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. We added Figure one to the appendix of the manuscript.
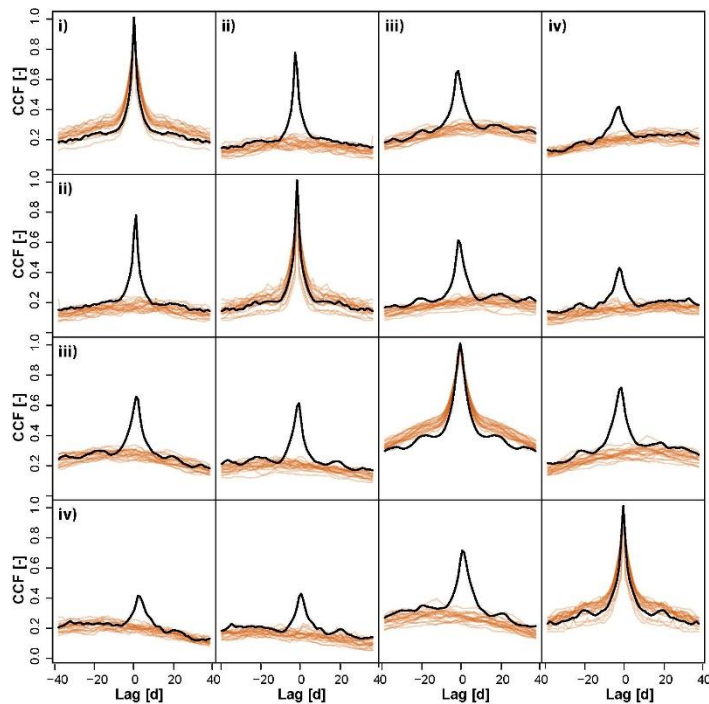**Modification: p.4, l.97; p.6, l.151, Figure A1**

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 and a note that we also tested the kappa distribution was added to the description of Step 1.*

Modification: p.6, l.154-155

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. This specification was added to the text.
Modification: p.4, l.120-121 and p.6 l. 156-158

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)). We added this information to the description of Steps 3 and 4 of the simulation procedure.*
Modification: p.8, l.179-186

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, which was added as an appendix to the manuscript, shows that spatial correlations are not retained if non-identical phases are used across all sites.*
Modification: p.9, l.210-211, Figure A1

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, which was added to the appendix of the manuscript, 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.
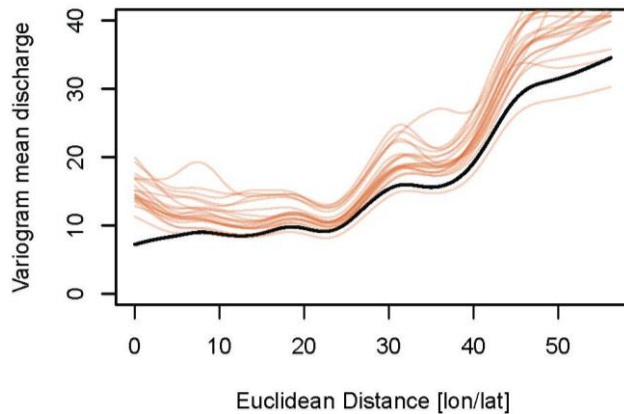**Modification: Figure A2**

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 $W_n(h)$ 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 $W_n(h)$ resulting from the transform are also complex numbers where the argument of the complex wavelet value $W_n(h)$ 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 $R(W_n(h_j))$. These explanations were added to the text.*
Modification: p.8, l.179-186

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 $h_j = h_0*2^{(j*dj)}$, 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.179

Reviewer 2: Dr. Sandhya Patidar

The paper presents a continuous wavelet-based phase randomisation approach for the stochastic generation of streamflow time series. This is an interesting paper that

substantially extends the original techniques developed in [1] for stochastic simulation of streamflow using Fourier transformation based phase randomisation. The original method, presented in [1] is associated with certain limitations such as application of Fourier transformation does not account for the non-stationarities in time series and resulted in an underestimation of spatial dependencies (e.g. cross-correlation) in both daily discharge and extreme events across the multiple sites. The present paper address some of the limitations observed in [1] by replacing Fourier transformation with a complex wavelet-based approach. The efficiency of original method has been evaluated to generate realistic series for distributional and temporal correlation characteristics and is validate through the application across four catchments in Switzerland. The proposed model is applied to a large dataset of 671 catchments in the contiguous United States and the efficiency of the model has been evaluated by assessing its ability to capture distributional and temporal characteristics at individual sites along with the spatial dependencies across the multiple sites, and extreme events (floods) including duration and volume.

Some general comments

Section 1 - Considering the theme of paper that signifies the application of wavelet approach, the Introduction section presents an interesting critical review on recently developed/applied modelling schematics that involves wavelet-based phase randomisation as a key approach.

Line 70-73: To add clarity it would be helpful if the authors' team add some brief explanation on how data normalisation procedure and back transformation impacts the spatial dependencies.

Reply: *Thank you for pointing out the need for clarification. We specify that 'This weakening is because phase randomization preserves the cross-correlation in the normal domain but not necessarily in the domain of the original distribution as linear correlation is not invariant under non-linear strictly increasing transformations.'*

Modification: p.3, l.71-73

Line 87-88: Please add some clarity on how continuous wavelet transform is more effective than a discrete wavelet transforms in minimising/overcoming issues around the long-term periodicities and/or non-stationarities. Is there any specific studies carried out to investigate such issues.

Reply: *The discrete wavelet transform only allows for real wavelet functions while the continuous transform allows for complex wavelet functions. Stochastic approaches randomizing only real-valued amplitudes have been shown to have problems with the reproduction of non-stationarities [Breakspear et al., 2003]. Chavez and Cazelles (2019) show that the randomization of phases (information stored in the complex-valued coefficients) allows for the generation of non-stationary time series. We rephrased the whole paragraph to clarify that the advantage of using a continuous instead of a discrete transform comes from the additional phase information gained when using complex wavelet functions which are only available for the continuous transform.*

Modification: p.3, l.83-83 and l.89-93

Section 2 - Theoretical background section provides suffiecient details on the wavelet decomposition approach.

Reply: *We are glad that you consider the background section to provide sufficient detail.*

Section 3 Data – For illustration and validation purposes, dataset are organised in three different region based on the general hydrological characteristics. It is not clear which specific properties has been used and how rigorously they have been appied. I think, this work could have benefitted if Authors' have considered using some form of clustering approaches (e.g. K-mean) based on key characterisetic for clustering the sites.

