Testing tests before testing data: an untold tale of compound events and binary dependence

Francesco Serinaldi\textsuperscript{1,2} \and Federico Lombardo\textsuperscript{3} \and Chris G. Kilsby\textsuperscript{1,2}

Accepted: 26 January 2022 / Published online: 5 April 2022
© The Author(s) 2022

Abstract
In any statistical investigation, we deal with the applications of probability theory to real problems, and the conclusions are inferences based on observations. To obtain plausible inferences, statistical analysis requires careful understanding of the underlying probabilistic model, which constrains the extraction and interpretation of information from observational data, and must be preliminarily checked under controlled conditions. However, these very first principles of statistical analysis are often neglected in favor of superficial and automatic application of increasingly available ready-to-use software, which might result in misleading conclusions, confusing the effect of model constraints with meaningful properties of the process of interest. To illustrate the consequences of this approach, we consider the emerging research area of so-called ‘compound events’, defined as a combination of multiple drivers and/or hazards that contribute to hydro-climatological risk. In particular, we perform an independent validation analysis of a statistical testing procedure applied to binary series describing the joint occurrence of hydro-climatological events or extreme values, which is supposed to be superior to classical analysis based on Pearson correlation coefficient. To this aim, we suggest a theoretically grounded model relying on Pearson correlation coefficient and marginal rates of occurrence, which enables accurate reproduction of the observed joint behavior of binary series, and offers a sound simulation tool useful for informing risk assessment procedures. Our discussion on compound events highlights the dangers of renaming known topics, using imprecise definitions and overlooking or misusing existing statistical methods. On the other hand, our model-based approach reveals that consistent statistical analyses should rely on informed stochastic modeling in order to avoid the proposal of flawed methods, and the untimely dismissal of well-devised theories.

Keywords Compound events \and Binary processes \and Binary dependence test \and Rainfall occurrence \and Serial dependence \and Cross correlation

“The problems caused by the indiscriminate use of recipes are compounded when obscure sophisticated techniques are used. It is fashionable to surprise the community with miraculous new techniques, even though the statistical model implicit in the method is often not understood”

[von Storch and Zwiers, 2003, p. 97]

\textsuperscript{1} School of Engineering, Newcastle University, Newcastle Upon Tyne NE1 7RU, UK
\textsuperscript{2} Willis Research Network, 51 Lime St, London EC3M 7DQ, UK
\textsuperscript{3} Corpo Nazionale dei Vigili del Fuoco, Ministero dell’Interno, Piazza del Viminale, 1, Rome 00184, Italy

1 Introduction
Similarly to every human activity and aspect of life, science is characterized by some cyclical dynamics, making topics and theories cyclically fashionable or neglected (e.g., Landauer 1997). Sometimes existing topics are revitalized by changing their name and/or slightly modifying definitions and context. Often, this generates confusion and leads to dismiss or forget part of the literature and results concerning those topics or strictly related subjects. Here, we consider the case of the so-called ‘compound events’ and their probabilistic and statistical treatment as an example of this phenomenon. Broadly speaking, compound events are sets of single events that are gathered and analyzed altogether for some reason. For example, analyzing the combination of multiple factors such as direct surface runoff, river discharge, storm surge, and waves can
improve flood risk assessment in coastal regions (e.g. Camus et al. 2021).

In the context of the study of compound events, De Michele et al. (2020) proposed a procedure to assess the mutual dependence of pairs of binary time series describing the rainfall occurrence (wet/dry conditions) observed at two rain gauge stations. This testing procedure aims to overcome the possible limitations of classical analysis based on Pearson correlation coefficient, and it attempts to provide a more detailed description of the dependence between pairs of binary time series. Hereinafter, we define such a test as JOP (joint occurrence probability) test.

In the spirit of seeking theoretical consistency in statistical analyses (Lombardo et al. 2012, 2014, 2017, 2019; Serinaldi and Kilsby 2016a; Serinaldi et al. 2015, 2020a, b), this study discusses the results of a so-called neutral (independent) validation/falsification analysis (see e.g. Popper 1959; Boulesteix et al. 2018, and references therein) to independently check theoretical properties and interpretation of the JOP test. We also discuss the supposed limitations of classical Pearson correlation analysis for binary processes that led to introduce the JOP procedure. Our inquiry relies on theoretical arguments, Monte Carlo simulations, and re-analysis of rainfall data previously studied by De Michele et al. (2020). We also propose a procedure to simulate binary pairs with negative correlation.

This study is organized as follows. In Sect. 2, we discuss the effects of introducing new research topics based on ambiguous definitions, with an emphasis on ‘compound events’. Section 3 summarizes the JOP algorithm, introducing concepts and symbols required to provide a self-contained discussion. We also present and motivate our alternative model-based approach relying on Pearson correlation. In Sect. 4, we set up and perform Monte Carlo simulations to assess the effective significance and power of the JOP procedure accounting for the interaction of serial and cross correlation as well as cyclical patterns of the rate of occurrence (seasonality). A subset of rainfall data studied by De Michele et al. (2020) is re-analyzed in Sect. 5, and JOP outcomes are compared with those corresponding to simulated time series with given properties mimicking rainfall data. In Sect. 5.2, we discuss theoretical aspects enabling the correct interpretation of JOP outcomes in terms of Pearson cross correlation and marginal rates of occurrence, which are shown to provide enough information to reproduce the behavior of rainfall co-occurrence. In Sect. 6, we discuss the lesson learned from our analysis in the broader perspective of a change of paradigm in the stochastic analysis of hydro-climatological data, moving from statistical testing to stochastic modeling. Conclusions are summarized in Sect. 7.

2 The ‘Brave New World’ of compound events

2.1 Ambiguity

The increasing interest in compound events in hydroclimatological data is partly related to relatively recent advancements of stochastic methods, which have enabled unprecedented analysis and modeling of multiple variables via convenient and easy-to-use techniques and software (e.g. Heffernan and Tawn 2004; Nelsen 2006; Salvadori et al. 2007; Davison et al. 2012; Asadi et al. 2015; Papalexiou 2018; Papalexiou and Serinaldi 2020; Papalexiou et al. 2021, and references therein). Despite such an interest, a rigorous mathematical definition of compound events is still lacking in risk analysis; the literature is fragmented, and the very different concepts of compound, interconnected, interacting, and cascading risks are often used as synonyms, thus causing redundancy and confusion (Pescaroli and Alexander 2018). Koutsoyiannis et al. (2018) emphasized that “absence of a definition entails domination of intuition over logic, dark over light, or uncritical acting over critical thinking”. Clarity and precision in the definition of new forms of risk are key issues because “the way we understand and describe risk strongly influences the way risk is analyzed and hence it may have serious implications for risk management and decision making” (Aven 2016). In this respect, Pescaroli and Alexander (2018) attempted to highlight differences and complementary aspects of compound, interconnected, interacting, and cascading risks in order to propose a holistic framework that can support focused research, avoiding redundancy. In this framework, compound risk mainly refers to the concurrence of natural events, while “institutional definitions tend to focus more narrowly on the hazard component of disaster risk”. For example, the cross-border scale effects of the eruption of the Icelandic volcano Eyjafjallajökull in April 2010 were due to the coincidence of such a geophysical event and weather conditions that caused the ash spread out over an area with a high concentration of essential transportation system nodes (Pescaroli and Alexander 2018). On the other hand, storms and related floods, or seaside earthquakes and related tsunamis are classified as interacting events as the former causes the latter in a chain of events reflecting the physical relationship of these phenomena. More generally, moving from compound risk to cascading risks in the
framework proposed by Pescaroli and Alexander (2018) means introducing additional degrees of complexity, including interactions between human, environmental, and technological systems, and therefore those among hazard, exposure, and vulnerability.

In this context, stochastic methods are the privileged approach for quantitative assessment of compound risks, such as the simultaneous occurrence of volcanic eruption and particular weather conditions. In fact, the inherent rarity and uncertainty of such events, reflecting the lack of knowledge about their possible physical relationships, prevent the use of deterministic methods. Therefore, it is easy to recognize that the analysis and modeling of compound events heavily rely on traditional stochastic methods devised for multivariate analysis of ‘joint variables’ and ‘joint events’. In particular, joint events have been the subject of the entire theoretical and applied literature dealing with multivariate statistics and spatio-temporal dependent processes.

However, as recognized by Pescaroli and Alexander (2018), the lack of a precise and clear definition of compound events generated confusion, which had a two-fold consequence: (i) a growing fashionable literature on compound events that seems to deal with a new topic, but addresses technical problems that do not differ from those already discussed in the existing literature on multivariate random processes; in fact, such a literature commonly uses the same stochastic approaches (often the simplest version) previously developed for joint events, such as multivariate distributions, generalized linear models, Markov chains, and their variants (e.g., Hao et al. 2018, for a review); and (ii) the apparent novelty of the topic seems to prevent the acknowledgment of its similarities to other topics. This seems to cause a sort of damnatio memoriae (i.e., condemnation of memory) affecting existing theories and methods, which are therefore neglected when approaching problems that were actually solved in the past.

2.2 Inconsistency

Before discussing in depth the case of joint occurrence processes, which is the specific subject of this study, we report some preliminary examples of the possible consequences of introducing seemingly new topics without precise definitions, preventing the construction of a consistent theory.

As mentioned above, compound events remain without an agreed mathematical definition; however, they are often introduced using the definition proposed by Leonard et al. (2014): “A compound event is an extreme impact that depends on multiple statistically independent variables or events”. This definition may be problematic. In fact, while interacting factors are reasonably related either functionally (if the physical dynamics are clear) or statistically (if the physical dynamics are complex and not well known), for compound events often it is not even possible to establish whether their forcing variables are correlated or not. As compound events are extremes according to the foregoing definition, they occur rarely and can result from a set of statistically independent and non dangerous forcing events that occur together with a frequency given by the product of their marginal probabilities. Therefore, the statistical dependence used in the definition above is questionable, as we can have compound events (extreme impact) with statistically independent forcing factors.

Moreover, statistically significant dependence/correlation does not mean or imply physical dependence, which is the dependence of interest. Indeed, the former depends on the sample size, and almost every test assigns statistical significance to physically negligible differences for very large samples (Nicholls 2001; Serinaldi et al. 2018). As discussed below, talking about ‘statistically significant dependence/correlation’ might be meaningless. In other words, we can have ‘an extreme impact that depends on multiple statistically independent variables or events’. We also note that sometimes dependence is incorrectly considered necessary for the occurrence of joint or compound events, as independence is confused with mutual exclusiveness. Ironically, mutually exclusive events are indeed dependent on each other.

Imprecise definitions generate paradoxes as well. For example, the joint occurrence of relatively low precipitation and high temperature characterizing the 2014 California drought was studied by AghaKouchak et al. (2014) as a compound event although those variables are statistically independent, as shown by Serinaldi (2016). In this case, either the 2014 California drought is not a compound event (as precipitation and temperature are statistically independent, contradicting the definition of Leonard et al. (2014)) or the 2014 California drought is a compound event according to a different but not specified definition.

