Water Resources Research

RESEARCH ARTICLE

10.1029/2018WR023055

Key Points:

- Spatiotemporal dependence makes hydroclimatic record-breaking events much less surprising than expected
- Poisson binomial and beta-binomial models describe the frequency of overthreshold and record-breaking events under space-time dependence
- BetaBitST, a new generator of correlated binary random fields mimicking the occurrence of extreme events over time and multiple locations, is introduced

Supporting Information:

- Supporting Information S1

Correspondence to:

F. Serinaldi,
francesco.serinaldi@ncl.ac.uk

Citation:

Serinaldi, F., & Kilsby, C. G. (2018). Unsurprising surprises: The frequency of record-breaking and overthreshold hydrological extremes under spatial and temporal dependence. *Water Resources Research*, 54, 6460–6487. <https://doi.org/10.1029/2018WR023055>

Received 31 MAR 2018

Accepted 22 JUN 2018

Accepted article online 29 JUN 2018

Published online 15 SEP 2018

Unsurprising Surprises: The Frequency of Record-breaking and Overthreshold Hydrological Extremes Under Spatial and Temporal Dependence

Francesco Serinaldi^{1,2} and Chris G. Kilsby^{1,2}

¹School of Engineering, Newcastle University, Newcastle upon Tyne, UK, ²Willis Research Network, London, UK

Abstract Record-breaking (RB) events are the highest or lowest values assumed by a given variable, such as temperature and precipitation, since the beginning of the observation period. Research in hydroclimatic fluctuations and their link with this kind of extreme events recently renewed the interest in RB events. However, empirical analyses of RB events usually rely on statistical techniques based on too restrictive hypotheses such as independent and identically distributed (*i/id*) random variables or nongeneral numerical methods. In this study, we propose some exact distributions along with accurate approximations describing the occurrence probability of RB and peak-over-threshold (POT) events under general spatiotemporal dependence, which enable analyses based on more appropriate assumptions. We show that (i) the Poisson binomial distribution is the exact distribution of the number of RB events under *i/id*, (ii) equivalent binomial distributions are accurate approximations under *i/id*, (iii) beta-binomial distributions provide the exact distribution of POT occurrences under spatiotemporal dependence, and (iv) equivalent beta-binomial distributions provide accurate approximations for the distribution of RB occurrences under spatiotemporal dependence. To perform numerical validations, we also introduce a generator of spatially and temporally correlated binary processes, called BetaBitST. As examples of application, we study RB and POT occurrences for monthly precipitation and temperature over the conterminous United States and reanalyze Mauna Loa daily temperature data. Results show that accounting for spatiotemporal dependence yields strikingly different conclusions, making the observed frequencies of RB and POT events much less surprising than expected and calling into question previous results reported in the literature.

1. Introduction

Emphatic adjectives like *unprecedented*, *surprising*, or *exceptional* are often used to describe extreme hydroclimatic events, such as storms, floods, and droughts (Colucci et al., 2017; Coumou & Rahmstorf, 2012; Thompson et al., 2017). However, such adjectives are usually applied in their colloquial sense, while discussions concerning physical phenomena, engineering, and decision-making should rely on words having a more precise meaning. Terms like *surprise* are intrinsically relative in the sense that they are related to the expectations of the observer and imply a comparison of the event under consideration with what is known based on the past experience (Itti & Baldi, 2009; Kjeldsen & Prosdociimi, 2018; Merz et al., 2015). The involvement of the observer's mindset in the definition and recognition of surprising events reveals the psychological nature of surprise. The notion of psychological surprise is not new in water science, where it was introduced by Fiering and Kindler (1984) and Matalas (2003), who built on the principles of the theory of investment decisions proposed by Shackle (1942). According to Shackle (1949), surprise is the state of mind following the occurrence of *counterexpected* or *unexpected events*, where the former occurs in a state of so-called *imperfect knowledge*, whereas the latter in a state of *incomplete knowledge*. Imperfect knowledge and counterexpected events refer to the case in which all possible outcomes of a decision process are known but some of them are deliberately excluded from the so-called *inner group* of credible (unsurprising) options (Shackle, 1942). On the other hand, incomplete knowledge and unexpected events concern the case in which all possible outcomes are not knowledgeable at the time a decision is made (Matalas, 2003).

Imperfect knowledge can be addressed by probability theory and (frequentist) statistical tools, whereas incomplete knowledge requires a more careful treatment since unexpected events fall outside the set of *imagined* options, and "Thus an important surprising event will require [an individual] more or less to create

afresh his structure of expectations . . . We shall call this process the assimilation of the event into the structure of expectations" (Shackle, 1942). This assimilation process, resulting in updated mental models, is basically a Bayesian procedure where new data observations update the prior distribution of beliefs, thus yielding posterior belief distributions accounting for the new information. In this context, the index of surprise proposed by Itti and Baldi (2009) enables the quantification of surprise of unexpected events in terms of Kullback-Leibler divergence (Kullback & Leibler, 1951), that is, the average of the log-odd ratio of posterior and prior beliefs. Conversely, surprise of counterexpected events can be measured by indices such as Weaver's index (Weaver, 1948) relying on the ratio of the *average amount of probability we can expect to realize per trial of the experiment in question* and the probability associated to the realized event. The definition of Weaver's index is genuinely frequentist and implies that the set of outcomes is exhaustive (i.e., the knowledge is imperfect but not incomplete).

Dealing with hydroclimatic variables, we can use historical observations and/or other (empirical and theoretical) sources of information to assign a given probability to every event whose magnitude exceeds a given value. Therefore, also the largest observed value of a hydroclimatic variable (i.e., the so-called *record* or *record event*) can be exceeded (*broken*) with a given probability. Moreover, under some assumptions discussed later, it can be shown that the probability of observing a record-breaking (RB) event over a future time window, such as the design life of an infrastructure, is exceedingly high, thus making such event expected. Therefore, to what extent a future RB event exceeds the magnitude of the most recent record matters more than how frequently they occurred or may occur in the future (Matalas, 1997). In this respect, Kjeldsen and Prosdocimi (2018) studied the level of surprise of RB flood events in the United Kingdom by applying an index of surprise based on the magnitude of the top ranked events (Solow & Smith, 2005). It should be noted that Solow-Smith's index is a frequentist metric dealing with counterexpected events (imperfect knowledge) rather than truly unexpected events (incomplete knowledge). However, despite the importance of event magnitude, several studies focused on the occurrence of RB events in hydroclimatology (Anderson & Kostinski, 2010; Bassett, 1992; Benestad, 2003, 2004; Finkel & Katz, 2017; Matalas, 1997; Meehl et al., 2009; Newman et al., 2010; Rahmstorf & Coumou, 2011; Vogel et al., 2001).

In light of the above remarks, both peak-over-threshold (POT) and RB events are counterexpected since their probability is low compared with that of other possible outcomes. However, as these extreme values are included in the set of possible (even if rare) outcomes, they should not be considered as unexpected (in Shackle's sense). Nonetheless, the analysis of the properties of POT and RB events (e.g., the cumulative number of RB values over a time window) is often used to infer the possible change of the frequency of occurrence of such events under climate change or other forcing factors. In other words, this type of analysis is often used to support a paradigm shift (e.g., from stationary to nonstationary representation) corresponding to an update of the prior belief, when the behavior of empirical values of RB and/or POT statistics computed under a specific model (e.g., *i/id*) shows *substantial* discrepancy with values expected under that model. This updating procedure implies implicitly or explicitly an attribution of the observed discrepancies to the factors (e.g., anthropogenic factors) embedded in the alternative model whose expectations are closer the empirical values computed under the new model. However, since a rigorous attribution requires the careful assessment of multiple lines of evidence leading to the identification, within a prespecified margin of error, of unique causes and exclusion of any other plausible alternative (Hasselmann, 1997; Mitchell et al., 2001; Serinaldi et al., 2018), performing POT and RB analysis under different reasonable prior models is paramount.

A possible alternative to *i/id* hypothesis is the assumption that the underlying process is stationary and correlated, that is, variables are nonindependent and identically distributed (*ni/id*). Correlation inflates the variability of the expected values and the width of confidence intervals (CIs). This behavior is related to information redundancy. Matalas and Langbein (1962) provided a first detailed discussion on the impact of both spatial and temporal correlations on the information content of the mean of a hydrological time series. Since then several studies focused on the effect of spatiotemporal correlation on different inference problems, such as analysis of variance (Jones, 1975), estimation of distribution quantiles (Koutsoyiannis, 2003), trend hypothesis testing (Bayazit & Önöz, 2007; Douglas et al., 2000; Hamed & Rao, 1998; Hamed, 2008, 2009a; Katz, 1988a; Kulkarni & von Storch, 1995; Serinaldi & Kilsby, 2016a; Serinaldi et al., 2018; Yue & Wang, 2002, 2004), and interaction between spatial and temporal correlations in various applications (Hamed, 2009b, 2011; Katz, 1988b; Katz & Brown, 1991), just to mention a few.

The *ni/id* hypothesis has also been considered as an alternative to *i/id* assumption in the context of RB analysis, in order to distinguish systematic changes and spurious fluctuations due to spatiotemporal dependence. Vogel et al. (2001) proposed analytical expressions (based on theoretical results and Monte Carlo simulations) of the moments of the number of RB events considering the average regional value of spatial correlation (under the assumption of multivariate Gaussian distribution). Benestad, (2003, 2004) explored the effect of temporal correlation on the RB statistics by Monte Carlo simulations, using a first-order autoregressive (AR) model, while spatial correlation was either introduced by sampling a set of time series many times (Benestad, 2003) or removed by resampling the original data set (Benestad, 2004). Newman et al. (2010) further explored the effect of autocorrelation by Monte Carlo simulations considering fractional Gaussian noise (fGn) with a power law decaying autocorrelation function (ACF) and power spectral density. Meehl et al. (2009) applied block bootstrap to preserve short-term (3-year) autocorrelation and spatial correlation in order to assess the uncertainty of observed and modeled statistics of RB temperatures, averaged over the conterminous United States (CONUS).

From a theoretical point of view, results mainly refer to the statistics of the standard RB process corresponding to an infinite sequence of independent identically distributed (*i/id*) observations (Arnold et al., 2011, p. 7), while nonstandard variations corresponding to nonindependent or nonidentically distributed variables are less developed and often involve specific dependence structures (Arnold et al., 2011, pp. 208–215) or models whose applicability is generally limited to low-dimensional cases (Nagaraja et al., 2002). Similar lack of general theoretical results also concerns the statistics of occurrence of events exceeding a given percentage threshold (hereinafter, *POT events*) under spatial and temporal dependence. In fact, while the number of POT events over a finite number of time steps and/or spatial locations is described by a binomial ($\mathcal{B}$) distribution under *i/id* assumption, the lack of independence is usually treated by using Monte Carlo methods (e.g., Renard & Lang, 2007).

In this study, we show that the Poisson binomial ($\mathcal{PB}$) distribution describes the occurrence of RB events under *i/id* conditions of the underlying process, overcoming the computational burden of computing Stirling numbers of the first kind, while equivalent binomial ($\mathcal{B}_e$) distributions provide accurate approximations under *i/id* hypothesis. Moreover, the beta-binomial ($\mathcal{BB}$) distribution describes the number of POT occurrences under general spatiotemporal dependence, while equivalent beta-binomial distributions ($\mathcal{BB}_e$) give accurate approximations of the distribution of RB occurrences under spatiotemporal dependence. We also show that the parameters of $\mathcal{B}_e$, $\mathcal{BB}$, and $\mathcal{BB}_e$ distributions only depend on the rate of occurrence and/or the spatiotemporal correlation function (STCF).

In order to check the accuracy of $\mathcal{BB}$ and $\mathcal{BB}_e$ distributions and their generality, we also introduce an extension of the so-called BetaBit algorithm proposed by Serinaldi and Lombardo (2017) to generate binary processes with given temporal correlation structure. This approach, called BetaBitST, enables the simulation of binary random fields with desired STCF under minimal assumptions to be used as reference benchmarks in Monte Carlo analysis. Altogether, $\mathcal{PB}$, $\mathcal{B}_e$, $\mathcal{BB}$, $\mathcal{BB}_e$, and BetaBitST represent a set of new practical tools to study and make inference on POT and RB occurrence processes.

This study is organized as follows. Section 2 introduces the heuristic reasoning that lead us to consider $\mathcal{BB}$, $\mathcal{BB}_e$ as suitable distributions for the number of POT and RB events, respectively, under spatiotemporal dependence. Section 3 discusses the role of $\mathcal{PB}$ and the $\mathcal{B}_e$ approximation under *i/id* hypothesis. Sections 4 and 5 provide technical details concerning $\mathcal{BB}$ and $\mathcal{BB}_e$ distributions. In section 6, we introduce the BetaBitST generator and use it to validate the $\mathcal{BB}$ and $\mathcal{BB}_e$ models described in sections 4 and 5. We therefore analyze RB and POT occurrences in monthly temperature and precipitation across the CONUS and reanalyze Mauna Loa daily temperatures in section 7. Discussion and conclusions are reported in section 8.

