Supplement of

Impact of spatio-temporal dependence on the frequency of precipitation extremes: Negligible or neglected?

Francesco Serinaldi[1,2]

[1]School of Engineering, Newcastle University, Newcastle Upon Tyne, NE1 7RU, UK
[2]Willis Research Network, 51 Lime St., London, EC3M 7DQ, UK

Correspondence: 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 occurence $\mu = (1-\rho_1)\cdot\lambda$, 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 $p = \alpha/(\alpha+\beta)$, 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 Var[Ψ] = 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., 2020)

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

where $\mathbb{V}$ is the expectation operator, $\rho_{jl} = \mathbb{C}[Y_j, Y_l]$ denotes the pairwise correlation of 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 element of the $q = n\cdot m$ space-time correlation matrix and Eq. S4 reads as

$$\rho_{\beta\mathcal{B}} = \frac{\sum\sum_{ji\neq lk}\rho_{ji,lk}}{q(q-1)}. \tag{S5}$$

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), 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 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., 2022). 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:

$$\begin{aligned}\zeta_k = \mathcal{F}_k[\{z_j\}_{j=0}^{n-1}] =& \sum_{j=0}^{n-1} z_j \cdot \left[\cos\left(\frac{2\pi}{n}jk\right) - \mathrm{i}\sin\left(\frac{2\pi}{n}jk\right)\right] \\ =& \sum_{j=0}^{n-1} z_j \cdot \mathrm{e}^{-\mathrm{i}\frac{2\pi}{n}jk} \\ =& A_k \mathrm{e}^{\mathrm{i}\varphi_k}, \end{aligned} \tag{S6}$$

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

$$\begin{aligned}\tilde{z}_j =& \mathcal{F}_j^{-1}[\{A_k \mathrm{e}^{\mathrm{i}\tilde{\varphi}_k}\}_{k=0}^{n-1}] \\ =& \mathcal{F}_j^{-1}[\{\tilde{\zeta}_k\}_{k=0}^{n-1}] \\ =& \frac{1}{n}\sum_{k=0}^{n-1} \tilde{\zeta}_k \cdot \mathrm{e}^{\mathrm{i}\frac{2\pi}{n}jk}. \end{aligned} \tag{S7}$$

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. In this study, we exploit the relationship between autocovariance and power spectrum to adjust the autocovariance for finite-sample bias and then beck-transforming it to power spectrum. 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)

$$\tilde{s}^2 := \frac{n-1}{n - n^{2H-1}} s^2, \tag{S8}$$

$$\tilde{g}_k := g_k + \frac{n-1}{n^{3-2H} - n} s^2, \tag{S9}$$

and

$$\tilde{r}_k := \frac{\tilde{g}_k}{\tilde{s}_k^2} = r_k \left(1 - \frac{1}{n^{2-2H}}\right) + \frac{1}{n^{2-2H}}, \tag{S10}$$

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). 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.

S4 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.

