Review report on “Stochastic simulation of streamflow time series using phase randomization”

by Demetris Koutsoyiannis

Journal:	Hydrol. Earth Syst. Sci. Discuss.
Journal’s Ref.:	HESS-2019-142
Title:	Technical note: Stochastic simulation of streamflow time series using phase randomization
Authors:	Manuela I. Brunner, András Bárdossy, and Reinhard Furrer (Switzerland, Germany)
Reviewer’s Ref.:	DK-JR-345
Date:	2019-05-12

Recommendation: Major revision

Reviewer’s assertion: It is my opinion that a shift from anonymous to eponymous (signed) reviewing would help the scientific community to be more cooperative, democratic, equitable, ethical, productive and responsible. Therefore, it is my choice (and aesthetic attitude) to sign my reviews.

1. The Technical Note by Brunner et al. (2019) implements a useful idea for easy stochastic simulation of daily streamflow, based on spectral representation and phase randomization. The method has several limitations (see below) but it is practical and useful, and it certainly deserves publication. I believe several issues can be improved before final publication and therefore I am providing some suggestions. I also appreciate the commentaries by Francesco Serinaldi, Ioannis Tsoukalas and Panayiotis Dimitriadis, who provided a lot of information to the authors. I think this information is useful to optimize their Note and also to put it in the context of modern and older literature, some of which is missing in the literature review. I believe that not everything suggested in the commentaries needs to be addressed, as this would change the orientation of the Note. However, with several changes in the formulations and a few expansions, rather than additional analyses, the Note could be improved. My own suggestions, which I list in the following points, fall in two categories: (a) recognition of the limitations of the method and (b) improvements in formulations, phraseology and terminology.

2. A first limitation, which in the current version is not stated clearly, is the severe dependence of the method on the sample size of observations. The synthetic series has the same length as the observed series. The authors properly recognize the importance of respecting long-range dependence (LRD) in simulation. However, to study its effect in hydrosystems we need synthetic series much longer than the observed. The use of ensembles of small-length time series may not be equivalent

with using a long time series as each member of the ensemble is independent from the others.

3. A second limitation is the absence of a model for time dependence. While the authors correctly adopt a model for the marginal distribution (e.g. they state "*Using the empirical distribution instead of the Kappa distribution would prevent us from obtaining values that go beyond the range of observed data…*") their method misses to do so for the dependence structure. The empirical autocorrelogram and periodogram are affected by significant bias and huge noise (see references provided by Panayiotis Dimitriadis) and if we do not use a model, then we reproduce a particular random realization, in terms of autocorrelogram and periodogram, in all our simulated series. I believe authors' statement "*The periodogram, the empirical counterpart of the power spectrum, shows high values at those frequencies which correspond to strong periodic components*" is only partly true and perhaps misleading. The periodogram could be regarded a realization of a stochastic process per se (on the frequency domain) and its peaks do not necessarily reflect a real peak in the "true" power spectrum. The same thing happens with the autocorrelogram. For example, the ups and downs in the empirical autocorrelograms in Fig. 5 may well be sampling artefacts, which we do not need to reproduce—but the method does reproduce them.[1]

4. A third limitation is the lack of parsimony of the entire methodology. From the statement "*We fit a separate distribution for each day to take into account seasonal differences in the distribution of daily streamflow values*" one can imagine that the overall method encompasses lots of parameters. Apparently, it is nowadays easy to do calculations with lots of parameters but, in my view, stochastics goes beyond calculations and algorithms. Parsimony in stochastic modelling is always important (see Koutsoyiannis 2016).

5. A final limitation for the particular time scale of modelling, i.e. daily, is the lack of explicit modelling of time irreversibility (an issue also mentioned in the comment by Francesco Serinaldi). This would not be an issue if the time scale was monthly or longer, but I suspect that it is for the daily scale (see Koutsoyiannis 2019 and also Müller et al. 2017). I clarify here that I do not suggest changing the method to overcome the limitations (e.g. to become more parsimonious or to take irreversibility into account). Rather, I just recommend stating them in a clear and explicit manner.

6. Now coming to the second category of my suggestions, I would recommend avoiding the name *kappa distribution* for the chosen distribution. It is true that in hydrological literature this name is in common use, but if we wish to facilitate communication with other disciplines, we should be aware that the name *kappa*

[1] If the authors have difficulty to accept my comment, I would suggest doing an experiment with a particular (smooth) autocorrelation function and see the ups and downs in the produced autocorrelogram and periodogram of a single realization.

distribution has another meaning in statistical thermodynamics—namely it is used to describe Cauchy-type (or Student-like) distributions in motion of particles (e.g. Olbert, 1968; Livadiotis and McComas 2013). The specific distribution used in the Note (which I do not think is a generalization of GEV as suggested by Ioannis Tsoukalas), is commonly (in most disciplines) referred to as the Dagum distribution—see https://en.wikipedia.org/wiki/Dagum_distribution. In addition, in terms of sign conventions in eqn. (4), I would suggest changing the signs of *k* and *h* and replacing the two minus signs in front of them with plus signs. This will make the expression more convenient and intuitive, and also complying to the standard notation used in other disciplines (e.g. as seen in the above web site).