2. Preliminary Remarks on the Occurrence of POT Events Under Independence and Nonstationarity

Since the main results discussed in the following sections build on previous findings reported by Obeysekera and Salas (2016), we first summarize their key results and use the models applied in their case study to support our discussion. Obeysekera and Salas (2016) studied the distribution of the number of events, Z , exceeding a given design (critical) value x_{d0} during the design (project) life spanning for instance n years, under the assumption that the discrete-time stochastic process $X = \{X_j\}_{j \in \mathbb{N}}$ (e.g., annual peak flow) is a sequence of independent nonidentically distributed (*i/nid*) random variables with univariate distributions

$F_j(x) = \mathbb{P}[X_j \leq x]$. Under these conditions, the occurrence or absence of an extreme event exceeding x_{d0} in year j is described by a Bernoulli process $Y = \{Y_j\}_{j \in \mathbb{N}}$ with state space $\{0, 1\}$, where $j (= 0, 1, 2, \dots)$ denotes discrete time, and $Y_j = 1$ if $X_j > x_{d0}$; otherwise $Y_j = 0$. Under nonstationary conditions, the variables Y_j are non-identically distributed with time-varying occurrence probability given by $p_j = \mathbb{P}[Y_j = 1]$. With this notation, the number of events exceeding a given design value x_{d0} during n time steps is defined as $Z(\mathbf{Y}) = \sum_{j=1}^n Y_j$, where $\mathbf{Y} = \{Y_1, Y_2, \dots, Y_n\}$. For *i/id* Bernoulli trials, Z is distributed as a $\mathcal{PB}$ distribution (Hong, 2013; Tejada & den Dekker, 2011; Wang, 1993; Zaigraev & Kaniovski, 2013) whose probability mass function (pmf) and cumulative distribution function (cdf) are, respectively,

$$f_{\mathcal{PB}}(z) = \mathbb{P}[Z = z] = \sum_{\mathbf{y}: Z(\mathbf{y})=z} \prod_{j=1}^n p_j^{y_j} (1 - p_j)^{1-y_j}, \quad (1)$$

$$F_{\mathcal{PB}}(z) = \mathbb{P}[Z \leq z] = 1 - \sum_{\mathbf{y}: Z(\mathbf{y}) \geq (z+1)} \prod_{j=1}^n p_j^{y_j} (1 - p_j)^{1-y_j}. \quad (2)$$

Hong (2013) introduced an exact formula with a closed-form expression to compute the $\mathcal{PB}$ distribution function, developed an algorithm for efficient implementation, and studied the advantages and disadvantages of various approximation methods in order to overcome the computational burden associated with extensive enumerations involved in equations (1) and (2), especially for high n (i.e., $n > 2,000$).

In order to show the applicability of $\mathcal{PB}$ distribution, Obeysekera and Salas (2016) used probabilities p_j resulting from a generalized extreme value (GEV) distribution describing the annual peak flows of the Assunpink Creek watershed in Trenton, New Jersey, with parameters varying along 100-year design life according to the following parametrization:

$$p_j = 1 - \exp \left\{ - \left[1 + \frac{\varepsilon}{\sigma} (x_{d0} - \mu_j) \right]^{-1/\varepsilon} \right\}, \quad (3)$$

where $\mu_j = 44.587 + 0.306(j - 1968.027)$ is a time-varying location parameter in year j and $\sigma = 16.617$ and $\varepsilon = 0.136$ are time-invariant scale and shape parameters, respectively. To support our discussion, Figure 1 reproduces Figure 2 in the work by Obeysekera and Salas (2016). Figure 1a shows the time-varying probabilities p_j (for each year $j \in \{1, 2, \dots, 100\}$ of the design life) corresponding to the values $p_0 \in \{0.04, 0.1, 0.2, 0.5\}$ for $j = 0$, while Figure 1b displays the cdfs of the number of POT events computed by $\mathcal{B}$ distributions for the *i/id* case with $p_0 \in \{0.04, 0.1, 0.2, 0.5\}$ and by the $\mathcal{PB}$ distribution for the *i/nid* case. It is worth noting that the GEV model in equation (3) is only used to highlight theoretical properties concerning the link between time-varying probabilities and $\mathcal{PB}$ distribution and their subsequent applicability to RB analysis. We refer the reader to the supplementary material of Serinaldi and Kilsby (2015) for a critical discussion of the use of the GEV distribution in equation (3) to describe annual peak flows of the Assunpink Creek.

Figure 1b highlights the positive shift in the number of occurrences introduced by the increasing values of GEV location parameter. Of course, an opposite shift (i.e., fewer POT events) would occur if the GEV location parameter decreased in time. It should be noted that the first two central moments of $\mathcal{PB}$ (i.e., mean and variance) are given by the relationships (Hong, 2013)

$$\mu_{\mathcal{PB}} := \mathbb{E}[Z] = \sum_{j=1}^n p_j, \quad (4)$$

$$\sigma_{\mathcal{PB}}^2 := \text{Var}[Z] = \sum_{j=1}^n p_j(1 - p_j). \quad (5)$$

Under *i/id* conditions, with $p_j = p_0$ for all j s, equations (4) and (5) yield $\mu_B = np_0$ and $\sigma_B^2 = np_0(1 - p_0)$, corresponding to mean and variance of a $\mathcal{B}$ distribution describing the conditions at the beginning of the design life. The mean and variance of $\mathcal{PB}$ distribution can also be written as (Edwards, 1960; Poisson, 1837; Wang, 1993)

$$\mu_{\mathcal{PB}} = n\bar{p}, \quad (6)$$

$$\sigma_{\mathcal{PB}}^2 = n\bar{p}(1 - \bar{p}) - n\sigma_p^2, \quad (7)$$

Figure 1. (a) The time-varying probabilities p_j , for $j \in \{1, 2, \dots, 100\}$, corresponding to the model in equation (3) discussed by Obeysekera and Salas (2016) and denoted as OS2016 case. Each curve refers to a given value $p_0 \in \{0.04, 0.1, 0.2, 0.5\}$ at time $j = 0$. (b) The cumulative distribution functions of the number of peak-over-threshold events computed by B distributions for the stationary case with $p_0 \in \{0.04, 0.1, 0.2, 0.5\}$ and by PB distribution for nonstationary case. (c and d) Similar to panels (a) and (b), respectively, but compare OS2016's model and the stationary case with $\bar{p} = \frac{1}{100} \sum_{j=1}^{100} p_j$. (e–h) Similar to panels (a) and (b) but for randomly varying probabilities p_j corresponding to GEV location and scale parameters randomly varying around the values of μ_0 and σ reported in the text and OS2016. GEV = generalized extreme value.

where $\bar{p} = \frac{1}{n} \sum_{j=1}^n p_j$ and $\sigma_p^2 = \frac{1}{n} \sum_{j=1}^n (p_j - \bar{p})^2$ are the mean and variance of the set of probabilities $\{p_1, \dots, p_n\}$, respectively. Therefore, σ_{PB}^2 increases as the set of probabilities $\{p_1, \dots, p_n\}$ tends to be more and more homogeneous and attains its largest value when all p_j values are identical (Tejada & den Dekker, 2011; Wang, 1993). From equations (6) and (7), it follows that one can approximate the PB distribution (describing a nonstationary Bernoulli process) with a simpler equivalent binomial B_e distribution describing the number of failures/successes over n trials for an equivalent stationary Bernoulli process with average rate of occurrence $\bar{p}$. The B_e model preserves exactly the mean of PB distribution, as $\mu_{B_e} = n\bar{p} = \mu_{PB}$, while the variance is at most overestimated by a quantity $n\sigma_p^2$, thus giving a cautionary estimate of the variability of the number of successes/failures over n trials. Referring to the GEV example mentioned above, Figure 1c shows the nonstationary GEV probabilities p_j (already shown in Figure 1a) and the average $\bar{p}$ corresponding to the equivalent stationary Bernoulli process, while Figure 1d highlights the remarkable agreement of PB and B_e models. In this example, the sequence $\{p_1, \dots, p_{100}\}$ with $p_0 = 0.5$ has $\bar{p}(1-\bar{p}) = 0.16$ and $\sigma_p^2 = 0.02$. Obviously, the ratio of the two components of variance in the right side of equation (7) depends on the absolute values of p_j and how they vary with time. However, this example and the case discussed in the next section support the use of an equivalent stationary Bernoulli process and the corresponding B_e distribution in practical applications.

To further explore this point, we also considered GEV distributions with location and scale parameters randomly varying around the values assumed at time $j = 0$ (i.e., $\mu_0 = 44.587$ and $\sigma = 16.617$). In particular, we assumed additive error models for both location and scale parameters such that $\mu_\varepsilon = \mu_0 + \varepsilon$, with $\varepsilon \sim \mathcal{N}(0, 8)$, and $\sigma_\varepsilon = \sigma + \varepsilon$, with $\varepsilon \sim \mathcal{N}(0, 5)$, where $\mathcal{N}$ denotes a Gaussian distribution. Figure 1e shows the exceedance probabilities corresponding to the model with randomly varying location around μ_0 along with the reference probabilities resulting from the GEV distribution with constant parameters. Figure 1f highlights that the random fluctuations do not influence the distribution of the number of events. The same conclusion holds for the GEV model with randomly varying scale parameter (Figures 1g and 1h). The insensitivity of the probability of failure to randomly fluctuating parameters is an important aspect since it highlights a critical characteristic of frequency analysis relying on models with parameters depending on covariates exhibiting random behavior, such as teleconnection indices. In these cases, the resulting models are not truly nonstationary distributions, as is often incorrectly referred to in the literature, but simply compound stationary distributions with negligible or no effect on the probability of failure in a given time window. Confusing compound distributions with nonstationary distributions (whose parameters depend on time via deterministic functions) can lead to misleading interpretations and conclusions. We refer to Serinaldi and Kilsby (2015) and Serinaldi et al. (2018) for further discussion of these aspects. These concepts and remarks will be used in the next sections in the context of the analysis of POT and RB events under different conditions of spatial and temporal dependence.

3. Updating the Analysis of RB Events: The Role of $\mathcal{PB}$ and $\mathcal{B}_\varepsilon$ Distributions

As mentioned in section 1, RB values are observations exceeding all previous observations of the same variable. The statistical theory of RB values was introduced by Chandler (1952), and the main theoretical results refer to an infinite sequence of *i.i.d* observations of a random variable X with cdf F , which is assumed to be nondegenerate in order to avoid the possibility of ties, that is, identical values in the sequence (Arnold et al., 2011; Glick, 1978). An observation x_j is a *RB high value* if $x_j > x_i$ for every $i < j$. An analogous definition holds for *RB low values*, whereby $x_j < x_i$ for every $i < j$. With no loss of generality, we will refer to RB high values if not otherwise specified. When dealing with RB events, several statistics can be of interest, for example, RB time (i.e., the time of occurrence of an RB value), RB magnitude, RB increment process (i.e., the difference of magnitude of subsequent RB values), and inter-RB time (i.e., the number of time steps between RB values). In this study, we focus on the number of RB events over a given time window.

Formally, given a sequence $\{x_1, \dots, x_n\}$ of observations of the variable X , the occurrence of RB values in the sequence defines a binary process such that $Y_j = 1$ if $X_j = \max\{x_1, \dots, x_j\}$; otherwise, $Y_j = 0$ (Vogel et al., 2001). Therefore, the number of RB events in n trials (e.g., n years of annual data) is $Z = \sum_{j=1}^n Y_j$. We use the same notation applied in section 2 because the occurrence of RB events is a nonstationary Bernoulli process $\{Y_1, Y_2, \dots, Y_n\}$ with mean and variance (Glick, 1978)

$$E[Y_j] = \mathbb{P}[Y_j = 1] = \mathbb{P}[X_j = \max\{x_1, \dots, x_j\}] = \frac{1}{j}, \quad (8)$$

$$\text{Var}[Y_j] = E[Y_j^2] - E^2[Y_j] = \frac{1}{j} - \frac{1}{j^2}, \quad (9)$$

while the mean and variance of the process Z are (Glick, 1978)

$$E[Z] = \sum_{j=1}^n \frac{1}{j} \approx \log(n) + \gamma, \quad (10)$$

$$\text{Var}[Z] = \sum_{j=1}^n \left(\frac{1}{j} - \frac{1}{j^2} \right) \approx \log(n) + \gamma - \frac{\pi^2}{6}, \quad (11)$$

where $\gamma = 0.5772$ is the Euler constant. It is worth noting that Z , along with any other counting statistic, is not affected by the parent cdf F , as the counting process relies on ranks rather than on absolute values. Unlike the example in section 2, here the time-varying probabilities p_j do not come from a nonstationary parent distribution resulting from fitting to data (i.e., induction) but from conceptual considerations (i.e., deduction) according to the following reasoning (Glick, 1978): “The first observation necessarily must be a record ... But, *prior to observing any values*, I know that the second of two numbers in random sequence has equal probability of being smaller or larger than the first. Hence the probability is exactly 50% that a second,

Figure 2. (a) Comparison between PB and Stirling distributions of the number of record-breaking events, Z , over n experiments (time steps). For $n = 1,000$, Stirling distribution is not available as it involves prohibitive combinatorial calculations. (b) Comparison between PB distribution and the asymptotic approximation reported by Glick (1978). (c) PB distribution versus the B_e approximation based on $\bar{p}$. (d) The ratio of the standard deviations of PB distribution and the corresponding B_e approximation as a function of the record length n . cdf = cumulative distribution function.

independent observation will be a new record high surpassing the initial record, assuming that there cannot be an exact tie ... From the same perspective, there is probability $1/3$ that a third trial will be a new maximum, since the last of three repeated observations will be equally likely to be smallest, middle, or largest ... Similarly, all 10 ranks are equally likely for the tenth observation; so maximum rank for the tenth observation has probability $1/10$." This deductive arguments support the use of tools developed for nonstationary processes in contrast to the application of nonstationary models relying on widespread but questionable data-driven inductive inference (see Serinaldi et al., 2018).