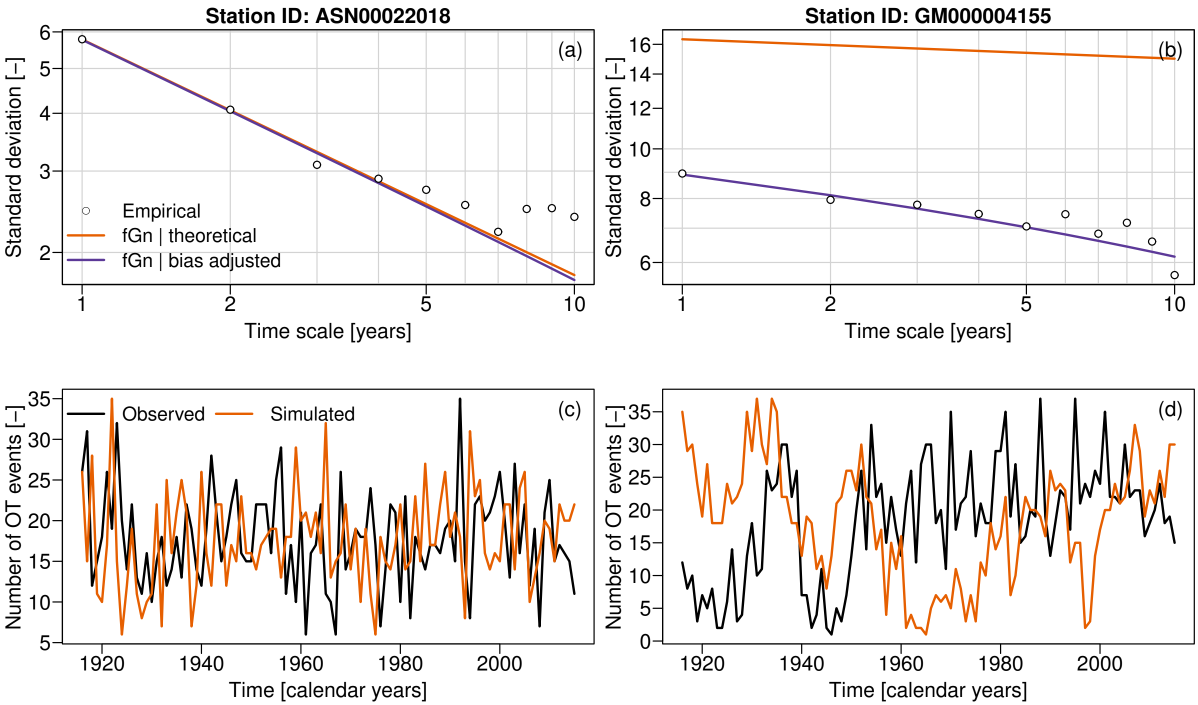

Figure S1. (a-b) Climacograms of two randomly selected Z time series highlighting the effect of bias adjustment. (c-d) Observed time series and one simulated time series shown for illustration purpose.

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.