Another example of the consequences of inconsistency when performing data analysis could be the following statement: “We note that for cascading hazards with independent drivers, we cannot simply multiply the probabilities of the hazard drivers. The reason is that the events are physically related, primarily through the actual impact (e.g., debris flow as a result of rain over burned area); however, the relationship cannot be simply described using the commonly used statistical metrics. More research should focus on developing methods for characterization of such events.” (AghaKouchak et al. 2020). This statement concerns the deadliest debris flow event in California’s history, which occurred in 2018 in Montecito as a result of heavy rainfall over a burned area previously affected by wildfires in 2017. In this context, ‘cascading events’ refer
to events delayed in time. While wildfires and heavy rainfall look like independent events (as there is no evidence to establish a link between them), obviously the extreme debris flow results from earlier joint occurrence of both forcing events. Then, why does not the product probability of the hazard drivers (heavy rainfall and wildfire) describe the probability of the target event (debris flow)? Because such a product is not the correct probabilistic model describing the problem at hand.

Let us derive suitable probabilistic models. As the problem involves three hazards, a probabilistic model must involve three random variables as well, that is, $R$ (rainfall), $W$ (wild fire), and $Q$ (debris flow). Assuming that we are only interested in the occurrence probability, we can assume that the state space of each variable is binary, that is, each hazard occurs or does not. Let $R_i(R_0), W_j(W_0)$, and $Q_1(Q_0)$ denote the event ($E_i$) corresponding to the occurrence (non-occurrence) of $R$, $W$, and $Q$, respectively. For a generic binary variable, the events $E_0$ and $E_1$ are mutually exclusive and exhaustive, i.e. $\mathbb{P}[E_0 \cap E_1] = 0$ and $\mathbb{P}[E_0] + \mathbb{P}[E_1] = 1$. Recalling the total probability theorem and the conditional probability theorem (e.g., VanMarcke, 1977; AghaKouchak et al., 2020), the occurrence probability of $Q$ (i.e. the probability of the event $Q_1$) is

$$\mathbb{P}[Q_1] = \sum_{i=0}^{1} \sum_{j=0}^{1} \mathbb{P}[Q_1, R_i, W_j]$$

$$= \sum_{i=0}^{1} \sum_{j=0}^{1} \mathbb{P}[Q_1|R_i, W_j] \cdot \mathbb{P}[R_i, W_j]$$

$$= \sum_{i=0}^{1} \sum_{j=0}^{1} \mathbb{P}[Q_1|R_i, W_j] \cdot \mathbb{P}[R_i] \cdot \mathbb{P}[W_j].$$

This model highlights that the occurrence probability of debris flow $Q$ is not (and cannot be) the simple product of the occurrence probabilities of the forcing factors $R$ and $W$. The conditional probability $\mathbb{P}[Q_1|R_i, W_j]$ actually accounts for the (unknown) physical relationship mentioned by AghaKouchak et al. (2020). If $R, W, Q$ are assumed to be random variables of continuous type (e.g. rainfall intensity, wild fire extension, debris flow height), the model in Eq. 1 assumes the following form giving the distribution function of $Q$:

$$F(Q \leq q) = \int_{0}^{q} \int_{\Omega_R} \int_{\Omega_W} f(q, r, w) dq dr dw$$

$$= \int_{0}^{q} \int_{\Omega_R} \int_{\Omega_W} f(q|r, w) \cdot f(r, w) dq dr dw$$

$$= \int_{0}^{q} \int_{\Omega_R} \int_{\Omega_W} f(q|r, w) \cdot f(r) \cdot f(w) dq dr dw$$

(2)

where $F$ denotes the distribution function, $f$ is the probability density function, and $\Omega_R$ and $\Omega_W$ are the state spaces of $R$ and $W$, respectively.

Even though the probabilistic formalization in Eqs. 1 and 2 does not require any definition of compound events, the framework proposed by Pescaroli and Alexander (2018) provides a useful interpretation. Here, rainfall and wildfire are compound events as they concur to set the conditions for debris flow. They are not necessarily statistically dependent, thus not fulfilling the definition by Leonard et al. (2014); they are delayed in time and not interacting, as there is no evidence for physical laws linking them. On the other hand, rainfall and debris flow are interacting as the former triggers the latter, while wildfire and debris flows are indirectly related, as the former is not the direct trigger of debris flow, but it is the direct cause of the antecedent soil conditions that emphasized rainfall effects. In this framework, $\mathbb{P}[R_i, W_j]$ describes the probability of a compound event (i.e. occurrence of rainfall and wildfire), while $\mathbb{P}[Q_1|R_i, W_j]$ provides a probabilistic representation of the causal chains, when we do not use (or do not know) a physical model to describe the interacting dynamics of rainfall, soil conditions, and debris flow. We also note the profound difference of the definitions of “cascading events/risks” given by AghaKouchak et al. (2020) and Pescaroli and Alexander (2018).

This example highlights that the use of inconsistent and ambiguous definitions could generate not only redundancy and confusion, but also incorrect risk assessment; therefore, vague definitions hinder correct formalization of the problem under investigation. On the other hand, correct use of basic principles of probabilistic reasoning can prevent the untimely requirement of new statistical metrics and methods. As further discussed below, before dismissing seemingly old-fashioned but theoretically solid concepts, one should be sure that their possible inconsistencies are not related to their misuse.
2.3 Rainfall occurrence: new compound or classical joint?

Occurrence processes have been widely studied in many disciplines from different perspectives for a long time, as they are related to the occurrence or non-occurrence of some natural or technological phenomena of interest, such as rainfall (Koutsoyiannis 2006; Ng and Panu 2010; Marchand 2012; Lombardo et al. 2017; Olson and Kleiber 2017; Serinaldi and Lombardo 2017b), floods (Bogachev and Bunde 2012; Serinaldi and Kilsby 2016b), droughts (Coats et al. 2016), wind storms (Mailier et al. 2006), tornados (Valente and Laurini 2020), earthquakes (Anagnos and Kiremidjian 1988; Ogata 1999), disease incidence (Serinaldi and Lombardo 2017b), floods (Bogachev et al. 2018; Papalexiou et al. 2021, among others).

Focusing on rainfall, occurrence (rain/no rain or wet/dry periods) recorded at a rain gauge network can be described as a binary spatio-temporal stochastic process defined over a fixed set of spatio-temporal coordinates, that is, rain gauge locations, and time steps identified by gauges’ temporal resolution along the recording period. As these processes are multidimensional, their description, modeling, and simulation include spatio-temporal dependence as a distinctive characteristic. Commonly, the study and modeling of dependence involve only second-order statistical properties, such as serial and cross (spatio-temporal) correlation functions (see e.g., Cox and Isham 1980; Emrich and Piedmonte 1991; Demirtas 2006; Macke et al. 2009; Diggle 2013; Serinaldi and Lombardo 2017a, b; Papalexiou 2018; Papalexiou et al. 2018; Serinaldi and Kilsby 2018; Papalexiou and Serinaldi 2020; Papalexiou et al. 2021, among others).

Based on remarks and examples in Sect. 2.2, the following research questions arise:

- Is standard analysis based on Pearson correlation actually unsuitable or insufficient to describe the joint occurrence of rainfall at location pairs? Does it have limitations? If yes, which ones?
- Is JOP better and more informative than methods based on Pearson correlation?

In the next sections, we try to answer these questions, highlighting the effects of re-framing joint occurrence processes as ambiguously defined compound events.

3 Changing perspective: from testing to modeling

3.1 JOP testing procedure

Let $X_i = \{X_{i,n}\}_{n \in \mathbb{N}}$ and $X_j = \{X_{j,n}\}_{n \in \mathbb{N}}$ be two discrete-time stochastic processes with state space $\{0, 1\}$, where $n (= 0, 1, 2, \ldots)$ denotes discrete time. Then, we define $\{x_i\}_{n \in \mathbb{N}}$ and $\{x_j\}_{n \in \mathbb{N}}$ as two sequences of binary random numbers respectively sampled from $X_i$ and $X_j$, for simplicity $\{x_i\}$ and $\{x_j\}$, taking values 1 and 0 with probabilities $\pi^X_i$ and $\pi^X_j$, where $l, k \in \{0, 1\}$, respectively (e.g., $\pi^X_0 = \mathbb{P}(X_i = 0)$ and $\pi^X_1 = \mathbb{P}(X_i = 1)$). For instance, $\{x_i\}$ and $\{x_j\}$ can be seen as daily time series describing the occurrence (‘1’) or non-occurrence (‘0’) of rainfall at two locations $i$ and $j$ (e.g. two rain gauge stations), whereby rainfall occurrence corresponds to daily precipitation amount exceeding a given threshold value (e.g. 0 mm, as for De Michele et al. (2020)).

For each $n$, we can have four possible combinations of the values assumed by the pair $(X_i, X_j)$: $(x_{i,n} = 1, x_{j,n} = 1)$ or wet-wet condition, $(x_{i,n} = 0, x_{j,n} = 0)$ or dry-dry, $(x_{i,n} = 1, x_{j,n} = 0)$ or wet-dry, $(x_{i,n} = 0, x_{j,n} = 1)$ or dry-wet. Following De Michele et al. (2020), let $\mathcal{P}_* = \mathcal{P}_{lk}$ denote the probability of the outcome $\bullet (l, k \in \{0, 1\})$, that is

\[
\mathcal{P}_{11} = \mathbb{P}(X_i = 1, X_j = 1) \quad (3a)
\]
\[
\mathcal{P}_{00} = \mathbb{P}(X_i = 0, X_j = 0) \quad (3b)
\]
\[
\mathcal{P}_{10} = \mathbb{P}(X_i = 1, X_j = 0) \quad (3c)
\]
\[
\mathcal{P}_{01} = \mathbb{P}(X_i = 0, X_j = 1) \quad (3d)
\]

If the processes $X_i$ and $X_j$ are independent of one another, the probability $\mathcal{P}_{lk}$ is the product of the marginal probabilities of occurrence, $\mathcal{P}_{lk}^{'} := \pi^X_i \pi^X_j$. The JOP testing procedure aims at assessing whether $\mathbb{P}(X_i = l, X_j = k) \neq \mathbb{P}(X_i = l) \mathbb{P}(X_j = k)$, meaning that the locations $i$ and $j$ tend to exhibit the joint condition $(l, k)$, e.g. wet-wet, with frequency lower or higher than that expected under the assumption of independence. To support our discussion and for the sake of completeness, we recall the JOP algorithm, which consists of the following six steps to be applied to each pair of time series $\{x_i\}$ and $\{x_j\}$, and each year of data (De Michele et al. 2020):

1. Calculate the sample estimate, $\hat{\mathcal{P}}_*$, of $\mathcal{P}_*$ for each of the four combinations (i.e., 00, 11, 01, and 10);
2. compare $\hat{\mathcal{P}}_*$ with the confidence interval (CI) under the hypothesis of independence (i.e., binomial
distribution). The limits of such a CI are given by Laplace approximation (Laplace 1812, p. 282)

\[
\begin{align*}
\left[ l_{\bar{P}_{j,2}}, u_{\bar{P}_{j,2}} \right] = & \bar{P}_j \pm z_{\alpha/2} \frac{\sqrt{ \bar{P}_j (1 - \bar{P}_j)}}{N} \\
\left[ l_{\bar{P}_{j,3}}, u_{\bar{P}_{j,3}} \right] = & \bar{P}_j \pm z_{\alpha/2} \frac{\sqrt{ \bar{P}_j (1 - \bar{P}_j)}}{N}
\end{align*}
\]

(4)

where \(z_{\alpha/2}\) is the \(\alpha/2\) quantile of the standard Gaussian distribution, \(N\) is the sample size, and \(\alpha\) is the significance level. The interval in Eq. 4 is also known as Wald interval (see e.g., Miao and Gastwirth 2004). If \(\bar{P}_j \in [l_{\bar{P}_{j,2}}, u_{\bar{P}_{j,2}}]\), then \(X_i\) and \(X_j\) are independent with respect to combination \(\bullet\).

