Supplement of Hydrol. Earth Syst. Sci., 28, 3191–3218, 2024
https://doi.org/10.5194/hess-28-3191-2024-supplement

Hydrology and Earth System Sciences
Open Access

Supplement of

Scientific logic and spatio-temporal dependence in analyzing extreme-precipitation frequency: negligible or neglected?

Francesco Serinaldi

Correspondence to: Francesco Serinaldi (francesco.serinaldi@ncl.ac.uk)

S1 Non-homogeneous Poisson (NHP) Process and First-order Poisson Integer Autoregressive (Poisson-INAR(1)) Process

Let P be the precipitation process sampled at given time scale (e.g., daily), and let us denote as Z the number of events (observations of P) exceeding a given value (e.g., a percentage threshold) in a specified time windows (e.g., 365 days = 1 year). Loosely speaking, the process $\{Z_j\}$ (with $j = 0,1,2,...$) follows a non-homogenous Poisson (NHP) process if $\{Z_j\}$ has Poisson distribution with time-varying rate of occurrence $\lambda(j)$. Under the assumption that $\lambda(j)$ varies linearly along the years, we have $\lambda(j) = \lambda_0 + \phi \cdot j$ (with $j = 0,1,2,...$), where λ_0 and ϕ are the intercept and slope parameters, respectively.

A process $\{Z_j\}$ (with $j = 0,1,2,...$) is first-order Poisson Integer Autoregressive (Poisson-INAR(1)) process if $Z_j = \rho_1 \circ Z_{j-1} + \varepsilon_j$ (with $j = 1,2,3,...$), where ρ_1 correspond to the lag-1 autocorrelation value of $\{Z_j\}$, the symbol '$\circ$' denotes the binomial thinning operator, and $\{\varepsilon_j\}$ is a sequence of independent Poisson random variables with rate of occurrence $\mu = (1-\rho_1)\cdot\lambda$, where λ is the rate of occurrence of the process $\{Z_j\}$. We refer to Farris et al. (2021) and references therein for further details about these processes and estimation of their parameters.

S2 Beta-Binomial distribution

Let $\{Y_j\}$ be a discrete-time Bernoulli process with state space $\{0,1\}$ and probability of success/failure in each trial $p = \mathbb{P}[Y_j = 1] \in [0,1]$, where j $(= 0,1,2,...)$ denotes discrete time. For daily P, the process $\{Y_j\}$ describes the binary time series resulting from the occurrence/non occurrence of over-threshold (OT) exceedances in each day of the period of record. With this notation, the number of OT events during n time steps (e.g., 365 days) is defined as $Z(Y_j) = \sum_{j=1}^{n} Y_j$. Among the distributions devised to describe Z, the Beta-Binomial ($\beta\mathcal{B}$) distribution plays a key role in the case of mutually dependent trials. The $\beta\mathcal{B}$ distribution is a compound distribution resulting from the ordinary Binomial ($\mathcal{B}$) distribution $f_{\mathcal{B}}(z) = \binom{n}{z}\psi^z(1-\psi)^{n-z}$, when ψ is assumed to be a random variable Ψ following a beta distribution $f_\beta(\psi) = \frac{\psi^{\alpha-1}(1-\psi)^{\beta-1}}{\mathrm{B}(\alpha,\beta)}$ with mean $\mathbb{E}[\Psi] = p$, where $\mathbb{E}$ is the expectation operator, B denotes beta function, and α and β are two positive shape parameters. The $\beta\mathcal{B}$ probability mass function can be written as (Skellam, 1948)

$$f_{\beta\mathcal{B}}(z) = \binom{n}{z}\frac{\mathrm{B}(z+\alpha, n-z+\beta)}{\mathrm{B}(\alpha,\beta)}, \tag{S1}$$

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

$$\mu_{\beta\mathcal{B}} := \mathbb{E}[Z] = np, \tag{S2}$$

and

$$\sigma^2_{\beta\mathcal{B}} := \mathbb{V}[Z] = np(1-p)\left[1+(n-1)\rho_{\beta\mathcal{B}}\right], \tag{S3}$$

where $\mathbb{V}$ is the variance operator, $p = \alpha/(\alpha+\beta)$, and $\rho_{\beta\mathcal{B}} = 1/(\alpha+\beta+1)$ is known as the 'intra class' or 'intra cluster' correlation. If the random variable Ψ has a degenerate distribution with probability 1 at a single point (or $\alpha \to \infty$ and $\beta \to \infty$), then $\mathbb{V}[\Psi] = 0$ and Z becomes binomial with $\mu_{\mathcal{B}} = p$ (Ahn and Chen, 1995). Being positive by definition, $\rho_{\beta\mathcal{B}}$ produces over-dispersion as it inflates the variance $np(1-p)$ of the original $\mathcal{B}$ distribution with constant p. On the other hand, $\rho_{\beta\mathcal{B}}$ does not affect the expected value, which is identical for $\beta\mathcal{B}$ and $\mathcal{B}$ models. For correlated experiments, we have (Serinaldi et al., 2020b)

$$\rho_{\beta\mathcal{B}} = \frac{\sum\sum_{j\neq l}\rho_{jl}}{n(n-1)}, \tag{S4}$$

where $\rho_{jl} = \mathbb{C}[Y_j, Y_l]$ denotes the pairwise correlation of Bernoulli experiment j and l in the parent process Y. The indices j and l can refer to two different time steps in a temporal process evolving over n time steps, or two locations in a spatial process over n locations. For a spatio-temporal process over n time steps and m locations, $\rho_{ji,lk} = \mathbb{C}[Y_{ji}, Y_{lk}]$ are the $q = n \cdot m^2 - m$ elements of the space-time correlation matrix among m time series up to lag n, and Eq. S4 reads as

$$\rho_{\beta\mathcal{B}} = \frac{\sum\sum_{ji \neq lk} \rho_{ji,lk}}{q}. \tag{S5}$$

For stationary spatio-temporal processes Y, ρ_{jl} and $\rho_{ji,lk}$ are independent of the specific spatio-temporal coordinates, and depend only on the distance, that is, the spatio-temporal lag, thus reading as $\rho_{|l-j|}$ and $\rho_{|l-j|,|k-i|}$, where $|l-j|$ $(i = 1, ..., n)$ denotes the time lag, and $|k-i|$ $(i, k = 1, ..., m)$ the distance between two spatial locations i and k. In these cases, $\rho_{|l-j|}$ and $\rho_{|l-j|,|k-i|}$ are the terms of the ACF and spatio-temporal correlation matrix of Y, respectively.

