Hydrol. Earth Syst. Sci. Discuss.,
https://doi.org/10.5194/hess-2019-142-SC2, 2019

Hydrology and Earth System Sciences Discussions

Open Access

HESSD

Interactive comment

Interactive comment on "Technical note: Stochastic simulation of streamflow time series using phase randomization" *by* Manuela I. Brunner et al.

Ioannis Tsoukalas

itsoukal@mail.ntua.gr

Received and published: 19 April 2019

I enjoyed reading this paper, which aims to address a challenging and critical task for many water resources studies, that of synthetic data generation. I also appreciated the availability of the PRSim R package. This is an additional short comment continuing the interesting comment posed by F. Serinaldi regarding the preservation of the ACF/CCF. As the Authors recognize, and F. Serinaldi mentions in his comment, the ACF/CCF in the Gaussian domain and the ACF/CCF in the actual domain (i.e., that of the distribution function), typically differ. In my view, this is mainly due to two reasons:

Printer-friendly version

Discussion paper

1. The final process (the one obtained after the use of the quantile function; i.e., the inverse cumulative distribution function; ICDF) results from the mapping of a Gaussian one (a concept related with the so-called Nataf's joint distribution model (Mardia, 1970; Nataf, 1962) – see below)

2. The use of Pearson's correlation coefficient to express the dependence structure (ACF/CCF) of the process.

Within hydrological domain, beyond the work of Papalexiou (2018) (also mentioned by F. Serinaldi), this mapping procedure has been also employed for the simulation of non-Gaussian univariate and multivariate processes in the works of Tsoukalas et al. (2019, 2018a, 2018b, 2017) and Tsoukalas (2018; for a Thesis-length treatment on the subject) which adopt the term "Nataf-based processes", and also discuss similarities with other approaches in hydrology and beyond (see also the work of Serinaldi and Lombardo (2017) for univariate binary processes). For instance, the work of Liu and Der Kiureghian (1986) and Lebrun and Dutfoy (2009), that regard the Nataf's joint distribution model per se – which in my understanding regards the core idea of the above mapping procedure (initially proposed for correlated random variables, and not processes).

Interestingly, and regarding processes, the concept of Nataf's joint distribution can be traced back in the early work of Grigoriu (1984) and later in sequel works, who used it to establish non-Gaussian stochastic process, the so-called "translation processes". Similar models for non-Gaussian stochastic processes have been proposed by Cario and Nelson (1996) and Biller and Nelson (2003), the so-called "To-Anything" models. More specifically, this type of models use (low-order) AR models to simulate an auxiliary (appropriately "inflated/adjusted") Gaussian process (Gp), which after its mapping (through the ICDF) to the actual domain, results into a process with the target distribution and correlation structure. Actually, it is noted, that any linear stochastic model (e.g., AR, MA, ARMA or other) can be used within such simulation schemes to simulate the auxiliary Gp (Tsoukalas et al., 2018b).

Regarding the preservation of the ACF/CCF from such methods (i.e., relying on this mapping procedure), it is recalled that the Pearson's correlation is a linear measure of dependence that uses first cross-product moments of the process, and is not invariant under non-linear monotonic transformations (such as those imposed by the ICDF). Please, also refer to section 4 in Tsoukalas et al. (2018c), and section 3.2.3 in Tsoukalas et al. (2018b), who highlight a delicate point related with the above mapping procedure and the use of alternative rank-based dependence measures (i.e., Kendall's tau and Spearman's rho). Further information can be found in the book of Embrechts et al. (1999), while a discussion related with this property, in the context of similarly-constructed stochastic processes for hydrological time series generation, is given in Tsoukalas et al. (2018a, 2018b).