While the first two moments of Z have a relatively simple expression, the exact pmf of Z was originally derived in terms of the Stirling numbers of the first kind S_n^z (David & Barton, 1962, pp. 178–183), where

$$\mathbb{P}[Z = z] = \frac{|S_n^z|}{n!}, \quad (12)$$

with

$$S_n^z = \sum_{k=1}^{n-z} (-1)^k \binom{n-1+k}{n-z+k} \binom{2n-z}{n-z+k} \left[\frac{1}{k!} \sum_{j=0}^k (-1)^{k-j} \binom{k}{j} j^{n-z+k} \right]. \quad (13)$$

Because the burden of combinatorial computations increases with n , Vogel et al. (2001) considered the asymptotic result for large sample size reported by Glick (1978)

$$\mathbb{P}[Z = z] = \frac{[\log(n)]^{z-1}}{n \cdot (z-1)!}, \quad (14)$$

which, however, does not provide a very accurate approximation (see Figure 2 in Vogel et al., 2001, and discussion below).

Recalling that the occurrence of RB events associated to an i/id parent process is a nonstationary Bernoulli process with $p_j = 1/j$, it follows that the pmf and cdf of the number of occurrences over n trials are described by a PB distribution, whose computation is readily available in statistical software (Hong, 2013). Moreover, according to remarks in section 2, we can also consider an equivalent Bernoulli process and the corresponding B_e distribution with $\bar{p} = \frac{1}{n} \sum_{j=1}^n \frac{1}{j}$. Figure 2 compares the PB cdf and those obtained by equation (12) (referred to as *Stirling* distribution), equation (14) (denoted as *Asy*), and B_e distribution for $n \in \{10, 50, 150, 1,000\}$. Figure 2a shows that the PB and Stirling cdfs are identical for $n \in \{10, 50, 150\}$, while Stirling is not available for $n = 1,000$ because of computational infeasibility. Therefore, PB provides the exact distribution of Z and it can also be used for (relatively) large n values, avoiding the computational problems of Stirling numbers. Figure 2b confirms that the asymptotic approximation gives biased results even for $n = 1,000$, thus exhibiting a slow rate of convergence. On the other hand, the B_e approximation is unbiased in terms of expectation $E[Z]$ (Figure 2c) and only biased in terms of variance $\text{Var}[Z]$, which is overestimated as per equation (7). Since the time-varying probabilities of occurrence $p_j = 1/j$ are known a priori, we can assess in advance this bias as a function of n . Figure 2d shows that σ_{B_e} overestimates σ_{PB} of the 23% at most for $n = 5$ (17% for $n = 100$,

and 12% for $n = 1,000$). Even if convergence in terms of standard deviation is slow, Figure 2c highlights that the difference of B_e and PB models is very small in terms of cdf shape and probabilities associated to a given value of Z , and the approximation is better than that of the asymptotic one, thus making B_e an additional tool for practical applications.

4. The Number of POT Events Under Spatiotemporal Dependence: Introducing the BB Distribution

In the previous section, we discussed the case of i/nid binary processes, namely, processes comprising independent Bernoulli trials with (deterministic) time-varying probability of occurrence p_j . In this section, we introduce basic results concerning the distribution of the number of occurrences (failures or successes) over n trials, under the assumption of constant p and mutually dependent trials, that is, the ni/id case. In this context, the BB distribution plays a key role. Denoting $\psi \in [0, 1]$, the probability of success/failure in n trials, the BB distribution is a compound distribution resulting from the ordinary B distribution $f_B(z) = \binom{n}{z} \psi^z (1 - \psi)^{n-z}$, when ψ is assumed to be a random variable Ψ following a beta distribution $f_\beta(\psi) = \frac{\psi^{\alpha-1} (1-\psi)^{\beta-1}}{B(\alpha, \beta)}$ with mean $E[\Psi] = p$, where B denotes beta function and α and β are two positive shape parameters. The BB pmf can be written as (Skellam, 1948)

$$f_{BB}(z) = \binom{n}{z} \frac{B(z + \alpha, n - z + \beta)}{B(\alpha, \beta)}, \quad (15)$$

while mean and variance are given by the formulas (Ahn & Chen, 1995)

$$\mu_{BB} := E[Z] = np, \quad (16)$$

and

$$\sigma_{BB}^2 := \text{Var}[Z] = np(1 - p) [1 + (n - 1)\rho_{BB}], \quad (17)$$

where $p = \alpha/(\alpha + \beta)$, and $\rho_{BB} = 1/(\alpha + \beta + 1)$ is known as the *intracluster* or *intracluster* correlation. If the random variable Ψ has a degenerate distribution with probability 1 at a single point (or $\alpha \rightarrow \infty$ and $\beta \rightarrow \infty$), then $\text{Var}[\Psi] = 0$ and Z becomes binomial with $\mu_B = p$ (Ahn & Chen, 1995). Being positive by definition, ρ_{BB} produces overdispersion as it inflates the variance $np(1 - p)$ of the original B distribution with constant p . On the other hand, ρ_{BB} does not affect the expected value, which is identical for BB and B . For correlated experiments, we have

$$\rho_{BB} = \frac{\sum_{j \neq l} \rho_{jl}}{n(n - 1)}, \quad (18)$$

where $\rho_{jl} = \text{Corr}[Y_j, Y_l]$ denotes the pairwise correlation of experiment j and l . The indices j and l can refer to two different time steps in a temporal process evolving over n time steps or two locations in a spatial process over n locations. For a spatiotemporal process over n time steps and m locations, $\rho_{ji, lk} = \text{Corr}[Y_{ji}, Y_{lk}]$ are the element of the $q \times q = nm \times nm$ space-time correlation matrix and equation (18) reads as

$$\rho_{BB} = \frac{\sum_{ji \neq lk} \rho_{ji, lk}}{q(q - 1)}. \quad (19)$$

Note that the BB distribution can arise in a number of ways (Hisakado et al., 2006; Moran, 1968) and is also known as the Polya or negative hypergeometric distribution (Griffiths, 1973). The above derivation (compounding B and beta distributions) is analogous to the derivation of the negative binomial distribution by compounding the Poisson with a gamma distribution (Moran, 1968, pp. 87–91). The Poisson model is the limiting form of the binomial distribution when n is large and p is small, while the negative binomial is the limiting form of the BB distribution when n and $\alpha + \beta$ are large (Hughes & Madden, 1993). The BB distribution has been used in several fields for various purposes such as the modeling of correlated failures and reliability of multiversion software (Nicola & Goyal, 1990), the description of plant disease incidence data (Hughes & Madden, 1993), the estimation of false discovery rates in multiple testing for significance with gene expression data from DNA microarray experiments (Tsai et al., 2003), or the estimation of the rejection rate in multiple trend testing for correlated stream flow data (Serinaldi et al., 2018).

Figure 3. Each panel shows the empirical distribution function of the number of peak-over-threshold events computed on binary time series with fractional Gaussian noise correlation structure and rate of occurrence p simulated by BetaBit, along with the corresponding BB distribution and the B distribution corresponding to i/id conditions. Each panel refers to a specific combination of Hurst parameter H , rate of occurrence p , and sample size n . cdf = cumulative distribution function.

For completeness, we also recall a simple and asymptotically efficient estimator of BB parameters based on the ratio of the first two factorial moments (Tripathi et al., 1994):

$$\begin{cases} \hat{\alpha} = \frac{\hat{\xi}_0(n-1-\hat{\xi}_1)}{\hat{\xi}_0+n(\hat{\xi}_1-\hat{\xi}_0)} \\ \hat{\beta} = \frac{(n-\hat{\xi}_0)(n-1-\hat{\xi}_1)}{\hat{\xi}_0+n(\hat{\xi}_1-\hat{\xi}_0)} \end{cases}, \quad (20)$$

where $\hat{\xi}_i = \frac{\hat{\mu}_{i+1}}{\hat{\mu}_i}$ and $\hat{\mu}_i$ denotes the i th sample factorial moment. Adopting the convention that $\hat{\mu}_0 = 1$, the estimators in equation (20) require only the computation of $\hat{\mu}_1 = \frac{1}{n} \sum_{j=1}^n y_j$ and $\hat{\mu}_2 = \frac{1}{n} \sum_{j=1}^n y_j(1 - y_j)$. Tripathi et al. (1994) showed that such estimators outperform other options, such as maximum likelihood, in terms of asymptotic relative efficiency.

In order to show the accuracy of the BB distribution, we performed a Monte Carlo experiment by simulating binary time series with $p \in \{0.01, 0.02, 0.05\}$, $n \in \{10, 50, 100\}$, and two different ACFs, namely, the one-parameter fGn and Markov dependence structures characterized by parameters $H \in (0.5, 1)$ and $\rho_1 \in (0, 1)$, respectively (see the appendix for further details). The range of p values represents relatively rare events, while the intervals of H and ρ_1 cover the admissible values for positively correlated stationary processes, where the lower limits correspond to independence (i.e., the i/id case). Since hydroclimatic variables generally show positive dependence, empirical estimates of H and ρ_1 usually fall in those intervals. For the sake of space, hereinafter, some results are reported only for representative values of H and ρ_1 (i.e., low, medium, and high positive correlation). Binary correlated samples are generated by BetaBit algorithm (Serinaldi & Lombardo, 2017). For each parameter configuration, we generated 5,000 replications, thus counting the number of occurrences z for each replication. This set of 5,000 z values was used to obtain the empirical cdf (ecdf) of Z . Hereinafter, results for fGn processes are reported in the text, while those for Markov models in the supporting information. Figures 3 and S1 compare ecdfs and fitted B and BB cdfs for some representative parameter configurations, highlighting how the B distribution can strongly underestimate the probability of z values smaller/higher than the expected values np . On the other hand, the BB distribution correctly describes the overdispersion introduced by the temporal dependence structures.

We also checked the accuracy of the estimators in equation (20) by comparing the parameters $\hat{\beta}$ and $\hat{\rho}_{BB}$ estimated on the simulated sequences with the theoretical values p and ρ_{BB} , where ρ_{BB} is computed from equation (18) by using the pairwise correlation terms ρ_{ij} resulting from the theoretical ACFs in equations (A1)

Figure 4. Each panel shows the theoretical parameters of the BB distribution of the number of peak-over-threshold events for binary processes with fractional Gaussian noise dependence structure and a specific combination of p and n as a function of H . Theoretical values are compared with estimates obtained by applying the factorial moment method (equation (20)) on simulated samples.

and (A2). Figures 4 and S2 show the agreement of theoretical and estimated parameters, confirming (i) the suitability of the BB distribution and its parametrization in equation (18) to describe the distribution of Z under ni/id hypothesis, and (ii) the performance of the estimators in equation (20). Moreover, these results further validate the performance of BetaBit in terms of a by-product variable, Z , which is not used in the algorithm structure.

5. The Number of RB Events Under Spatiotemporal Dependence

5.1. Introducing Nonindependent and Nonidentically Distributed (ni/id) Binary Variables

In sections 3 and 4, we discussed respectively the cases i/id (independent binary processes with time-varying rate of occurrence p_j) and ni/id (dependent binary processes with constant rate of occurrence p) focusing on temporal processes. However, in practical applications, we are more often interested in spatiotemporal processes, such as the number of POT or RB events occurring over multiple locations in a given time window

under possible spatiotemporal dependence (e.g., Benestad, 2003, 2004; Renard & Lang, 2007). For POT events, the focus can be, for instance, on the number of flood events exceeding the at-site value corresponding to a given annual exceedance probability p (or return period) in agreement with specified standard of protection (e.g., Interagency Committee on Water Data, 1982; Kjeldsen et al., 2008). The BB model provides the exact distribution of Z even though, to the best of our knowledge, it was never considered before for these type of problems, where simulation approaches have generally been preferred (e.g., Renard & Lang, 2007).