3. otherwise, if \(\bar{P}_j \notin [l_{\bar{P}_{j,2}}, u_{\bar{P}_{j,2}}]\), then calculate the CI under hypothesis of independence via Monte Carlo simulations assuming that \(X_i\) and \(X_j\) can be modeled by two-state Markov chains, to account for the effect of possible serial dependence. If \(\bar{P}_j \in [l_{\bar{P}_{j,3}}, u_{\bar{P}_{j,3}}]\), then \(X_i\) and \(X_j\) are independent with respect to combination \(\bullet\).

4. If \(\bar{P}_j > u_{\bar{P}_{j,3}}\) (or \(\bar{P}_j < l_{\bar{P}_{j,3}}\)), one checks the behavior of the other combinations (denoted as \(\bullet\)) different from the tested \(\bullet\), as \(\sum \bar{P}_j = 1\) and the increase of one \(\bar{P}_j\) causes the decrease of one or more \(\bar{P}_o\). If there is at least a combination such that \(\bar{P}_o > u_{\bar{P}_{o,3}}\) (or \(\bar{P}_o < l_{\bar{P}_{o,3}}\)), therefore \(X_i\) and \(X_j\) are judged dependent with respect to \(\bullet\) (or \(\circ\)).

5. Conversely, one makes the same check shrinking the CI, considering a level of significance equal to \(1.5\alpha\), as the reduction of \(\bar{P}_o\) (or \(\bar{P}_j\)) could be not significant at a level \(\alpha\). If there is a combination such that \(\bar{P}_o > u_{\bar{P}_{o,1.5\alpha}}\) (or \(\bar{P}_o < l_{\bar{P}_{o,1.5\alpha}}\)), therefore \(X_i\) and \(X_j\) are judged dependent with respect to \(\bullet\) (or \(\circ\)).

6. If not, then one checks the \(p\) value of \(\bar{P}_j > u_{\bar{P}_{j,3}}\). If its \(p\) value is \(< 0.0025\), therefore \(X_i\) and \(X_j\) are considered dependent with respect to \(\bullet\), otherwise they are considered independent.

The following preliminary remarks are needed to better investigate the properties of the JOP algorithm described above:

- Similar to many other works developed under the umbrella of ‘compound events’, JOP does not refer to any true risk analysis involving hazard, exposure, vulnerability, interactions between human, environment, and technological systems, disaster escalation, and appropriate cost-benefit and/or multi-criteria analysis. Actually, JOP simply deals with the study of joint behavior of two binary random processes, that is, a classical topic of spatio-temporal statistics. Thus, referring to ‘compound’ context does not add any information to problem definition and formalization. Recognizing this fact is fundamental to review the existing literature and properly address the available theories and operational tools before proposing something new (see discussion in Sect. 2.2).

- De Michele et al. (2020) suggest using quite a small significance level \(\alpha = 0.005\) to reduce the probability of Type I errors. However, while this value is valid for steps 2 and 3 of the JOP algorithm, it does not provide the effective statistical significance of the overall procedure, as the step 4 requires multiple comparisons involving probabilities different from the target \(\bar{P}_j\) (resulting in a multiple testing procedure). Moreover, step 5 implies shrinking the CIs of the combinations \(\circ\), while step 6 resorts to \(p\) values. Therefore, the effective significance of the overall procedure is basically unknown.

- Type I and Type II errors cannot be both arbitrarily small, because the probability of the latter tends to increase as that of the former decreases, and vice versa. Therefore, statistical hypothesis tests should be applied balancing between the two types of errors or focusing on one of them. However, this is possible only if we know the effective significance, and we have insights into the power of the test, i.e., the rejection rate under the alternative hypothesis (here, some form of dependence between the two processes \(X_i\) and \(X_j\)). Then, before applying the JOP test to real-world data, we need to know its properties and performance under controlled conditions.

### 3.2 Pearson-based modeling

Investigating significance and power of the JOP testing procedure requires suitable models enabling the simulation of processes with prescribed properties, such as serial dependence, marginal (at-site) rate of occurrence, cross dependence, and seasonality. On the other hand, the availability of such models capable of reproducing the key properties of an observed process makes tests quite superfluous, especially if the tested null hypothesis \(H_0\) is rather unrealistic, such as the lack of cross dependence among rainfall time series recorded over areas affected by similar or contrasting weather patterns. Referring to Sect. 6 for a deeper discussion of inherent inconsistencies of statistical testing procedures, a model-based method does not focus on rejecting an hypothesis that is often trivial and already untenable, but on reproducing something that is theoretically sound and empirically consistent with the observations.

Therefore, our model-based approach starts from existing (prior) knowledge of key features of rainfall fields, including the existence of spatial homogeneity at given
spatio-temporal scales, and possible dual behavior (wet/dry) of different areas, related for instance to fluctuations in the difference of atmospheric pressure at sea level (such as North Atlantic Oscillation, or Southern Oscillation). Thus, we discard the hypothesis of independence, which is too restrictive and unrealistic according to (conditioned on) our prior knowledge. Therefore, according to the principle of parsimony, we adopt a modeling strategy that includes prior knowledge. Therefore, according to the principle of restrictive and unrealistic according to (conditioned on) our we discard the hypothesis of independence, which is too North Atlantic Oscillation, or Southern Oscillation). Thus, the difference of atmospheric pressure at sea level (such as dry) of different areas, related for instance to fluctuations in spatio-temporal scales, and possible dual behavior (wet/dry) of different areas, related for instance to fluctuations in the difference of atmospheric pressure at sea level (such as North Atlantic Oscillation, or Southern Oscillation). Thus, we discard the hypothesis of independence, which is too restrictive and unrealistic according to (conditioned on) our prior knowledge. Therefore, according to the principle of parsimony, we adopt a modeling strategy that includes prior knowledge. Therefore, according to the principle of restrictive and unrealistic according to (conditioned on) our we discard the hypothesis of independence, which is too

4 Monte Carlo experiments

In this section, we investigate the effective significance and power of the JOP test as well as the combined effects of serial correlation and seasonally varying rates of occurrence \( \pi_1^X \) and \( \pi_1^Y \) on the JOP outcomes, and their interpretation.

4.1 Significance of JOP test and effects of serial correlation

Firstly, we investigate the effective significance of the JOP test, i.e. the actual rejection rate when the null hypothesis, \( H_0 : \mathbb{P}_v = \mathbb{P}_w \), is true. As mentioned in Sect. 3.1, the JOP test accounts for possible serial dependence in the processes \( X_t \) and \( Y_t \) assuming that they can be modeled by two-state Markov chains. Here, we use binary processes characterized by (i) an exponentially decaying autocorrelation function (ACF) corresponding to the first-order autoregressive model, AR(1), with parameter \( \rho_1 \) (e.g. Jacobs and Lewis 1983; Jentsch and Reichmann 2019), which is similar to the ACF of Markov chains, and alternatively (ii) power-law decaying ACF corresponding to fractional Gaussian noise (fGn) with parameter \( H \) (see e.g. Lowen and Teich 2005; Gong et al. 2011, and Appendix B for ACF formulas). Complementing AR(1) with fGn allows for a more comprehensive assessment of JOP behavior for a process implying long range persistence and remarkable over-dispersion (Serinaldi and Lombardo 2017a, b, 2020; Serinaldi and Kitsby 2018). Simulations consist of generating independent time series \( \{x_t\} \) and \( \{y_t\} \) of size \( N = 365 \) with \( \rho_1 \) ranging from 0 to 0.9 by steps of 0.1, \( H \in \{0.55, 0.65, 0.75, 0.85\} \), and pairs \( (x_1^X, x_1^Y) \in \{(0.1, 0.1), (0.1, 0.3), (0.1, 0.5), (0.3, 0.3), (0.3, 0.5), (0.5, 0.5)\} \), and then applying the JOP test.

This setting mimics the analysis of daily records on an annual basis in terms of sample size. The values of \( \rho_1 \) and \( H \) cover a wide range of serial dependence scenarios, from independence to fairly strong persistence, while the values of \( x_1^X \) and \( x_1^Y \) are consistent with the possible combinations of average probabilities of occurrence of daily rainfall in various climates (e.g., Harrold et al. 2003; Robertson et al. 2004; Serinaldi 2009; Olson and Kleiber 2017). For each combination of parameters, simulation and JOP test are repeated 10000 times in order to compute the empirical rejection rate. For each simulated series, Monte Carlo CIs
are computed using 5000 auxiliary simulated series with rates of occurrence, $\rho_1$, and $H$ estimated on the parent series, and the same sample size (365).

For nominal rejection rates $\alpha = 0.5\%, 5\%$, and $10\%$, the effective rejection rates are much lower than the nominal ones, with maximum values equal to 0.04%, 0.39%, and 1.43%, respectively, for binary processes with AR(1) ACF, and 0.02%, 0.2%, and 0.48% for binary processes with fGn ACF (see Fig. 1). Moreover, rejection rates tend to decrease as $\rho_1$ and $H$ increase. The cases corresponding to binary processes with AR(1) ACF and nominal significance equal to 5%, and 10% show that the effective rejection rates decrease as the rates of occurrence $p_{X_i}$ and $p_{X_j}$ increase. This behavior is not so evident for the other cases because of the very low effective rejection rate. These results indicate that the JOP test almost never rejects the null hypothesis when it is true (that is, the test is very conservative), and it is not generally possible to control the effective rejection rate modulating the nominal one. In fact, the suggested use of the nominal $\alpha = 0.5\%$ to avoid Type I errors does not seem to be grounded on theoretical or empirical evidence, and this choice can be problematic in terms of power of the test, as discussed below.

