sen, P.K. (1968). Estimates of the regression coefficient based on Kendall's tau. Journal of the American Statistical Association. 63:1379-1389.

Thiel, H. (1950). A rank-invariant method of linear and polynomial regression analysis, Part 3.Proceedings of Koninalijke Nederlandse Akademie van Weinenschatpen A. 53:1397-1412.

Viessman, .W, Krapp, J.W and Harbough, T. E. (1989). Introduction to Hydrology, Third edition, Harper and Row Publishers Inc., New York.