The $\beta\mathcal{B}$ distribution has been used in several fields for various purposes (see e.g., Nicola and Goyal, 1990; Hughes and Madden, 1993; Tsai et al., 2003), including the estimation of the number of rejections in multiple tests for trend in spatially dependent stream flow records (Serinaldi et al., 2018), and the estimation of the number of OT events under spatio-temporal dependence (Serinaldi and Kilsby, 2018).

S3 Stationarity, nonstationarity, and trends

As stressed by Serinaldi et al. (2018), one of the key problems affecting the literature on trend analysis is the confusion between population properties that are deduced logically from the theory and the sample properties that are determined empirically from observations. Even though these concepts are reported in every introductory handbook of applied statistics, and should be clear to any data analyst, the existing literature indicates widespread confusion resulting in sentences such as 'finite value of the mean does always exist as we can always calculate it from a sample'. Of course, such kind of sentences misses the difference between the sample estimates of the mean and its population counter part, which cannot exist for Cauchy random variables, for instance. In other words, the meaningfulness of sample means depends on the assumption we make about the underlying population.

Similar remarks hold for 'trends'. The word 'trend' is used with different meanings not only in different disciplines but also in the same discipline. For example, in finance, when we deal with econometric models, trend and nonstationarity have necessarily the meaning resulting from the formal definition of stationarity given by Khinchin, that is, "*... a stationary stochastic process in the sense of Khinchin ... is a set of random variables X_t depending on the parameter t, $-\infty < t < +\infty$, such that the distributions of the systems $(X_{t_1}, X_{t_2}, ..., X_{t_n})$ and $(X_{t_1+\tau}, X_{t_2+\tau}, ..., X_{t_n+\tau})$ coincide for any n, t_1, t_2, ...,t_n, and τ*" (Koutsoyiannis and Montanari, 2015). Otherwise, econometric models would not be compliant with mathematics, asymptotic derivations, etc. However, in finance, the word 'trend' can have a different meaning when dealing with technical (visual) analysis, which is often used to support investment and trading strategies. In that context, upward (downward) trends are observed patterns of stock price records where local minimum and maximum values are followed larger (smaller) local minimum and maximum values. Obviously, this definition of 'trend' reflects just a sampling property. It may be useful in some contexts, but it does not allow any theoretical derivation, model definition, or formal testing procedure.

The hydro-climatological literature dealing with trend testing often confuses sample patterns observed for instances over a few decades of annual records with population properties, which in turn must be valid for any time $-\infty < t < +\infty$. Building on Koutsoyiannis and Montanari (2015) and elaborating around the formal definition of stationarity and ergodicity, Serinaldi et al. (2018)suggested the following definition of stochastic process with trend:

$$\mathrm{G}[X_t] = d_t + \mathrm{G}[\mathcal{E}_t], \tag{S6}$$

where, $\{X_t\}$ is a stationary stochastic process, that is, a set of random variables X_t depending on the parameter t, $-\infty < t < +\infty$, $\mathrm{G}[\cdot]$ is a generic operator, and d_t is a *deterministic function of time* $d_t \equiv d(t)$. According to Koutsoyiannis and Montanari (2015), a deterministic function of time is "*precisely known and perfectly predictable*" meaning that a system input corresponds to a single system response, contrasting stochastic dynamics where a single input could result in multiple outputs. Therefore a trend is defined as "*time-dependent deterministic and therefore predictable change d_t of the properties of a process X_t, where the term 'deterministic' implies prediction variance equal to zero (one-to-one relationship). This definition* *highlights that trends (and nonstationarity) refer to the underlying process, and attempting to infer nonstationarity requires both detection and attribution based on a combination of deductive reasoning, which supports and justifies the existence of a time-dependent deterministic function (i.e., trends and nonstationarity) and inductive reasoning, which provides (i) preliminary knowledge by exploratory data analysis, and (ii) quantification/parametrization of dt by confirmatory/disproving analysis and modeling. We stress that 'prediction variance equal to zero' does not refer to the specific parametrization of dt but its existence and its overall evolution. For example, the parametrization of seasonal trends (components) obviously varies even for the same process observed at different locations, but the existence of the seasonal cycle and its effects in terms of alternation of wet/dry and cold/warm conditions along the calendar year are predictable with (almost) no uncertainty. Other forms of trend/nonstationarity are allowed only if they are supported by the same kind of deductive and inductive arguments.*" (Serinaldi et al., 2018).

Therefore, in the context of the analysis reported in the paper and most of the literature on the topic, a trend should be intended as a deterministic function or rule linking a variable of interest (or one or more of its properties) to another variable describing a parametric support. For example, the NHP model (Section S1) assumes that the rate of occurrence λ is a log-linear function of time steps j:

$$\log(\lambda(j)) = \lambda_0 + \phi \cdot j, \quad j = 0, 1, 2, ... \tag{S7}$$

On the other hand, the tested trend in Mann-Kendall test is defined as follows (Mann, 1945): "*The hypothesis of a downward trend may be defined in the following way: The sample is still a random sample but X_i has the continuous cumulative distribution function F_i and $F_i(X) < F_{i+k}(X)$ for every i, every X, and every $k > 0$. An upward trend is similarly defined with* $F_i(X) > F_{i+k}(X)$". In this case, the property of concern is the so-called stochastic ordering and stochastic dominance. Thus, trends correspond to well-defined rules implying that the distribution of X at time step i stochastically dominates the distribution of X at any subsequent time $i+k$, or vice versa. In other words, it assumes that stochastic dominance is on play, and it depends on time. In this case $\mathrm{G}[X_t]$ is a precise rule, that is, systematic stochastic dominance over time. Since stochastic dominance often results in monotonic patterns of central tendency measures, Mann-Kendall test is often (but incorrectly) referred to as a test for 'trends in the median/mean values' (see Serinaldi and Kilsby, 2016; Serinaldi et al., 2018, for further discussion).

Therefore, different statistical methods and tests rely on their own definition of 'trend' that refer to population properties and not to sample properties. When tests are well devised, such definitions are clearly stated as they are necessary to set up a suitable and coherent test statistic and derive its properties. However, end-users often neglect the different definitions of trend corresponding to different tests/methods, which become just algorithms to be run over large data sets. The consequence is not only merging methods that are devised to answer different questions (without recognizing such a difference) but also logical paradoxes. For example, focusing on the analysis in the paper, some of the NHP distributions (fitted under the assumption of non-stationarity) show negative slope ϕ, meaning decreasing rate of occurrence. Some of these models yield λ equal to zero in few decades. This has very practical and inconvenient consequences:

- Decreasing λ entails decreasing variance, which depends on time as well. This means that a single population variance does not exist, and autocovariance and autocorrelation are ill-defined as well. In other words, sample ACF is as meaningless as the sample mean under the assumption that data come from a Cauchy distribution, for instance. Therefore, all ACFs or lag-1 ACF terms reported in the paper and in Farris et al. (2021) under the assumption of non-stationarity do not correspond to any population counterpart. They are just numbers that do not allow any inference because the supposed inferred population counter part does not even exist.