Reply: *Thank you very much for this suggestion. We indeed applied a clustering procedure to define these regions which are similar in terms of their flood characteristics. The clustering was applied on a distance matrix computed from the F-madogram, which is a measure of extremal dependence, between pairs of stations. The clustering was applied to the 671 catchments and resulted in 15 clusters among which we selected 3 for illustration purposes. We added a short description of the clustering procedure to the text. Our manuscript, where we describe this clustering procedure and the resulting clusters, is currently under review and we will include its reference here if it receives a DOI before this manuscript eventually goes into press.*
Modification: p.5, l. 142-154

Section 3 Method - It seems that the model interconnect different site only as part of step 1 (phase randomisation/perturbation applied throught the medium of white noise). All the remaing steps (1-4 steps) are applied independently across all the sites. I think the approach is appropriate. A separate Kappa distribution is fitted for each day for a 30-day window to factor in seasonal differences. I have a minor concern here, What is the motivation for the selection of a 30 day window, how does it effect data with different seasonal periods across different sites (e.g. may be one site having monthly seasonality but other having weekly seasonal characteristics or say over a three months period).
Reply: *Thank you for acknowledging the appropriateness of our approach. We chose a window of 30 days to temporally pool data prior to the estimation of the daily parameters of the kappa distribution. The idea of the smoothing is to reduce the effect of sampling uncertainty and to reduce day-to-day variability in parameters under the assumption that the distribution of flow on day x is unlikely to be substantially different from the one on day x+1. A value of 30 days was chosen for the moving window to enable sufficient smoothing in the parameter space and to ensure a sufficient sample size for the estimation of the four parameters. The moving-window nature of the approach allows for capturing seasonal and partly also weekly variations. We clarified in the method description that this approach corresponds to a moving window approach.*
Modification: p.8, l.171

Section 4
A robust evaluation has been conducted that includes careful seletion of sites (distinct and representative). Statistics used for comparision are appropriate and results are well explained.
Reply: *Thank you for appreciating the representativeness of our evaluation.*

Some minor comments
Figure 9 – Visually observed and simulated looks in good agreement for occurance of POT events but for a robust comparision some measurements should bave been used in parallel.
Reply: *We computed differences in the mean inter-event duration (i.e. time elapsing between two successive events) of observed and simulated series and the standard deviation of inter-event duration for events where 1, 2, and 3 stations were jointly affected, respectively. We find that over all three regions, relative differences in mean and standard deviation of inter-event duration lie mostly below 10%. However, we find that a visual comparison is most effective here to demonstrate the value of the simulation approach.*

Figure 10 gives a good illustration of how spatial dependencies could be effected among th sites with respect ot the Euclidean distance. However, for the readers benefit it would be appreciate if Authors' consider to provide few sentences to explain F-madograms plots, specifically, what should a relative difference of 0.05 in observed and simulated values should be interpreted.
Reply: *We specified in the figure caption that the F-madogram is a measure of extremal dependence. We also state that an overestimation of spatial dependence means that a pair of stations co-experiences more joint floods than in the observations. What a difference of 0.05 means in terms of differences in regional hazard estimates is hard to say and would need to be investigated in a proper*

study where the stochastic simulations are used to estimate regional flood hazard.
Modification: p.17, caption of Figure 10

Section 5 and 6 – Overall good and capture key aspects of the paper.
Reply: *Thank you.*

Reference
[1] Manuela I. Brunner, András Bárdossy, and Reinhard Furrer, Technical note: Stochastic simulation of streamflow time series using phase randomization, Hydrol. Earth Syst. Sci., 23, 3175–3187, 2019.

Commentator Tsoukalas

Introduction
This is an interesting work aiming to provide a method (as well as implement it in an R package; function called PRSim.wave within PRsim R package) for the simulation of multivariate hydrological processes (for now, focusing on streamflow) - which according to the results presented by the Authors has a good potential, requiring yet some improvements.
In general, I find the manuscript well-organized and straightforward to understand, yet in my view, there are several points that require the Authors attention.
All comments and suggestions are meant to be constructive and aim to improve the quality of the manuscript, as well as the findings obtained.
Reply: *Thank you for your detailed comments, which we address below.*

Comments
- **L6-7.** I suggest to write: "To do so, we propose the stochastic simulation approach called Phase Randomization Simulation using wavelets (here after called *PRSim.wave*) which combines… ".
  Reply: *We added brackets around PRSim.wave.*
  Modification: p.1, l.7

- **L11.** To avoid confusing the reader, and provided that a few lines above it is mentioned that "We apply and evaluate *PRSim.wave* on a large set of 671 catchments in the contiguous United States.", I suggest to write: "…at multiple sites (up to four)…"
  Reply: *We mean to say that we evaluate the set on 671 catchments. However, it is correct that we focus on three sets of four stations each in order to present a few more detailed results. Since we show that the approach works for very large datasets, we think that a generalization to 'multiple sites' is justified.*

- **L34.** I wonder what are the "potential non-stationarities" mentioned by the Authors? Are you refereeing to the typical cyclostationary behavior exhibited by hydrological processes? Given the opportunity, and as a side note, I would like highlight that stationarity is an essential tool for inferencing from data (e.g., model fitting). Stationarity should not be seen as a shortcoming, nor dead. Non-stationarity implies non-ergodicity, which in turn makes inference from observed data impossible, unless of course the deterministic dynamics of the process are known; which in my understanding, is never the case in hydrological sciences. On this topic, I recommend the recent work of Serinaldi et al. (2018), with emphasis on section 4.2, as well as the works of Koutsoyiannis and Montanari, (2007), (2015), Lins and Cohn (2011), Matalas (Matalas, 2012), and Montanari and Koutsoyiannis (2014), that argue in favor of stationarity. See also the very interesting, note of Harry F. Lins. As Harry F. Lins concludes his note:

Stationarity ≠ static

Non-stationarity ≠ change (or trend)

Reply: *What we state in this manuscript is that PRSim.wave is able to reproduce non-stationarities in the data. This is shown in Figures 4e. It is shown there that the wavelet power corresponding to different scales varies over time, i.e. the importance of different cyclical features varies over time. We are neither saying that stationarity is dead nor that it is a shortcoming. In contrast, we point out that developing stochastic models that can handle non-stationarities is important.*

L36. I am not sure what is the meaning of "continuous" here? Can you please elaborate/specify? Also, some references would be useful.

Reply: *We specify that continuous approaches correspond to discrete-time models in the stochastic literature.*

Modification: p.2, l.36-37

- **40.** Although I understand its rationale, I am not a big fan of the now-typical classification of stochastic models on "parametric and non-parametric", since in my view, there is no model without parameters. Typically, the literature uses the term "non-parametric" to refer to approaches that use some kind of resampling mechanism (e.g., k-nn algorithm) and/or "non-parametric" distribution functions (e.g., kernel-based approximation of the density function) to generate synthetic data. But, one should take a moment and think, are these really non-parametric? Isn't the *k* (i.e., the number of nearest neighbors) in k-nn algorithm a parameter? Isn't the choice of the kernel smoothing function (e.g., normal, epanechnikov, box, triangle) a parameter? Isn't the bandwidth of the kernel a parameter (also called the smoothing parameter)? Aren't the data *per se* used as parameters (e.g., when a non-parametric method relies on the sampling from the empirical CDF or kernel-based CDF. What if we have new data or alter a few? Does the model change?)? Having said these, I suggest to the Authors to reconsider using the employed classification, as well as review the recent literature (e.g., Serinaldi and Kilsby, 2014; Tsoukalas et al., 2019) for finding an alternative classification.

  Reply: *We here understand a parametric approach as an approach where a model is fitted to the data. We agree, however, that the choice of k in a nearest-neighbor algorithm might also be viewed as a parameter in a wider sense. We would, however, not go as far as to say that the data are parameters. However, they are subject to sampling uncertainty, which influences model fitting.*

- **L41-43.** The Authors write: "…and temporal disaggregation models such as fractional Gaussian noise models (Mandelbrot, 1965), fast fractional Gaussian noise models (Mandelbrot, 1971), broken line models (Mejia et al., 1972), and fractional autoregressive integrated moving average models (Hosking, 1984).".