7. The phrase "*Stochastically generated time series mimic the characteristics of observed data and represent sets of plausible but* ***as yet unobserved*** *streamflow sequences*" (my emphasis) may distort the meaning of what stochastic simulation is. It is not a matter of something that is "yet unobserved" but expected to be observed in the future. It is a matter of producing artificial "realizations" from the stochastic model. A model, stochastic or otherwise, is not identical to the real world.

8. The term "deseasonalization" needs to be used with care and clarification; otherwise it may mislead people to think that, by techniques like that used in the Note, we can get rid of seasonality. This, however, is quite difficult—if ever possible. With transformations of the time series, either linear (as in standardizing by mean and variance of each period) or nonlinear (as in fitting a separate distribution for each period, like what is done in this Note), we can only remove the seasonal effect on the marginal distribution, not that of the joint distribution of a cyclostationary stochastic process. (For example, differences in autocorrelation coefficients in different seasons are not removed by techniques such as the above mentioned). Therefore I suggest replacing "deseasonalization" with "deseasonalization of the marginal distribution."

9. The notion of "nonparametric" techniques referred to in the literature review is, in my opinion, problematic when we deal with stochastic processes with time dependence. As opposite to iid statistics, in which the first "i" (independent) is taken for granted, in stochastics there cannot be "nonparametric" methods; something of parametric type is always present, albeit sometimes hidden. Furthermore, the "bootstrap approaches" also mentioned in the Note are unsuitable for stochastic processes as they distort the stochastic structure—particularly in the presence of LRD. Therefore, I suggest making these clarifications and limiting the references to such types of models (as well as to ARMA-type models whose value is only historical, I believe). Instead, I suggest extending the review to other models, more appropriate for hydrological applications, such as those suggested by other commenters.

10. Could the authors double check their equations? Is an imaginary unit missing somewhere in equation (2)? Could they correct the notation in eqn (3)? (Is 'rand' meant to be a subscript?).

11. Finally, I uphold the other commenters in congratulating the authors and I particularly second Panayiotis Dimitriadis in congratulating them for using modest phraseology. I would add in the reasons for congratulation the fact that they do not follow the clichés and fashionable paths: for example they limit their mentions to climate impacts and nonstationarity, a notion that has become a must in hydrological papers—often by authors who do not know what it actually is (see Koutsoyiannis and Montanari 2015; Serinaldi and Kilsby, 2015).

References

Brunner, M.I., Bárdossy, A., and Furrer, R. (2019). Technical note: Stochastic simulation of streamflow time series using phase randomization, Hydrol. Earth Syst. Sci. Discuss., doi: 10.5194/hess-2019-142.

Dimitriadis, P., and Koutsoyiannis, D. (2015). Climacogram versus autocovariance and power spectrum in stochastic modelling for Markovian and Hurst–Kolmogorov processes, *Stochastic Environmental Research & Risk Assessment*, 29 (6), 1649–1669, doi: 10.1007/s00477-015-1023-7.

Koutsoyiannis, D. (2016). Generic and parsimonious stochastic modelling for hydrology and beyond, *Hydrological Sciences Journal*, 61 (2), 225–244, doi: 10.1080/02626667.2015.1016950.

Koutsoyiannis, D. (2019). Time's arrow in stochastic characterization and simulation of atmospheric and hydrological processes, *Hydrological Sciences Journal*, doi: 10.1080/02626667.2019.1600700,.

Koutsoyiannis, D., and Montanari, A. (2015). Negligent killing of scientific concepts: the stationarity case, *Hydrological Sciences Journal*, 60 (7-8), 1174–1183, doi: 10.1080/02626667.2014.959959,.

Livadiotis, G., and McComas, D.J. (2013). Understanding kappa distributions: A toolbox for space science and astrophysics. *Space Science Reviews*, 175(1-4), 183-214.

Müller, T., Schütze, M. and Bárdossy, A. (2017). Temporal asymmetry in precipitation time series and its influence on flow simulations in combined sewer systems. *Advances in Water Resources*, 107, 56-64.

Olbert S. (1968) Summary of Experimental Results from M.I.T. Detector on IMP-1. In: Carovillano R.L., McClay J.F., Radoski H.R. (eds) Physics of the Magnetosphere. Astrophysics and Space Science Library (A Series of Books on the Recent Developments of Space Science and of General Geophysics and Astrophysics Published in Connection with the Journal Space Science Reviews), vol 10. Springer, Dordrecht.

Serinaldi, F., and Kilsby, C.G. (2015). Stationarity is undead: Uncertainty dominates the distribution of extremes, *Advances in Water Resources*, 77, 17-36.