Reviewer 1

GENERAL COMMENTS

The manuscript proposes a wavelet-based phase randomization approach to generate multisite hydrological extremes, which is argued to be able to capture both spatial dependencies and non-stationarities. The stochastic simulated data is reconstructed with the same values as original data in another temporal order, and this assures the reproduction of temporal dependence, extremes and nonstationarities.

In addition, the multiplication of the same set of random phases to all the investigated sites ensures spatial dependences. The paper is clearly written, and the idea is well explained. However, I have listed some comments below which needs to be addressed before the manuscript is considered further for publications.

Reply: *Thank you very much for your comments, which we address point-by-point below.*

Major comments:

COMMENT1: The spatial dependence can be simulated by using the same randomized phases for multiple time series among sites. You claimed this in your introduction and Section 3.2 step 1 without detailed explanation and references. From my understanding, the phases represent the time of the timing of changes or variations (in your case, it is extreme events), so if you use the same randomized phases, it is expected that there will be spatial dependence of extremes among the sites you investigated. Thus, could you give explanations or illustrations on this statement.

Reply: *Prichard and Theiler (1994) and Schreiber and Schmitz (2000) describe that the phases of all sites have to be randomized in the same way to preserve cross-correlations for multivariate time series. We added the two references to the description of Step 1. Chavez and Cazelles* (2019) *later proposed a surrogate approach for multivariate time series for a dataset of weekly measles notifications and an electroencephalographic recording where they apply a phase randomization procedure using random phases extracted from a random, normally distributed time series. We also added this reference to the corresponding section.*

Figure 1 in this response to the reviewer shows on the example of the four stations in the Pacific Northwest (see Figure 1 in paper) that the observed cross-correlation would not be captured by the simulations if the phases were randomized for each catchment individually instead of using the same set of randomized phases across all catchments.

Modification: p.4, l.97; p.6, l.151

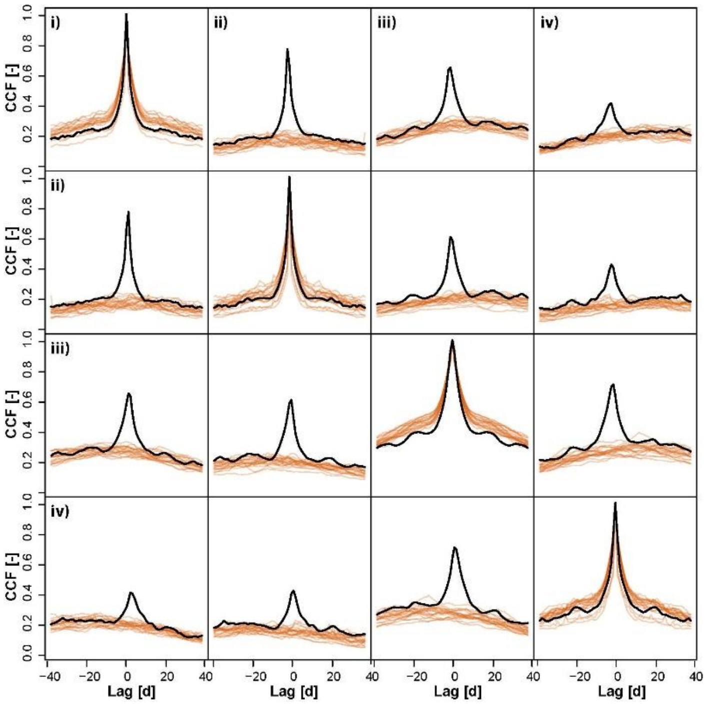

Figure 1: Comparison of observed (black) and simulated (orange) cross-correlation functions (ccfs) for the daily discharge values of the four catchments (i--iv) in the Pacific Northwest. 20 simulations were generated for each site individually, neglecting spatial dependence.

COMMENT2: Section 3.2, step 1 shows how the random phases are computed from a random discharge time series from a normal distribution with mean 0 and standard deviation 1. Could you explain the specific reason to choose normal distribution or have you tested with other distribution for example, gamma and kappa distribution?

