Mann-Kendall Test - A Novel Approach for Statistical Trend Analysis

Neel Kamal[#1], Dr. Sanjay Pachauri[*2]

[#] *Research Scholar, Faculty of Computing and Information Technology, Himalayan University Arunachal Pradesh*

[*2]*Head of Department, CSE/IT, IIMT College of Engineering,Greater Nodia, U.P.*

Abstract

Trend Analysis is aimed at projecting both current and future movement of observations through the use of time series data analysis which involves comparison of data over a sequential period of time to spot a pattern or trend. Mann-Kendell test is one of the most popular non-parametric trend test based on ranking of observations. The current paper describes Mann Kendall Test in the context of time series data analysis. It also presents a case study to demonstrate the implementation and advantage of using Mann Kendall Test over other trend analysis techniques

Keywords — *Trend Analysis, Time Series Data, Non-parametric Test, Mann Kendall Test.*

I. INTRODUCTION

Continuous data arise in most areas of life. Familiar domain examples include meteorology, biological science, water quality, ecosystem behavioural properties, medicine, machine measurements etc. (Tiwari et. al., 2014). Techniques for evaluating time series datasets falls into two major classes, categorized by either they make hypothesis about the distribution of the data. Standard deviation and the variance are the two main parametric techniques used statistically evaluate the distribution of the continuous variables. Technique which implement distributional hypothesis is comes under the umbrella of parametric techniques; some of the mostly used parametric techniques are t-tests and analysis of variance for comparing clusters, and least squares regression and correlation for studying the relation between variables. All of the common parametric techniques (“t techniques”) assume that in some way the data follow a normal distribution and also that the spread of the data (variance) is uniform either between groups or across the range being studied. Alternative Techniques which do not require us to make distributional assumptions about the data, such as the rank Techniques, are called non-parametric techniques. The term non-parametric applies to the statistical Method used to analyse data, and is not a property of the data. As tests of significance, rank Techniques have almost as much power as t techniques to detect a real difference when samples are large, even for data which meet the distributional requirements. Non-parametric techniques are most often used to analyse data which do not meet the distributional requirements of parametric techniques. Parametric Techniques are mostly based on linear models and normal theory and hence to the assumptions of normal and independent distributed residuals. Most nonparametric Techniques for trend detection in long time series are based on the Mann-Kendall test for trend (Mann, 1945; Kendall, 1975). In the light of relative efficiency as measured by the power of a test at a given significance level, it is known that parametric techniques are the most powerful if the residuals are normally distributed (Hirsch et al., 1982; Hirsch and Slack, 1984; Lettenmaier, 1988; Zetterqvist, 1988; Hirsch et al., 1991; Shoab et al., 2013; Tiwari and Jain 2017). However, the marginal distributions of time series values are frequently skewed. Nonparametric tests can have higher powers than parametric tests in case of normality, provided the sample size is large enough (Loftis et al., 1991; Berryman et al., 1988). Furthermore, the loss in power for normally distributed data is rather small (Hirsch et al., 1982; Berryman et al., 1988). Since in most cases it is not known a priori whether the data is normally distributed or not, it has been recommended to use nonparametric Techniques as a general approach for the detection of trends in long time series data (McLeod et al., 1983; Lettenmaier, 1976; Hirsch et al., 1982, Berryman et al., 1988; Hirsch et al., 1991).

Other arguments to the advantage of nonparametric over parametric techniques are the relative robustness against missing values and irregularly spaced observations, which is actually a special case of missing values (van Belle and Hughes, 1984; Zetterqvist, 1991; Harcum et al., 1992). Furthermore, outliers do not present a problem for nonparametric tests. Even truncated observations, which can be due to a detection limit of the measurement method, can be handled appropriately using nonparametric tests (Berryman et al., 1988). Since time series vales frequently show these limiting properties, nonparametric techniques seem in general more appropriate than their parametric counterparts.