  To clarify, these are not disaggregation models, but models able to simulate processes exhibiting long-range dependence (particularly, designed to simulate fractional Gaussian noise (fGn) processes or else, processes exhibiting Hurst behavior). See a similar discussion in the introduction section of Tsoukalas et al. (2018b).

  Reply: *We rephrased the sentence as follows: 'Parametric models include autoregressive moving average (ARMA) models and their modifications (Stedinger1982, Paplexiou2018), and fractional Gaussian noise models (Mandelbrot1965) comprising fast fractional Gaussian noise models (Mandelbrot1971), broken line models (Mejia1972), and fractional autoregressive integrated moving average models (Hosking1984).'*

  Modification: p.2, l.42

- **L44-45.** The Authors write: "Nonparametric models are based on disaggregation and resample from the data with perturbations and include…".
  I think that this statement can be confusing, and needs some refinement. "Non-parametric" models are not necessarily based on the notion of disaggregation. Of course, the literature offers "non-parametric" disaggregation methods (e.g., Lee et al., 2010; Tarboton et al., 1998), but this does not makes all "non-parametric" methods, methods that "based on disaggregation and resample". Further details on disaggregation methods can be found on the seminal works of Valencia and Schakke (1973), and Mejia and Rousselle (1976), as well in the work of Koutsoyiannis (2001) who provide a detailed overview on the subject. For a more recent overview and discussion on the topic of disaggregation and multi-temporal simulation see also work of Tsoukalas et al. (2019).
  Reply: *Thank you for pointing out that this statement is confusing. We rephrased the sentence to 'Nonparametric models include kernel density estimation (Lall1996, Sharma1997) and various bootstrap approaches…'*
  Modification: p.2, l.44-45

- **L48-49.** The Authors write: "…but none of these time domain methods can capture the spectral properties of the observed time series (Erkyihun et al., 2017).".
  In my view this statement is a bit confusing, requiring the Authors attention, for two reasons.
  1) A timeseries (i.e., a sequence of observations ordered in time) does not has spectral properties, it exhibits some form of dependence structure (which can be quantified using statistics/stochastics, e.g., through the empirical correlation coefficients and the empirical spectrum). What has spectral properties is the stochastic process that it is assumed that generated the observed timeseries.
  2) Having said the above, and since correlation and spectrum are interrelated quantities, if a model is capable of reproducing the process's correlation structure it also reproduces its spectrum (and vice versa). For further details and references, see my previous comment (Tsoukalas, 2019) on a recent work co-authored by the first Author of this work.
  Reply: *Thank you for pointing out that this statement was confusing. We removed the first part of this sentence and now state 'and the representation of spatial dependence in such time-domain models is challenging.'*
  Modification: p.2, l.48

- **L49-50.** In my view the sentence "Furthermore, these time-domain models struggle with the representation of spatial dependence" is a bit "strict", since as far as I see it, there is no struggle, but many research efforts (past, and new).
  The stochastic hydrology literature offers several "time-domain" models that can simulate parsimoniously multivariate processes, including both stationary and cyclostastionary processes (e.g., Efstratiadis et al., 2014; Koutsoyiannis, 2001, 2000), reproducing also the moments of the observed processes (typically up to third order). Further to these models/methods, more recent approaches allows the parsimonious simulation of multivariate stationary and cyclostationary processes with any marginal distribution and correlation structure (Kossieris et al., 2019; Tsoukalas, 2018; Tsoukalas et al., 2018a, 2018b), also in a multi-scale context (Tsoukalas et al., 2019). Apart from the last work, for another multi-scale and multivariate simulation study involving daily rainfall at 4 sites the Authors are referred to Appendix D, section D.2, of Tsoukalas (2018). Therefore, taking into consideration the above-mentioned works I would suggest the Authors to revise the sentence accordingly, as well as provide some references.
  Reply: *We rephrased the sentence to 'and the representation of spatial dependence in such time-domain models is challenging.'*
  Modification: p.2, l.48

- **L54-55.** The Authors write: "In contrast to time-domain models, frequency-domain models allow for the simulation of surrogate data with the same Fourier spectra as the raw data". This can be also true for time-domain methods (but not a good modelling practice in either of the two cases; see below). For instance, if one employs an AR or MA model of high order can simulate a realization of a process exhibiting exactly the empirical autocorrelation coefficients up to the order dictated by the model. However, this is not a good modeling practice since it is well-known that the empirical estimators of auto- (and cross-) correlation coefficients are (downward) biased (Beran, 1994; Koutsoyiannis, 2003, 2000), especially in the case of long-range dependence, short samples, and large lags. See also Matalas (1967 p. 945) who remark that:

  "*Parameters that are determined in terms of high order moments of large time lags are subject to large standard errors and consequently large operational biases. Operational biases can never be eliminated, but they can be minimized by the use of regionalization to account for the temporal and spatial variations inherent in the historic sequences…*".

  Of course, the same applies for the empirical estimators of spectrum (see the comparative work of (Dimitriadis and Koutsoyiannis (2015)). Note that this kind of approaches are not parsimonious (since all the empirical estimates used in model fitting are essentially model parameters). To cope with these, the recent literature (Kossieris et al., 2019; Tsoukalas et al., 2019, 2018b), as well some works already cited in the manuscript (i.e., Papalexiou (2018)), has leaned towards the use of parametric models (e.g., with two or three parameters) to parsimoniously describe the dependence structure of the processes. The Authors are referred to the work of Koutsoyiannis (2000) which in my view popularized that idea in hydrological domain, also introducing a parsimonious two-parameter auto-correlation structure. It is also interesting to note the work of Papalexiou (2018) (already cited in the manuscript), who employed the functional form provided by the survival function of a distribution to define several auto-correlation structures.

  Reply: *We weakened the statement and rephrased the sentence to: 'In contrast to most time-domain models,..'*
  Modification: p.2, l.53

- **L70-72.** The Authors write: "In addition, it may help to improve the representation of spatial dependencies because it does not require a transformation to the normal distribution and back to the original, skewed distribution, which usually weakens spatial correlations (Embrechts et al., 2010)."

  First, the comment on "weaken spatial correlations" applies for all "types" of correlations (that emerge from the mapping/transformation from the Gaussian to the actual domain) – not only spatial. Particularly, in the case of stochastic processes, it also applies for the auto-correlation structure of a stationary processes, as well as for the season-to-season correlations of a cyclostationary process (Tsoukalas et al., 2018a, 2017). It also holds for multivariate cases. However, I am afraid that I cannot see the improvement of the representation of spatial dependencies mentioned above by the Authors. The cross-correlations as well as the auto-correlation are still not accurately reproduced (see my comments below on the results/plots). It is my understanding that a previous comment of mine (Tsoukalas, 2019) on a recent work co-authored by the first Author of this work holds also for this method. This is due to the following:

  Reply: *We here specifically talk about the effect of back transformation on spatial correlation because this is the focus of this study. As mentioned in the manuscript the wavelet transform employed here does, compared to the Fourier transform, not necessitate a transformation to the normal domain, and a back transformation to the domain of the skewed distribution, which improves the representation of spatial correlations as compared to the Fourier transform based approach [Brunner et al., 2019]. Figure 1 shown in the response to reviewer 1 was added to the manuscript (A1) and illustrates that spatial correlations are not well*

represented if phases are randomized for each station individually. In contrast, the use of the
same random phases for all the stations as proposed in this manuscript leads to a nice
reproduction of spatial dependencies as illustrated e.g. by the cross-correlation functions
shown in Figure 7 in the manuscript.
Modification: Figure A1