Reply: *It is correct that we compute the random phases from a normal distribution. We have also tested the kappa distribution, which did not change the results. We decided to go with the normal distribution as such a distribution was also used by Chavez and Cazelles* (2019) *who proposed a surrogate approach for multivariate time series for a dataset of weekly measles notifications and an electroencephalographic recording. This reference was added to the description of Step 1.*

Modification: p.6, l.151

Another similar question is about the selection of wavelet family and scale. In this study, you use Morlet wavelet, what is the specific reason to use this wavelet family, and how about other wavelet families, e.g., Paul, DOG and Marr (Torrence and Compo, 1998)? The sensitivity of your approach to the selection of wavelet family and scale? For the application of wavelet method in real world, the selection of wavelet family and scales is of great importance.

Reply: *Because of the phase randomization step, our application requires a complex wavelet with an imaginary part in addition to a real part. The DOG wavelet family is real valued while the Morlet and Paul wavelets are complex valued [Torrence and Compo, 1998]. Among the complex valued wavelets, we use the Morlet wavelet because it has been found suitable for hydrological applications in previous studies [Lafrenière and Sharp, 2003; Labat et al., 2005; Schaefli et al., 2007]. It provides a better frequency localization than other complex wavelets such as the Paul wavelet [Torrence and Compo, 1998]. We added the additional references (Lafreniere et al. 2003 and Labat et al. 2005) to the text. It is important that the number of scales is chosen high enough to allow for a fine resolution* [*Torrence and Compo*, 1998]. *Using e.g. only 20 scale results in reconstructed time series that do not reflect all the necessary detail. We chose 100 scales because further increasing the number of scales no longer improves reconstruction performance.*

Modification: p.4, l.120-121

Additionally, I am unable to visualise where exactly the random phases go back into equation 3?
Reply: *The complex part of the wavelet transform Wn(h) comes in via the complex conjugate * in Equation 2. During back-transformation in Equation 3, it comes back in when deriving the real part at each scale h through R(Wn(hj)).*

What should have been the values of these phases with the observed data in the first place?
Reply: *The phases are uniformly distributed over the range of -π to π, which was specified in the text.*
Modification: p.7, caption of Figure 2

Why should these phases be the same? I suspect the only impact these phases have is on spatial dependence. In that case why are the majority of results that are presented focusing on temporal dependence attributes? Figure 6 gives some flavour of the multi-suite stochastic generation. It looks troublingly similar to the observations. What would happen if non-identical phases used across all the sites? These questions are important to address to establish the contribution here is an improved representation of spatial dependence compared to other alternatives.
Reply: *The random phases have to be the same across all stations to retain spatial dependence, which is mentioned in the text and is why the spatial aspect receives a lot of attention in the validation part of the manuscript (Figures 6-11, i.e. half of the figures). Figure 6 gives an impression of what the simulated time series look like for three regions with four stations each. While the general distributional and temporal characteristics of the series are retained by the simulations, the simulations do not reproduce observed events but instead generate new series of potential successions of events. Figure 7 shows the cross-correlation functions among pairs of stations for the four catchments in the Pacific Northwest. If non-identical phases were used, these spatial correlations would not be reproduced as demonstrated in Figure 1 shown above. Figure 8 summarizes spatial dependencies over the whole distribution for all pairs of stations (671 times 671) in a variogram. Figure 9 gives an overview of spatio-temporal characteristics of simulated flood events and Figure 10 summarizes spatial dependencies for all pairs of stations (671 times 671) via the F-madogram, a measure of extremal dependence. Also the last Figure (11) focuses on the spatial dependence aspect by comparing observed with simulated tail dependence coefficients, which are again computed for all pairs of stations. We clarify the description of the evaluation procedure by specifying that the general spatial dependence structure and the spatial dependence in high extremes is evaluated. Figure 1 in this response to the reviewer shows that spatial correlations are not retained if non-identical phases are used across all sites.*
Modification: p.9, l.199-200

Regarding other alternatives, is there a possibility of comparing this method with the commonly used spatial stochastic generators such as SPIGOT or other equally worthy alternatives? I would like to see the variogram figure possibly expanded to also show alternate outcomes using other formulations of spatial dependence (i.e. different ways of specifying phases).
Reply: *We rerun the simulations for the whole set of stations without using the same phases across all stations but generating a set of random phases for each station individually. Figure 2 in this response to the reviewer shows that spatial dependence is modeled much worse than when randomizing the phases for all catchments in the same way (Figure 8 in manuscript).*
We agree that a comparison of different spatial stochastic generators with respect to how they represent spatial extremes would be interesting and valuable. In order to make such a model comparison study beneficial for the community, ideally a broad range of models ranging from continuous to event-based models should be compared as there exists no commonly accepted benchmark/reference model. Such a comparison goes beyond the scope of this manuscript and should be addressed in a separate study.