II. STATISTICAL TREND ANALYSIS: MANN-KENDALL TEST

The statistical importance of trend in long time series datasets are evaluated using Mann Kendall trend analysis technique. For a series of non-parametric observations over time (like rainfall, temperature etc.), it is important to understand whether the time series dataset is going upward, downward, or staying in the same direction. The non-parametric Mann-Kendall test (Mann 1945; Kendall 1955) is commonly employed by various researchers (Sneyers et al 1990; Zhang et al, 2000; Yue et al, 2003; Aziz & Burn 2006; Cannarozzo et al. 2006; Wilks 2006; Chandler and Scott 2011) to detect monotonic trends in long time series datasets belong to different domain of science, medicine and meteorology.

The Mann-Kendall is applicable in cases when the data values xi of a time series can be assumed to obey the model

$$x_i = f(t) + \varepsilon_i \qquad \text{-- Equation 3}$$

Where $f(t)$ is a continuous monotonic increasing or decreasing function of time and the residual ε_i can be assumed to be from the same distribution with zero mean. It is therefore assumed that the variance of distribution is constant in time. Mann-Kendall test statistic S is calculated using given equation

$$s = \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} \mathrm{sgn}(x_j - x_i) \qquad \text{-- Equation 4}$$

Where n is the number of data points x_j and x_i are the data value in time series i and j (j >i), respectively and $\mathrm{sgn}(x_j - x_i)$ is the sign function as:

$$\mathrm{sgn}(x_j - x_i) = \begin{cases} +1, & \text{If } x_j - x_i > 0 \\ 0, & \text{if } x_j - x_i = 0 \\ -1, & \end{cases} \qquad \text{-- Equation 5}$$

The variance is computed when n ≥ 10:

$$Var(s) = \frac{n(n-1)(2n+5) - \sum_{i=1}^{m} t_i(t_i - 1)(2t_i + 5)}{18} \qquad \text{-- Equation 6}$$

Where total number of data points in the time series dataset are represented by n. m is representation for cluster of data points having same data value and t_i depicts the total equal observation of extent i at a particular timestamp. A tied group is a set of sample data having the same value. When total number of sample size is more than 10, following equation of 7 is used to compute Z_s value:

$$z_s = \begin{cases} \dfrac{s-1}{\sqrt{Var(s)}} & \text{If } S > 0 \\ 0 & \text{If } S = 0 \\ \dfrac{s+1}{\sqrt{Var(s)}} & \text{If } S < 0 \end{cases}$$

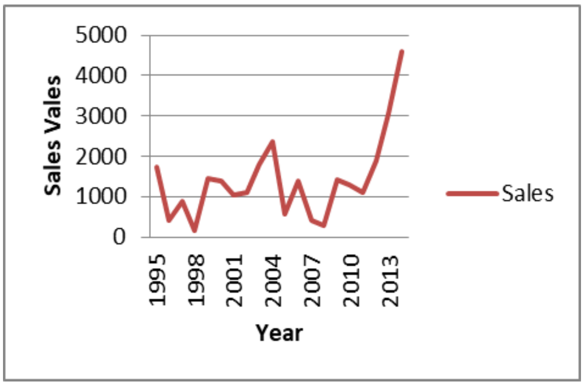
A positive value of Z_s represent trend in upward direction while negative value of Z_s shows downward trend direction. A significance value α is also computed for checking whether an upward or downward trend is found in the dataset.

III. CASE STUDY

Mann-Kendall test is a flexible tool for significant trends in time series. In the current paper we have selected a case study to find out; whether there is any trend exists in the sales record of a company X or not. Starting from 1995 to 2014 annual sales record for each year of the company x is presented in the table 1. The trend in the series should be monotonous, i.e. mainly going up or down during the observation period. The test is calculated by first ranking the observations. If two or more observations have the same value, they are assigned the same rank. In the next step, (for each observation), how many observations in the time before this observation have a lower value (a lower rank) and how many have a higher value (a higher rank) and sum these numbers over all observations to get two numbers. The difference between these two numbers constitutes the test statistics for the Mann-Kendall test. A high positive value indicates an upward trend, while a high negative value indicates a downward trend. Statistical conclusion is based on one standardized version of this test statistic, which is approximately normalized if we have more than 10 observations (where the observations must be so far apart in order to exclude that they are auto-correlated).