On the other hand, the case of RB events over several locations under spatiotemporal dependence is more challenging as it corresponds to nonindependent and nonidentically distributed (ni/nid) hypothesis for Y . As mentioned above, our review of theoretical literature highlights that some attempts have been made in the past, considering extensions devised for specific problems and resulting in rather complex models that are also difficult to apply/adapt to different applications. Therefore, we propose a general purpose approximation of the distribution of Z under ni/nid that derives from conceptual reasoning by merging the results and arguments reported sections 3 and 4. Our deduction is not analytical but consists of a working hypothesis to be validated *ex post* by simulating binary processes with specified properties, namely, prescribed time-varying p_j and spatiotemporal dependence structure. This approach requires some conceptual elaboration and the availability of a generator of binary random fields with given spatiotemporal correlation structure, which we introduce in the following. We aim to provide a simple model approximating the distribution of Z under ni/nid reasonably well, thus avoiding extensive Monte Carlo simulations.

5.2. BB_e : A BB Distribution for an Equivalent ni/id Process

If the rate of occurrence p is constant, the BB distribution can account for persistence by the parameter ρ_{BB} representing the average of the correlation matrix. On the other hand, the PB model can account for time-varying p_j but not for dependence. However, an i/nid binary process can be approximated by an equivalent i/id process with parameter $\bar{p}$ equal to the average of p_j . Therefore, our assumption is to approximate the distribution of Z under ni/nid by a BB_e model with parameters ρ_{BB_e} and ρ_{BB_e} , such that

$$\rho_{BB_e} = \frac{1}{n} \sum_{j=1}^n \frac{1}{j}, \quad (21)$$

and

$$\rho_{BB_e} = \phi \rho_{BB}, \quad (22)$$

where ϕ is a correction factor to be defined according to the law of variation of p_j . For $p_j = 1/j$, extensive simulations based on the binary generator described in section 6 yield $\phi = 0.75$. The BB_e model corresponds to an equivalent ni/id binary process with average rate of occurrence ρ_{BB_e} and prescribed (spatiotemporal) correlation function.

For an RB process with $p_j = 1/j$, we have $j \geq 2$ in equation (21) because the first observation of a time series is always an RB event by definition, and therefore, its occurrence is fully deterministic. Since the BB_e model only describes the stochastic part of the process with $p_j < 1$, the final distribution of Z over n time steps and m locations is a shifted BB_e distribution $BB_e(z + m; \rho_{BB_e}, \rho_{BB_e})$.

We stress that the probability p_j is generally different from $1/j$ under dependence because this property implies that large (small) values follows large (small) values. In this respect, Benestad (2003) showed that the weak dependence resulting from a first-order AR model, namely, AR(0.3), yields p_j values no significantly different from $1/j$, while Newman et al. (2010) showed that increasing dependence yields $p_j > 1/j$ when the parent process X is fGn or Brownian walk parametrized by the Hurst coefficient H . However, such differences are very small for $H \leq 0.75$ and become significant only for strong dependence ($H \geq 0.875$). Therefore, assuming $p_j = 1/j$ under persistence is acceptable in a wide range of practical applications where X exhibits weak or moderate dependence. Moreover, we focus on the dependence structure of the binary process Y rather than on that of X . Our aim is to show the potential impact of dependence when its nature and magnitude is potentially concealed by small sample sizes that commonly characterize hydrological time series. In fact, in some cases, observations x can appear approximately independent while z exhibits some anomalous behavior contrasting with the reasonable assumption $p_j = 1/j$ supported by the apparent independence. The BB_e distribution approximates the distribution of Z for a binary process that retains the assumption $p_j = 1/j$ (and therefore $E[Z]$) but shows inflated variability and clustering of events due to dependence.

Table 1

Relationships Between the Nature of the Parent Process X (i.e., i/id and ni/id) and That of the Corresponding POT and RB Occurrence Processes Y , Along With the Rates of Occurrence, and Distributions F_Z of the Number of Occurrences Z Over Generic Spatiotemporal Windows

Parent	X	i/id	ni/id	i/nid	ni/nid
	Y	i/id	ni/id	i/nid	ni/nid
POT	Rate	p	p	p_j	p_j
	F_Z	B	BB	PB or B_e	BB_e^*
	Y	i/nid	ni/nid	i/nid	ni/nid
RB	Rate	$1/j$	$\approx 1/j$	p'_j	p'_j
	F_Z	PB or B_e	BB_e	PB or B_e	BB_e^*

Note. For i/nid and ni/nid X , p_j and p'_j denote generic time-varying and generally different rates of occurrence of POT and RB events, respectively. Parametrization (correction factor) is case specific and should be assessed by preliminary simulation.

This process is therefore an alternative to i/id option to be used as a benchmark to check the behavior Z without introducing demanding assumptions related, for instance, to the nonstationarity of the parent process X (Serinaldi et al., 2018).

Table 1 summarizes the relationships between the type of the parent process X (i.e., i/id and ni/id) and that of the corresponding POT and RB occurrence processes Y , along with the rates of occurrence, and distributions F_Z of the number of occurrences Z over generic spatiotemporal windows. It should be noted that the results derived in this study strictly refer to POT and RB cases corresponding to i/id and ni/id parent processes X spanning n time steps over m locations for which the rate of occurrence is constant or $1/j$ (i.e., the cases in the first two columns of Table 1). When X are i/nid , ni/nid , general results are not available as the temporal evolution of the rate of occurrence is case specific and can be different for each location. We also stress that this type of POT and RB spatiotemporal analysis makes sense if we have simultaneous observations over all time steps and locations. Our results are not devised for i/nid , ni/nid X processes with arbitrarily evolving rates of occurrence. In this circumstances, we do not recommend the use of B_e and BB_e models without the support of accurate and extensive preliminary Monte Carlo simulations starting from the process X . However, in these cases, analytical models lose their usefulness as the necessary Monte Carlo experiments already provide the required information about the (case-specific) distributions of Z .

6. BetaBitST: An Extension of BetaBit Algorithm to Simulate Binary Processes With Given Spatial and Temporal Dependence

To check the performance of the BB_e distribution and assess the correction factor ϕ , we need a generator of binary time series and random fields preserving the required p_j and spatiotemporal dependence structure. To accomplish this task, we propose an extension of BetaBit (Serinaldi & Lombardo, 2017), which is an efficient and relatively simple generator of binary time series based on parent Gaussian bivariate distributions (see also Papalexou, 2018, for a detailed discussion of this class of models). We briefly recall the rationale of BetaBit to introduce its spatiotemporal version denoted as BetaBitST.

BetaBit generates a correlated sequence of random numbers $\{y_j\}_{j \in \mathbb{N}}$, for simplicity $\{y_j\}$, taking values 1 and 0 with probability p and $1 - p$, respectively, by generating a sequence of n random numbers $\{x_j\}_{j=0}^{n-1}$ for an auxiliary process X with the desired ACF (e.g., exponential or power law) and standard Gaussian marginal cdf $\mathcal{N}$, and then transforming the marginal distributions into Bernoulli marginals by

$$y_j = \begin{cases} 1, & \text{if } x_j < \mathcal{N}^{-1}(p) \\ 0, & \text{otherwise} \end{cases}, \quad (23)$$

where $\mathcal{N}^{-1}$ denotes the inverse of $\mathcal{N}$, that is, the quantile function. Even though this dichotomization does not preserve linear correlation, the ACF terms of the process X , $\rho_X(\tau) = \text{Corr}[X_j, X_{j+\tau}]$, are related to those of the process Y , $\rho_Y(\tau) = \text{Corr}[Y_j, Y_{j+\tau}]$, by an implicit analytical relationship, $\rho_X = \zeta(\rho_Y)$, that can be approximated with negligible error by a beta cdf F_β with parameters α_p and β_p depending on p (Serinaldi & Lombardo, 2017)

Figure 5. (a and b) An example of the theoretical spatiotemporal correlation functions (STCFs) of the parent Gaussian process X and the corresponding STCF of the binary process Y , respectively, assuming Markov dependence structures in space and time with parameters $\rho_{1,s} = \rho_{1,t} = 0.8$ and rate of occurrence $p = 0.05$. (c) The empirical STCF of a binary random field simulated by BetaBitST using the theoretical STCF in panel (b). As the empirical STCF in panel (c) refers to a single simulated random field, discrepancies between empirical STCF in panel (c) and theoretical STCF in panel (b) are due to finite sample effects (i.e., sampling uncertainty). Note that the ensemble STCF matches the theoretical STCF in panel (b) as expected (figure not shown).

$$\rho_X = \zeta(\rho_Y) \approx F_\beta(\rho'_Y; \alpha_p(p), \beta_p(p)) = \frac{1}{B(\alpha_p, \beta_p)} \int_0^{\rho'_Y} s^{\alpha_p-1} (1-s)^{\beta_p-1} ds, \quad (24)$$

where $\rho'_Y = 2/\pi \sin^{-1}((2/\pi \sin^{-1}(\rho_Y^{0.25}))^{0.25})$ and

$$\begin{cases} \alpha_p = 2.281 + 27.541p^{0.167+0.07p}e^{-0.313p} \\ \beta_p = 0.11 + 14.129p^{0.063+0.177p}e^{-0.147p} \end{cases} \quad (25)$$

Therefore, BetaBit algorithm generates a correlated binary sequence $\{y\}$ as follows:

1. Compute α_p and β_p from equation (25) based on the desired value of p ;
2. Use equation (24) to inflate the terms of the ACF of the auxiliary process X , $\rho_X(\tau)$, in order to obtain the target process Y with the desired ACF, $\rho_Y(\tau)$;
3. Generate a standard Gaussian time series $\{x\}$ with the inflated ACF by algorithms allowing the explicit use of the ACF in the simulation process, such as the method proposed by Davies and Harte (1987) and used in this study;
4. Apply the dichotomization in equation (23).

The BetaBit structure allows a straightforward spatiotemporal extension by using cross-correlated innovations to simulate standard Gaussian time series with given ACF. In particular, cross-correlated (but temporally independent) innovations used in Davies-Harte's algorithm are drawn from a multivariate Gaussian distribution with desired cross-correlation function. This approach is quite standard (e.g., Podgórski & Wegener, 2012) and yields Gaussian processes with separable STCF $\rho_{X,st}(\mathbf{u}, \tau) = \rho_{X,s}(\mathbf{u}) \cdot \rho_{X,t}(\tau)$, where $\rho_{X,s}(\mathbf{u})$ is the cross-correlation function and $\mathbf{u}$ is the distance between two generic locations. The hypothesis of STCF separability means that the STCF can be expressed as the product of the spatial and temporal correlation functions. Intuitively, this hypothesis corresponds to assume that the cross correlation (spatial correlation) can be studied independently of the autocorrelation (temporal correlation; see, e.g., Genton, 2007; Gneiting et al., 2006, and references therein for a technical discussion). Even though more general types of STCFs do exist (Genton & Kleiber, 2015; Gneiting et al., 2006), separable STCFs are the simplest way to account for spatiotemporal dependence especially in cases, such as the occurrence of rare events, where there are often not enough data to justify the use of more complex dependence structures.

Recalling that $\rho_X = \zeta(\rho_Y)$, the STCF of the binary process Y resulting from BetaBitST reads as

$$\rho_{Y,st}(\mathbf{u}, \tau) = \zeta^{-1}(\zeta(\rho_{Y,s}(\mathbf{u})) \cdot \zeta(\rho_{Y,t}(\tau))). \quad (26)$$

Figure 6. Sequences of 20 random fields simulated by BetaBitST with fractional Gaussian noise dependence structures, $p = 0.1$, and various combinations of spatial and temporal H parameters denoted as H_s and H_t , respectively. For a given H_s , subsequent snapshots are more and more similar as H_t increases, implying stronger temporal dependence, while spatial events tend to cluster more and more within each snapshot as H_s increases, denoting stronger spatial dependence.

Since ζ is a nonlinear transformation, the spatiotemporal binary process Y shows the desired ACF, as $\rho_{Y,st}(\mathbf{0}, \tau) = \zeta^{-1}(\mathbf{1} \cdot \zeta(\rho_{Y,t}(\tau))) = \rho_{Y,t}(\tau)$, and lag-0 cross correlation, as $\rho_{Y,st}(\mathbf{u}, 0) = \zeta^{-1}(\zeta(\rho_{Y,s}(\mathbf{u})) \cdot 1) = \rho_{Y,s}(\mathbf{u})$, while the lagged cross-correlation terms $\rho_{Y,st}(\mathbf{u}, \tau) \neq \rho_{Y,s}(\mathbf{u}) \cdot \rho_{Y,t}(\tau)$, thus meaning that the STCF of Y is not separable. Figure 5a shows an example of the separable STCF of X resulting from the product $\rho_{X,s}(\mathbf{u}) \cdot \rho_{X,t}(\tau)$ of spatial and temporal Markovian correlation functions (equation (A2)) with parameters $\rho_{1,s} = \rho_{1,t} = 0.8$, while Figure 5b illustrates the actual STCF (from equation (26)) of a binary process Y (with $p = 0.05$) resulting from BetaBitST by using the same (marginal) spatial and temporal Markovian correlation functions. Even though both STCFs share the same (marginal) spatial and temporal correlation functions, the mixed terms (corresponding to $\mathbf{u} \neq \mathbf{0}$ and $\tau \neq 0$) are affected by the nonlinear backward transformation ζ^{-1} , which slightly strengthens the lagged cross-correlation terms. However, this discrepancy is not a shortcoming for our purposes, as we are not interested in the separability of the STCF of Y but in knowing its exact form to compute the correlation matrices in equations (19) and (22). Note that the only condition required by BetaBitST is that the STCF is positive definite in the range of space and time lags m and n of interest.