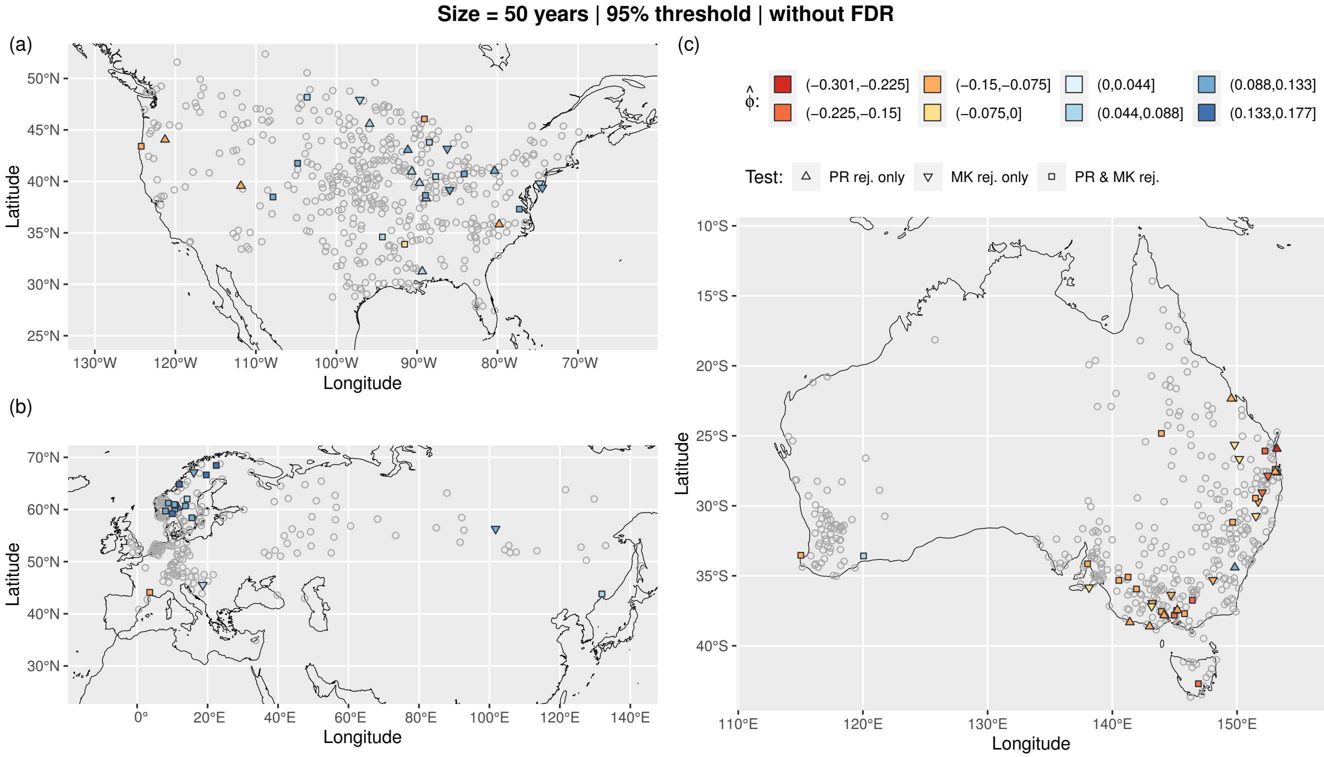

Figure S2. Maps of statistically significant trends at the GHCN gauges of the three regions North America (a), Eurasia (b), and Australia (c). Results refer to MK and PR tests applied to Z time series for 50-year sample size and the 95% threshold without FDR. Statistical tests are performed at the local 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.

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.

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.

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.