- Such models hinder the calculation of any index or summary statistics requiring integration over time, such as return periods and return levels. Indeed, if the process does no longer exist after e.g. 80 years (because the rate λ becomes equal to zero after 80 years), calculating return levels over longer return periods is just meaningless.

Moving from formal definitions to applications, the keyword in trend analysis of hydro-climatic data, and more generally in the application of statistical tests, is 'repeatability'. A key aspect of science is experimentation: a result is scientifically sound if it can be replicated. Hydro-climatic processes cannot be repeated; however, some of them repeat themselves in time over time scales that allow multiple observations of the same phenomenon. In these cases, we implicitly or explicitly replace multiple independent experiments with multiple observations of the same process in time. In this respect, seasonality is a trend as the distribution of a hydro-climatic variable is assumed to be a function of calendar days, or months, or similar. The difference between seasonal trends and ‘monotonic trends’ spanning the record length is ‘repeatability’: the formers have been observed many times (let’s say, Earth and Sun are so kind to repeat the experiment for us every year), whereas ‘monotonic trends’ are not. The latter are unique. Making inference on them is like assessing if a die is loaded by throwing it just once.

Therefore, when dealing with ‘trends’, we should consider three aspects at least:

- Context: type of process we are dealing with, and we want to model (e.g., hydrological, biological, financial, etc.).
- Formal definition of ‘trend’: it refers to population properties (not sampling fluctuations) and is required to develop any sound test or model. If such a definition is missing, the resulting methodology might be formally meaningless and useless, and numerical outputs are just nonsense numbers (see e.g., Serinaldi and Kilsby, 2016; Serinaldi et al., 2018, 2020a, 2022b, for examples of flawed algorithms).
- Type of data: experimental (and repeatable) or unrepeatable.

S4 Iterative Amplitude Adjusted Fourier Transform (IAAFT) and bias adjustment of power spetrum estimates

The Iterative Amplitude Adjusted Fourier Transform (IAAFT) is a simulation technique belonging to the class of Fourier Transform (FT) methods, which have been widely used in several disciplines to generate time series with desired properties (see e.g., Theiler et al., 1992; Schreiber and Schmitz, 2000; Venema et al., 2006; Maiwald et al., 2008; Keylock, 2010; Serinaldi and Lombardo, 2017; Lancaster et al., 2018; Serinaldi et al., 2022a). In particular, IAAFT allows the simulation of synthetic time series that preserve the empirical marginal distribution and, to some error level, the empirical power spectrum of the original data. For a given discrete-time process and regular time intervals, $\{z_j\}_{j=0}^{n-1}$, where n is the sample size, the discrete FT is:

$$\begin{aligned}\zeta_k = \mathcal{F}_k[\{z_j\}_{j=0}^{n-1}] =& \sum_{j=0}^{n-1} z_j \cdot \left[\cos\left(\frac{2\pi}{n}jk\right) - \mathrm{i}\sin\left(\frac{2\pi}{n}jk\right)\right] \\ =& \sum_{j=0}^{n-1} z_j \cdot \mathrm{e}^{-\mathrm{i}\frac{2\pi}{n}jk} \\ =& A_k \mathrm{e}^{\mathrm{i}\varphi_k},\end{aligned} \tag{S8}$$

where ζ_k is the k-th sinusoid component of the FT of $\{z_j\}$, $\mathrm{i} = \sqrt{-1}$, $A_k = \left|\sum_{j=0}^{n-1} z_j \cdot \mathrm{e}^{-\mathrm{i}\frac{2\pi}{n}jk}\right| = \sqrt{[\mathrm{Re}(\zeta_k)]^2 + [\mathrm{Im}(\zeta_k)]^2}$ are the Fourier amplitudes, and $\varphi_k = \tan^{-1}[\mathrm{Im}(\zeta_k)/\mathrm{Re}(\zeta_k)]$ are the phases (or phase angles). Since A_k^2 are the power spectrum values, a synthetic time series preserving the power spectrum can be generated by randomizing the phases. Phase randomization works as follows: the phases φ_k are replaced by random values $\tilde{\varphi}_k$ ranging in $[0, 2\pi)$, then a phase-randomized FT is created as $\tilde{\zeta}_k = A_k \mathrm{e}^{\mathrm{i}\tilde{\varphi}_k}$, and finally a synthetic time series is given by the inverse discrete FT

$$\begin{aligned}\tilde{z}_j =& \mathcal{F}_j^{-1}[\{A_k \mathrm{e}^{\mathrm{i}\tilde{\varphi}_k}\}_{k=0}^{n-1}] \\ =& \mathcal{F}_j^{-1}[\{\tilde{\zeta}_k\}_{k=0}^{n-1}] \\ =& \frac{1}{n}\sum_{k=0}^{n-1} \tilde{\zeta}_k \cdot \mathrm{e}^{\mathrm{i}\frac{2\pi}{n}jk}.\end{aligned} \tag{S9}$$

While the power spectrum of $\tilde{z}_j$ is equal to that of z_j by construction, phase randomization yields a marginal distribution different from the observed one. Therefore, the generated new values $\tilde{x}_j$ are replaced by the values in the original time series with the same rank (i.e. the same position in the time series sorted in ascending or descending order), according to a rank-order matching procedure (e.g. Schreiber and Schmitz, 2000). As this replacement modifies the power spectrum for the synthetic series, the procedure is repeated starting from a new sequence $\tilde{\zeta}_k$ that is built using the original amplitudes A_k and the phases resulting from the last iteration. Iterations stop when a convergence criterion is satisfied (Schreiber and Schmitz, 1996; Kugiumtzis, 1999; Keylock, 2012).

IAAFT relies on the power spectrum estimated by the periodogram via FT. For small/finite sample sizes such as those of Z data, the periodogram is known to be one of the most biased estimators of linear dependence properties among other estimators such as correlogram and climacogram for small sample sizes (Stoica et al., 2005; Koutsoyiannis, 2010; Dimitriadis and Koutsoyiannis, 2015). However, the definition the power spectrum as the Fourier transform of the autocovariance does not allow derivation of an analytical formula for the estimation bias. Moreover, bias adjustment depends on the underlying model assumed to describe the temporal dependence structure.