To confirm the correctness of equation (26), we simulated a binary sample spanning 10,000 time steps and 100 equally spaced locations and the corresponding empirical STCF (Figure 5c). Leaving aside fluctuations due to finite size effects, the empirical STCF matches the expected theoretical STCF in Figure 5b. Figure 6 provides some examples of binary random fields simulated by BetaBitST over a grid 20×30 and 20 time steps, with $p = 0.1$ and fGn spatial and temporal dependence with nine combinations of parameters H_s and H_t ranging in $\{0.6, 0.75, 0.9\}$. For each value of H_s , the increasing value of H_t yields increasing persistence

Figure 7. Each panel shows the empirical distribution function of the number of record-breaking values computed on binary time series with fractional Gaussian noise correlation structure and rate of occurrence $p_j = 1/j$ simulated by BetaBit, along with the corresponding BB_e distribution and the PB distribution corresponding to i/id conditions. Each panel refers to a specific combination of Hurst parameter H and sample size n . cdf = cumulative distribution function.

of the field structure across subsequent time steps. On the other hand, the scattering of the random fields decreases as H_s increases. For high values of both H_s and H_t , the random fields exhibit clustered spatial patterns (reflecting spatial persistence across adjacent grid boxes) preserving their structure in time due to temporal dependence.

BetaBitST enables the validation of BB and BB_e as suitable distributions of Z under ni/id and ni/nid conditions, respectively, as well as the evaluation of the correction factor ϕ in equation (22) for spatiotemporal processes spanning n time steps over m locations. Figure 7 (S3) compares ecdfs of Z and BB_e cdfs corresponding to binary temporal processes with different sizes n and values of H_t (ρ_t), and $p_j = 1/j$. Empirical and theoretical models show an excellent agreement for all combinations of parameters with $\phi = 0.75$. This value of ϕ results from extensive Monte Carlo simulation. Specifically, we considered several combinations of values of $H_t \in (0.5, 1)$ ($\rho_t \in (0, 1)$), H_s (ρ_s), $n \in [10, 100]$, and $m \in [10, 100]$. For each combination, we simulated 5,000 random fields with $p_j = 1/j$ and computed the number of RB events for each sample. The resulting

Figure 8. Each panel shows the empirical distribution function of the number of peak-over-threshold events computed on sequences of binary random fields with fractional Gaussian noise spatial and temporal correlation structures and rate of occurrence p simulated by BetaBitST. The corresponding BB distribution and the B distribution corresponding to i/iid conditions are also shown. Each panel refers to a specific combination of Hurst parameters H_t and H_s for the spatial and temporal dependence structures, rate of occurrence p , sample size n , and number of locations $m = 50$. cdf = cumulative distribution function.

5,000 values of Z are used to build the ecdf of Z and to estimate ρ_{BB_e} . After checking for substantial independence of ρ_{BB_e} estimates, $\hat{\rho}_{BB_e}$, from BetaBitST parameters, we computed the ϕ value minimizing the sum of squared relative errors $\sum (1 - \phi \cdot \rho_{BB_e} / \hat{\rho}_{BB_e})^2$. We stress that these results holds for $p_j = 1/j$. Other cases, such as those listed in the last two columns of Table 1, must to be addressed by case-specific Monte Carlo simulations. Figure 8 (S4) shows the perfect matching of ecdfs and BB distributions using different combinations of H_t ($\rho_{1,t}$), H_s ($\rho_{1,t}$), m , and n for spatiotemporal binary processes describing POT events with different rates of occurrence p . Figure 9 (S5) shows BB_e cdfs approximating the ecdfs of Z corresponding to ni/nid binary processes with different spatiotemporal dependence structures and time-varying $p_j = 1/j$ (with $\phi = 0.75$).

All cases highlight that the BB and BB_e distributions provide an exact or very accurate approximation of the distribution of Z under different assumptions concerning (in)dependence and (non)stationarity. These results also stress the variance-inflating effect of spatiotemporal dependence on the distribution of Z . Compared with the i/iid case, we have higher probabilities of observing a number of events smaller and greater than the theoretical average over n time steps and m locations. This inflated variability corresponds to greater uncertainty that should be taken into account when drawing conclusions about the surprising nature of POT and RB events or their clustering in space and time. These aspects are further discussed in the subsequent analysis of real world data.

7. Analysis of Temperature and Precipitation Data

7.1. Temperature and Precipitation Data Sets

As mentioned in section 1, previous studies attempted to study the rate of occurrence of RB events accounting for spatial and/or temporal correlation. However, they did not introduce a general framework allowing simulation and accurate description under fully specified spatiotemporal dependence structures, including short- and long-range dependence. In this section, we show the impact of spatiotemporal dependence on the rate of RB and POT events by studying temperature and precipitation data across CONUS and reanalyzing the Mauna Loa temperature data previously studied by Newman et al. (2010).

Figure 9. Each panel shows the empirical distribution function of the number of record-breaking values computed on sequences of binary random fields with fractional Gaussian noise spatial and temporal correlation structures and rate of occurrence $p_j = 1/j$ simulated by BetaBitST. The corresponding BB_e distribution is also shown. Each panel refers to a specific combination of Hurst parameters H_s and H_t for the spatial and temporal dependence structures, sample size n , and number of locations $m = 50$. cdf = cumulative distribution function.

The CONUS data set consists of monthly precipitation, and minimum and maximum temperature anomalies (R , $T_{\min}$, and $T_{\max}$, respectively), computed with respect to 1901–2000 average, from January 1895 to December 2017 (123 years) over the 344 divisions of the CONUS. This database was obtained from area-weighted averages of grid point estimates resulting from station data gridded via climatologically aided interpolation (Karl & Koss, 1984; Vose et al., 2014). The divisional scale yields sufficiently refined spatially smoothed results that allow the recognition of physically coherent hydroclimatological patterns (e.g., McCabe & Wolock, 2002; Wolock & McCabe, 1999). Data are provided by the National Oceanic and Atmospheric Administration (NOAA) through the U.S. Climate Divisional Database (nClimDiv data set; <http://www.ncdc.noaa.gov/cag/time-series/us>). Data were preliminarily deseasonalized by subtracting the calendar-month average and dividing by the calendar-month standard deviation. This procedure removes the effect of seasonal fluctuation of the first two marginal moments and allows for the study of the underlying homogeneous process, thus avoiding seasonal stratification and enabling better recognition of long-range (low-frequency) fluctuations.

The data set used by Newman et al. (2010) comprises 30 years of maximum and minimum daily temperatures at the NOAA Mauna Loa Observatory on the Big Island, Hawaii, recorded from 1 January 1977 to 31 December 2006. These time series are extracted from high-quality hourly temperature data measured at 2 m above ground level (Malamud et al., 2011). Data are freely available from NOAA Earth System Research Laboratory Global Monitoring Division (<https://www.esrl.noaa.gov/gmd/dv/data/>). For the sake of comparison, data were preprocessed by removing 29 February of leap years and infilling missing values by the same methodology used by Newman et al. (2010): “if one to three successive days were missing, the values for

Figure 10. Cross-correlation matrices of the residuals of fractional Gaussian noise models fitted on the 344 divisional time series of $T_{\max}$ (panel a), $T_{\min}$ (panel b), and AR models fitted on precipitation R (panel c).

the adjacent days were averaged. If more days were missing, the data for the adjacent years for that day were averaged."

7.2. RB Analysis of CONUS Divisional Temperature and Precipitation

Our analysis focuses on the occurrence of RB events in monthly $T_{\max}$, $T_{\min}$, and R both globally at the CONUS scale and locally at divisional scale. The research question is: is the occurrence of observed RB events consistent with a spatially and temporally correlated random process? To answer this question, we studied the cumulative average number of records occurred across the 344 divisions of the CONUS, following the rationale of the analyses reported by Vogel et al. (2001) and Newman et al. (2010). The empirical values are compared with those expected from a stochastic process reproducing spatial and temporal correlations of precipitation and temperature measurements. In particular, we simulated 1,000 standard Gaussian random fields of size $m \times n$ equal $344 \times (12 \cdot 123)$ with separable spatiotemporal correlation in the same spirit of BetaBitST (see also Serinaldi & Kilsby, 2017, for more details on this modeling approach). The temporal dependence structure of divisional $T_{\max}$ and $T_{\min}$ was modeled by fGn processes whose parameter H was estimated by the *least square based on variance* method (Tyalis & Koutsoyiannis, 2011), while the ACF of divisional R by AR models of order up to 3, which were selected by an automatic selection procedure based on the Akaike information criterion (Akaike, 1974). The choice of using fGn for $T_{\max}$ and $T_{\min}$ and AR for R is based on preliminary exploratory analysis of empirical ACFs and diagnostic plots of variance versus aggregation scale. Since this analysis recognized a persistent behavior in temperature data and short-range dependence in rainfall data, which is in agreement with the literature (e.g., Kantelhardt et al., 2006; Franzke, 2012), fGn and AR models provide reasonable and parsimonious descriptions of time series properties. To reproduce the spatial correlation, the divisional time series models were fed with temporally independent but spatially correlated innovations sampled for a multivariate Gaussian distribution with correlation matrix equal to the near positive definite version of the empirical correlation matrix (Higham, 2002) computed on the residuals of the divisional temporal models. It should be noted that we do not need to reproduce the exact marginal distributions of precipitation and temperature as the RB occurrences are rank-based statistics not affected by the absolute values of the observations. Figures 10 and 11 show the cross-correlation matrices for all variables and the spatial distribution of

Figure 11. Spatial patterns of Hurst parameter H for $T_{\max}$ (panel a) and $T_{\min}$ (panel b).

Figure 12. Each panel shows the empirical cumulative average number of RB events across the conterminous United States, where the cumulative sum is taken over time, and the average across the 344 division at each time step. Empirical estimates on the observed data are complemented with the theoretical expectations and 95% CIs under independence and 95% CIs under spatial and temporal dependence. The latter CIs are obtained by Monte Carlo simulation described in the text. Results refer to record-breaking high $T_{\max}$ and $T_{\min}$ (panels a and b), and record-breaking low and high precipitation (panels c and d). CI = confidence interval; RB = record-breaking.

H parameter for $T_{\max}$ and $T_{\min}$. It is worth noting that $T_{\max}$ and $T_{\min}$ exhibit a spatial correlation stronger than R (as expected), while the H values show coherent spatial patterns.

Using the simulated Gaussian fields, we computed the total (cumulative) number of records z_{ji} occurred at each division $i \in \{1, \dots, 344\}$ until the time step $j \in \{1, \dots, 12 \cdot 123\}$, and then the average $\bar{z}_j = \frac{1}{344} \sum_{i=1}^{344} z_{ji}$. We considered the RB high values for $T_{\max}$ and $T_{\min}$, since warming is usually of major interest in hydroclimatic research, and both RB high and RB low values for R . Figure 12 shows the $\bar{z}_j$ values estimated on the observed temperature and precipitation, along with the expected average number of records under independence, and the 95% pointwise CIs corresponding to independence and spatiotemporal dependence. Even though there is substantial discrepancy between the estimated and theoretical values of $\bar{z}_j$ for $T_{\max}$ and $T_{\min}$, the estimated $\bar{z}_j$ is still within the CIs when spatiotemporal dependence is accounted for. It should be noted the substantial difference between the width of CIs under independence and spatiotemporal dependence. Vogel et al. (2001) already discussed the effect of spatial correlation on the uncertainty of the rate of occurrences of annual maximum daily flow records; however, our results concerning temperature and precipitation emphasize how large can actually be the difference in cases where spatial and temporal dependence is not negligible, and it is therefore properly incorporated in the analysis. In this respect, we also note that focusing on extreme values, such as annual maximum daily flow, can lead to strong underestimation of the dependence characterizing the underlying process and therefore to too narrow CIs and different conclusions (see Serinaldi et al., 2018, for further discussion). For R , the estimated values of $\bar{z}_j$ are closer to the theoretical expectation but still outside the i/id CIs for RB high values. Even though R shows spatial and temporal correlations weaker than that of $T_{\max}$ and $T_{\min}$, also in this case, CIs under spatiotemporal dependence are much wider than those corresponding to independence, making observed patterns compatible with a stationary random process.

Figure 13. Maps show the year in which the empirical cumulative number of RB events in each division cross and fall outside the limits of the 95% Monte Carlo confidence intervals under spatiotemporal dependence. Results refer to record-breaking high $T_{\max}$ and $T_{\min}$ (panels a and b) and record-breaking low and high precipitation (panels c and d). RB = record-breaking.

