Journal of Hydrology 548 (2017) 536–551

Contents lists available at ScienceDirect

Journal of Hydrology

journal homepage: www.elsevier.com/locate/jhydrol

Research papers

Ordinal pattern dependence between hydrological time series

Svenja Fischer[a,*], Andreas Schumann[a], Alexander Schnurr[b]

[a] Institute of Hydrology, Ruhr-University Bochum, 44801 Bochum, Germany
[b] Department Mathematik, Universität Siegen, 57068 Siegen, Germany

ARTICLE INFO

Article history:
Received 22 December 2016
Received in revised form 22 February 2017
Accepted 14 March 2017
Available online 18 March 2017
This manuscript was handled by Corrado Corradini, Editor-in-Chief

Keywords:
Ordinal patterns
Discharge correlation
Homogeneous groups

ABSTRACT

Ordinal patterns provide a method to measure dependencies between time series. In contrast to classical correlation measures like the Pearson correlation coefficient they are able to measure not only linear correlation but also non-linear correlation even in the presence of non-stationarity. Hence, they are a noteworthy alternative to the classical approaches when considering discharge series. Discharge series naturally show a high variation as well as single extraordinary extreme events and, caused by anthropogenic and climatic impacts, non-stationary behaviour. Here, the method of ordinal patterns is used to compare pairwise discharge series derived from macro- and mesoscale catchments in Germany. Differences of coincident groups were detected for winter and summer annual maxima. Hydrological series, which are mainly driven by annual climatic conditions (yearly discharges and low water discharges) showed other and in some cases surprising interdependencies between macroscale catchments. Anthropogenic impacts as the construction of a reservoir or different flood conditions caused by urbanization could be detected.

1. Introduction

Catchments and river basins are the spatial units considered in hydrology. They are conceptualized as complex dynamical systems where deterministic and stochastic processes occur simultaneously. It is recognized that there is a strong need for the classification of catchments and hydrological phenomena (Blöschl et al., 2013) in the framework of comparative hydrology. The similarity of discharge regimes is a fundamental criterion for regionalisation. For this purpose, time series of characteristic runoff values (averages, upper and lower extremes), which are derived from discharge series have to be compared to analyse the impacts of spatial heterogeneous distributed climate characteristics as well as of differences of hydrological processes at the catchment scale. The spatial covariance (correlation) is often applied as a measure of the interrelationship between time series. It is the statistical basic tool for interpolation and consistent mapping of runoff and its statistical descriptors. One problem of this approach consists in the non-linearity of runoff processes. To give an example: an urbanized catchment will react more directly on a flood inducing rainfall than a natural one. If we consider two rain events, which differ in size, the urbanized catchment will produce more runoff from the higher amount of rainfall in relationship to the natural catchment. If we compare the runoff data from both catchments we see the same tendencies (one event is higher than the other), but the quantitative relationships between both events differ between the two catchments. As a result, a linear regression model would not be appropriated to describe the statistical relationship between the two runoff series. Ordinal patterns (Bandt and Pompe, 2002) are a simple approach to characterize the synchronicity of time series without quantification of the variances of the time series, which are affected by non-linearities or scale effects and not comparable in many cases. Ordinal patterns were applied in hydrology for time series analyses, e.g. to separate deterministic and stochastic parts of daily discharge series (Lange et al., 2013). Lange et al. (2013) estimated two indices, the permutation entropy and the permutation statistical complexity to quantify order pattern distributions by their information content and complexity. Ordinal patterns have been used in other fields of science for pattern recognition e.g. to analyse EEG data (Keller and Sinn, 2005), sunspot numbers (Bandt and Shiha, 2007), speech signals (Bandt, 2005) and chaotic maps (Bandt and Pompe, 2002). Further applications include estimation of the Hurst parameter in long-range-dependent data (Sinn and Keller, 2011) and the approximation of the Kolmogorov-Sinai entropy (Keller et al., 2013). Let us emphasize that before Schnurr (2014) appeared, all of the above mentioned authors used ordinal pattern analysis only difference, here the ordinal patterns flood peaks between gauges at different spatial

* Corresponding author.
E-mail addresses: svenja.fischer@rub.de (S. Fischer), andreas.schumann@rub.de (A. Schumann), schnurr@mathematik.uni-siegen.de (A. Schnurr).

http://dx.doi.org/10.1016/j.jhydrol.2017.03.029
0022-1694/

scales. Its potential to identify anthropogenic changes in runoff series is shown as well as a comparison with classical correlation methods.

2. Methodology

When comparing time series one often has to face the following problem: two data series show interdependencies but are not correlated in the mathematical sense of the word. Let us explain this in detail: in every-day-life, one would say that two data series are positively correlated if the following holds true:

$$\text{if} X \left\{ \begin{array}{c} \text{increases} \\ \text{decreases} \end{array} \right\}, \text{then} Y \text{is likely to} \left\{ \begin{array}{c} \text{increase} \\ \text{decrease} \end{array} \right\}, \text{too}.$$

Negative correlation would then mean that

$$\text{if} X \left\{ \begin{array}{c} \text{increases} \\ \text{decreases} \end{array} \right\}, \text{then} Y \text{is likely to} \left\{ \begin{array}{c} \text{decrease} \\ \text{increase} \end{array} \right\}.$$

Admittedly, in the context of certain models this behaviour is caught nicely by the mathematical correlation between time-series (or their increments). On the other hand it is well known that mathematical correlation measures linear dependence. If data is correlated, but not in a linear way, mathematical correlation might not be the method of choice. Furthermore, in order to deal with mathematical correlation, both time series have to have second moments, that is, the variance has to be finite. Several interesting models, like so called α-stable random variables, do not have this property. This means between two α-stable time series, we cannot use mathematical correlation. Last but not least in all classical approaches the time series have to be stationary from the beginning (or after a careful pre-processing).