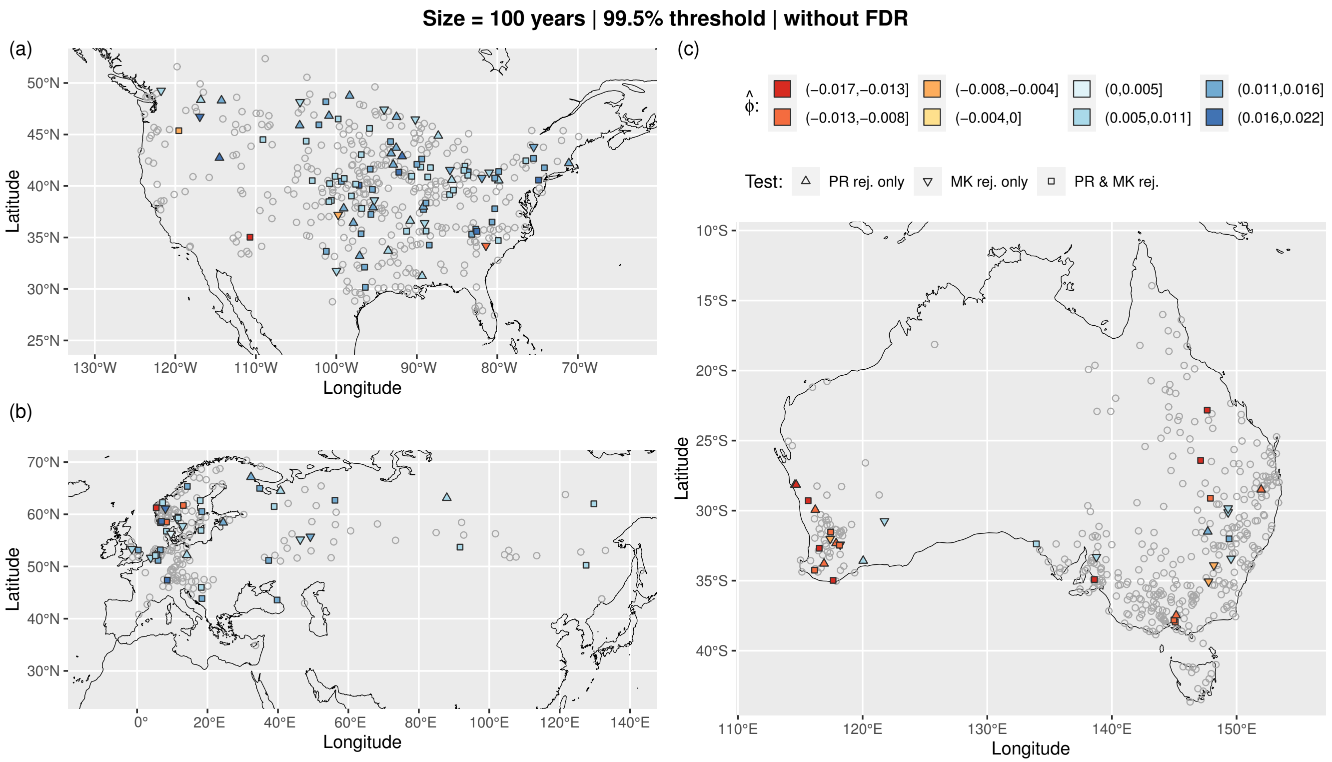

Figure S3. Similar to Figure S1, but for 100-year sample size, and the 99.5% threshold.

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., 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, 2020.

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, 2022.

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.

