hydrology

MDPI

Article

Supplementary material to:

Onyutha, C. (2017). On rigorous drought assessment using daily time scale: Frequency analyses considering non-stationarity, and a new method to yield non-parametric indices.

1. Stationarity *versus* non-stationarity

1.1. Prerequisites for frequency analyses

a) Data transformation

In flood frequency analyses, no transformation of flow or rainfall is needed. In other words, both river flow or rainfall of daily time scale can be used in their original forms after quality control.

For low flow analyses, river flow series *H* can be transformed by (1/*H*). Precipitation insufficiency (precipitation minus potential evapotranspiration) can be transformed using (-*H*).

b) Independence of the extreme events

For flood frequency analyses, the extraction of independent extreme events is conducted with original daily series (i.e. data which is not transformed). However, for low flow analyses, extraction of the extreme events is done using transformed data i.e. (1/*H*) flow or (-*H*) precipitation insufficiency.

Either the annual maxima method (AMM) or the peak over threshold (POT) approach can be used. For the AMM, the time slice is taken as one hydrological year. For the POT, independence criteria should be used.

For the POT approach, it is vital to note that both the POT events, and their corresponding times of observation should be extracted. The times of observations for the extreme events can be used for computing trend magnitude. For the ease of analysis, the time of observation (e.g. date) can be converted to number format using e.g. the DATEVALUE() function of Ms Excel. Alternatively, consider the 1st January 1900, 2nd January 1900, ... to have serial numbers 1, 2, ..., etc. Thus, if the first date of the original series is, say, 1/1/1965, it becomes 23743. Furthermore, if the extracted POT events occur on 4/1/1965, 12/27/1965, 3/13/1966, ..., 5/30/2002, their corresponding times of observations (in a number format) become 23833, 24103, 24179, ..., 37406, respectively. Using this same example, by unitizing the date of the first value from the original series, the Required Times of Observation (RTO) for the corresponding extreme events can be computed by e.g. (23833-23743)+1=91, (24103-23743)+1=361, (24179-23743)+1=437, (37406-23743)+1=13664, respectively. For the annual maxima series, there is no need to convert the date to number because the year can be used as the time of observation.

c) Significance of trend in the data

- Trend magnitude

The significance of the trend slope *m* can be tested in both parametric and non-parametric approach. In this study, the least squares approach was adopted

To assess the significance of the linear trend slope *m* (see Eq. (1) of the main paper), the following steps as found in [1] can be taken. After confirming that the data has minimal or no influence from the possible outliers, the H_0 (trend magnitude is not time-dependent) i.e. H_0: m = 0 and the alternative hypothesis H_1: $m \neq 0$ can be tested at α_s% using the following procedure. Compute the Pearson product-moment correlation coefficient d between the extreme events (i.e. POTs in this case) and RTOs. If n denotes the sample size of the POTs, S_y is the standard deviation of the POTs, and S_x is the standard deviation of RTOs such that $S_{yx}=S_y\times[(1-d^2)\times(n-1)/(n-2)]^{0.5}$, the standard error of estimate S_e becomes $S_e=S_{yx}/[S_x\times(n-1)^{0.5}]$. Eventually, the statistic (S_t) can be given by $S_t=m/S_e$. Calculate the probability value (p-value) based on Student's t–distribution. Here, in Ms Excel, the function TDIST(value, df, tails) can be used where df refers to the degree of freedom. The two-tailed-based p-value can be given by p = TDIST($|S_t|$, n–2, 2). Compare the computed p-value with the selected α_s%. If the α_s in decimal (i.e. α_s%/100) is less than the computed p-value, the H_0: m = 0 can be accepted at α_s%. If the estimated p-value ≤ (α_s%/100), H_1: $m \neq 0$ cannot be rejected.

- Trend direction

Several nonparametric methods exist for detection of trend direction. The well-known methods include the Mann-Kendall (MK) [2,3], and the Spearman's Rho (SMR) [4-6] test. Recent methods advanced towards trend detection exist e.g. the Cumulative Sum of rank Difference (CSD) [1, 7-8] tests. The detailed description of the CSD trend test is given in the main paper. However, the MK test is described in this supplementary material.