Table1. Annual sales record of company x from 1995 to 2014

Year	**1995**	**1996**	**1997**	**1998**	**1999**	**2000**	**2001**	**2002**	**2003**	**2004**
Annual Sale	1733	400	875	164	1464	1399	1039	1096	1812	2347
Year	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014
Annual Sale	575	1393	411	283	1407	1294	1102	1886	3058	4587



Graph1. Sales values from 1995 to 2014 for company x

G Mann Kendall analysis when performed for the current dataset, it is found that for S (variance) = 52, Z = 1.6547. Table 2 depicts trend values at 10%, 5% and 1% significance level. This can be seen that at 10% significance level Z value is under the acceptable limit and sales record are showing trend while at 1 % and 5 % significance level data values does not present any trend. Thus with 90% confidence level one can state that the current sales record possess a significant trend. In order to understand the type of trend (positive or negative) slope estimates techniques are required to be combined with trend analysis.raph1. Sales values from 1995 to 2014 for company x

Table2. Z values at 10%, 05% and 01% significance level

=	1.6547	>	1.645	at	10 % significance level
=	1.6547	<	1.960	at	05 % significance level
=	1.6547	<	2.576	at	01 % significance level

IV. DISCUSSION

A number of statistical tests are available to identify and quantify monotonic trends in a way that is defensible and repeatable. Among all the existing trend analysis techniques, Mann Kendall test is a statistical test widely used for the analysis of trend in climatologic and in hydrologic time series. The Mann-Kendall test provides the following advantages:

- Mann-Kendall does not require data to be arranged as per the bell curve or follow normal distribution.
- This trend analysis method is not majorly disturbed by missing values when numbers of data points are significantly high.
- Mann-Kendall does not get affected by any irregularity in occurrence of the time points of measurement.
- Mann-Kendall does not dependent upon by the length of the time series data.

Besides the discussed advantages this method has some limitations. The major limitations are as follows:

- This test is not suitable for the time series dataset where seasonal measurements are used to collect time dependent data values.
- This test results more negative outcomes when length of time series dataset is short. Thus longer the dataset higher accuracy of trend analysis will results into final outcomes.

In Mann-Kendall based trend analysis it is always recommended to maintain the higher length of the dataset and seasonal dependency must not be removed before perfuming this test.

V. CONCLUSION

The study concludes that Mann-Kendell statistical tests in TREND are relatively easy to understand and user can gain a good implementation of the tests by following the descriptions and the examples presented in the study. Trend analysis based up on Mann-Kendall statistics helps to understand how underlying function has performed and predict where current state of the selected function will guide the dataset. It will give a clear insight about where and how much change is required to move the selected function in the right direction. One can use trend analysis to help improve the selected domain of the study by:

- Identifying section where observed dataset is performing well the criteria governing the success of dataset can duplicate success
- Identifying section where observed dataset is underperforming
- Providing evidence to inform your decision making.

REFERENCES

[1]. Aziz, O.I.A. and Burn, D.H., 2006. Trends and variability in the hydrological regime of the Mackenzie River Basin. Journal of hydrology, 319(1-4), pp.282-294.

[2]. Belle, G. and Hughes, J.P., 1984. Nonparametric tests for trend in water quality. Water resources research, 20(1), pp.127-136.

[3]. Berryman, D., Bobée, B., Cluis, D. and Haemmerli, J., 1988. Nonparametric tests for trend detection in water quality time series 1. JAWRA Journal of the American Water Resources Association, 24(3), pp.545-556.