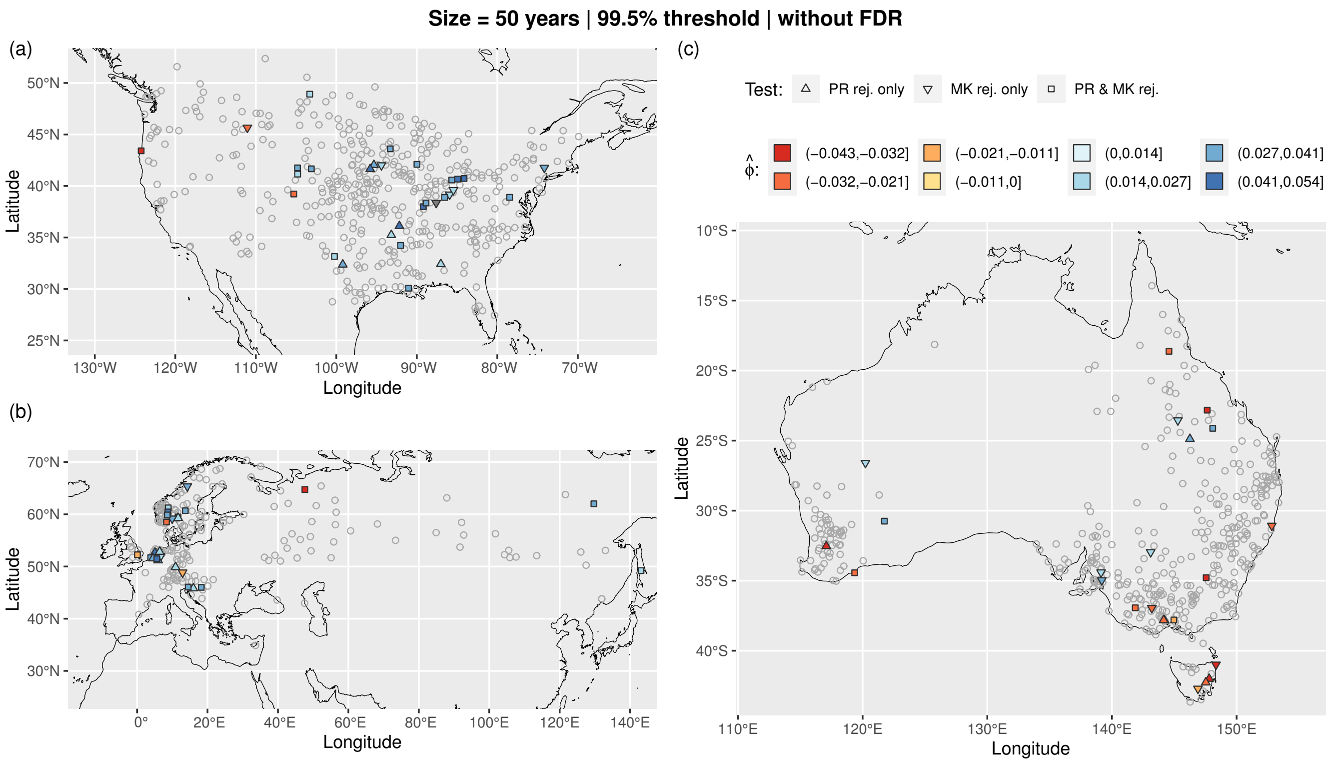

Figure S4. Similar to Figure S1, but for 50-year sample size, and the 99.5% threshold.

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.