Following Iliopoulou and Koutsoyiannis (2019), we use the fractional Gaussian noise (fGn), which is also known as Hurst-Kolmogorov (HK) process (Koutsoyiannis, 2010). The fGn process is characterized by the Hurst coefficient $H \in (0,1)$, where $H = 0.5$ corresponds to independence. For fGn processes, bias adjusted estimators of variance, and lag k covariance and autocorrelation are (Koutsoyiannis, 2003)

$$\tilde{s}^2 := \frac{n-1}{n-n^{2H-1}} s^2, \tag{S10}$$

$$\tilde{g}_k := g_k + \frac{n-1}{n^{3-2H}-n} s^2, \tag{S11}$$

and

$$\tilde{r}_k := \frac{\tilde{g}_k}{\tilde{s}_k^2} = r_k \left(1 - \frac{1}{n^{2-2H}}\right) + \frac{1}{n^{2-2H}}, \tag{S12}$$

where n is the sample size, while s^2, g_k, and r_k denote the standard estimators of variance, and lag k covariance and autocorrelation under independence, respectively. Here, H is estimated by the 'least squares based on variance' (LSV) method proposed by Tyralis and Koutsoyiannis (2011).

For the the fGn (or HK) process, Dimitriadis and Koutsoyiannis (2015, Table 3, and Table 1 in Supplementary material) derived a semi-analytical formula for the estimation bias of the power spectrum (through the definition of the autocovariance) based on the climacogram. In this study, we exploit the relationship between autocovariance and power spectrum to adjust the autocovariance for finite-sample bias (via Eq. S11) and then beck-transforming it to power spectrum.

Figures S1a and S1b show the climacograms of two randomly selected Z time series highlighting the effect of bias adjustment, while Figures S1c and S1d show the observed time series along with one simulated time series for each case, which is displayed for illustration purpose. The selected time series are representative of the two extreme cases of weak and strong dependence, and their effect on bias correction.

S5 Maps of PR and MK rejections

Figures S2, S3 and S4 show the maps of statistically significant trends at the GHCN gauges of the three regions North America (a), Eurasia (b), and Australia (c). Figure S2 refers to MK and PR tests applied to Z time series for 50-year sample size and the 95% threshold without FDR. Figure S3 refers to time series for 100-year sample size and the 99.5% threshold without FDR, while Figure S4 to 50-year sample size and the 99.5% threshold without FDR. Figures S5, S6 and S7 correspond to Figures S2, S3 and S4, respectively, but with FDR.

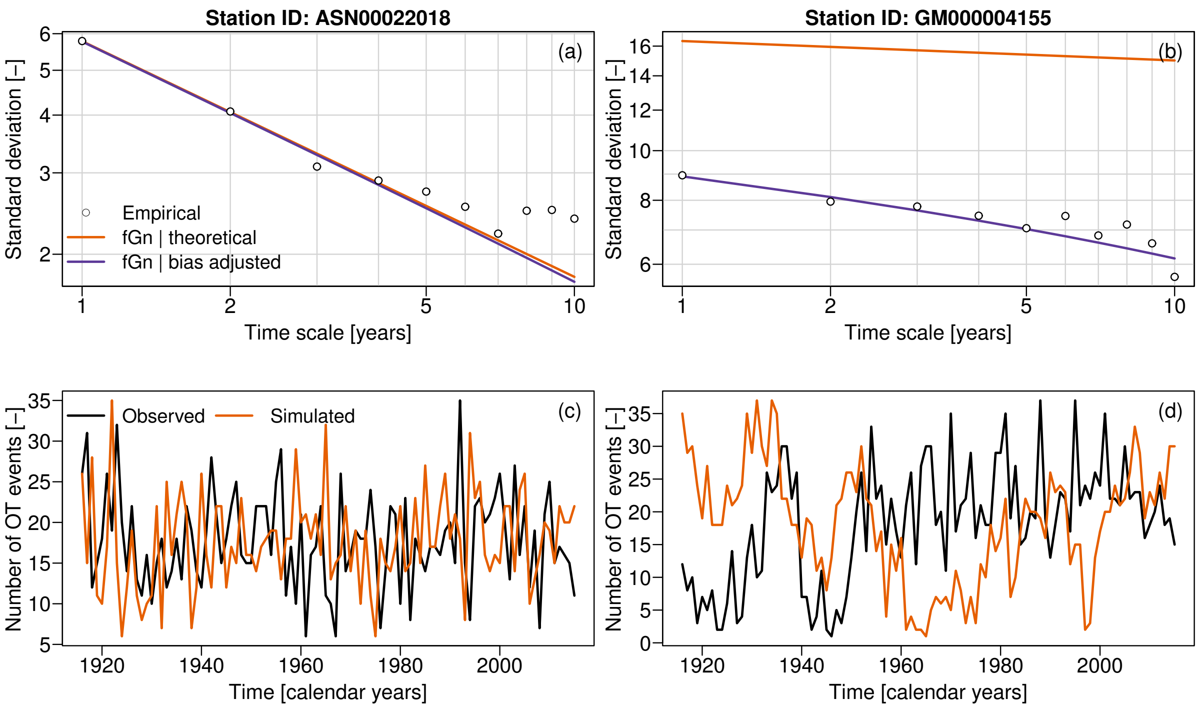

Figure S1. (a-b) Climacograms of two randomly selected Z time series highlighting the effect of bias adjustment. (c-d) Observed time series and one simulated time series shown for illustration purpose.

