sustainability MDPI

Introduction

The supporting information contains supplemental material for the methodology of the Standardized Precipitation Index, Homogeneity Tests, and Relative Operating Characteristic (ROC) .

1. Supplemental Material for the Standardized Precipitation Index (SPI)

The SPI was chosen for this study because of its simplicity and being based solely on the accessible precipitation data. The Standardized Precipitation Index (*SPI*) is based on an equi-probability transformation of aggregated monthly precipitation into a standard normal variable [1] and recommended by the World Meteorological Organization as a standard to characterize meteorological droughts [2]. McKee assumed an aggregated precipitation gamma distributed and used a maximum likelihood method to estimate the parameters of the distribution. In the most cases, the Gamma distribution is the distribution that best models observed precipitation data. The density probability function for the Gamma distribution is given by the expression [3]:

$$g(x) = \frac{1}{\beta^{\alpha}\Gamma(\alpha)} x^{\alpha-1} \quad for\ x > 0 \tag{1}$$

Where $\alpha \succ 0$ is a shape parameter, $\beta \succ 0$ is a scale parameter and $x \succ 0$ is the precipitation amount $\Gamma(\alpha)$ is the Gamma function and defined by:

$$\Gamma(\alpha) = \int_0^{\infty} x^{\alpha-1} e^{-x} dx \tag{2}$$

Computation of the SPI involves the fitting of a gamma probability density function to a given frequency distribution of precipitation totals for a station. The alpha and beta parameters of the gamma probability density function are estimated for each station, for each time scale of interest (1month, 3 months, 12 months, 48 months, etc.) and for each month of the year. After estimating coefficient alpha and beta the density of probability function *g(x)* is integrated with respect to *x* and we obtain an expression for cumulative probability *G(x)* that a certain amount of rain has been observed for a given month and for a specific time scale.

$$G(x) = \int_0^x g(x) \rightleftarrows dx = \frac{1}{\beta^{\alpha}\Gamma(\alpha)} \int_0^x x^{\alpha-1} e^{\frac{-x}{\beta}} dx \tag{3}$$

The Gamma function is not defined by *x*= 0 and since there may be no precipitation the cumulative probability becomes:

$$H(x) = q + (1-q)G(x) \tag{4}$$

Where *q* is the probability of no precipitation. The cumulative probability is then transformed into a normal standardized distribution with null average and unit variance from which we obtain the *SPI* index [4]. The SPI Index is then defined as:

$$SPI = -\left(t - \frac{c_0 + c_1 t + c_2 t^2}{1 + d_1 t + d_2 t^2 + d_3 t^3}\right) \ for\ 0 < H(x) < 0.5 \tag{5}$$

$$SPI = +\left(t - \frac{c_0 + c_1 t + c_2 t^2}{1 + d_1 t + d_2 t^2 + d_3 t^3}\right) \ for\ 0.5 < H(x) < 1 \tag{6}$$

where

$$t = \sqrt{ln\left[\frac{1}{(H(x))^2}\right]} \ for\ 0 < H(x) < 0.5 \tag{7}$$

$$t = \rightleftarrows \sqrt{\ln\left[\frac{1}{(1-H(x))^2}\right]} \ for\ 0.5 < H(x) < 1 \tag{8}$$

where c_0, c_1, c_2, d_1, d_2 and d_3 are constants with the following values:

$c_0 = 2.515517$ $c_1 = 0.802853$ $c_2 = 0.010328$

$d_1 = 1.432788$ $d_2 = 0.189269$ $d_3 = 0.001308$

2. Homogeneity Tests

2.1. Buishand Range Test

It is applied for testing homogeneity of the datasets and it depends on the rescaled adjusted partial sums for the time series X_i as follow [5]:

$$S_k^* = \sum_i^k (X_i - \overline{X})^2 \qquad k = 1, 2, \ldots\ldots\ldots, N \; S_0^* = 0 \tag{9}$$

In case of homogeneous records the values of S_k^* fluctuates around zero, because no systematic deviations of the X_i values from their average will detected. If a break is identified in year , k, then S_k^* reaches a maximum negative- shift or minimum positive-shift near the year , $k = K$ .I t is possible to determine the significance of this shift the rescaled adjusted range R :

$$R = (\max_{0 \ll k \ll n} S_k^* - \max_{0 \ll k \ll n} S_k^*)/S_d \tag{10}$$

Here, S_d is the standard deviation. Critical limits for the test-statistic are found in [5].

2.2. Von Neumann Ratio Test

The Von Neumann ratio is described as:

$$N_V = \sum_{i=1}^{n-1} (X_i - X_{i+1})^2 \Big/ \sum_{i=1}^{n} (X_i - \overline{X})^2 \tag{11}$$

In which $\bar{X}$ represents the average of the $Xi's$. If the data includes changes or breaks, then N_V will be less than the critical value. If the sample changes rapidly with respect to its average, then values of N_v may increase above 2 [6]. This test, however, does not provide any details about the years of breaks.