As noted in F. Serinaldi's comment, and also discussed by others (e.g., Koutsoyiannis, 2016, 2000; Papoulis, 1991 pp. 118) the power spectrum and the ACF (or CCF) of a process are interrelated quantities. Hence, it is reasonably to expect that a spectrum-based simulation method (such as the one proposed by the Authors) will inherit the properties of Pearson's correlation coefficient. To elaborate, let Z_t be a univariate stationary standard (with mean zero and unit variance) Gaussian process with autocorrelation $\widetilde{\rho}_\tau = \text{Corr}\left(X_t, X_{t+\tau}\right)$, where τ denotes the time lag, and X_t be a process obtained by the mapping operator $X_t = F^{-1}\left(\Phi\left(Z_t\right)\right)$, where $F^{-1}()$ denotes the ICDF of the target distribution (with finite variance) and $\Phi()$ denotes the Gaussian cumulative distribution function. It can be shown that the autocorrelation $\rho_\tau = \text{Corr}\left(X_t, X_{t+\tau}\right)$ of the final process is related to the Gaussian one by (see the abovementioned papers, and references therein),

$$\rho_\tau = \frac{\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} F_X^{-1}\left(\Phi\left(z_t\right)\right) F_X^{-1}\left(\Phi\left(z_{t+\tau}\right)\right)\ \varphi_2\left(z_t, z_{t+\tau},\ \widetilde{\rho}_\tau\right) \mathrm{d}z_t \mathrm{d}z_{t+\tau} - \left(\mathrm{E}\left[X\right]\right)^2}{\text{Var}\left[X\right]} 1 \quad (1)$$

where $\varphi_2\left(z_t, z_{t+\tau},\ \widetilde{\rho}_\tau\right)$ is the bivariate standard normal probability density function, which contains $\widetilde{\rho}_\tau$.

Furthermore let $\gamma_\tau \ = \text{Cov}\,(X_t, X_{t+\tau}) = \rho_t\gamma_0$ denote the autocovariance of the process, where γ_0 stands for its variance.

Let also recall that the power spectrum and the autocorrelation function of a process are related by (e.g., Koutsoyiannis, 2016, 2000; Papoulis, 1991 pp. 118),

$$S_{\gamma(\omega)} = 2\gamma_0 + 4\sum_{\tau=1}^{\infty}\gamma_\tau \cos\left(2\pi\tau\omega\right) \ = 2\sum_{\tau=-\infty}^{\infty}\gamma_\tau \cos\left(2\pi\tau\omega\right), \quad \omega \in [0,\ 1/2]\ \ 2 \qquad (2)$$

Therefore, by using $\gamma_\tau \ = \ \rho_t\gamma_0$ and substituting Eq. (1) into Eq. (2), it is shown that the power spectrum of the process X_t is related to the autocorrelation $\widetilde{\rho}_\tau$ of Z_t, and hence on its power spectrum $S_{\widetilde{\gamma}(\omega)} \ = \ 2\sum_{\tau=-\infty}^{\infty}\widetilde{\gamma}_t \cos\left(2\pi\tau\omega\right)$, where $\widetilde{\gamma}_t \ = \ \widetilde{\rho}_\tau$ (since the Gp has unit variance).

For completeness, it is mentioned that the autocovariance γ_τ can be obtained from a known power spectrum $S_{\gamma(\omega)}$ by,

$$\gamma_\tau = \int_0^{1/2} S_{\gamma(\omega)} \cos\left(2\pi\tau\omega\right) \mathrm{d}\omega\ , \quad j = 0, 1, 2, \ldots .3 \qquad (3)$$

An example of a spectrum-based method that uses this mapping procedure (i.e., through the ICDF) in combination with a suitably inflated spectrum for the auxiliary (or parent) Gaussian process is given by Deodatis and Micaletti (2001). This method avoids the underestimation of correlation coefficients and manages to preserve the distribution function of the process.