Nevertheless, hydrological time series often are modified by anthropogenic or climatic impacts. Climate variability as well as water management or land-use changes lead to a non-stationary behaviour like changing means or variances in hydrological time series. A reservoir for example has the aim to compensate the fluctuations of runoff at the annual time scale. In this case, especially the upper and lower extremes may show significant changes. Hence, many recent results consider the non-stationarity of the data (see Clarke, 2007; Salas and Obeysekera, 2014; Liu et al., 2015; Serinaldi and Kilsby, 2015 and the references therein).

Here, we suggest a simple approach to describe dependencies between time series by ordinal patterns (Schnurr, 2014; Schnurr and Dehling, 2016), where the probability measures do not need to have second moments. The time series we consider do not have to be stationary. And: in a certain sense we are able to measure non-linear correlation. The basic idea is to reduce the data to so called ordinal patterns and then count, how often one finds pairwise the same patterns at the same instants of time in two data sets.

2.1. The specification of ordinal patterns

For a fixed number of consecutive data points n, their ordinal pattern describes the relative positions of the points. Let $x_1, x_2, \ldots$ be the realized data of a time series. Fix the number of considered data points $n = 4$ (often $n \in \{2, 3, 4, 5, 6\}$), respectively the number of increments $h = n - 1$. Let us consider the first four data points x_1, x_2, x_3, x_4 and assume that the four values are pairwise different, e.g. $x_1 = 2, x_2 = 9, x_3 = 3$ and $x_4 = 11$. Order them top-to-bottom: $x_4 > x_2 > x_3 > x_1$. Then write down the indices of the data points in that order: $(4, 2, 3, 1)$. This vector in $\mathbb{N}^n$ is called the ordinal pattern of (x_1, x_2, x_3, x_4). We write $(r_1, \ldots, r_4)$ for this vector. The pattern $(r_1, \ldots, r_4)$ contains the whole information of relative positions of the data points, but nothing more. Therefore, the information is reduced significantly. For each time point t one now has to consider $(x_{t+1}, x_{t+2}, x_{t+3}, x_{t+4})$ in the same way. For each starting point t we obtain an n-dimensional vector consisting of the entries $1, 2, 3, \ldots, n$. A vector of this kind is called a permutation.

It could be a problem that for different i and j the measured values of x_i and x_j do coincide. In order to have a unique representation, we demand in addition:

$$\text{if} i < j \quad \text{and} \quad x_i = x_j \quad \text{then} \quad r_i < r_j.$$

For example, in the case $(x_{t+1}, x_{t+2}, x_{t+3}, x_{t+4}) = (7, 10, 7, 5)$ we would obtain $(2, 1, 3, 4)$ as the ordinal pattern at time t.

In order to get a better intuition of the meaning of the permutations, one could in fact think of the patterns as an archetype structure as in the Fig. 1.

Instead of the two data sets x and y we, from now on, consider only the sequence of patterns in both time series.

We count how often we find coincident patterns in two series. Coincident patterns do mean that for the given length n (of the time windows) the up-and-down behaviour of the two time series is similar within the two synchronous windows (Fig. 2). E.g. if we have $(4, 2, 3, 1)$ this means we start on a low value, increase, go back to a point in between the first two and then have the highest value in the end.

2.2. A measure to assess the significance of coincidences of ordinal patterns between two times series

In the second step, we estimate a measure to compare the number of coincident patterns with its random value. This comparison value is obtained in the following way: we assume for a moment that the two time series are independent. Let us denote by $P_X(r)$ the probability that the pattern r appears in the time series X (same with Y). In the case $h = 3$ we would have the 24 different patterns that are shown in Fig. 1. If the time series were independent, the probability that r appears in both time series at the same time would be $P_X(r) \cdot P_Y(r)$ and the overall probability to find the same pattern in both time series at a given time would be:

$$q := \sum_{r=(r_1,\ldots,r_n)} P_X(r) \cdot P_Y(r) \tag{1}$$

where we sum over all patterns r of length n. This is only a theoretical construct. Caused e.g. by seasonality, some patterns will occur more often in hydrological time series than others. In practice, for each time series we estimate the empirical probabilities of the single patterns by their relative frequencies. The comparison value v_c is the estimator of q, based on relative frequencies of the patterns in both time series multiplied with $(N - h)$, where N is the number of observations. This means: The comparison value v_c is the number of coincident patterns which we would expect if the time series were independent.

$$v_c = (N - h) \sum_{r=(r_1,\ldots,r_{h+1})} \widehat{P}_X(r) \cdot \widehat{P}_Y(r), \tag{2}$$

where $\widehat{P}_X(r)$ denotes the relative frequency of the pattern r in the sample X.

Let us recall some of the advantages of the method which have been emphasized in Schnurr and Dehling (2016): the whole analysis is stable under monotone transformations of the state space. The ordinal structure is not destroyed by measurement errors or small perturbations of the data. Structural breaks in a single time series do not effect the ordinal pattern dependence significantly. There are fast algorithms to analyse the relative frequencies of ordinal patterns in given data sets (cf. Keller et al., 2007, Section 1.4). Furthermore, let us again emphasize that unlike other concepts which are based on mathematical correlation, we do not