Mann kendall trend test

Often, we need to evaluate the data to see if the data is following a trend and if the trend is increasing or decreasing. Some regulators now require that data be evaluated for trends. The Mann-Kendall test is the test often used and/or required. The following is information about the method and includes spreadsheets that can be used to perform calculations. The following general information is adapted from the Guide on Natural Mitigation for Oil Emissions by the Wisconsin Natural Resources Department available online at www.dnr.state.wi.us/Org/aw/rr/archives/pubs/RR614.pdf (also attached below). The Mann-Kendall (M-K) test is a simple test for trend. Mann-Kendall is a non-parametric test and as such does not depend on: • The extent of the data, • Distribution assumptions (must not have a normal distribution/bell form), • Missing data or • Irregular monitoring periods. Mann-Kendall assesses whether a time-ordered data set shows a tendency to increase or decrease in a predetermined level of significance. Mann-Kendall Limitations Non-seasonal Effects – The M-K test requires 4 to 10 rounds of data (note there are changes that allow larger data sets) that are NOT influenced by seasonal effects. This means that the investigator must either determine that the data are not seasonally influenced or collect data from the same season of the year for at least 4 years. Another option is the use of the Seasonal Kendall test (Gibbons, 1994), which includes seasonal variability in the analysis. When data appear to be affected seasonally, one way to reduce seasonal biases in the M-K test is to include only data collected from that season with the highest concentrations of contaminants. This would require a longer monitoring period to collect sufficient data. Note: One method of assessing whether data follows a seasonal trend is to trace the concentration or concentration of the log relative to the altitude of groundwater to determine whether there is a trend. More is better - Statistical confidence increases with the number of data points available. The more data there is, the more likely the M-K test is to discern a trend. While the test can use as little as 4 data points, often they are not enough data to detect a trend. A trendless finding with few data points does not always indicate a stable data set. This situation more likely indicates that there is too little data to determine a trend. Therefore, it is highly recommended that at least 6 or more rounds of data be collected before using the M-K test to assess data trends. Censored Data / Non-Detects - you have censored data (Non-Detects) and these are different values, you have to address this otherwise it will be measuring the trend of reporting limits. Normally, replacing censored data with 1/2 of the lowest reporting limit the data set works best. Find out before use!!! I personally think it's always best to make a 1st hand method so you know it works and how data changes affect it. The Mann-Kendall method is very easy to do by hand and I would recommend that you see the examples in the EPA - Statistical analysis of groundwater monitoring data at RCRA installations. THE IDEM's Mann-Kendall Spreadsheet Indiana Department of Environmental Management (DITEM) has a spreadsheet that evaluates data for trends using Mann-Kendall. It allows you to determine if there is a trend and if there is a trend, whether it is increasing or decreasing. The web page is (also attached below). It is an excellent tool because it will make 5 fountains at a time and evaluate data points, as well as generating data graphs: • Quarters 1-8, 5-13, 7-16, • Quarters 1-12, 1-16 • Year 1-2, Year 2-3, Year 3-4, Year 1-3, and Year 1-4 Please find attached an original version and a version that we have modified and an example output. The Wisconsin Department of Natural Resources has a spreadsheet that evaluates between 4 and 10 data points. A copy of the spreadsheet can be downloaded from: (also attached below) The spreadsheet information is covered in the www.dnr.state.wi.us/Org/aw/rr/archives/pubs/RR614.pdf (also attached below). The Mann-Kendall Test Python module to calculate the Mann-Kendall test for the trend of data in the time series. This module contains a single test function that implements the Mann-Kendall test for a linear trend in a given time series. The Mann-Kendall test is used to determine whether or not there is a monotonous linear trend in a given data series. It is a non-parametric trend closely related to the concept of the kendall correlation coefficient [1]. The null hypothesis, \(H_0\), states that there is no monotonous tendency, and this is tested against one of the three possible alternative hypotheses, \(H_a\): (i) there is an upward monotonous tendency, (ii) there is a descending monotonous tendency, or (iii) there is either an ascending monotonous tendency or a descending monotonous tendency. It is a robust test for trend detection widely used in the analysis of financial, climatological, hydrological and environmental time series. The Mann-Kendall test involves the following assumptions [2] regarding data from the given time series: 1. In the absence of a trend, the data is distributed independently and identically (iid). 2. Measurements represent the actual states of observables at the time of measurements. 3. Methods used for the collection of samples, measurements and data manipulation are impartial. The Mann-Kendall test offers the following advantages: 1. Do not assume that the data to be distributed according to a particular rule, for example, does not require that the data be Distributed. 2. This shall not be achieved by lack of data other than the fact that the number of sampling points is reduced and could therefore have a negative effect on statistical significance. 