Just a few quick comments:

1. Section 3.1 (step 1): Since the authors employ the Kappa distribution (a general-

ization of the GEV distribution; which is bounded from below or above, depending on the parameters) to model the historical data, it could be insightful to mention that under certain parameter combinations, this distribution may lead to infinite moments. This can be a delicate issue, since if the fitted distribution exhibits infinite variance then the Pearson's correlation cannot be defined (the denominator contains the variance), and thus the proposed model (as well as many other models) cannot be used. This situation is discussed in section 3.4 of Tsoukalas et al. (2018b; and references therein), where it is advocated (based on empirical, as well as theoretical reasoning) that physical processes are characterized by finite variance (Koutsoyiannis, 2016).

Particularly, if X is a Kappa-distributed random variable, and $\mu_r = E\left[X^r\right]$ denotes the r^{th} raw moment, as discussed in Hosking (1994), and elsewhere, the existence of the r^{th} depends on the values of h and k. Specifically, the moments exist:

$$\text{for all } r \quad \text{if} \quad h \geq 0 \text{ and } k \geq 0$$

$$\text{for } r < -1/hk \;\; \text{if} \;\; h < 0 \text{ and } k \geq 0, \text{ and}$$

$$\text{for } r < -1/k \;\; \text{if } k < 0$$

It is also interesting to mention that Hosking (1994) notes that the first four moments cannot uniquely determine the parameters of the distribution, since some combinations of moments (expressed by skewness and kurtosis coefficients) correspond to different pairs of h and k.

1. Section 3.1 (step 1): Can you please provide more details on the employed fitting method. The fitting was performed using classical product-moments, L-moments, maximum likelihood, or another method?

2. Section 3.1 (Introduction and step 2): I believe that it would be useful to mention that the recent literature contains alternative models, such as the Stochastic Periodic Autoregressive To Anything (SPARTA) model of Tsoukalas et al. (2018a, 2017), that do not employ de-seasonalization techniques and are able to simulate cyclostationary processes (univariate and multivariate), accounting for many of its facets such as, seasonally varying marginal distributions and correlations.

3. Section 3.1 (step 6): The Authors mention that: "Negative simulated values are replaced by 0, which corresponds to the lower boundary of the Kappa distribution". As far as I am aware the left support of Kappa distribution is not necessarily zero (e.g., when $k = 0$ and $h \leq 0$, then the supports of the distribution are, $-\infty < x < \infty$; see Hosking (1994)). In any case, the generation of negative values can be eliminated by using a distribution function defined in the positive real line. Particularly, I would suggest the investigation/use of the Generalized Gamma and Burr type-XII distributions, which are more parsimonious (they entail three parameters) and were found adequate for modelling of hydrometeorological variables; particularly rainfall (e.g., Papalexiou and Koutsoyiannis, 2016). Examples of their use within the context of stochastic modelling can be found the work Papalexiou (2018), as well as in Tsoukalas et al. (2019, 2018b) and Tsoukalas (2018).

Regards,

Ioannis Tsoukalas

References

Biller, B., Nelson, B.L., 2003. Modeling and generating multivariate time-series input processes using a vector autoregressive technique. ACM Trans. Model. Comput. Simul. 13, 211–237. https://doi.org/10.1145/937332.937333

Cario, M.C., Nelson, B.L., 1996. Autoregressive to anything: Time-series input processes for simulation. Oper. Res. Lett. 19, 51–58. https://doi.org/10.1016/0167-6377(96)00017-X

Deodatis, G., Micaletti, R.C., 2001. Simulation of highly skewed non-Gaussian stochastic processes. J. Eng. Mech. 127, 1284–1295.