Figures 12a and 12b also show that the estimated curves tend to depart from the theoretical ones at the beginning of the time series, thus indicating clustering of records in the first years of observation. To further investigate this aspect, we identified the divisions where the estimated z_{ji} fall outside the divisional 95% CIs, and the year in which the z_{ji} curves cross the CI limits. Figure 13 shows that the crossing points tend to occur in the first years in most of cases with few exceptions, where crossing occurs after 1950. As expected, crossing points are more clustered in space for $T_{\max}$ and $T_{\min}$ than for R because of the stronger spatial correlation.

Finally, we used the BB_e to answer the following question: given the information available until 1980, what is the probability of the total number of records observed until 2017? In other words, we attempt to understand if the occurrence of the most recent records is consistent with a spatially and temporally correlated random process. To provide an answer, ρ_{BB_e} is estimated on data from 1895 to 1979, and BB_e is used to compute the probability of the number of records observed until 2017. Figure 14 shows that these probabilities are generally below the the upper limit of the 95% prediction interval. In agreement with Figure 12, probabilities are closer to the upper limit for $T_{\max}$ and $T_{\min}$ than for R . For $T_{\max}$ and $T_{\min}$, limit exceedances tend to cluster in space because of spatial correlation, reflecting the common behavior of groups of neighboring divisions. Moreover, when we integrate spatial correlation out, that is, we look at the average behavior over the entire area, the overall number of records does not exceed the confidence limits (as shown in Figures 12a and 12b). Therefore, even though $T_{\max}$ and $T_{\min}$ exhibit a larger number of local exceedances compared with R , this is consistent with and can be described by spatial dependence. It is worth noting that we do not say that the number of temperature RB events are not higher than expected (as this is obvious from Figures 12a and 12b) but that the departures started at the beginning of the observation period (i.e., the first decades of the nineteenth century), and they are still consistent with stationary correlated random processes.

7.3. POT Analysis of CONUS Divisional Temperature and Precipitation

In this section, we complement the RB analysis of $T_{\max}$, $T_{\min}$, and R across CONUS divisions with a POT analysis focusing on the effect of spatial correlation on the occurrence of events with frequency exceeding a given percentage threshold. In more detail, we want to know how many simultaneous POT events we can expect to observe across the CONUS at each time step (month) and their distribution. Recalling that the data set comprises 123 years of (deseasonalized) monthly values of $T_{\max}$, $T_{\min}$, and R across 344 divisions, each value x_{ji} occurred at each division $i \in \{1, \dots, 344\}$ until the time step $j \in \{1, \dots, 12 \cdot 123\}$ is transformed into y_{ji} , such that $y_{ji} = 0$ if $x_{ji} \leq x_{i,p}$ and $y_{ji} = 1$ otherwise, where $x_{i,p}$ is the empirical quantile of X with exceedance

Figure 14. Maps show the spatial pattern of the nonexceedance probability of the total number of RB events observed from 1895 to 2017 in each division computed by the BB_e model with parameters estimated on data from 1895 to 1980. Divisions where the probability exceeds the value of 0.975, that is, the upper limit of the 95% prediction interval are highlighted by bold black boundaries. Results refer to record-breaking high $T_{\max}$ and $T_{\min}$ (panels a and b), and record-breaking low and high precipitation (panels c and d). RB = record-breaking.

probability p for the division i . The resulting binary sequences are therefore summed up yielding $z_j = \sum_{i=1}^{344} y_{ji}$, that is, the number of simultaneous POT events over the CONUS divisions at each time step. This sequence is used to build the ecdf of Z , which can be compared with B and BB distributions. We recall that the B model describes the number of successes over 344 trials for an experiment with success rate equal to p under i/id assumption, while BB model accounts for spatial correlation via the parameter ρ_{BB} . ρ_{BB} is estimated by equation (18) where the correlation value $\rho_{Y,ik}$ of a pair of binary processes (Y_i, Y_k) is computed from the correlation $\rho_{X,ik}$ of the parent processes (X_i, X_k) by the relationship (Emrich & Piedmonte, 1991)

$$\rho_{Y,ik} = \frac{\Phi_2(\Phi^{-1}(p), \Phi^{-1}(p); \rho_{X,ik}) - p^2}{p(1-p)} \approx \zeta^{-1}(\rho_{X,ik}), \quad (27)$$

where Φ_2 denotes the bivariate Gaussian cdf of (X_i, X_k) . It should be noted that a bivariate Gaussian cdf is a good approximation for monthly $T_{\max}$, $T_{\min}$, and R , since the marginal distributions of these variables are approximately Gaussian or bell shaped and weakly skewed. Figure 15 shows that the BB distributions fit the ecdfs of all variables for the thresholds $p \in \{0.01, 0.05\}$. Accounting for spatial correlation enables the modeling of discrepancies between ecdfs and the B distributions without introducing further demanding assumptions on nonstationarity, for instance. In this respect, is worth noting that the spatial correlation yields not only a larger number of simultaneous POT events over a given area but also a higher probability to observe no events over the same area. For example, this probability is close to 0 under i/id according to B model and is $\approx 40\%$ (80%) for the observed $T_{\max}$ and $T_{\min}$ POT events with $p = 0.05$ (0.01). This means that the alternation of periods with widespread POT events and no events should not be considered as exceptional but the rule under the ni/id assumption, which in turn is sufficient to properly describe this behavior.

7.4. RB Analysis of Mauna Loa Temperature

In this section, we replicate one of the analyses on RB events reported by Newman et al. (2010). Recalling that the data span 30 years from 1977 to 2006, this analysis considers the time series of 30 maximum and 30 minimum temperatures for each calendar day, resulting in 365 time series of size 30 (e.g., the first time series is the sequence of the 30 temperature values recorded on 1 January from 1977 to 2006). For each time series, we computed the cumulative numbers of RB temperatures, $z_{\max,ji}$ and $z_{\min,ji}$, with $j \in \{1, \dots, 30\}$ and $i \in \{1, \dots, 365\}$. For each year j , the 365 values are averaged obtaining $\bar{z}_{\max,j} = \frac{1}{365} \sum_{i=1}^{365} z_{\max,ji}$ and $\bar{z}_{\min,j} = \frac{1}{365} \sum_{i=1}^{365} z_{\min,ji}$ as functions of year j from 1977 to 2006.

Figure 15. Each panel shows empirical cdfs of the number of simultaneous POT events recorded across the 344 divisions of the conterminous United States at each time step (month) for percentage thresholds $p = \mathbb{P}[X > x] \in \{0.01, 0.05\}$. Theoretical B and BB cdfs are also shown. cdf = cumulative distribution function; POT = peak-over-threshold.

Since the sampling distribution of the number of RB values is approximately Gaussian with mean and variance, respectively, given by equations (8) and (9) under i/id hypothesis, Newman et al. (2010) used such a distribution to assess the agreement of the estimated $\bar{z}_{\max,j}$ and $\bar{z}_{\min,j}$ with theoretical results, accounting for sampling uncertainty. As $\bar{z}_{\max,j}$ and $\bar{z}_{\min,j}$ averages are taken over 365 values, their sampling standard deviation is expected to be $\sqrt{365}$ times smaller than that of $z_{\min,j}$, under i/id hypothesis. Under these assumptions, Newman et al. (2010) concluded that the observed values of $\bar{z}_{\max,j}$ and $\bar{z}_{\min,j}$ are not consistent with an i/id random process.

However, some questions arise: Are data really independent? Is the i/id hypothesis defensible/credible? If not, what is its effect on results? As far as temporal dependence is concerned, inference on time series of size 30 is generally speculative, as much longer time series are required to obtain reliable estimates. For example, clustering of extreme events in time is a typical characteristic of temporally correlated stochastic processes. This property can result in observed sequences of low and high summary statistics (e.g., block minima, block maxima, and POT values) that appear, however, approximately uncorrelated because these statistics do not provide enough information to assess the actual dependence of the underlying process. Lack of apparent dependence can lead to interpret low and high regimes as lack of stationarity, while they can be explained by the underlying dependence, which is however concealed by the sampling procedure (Serinaldi & Kilsby, 2016b; Serinaldi et al., 2018). In this respect, focusing on monthly, seasonal, and annual values, as done in the literature mentioned above, and using resampling methods that preserve only approximately a fraction of the actual correlation can invalidate preliminary analysis and modeling efforts. This further explains why we used deseasonalized time series instead of seasonally stratified data in the analysis reported in section 7.2. Based on our analysis of CONUS data and previous studies on temporal dependence in temperature records (e.g.,

Figure 16. Cross-correlation matrices of size 365×365 showing the pairwise correlation values of 30-year temperature time series recorded in each pair of calendar days at Mauna Loa from 1977 to 2006.

Figure 17. Each panel shows the empirical cumulative average number of RB events for $T_{\max}$ and $T_{\min}$ time series recorded at Mauna Loa, where the cumulative sum is taken over the 30 calendar years from 1977 to 2006 and the average across the 365 calendar days. Empirical estimates on the observed data are complemented with the theoretical expectations and 95% CIs under independence and 95% CIs preserving dependence between the values recorded at different calendar days. The latter CIs are obtained by Monte Carlo simulation described in the text. Results refer to record-breaking low $T_{\max}$ (panel a) and high $T_{\min}$ (panel b). RB = record-breaking; CI = confidence interval.

Koutsoyiannis, 2003; Maraun et al., 2004; Percival et al., 2001; Stephenson et al., 2000; Vyushin & Kushner, 2009), we cannot exclude that Mauna Loa temperatures are temporally dependent. However, we conservatively assume that the *i/id* hypothesis holds true in time for the year-to-year calendar-day values.

Nonetheless, we cannot overlook the cross correlation of the 365 time series used to compute $\bar{z}_{\max,j}$ and $\bar{z}_{\min,j}$. In fact, the cross-correlation matrices in Figure 16 show that the pairwise correlation is not negligible, especially for neighboring calendar days (values around the secondary diagonal). In other words, the 30 observations for a given calendar day (e.g., 1 January) are correlated with the 30 values of neighboring calendar days (e.g., 31 December and 2 January).

To assess the impact of the cross correlation, we simulated samples of size 30 from a 365-dimensional standard multivariate Gaussian distribution with correlation matrix equal to the near positive definite version of the empirical correlation matrices. Figure 17 confirms the conclusions of Newman et al. (2010) under *i/id* hypothesis, showing that the values of $\bar{z}_{\max,j}$ and $\bar{z}_{\min,j}$ fall outside the *i/id* CIs. However, conclusions dramatically change under cross correlation, which substantially inflates the variability of the expected values and the width of CIs. Accounting for dependence, the observed $\bar{z}_{\max,j}$ and $\bar{z}_{\min,j}$ are no longer incompatible with fluctuations of a stationary process. We do not claim that dependence *caused* the observed behavior of RB occurrences but only that such a behavior is consistent with a stationary process different from *i/id*. Therefore, the existence of an alternative plausible description of the observed patterns of $\bar{z}_{\max,j}$ and $\bar{z}_{\min,j}$ calls into question conclusions and possible attributions previously reported in the literature mentioned above and indicates that a more careful investigation and use of statistical analyses are required.

8. Discussion and Conclusions

8.1. Discussion

In this study, we have studied analytical distributions of the number of POT and RB events over specified temporal windows starting from the beginning of the period of record. While this is useful for exploratory analysis and general modeling, for design purposes it can be of interest to know the distribution of Z over a future design life given the information collected in the period of record. This requires the development of conditional distributions. Since POT and RB events are characterized by several properties, such as magnitude, differences of magnitude between consecutive observations, interarrival times, and number of events, different conditional distributions can be defined. For instance, defining the distribution of Z conditioned on the magnitude of the last recorded event requires information on the distributions of the absolute values of the process X . To our knowledge, in these cases analytical solutions are not straightforward even for relatively simple cases of *i/id* X . Of course, the problem can be addressed by case-specific Monte Carlo

simulations. Examples of this approach concerning conditional interarrival times of POT events and conditional block maxima under *ni/id* assumption are given by Eichner et al. (2011, and references therein).

Even though this work describes results for POT and RB processes referring to the case *ni/id* and specific dependence structures (i.e., fGn and Markov), we do not endorse any particular model or assumption. As explicitly stated at the end of the case studies, our message is that the empirical observations can be consistent with *ni/id* as well as *i/nid* models reported in the previous literature, where the latter models are justified by inconsistency between observations and *i/id* or alternative assumptions that do not properly account for correlation. Since rejection of *i/id* cannot lead to acceptance of *i/nid* because other options are available, and we cannot discriminate among alternative descriptions, attribution procedures based on the logic of mutually exclusive frameworks (in the spirit of common interpretation of null hypothesis statistical testing) might be inconclusive and should be taken with great care, if not discarded (e.g., Serinaldi et al., 2018, and references therein). In other words, when we observe discrepancies between observed and theoretical patterns, we should bear in mind that the theoretical values refer to a specific (not necessarily well-devised) reference hypothesis (e.g., *i/id*) and that statements such as “*a* causes *b*” can make sense only if we can exclude any other reasonable explanation. Until we can say “*a* can be compatible with *b*, *c*, *d*,...,” we can only choose one of the possible options, bearing in mind parsimony, and other (common sense) criteria.