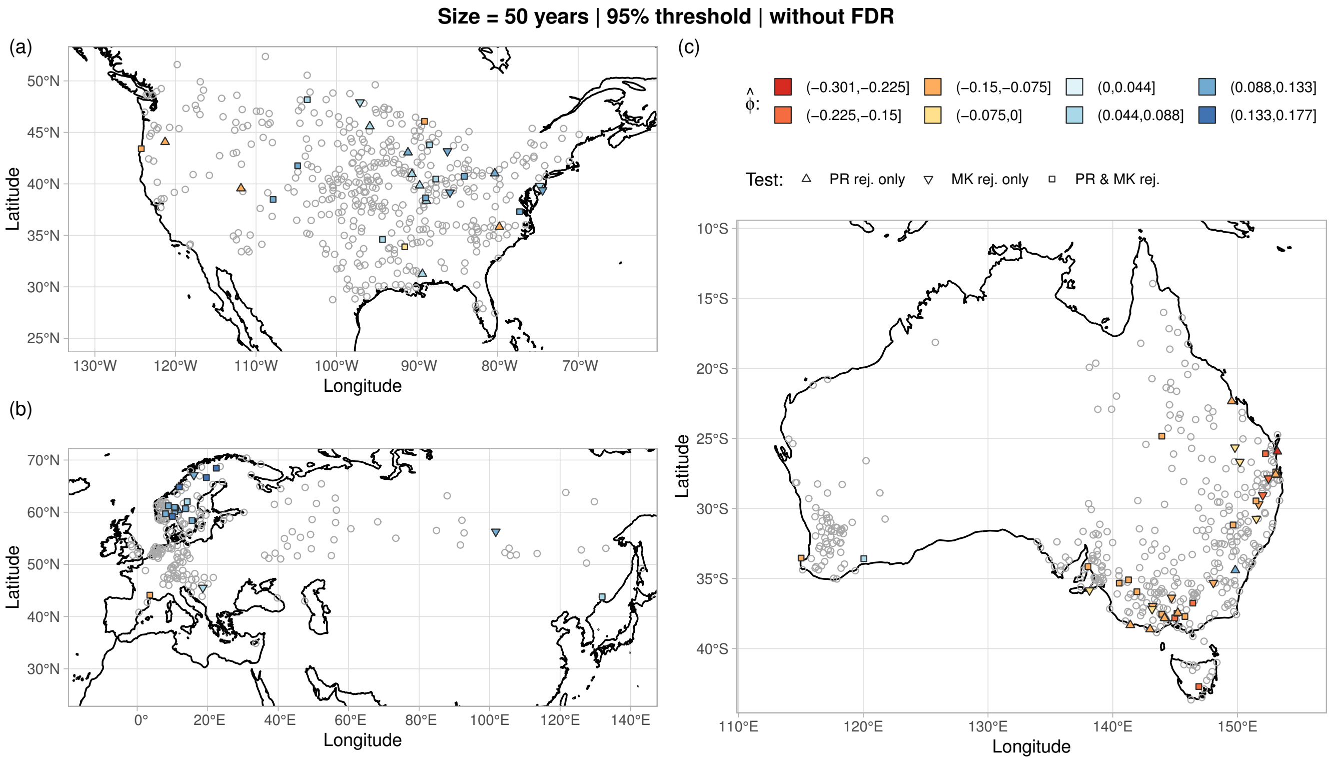

Figure S2. Maps of statistically significant trends at the GHCN gauges of the three regions North America (a), Eurasia (b), and Australia (c). Results refer to MK and PR tests applied to Z time series for 50-year sample size and the 95% threshold without FDR. Statistical tests are performed at the local 5% significance level without applying FDR. The distributions of test statistics (and therefore critical values) are estimated from 10,000 IAAFT samples. Gray circles '○' denote lack of rejection by both tests.

Size = 100 years | 99.5% threshold | without FDR

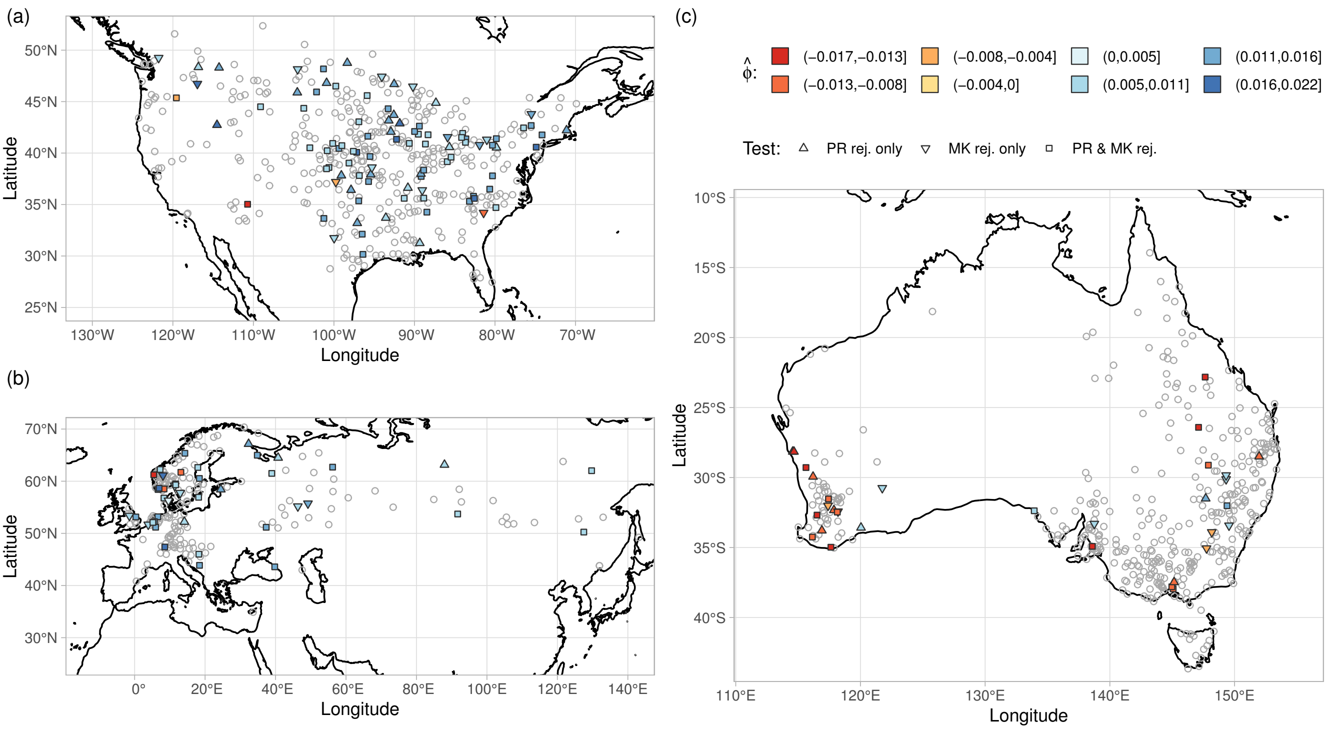

Figure S3. Similar to Figure S1, but for 100-year sample size, and the 99.5% threshold.