- The Mann-Kendall (MK) [2,3] test

For the MK test, the trend statistic S is defined as:

$$S_{MK}=\sum_{i=1}^{n-1}\sum_{j=i+1}^{n}\operatorname{sgn}\left(x_j-x_i\right) \tag{S1}$$

where x_j and x_i are the sequential data values in a sample of size n, and

$$\operatorname{sgn}\left(x_j-x_i\right)=\begin{cases}1 & \text{if } \left(x_j-x_i\right)>0\\ 0 & \text{if } \left(x_j-x_i\right)=0\\ -1 & \text{if } \left(x_j-x_i\right)<0\end{cases} \tag{S2}$$

For $n \geq 8$, the distribution of S_{MK} is approximately normal with the mean E(S) = 0 while the variance $V(S)$ is given by [2,3]:

$$V\left(S_{MK}\right)=\frac{1}{18}n\left(n-1\right)\left(2n+5\right) \tag{S3}$$

When tied data points exist, $V(S_{MK})$ becomes:

$$V\left(S_{MK}\right)=\frac{1}{18}\left(n\left(n-1\right)\left(2n+5\right)-\sum_{j=1}^{u}q_j\left(j-1\right)\left(2j+5\right)\right) \tag{S4}$$

where u is the number of tied groups, and q_j denotes the number of observations in the j^{th} group. The standardized statistic of the MK test Z_{MK} which follows the standard normal distribution with the mean = 0 and variance =1 is given by:

$$Z_{MK} = \begin{cases} \dfrac{S_{MK} - 1}{\sqrt{V\left(S_{MK}\right)}} & \text{for } S_{MK} > 0 \\ 0 & \text{for } S_{MK} = 0 \\ \dfrac{S_{MK} + 1}{\sqrt{V\left(S_{MK}\right)}} & \text{for } S_{MK} < 0 \end{cases} \tag{S5}$$

Increasing and decreasing trends are indicated by positive and negative values of S_{MK} respectively. The procedure suggested by [9] can be used for auto-correlated time series. If $|Z_{MK}|$ is greater than or equal to $Z_{\alpha s/2}$, the H_0 (no trend) is rejected.

d) Decision to adopt stationarity or non-stationarity

Here, decision on whether to adopt stationarity or non-stationarity for analyzing frequency of the extreme events can be made based on the significance of both trend magnitude and direction (see information from Table 2 of the main paper). As a side note, for flood frequency analysis (where data is not transformed), there is more concern for increasing than decreasing trend in the peak high events.

References

1. Onyutha, C. Statistical uncertainty in hydro-meteorological trend analyses. *Adv. Meteorol*. **2016**, *2016(Article ID 8701617)*, 1-26, DOI:10.1155/2016/8701617
2. Mann, H.B. Nonparametric tests against trend. *Econometrica* **1945**, *13(3)*:245–259.
3. Kendall, M.G. *Rank Correlation Methods*. Charles Griffin, London, UK, 4th edition, 1975.
4. Spearman, C. The proof and measurement of association between two things. *Am. J. Psychol*., **1904**, 15: 72–101.
5. Lehmann, E.L. *Nonparametrics, statistical methods based on ranks*. Holden-Day Inc., Carlifornia, 1975.
6. Sneyers, R. *On the statistical analysis of series of observations*. Technical Note no. 143, WMO no. 415. Geneva, Switzerland: Secretariat of the World Meteorological Organization, 1990.
7. Onyutha, C. Identification of sub-trends from hydro-meteorological series. *Stoch. Environ. Res. Risk Assess*. **2016**, *30*,189–205.
8. Onyutha, C. Statistical analyses of potential evapotranspiration changes over the period 1930-2012 in the Nile River riparian countries. *Agr. Forest. Meteorol*. **2016**, *226-227*, 80–95.
9. Yue, S.; Wang,C. The Mann-Kendall test modified by effective sample size to detect trend in serially correlated hydrological series. *Water Resour. Manage*. **2004**, 18 (3), 201-218.

Conflict of Interest

The author declares no conflict of interest and no competing financial interests.