- **L145.** The Authors write: "Derivation of random phases: A random discharge time series (white noise) of the same length as the input series is sampled from a normal distribution with mean 0 and standard deviation 1."
- **L170-173.** The Authors write: "Transformation to kappa distribution: The simulated values are transformed to the kappa domain using the fitted daily kappa distributions from Step 2. For each day, a random sample is generated from the fitted, daily kappa distribution. The simulated values are replaced by the values generated from the kappa distribution using rank-ordering. This procedure is repeated for each day in the year."
  Based on the above the method presented herein depends on an auxiliary Gaussian process and uses the target ICDFs, as well as the rank-correlations to establish the (auto- and cross-) dependence structure. It is reminded that such a procedure will preserve the ranks correlation coefficients (which do not depend on the marginals) but not the Pearson's, (which depends on the marginals; since it involves the cross-product moment of the among the variables). For further details the Authors are referred to the comment mentioned above, as well as in the references therein. It is my understanding that the mechanics of the method that dictate the preservation of ranks is the reason why the auto- (and cross-) correlations are not so well reproduced by the proposed method.
  Reply: *The results shown throughout Figures 7-11 demonstrate that the wavelet-based phase-randomization approach we present in this manuscript is well capable of reproducing observed spatial dependencies which is not the case when randomizing the phases of the Fourier transform as shown in Brunner et al.* (2019)*.*

- **L149 (and elsewhere)**: The use of Kappa distribution. As mentioned in a previous comment of mine in HESS related with a work co-authored by an Author of this manuscript there are few complications worth considering when using the Kappa distribution. The following comments are excerpted with minimum or no modifications at all from Tsoukalas (Tsoukalas, 2019).
  1. Since you are using the Kappa distribution 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 $\mu_r=[X_r]$ 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: for all r if $h\geq0$ and $k\geq0$ for $r<-1hk/$ if $h<0$ and $k\geq0$,and for $r<-1k/$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.
  2. How do you handle negative values? As far as I am aware the left (and right) support of Kappa distribution is not necessarily zero (e.g., when $k=0$ and $h\leq0$, 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; instead of four as in Kappa) 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).

Reply: *We tested the Burr type XII (validation results see Figure 1 in https://www.hydrol-earth-syst-sci-discuss.net/hess-2019-142/hess-2019-142-AC1-supplement.pdf) and generalized Gamma distributions, which were, however, not flexible enough to model the marginal distributions of the daily discharge values and led to unrealistically extreme high flows in the simulations and will therefore not be considered as alternatives to the kappa distribution. The R-package PRSim, however, allows for the implementation of any distribution chosen by the user as specified in the section code and data availability. We specify that negative simulated values are replaced by 0, which corresponds to the use of a bounded kappa distribution.*

Modification: p.8, l.189

- **L158-160.** The Authors write: "We fit a separate distribution for each day to take into account seasonal differences in the distribution of daily streamflow values. To do so, we use the daily values in a 30-day window around the day of interest."
  Can you please elaborate on this? Just to be sure, for each "site" and for each day of the year a Kappa distribution has been fitted with different parameters? If this is the case, just for the marginal behavior, and for each site you fitted Kappa 365 times, which implies that the model has 365×4 (the number of parameter of Kappa)=1460× the number of sites, parameters (not accounting those for the specification of the auto and cross-dependence structure of the process). If this is the case, I am afraid that this is not a parsimonious model, something that should be clearly stated in the manuscript (also mentioning the total number of its parameters).
  Also, I don't think that it is reasonable to assume that days belonging in the same month (e.g., the 19_{th} and the 20_{th} of August) have different marginal distribution (although, I have seen stochastic simulation related works following that approach, I am not aware of any paper supporting this assumption). The standard approach for daily (or finer time) scales is to consider stationarity within the monthly interval (i.e., in the case of daily data consider that all days belonging in the month have the same marginal distribution). An arguably more parsimonious approach, since in this case the total number of parameters for the marginal behavior would be 12×4 (the number of parameter of Kappa)=48× the number of sites. Note that the number of parameters could be further reduced by using, instead of Kappa, alternative 2- or 3-parameter distribution models.
  Reply: *Yes, the kappa distribution was fitted for each day of the year. A moving window approach was used to reduce the effect of sampling uncertainty and to reduce day-to-day variability in parameters under the assumption that the distribution of flow on day x is unlikely to be substantially different from the one on day x+1. We acknowledge in the discussion section that ' it [the model] requires the fitting of many parameters, which make the model non-parsimonious (Koutsoyiannis2016).*

- **L162-164.** The Authors write: "In a few regions with many zero discharge values (e.g. some catchments in the Great Plains) fitting the kappa distribution is not possible and we therefore use the empirical distribution instead."
  This is a work-around that could work, but I wonder, why not use an alternative distribution model (e.g., zero-inflated or mixed) that can model simultaneously both the discrete (i.e., probability of no discharge) and continuous part (i.e., distribution of non-zero discharge) of the process? Also, can you provide an estimate on the number of cases where the empirical distribution is employed instead of Kappa?

Reply: *This work around was necessary for those stations in the Great Plains showing an intermittent flow regime with many 0 values. If desired, the R-package PRSim allows for the implementation of an alternative distribution as specified in the data and code availability section.*

- **L218-219.** The Authors write: "the seasonal (3d) and monthly distributions (3e–g) are well captured by the simulations."
  This is a confusing description of the plots. The (3e–g) plots show that some seasonal summary statistics are reproduced (i.e., monthly mean, monthly maxima, monthly minima), not the seasonal distributions (to do so you need alternative plots, comparing the empirical distribution of each month with the corresponding theoretical one). Also, I don't understand what plot 3d shows? What does "seasonal statistics" means in the title of this plot? What does the phrase "the seasonal discharge distribution" means at the legend of Figure 3? These are box-plots, and by no means should be used to compare distribution functions (they provide way too few information - about specific quantiles).
  Reply: *We rephrased this sentence to 'the seasonal and monthly distribution characteristics'. The boxplots are used here to give an impression of the good match between median, upper, and lower quartile. We see the information provided by the boxplots as a proxy for important distribution characteristics. Tails of the observed and simulated distributions are not expected to perfectly match as the theoretical distribution allows for the extrapolation to unobserved values.*
  Modification: p.9, l.236

- **L220-222.** The Authors write: "the temporal correlation characteristics (4a–c), …. are well captured by the simulations as well."
  In my view there is an important difference between the simulated and empirical autocorrelation coefficients. For instance, by eye-balling the median of the orange lines (simulated) for time lag 10 we get a value about equal to 0.5, while the observed one is 0.4. This should be clearly stated in the manuscript. For the readers convenience, I also suggest the inclusion of a line depicting the median of the simulated quantities (orange lines). Providing lines for a low and high quantile (say 0.05 and 0.95) would be nice also. This comment applies for all similar plots throughout this manuscript.
  Reply: *As illustrated by Figure 4, the simulated temporal correlation characteristics are very close to the observed ones. We depict individual lines for individual realizations to communicate the variability inherent in the simulations to the reader.*

- **L223-225.** The Authors write: "Both high- and low- extremes are realistically modeled as illustrated by the distributions of the above and below threshold events of the four catchments in the Pacific Northwest (Fig. 5)."
  To avoid confusion with actual distribution functions, e.g., the Kappa, instead of using the phase "as illustrated by the distributions of the above and below threshold events", I would suggest the use of phrase similar to: "as illustrated box-plots of Fig.5, constructed by the values of the above and below threshold events".
  Reply: *We rephrased the sentence to 'as illustrated by the boxplots depicting the distribution of the above and below threshold events.'*
  Modification: p.13, l.240

- **L223-225.** Please remind the reader the selected threshold values.
  Figure 6 (legend). I assume that you wanted to write: "… (b) for the **three** catchments in the….".