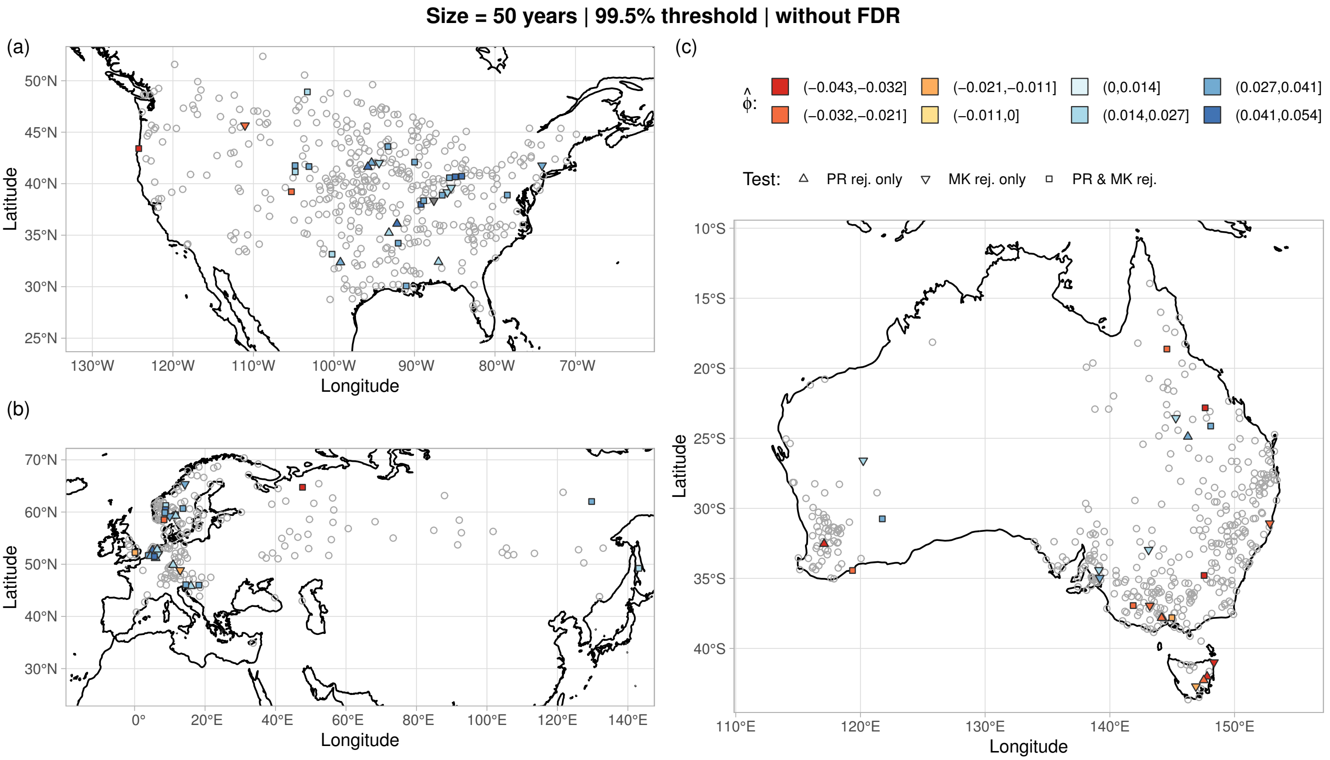

Figure S4. Similar to Figure S1, but for 50-year sample size, and the 99.5% threshold.

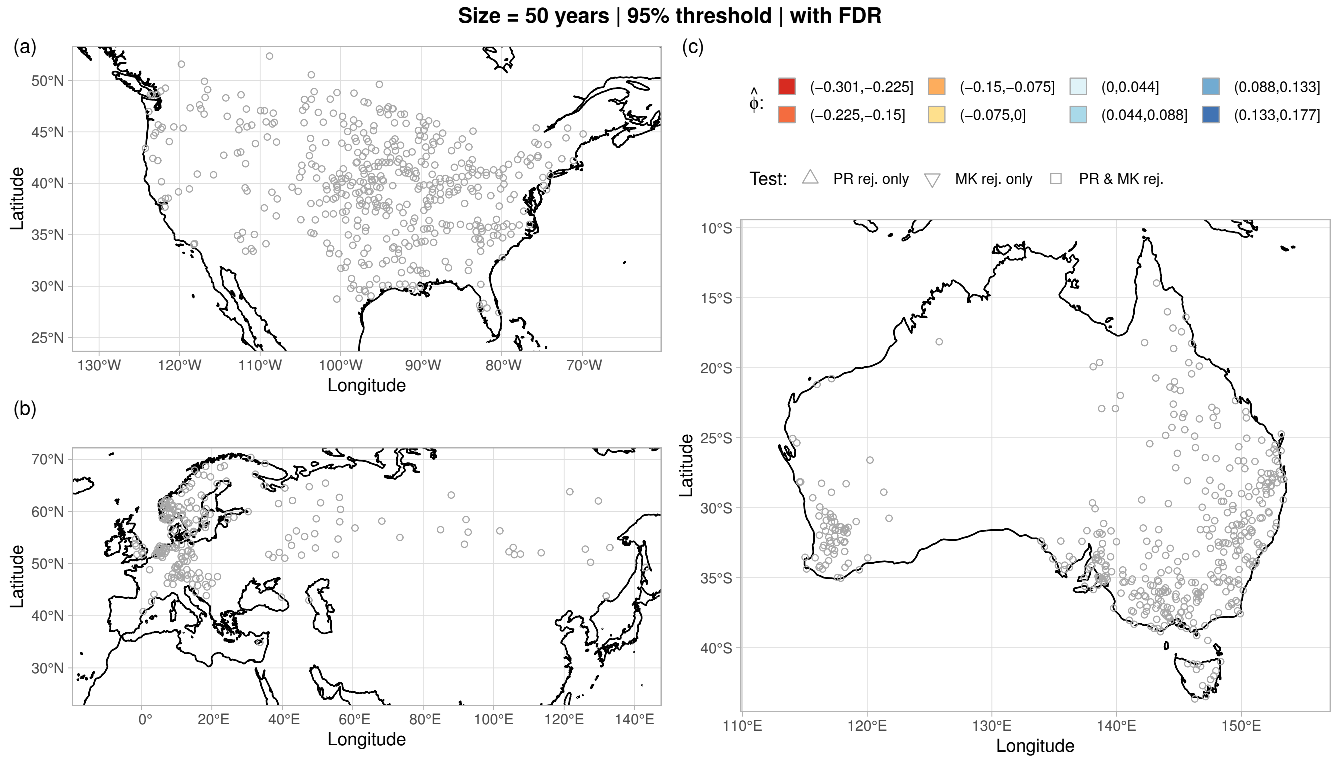

Figure S5. Similar to Figure S1, but for 50-year sample size, and the 95% threshold with FDR.