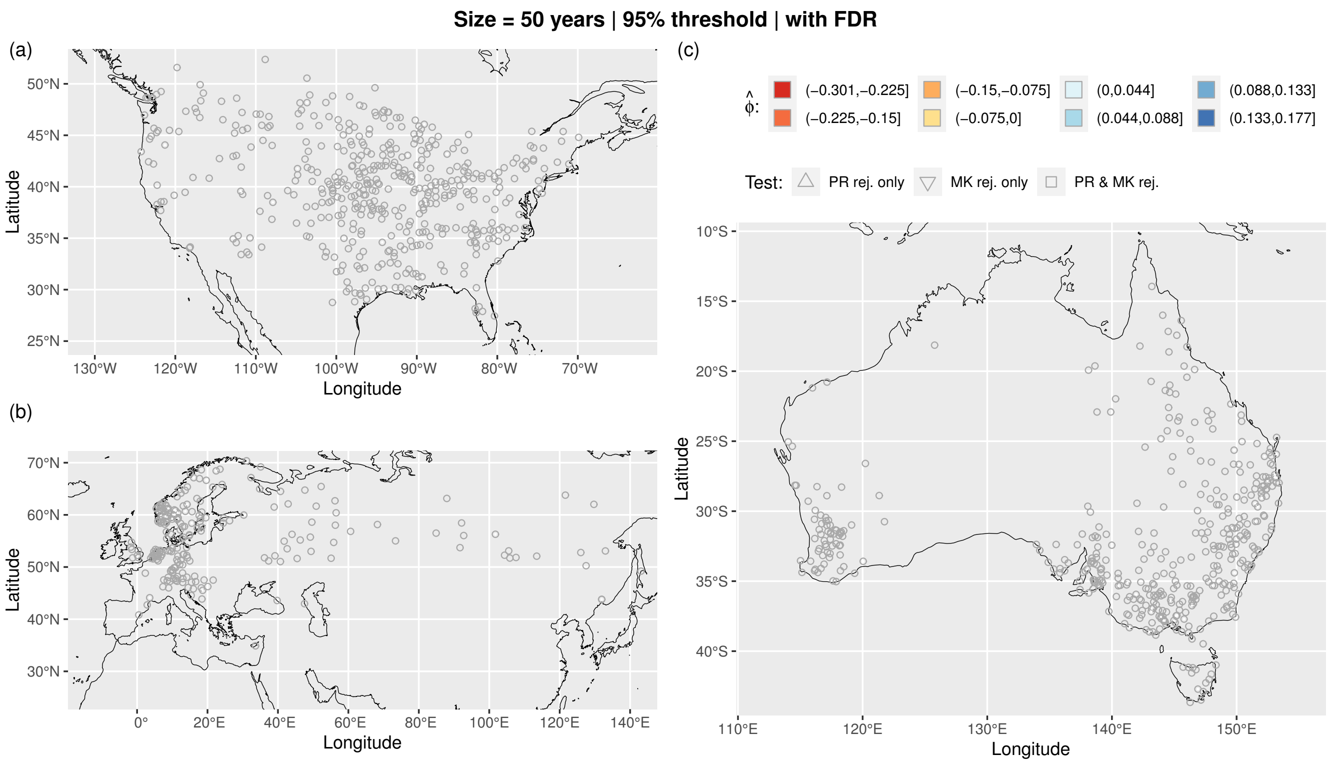

Figure S5. Similar to Figure S1, but for 50-year sample size, and the 95% threshold with FDR.

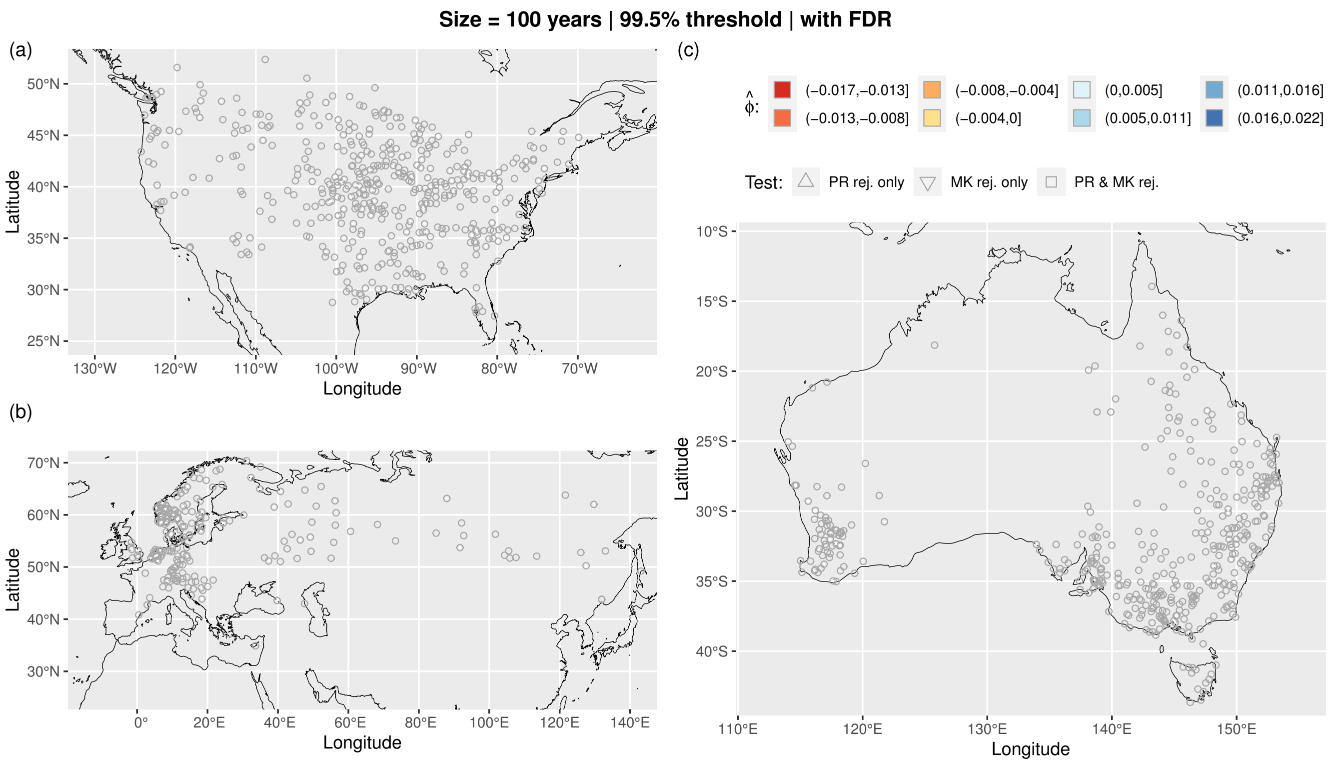

Figure S6. Similar to Figure S1, but for 100-year sample size, and the 99.5% threshold with FDR.