Embrechts, P., McNeil, A.J., Straumann, D., 1999. Correlation and Dependence in Risk Management: Properties and Pitfalls, in: Dempster, M.A.H. (Ed.), Risk Management. Cambridge University Press, Cambridge, pp. 176–223. https://doi.org/10.1017/CBO9780511615337.008

Grigoriu, M., 1984. Crossings of Non-Gaussian Translation Processes. J. Eng. Mech. 110, 610–620. https://doi.org/10.1061/(ASCE)0733-9399(1984)110:4(610)

Hosking, J.R.M., 1994. The four-parameter kappa distribution. IBM J. Res. Dev. 38, 251–258. https://doi.org/10.1147/rd.383.0251

Koutsoyiannis, D., 2016. Generic and parsimonious stochastic modelling for hydrology and beyond. Hydrol. Sci. J. 61, 225–244. https://doi.org/10.1080/02626667.2015.1016950

Koutsoyiannis, D., 2000. A generalized mathematical framework for stochastic simulation and forecast of hydrologic time series. Water Resour. Res. 36, 1519–1533. https://doi.org/10.1029/2000WR900044

Lebrun, R., Dutfoy, A., 2009. An innovating analysis of the Nataf transformation from the copula viewpoint. Probabilistic Eng. Mech. 24, 312–320. https://doi.org/10.1016/j.probengmech.2008.08.001

Liu, P.L., Der Kiureghian, A., 1986. Multivariate distribution models with prescribed marginals and covariances. Probabilistic Eng. Mech. 1, 105–112. https://doi.org/10.1016/0266-8920(86)90033-0

Mardia, K. V, 1970. A Translation Family of Bivariate Distributions and Frechet's Bounds. Sankhya Indian J. Stat. Ser. A 32, 119–122.

Nataf, A., 1962. Statistique mathematique-determination des distributions de probabilites dont les marges sont donnees. C. R. Acad. Sci. Paris 255, 42–43.

Papalexiou, S.M., 2018. Unified theory for stochastic modelling of hydroclimatic processes: Preserving marginal distributions, correlation structures, and intermittency. Adv. Water Resour. https://doi.org/10.1016/j.advwatres.2018.02.013

Papalexiou, S.M., Koutsoyiannis, D., 2016. A global survey on the seasonal variation of the marginal distribution of daily precipitation. Adv. Water Resour. 94, 131–145. https://doi.org/10.1016/j.advwatres.2016.05.005

Papoulis, A., 1991. Probability, Random Variables, and Stochastic Processes, Third edit. ed. McGraw-Hill Series in Electrical Engineering. New York City, New York, USA: McGraw-Hill.

Serinaldi, F., Lombardo, F., 2017. BetaBit: A fast generator of autocorrelated binary processes for geophysical research. EPL (Europhysics Lett. 118, 30007. https://doi.org/10.1209/0295-5075/118/30007

Tsoukalas, I., 2018. Modelling and simulation of non-Gaussian stochastic processes for optimization of water-systems under uncertainty. PhD Thesis, Department of Water Resources and Environmental Engineering, National Technical University of Athens (Defence date: 20 December 2018).

Tsoukalas, I., Efstratiadis, A., Makropoulos, C., 2019. Building a puzzle to solve a riddle: A multi-scale disaggregation approach for multivariate stochastic processes with any marginal distribution and correlation structure. J. Hydrol (in review).

Tsoukalas, I., Efstratiadis, A., Makropoulos, C., 2018a. Stochastic Periodic Autoregressive to Anything (SPARTA): Modeling and simulation of cyclostationary processes with arbitrary marginal distributions. Water Resour. Res. 54, 161–185. https://doi.org/10.1002/2017WR021394

Tsoukalas, I., Efstratiadis, A., Makropoulos, C., 2017. Stochastic simulation of periodic processes with arbitrary marginal distributions, in: 15th International Conference on Environmental Science and Technology. CEST 2017. Rhodes, Greece.

Tsoukalas, I., Makropoulos, C., Koutsoyiannis, D., 2018b. Simulation of stochastic processes exhibiting any-range dependence and arbitrary marginal distributions. Water Resour. Res. https://doi.org/10.1029/2017WR022462

Tsoukalas, I., Papalexiou, S., Efstratiadis, A., Makropoulos, C., 2018c. A Cautionary Note on the Reproduction of Dependencies through Linear Stochastic Models with Non-Gaussian White Noise. Water 10, 771. https://doi.org/10.3390/w10060771