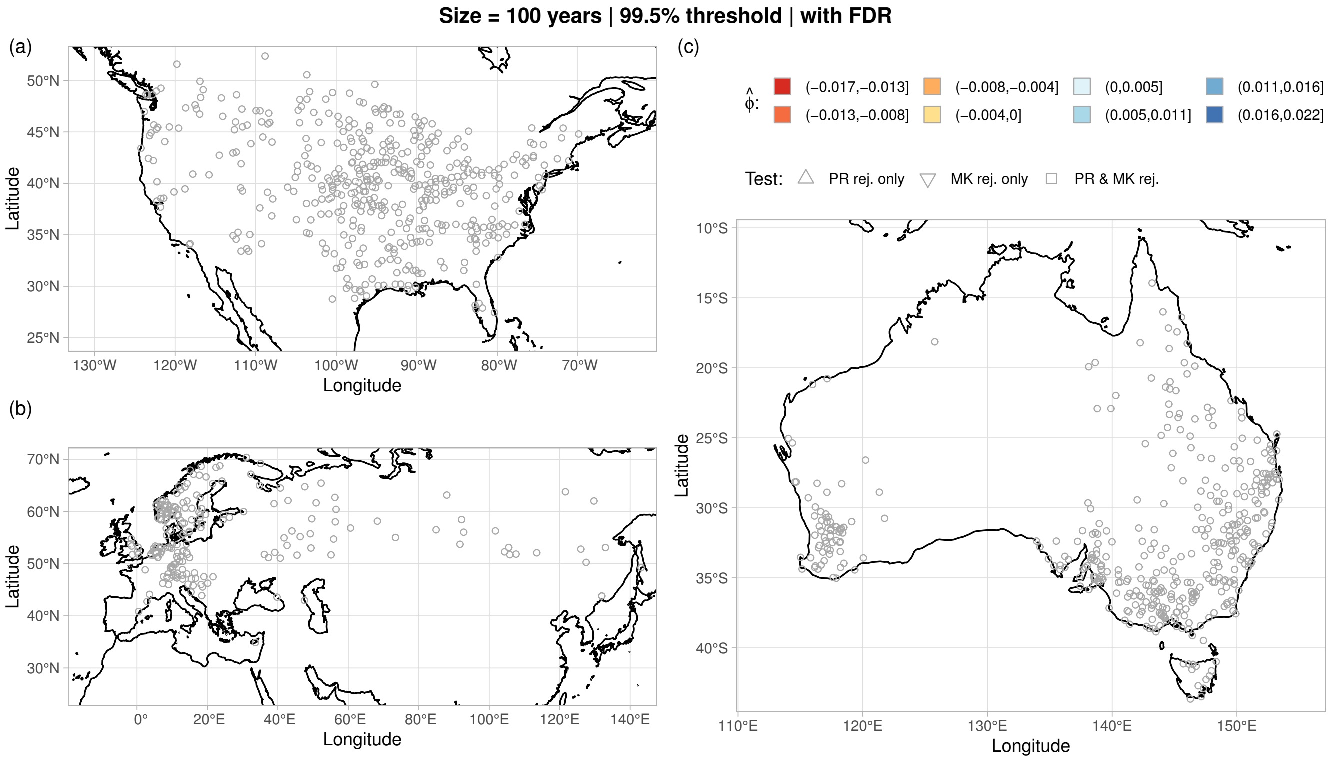

Figure S6. Similar to Figure S1, but for 100-year sample size, and the 99.5% threshold with FDR.

Size = 50 years | 99.5% threshold | with FDR

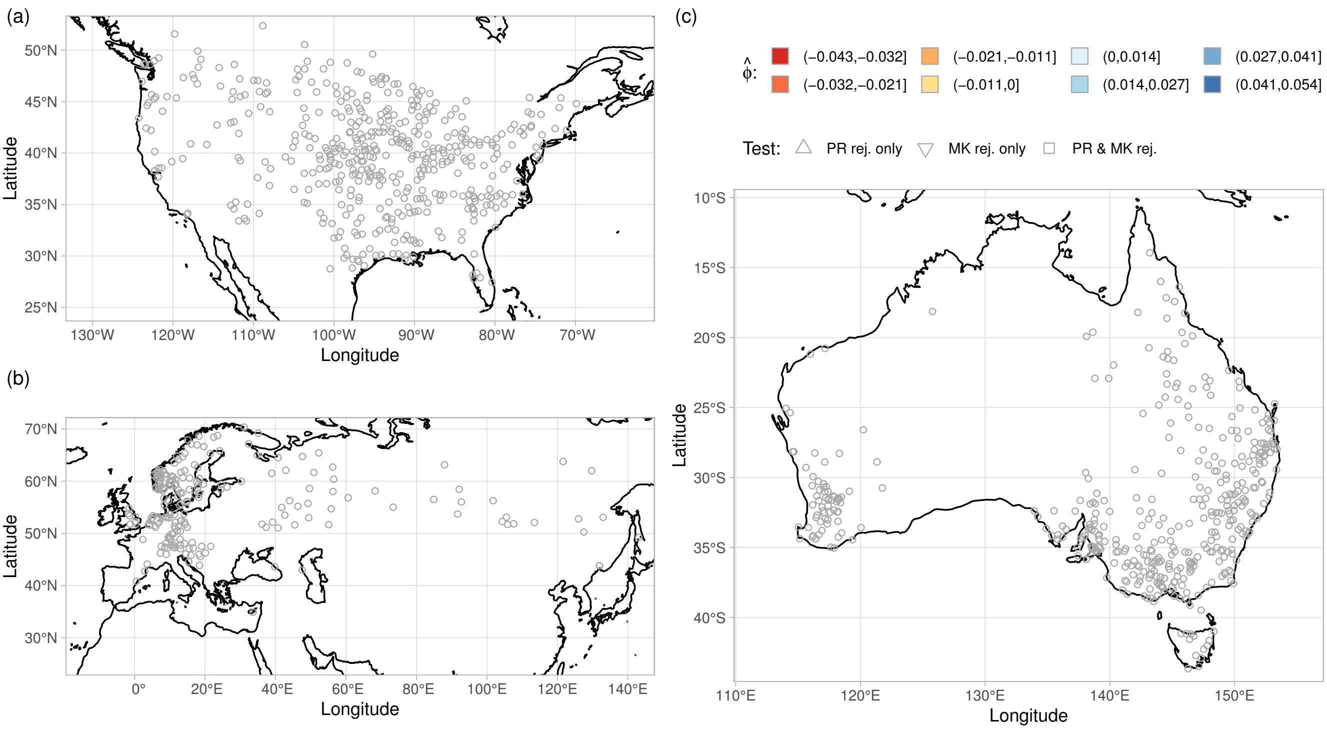

Figure S7. Similar to Figure S1, but for 50-year sample size, and the 99.5% threshold with FDR.