For a correct interpretation of our empirical results, it is worth recalling the two main step of the statistical inference well summarized by von Storch and Zwiers (2003):

1. “A statistical model is adopted that supposedly describes both the stochastic characteristics of the observed process and the properties of the method of observation. It is important to be aware of the models implicit in the chosen statistical method and the constraints those models necessarily impose on the extraction and interpretation of information.”
2. “The observations are analyzed in the context of the adopted statistical model.”

According to these principles, it is clear, for instance, that our estimates of H or ρ_1 are based on the assumption that the underlying processes are fGn or Markov. More generally, every inferential procedure relies on an assumed model (e.g., *i/id*, *ni/id*, linear, and nonlinear). It is also obvious that the inferential results are generally biased if the assumed theoretical process is misspecified. For example, we can obtain a full range of H and ρ_1 values for time series drawn from correlated processes different from fGn and Markov. These estimates make sense only under the assumption that fGn and Markov provide a reasonable description of the observed process. In this context, our emphasis on the *ni/id* framework versus *i/nid* relies on the fact that the latter requires much stronger assumptions, involves additional model uncertainties, and makes many results of classical statistical inference questionable, or even meaningless/undefined (e.g., Serinaldi & Kilsby, 2015; Serinaldi et al., 2018). Therefore, the *ni/id* should always be considered as a parsimonious alternative to *i/nid* to describe possible discrepancies from *i/id*, and when observations agree with both *ni/id* and *i/nid* descriptions, we can only conclude that the evidence is not enough for a clear choice and unambiguous attribution.

The remarks above are related to the problem of recognizing and identifying dependence, nonstationarity, and other properties from data. For example, it is well known that very large sample sizes are required to reliably identify long-range persistence and estimate H from observed time series (e.g., Koutsoyiannis & Montanari, 2007). Inference becomes even more difficult when dealing with binary processes Y describing POT and RB occurrences, as the dichotomization removes the information of the absolute values of the parent process X . To mitigate this effect, we did not apply monthly or seasonal stratification in the analysis of CONUS temperature and precipitation, and spatiotemporal dependence structures were estimated on the parent process X and then suitably deflated by equation (27) to obtain those of Y . Moreover, it is known that correlation can influence the statistics of extreme values yielding, for instance, spatiotemporal clustering of RB, POT, or block maxima (e.g., Bogachev & Bunde, 2012; Eichner et al., 2011; Serinaldi & Kilsby, 2016b). On the other hand, it is generally difficult to retrieve the underlying correlation structures only from extreme events, which often appear to be approximately independent because of downsampling effects of data selection and consequent removal of nonextreme data providing information on correlation (e.g., Serinaldi et al., 2018). In these cases, *ni/id* can be confused with *i/nid* and *vice versa*, as the data are not enough to draw conclusions, and further information does need to be collected. As shown in the case studies, in these circumstances, endorsing a particular model or assumption and their consequences can be questionable, while a more balanced approach implies the use of multiple schemes, bearing in mind their properties, shortcomings, and suitability for the problem at hand.

8.2. Concluding Remarks

The occurrence of RB events and, more generally, extreme values of hydroclimatic variables such as temperature, precipitation, or stream flow attract attention from the media (King, 2017), especially in light of the possible connection with anthropogenic influence. This explains the recent rapid evolution of methods devised to assess possible systematic changes of rate of extreme events. Among these methods, we can mention the so-called *event attribution*, which “compares the occurrence probability of an event in the present, factual climate with its probability in a hypothetical, counterfactual climate without human-induced climate change” (Hauser et al., 2017). Detection and attribution heavily rely on climate model simulations and are therefore subject to their limitations (King, 2017). In particular, the choice of model, counterfactual climate, and boundary conditions can lead to contradicting conclusions (Hauser et al., 2017). In this context, uncertainty and the nature of statistical methods used to summarize information play a key role. Uncertainty can refer to multiple aspects, going to sample uncertainty (i.e., limited size of sequences of reliable systematic observations, which call the assessment of the so-called *natural variability* into question), to model uncertainty (including model structure and underlying assumptions).

In this study we have therefore attempted to propose a set of tools enabling an analysis of POT and RB occurrences overcoming too restrictive *i/id* hypothesis and accounting more consistently for spatial and temporal dependence. In particular, we showed that the BB and BB_e distributions allow for readily assessing the probability of the number of POT and RB events, respectively, over n time steps and m locations, with minimal information and avoiding extensive simulation. On the other hand, the BetaBitST generator enables the simulation of binary random processes with general spatial and temporal correlation structures, resulting in binary sequences/fields characterized by spatial and temporal clustering, which is often attributed too superficially to external factors.

Our validation of BB and BB_e showed that spatiotemporal dependence has very limited or no impact on the expected number of POT and RB events but strongly affects the shape of their distribution, resulting in so-called overdispersion, that is, a substantial increase of the probability to observe a number of events larger (smaller) than the mean compared with the *i/id* reference. Therefore, for a given location, we can expect, for instance, periods showing clusters of POT and RB events followed by periods with no events. Similarly, for a fixed date, POT and RB events will tend to occur in nearby locations forming hot spot and event-free areas.

As shown in the case studies, the use of different benchmark assumptions involving spatiotemporal dependence lead to conclusions completely different from those resulting from analyses based on *i/id* hypothesis or underrepresented dependence, thus making the departure from theoretical averages less surprising than expected. Referring to 2017 Houston flood caused by Hurricane Harvey, Montz (2017) posed the question “While the flooding may have been unprecedented, was it unexpected?” As mentioned above, the answer to this question depends on many factors and “uncertainty abounds and the generally available information ... does not address the uncertainty well, if at all” (Montz, 2017). However, a key aspect is how we use the available information and what statistical models and underlying assumptions we use to quantify uncertainty. Inefficient use of the data can prevent the retrieval of valuable information, while the choice of models based on inappropriate or too restrictive hypotheses (e.g., *i/id*) can lead to incorrect conclusions. Following the rationale of Serinaldi and Kilsby (2015) and Serinaldi et al. (2018), this study showed the huge impact of these assumptions and provides statistical tools to analyze POT and RB occurrences under reasonable and more challenging reference hypotheses.

Appendix A: Dependence Structures

The performance of the BB distribution is tested by generating binary sequences with ACF corresponding to two widely used stationary processes, that is, the fGn and the Markov process. The former, also known as Hurst-Kolmogorov process (e.g., Koutsoyiannis, 2010), is characterized by the following ACF:

$$\rho_X(\tau) = \frac{1}{2}(|\tau + 1|^{2H} - 2|\tau|^{2H} + |\tau - 1|^{2H}), \quad (A1)$$

which exhibits a power law decay $\rho_X(\tau) \propto |\tau|^{2H-2}$. For $0.5 < H < 1$ the process is positively correlated and exhibits long-range dependence, while it reduces to white noise for $H = 0.5$. As a second example, we consider a process with short-range Markovian dependence, which is characterized by exponentially decaying ACF of the form

$$\rho_X(\tau) = \exp(-\gamma|\tau|) = \rho_1^{|\tau|}, \quad (A2)$$

where $1/\gamma$ is the correlation radius and $\rho_1 = \exp(-\gamma)$ is the lag-one autocorrelation coefficient.

Acknowledgments

The authors were supported by the Engineering and Physical Sciences Research Council (EPSRC) grant EP/K013513/1 "Flood MEMORY: Multi-Event Modelling Of Risk & recovery", and Willis Research Network. CONUS data are provided by the National Oceanic and Atmospheric Administration (NOAA) through the U.S. Climate Divisional Database (nClimDiv data set; <http://www.ncdc.noaa.gov/cag/timeseries/us>), while Mauna Loa data are freely available from NOAA Earth System Research Laboratory Global Monitoring Division (ESRL-GMD; <https://www.esrl.noaa.gov/gmd/dv/data/>). The authors wish to thank the eponymous reviewer Simon-Michael Papalexiou (University of California, Irvine) and two anonymous reviewers for their insightful remarks and constructive criticisms. In particular, we are grateful to one of the reviewers for introducing us to the work of George L.S. Shackle and some seminal papers of Nicholas C. Matalas. The analyses were performed in R (R Development Core Team, 2018).

References

- Ahn, H., & Chen, J. J. (1995). Generation of over-dispersed and under-dispersed binomial variates. *Journal of Computational and Graphical Statistics*, 4(1), 55–64.
- Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, 19(6), 716–723.
- Anderson, A., & Kostinski, A. (2010). Reversible record breaking and variability: Temperature distributions across the globe. *Journal of Applied Meteorology and Climatology*, 49(8), 1681–1691.
- Arnold, B. C., Balakrishnan, N., & Nagaraja, H. N. (2011). *Records*, Wiley Series in Probability and Statistics. New York: Wiley.
- Bassett, G. W. (1992). Breaking recent global temperature records. *Climatic Change*, 21(3), 303–315.
- Bayazit, M., & Önöz, B. (2007). To prewhiten or not to prewhiten in trend analysis? *Hydrological Sciences Journal*, 52(4), 611–624.
- Benestad, R. E. (2003). How often can we expect a record event? *Climate Research*, 25(1), 3–13.
- Benestad, R. E. (2004). Record-values, nonstationarity tests and extreme value distributions. *Global and Planetary Change*, 44(1), 11–26.
- Bogachev, M. I., & Bunde, A. (2012). Universality in the precipitation and river runoff. *EPL (Europhysics Letters)*, 97(4), 48011.
- Chandler, K. N. (1952). The distribution and frequency of record values. *Journal of the Royal Statistical Society. Series B (Methodological)*, 14(2), 220–228.
- Colucci, R. R., Giorgi, F., & Torma, C. (2017). Unprecedented heat wave in December 2015 and potential for winter glacier ablation in the eastern Alps. *Scientific Reports*, 7, 7090.
- Coumou, D., & Rahmstorf, S. (2012). A decade of weather extremes. *Nature Climate Change*, 2(7), 491–496.
- David, F. N., & Barton, D. E. (1962). *Combinatorial chance*. New York: Hafner Publishing Company.
- Davies, R. B., & Harte, D. S. (1987). Tests for Hurst effect. *Biometrika*, 74(1), 95–101.
- Douglas, E., Vogel, R., & Kroll, C. (2000). Trends in floods and low flows in the United States: Impact of spatial correlation. *Journal of Hydrology*, 240(12), 90–105.
- Edwards, A. W. F. (1960). The meaning of binomial distribution. *Nature*, 186, 1074.
- Eichner, J. F., Kantelhardt, J. W., Bunde, A., & Havlin, S. (2011). The statistics of return intervals, maxima, and centennial events under the influence of long-term correlations. In J. Kropp & H.-J. Schellnhuber (Eds.), *Extremis* (pp. 2–43). Berlin, Heidelberg: Springer.
- Emrich, L. J., & Piedmonte, M. R. (1991). A method for generating high-dimensional multivariate binary variates. *The American Statistician*, 45(4), 302–304.
- Fiering, M., & Kindler, J. (1984). Surprise in water? Resource design. *International Journal of Water Resources Development*, 2(4), 1–10.
- Finkel, J. M., & Katz, J. I. (2017). Changing US extreme temperature statistics. *International Journal of Climatology*, 37(13), 4749–4755.
- Franzke, C. (2012). Nonlinear trends, long-range dependence, and climate noise properties of surface temperature. *Journal of Climate*, 25(12), 4172–4183.
- Genton, M. G. (2007). Separable approximations of space-time covariance matrices. *Environmetrics*, 18(7), 681–695.
- Genton, M. G., & Kleiber, W. (2015). Cross-covariance functions for multivariate geostatistics. *Statistical Science*, 30(2), 147–163.
- Glick, N. (1978). Breaking records and breaking boards. *The American Mathematical Monthly*, 85(1), 2–26.
- Gneiting, T., Genton, M. G., & Guttorp, P. (2006). Geostatistical space-time models, stationarity, separability, and full symmetry. In B. Finkenstadt, L. Held, & V. Isham (Eds.), *Statistical methods for spatio-temporal systems* (pp. 151–175). New York: Chapman & Hall/CRC.
- Griffiths, D. A. (1973). Maximum likelihood estimation for the Beta-Binomial distribution and an application to the household distribution of the total number of cases of a disease. *Biometrics*, 29(4), 637–648.
- Hamed, K. H. (2008). Trend detection in hydrologic data: The Mann-Kendall trend test under the scaling hypothesis. *Journal of Hydrology*, 349(3–4), 350–363.
- Hamed, K. H. (2009a). Enhancing the effectiveness of prewhitening in trend analysis of hydrologic data. *Journal of Hydrology*, 368(1–4), 143–155.
- Hamed, K. H. (2009b). Effect of persistence on the significance of Kendall's tau as a measure of correlation between natural time series. *The European Physical Journal Special Topics*, 174(1), 65–79.
- Hamed, K. H. (2011). The distribution of Kendall's tau for testing the significance of cross-correlation in persistent data. *Hydrological Sciences Journal*, 56(5), 841–853.
- Hamed, K. H., & Rao, A. R. (1998). A modified Mann-Kendall trend test for autocorrelated data. *Journal of Hydrology*, 204(1–4), 182–196.
- Hasselmann, K. (1997). Multi-pattern fingerprint method for detection and attribution of climate change. *Climate Dynamics*, 13(9), 601–611.
- Hauser, M., Gudmundsson, L., Orth, R., Jézéquel, A., Hausteine, K., Vautard, R., et al. (2017). Methods and model dependency of extreme event attribution: The 2015 European drought. *Earth's Future*, 5(10), 1034–1043.
- Higham, N. J. (2002). Computing the nearest correlation matrix—A problem from finance. *IMA Journal of Numerical Analysis*, 22(3), 329–343.
- Hisakado, M., Kitsukawa, K., & Mori, S. (2006). Correlated binomial models and correlation structures. *Journal of Physics A: Mathematical and General*, 39(50), 15365.
- Hong, Y. (2013). On computing the distribution function for the Poisson binomial distribution. *Computational Statistics and Data Analysis*, 59(Supplement C), 41–51.
- Hughes, G., & Madden, L. (1993). Using the Beta-Binomial distribution to describe aggregated patterns of disease incidence. *Phytopathology*, 83(7), 759–763.
- Interagency Committee on Water Data (1982). Guidelines for determining flood flow frequency: Bulletin no. 17B, Hydrology Committee, U.S. Water Resources Council, Washington, DC.
- Itti, L., & Baldi, P. (2009). Bayesian surprise attracts human attention. *Vision Research*, 49(10), 1295–1306.
- Jones, R. H. (1975). Estimating the variance of time averages. *Journal of Applied Meteorology (1962-1982)*, 14(2), 159–163.
- Kantelhardt, J. W., Koscielny-Bunde, E., Rybski, D., Braun, P., Bunde, A., & Havlin, S. (2006). Long-term persistence and multifractality of precipitation and river runoff records. *Journal of Geophysical Research*, 111, D01106. <https://doi.org/10.1029/2005JD005881>
- Karl, T. R., & Koss, W. J. (1984). *Regional and national monthly, seasonal, and annual temperature weighted by area, 1895–1983*, Historical Climatology Series, 4-3. Asheville, N.C.: National Climatic Data Center.
- Katz, R. W. (1988a). Statistical procedures for making inferences about climate variability. *Journal of Climate*, 1(11), 1057–1064.
- Katz, R. W. (1988b). Use of cross correlations in the search for teleconnections. *Journal of Climatology*, 8(3), 241–253.
- Katz, R. W., & Brown, B. G. (1991). The problem of multiplicity in research on teleconnections. *International Journal of Climatology*, 11(5), 505–513.
- King, A. D. (2017). Attributing changing rates of temperature record breaking to anthropogenic influences. *Earth's Future*, 5(11), 1156–1168.
- Kjeldsen, T. R., Jones, D. A., & Bayliss, A. C. (2008). *Improving the FEH statistical procedures for flood frequency estimation*. Wallingford, UK: Environment Agency.