Reply: *We actually meant to write for the four catchments in the three example regions. The 'three' was added for clarification.*
Modification: p.14, caption of Figure 6

- **L230-233.** The Authors write: "This visual impression of a good performance with respect to the reproduction of spatial correlations in daily discharge data is confirmed by comparing observed and stochastically simulated cross-correlation functions for the catchments in the Pacific Northwest (Fig. 7). Both the shape and magnitude of the cross-correlation functions are well simulated".
  I am afraid that this sentence is needs some refinement, since as with the case of auto-correlation coefficients, Fig. 7 shows a significant deviation of the simulated cross-correlation coefficients from the observed ones. This should be reflected in the manuscript. For the readers convenience, I would also suggest the inclusion of a line depicting the median of all simulations (orange lines). Providing lines for a low and high quantile (say 0.05 and 0.95) would be nice also.
  Reply: *Figure 7 shows that the ccfs derived from the simulations are very close to those derived from the observations. For a reference we provide Figure 1 in response to reviewer 1, which shows that a model where phases are randomized for each catchment individually would result in a poor reproduction of observed ccfs.*

- **Figure 7.** There is something wrong with the labeling of the panels (i.e., multiple panels are labeled as ii, iii, iv, while some others are completely unlabeled).
  Reply: *Each row and column receives a label for the catchment it refers to. These are pairwise plots where cross-correlations are computed for pairs of stations. This specification was added to the caption.*
  Modification: p.15, caption of Figure 7

- **239.** The Authors write: "…but also for extreme values as illustrated by the peak-over-threshold (POT) values for the different stations in the three illustration regions (Fig. 9). These results show that besides regional flood co-occurrences, the temporal clustering behavior of events is also reproduced."
  In my view, Figure 9 is not very informative (the y-axis has been omitted intentionally?). Can you please provide an alternative figure, as well as a quantitative metric, quantifying the reproduction of temporal clustering behavior of events?
  Reply: *We labeled the y-axis with 'event occurrences'. We think that the figure gives a good impression on whether the temporal clustering features of the observations are reproduced by the simulations. Finding intuitive quantitative measures for this is challenging. However, we computed differences in the mean inter-event duration (i.e. time elapsing between two successive events) of observed and simulated series and the standard deviation of inter-event duration for events where 1, 2, and 3 stations were jointly affected, respectively. We find that over all three regions, relative differences in mean and standard deviation of inter-event duration lies mostly below 10%. However, we find that a visual comparison is most effective here to demonstrate the value of the simulation approach.*
  Modification: p.17, Figure 9

- **Discussion section.** In my view all the above points should be discussed in this section, highlighting also the limitations of the presented method. Further to these, it should be noted that the proposed model has been tested for multivariate problems involving 4 processes, as well as the method is capable of generating synthetic timeseries with length equal to the observed one (I haven't read how to handle the case where one wants to generate longer timeseries – is it possible to generate synthetic timeseries with length different than the observed one?).

Reply: *In the discussion we address the limitations of our approach by stating that the approach only allows for the simulation of time series of the same length as the observed series (l 257-259). We would like to highlight here that the modeled range of dependence is also limited to the one in the observed series if one very long time series is generated. We also discuss the disadvantages related to working with distributions fitted to daily values.*

- **L258-259.** The Authors write: "This difference between methods may be related to the fact that the wavelet transform compared to the Fourier transform does not necessitate a transformation to the normal domain, and a back transformation to the domain of the skewed distribution, which has been shown to weaken spatial correlations".
  I think that the answer is is much simpler than the one stated above. The former method is simply designed for the simulation of univariate processes, i.e., not to account for the cross-correlations (or cross-spectrum) among processes. I suggest the Authors to consider more carefully the "mechanics" of the aforementioned methods, and revise the sentence accordingly. Also, as mentioned before, the comment on "weaken spatial correlations" applies for all "types" of correlations (that emerge from the mapping/transformation from the Gaussian to the actual domain) – not only spatial. Particularly, in the case of stochastic processes, it also applies for the auto-correlation structure of a stationary processes, as well as for the season-to-season correlations of a cyclostationary process (Tsoukalas et al., 2018a, 2017). It also holds for multivariate cases.
  Reply: *We here specifically focused on the spatial aspect because the focus of our study was on generating spatially consistent time series and event sets. The Fourier-based method results in much weaker spatial correlations than the wavelet-based approach even if phases are randomized in the same way across catchments.*

- **L270-271.** The Authors write: "Thanks to a spatio-temporal model based on phase randomization, temporal short- and long range dependencies, non-stationarities, and spatial dependencies are reproduced."
  Please consider my comments on the parameterization (i.e., number of parameters), as well as on the performance of the model and revise this sentence accordingly.
  Reply: *We think that this conclusion is supported by the results presented in our model evaluations throughout the results section, which is also acknowledged by Reviewer 2.*

- **A general comment.** A final comment regards the title of the manuscript, which is: "Stochastic simulation of streamflow and spatial extremes: a continuous, wavelet-based approach". By reading the paper I see that the Authors pay special focus on the reproduction of extremes, but it is not clear to me why this model is different from any other in that aspect (e.g., see those mentioned above)? What makes this model suitable when aiming to reproduce extremes? Other similarly parameterized models behave differently? If yes, why? I believe that a theoretical justification or even an empirical comparison with alternative model(s) would be particularly useful and an added value for the paper.
  Regards,
  Ioannis Tsoukalas
  Reply: *We show that PRSim.wave is capable of reproducing spatial dependencies in extremes which has been shown to be challenging in previous studies [Sharma et al., 1997; Caraway et al., 2014]. Our wavelet-based approach has the advantage of being non-parametric in terms of the spatio-temporal model. It avoids assumptions on the distribution and dependence structure of the data. We combine this non-parametric spatio-temporal model with a flexible four-parameter distribution allowing for the representation of a wide range of tail behaviors. 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.

Stochastic simulation of streamflow and spatial extremes: a continuous, wavelet-based approach.

Manuela I. Brunner[1] and Eric Gilleland[1]

[1]Research Applications Laboratory, National Center for Atmospheric Research, Boulder CO, USA

Correspondence: Manuela I. Brunner (manuelab@ucar.edu)

Abstract. Stochastically generated streamflow time series are used for various water management and hazard estimation applications. They provide realizations of plausible but yet unobserved streamflow time series with the same temporal and distributional characteristics as the observed data. However, the representation of non-stationarities and spatial dependence among sites remains a challenge in stochastic modeling. We investigate whether the use of frequency-domain instead of time-domain models allows for the joint simulation of realistic, continuous streamflow time series at daily resolution and spatial extremes at multiple sites. To do so, we propose the stochastic simulation approach called Phase Randomization Simulation using wavelets (*PRSim.wave*) which combines an empirical spatio-temporal model based on the wavelet transform and phase randomization with the flexible four-parameter kappa distribution. The approach consists of five steps: (1) derivation of random phases, (2) fitting of kappa distribution, (3) wavelet transform, (4) inverse wavelet transform, and (5) transformation to kappa distribution. We apply and evaluate *PRSim.wave* on a large set of 671 catchments in the contiguous United States. We show that this approach allows for the generation of realistic time series at multiple sites exhibiting short- and long-range dependence, non-stationarities, and unobserved extreme events. Our evaluation results strongly suggest that the flexible, continuous simulation approach is potentially valuable for a diverse range of water management applications where the reproduction of spatial dependencies is of interest. Examples include the development of regional water management plans, the estimation of regional flood or drought risk, or the estimation of regional hydropower potential among others.