References

Ahn, H. and Chen, J. J.: Generation of Over-Dispersed and Under-Dispersed Binomial Variates, Journal of Computational and Graphical Statistics, 4, 55–64, 1995.

Dimitriadis, P. and Koutsoyiannis, D.: Climacogram versus autocovariance and power spectrum in stochastic modelling for Markovian and Hurst–Kolmogorov processes, Stochastic environmental research and risk assessment, 29, 1649–1669, 2015.

Farris, S., Deidda, R., Viola, F., and Mascaro, G.: On the Role of Serial Correlation and Field Significance in Detecting Changes in Extreme Precipitation Frequency, Water Resources Research, 57, e2021WR030 172, 2021.

Hughes, G. and Madden, L.: Using the Beta-Binomial distribution to describe aggegated patterns of disease incidence, Phytopathology, 83, 759–763, 1993.

Iliopoulou, T. and Koutsoyiannis, D.: Revealing hidden persistence in maximum rainfall records, Hydrological Sciences Journal, 64, 1673–1689, https://doi.org/10.1080/02626667.2019.1657578, 2019.

Keylock, C. J.: Characterizing the structure of nonlinear systems using gradual wavelet reconstruction, Nonlinear Processes in Geophysics, 17, 615–632, 2010.

Keylock, C. J.: A resampling method for generating synthetic hydrological time series with preservation of cross-correlative structure and higher-order properties, Water Resources Research, 48, W12 521, 2012.

Koutsoyiannis, D.: Climate change, the Hurst phenomenon, and hydrological statistics, Hydrological Sciences Journal, 48, 3–24, 2003.

Koutsoyiannis, D.: HESS Opinions "A random walk on water", Hydrology and Earth System Sciences, 14, 585–601, 2010.

Koutsoyiannis, D. and Montanari, A.: Negligent killing of scientific concepts: the stationarity case, Hydrological Sciences Journal, 60, 1174–1183, 2015.

Kugiumtzis, D.: Test your surrogate data before you test for nonlinearity, Physical Review E, 60, 2808–2816, 1999.

Lancaster, G., Iatsenko, D., Pidde, A., Ticcinelli, V., and Stefanovska, A.: Surrogate data for hypothesis testing of physical systems, Physics Reports, 748, 1–60, 2018.

Maiwald, T., Mammen, E., Nandi, S., and Timmer, J.: Surrogate data – A qualitative and quantitative analysis, in: Mathematical Methods in Signal Processing and Digital Image Analysis, edited by Dahlhaus, R., Kurths, J., Maass, P., and Timmer, J., pp. 41–74, Springer, Berlin Heidelberg, 2008.

Mann, H. B.: Nonparametric Tests Against Trend, Econometrica, 13, 245–259, 1945.

Nicola, V. F. and Goyal, A.: Modeling of correlated failures and community error recovery in multiversion software, IEEE Transactions on Software Engineering, 16, 350–359, 1990.

Schreiber, T. and Schmitz, A.: Improved Surrogate Data for Nonlinearity Tests, Physical Review Letters, 77, 635–638, 1996.

Schreiber, T. and Schmitz, A.: Surrogate time series, Physica D: Nonlinear Phenomena, 142, 346–382, 2000.

Serinaldi, F. and Kilsby, C. G.: The importance of prewhitening in change point analysis under persistence, Stochastic Environmental Research and Risk Assessment, 30, 763–777, 2016.

Serinaldi, F. and Kilsby, C. G.: Unsurprising Surprises: The Frequency of Record-breaking and Overthreshold Hydrological Extremes Under Spatial and Temporal Dependence, Water Resources Research, 54, 6460–6487, 2018.

Serinaldi, F. and Lombardo, F.: General simulation algorithm for autocorrelated binary processes, Physical Review E, 95, 023 312, 2017.

Serinaldi, F., Kilsby, C. G., and Lombardo, F.: Untenable nonstationarity: An assessment of the fitness for purpose of trend tests in hydrology, Advances in Water Resources, 111, 132–155, 2018.

Serinaldi, F., Chebana, F., and Kilsby, C. G.: Dissecting innovative trend analysis, Stochastic Environmental Research and Risk Assessment, 34, 733–754, 2020a.

Serinaldi, F., Lombardo, F., and Kilsby, C. G.: All in order: Distribution of serially correlated order statistics with applications to hydrological extremes, Advances in Water Resources, 144, 103 686, 2020b.

Serinaldi, F., Briganti, R., Kilsby, C. G., and Dodd, N.: Sailing synthetic seas: Stochastic simulation of benchmark sea state time series, Coastal Engineering, 176, 104 164, 2022a.

Serinaldi, F., Lombardo, F., and Kilsby, C. G.: Testing tests before testing data: an untold tale of compound events and binary dependence, Stochastic Environmental Research and Risk Assessment, 36, 1373–1395, 2022b.

Skellam, J. G.: A probability distribution derived from the binomial distribution by regarding the probability of success as variable between the sets of trials, Journal of the Royal Statistical Society. Series B (Methodological), 10, 257–261, 1948.

Stoica, P., Moses, R. L., et al.: Spectral analysis of signals, Pearson Prentice Hall, Upper Saddle River, New Jersey, 2005.

Theiler, J., Eubank, S., Longtin, A., Galdrikian, B., and Doyne Farmer, J.: Testing for nonlinearity in time series: the method of surrogate data, Physica D: Nonlinear Phenomena, 58, 77–94, 1992.

Tsai, C.-A., Hsueh, H.-m., and Chen, J. J.: Estimation of False Discovery Rates in Multiple Testing: Application to Gene Microarray Data, Biometrics, 59, 1071–1081, 2003.

Tyralis, H. and Koutsoyiannis, D.: Simultaneous estimation of the parameters of the Hurst–Kolmogorov stochastic process, Stochastic Environmental Research and Risk Assessment, 25, 21–33, 2011.

Venema, V., Bachner, S., Rust, H. W., and Simmer, C.: Statistical characteristics of surrogate data based on geophysical measurements, Nonlinear Processes in Geophysics, 13, 449–466, 2006.