### 4.2 Effects of seasonal variation under mutual independence

Laplace CIs and Monte Carlo CIs (relying on first-order Markov chains) in steps 2 and 3 of the JOP algorithm (Sect. 3.1) implicitly assume that the theoretical values of $p_{X_i}$ and $p_{X_j}$ (and $p_{X_0}$ and $p_{X_0}$) are constant over the period of analysis. However, the seasonal patterns of rainfall occurrence (e.g., Harrold et al. 2003; Robertson et al. 2004; Serinaldi 2009; Mehrotra et al. 2012; Olson and Kleiber 2017) make this assumption quite unrealistic for daily data over annual time windows, which is the time interval suggested by De Michele et al. (2020) for JOP analysis. Therefore, it is important to understand the effect of seasonally varying $p_{X_i}$ and $p_{X_j}$ on the outcome of the JOP procedure. To this aim, we performed a Monte Carlo experiment with the following settings: sample size $N = 365$, AR(1) ACF with $\rho_1$ ranging from 0 to 0.9 by steps of 0.1, fGn ACF with $H \in \{0.55, 0.65, 0.75, 0.85\}$, and $p_{X_i} = p_{X_j} = 0.16 \cos(\frac{2\pi}{365} n) + 0.25$, $n = 1, ..., N$, such that the rate of occurrence varies in the range [0.09, 0.41] mimicking the smooth transition between wet and dry season. Similarly to the experiment in Sect. 4.1, the JOP analysis is repeated 10000 times for each value of $\rho_1$. For

![Fig. 1](https://example.com/f1.png)
AR(1) correlation structure, Fig. 2 shows that the seasonal fluctuations of $p_{i1}$ and $p_{j1}$ yield rejection rates higher than those obtained without seasonality (Fig. 1). There is quite a clear dependence of rejection rate on $\rho_1$ for all $\alpha$ values. Seasonality tends to generate spurious correlation that inflates the rejection rate. On the other hand, increasing serial correlation tends to conceal the seasonal patterns of $p_{i1}$ and $p_{j1}$, and therefore the spurious cross correlation. In agreement with results described in Sect. 4.1, long range dependence causes a decrease of the effective rejection rate, compared to the case of AR(1) dependence structure. Long range dependence conceals the effects of seasonality, preventing the appearance of spurious correlations, and inflation of rejection rate. These results show that the type of correlation structure is an additional factor influencing the JOP outcome. Therefore, establishing a clear relationship between effective and nominal rejection rates is even more problematic due to the dependence of effective rejection rate on serial correlation type and intensity.

4.3 Power of JOP test and compound effects of serial and cross correlation

Besides significance, a complementary aspect to be investigated before applying a testing procedure is its power, i.e. its rejection rate under the assumption that the alternative hypothesis is true (here, $H_1 : P_\star \neq P_1$). In other words, the power of the test is the probability that we reject $H_0$ when false. Based on results of Sect. 4.1, we performed a Monte Carlo experiment with the following settings: sample size $N = 365$, cross correlation $\rho_c$ ranging from 0 to 0.9 by steps of 0.1, $\pi_{i1}^X = \pi_{j1}^X = 0.3$, and $\rho_1 \in \{0.1, 0.5, 0.8\}$. The analysis is repeated 1000 times for each combination of parameters. Values of $\pi_{i1}^X$ and $\pi_{j1}^X$ are chosen as averages of real-world daily rainfall rates of occurrence. On the other hand, the experiments discussed in Sect. 4.1 showed that the JOP test is very conservative independently of the rates of occurrence. As mentioned above, bivariate binary series characterized by desired serial and cross correlation are simulated by BetaBit algorithm and its extension BetaBitST (Serinaldi and Lombardo 2017a; Serinaldi and Kilsby 2018).

Results of power analysis are reported in Fig. 3. As the JOP test is very conservative, its power is rather low when
in this case, Monte Carlo experiments are set up as follows: sample size $N = 365$, cross correlation $\rho_c$ ranging from 0 to 0.9 by steps of 0.1, and $\pi_1^X = \pi_1^Y = 0.16 \cos \left( \frac{2\pi}{365} n \right) + 0.25$, $n = 1, \ldots, N$, and $\rho_1 \in \{0.1, 0.5, 0.8\}$. The analysis is repeated 1000 times for each parameter combination.

Comparing Fig. 4 to 3, we can note that the synchronous cyclic rate of occurrence inflates the power, as it combines with cross correlation, strengthening the frequency of simultaneous occurrence of the pairs ’00’ and ‘11’. However, a closer look reveals that we have an opposite effect when $\rho_1 = 0.8$. This behavior is more evident for $\alpha = 0.5\%$, and moderate/high values of cross correlation, and denotes quite a complex interaction of serial correlation and cyclic rate of occurrence. Of course, a deeper analysis including seasonal patterns of $\pi_1^X$ and $\pi_1^Y$ with different amplitude, period, and phase, and different values of $\rho_1$ for $X_i$ and $X_j$ would be feasible but non-exhaustive in any case, as the possible combinations of parameters and conditions is almost infinite. Moreover, additional investigation is also not very informative as the JOP analysis suffers from shortcomings more serious than the effects of confounding factors on significance and power. These drawbacks are discussed below in detail.

## 5 Rainfall data re-analysis

For illustration, De Michele et al. (2020) reported graphical representation of JOP outcomes, and corresponding interpretation, for some pairs of European daily rainfall time series. We re-analyzed 10 of those pairs accounting for the results of the foregoing Monte Carlo experiments. In particular, we performed the JOP analysis on an annual basis, and using monthly stratification. Similar to seasonal analysis, here, ‘monthly stratification’ means concatenating all the observations of each specific calendar month (January, February, etc.) to create twelve time series of size equal to $n_m \times n_y$, where $n_m$ is the number of days of a specific month (28/29, 30, or 31) and $n_y$ is the number of years of record. Annual analysis allows direct comparison and independent check of the original results, while monthly stratification filters the concealing effects of seasonality out, and guarantees (i) seasonally-filtered rates of
occurrence fulfilling (at least approximately) the required assumption of stationarity for \( p_{X1} \) and \( p_{Xj} \), and (ii) suitable sample size to reliably estimate the properties of the zero inflated rainfall process, and conduct meaningful analysis. Data come from two freely available databases: the Global Historical Climatology Network-Daily (GHCN-D; Menne et al. 2012a, b) (https://www.nci.noaa.gov/data/global-historical-climatology-network-daily/), and the European Climate Assessment & Dataset (ECA&D; Klein Tank et al. 2002) (https://www.ecad.eu). The GHCN-D data set was retrieved and handled by the R contributed package rnoaa (Chamberlain 2020).

5.1 Annual and monthly JOP analysis

Figures 5 and 6 report the JOP results for two pairs of stations: SOUID100810 (Wick, UK) and SOUID105863 (Hwardsen Bridge, UK), and SOUID105474 (Glomfjord, Norway) and SOUID110384 (Tuddal, Norway), respectively. Results for the other eight pairs are reported in Figs. S1–S8, to allow a direct comparison with results in De Michele et al. (2020). The analysis was performed using the same setting of the original work, focusing on lag-0 cross correlation with nominal significance level \( \alpha = 0.5\% \) (even if we showed that the effective significance is much lower). Figures display the values of \( \hat{P} \) along with Laplace and Monte Carlo CIs of \( P_{00} \) (i.e., under the hypothesis of independence). Laplace CIs are the Bernoulli CIs in the original work. As for De Michele et al. (2020), the annual analysis was performed on years with no missing values; however, our data set has generally more years with complete records than that considered in the original work. This can be due to the use of a different versions of the GHCN-D and ECA&D databases, which are regularly updated.

De Michele et al. (2020) report detailed results for the two pairs SOUID100810-SOUID105863, and SOUID105474-SOUID110384, as examples of how JOP outcomes should be interpreted. Namely, the annual analysis in Fig. 5 was reported as an example of ‘dry–dry type of dependence’ as there are several years with \( \hat{P}_{00} \) falling outside the CIs, and the JOP test suggests rejection of \( H_0 : P_{00} = P_{I00} \). Yet, \( \hat{P}_{11} \) values are systematically greater than those expected under independence, but generally within CIs, i.e., \( \hat{P}_{11} \) values are close to but lower than the upper limit of Laplace and Monte Carlo CIs. Conversely, \( \hat{P}_{10} \) and \( \hat{P}_{01} \) values are systematically lower than those expected under independence. On the other hand, results of annual analysis for the pair SOUID105474-SOUID110384 (Fig. 6) were reported as a typical example of ‘dry–wet type of dependence’, as there are several years with \( \hat{P}_{01} \) values falling outside the CIs, and the JOP test suggests rejection of \( H_0 : P_{01} = P_{I01} \). Yet, compared with the first example, in this case, \( \hat{P}_{10} \) values are systematically greater than those expected under independence, while \( \hat{P}_{11} \) and \( \hat{P}_{00} \) values are systematically lower.

Based on the above examples and the others shown in Figs. S1–S8, De Michele et al. (2020) concluded that the JOP procedure can discriminate different types of dependence corresponding to wet–wet, dry–dry, wet–dry, dry–wet conditions or their possible combinations. However, filtering the seasonal effect out, results of the monthly analysis tell a different story. For almost all pairs of stations and months, probabilities \( \hat{P}_{11}, \hat{P}_{00}, \hat{P}_{10}, \text{ and } \hat{P}_{01} \) fall outside the CIs, and the JOP test rejects \( H_0 : P_{11} = P_{00} \), with few exceptions corresponding to the pairs (SOUID100039 (Salzburg, Austria), SPE00120602 (Valladolid, Spain)), and (SOUID110002 (Nes Pa Hedmark, Norway), SOUID109403 (Tromso, Norway)). Therefore, filtering
Fig. 5 Results of annual and monthly JOP analysis for the pair of rainfall series (SOUID100810, SOUID105863). Lines represent confidence intervals at the 99.5% confidence level.

Fig. 6 Results of annual and monthly JOP analysis for the pair of rainfall series (SOUID105474, SOUID110384). Lines represent confidence intervals at the 99.5% confidence level.
seasonal effects out, the monthly analysis highlights three main points that are discussed in detail in Sect. 5.2:

1. The JOP procedure cannot discriminate different types of dependence. When the probability of one combination, e.g. $P_{11}$ falls outside the CI, and it does not pass the JOP test, the same should hold for the other three combinations as well.

2. For pairs of binary variables, we can have only two points (1) and (2) listed at the end of Sect. 5.1. Let vide the theoretical explanation of the empirical results in 5.2 Revising the interpretation of JOP output: a matter of total probability theorem. In order to have a valid joint distribution consistent with marginal probabilities, it must be $P_{10} < P_{01}$, $P_{01} > P_{10}$, $P_{00} > P_{11}$, and $P_{01} > P_{00}$ according to the constraints in Eqs. 8, 9, and 10, respectively. Recalling the definition of Pearson correlation coefficient of two binary variables, which is known as Pearson $\phi$ coefficient or mean square contingency (Cramér 1946, pp. 281-283)

$$\rho_c \equiv \phi = \frac{P_{11}P_{00} - P_{10}P_{01}}{\sqrt{\pi_1^X\pi_0^X \pi_1^X \pi_0^X}} = \frac{P_{11}P_{00} - P_{10}P_{01}}{\sqrt{\pi_1^X(1 - \pi_1^X)\pi_1^X(1 - \pi_1^X)}}$$

it follows that such a case corresponds to positive cross correlation, $\rho_c > 0$, which is the scenario in Equation 5.