Keypoints:

1. Stochastic simulation of continuous streamflow time series using an empirical, wavelet-based, spatio-temporal model in combination with the parametric kappa distribution.

2. Generation of stochastic time series at multiple sites showing temporal short- and long range dependence, non-stationarities, and spatial dependence in extreme events.

3. Implementation of *PRSim.wave* in R-package *PRSim*: Stochastic Simulation of Streamflow Time Series using Phase Randomization.

1 Introduction

Stochastic models are used to generate long time series or large event sets showcasing the full variability of a phenomenon. In hydrology, we use stochastically generated time series or event sets to refine water management plans, to get a better idea of potential reservoir inflows, or to develop suitable adaptation strategies for droughts and floods. If the focus is on such extreme events, event-based instead of continuous simulation approaches are often employed (e.g. Bracken et al., 2016; Diederen et al., 2019; Quinn et al., 2019). This strategy requires an *a priori* definition of extreme events and leads to a loss of temporal information, e.g. on the season of occurrence. In contrast, continuous simulation approaches allow for the simulation of time series including, but not limited to, extreme events which are provided together with their time of occurrence. Such continuous approaches enable the investigation of drought and flood characteristics going beyond magnitude and extending to timing, duration, or volume. However, the model representation of this additional information on timing adds some challenges because the temporal characteristics of the data need to be represented in addition to their distribution. These temporal characteristics include fluctuations on short and long timescales (Rajagopalan et al., 2010) and potential non-stationarities in the data (Nowak et al., 2011).

There exists a variety of continuous, stochastic modeling approaches (corresponding to discrete-time models in the stochastic literature) which differ in their capability of representing distributional and/or temporal characteristics in the data. Here, we focus on direct modeling approaches that directly simulate streamflow using a stochastic model as opposed to indirect approaches which use a hydrological model to transform stochastically generated precipitation to streamflow. The most commonly used stochastic simulation approaches belong to the two classes of parametric and nonparametric models. Parametric models include autoregressive moving average (ARMA) models and their modifications (Stedinger and Taylor, 1982; Papalexiou, 2018), and fractional Gaussian noise models (Mandelbrot, 1965) comprising fast fractional Gaussian noise models (Mandelbrot, 1971), broken line models (Mejia et al., 1972), and fractional autoregressive integrated moving average models (Hosking, 1984). Nonparametric models include kernel density estimation (Lall and Sharma, 1996; Sharma et al., 1997) and various bootstrap approaches such as simple bootstrap, moving block-bootstrap, nearest-neighbor bootstrap (Salas and Lee, 2010; Herman et al., 2016), matched-block bootstrap (Srinivas and Srinivasan, 2006), or maximum-entropy bootstrap (Srivastav and Simonovic, 2014). These parametric and non-parametric methods have different strengths and weaknesses as e.g. discussed in Rajagopalan et al. (2010) and the representation of spatial dependence in such time-domain models is challenging. For parametric models, the number of parameters grows rapidly with the number of locations (Caraway et al., 2014) and the spatial dependence structure is similar for high, medium, and low flow values, which is not the case for observed data (Sharma et al., 1997). Similarly, non-parametric approaches are not effective for multiple-site streamflow generation because of the high dimension of the problem (Nowak et al., 2010).

In contrast to most time-domain models, frequency-domain models allow for the simulation of surrogate data with the same Fourier spectra as the raw data (Theiler et al., 1992) and can easily be extended to multiple sites (Prichard and Theiler, 1994; Schreiber and Schmitz, 2000). Such methods are based on the randomization of the phases of the Fourier transform and are known as the amplitude-adjusted Fourier transform (AAFT) (Lancaster et al., 2018). They have only recently been used in

hydrology for other applications besides hypothesis testing, trend detection (Radziejewski et al., 2000), and the identification of nonlinearities in time series (Schmitz and Schreiber, 1996; Kugiumtzis, 1999; Venema et al., 2006; Maiwald et al., 2008). Serinaldi and Lombardo (2017) used an iterative AAFT method to generate binary series of rainfall occurrence and non-

60 occurrence. Brunner et al. (2019) used a phase randomization approach in combination with the flexible four-parameter kappa distribution (Hosking, 1994) to simulate continuous discharge time series including unobserved extremes. The approach is implemented in the R-package Stochastic Simulation of Streamflow Time Series using Phase Randomization *PRSim* (Brunner and Furrer, 2019) and can be applied to both individual and multiple sites. This phase randomization approach has been shown to reproduce the distributional and temporal characteristics of the data at individual sites well (Brunner et al., 2019; Brunner

65 and Tallaksen, 2019). However, the approach has some deficiencies when applied to multiple sites because spatial dependencies in both daily discharge and extreme events are underestimated. In addition, the approach does not allow for the consideration of non-stationarities.

In contrast to the Fourier transform, the wavelet transform allows for the representation of non-stationarities in time series (Rajagopalan et al., 2010). For a short introduction to the wavelet transform see Sect. 2. In addition, it may help to improve

70 the representation of spatial dependencies because it does not require a transformation to the normal distribution and back to the original, skewed distribution, which usually weakens spatial correlations (Embrechts et al., 2010). This weakening is because phase randomization preserves the cross-correlation in the normal domain but not necessarily in the domain of the original distribution as linear correlation is not invariant under non-linear strictly increasing transformations. Because of its favorable properties, the wavelet transform has been used in stochastic time series generation in various ways. Kwon et al.

75 (2007) proposed a wavelet-based autoregressive modeling approach (WARM) suitable for systems with a quasi-periodic long memory behavior. They used the continuous wavelet transform to decompose a time series into several statistically significant components. Each of these components was fitted using a linear AR model which was subsequently used for simulation. Later, Nowak et al. (2011) adapted this WARM approach such that it can handle non-stationarities. Another possibility of handling non-stationarities is the wavelet-based time series bootstrap model introduced by Erkyihun et al. (2016) which generates

80 wavelet-derived signal components with a block resampling approach therefore replacing the AR component of the WARM.

An alternative to these approaches where only certain signal components are modified are approaches that randomize the wavelet coefficients for all components. These approaches typically perform a discrete wavelet decomposition using real wavelet functions (as opposed to complex wavelet functions), randomize the real-valued wavelet coefficients (i.e. amplitudes) and then invert the transform to produce a new realisation of a time series (Breakspear et al., 2003; Keylock, 2007; Wang

85 et al., 2010; Lancaster et al., 2018). However, a completely random shuffling of coefficients destroys their periodicity. To overcome this drawback, Breakspear et al. (2003) used block resampling on coefficients and Keylock (2007) introduced a threshold criterion that is used to pin particular wavelet coefficients to the same position in the wavelet domain for both the original and the surrogate data. Despite these workarounds, some of the long-term periodicities and/or non-stationarities may not be preserved (Breakspear et al., 2003). Using a continuous instead of a discrete wavelet transform allows for the use of

90 complex wavelet functions which provide complex-valued coefficients containing information on the phase in addition to amplitude. The use of this additional phase information in the randomization process instead of the amplitude can prevent the

loss of non-stationarities. To generate non-stationary surrogate time series, Chavez and Cazelles (2019) extended the classic phase-randomised surrogate data algorithm to the time-frequency domain using a dataset of weekly measles notifications and an electroencephalographic recording. They randomized the phases resulting from the continuous wavelet transform instead of the real valued amplitudes. This approach has the advantage of being non-parametric, avoids assumptions on the distribution and dependence structure of the data, allows for multiple realizations, and can be extended to multiple sites by randomizing the phases for multiple time series in the same way (Prichard and Theiler, 1994; Schreiber and Schmitz, 2000).