- Kjeldsen, T. R., & Prosdocimi, I. (2018). Assessing the element of surprise of record-breaking flood events. *Journal of Flood Risk Management*, 11, S541–S553.
- Koutsoyiannis, D. (2003). Climate change, the Hurst phenomenon, and hydrological statistics. *Hydrological Sciences Journal*, 48(1), 3–24.
- Koutsoyiannis, D. (2010). HESS opinions “A random walk on water”. *Hydrology and Earth System Sciences*, 14(3), 585–601.
- Koutsoyiannis, D., & Montanari, A. (2007). Statistical analysis of hydroclimatic time series: Uncertainty and insights. *Water Resources Research*, 43, W05429. <https://doi.org/10.1029/2006WR005592>
- Kulkarni, A., & von Storch, H. (1995). Monte Carlo experiments on the effect of serial correlation on the Mann–Kendall test of trend. *Meteorologische Zeitschrift*, 4(2), 82–85.
- Kullback, S., & Leibler, R. A. (1951). On information and sufficiency. *The Annals of Mathematical Statistics*, 22(1), 79–86.
- Malamud, B. D., Turcotte, D. L., & Grimmer, C. S. B. (2011). Temperature trends at the Mauna Loa observatory, Hawaii. *Climate of the Past*, 7(3), 975–983.
- Maraun, D., Rust, H. W., & Timmer, J. (2004). Tempting long-memory—On the interpretation of DFA results. *Nonlinear Processes in Geophysics*, 11(4), 495–503.
- Matalas, N. C. (1997). Stochastic hydrology in the context of climate change. *Climatic Change*, 37(1), 89–101.
- Matalas, N. C. (2003). Vulnerability of water systems to acts of terrorism and acts of nature risk-based decision making in water resources X. In *American Society of Civil Engineers, 10th United Engineering Foundation Conference November 3–8, 2002* (pp. 82–90). Santa Barbara, CA.
- Matalas, N. C., & Langbein, W. B. (1962). Information content of the mean. *Journal of Geophysical Research*, 67(9), 3441–3448.
- McCabe, G. J., & Wolock, D. M. (2002). Trends and temperature sensitivity of moisture conditions in the conterminous United States. *Climate Research*, 20(1), 19–29.
- Meehl, G. A., Tebaldi, C., Walton, G., Easterling, D., & McDaniel, L. (2009). Relative increase of record high maximum temperatures compared to record low minimum temperatures in the U.S. *Geophysical Research Letters*, 36, L23701. <https://doi.org/doi:10.1029/2009GL040736>
- Merz, B., Vorogushyn, S., Lall, U., Viglione, A., & Blöschl, G. (2015). Charting unknown waters—On the role of surprise in flood risk assessment and management. *Water Resources Research*, 51, 6399–6416. <http://doi.org/10.1002/2015WR017464>
- Mitchell, J. F. B., Karoly, D. J., Hegerl, G. C., Zwiers, F. W., Allen, M. R., & Marengo, J. (2001). *Detection of climate change and attribution of causes* (pp. 393–425). Cambridge, UK: Cambridge University Press.
- Montz, B. (2017). When “unprecedented” becomes the norm. *Journal of Flood Risk Management*, 10(4), 413–414.
- Moran, P. A. P. (1968). *An introduction to probability theory Oxford science publications*. New York: Oxford Science Publications, Oxford University Press.
- Nagaraja, H. N., Choudhary, P. K., & Matalas, N. C. (2002). Number of records in a bivariate sample with application to Missouri River flood data. *Methodology And Computing In Applied Probability*, 4(4), 377–391.
- Newman, W. I., Malamud, B. D., & Turcotte, D. L. (2010). Statistical properties of record-breaking temperatures. *Physical Review E*, 82, 066111.
- Nicola, V. F., & Goyal, A. (1990). Modeling of correlated failures and community error recovery in multiversion software. *IEEE Transactions on Software Engineering*, 16(3), 350–359.
- Obeyskera, J., & Salas, J. D. (2016). Frequency of recurrent extremes under nonstationarity. *Journal of Hydrologic Engineering*, 21(5), 04016005.
- Papalexiou, S. M. (2018). Unified theory for stochastic modelling of hydroclimatic processes: Preserving marginal distributions, correlation structures, and intermittency. *Advances in Water Resources*, 115, 234–252.
- Percival, D. B., Overland, J. E., & Mofjeld, H. O. (2001). Interpretation of North Pacific variability as a short- and long-memory process. *Journal of Climate*, 14(24), 4545–4559.
- Podgórski, K., & Wegener, J. (2012). Velocities of a spatial-temporal stochastic field with embedded dynamics. *Environmetrics*, 23(3), 238–252. <https://doi.org/10.1002/env.2135>
- Poisson, S. D. (1837). *Recherches sur la probabilité des jugements en matière criminelle et en matière civile: précédées des règles générales du calcul des probabilités*. Paris: Bachelier.
- R Development Core Team (2018). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing.
- Rahmstorf, S., & Coumou, D. (2011). Increase of extreme events in a warming world. *Proceedings of the National Academy of Sciences*, 108(44), 17,905–17,909.
- Renard, B., & Lang, M. (2007). Use of a Gaussian copula for multivariate extreme value analysis: Some case studies in hydrology. *Advances in Water Resources*, 30(4), 897–912.
- Serinaldi, F., & Kilsby, C. G. (2015). Stationarity is undead: Uncertainty dominates the distribution of extremes. *Advances in Water Resources*, 77, 17–36.
- Serinaldi, F., & Kilsby, C. G. (2016a). The importance of prewhitening in change point analysis under persistence. *Stochastic Environmental Research and Risk Assessment*, 30(2), 763–777.
- Serinaldi, F., & Kilsby, C. G. (2016b). Understanding persistence to avoid underestimation of collective flood risk. *Water*, 8(4), 152.
- Serinaldi, F., & Kilsby, C. G. (2017). A blueprint for full collective flood risk estimation: Demonstration for European river flooding. *Risk Analysis*, 37(10), 1958–1976.
- Serinaldi, F., Kilsby, C. G., & Lombardo, F. (2018). Untenable nonstationarity: An assessment of the fitness for purpose of trend tests in hydrology. *Advances in Water Resources*, 111, 132–155.
- Serinaldi, F., & Lombardo, F. (2017). Betabit: A fast generator of autocorrelated binary processes for geophysical research. *EPL (Europhysics Letters)*, 118(3), 30007.
- Shackle, G. L. S. (1942). A theory of investment-decisions. *Oxford Economic Papers*, 6, 77–94.
- Shackle, G. L. (1949). *Expectation in economics*. London, UK: Cambridge at The University Press.
- Skellam, J. G. (1948). A probability distribution derived from the binomial distribution by regarding the probability of success as variable between the sets of trials. *Journal of the Royal Statistical Society. Series B (Methodological)*, 10(2), 257–261.
- Solow, A. R., & Smith, W. (2005). How surprising is a new record? *The American Statistician*, 59(2), 153–155.
- Stephenson, D. B., Pavan, V., & Bojariu, R. (2000). Is the North Atlantic Oscillation a random walk? *International Journal of Climatology*, 20(1), 1–18.
- Tejada, A., & den Dekker, A. J. (2011). The role of Poisson’s binomial distribution in the analysis of TEM images. *Ultramicroscopy*, 111(11), 1553–1556.
- Thompson, V., Dunstone, N. J., Scaife, A. A., Smith, D. M., Slingo, J. M., Brown, S., & Belcher, S. E. (2017). High risk of unprecedented UK rainfall in the current climate. *Nature Communications*, 8, 107.
- Tripathi, R. C., Gupta, R. C., & Gurland, J. (1994). Estimation of parameters in the beta binomial model. *Annals of the Institute of Statistical Mathematics*, 46(2), 317–331.

- Tsai, C. A., Hsueh, H. M., & Chen, J. J. (2003). Estimation of false discovery rates in multiple testing: Application to gene microarray data. *Biometrics*, 59(4), 1071–1081.
- Tyralis, H., & Koutsoyiannis, D. (2011). Simultaneous estimation of the parameters of the Hurst–Kolmogorov stochastic process. *Stochastic Environmental Research and Risk Assessment*, 25(1), 21–33.
- Vogel, R. M., Zafirakou-Koulouris, A., & Matalas, N. C. (2001). Frequency of record-breaking floods in the united states. *Water Resources Research*, 37(6), 1723–1731.
- von Storch, H., & Zwiers, F. W. (2003). *Statistical analysis in climate research* Cambridge University Press. New York: US.
- Vose, R. S., Applequist, S., Squires, M., Durre, I., Menne, M. J., Williams-Jr, C. N., et al. (2014). Improved historical temperature and precipitation time series for U.S. climate divisions. *Journal of Applied Meteorology and Climatology*, 53(5), 1232–1251.
- Vyushin, D. I., & Kushner, P. J. (2009). Power-law and long-memory characteristics of the atmospheric general circulation. *Journal of Climate*, 22(11), 2890–2904.
- Wang, Y. H. (1993). On the number of successes in independent trials. *Statistica Sinica*, 3(2), 295–312.
- Weaver, W. (1948). Probability, rarity, interest, and surprise. *The Scientific Monthly*, 67(6), 390–392.
- Wolock, D. M., & McCabe, G. J. (1999). Explaining spatial variability in mean annual runoff in the conterminous United States. *Climate Research*, 11(2), 149–159.
- Yue, S., & Wang, C. Y. (2002). Regional streamflow trend detection with consideration of both temporal and spatial correlation. *International Journal of Climatology*, 22(8), 933–946.
- Yue, S., & Wang, C. Y. (2004). The Mann–Kendall test modified by effective sample size to detect trend in serially correlated hydrological series. *Water Resources Management*, 18(3), 201–218.
- Zaigraev, A., & Kaniovski, S. (2013). A note on the probability of at least k successes in n correlated binary trials. *Operations Research Letters*, 41(1), 116–120.