3. The JOP analysis should be performed on series (or sub-samples) fulfilling, at least approximately, the implicit assumption of stationary $\pi_1^X$ and $\pi_1^X$, in order to avoid the concealing effects of their seasonal fluctuations, which in turn lead to confuse numerical artefacts with physical properties.

5.2 Revising the interpretation of JOP output: a matter of total probability theorem

Elementary properties of bivariate binary processes provide the theoretical explanation of the empirical results in points (1) and (2) listed at the end of Sect. 5.1. Let $X_i$ and $X_j$ be two mutually independent binary processes with marginal rates of occurrence $\pi_i^{X_j}$ and $\pi_i^{X_j} = 1 - \pi_i^{X_j}$. The four joint probabilities $P_{ij}$ describe a valid joint distribution if they fulfill not only the relationship

$$P_{00}^i + P_{01}^i + P_{10}^i + P_{11}^i = 1$$

mentioned by De Michele et al. (2020), but also at least three additional relationships that uniquely define the four probabilities $P_{ij}$, ensuring that they are consistent with the marginal probabilities $\pi_i^{X_j}$ and $\pi_i^{X_j}$. By total probability theorem, we have for example:

$$P_{10}^i + P_{11}^i = \pi_1^{X_j}$$

(8)

$$P_{01}^i + P_{11}^i = \pi_1^{X_j}$$

(9)

$$P_{01}^i + P_{00}^i = \pi_0^{X_j}$$

(10)

Now, let us consider two binary processes with marginal rates of occurrence equal to those of the foregoing independent processes, and some unknown form of dependence that impacts on the joint probabilities. Without loss of generality, let us assume that such an unknown form of dependence results in $P_{11} > P_{11}^i$ (i.e. the so-called ‘11’ dependence in JOP terminology). In order to have a valid joint distribution consistent with marginal probabilities, it must be $P_{10} < P_{10}^i$, $P_{01} > P_{01}^i$, and $P_{00} > P_{00}^i$ according to the constraints in Eqs. 8, 9, and 10, respectively. Starting the same reasoning from any $P_{ij}$, it yields the same conclusions.

Using JOP somewhat misleading language, ‘11’ dependence would necessarily imply ‘00’ dependence, meaning that both $P_{11}$ and $P_{00}$ are necessarily greater than $P_{11}^i$ and $P_{00}^i$, respectively. In this case, we also necessarily have a form of ‘01’ and ‘10’ dependence corresponding to systematic inequalities $P_{10} < P_{10}^i$, and $P_{01} < P_{01}^i$. On the other hand, ‘01’ dependence would imply ‘10’ dependence (i.e. simultaneous $P_{10} > P_{10}^i$, and $P_{01} > P_{01}^i$), and necessarily a form of ‘11’ and ‘00’ dependence corresponding to systematic inequalities $P_{00} < P_{00}^i$, and $P_{11} < P_{11}^i$.

In the ‘new world’ of vaguely defined compound events, JOP implicitly assumes that only the relationship in Equation 7 links the four joint probabilities $P_{ij}$, allowing three of them to move arbitrarily. This misconception leads to the incorrect belief that $P_{ij}$ can be tested individually without accounting for the theoretical constraints driving their joint behavior. On the other hand, going back to the ‘classical world’ of well-devised probability theory, we can
recover the missed elementary properties of joint binary stochastic processes, which allow meaningful analysis, and correct interpretation. In the case of JOP procedure, probability theory says that two binary processes can only be positively/negatively correlated (or independent, of course), and the four probabilities $P_\pi$ must fulfill the inequalities in Eqs. 5 and 6, which are the only admissible scenarios. Any analysis that does not recognize this behavior is theoretically inconsistent.

As far as the point (3) is of concern, misinterpretation of JOP output is further exacerbated by performing the analysis on time series with cyclostationary (seasonally varying) $\pi_1^X$ and $\pi_1^Y$, with no preliminary check of its effect, as well as significance and power. Therefore, JOP analysis does not only rely on an algorithm that neglects a fundamental theorem of probability theory, but it is also applied neglecting its own implicit assumption of stationarity for $\pi_1^X$ and $\pi_1^Y$. Filtering the seasonal variability of $\pi_1^X$ and $\pi_1^Y$ out, our monthly analysis reveals the expected behavior consistent with the theory by recognizing the two scenarios in Eqs. 5 and 6, and indicating that all the probabilities $P_\pi \neq P_\pi^*$ under dependence. As a consequence, identification of multiple types of dependence (i.e. ‘011’, ‘10’, ‘01&10’, etc.) has neither theoretical nor empirical basis, as it comes from the incorrect interpretation of numerical output resulting from a theoretically flawed algorithm.

To further support the foregoing statements, here we present a simple Monte Carlo simulation showing that the results of the annual analysis are spurious numerical artefacts, which cannot be used to draw conclusions about supposed types of dependence different from the overall (positive or negative) dependence summarized by $\rho_C$. We simulated two pairs of binary time series $\{x_i\}$ and $\{x_j\}$ with size $N = 365 \cdot 50$, AR(1) serial dependence with $p_1 = 0.3$, cross correlation $\rho_C \in \{-0.3, 0.3\}$, and cyclically varying $\pi_1^X = \pi_1^Y = 0.16 \cos(\frac{2\pi}{365} n) + 0.25$, $n = 1, \ldots, N$. This setting is similar to that used in Sect. 4.2, and mimics 50 years of daily data. Figure 7 and S9 show JOP results for pairs with positive and negative cross correlation, respectively. According to the original interpretation of JOP annual analysis, one should conclude that the positively cross-correlated pair exhibits ‘11’ dependence, while the negatively cross-correlated pair exhibits ‘10’ dependence. However, monthly analysis clearly shows that (i) all the four probabilities $P_\pi$ generally fall outside their CIs, (ii) null hypothesis $H_0 : P_\pi = P_\pi^*$ is almost always rejected, (iii) lack of systematic rejection for some of the probabilities $P_\pi$ is a spurious effect of the interaction of cross correlation and seasonality, and (iv) results fall within the two scenarios in Eqs. 5 and 6.

Note that we purposely used relatively low values of $\rho_C$ to emphasize the confounding effect of cyclically varying $\pi_1^X$ and $\pi_1^Y$, and obtain JOP results resembling those corresponding to the observed series with supposed ‘11’ and ‘01’ dependence. Therefore, some cases of lack of rejection in the monthly analysis are expected as well. Modulating $\rho_C$, $\pi_1^X$, and $\pi_1^Y$, we can also build synthetic examples so that JOP yields other supposed types of dependence (‘00’, ‘00&11’, etc.). Of course, the cases ‘00&11’ and ‘01&10’ can be obtained using higher absolute values of $\rho_C$ or reducing the amplitude of the seasonal variation of $\pi_1^X$ and $\pi_1^Y$. In fact, both strategies reduce the concealing effect of cyclostationary $\pi_1^X$ and $\pi_1^Y$ on annual JOP analysis. In all cases, monthly analysis better highlights the expected theoretical patterns in Eqs. 5 and 6, revealing the existence of only two types of admissible scenarios alternative to independence.

5.3 Rainfall re-analysis from modeling perspective

To further stress the inconsistency of testing dependence by using the single probabilities $P_\pi$, we show how a model-based approach involving cross correlation $\rho_C$, and marginal probabilities $\pi_1^X$ and $\pi_1^Y$ can effectively describe the patterns of $P_\pi$ for the pairs of rainfall time series analyzed in Sect. 5.1. In fact, unlike the single probabilities $P_\pi$, the three quantities $\rho_C$, $\pi_1^X$, and $\pi_1^Y$ provide sufficient information to describe a bivariate binary process, and set up a model that can be used to obtain CIs under dependence. Precisely, these Monte Carlo CIs are obtained simulating cross correlated binary time series with parameters $\rho_C$, $\pi_1^X$, and $\pi_1^Y$ estimated from the observed binary series describing rainfall occurrence. For direct comparison with results in Sect. 5.1, the analysis was performed on an annual basis and via monthly stratification. Figures 8 and 9 (and S10-S17), corresponding to Figs. 5 and 6 (and S1–S8), report the Laplace CIs and Monte Carlo CIs under dependence. In all cases, observed probabilities $P_\pi$ fall in the center of Monte Carlo CIs, as $\rho_C$, $\pi_1^X$, and $\pi_1^Y$ summarize all the information required to describe and reproduce all $P_\pi$ patterns. This confirms that the supposed multiple types of dependence identified by JOP (testing the single $P_\pi$ at an inappropriate time scale) do not really exist and are only the expected effect of the overall correlation of two binary random processes.

It is worth noting that the interpretation of the JOP test relies on the condition $P_\pi > P_\pi^*$ (equation 2 in De Michele et al. (2020)), meaning that a pair of stations is classified as wet–wet if $P_{11} > P_{11}^*$, for instance. Theoretical discussion
Fig. 7 Results of annual and monthly JOP analysis for a pair of simulated binary series with synchronous cyclic rate of occurrence and positive cross correlation $\rho_c = 0.3$. Lines represent confidence intervals at the 99.5% confidence level.

Fig. 8 Results of annual and monthly $\rho_c$ analysis for the pair of rainfall series (SOUID100810, SOUID105863). Lines represent confidence intervals at the 99.5% confidence level. Monte Carlo confidence intervals ('MC CIs') are obtained from models incorporating cross correlation.
in Sect. 5.2, and modeling results highlight that the foregoing condition and corresponding interpretation are misleading, as the four probabilities $P_i$ are inherently related, and their values necessarily fulfill the conditions corresponding to the two scenarios in Eqs. 5 and 6. Therefore, the actual condition to be checked would be at most $P_i = P_i'$, rather than $P_i > P_i'$. Actually, the structure of a possible statistical test should check the four probabilities $P_i$ simultaneously for the right direction (i.e. $P_i \geq P_i'$) according to the expected theoretical inequalities in Eqs. 5 and 6. Direction and magnitude of such inequalities is identified by the sign and modulus of $q_c$, respectively, while the asymmetries of $P_{i1}$ and $P_{i0}$, and $P_{i0}'$ and $P_{i1}'$ are the effect of the different values of $\pi_1$ and $\pi_1'$. Note that JOP does not (and cannot) provide information about the magnitude of the difference between $P_*$ and $P_*'$. Such a difference could be statistically significant even if its value seemed numerically negligible. In fact, the outcome of a statistical test depends on the width of the distribution of the test statistic under the null hypothesis $H_0$. If the sample size is large enough, any statistical test always rejects the null hypothesis for very small departures from the expected value of the test statistics under $H_0$. In other words, JOP cannot provide information about the physical significance of the difference between $P_*$ and $P_*'$. If the magnitude of such a difference is required for design purposes, the analysis cannot rely on statistical significance, but it should involve the actual values assumed by the probabilities $P_*$. A model-based approach is effective because it simultaneously involves a triplet of summary statistics ($\rho_c$, $\pi_1^X$, $\pi_1^Y$) that provide sufficient information to correctly describe and reproduce the joint behavior of two binary processes, including the patterns of the four probabilities $P_*$. Of course, three out of the four probabilities $P_*$ (e.g. $P_{00}$, $P_{01}$, and $P_{11}$) provide the same amount of information when considered simultaneously. In fact, $\rho_c$, $\pi_1^X$, and $\pi_1^Y$ can be written in terms of three probabilities $P_*$. In this respect, we stress again that the theoretical inconsistency of JOP is not related to use of $P_*$ per se, but to the incorrect assumption that three out of the four probabilities $P_*$ are independent of each other, and can be taken and tested separately. As mentioned in Sect. 3.2 and confirmed by the foregoing rainfall re-analysis, using the triplet ($\rho_c$, $\pi_1^X$, $\pi_1^Y$) has some advantages: (i) it provides a straightforward summary of bivariate binary processes, clearly distinguishing dependence and marginal properties, and avoiding possible misinterpretations; and (ii) it allows for easily setting models that enable simulation of multivariate binary processes, and thus uncertainty analysis of any characteristic.
of interest, including $P_\bullet$. As far as we know, no other triplet of summary statistics is so effective in terms of descriptive clarity, analysis’ efficiency, and ease of building models.