We investigate whether such a wavelet-based phase randomization approach allows for a realistic representation of spatial dependence in both continuous streamflow time series and spatial extremes. To do so, we propose a continuous wavelet-based approach for the stochastic generation of streamflow time series, hereafter referred to as *PRSim.wave* which is based on the empirical spatio-temporal model used by Chavez and Cazelles (2019), i.e. wavelet-based phase randomization. This empirical approach is supposed to overcome difficulties in modeling spatial dependence over a large domain as occurs when using parametric models (Caraway et al., 2014). We combine this empirical spatio-temporal model with a parametric distribution to enable the generation of hydrological extremes exceeding the range of the observed values. By doing so, we extend the approach by Brunner et al. (2019) from the Fourier to the wavelet domain therefore improving the representation of spatial dependencies and non-stationarities.

We implement *PRSim.wave* in the R-package *PRSim* (Brunner et al., 2019) and apply and evaluate it on a large dataset of 671 catchments in the contiguous United States. Our evaluation results indicate that the flexible, continuous simulation approach can be used for a diverse range of water management applications requiring continuous discharge time series at multiple sites or for regional drought and flood hazard assessments not limited to peak magnitude or maximum deficit but extending to event duration and volume.

2 Theoretical background

Wavelet decomposition transforms a one-dimensional time series to a two-dimensional time-frequency space (Daubechies, 1992; Torrence and Compo, 1998) using either a discrete or continuous wavelet transform. Both transforms decompose a hydrological series into a set of coefficients, representing different scales or frequency bands (Sang, 2012). The coefficients of the discrete transform are real numbers representing amplitudes. In contrast, the coefficients derived from the continuous transform have a real and an imaginary part corresponding to an amplitude and phase, respectively. This additional information on the phases makes complex wavelet functions more suitable for capturing oscillatory behavior (Torrence and Compo, 1998).

The wavelet function used for the transform should reflect the features present in the time series. Because of its smooth features, the Morlet wavelet has often been used in hydrological applications (Labat et al., 2005; Lafrenière and Sharp, 2003; Schaefli et al., 2007) and is given by (Torrence and Compo, 1998):

$$\psi_0(\eta) = \pi^{-1/4} e^{i\omega_0 \eta} e^{-\eta^2/2}, \tag{1}$$

where η is a non-dimensional time parameter, ω_0 is the non-dimensional frequency, and $i = \sqrt{-1}$ is the imaginary unit. The continuous wavelet transform is defined as the convolution of a time series x_n of length n with a scaled version of $\psi_0(\eta)$:

$$W_n(h) = \sum_{n'=0}^{N-1} x_{n'} \psi^* \left[\frac{(n'-n)\delta t}{h} \right],$$

(2)

where the (*) indicates the complex conjugate. Varying the wavelet scale h and translating along the localized time index n allows for showing the amplitude of certain features versus scale and how the amplitude varies with time and scale. An inverse filter can be used to reconstruct the original time series as the sum of the real part of the wavelet transform over all scales $(h_1, ..., h_J)$:

$$x_n = \frac{\delta j \delta t^{1/2}}{C_\delta \psi_0(0)} \sum_{j=0}^{J} \frac{R(W_n(h_j))}{h_j^{1/2}},$$

(3)

where the factor $\psi_0(0)$ removes the energy scaling, $h_j^{1/2}$ converts the wavelet transform to an energy density, and the factor C_δ is a constant for each wavelet function (0.776 for the Morlet wavelet).

3 Data and Methods

We develop and apply the stochastic simulation approach *PRSim.wave* using a large-scale dataset of 671 stations in the contiguous United States (CONUS). We evaluate the approach with respect to distributional and temporal characteristics at individual sites and with respect to spatial dependencies across multiple sites in general and for floods in particular.

3.1 Study area

The 671 catchments in the United States cover a wide range of discharge regimes minimally influenced by human activity (Newman et al., 2015; Addor et al., 2017). Daily discharge data were downloaded for the period 1981–2018 from the USGS water information system (USGS, 2019) using the R-package dataRetrieval (De Cicco et al., 2018).

For illustration and validation purposes, we select three regions which are distinct in terms of their hydrological regimes and their flood behavior. Flood similarity regions are determined using hierarchical clustering (Gordon, 1999) on a distance matrix computed from the F-madogram, which is a measure of extremal dependence between pairs of stations (Cooley et al., 2006). The clustering is applied to the 671 catchments and results in 15 clusters among which we select 3 for illustration purposes: (1) catchments in the lower-elevation coastal Pacific Northwest characterized by high mean annual precipitation and a strong discharge seasonality experiencing floods mainly in December and January (2) catchments in Texas with low mean discharge, weak seasonality, and flood occurrence in spring to fall, and (3) catchments in the Mid-Atlantic coastal plain and central Appalachian Mountains with a strong streamflow seasonality showing flood occurrence for much of the year except early-mid summer. For each of these regions, four catchments are chosen for illustration purposes (Fig. 1).

5
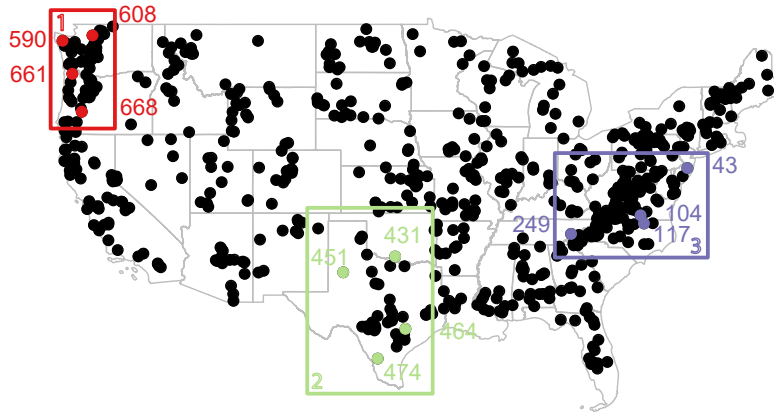
Figure 1. Location of 671 stations in the dataset and of four catchments chosen per example region: (1) Pacific Northwest (red; 590, 608, 661, and 668), (2) Texas (light green; 431, 451, 464, and 474), and (3) Mid-Atlantic (purple; 43, 104, 117, 249).