2.3. Pettitt Test

It is based on the Wilcoxon test [7] and it can also be derived using the Mann–Whitney U-test. The ranks $r_1, \ldots r_n$ of the $X_1, \ldots X_n$ are used to calculate the statistics:

$$X_k = 2\sum_{i=1}^{k} r_i - k(n + 1) \quad k=1, \ldots\ldots, n \tag{12}$$

If a break appeared in any year G, the absolute value of X_k reaches to its maximum.

$$X_G = \max_{0 \ll k \ll n} |X_k| \tag{13}$$

Critical values were given by Pettitt [7].

2.4. Standard Normal Homogeneity Test

Alexandersson [8] illustrates a statistic $T(k)$ in order to compare the mean of the first k years of the records with that of the last $n - k$ years:

$$T(k) = k\,z_1^2 + (n - k)z_2^2 \tag{14}$$

where:

$$z_1^2 = \frac{1}{k}\sum_{i=1}^{k}(X_i - \bar{X}) / s \;\; and \;\; z_2^2 = \frac{1}{n-k}\sum_{i=k+1}^{n}(X_i - \bar{X}) / s \tag{15}$$

If a break is detected at any year k, then $T(k)$ reaches a maximum near the year $k = K$. The $T(k)$ is depicted in the graphs representing the results of this test. The test statistic$T_0$is calculated as:

$$T_0 = \max_{1 \ll k \ll n} T(k) \tag{16}$$

The null hypothesis will be rejected if T_0 is greater than the significant level which depends on the sample length.

3. Relative Operating Characteristic (ROC)

ROC is a graph that could be constructed for any event such as drought forecasting, was used to provide information on the hit rates and false alarm rates that can be expected from use of different probability thresholds to trigger advisory action. *ROC* is conditioned on the observations, answering the question given that an event *Y* occurred, what was the corresponding forecast? It therefore measures the ability of the forecasting system to discriminate between events and non-events, i.e. the resolution of the forecast. *ROC* is calculated by means of a 2x2 contingency table for each probability as in Table 2, which counts the number of forecast hits (*HT*), the number of misses (*MS*), the number of false alarms (*FA*) and the number of correct rejections (*CJ*). *ROC* is then probability of detection (*PoD*) as a function of the false alarm rate (*FAR*), where

$$PoD = \frac{HT}{HT + MS} \tag{17}$$

$$FAR = \frac{FA}{FA + HT} \tag{18}$$

A *ROC* curve is plotted as a curve joining the *PoD* as a function of the *FAR* for all forecast probabilities. The area under the *ROC* curve gives a measure of the skill of the forecast.

Table S1. 2*2 contingency table for relative operating characteristic (ROC) calculation.

		Forecasted Event	
		Yes	**No**
Observed Event	Yes	Hit (HT)	Miss (MS)
	No	False Alarm (FA)	Correct Rejection (CJ)

References

1. McKee, T.B.; Doesken, N.J.; Kleist, J. The relationship of Drought Frequency and Duration to Time Scales. In Proceedings of the 8th Conference on Applied Climatology, Anaheim, CA, USA, 17–22 January 1993.
2. Dutra,E.; Di Giuseppe, F.; Wetterhall, F. Pappenberger, F. Seasonal forecasts of droughts in African basins using the Standardized Precipitation Index. *Hydrol. Earth Syst. Sci.* **2013**, *17*, 2359–2373. https://doi.org/10.5194/hess-17-2359-2013
3. Cacciamani, C.; Morgillo, A.; Marchesi, S.; Pavan, V. *Monitoring and Forecasting Drought on a Regional Scale: Emilia-Romagna Region*; Springer: Dordrecht, Netherlands, 2007; Volume 62.
4. Khadr, M.; Morgenschweis, G.; Schlenkhof, A. Analysis of Meteorological Drought in the Ruhr Basin by Using the Standardized Precipitation Index. In Proceedings of the International Conference on Sustainable Water Resources Management (SWRM2009), Amsterdam, Netherland, 9–11 September 2009.
5. Buishand, T., Some Methods for Testing the Homogeneity of Rainfall Records. *J. Hydrol.* **1982**, *58*, 11–27.
6. Sahin, S.; Cigizoglu, H.K. Homogeneity Analysis of Turkish Meteorological Data set. *Hydrol. Process.* **2010**, *24*, 981-992.
7. Pettitt, A.N., A non-parametric approach to the change-point detection. *Appl. Stat.* **1979**, *28*, 126-135.
8. Riabova, S. Application of wavelet analysis to the analysis of geomagnetic field variations. *J. Phys. Conf. Ser.* **2018**, *1141*, 012146. https://doi.org/10.1088/1742-6596/1141/1/012146