6 Discussion

6.1 Pearson correlation: properties and use

Since JOP procedure relies on the probabilities $P_\bullet$, it is claimed that it “has no problems of existence (like the Pearson correlation coefficient) and has the advantage (respect the existing techniques available in literature) to distinguish between the different types of dependence within the network [with respect to wet–wet, dry–dry, wet–dry, and dry–dry cases], of interest for compound events perspective.” (De Michele et al. 2020). We discussed the issue of ‘types of dependence’ in the previous sections. Here, we focus on the properties of Pearson correlation and its possible limitations in describing mutual dependence of binary processes.

Concerning the problem of existence, $\rho_\bullet$ is not defined if one of the probabilities in the denominator of Eq. 11 is equal to zero. This can happen in two cases: (i) one of the marginal probabilities ($p_1^X$, $p_0^X$, $p_0^X$, and $p_0^X$) is theoretically equal to zero, or (ii) the estimate of one of them is equal to zero. The first case only occurs if a row or column of the theoretical contingency table describing the joint distribution of two binary processes contains only zeros, which in turn means that one of the processes assumes values ‘1’ or ‘0’ with probability equal to one. This case is theoretically meaningless because defining $\rho_\bullet$ would mean attempting to quantify dependence between a random variable and a constant. In this respect, the analysis of $P_\bullet$ leads to identical conclusions from a different perspective. For example, if $p_1^X = 0$, then $P_{11} = P_{10} = 0$, $X_i \equiv 0$, $P_{01} = p_1^X$, and $P_{00} = p_0^X = 1 - p_0^X$. This means that the problem is purely univariate, and no dependence analysis applies or makes sense, as there are no longer two random processes to analyze. The same argument holds for the other cases ($p_0^X = 0$, $p_0^X = 0$, and $p_0^X = 0$), mutatis mutandis. The second case can happen if we have two true binary random processes with marginal probabilities theoretically different from zero, but the estimate $\hat{\rho}_\bullet$ is not defined because one of the estimates $\hat{p}_1^X$, $\hat{p}_1^X$, $\hat{p}_0^X$, $\hat{p}_0^X$ is equal to zero. If the binary process is actually bivariate, undefined $\hat{\rho}_\bullet$ means that one of the time series $\{x_i\}$ or $\{x_j\}$ is a sequence of 0’s or 1’s, which is not representative of the actual marginal properties of the binary processes. Therefore, undefined $\hat{\rho}_\bullet$ indicates that the sample size is insufficient to explore the sample space, and the calculation must be done using longer time series. In other words, the emergence of undefined $\hat{\rho}_\bullet$ values indicates that the analysis is performed improperly on a sample that is not representative of the underlying process. This case can be common when one analyzes too short time series of zero-inflated and potentially persistent (over-dispersed) occurrence processes, such as daily rainfall, especially in arid climates. For this type of data, undefined $\hat{\rho}_\bullet$ is a warning message recalling that statistics should be used wisely, accounting for the nature of processes, available data, and aim of the analysis.

Another characteristic of Pearson correlation coefficient of continuous or discrete (including binary) processes is the dependence of its admissible values on the marginal distributions (e.g. Yule 1912; Lancaster 1957; Embrechts et al. 2002). In particular, for bivariate binary processes, the greatest possible positive (negative) value of $\rho_\bullet$ is $\max\{-(p_1^X/p_0^X)^{1/2}, -(p_0^X/p_1^X)^{1/2}\} \min\{p_1^X, p_0^X/p_0^X/p_1^X\}^{1/2}, (p_0^X/p_1^X/p_1^X/p_1^X)^{1/2}\} \min\{p_1^X, p_0^X/p_0^X/p_1^X\}^{1/2}, (p_0^X/p_1^X/p_1^X/p_1^X)^{1/2}\}$ (e.g. Prentice 1986; Emrich and Piedmonte 1991; Demirtas and Hedeker 2011; Serinaldi and Lombardo 2017a, and references therein). These range restrictions are required to ensure that the joint mass function for the binary outcomes be non-negative for all outcomes (Emrich and Piedmonte 1991).

The boundaries of $\rho_\bullet$ can be problematic in three main circumstances: (i) when $\rho_\bullet$ is used to directly quantify the degree of correlation of two binary processes; (ii) when $\rho_\bullet$ values are used to compare the degree of correlation of different pairs of binary processes with different marginal rates of occurrence (in this case, two $\rho_\bullet$ values are not generally comparable, as they have different admissible ranges); and (iii) when we want to simulate negatively correlated binary processes. The point (iii) is addressed in Appendix A, while the issues (i) and (ii) are directly related, but do not play any relevant role in the context of JOP testing and model-based analysis. In fact, neither JOP nor our model-based approach aim at quantifying dependence and make a comparison of dependence measures. JOP procedure checks whether independence can be discarded or not, and JOP outcome does not depend on the values of the probabilities $P_\bullet$, but on the significance of their difference from $P_\bullet$. On the other hand, our method allows us building models that match/reproduce observed properties, implicitly accounting for their admissible ranges. In other words, for fixed marginal rates of occurrence, the significance of a test and the coverage of confidence intervals are consistent with the admissible ranges of the target statistic, such as $\rho_\bullet$ or $P_\bullet$. For example, for $p_1^X = 0.1$, and $p_1^X = 0.3$, we have $\rho_\bullet \in (-0.048, 0.259)$, and such a range is consistent with all the admissible values of $P_\bullet$. This explains why our models based on $\rho_\bullet$, $p_1^X$, and
\( \pi^X_1 \), are capable of precisely reproducing all observed \( \mathcal{P} \) values.

To summarize, the properties of \( \rho_C \) for bivariate binary processes can be a shortcoming in applications involving statements on their values. Since JOP test makes only statement on rejection/no rejection, and the distribution of the test statistics under \( H_0 \) accounts automatically for the admissible range, it follows that such a range does not matter in this type of analysis, and therefore JOP does not provide any practical improvement compared to a test based on \( \rho_C \). On the contrary, JOP leads to incorrect conclusions about the nature of dependence because it tests the four probabilities \( \mathcal{P} \) independently, missing their mutual relationships.

### 6.2 Avoiding statistical tests to avoid misinterpretation

Most of the problems raising from the use and interpretation of the JOP test are due to the inherent limits (and inconsistencies) characterizing the hybrid Fisher-Neymann-Pearson statistical testing procedures, which are widely discussed in the theoretical and applied literature (Gigerenzer et al. 1989; Flueck and Brown 1993; McBride et al. 1993; Meehl 1997; Gill 1999; Johnson 1999; Nicholls 2001; Levine et al. 2008; Clarke 2010; Beninger et al. 2012; Ellison et al. 2014; Nuzzo 2014; Briggs 2016; Greenland et al. 2016; Wasserstein and Lazar 2016; Serinaldi et al. 2018; Wasserstein et al. 2019, and references therein). In the context of climate literature, Ambaum (2010) highlights how statistical tests very rarely provide useful quantitative information on correlation, for instance. In particular, Ambaum (2010) focuses on a common error called the error of transposed conditional (also known as converse inequality argument or inverse probability problem (Pollard and Richardson 1987; Gill 1999; Krämer and Gigerenzer 2005; Beninger et al. 2012; Serinaldi et al. 2018)), which consists of confusing conditional and conditioning events, so that “we want to devise a way to assign a probability to the validity of the null hypothesis, given the observed correlation. But what we end up doing is calculating the probability for a correlation at least as big as the observed correlation when the null hypothesis is assumed to be true.” (Ambaum 2010). Even though these probabilities are related by Bayes’ theorem, they are different. Moreover, rejection of \( H_0 \) (and thus low \( p \) value), “is not indicative of much at all except that the observed correlation is not very probable if the null hypothesis were true... But... any quantitative information [on the posterior probability for \( H_1 \) given the observations] assuming the null hypothesis is quite irrelevant” (Ambaum 2010). Indeed, stating that ‘correlation is significant’ makes little sense in any practical sense, as this statement is a category error occurring when a property is wrongly ascribed to something that cannot have this property. In this case, the hypothesis test, and related \( p \) values and rejection rules, refer to unrelated processes (i.e. \( H_0 \)), and they do not (and cannot) say anything about related processes (i.e. \( H_1 \)).

For example, the outcome of the JOP test (or any test checking a hypothesis \( H_0 \)) can only be a statement on ‘rejection’ or ‘no rejection’ of independence. However, rejection does not allow any statement about possible dependence, because dependence is not a property of the model \( H_0 \) assumed to perform the test. In other words, ‘rejection’ can be due to something that is unknown and different from dependence. Analogously, ‘no rejection’ resulting from annual JOP might be related to violation of the implicit assumption of stationary rates of occurrence, i.e. the introduction of a exogenous factor (cyclical non-stationarity of \( \pi^X_1 \), and \( \pi^Y_1 \)) that is not included in the list of assumptions of the model \( H_0 \). To make statements on dependence, we need a model incorporating dependence, thus enabling comparison of model output and observations. If we call this model \( H_1 \), in principle, we can build a test checking \( H_1 \). However, also in this case, such a test can only provide statements (‘rejection’ or ‘no rejection’) on \( H_1 \), and nothing else. Moreover, as shown in Sect. 5.3, if the model is informative, testing \( H_1 \) is no very useful, as the direct comparison of model output and observations is sufficient to assess the model performance and the suitability of the underlying assumptions.

The foregoing inconsistencies affecting statistical tests contributed to the misinterpretation of the JOP output, as it does not provide information about anything different from independence of pairs of binary processes. Conversely, as discussed in Sect. 3.2, our model-based approach is actually a change of paradigm in this context, as it moves from ubiquitous uninformative statistical testing to informed, parsimonious, and scientifically tenable stochastic modeling, relying on solid theory and extensive validation/falsification under challenging conditions, rather than untenable/unrealistic null hypotheses.