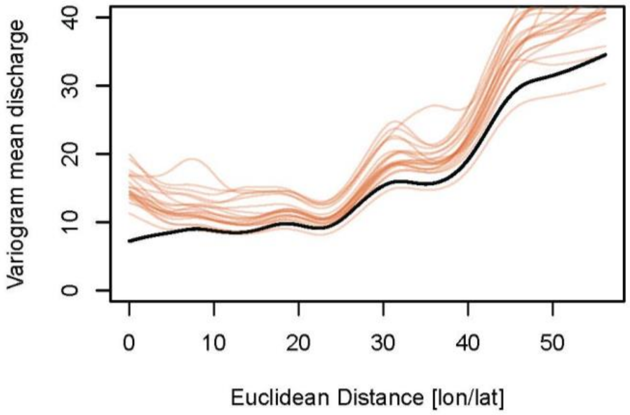

Figure 2: Comparison of observed (black) and simulated (orange) variograms for 20 simulation runs where the phases were randomized for each station individually.

COMMENT3: Section 3.2, step 4 mentions how the reconstruction is done by using the inverse wavelet transform (Eq. (3)) combining the derived random phases and the amplitudes in previous steps. Could you write the combination explicitly in the from of the equation (i.e., include the random phases in the Eq. (3))?

Reply: *The complex part of the wavelet transform Wn(h) comes in via the complex conjugate * in Equation 2. The complex conjugate is derived using the modulus and argument, i.e. the phases, of the complex numbers contained in the Morlet wavelet (Equation 1). The wavelet coefficients Wn(h) resulting from the transform are also complex numbers where the argument of the complex wavelet value Wn(h) corresponds to the wavelet phase. During back-transformation in Equation 3, the phases come back in when deriving the real part at each scale h through R(Wn(hj)).*

Minor comments: Line 147: h=100 wavelet scales, I think you mean number of wavelet scales is 100, not the scale itself equals 100. Wavelet scales should be hj = h0*2ˆ(j*dj), j=0,1,: : :J. Or is this some parameter that is specific to a continuous wavelet transform? Why is it of relevance, what impact does it make, and why use the same value for all locations is something that should be discussed.

TORRENCE, C. & COMPO, G. P. 1998. A Practical Guide to Wavelet Analysis. Bulletin of the American Meteorological Society, 79, 61-78.

Reply: *Thank you for pointing this out, yes, we intended to refer to the number of wavelet scales. This was adjusted in the text. The number of scales can be freely chosen in the case of the continuous wavelet transform. A high number of scales compared to a low number of scales allows for a finer resolution* [*Torrence and Compo*, 1998] *and for a detailed reconstruction of the original time series.*

Modification: p.8, l.173

References used in this response to the reviewer

Chavez, M., and B. Cazelles (2019), Detecting dynamic spatial correlation patterns with generalized wavelet coherence and non-stationary surrogate data, *Sci. Rep.*, *9*(1), 1–9, doi:10.1038/s41598-019-43571-2.

Labat, D., J. Ronchail, and J. L. Guyot (2005), Recent advances in wavelet analyses: Part 2 - Amazon, Parana, Orinoco and Congo discharges time scale variability, *J. Hydrol.*, *314*(1–4), 289–311, doi:10.1016/j.jhydrol.2005.04.004.

Lafrenière, M., and M. Sharp (2003), Wavelet analysis of inter-annual variability in the runoff regimes of glacial and nival stream catchments, Bow Lake, Alberta, *Hydrol. Process.*, *17*(6), 1093–1118, doi:10.1002/hyp.1187.

Prichard, D., and J. Theiler (1994), Generating surrogate data for time series with several simultaneously measured variables, *Phys. Rev. Lett.*, *73*(7), 951–954.

Schaefli, B., D. Maraun, and M. Holschneider (2007), What drives high flow events in the Swiss Alps? Recent developments in wavelet spectral analysis and their application to hydrology, *Adv. Water Resour.*, *30*(12), 2511–2525, doi:10.1016/j.advwatres.2007.06.004.

Schreiber, T., and A. Schmitz (2000), Surrogate time series, *Phys. D Nonlinear Phenom.*, *142*(3–4), 346–382, doi:10.1016/S0167-2789(00)00043-9.

Torrence, C., and G. P. Compo (1998), A practical guide to wavelet analysis, *Bull. Am. Meteorol. Soc.*, *79*(1), 61–78.