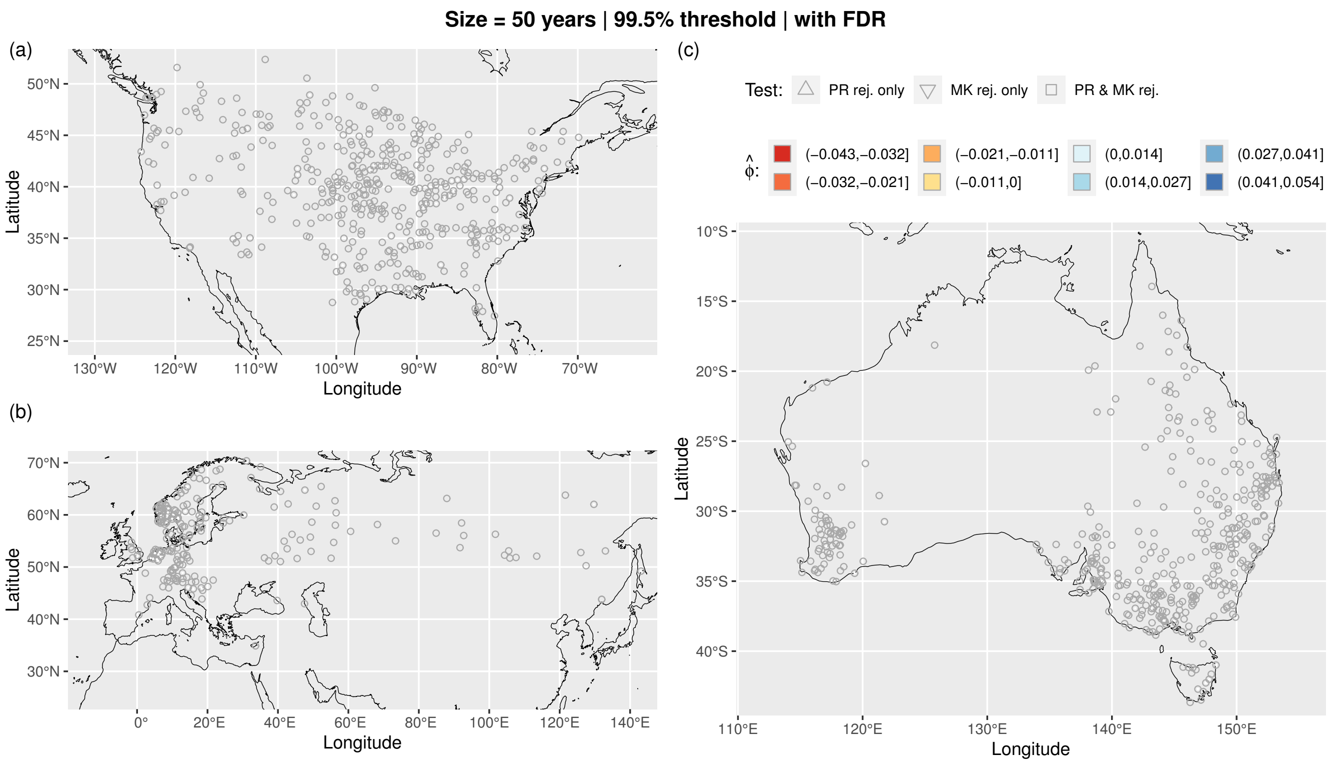

Figure S7. Similar to Figure S1, but for 50-year sample size, and the 99.5% threshold with FDR.