### 7 Conclusions

This study presented an inquiry on the potential problems and inconsistencies related to the introduction of new topics or relabeling known concepts without the necessary rigor and attention for the existing theories. We considered the emerging research area of so-called ‘compound events’, focusing on a specific test proposed by De Michele et al. (2020) to study the joint occurrence of rainfall at pairs of locations. This method, denoted as JOP test throughout the
text, relies on a separate analysis of the single probabilities $P_i$ (i.e., $P_{00}$, $P_{11}$, $P_{10}$, and $P_{01}$), and was proposed as an improved technique overcoming the supposed shortcomings of classical analysis based on Pearson correlation.

However, our _neutral_ validation led to an opposite conclusion. In fact, the probabilities $P_i$ are inherently related, they are not enough to describe the effects of dependence when taken individually, and their individual analysis can lead to incorrect interpretations if their relationship is not accounted for. As a consequence, the four types of dependence originally introduced to interpret JOP output are numerical artefacts that result from overlooking the available theory, the expected effects of dependence and marginal rates of occurrence, and the interaction with confounding factors, such as seasonality and serial correlation. We also discussed why the properties of Pearson correlation are not a true shortcoming in the context of statistical testing, and why JOP test cannot provide any improvement in this respect. Therefore, the rather convoluted JOP algorithm cannot provide the information we expect on the joint behavior of bivariate binary processes. On the other hand, classical Pearson correlation coefficient and the marginal probabilities of occurrence provide a rigorous description and clear interpretation of the behavior of bivariate binary rainfall observations, and allow for setting up models capable of faithfully reproducing the observed fluctuations of the single probabilities $P_i$. Moreover, these models can be used to check and validate any statistical test under challenging conditions.

We presented the case of the JOP algorithm in a wider perspective of quite a common attitude in many human activities, consisting of revitalizing a research topic or revamping businesses by re-labeling concepts or rebranding products, without introducing substantial changes. While a rigorous approach to compound, interacting, interconnected, and cascading risks is surely of interest, the concept of ‘compound events’ is often ill-defined, superficially used, and conceals the re-proposal of classical problems already addressed in the past. When new names do not correspond to new concepts, the main effect is a hiatus between new and existing literature, whereby the former neglects the latter, which in turn provides theories and solutions for the same (or very close) subjects under different names. In this context, a method like the JOP test should have been considered only after double checking that the classical analysis based on Pearson correlation and related testing and modeling do not provide satisfactory results. This would have been easier by recognizing that the study of rainfall occurrence process at two locations concerns the theory of bivariate binary processes. Proper study of such processes does not require any ‘compound events’ re-labeling, which in turn might introduce confusion and theoretically inconsistent statements transforming ‘compound events’ into ‘confounding events’. Finally, we argue that the use of different names to denote the same well-defined concept already causes lack of communication among different disciplines, slowing science advances down. When new names do not even correspond to well-defined concepts, their ambiguity can cause the paradoxes discussed in this study. The best way to advance science is to introduce new concepts rather than names.

### Appendix A: Simulation of binary processes with negative correlation

BetaBit and BetaBitST simulate positively correlated binary random processes using a parent Gaussian process (Serinaldi and Lombardo 2017a; Serinaldi and Kilsby 2018). Referring to Papalexiou (2018) and references therein for a general discussion, these methods rely on the relationship between the correlation coefficient, $\rho_Y$, of a parent bivariate Gaussian processes $(Y_i, Y_j)$ and that of the target bivariate binary process, $\rho_X$:

$$
\rho_X = \Phi(y(\pi_1^X), y(\pi_1^X); \rho_Y) - \pi_1^X \pi_1^X,
$$

(A.1)

where $y(\pi)$ denotes the $\pi$th quantile of the standard Gaussian distribution, and $\Phi$ the bivariate Gaussian distribution. Solving Eq. A.1 yields the value of $\rho_Y$ required to simulate a bivariate binary random process with desired $\rho_X$, transforming the parent standard Gaussian variables into binary by

$$
x = \begin{cases} 
1 & \text{if } y < \Phi^{-1}(\pi_1^X) \\
0 & \text{otherwise} 
\end{cases}
$$

(A.2)

However, as mentioned in Sect. 6.1, $\rho_X$ resulting from the transformation in Eq. A.1 is not free to range in [-1,1]. In more detail, as $\pi_1^X$ and $\pi_1^X$ decrease and are lower than 0.5, the lower limit of admissible negative values of $\rho_X$ tends to zero according to the theoretical lower bound

$$
\max\left\{ -\left(\frac{\pi_1^X}{\pi_1^X/\pi_0^X}\right)^{1/2}, -\left(\frac{\pi_0^X/\pi_1^X}{\pi_1^X}\right)^{1/2} \right\} 
$$

(Emrich and Piedmonte 1991). While this problem is not serious for geophysical processes and applications characterized by positive correlation and persistence, it prevents the simulation of negatively correlated rainfall occurrence processes considered in this study.

Here, we suggest a method to overcome such limitations and simulate negatively correlated binary processes. Let $\rho_{X-}$ be the correlation coefficient of a negatively correlated binary processes, and $\pi_1^X$ and $\pi_1^X$ the marginal rates of occurrence. As we cannot generally use Eq. A.1 to define the required value of $\rho_{Y-}$ of the parent Gaussian process,
we can circumvent the problem by focusing on a dual positively correlated binary process with \( \rho_{X+}, \pi_{1+}^{X}, \) and \( \pi_{1-}^{X}, \) and the corresponding parent Gaussian process with \( \rho_{Y+}. \) Such dual binary processes are linked by the following relationships:

\[
\begin{align*}
\rho_{X+} &= -\rho_{X-} \\
\pi_{1+}^{X} &= \mathcal{P}_{00-} + \mathcal{P}_{10-} = \pi_{0-}^{X} \\
\pi_{1-}^{X} &= \mathcal{P}_{11-} + \mathcal{P}_{10-} = \pi_{1-}^{X}
\end{align*}
\]  
(A.3)

These relationships derive from observing that \( \rho_{C} \) in Eq. 11 depends on \( \mathcal{P}_{\star} \), and the sign of \( \rho_{C} \) changes from positive to negative when the off diagonal probability mass becomes greater than that along the diagonal, and vice versa. Therefore, for bivariate binary pairs with negative correlation, we can define a dual process with positive correlation whose properties are obtained by circular rotation of the probability masses \( \mathcal{P}_{\star} \), writing for instance

\[
\begin{align*}
\mathcal{P}_{00+} &= \mathcal{P}_{01-} \\
\mathcal{P}_{10+} &= \mathcal{P}_{00-} \\
\mathcal{P}_{11+} &= \mathcal{P}_{10-} \\
\mathcal{P}_{01+} &= \mathcal{P}_{11-}
\end{align*}
\]  
(A.4)

Figure 10 provides a visualization of counterclockwise rotation along with an illustration of parent Gaussian samples identifying the masses of probability in each of the four quadrants corresponding to the four possible states of a binary pair. Therefore, the simulation procedure of a process with desired \( \rho_{X-}, \pi_{1-}^{X}, \) and \( \pi_{1-}^{Y} \) consists of the following steps:

1. Given \( \rho_{X-}, \pi_{1-}^{X}, \) and \( \pi_{1-}^{Y}, \) compute \( \rho_{X+}, \pi_{1+}^{X}, \) and \( \pi_{1+}^{Y} \) by Eq. A.3.

2. Simulate a dual positively correlated bivariate binary process via BetaBitST.

3. Back transform the positively correlated binary pairs into the corresponding negatively correlated pairs according to the correspondences in Eq. A.4; this means replacing 0’s (1’s) with 1’s (0’s) as follows

\[
\begin{align*}
\begin{array}{ll}
& x_{i} = 1 \\
& x_{i} = 0 \\
& x_{i} = 1
\end{array}
\end{align*}
\]  
(A.5)

Of course, a dual positively correlated bivariate binary process can be obtained by clockwise rotation as well. In this case, Eqs. A.3, A.4, and A.5 become respectively

\[
\begin{align*}
\rho_{X+} &= -\rho_{X-} \\
\pi_{1+}^{X} &= \mathcal{P}_{11-} + \mathcal{P}_{01-} = \pi_{1-}^{Y} \\
\pi_{1+}^{X} &= \mathcal{P}_{00-} + \mathcal{P}_{01-} = \pi_{0-}^{X}
\end{align*}
\]  
(A.6)

and

\[
\begin{align*}
\begin{array}{ll}
& x_{i} = 1 \\
& x_{i} = 0 \\
& x_{i} = 1
\end{array}
\end{align*}
\]  
(A.7)

get pairs \( \{00+, 01-, 10+, 11-, 01+, 10-, 11+, 00-\} \) and set \( x_{i} = 1 \) to obtain \( \{11-, 10-, 01-, 00-\} \).

get pairs \( \{00+, 01-, 10+, 11-, 01+, 10-, 11+, 00-\} \) and set \( x_{i} = 1 \) to obtain \( \{11-, 10-, 01-, 00-\} \).

Fig. 10 Visual representation of the mechanism underlying the simulation of negatively correlated binary processes.
Appendix B: ACF of AR(1) and fGn processes

Denoting the lag time as $\tau$, the discrete-time AR(1) process is characterized by exponentially decaying autocorrelation function of the form

$$p_X(\tau) = \exp(-\gamma|\tau|) = \rho_1^{|\tau|}, \quad (B.1)$$

where $1/\gamma$ is the correlation radius and $\rho_1 = \exp(-\gamma)$ is the lag-one autocorrelation coefficient. The fractional Gaussian noise (fGn) also known as Hurst-Kolmogorov process (Koutsoyiannis 2010), is characterized by the following ACF

$$p_X(\tau) = \frac{1}{2}(|\tau + 1|^{2H} - 2|\tau|^{2H} + |\tau - 1|^{2H}), \quad (B.2)$$

which exhibits a power-law decay $p_X(\tau) \propto |\tau|^{2H-2}$. For $0.5 < H < 1$ the process is positively correlated and exhibits long-range dependence, while it reduces to white noise for $H = 0.5$.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s00477-022-02190-6.

Acknowledgements Francesco Serinaldi and Chris G. Kilsby acknowledge the support from the Willis Research Network. Federico Lombardo is grateful to the Italian National Fire and Rescue Service for the continuous support. The authors thank the two anonymous reviewers, Dr. Carlo De Michele (Politecnico di Milano, Italy) and Dr. Giuseppe Mascaro (Arizona State University, US), and an anonymous reviewer for their insightful remarks and criticisms that helped greatly improve the original manuscript. The analyses were performed in R (R Development Core Team 2021).

Funding No funding was received for conducting this study.

Availability of data and materials Data are freely available from the Global Historical Climatology Network-Daily (GHCN-D; Menne et al. 2012a, b) (https://www.ncdc.noaa.gov/data/global-historical-climatology-network-daily/), and the European Climate Assessment Dataset (ECA&D; Klein Tank et al. 2002) (https://www.ecad.eu).