3. It shall not be carried out by irregular spacing of the measuring points. This is not done according to the duration of the time series. The following limitations should be considered: 1. The Mann-Kendall test is not suitable for periodicity data (e.g. seasonal effects). For the test to be effective, it is recommended to remove all known periodic effects from the data at a preprocessing stage before the Mann-Kendall test is calculated. 2. The Mann-Kendall test tends to give more negative results for shorter datasets, i.e., the more efficient the time series, the more effective the trend detection calculation is. The first step in the Mann-Kendall test for a series of time \(x_1, x_2, \dots, x_n\) of length \(n\) is to calculate the indicator function \(sgn(x_i - x_j)\) so that: \[\begin{split}sgn(x_i - x_j) &= \begin{cases} 1, & x_i - x_j > 0\\ 0, & x_i x_j - x_j = 0\\ -1, & x_i - ,\end{split}\ that tells us whether the difference between timely \(i\) and \(j\) metrics is positive, negative, or zero. Then we calculate the average and the variance of the above quantity. The media \(E[S]\) is given by: \[E[S] = \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} sgn(x_i - x_j)\] and the variance \(VAR(S)\) is given by : \[VAR(S) = \frac{1}{18} \Big(n-1)(2n+5) - \sum_{k=1}^p q_k(q_k-1)(2q_k+5) \Big),\] where \(p\) is the total number of tie groups in the data , and \(q_k\) is the number of data points contained in the tie group \(k\)-th. For example, if the time series measurements were {12, 56, 23, 12, 67, 45, 56, 56, 10}, we would have two tie groups for measurements 12 and 56, i.e. \(p=2\), and the number of data points in these tie groups would \(q_1=2\) for the tie group with {12}, and \(q_2=3\) for the tie group with {56}. Using the average \(E[S]\) and the variance \(VAR(S)\) we calculate the Mann-Kendall test statistics, using the following transformation, which ensures that for large sample sizes, test statistics \(Z_{MK}\) are distributed approximately normally: \[\begin{split}Z_{MK} &= \begin{houses} \frac{E[S] - 1} {\sqrt{VAR(S)}}, & E[S] > 0\\ 0, & E[S] = 0\\ \frac{E[S] + 1} {\sqrt{VAR(S)}} , & E[S] < 0\\ \end{houses}.\end{split}]] At a significance level \(\alpha\) of the test, which is also the type I error rate, we calculate whether or not to accept the alternative hypothesis \(H_a\) for each variant of \(H_a\) separately: \(H_a\): There is an upward monotonous trend If \(Z_{MK} \geq Z_{1 - \alpha}\) then accept \(H_a\) , where the notation \(Z_{1 - \alpha}\) denotes the percentile \(100(1-\alpha)\)-a of the distribution Standard. \(H_a\): There is a downward monotonous trend If \(Z_{MK} \leq -Z_{1 - \alpha}\) then you accept \(H_a\). \(H_a\): There are a monotonous upward or downward trend If \(| Z_{MK}}| \geq Z_{1 - \alpha/2}\) and then accept \(H_a\), where the notation \(|\cdot|\) is used to designate the absolute value function. In addition to the Mann-Kendall test result, which is in the form of a string indicating whether or not to accept the alternative hypothesis, the test() function also returns some additional estimates related to the estimation of a monotonous trend in the time series. The slope \(m\) and the interception of a straight line mounted by time series data are estimated as follows: \[m = r_{x,t} \frac{\sigma_x}{\sigma_t},\ where r_{x,t} is The Cross Correlation Coefficient of Pearson between \(x\) and \(t\). where \(\mu\) denotes the average of both variables. The test() function also returns the \(p\)-values for the given data set under the various alternative assumptions. Note that estimating the \(p\) value is not essential for calculating the test results, as formulated above. The values \(p\)-must be estimated separately depending on the type of alternative hypothesis used and the sign \(E[S]\). Denotes \(f(u)\) as a probability density function of the standard normal distribution, we can note the values \(p\)-as: \(H_a\): There is an ascending monotonous trend \[\begin{split}p_{Z_{MK}} &= \begin{houses} \int_{Z_{MK}}\infty} f(u) \mathrm{d E[S]|>\varepsilon\\ 0.5, & | E[S]| \leq \varepsilon\\ \end{houses}.\end{split}\] \(H_a\): There is a monotonous downward trend \[\begin{split}p_{Z_{MK}} &= \begin{houses} \int^{Z_{MK}}{{-\infty} f(u) \mathrm{d}u,&; | E[S]|>\varepsilon\\ 0.5, & | E[S]| \leq \varepsilon\\ \end{houses}.\end{split}\] \(H_a\): There is either an upward monotonous trend, either a monotonous downward trend \[\begin{split}p_{Z_{MK}} &= 0.5 \begin{houses} \int_{Z_{MK}}{\infty} f(u) \mathrm{d}u,and E[S]>\varepsilon\ 1, & | E[S]| \leq \varepsilon\\ \int^{Z_{MK}}{-\infty} f(u) \mathrm{d}u,& E[S]<-\varepsilon\\ \end{houses}.\end{split}[] Mkt.test(t, x, eps=None, alpha=None, Ha=None)[source]¶ Runs the Mann-Kendall trend for the trend in time series data. Parameters: t (1D numby.ndararay) – matrix of the time points of measurements x (1D numby.ndararay) – matrix containing measurements corresponding to t eps inputs (scaler, float, greater than zero) – the smallest measurement error that helps to determine links in alpha data (scaler, float, higher than zero) – the significance level of the statistical test (type I error) Ha (string options, string options include top, bottom, upordown) – test type: one-sided (up or down) Returns: MK (string) – the result of the statistical test indicating whether or not to accept the alternative Ha m hypothesis (scaler, float) – slope of linear fit to the data c float) – intercept of linear match to p data (scaler, float, float, than zero) – the p value of the Z-score statistic obtained for the Mann-Kendall Test Lift: AssertionError : error – the lowest number of eps measurement error is not given AssertionError : error – significance level of alpha test is not given AssertionError : error – alternative ha hypothesis is not given