[4]. Cannarozzo, M., Noto, L.V. and Viola, F., 2006. Spatial distribution of rainfall trends in Sicily (1921–2000). Physics and Chemistry of the Earth, Parts A/B/C, 31(18), pp.1201-1211.

[5]. Chandler, R. and Scott, M., 2011. Statistical Techniques for trend detection and analysis in the environmental sciences. John Wiley & Sons.

[6]. Harcum, J.B., Loftis, J.C. and Ward, R.C., 1992. Selecting trend tests for water quality series with serial correlation and missing values 1. JAWRA Journal of the American Water Resources Association, 28(3), pp.469-478.

[7]. Hirsch, R.M., Slack, J.R. and Smith, R.A., 1982. Techniques of trend analysis for monthly water quality data. Water resources research, 18(1), pp.107-121.

[8]. Hirsch, R.M. and Slack, J.R., 1984. A nonparametric trend test for seasonal data with serial dependence. Water Resources Research, 20(6), pp.727-732.

[9]. Hirsch, R.M., Alexander, R.B. and Smith, R.A., 1991. Selection of Techniques for the detection and estimation of trends in water quality. Water resources research, 27(5), pp.803-813.

[10]. Kendall, M.G., 1975. Rank Correlation Techniques, Charles Griffen. London ISBN, 195205723.

[11]. Kendall, M.G., 1955. Further contributions to the theory of paired comparisons. Biometrics, 11(1), pp.43-62.

[12]. Lettenmaier, D.P., 1988. Multivariate nonparametric tests for trend in water quality 1. JAWRA Journal of the American Water Resources Association, 24(3), pp.505-512.
[13]. Loftis, J.C., 1996. Trends in groundwater quality. Hydrological processes, 10(2), pp.335-355.
[14]. Mann, H.B., 1945. Nonparametric tests against trend. Econometrica: Journal of the Econometric Society, pp.245-259.
[15]. McLeod, A.I., Hipel, K.W. and Comancho, F., 1983. Trend assessment of water quality time series 1. JAWRA Journal of the American Water Resources Association, 19(4), pp.537-547.
[16]. Shoab, M., Jain, K. and Shashi, M., 2013. GNSS Based Real Time Train Monitoring: A Web Approach. International Journal of Computer Applications, 73(14).
[17]. Sneyers, R., Vandiepenbeeck, M., Vanilierde, R. and Demarée, G.R., 1990. Climatic changes in Belgium as appearing from the homogenized series of observations made in Brussels–Uccle (1933-1988) In: SCHIETECAT, GD. Contributions à l'etude des changements de climat. Bruxelles: Institute Royal Meteorologique de Belgique, Publications Série, 124, pp.17-20.
[18]. Tiwari, A. and Jain, K., 2017. Concepts and Applications of Web GIS. Nova Publications, USA.
[19]. Tiwari, A., Suresh, M. and Rai, A.K., 2014. Ecological Planning for Sustainable Development with a Green Technology: GIS. International Journal of Advanced Research in Computer Engineering & Technology (IJARCET), 3(3), pp.2278-1323.
[20]. Wilks, D.S., 2006. On "field significance" and the false discovery rate. Journal of applied meteorology and climatology, 45(9), pp.1181-1189.
[21]. Yue, S. and Hashino, M., 2003. Long term trends of annual and monthly precipitation in Japan 1. JAWRA Journal of the American Water Resources Association, 39(3), pp.587-596.
[22]. Zetterqvist, L., 1991. Statistical estimation and interpretation of trends in water quality time series. Water Resources Research, 27(7), pp.1637-1648.
[23]. Zetterqvist, L., 1988. Asymptotic distribution of Mann's test for trend for m-dependent seasonal observations. Scandinavian journal of statistics, pp.81-95.
[24]. Zhang, X., Vincent, L.A., Hogg, W.D. and Niitsoo, A., 2000. Temperature and precipitation trends in Canada during the 20th century. Atmosphere-ocean, 38(3), pp.395-429.