Declarations

Competing interest The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

CRediT authorship contribution statement Francesco Serinaldi: Conceptualization, Methodology, Software, Formal analysis, Writing-original draft, Writing-review & editing, Visualization. Federico Lombardo: Conceptualization, Methodology, Writing-review & editing. Chris G. Kilsby: Conceptualization, Writing-review & editing.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

References

AghaKouchak A, Cheng L, Mazdiyasni O, Farahmand A (2014) Global warming and changes in risk of concurrent climate extremes: insights from the 2014 California drought. Geophys Res Lett 41(24):8847–8852

AghaKouchak A, Chiang F, Huning LS, Love CA, Mallakpour I, Mazdiyasni O, Moftakhari H, Papalexiou SM, Rago E, Sadegh M (2020) Climate extremes and compound hazards in a warming world. Annu Rev Earth Planet Sci 48(1):519–548

Ambaum MHP (2010) Significance tests in climate science. J Clim 23(22):5927–5932

Anagnos T, Kiremidjian AS (1988) A review of earthquake occurrence models for seismic hazard analysis. Probab Eng Mech 3(1):3–11

Asadi P, Davison AC, Engelke S (2015) Extremes on river networks. Ann Appl Stat 9(4):2023–2050

Aven T (2016) Risk assessment and risk management: review of recent advances on their foundation. Eur J Oper Res 253(1):1–13

Beninger PG, Boldina I, Katsanevakis S (2012) Strengthening statistical usage in marine ecology. J Exp Mar Biol Ecol 426–427:97–108

Bogachev MI, Bunde A (2012) Universality in the precipitation and river runoff. Europhys Lett 97(4):48011

Boulesteix A, Binder H, Abrahamowicz M, Sauerbrei W (2018) for the simulation panel of the STRATOS initiative, On the necessity and design of studies comparing statistical methods. Biometrical J 60(1):216–218

Briggs W (2016) Uncertainty; the soul of modeling, probability & statistics. Springer, New York

Brusco M, Cradit JD, Steinley D (2021) A comparison of 71 binary similarity coefficients: the effect of base rates. Plos One 16(4):e0247751

Camus P, Haigh ID, Nasr AA, Wahl T, Darby SE, Nicholls RJ (2021) Regional analysis of multivariate compound coastal flooding potential around Europe and Environs: sensitivity analysis and spatial patterns. Nat Hazard 21(7):2021–2040

Chamberlain S (2020) rnoaa: ‘NOAA’ Weather Data from R. https://CRAN.R-project.org/package=rnoaa, r package version 1.2.0

Clarke RT (2010) On the (mis)use of statistical methods in hydro-climatological research. Hydrol Sci J 55(2):139–144

Coats S, Smerdon JE, Karnauskas KB, Seager R (2016) The improbable but unexceptional occurrence of megadrought clustering in the American West during the Medieval Climate Anomaly. Environ Res Lett 11(7):074025

Cox DR, Isham V (1980) Point processes. Chapman & Hall/CRC Monographs on Statistics & Applied Probability, Taylor & Francis, London

Cramer H (1946) Mathematical methods of statistics. Landmarks in Mathematics, Princeton University Press, New Jersey

Davison AC, Padoan SA, Ribatet M (2012) Statistical modeling of spatial extremes. Stat Sci 27(2):161–186
De Michele C, Meroni V, Rahimi L, Deidda C, Ghezzi A (2020) Dependence types in a binarized precipitation network. Geophys Res Lett 47(23):e2020GL090196

Demirtas H (2006) A method for multivariate ordinal data generation given marginal distributions and correlations. J Stat Comput Simul 76(1):1017–1025

Demirtas H, Hedeker D (2011) A practical way for computing approximate lower and upper correlation bounds. Am Stat 65(2):104–109

Diggle PJ (2013) Statistical analysis of spatial and spatio-temporal point patterns, 3rd edn. CRC Press, Boca Raton

Diggle PJ, Rowlingson B, Su T (2005) Point process methodology for on-line spatio-temporal disease surveillance. Environmetrics 16(5):423–434

Ellison AM, Gotelli NJ, Inouye BD, Strong DR (2014) P values, hypothesis testing, and model selection: it’s déjà vu all over again. Ecology 95(3):609–610

Embrechts P, McNeil AJ, Straumann D (2002) Correlation and dependence in risk management: properties and pitfalls. Cambridge University Press, Cambridge, pp 176–223

Emrich LJ, Piedmonte MR (1991) A method for generating high-dimensional multivariate binary variates. Am Stat 45(4):302–304

Flueck JA, Brown TJ (1993) Criteria and methods for performing and evaluating solar-weather studies. J Clim 6(2):373–385

Gigerenzer G, Swijtink Z, Porter T, Daston L, Beatty J, Krüger L (1989) The Empire of chance: how probability changed science and everyday life. New York

Gill J (1999) The insignificance of null hypothesis significance testing. Polit Res Q 52(3):647–674

Gong L, Zhou Z, Tong P, Zhao S (2011) Statistical properties of one-dimensional binary sequences with power-law power spectrum. Physica A 390(17):2977–2986

Greenland S, Senn SJ, Rothman KJ, Carlin JB, Poole C, Goodman SN, Altman DG (2016) Statistical tests, P values, confidence intervals, and power: a guide to misinterpretations. Eur J Epidemiol 31(4):337–350

Hao Z, Singh V, Hao F (2018) Compound extremes in hydroclimatology: a review. Water Resour Res 10(6):718

Harrold TI, Sharma A, Sheather SJ (2003) A nonparametric model for stochastic generation of daily rainfall occurrence. Water Resour Res 39(10):1300

Heffernan JE, Tawn JA (2004) A conditional approach for multivariate extreme values (with discussion). J R Stat Soc: Ser B (Stat Methodol) 66(3):497–546

Jacobs PA, Lewis PAW (1983) Stationary discrete autoregressive-moving average time series generated by mixtures. J Time Ser Anal 4(1):19–36

Jentsch C, Reichmann L (2019) Generalized binary time series models. Econometrics 7(4):47

Johnson DH (1999) The insignificance of statistical significance testing. J Wildl Manag 63(3):763–772

Klein Tank AMG, Wijngaard JB, Können GP, Böhm R, Demarée G, Gocheva A, Mileta M, Pashiardis S, Hejkrlik L, Kern-Hansen C, Heino R, Bessemoulin P, Müller-Westermann G, Tzanakou M, Szolgay J, Šalai J, Pálföldi T, Fritschy J, Rubin S, Capaldo M, Maugeri M, Leitass A, Bukiants A, Aberfeld R, van Engelen AFV, Forland E, Mietus M, Coelho F, Mares C, Razuvaev V, Nieplova E, Cegnar T, Antonio López J, Dahlström B, Moberg A, Kirchoffer W, Ceylan A, Pachaliuk O, Alexander LV, Petrovic P (2002) Daily dataset of 20th-century surface air temperature and precipitation series for the European Climate Assessment. Int J Climatol 22(12):1441–1453

Koutsoyiannis D (2006) An entropic-stochastic representation of rainfall intermittency: the origin of clustering and persistence. Water Resour Res 42(1):W01401

Koutsoyiannis D (2010) HESS Opinions “A random walk on water”. Hydrol Earth Syst Sci 14(3):585–601

Koutsoyiannis D, Dimitriadi P, Lombardo F, Stevens S (2018) From fractals to stochastics: seeking theoretical consistency in analysis of geophysical data. In: AT (ed) Advances in nonlinear geosciences. Springer, pp 237–278

Krämer W, Gigerenzer G (2005) How to confuse with statistics or: the use and misuse of conditional probabilities. Stat Sci 20(3):223–230

Lancaster HO (1957) Some properties of the bivariate normal distribution considered in the form of a contingency table. Biometrika 44(1/2):289–292

Landauer R (1997) Fashions in science and technology. Phys Today 50(12):61–62

Laplace PS (1812) Theorie analytique des probabilites. Courcier, Paris

Landauer R (1997) Fashions in science and technology. Phys Today 50(12):61–62

Leonard M, Westra S, Phatak A, Lambert M, van den Hurk B, McInnes K, Risbej J, Schuster S, Jakob D, Stafford-Mitch M (2014) A compound event framework for understanding extreme impacts. WIREs Clim Change 5(1):113–128

Levine TR, Weber R, Hullett C, Park HS, Lindsey LLM (2008) A critical assessment of null hypothesis significance testing in quantitative communication research. Hum Commun Res 34(2):171–187

Lombardo F, Volpi E, Koutsoyiannis D (2012) Rainfall downscaling in time: theoretical and empirical comparison between multifractal and Hurst–Kolmogorov discrete random cascades. Hydrol Sci J 57(6):1052–1066

Lombardo F, Volpi E, Koutsoyiannis D, Papalexiou SM (2014) Just two moments! a cautionary note against use of high-order moments in multifractal models in hydrology. Hydrol Earth Syst Sci 18(1):243–255

Lombardo F, Volpi E, Koutsoyiannis D, Serinaldi F (2017) A theoretically consistent stochastic cascade for temporal disaggregation of intermittent rainfall. Water Resour Res 53(6):4586–4605

Lombardo F, Napolitano F, Russo F, Koutsoyiannis D (2019) On the exact distribution of correlated extremes in hydrology. Water Resour Res 55(12):10405-10423

Lowen SB, Teich MC (2005) Fractal-based point processes. Wiley, Hoboken, New Jersey

Macke JH, Berens P, Ecker AS, Tolias AS, Bethge M (2009) Generating spike trains with specified correlation coefficients. Neural Comput 21(2):397–423

Mailier PJ, Stephenson DB, Ferro CAT, Hodges KJ (2006) Serial clustering of extratropical cyclones. Mon Weather Rev 134(8):2224–2240

Marchaud R (2012) Spatial correlation of hydrometeor occurrence, reflectivity, and rain rate from cloudsat. J Geophys Res: Atmospheres 117(D6):D06202

McBride GB, Loftis JC, Adkins NC (1993) What do significance tests really tell us about the environment? Environ Manage 17(4):423–432

Meehl PE (1997) The problem is epistemology, not statistics: replace significance tests by confidence intervals and quantify accuracy of risky numerical predictions. Erlbaum, Mahwah, pp 393–425

Mehrotra R, Westra S, Sharma A, Srikanth R (2012) Continuous rainfall simulation: 2. A regionalized daily rainfall generation approach. Water Resour Res 48(1):W01536

Menne MJ, Durre I, Korzeniewski B, McNeal S, Thomas K, Yin X, Anthony S, Ray R, Vose R, Gleason BE, Houston TG (2012a) Global historical climatology network - daily (ghcn-daily), version 3.26

Menne MJ, Durre I, Vose RS, Gleason BE, Houston TG (2012) An Overview of the Global Historical Climatology Network-Daily Database. J Atmos Oceanic Tech 29(7):897–910