3.2 Simulation procedure

The stochastic simulation procedure *PRSim.wave* for multiple sites consists of 5 main steps (Fig. 2), which can be run p times to generate p spatially consistent time series over n sites at daily resolution:

1. **Derivation of random phases**: A random discharge time series (white noise) of the same length as the input series is sampled from a normal distribution with mean 0 and standard deviation 1 (Chavez and Cazelles, 2019). We also tested the kappa distribution, which did, however, not significantly change the model performance. The wavelet transform is applied to the white noise series using the Morlet wavelet (Eq. (1)) for $h = 100$ wavelet scales. 100 scales are used as using only few scales (e.g. 20) results in reconstructed time series that do not reflect all the necessary detail while further increasing the number of scales no longer improves reconstruction performance. The phases of the wavelet transform are derived. These same phases are used for all the sites considered to retain spatial dependencies among sites (Prichard and Theiler, 1994; Schreiber and Schmitz, 2000).

2. **Fitting of kappa distribution**: The flexible four-parameter kappa distribution (Hosking, 1994) is fit to the daily values of the observed input time series using L-moments. These daily distributions will be used for the transformation in Step 5 to simulates extreme values going beyond the empirical distribution. The cumulative distribution function of the kappa

6
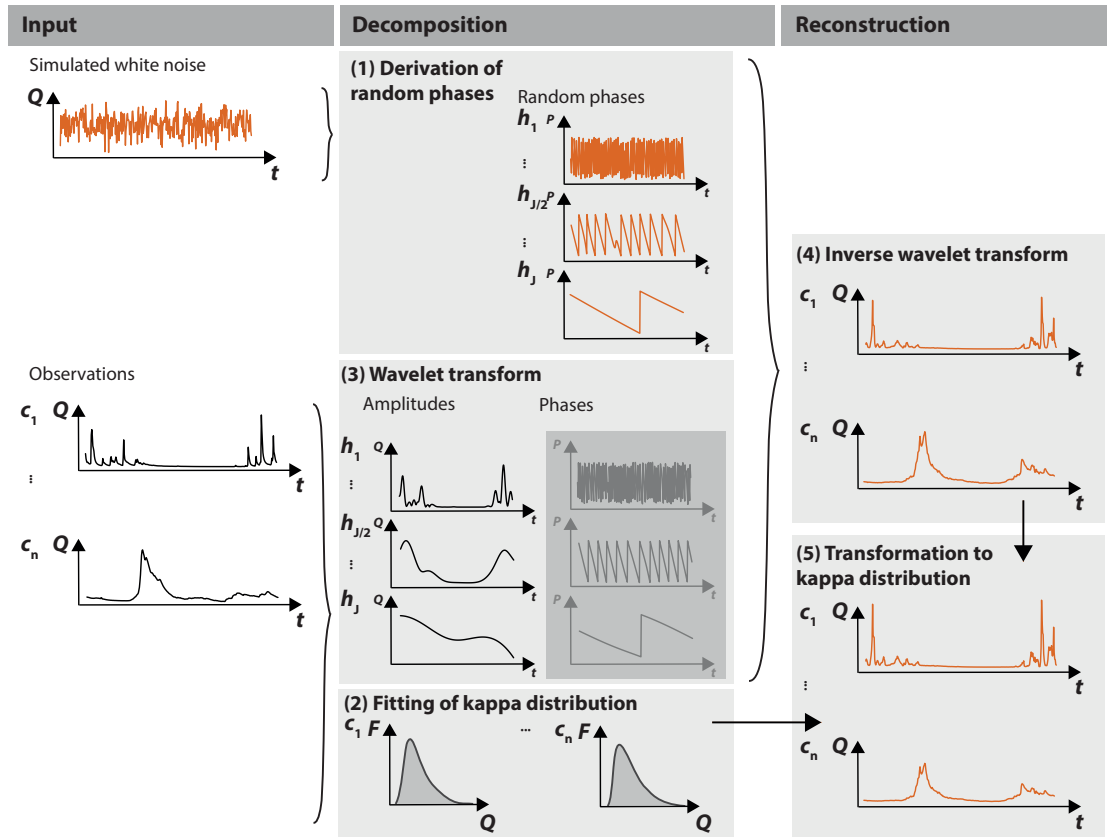
Figure 2. Illustration of stochastic simulation approach *PRSim.wave*: (1) Derivation of random phases which are uniformly distributed over the range of $-\pi$ to π, using a white noise time series to be used for the simulation at all stations; for each station: (2) Fitting of kappa distribution to the observed streamflow time series using L moments; (3) Wavelet transform to derive the amplitudes and phases of the time series; (4) Inverse wavelet transform recomposing a time series using the random phases from Step 1 in combination with the amplitudes of Step 3; and (5) Transformation to the kappa distribution using the kappa distribution fitted in Step 2. Steps 1–5 are repeated p times to generate p time series.

distribution is expressed as

$$F(x) = \left\{ 1 - \xi[1 - \kappa(x - \mu)/\sigma]^{1/\kappa} \right\}^{1/\xi}, \qquad (4)$$

where μ is the location parameter, σ is the scale parameter which must be positive, and κ and ξ are the shape parameters (R-package homtest Viglione, 2009).

The kappa distribution was found suitable for fitting observed streamflow data in U.S. catchments (Blum et al., 2017). A suitable fit is also found for our data as confirmed by the Kolmogorov–Smirnov and Anderson–Darling tests (Chernobai et al., 2015) which did not reject the null hypothesis at $\alpha = 0.05$ for most catchments. We fit a separate distribution for each day using a moving window approach to take into account seasonal differences in the distribution of daily streamflow values. To do so, we use the daily values in a 30-day window around the day of interest. This procedure guarantees a large enough sample for the parameter fitting procedure, and allows for smoothly changing distributions along the year. For leap years, flows from February 29 are removed to maintain constant sample sizes across years as in Blum et al. (2017). In a few regions with many zero discharge values (e.g. some catchments in the Great Plains) fitting the kappa distribution is not possible and we therefore use the empirical distribution instead. After fitting the daily distributions, the mean discharge is subtracted from the observed discharge values to center the observations.

3. **Wavelet transform**: The wavelet transform (Eq. (2)) is applied to the input time series using the complex-valued Morlet wavelet (Eq. (1)) to derive the amplitudes for $h = 100$ number of wavelet scales. The complex part of the wavelet transform $W_n(h)$ comes in via the complex conjugate $*$ in Eq. 2. The complex conjugate is derived using the modulus and argument, i.e. the phases, of the complex numbers contained in the Morlet wavelet (Eq. 1). The wavelet coefficients $W_n(h)$ resulting from the transform are also complex numbers, where the argument of the complex wavelet value $W_n(h)$ corresponds to the wavelet phase.

4. **Inverse wavelet transform**: Reconstruction of the series in the time domain by using the inverse wavelet transform (Eq. (3)) combining the random phases derived in Step 1 and the amplitudes derived in the previous step. During back-transformation in Eq. 3, the phases come back in when deriving the real part at each scale h through $R(W_n(h_j))$.

5. **Transformation to kappa distribution**: The simulated values are transformed to the kappa domain using the fitted daily kappa distributions from Step 2. For each day, a random sample is generated from the fitted, daily kappa distribution. Negative simulated values are replaced by 0. The simulated values are replaced by the values generated from the kappa distribution using rank-ordering. This procedure is repeated for each day in the year.

Step 1 is run only once per iteration to maintain spatial dependencies in the data while Steps 2–5 are run for each station separately.

3.3 Evaluation

We run the stochastic simulation algorithm for the 671 catchments in the dataset $n = 100$ times to generate 100 time series of the same length as the observed time series, i.e. 38 years. We look at (1) individual sites to evaluate the general distributional

and temporal correlation characteristics as well as the reproduction of high- and low flows; (2) three sets of four stations each to evaluate the spatial consistencies in daily discharge and floods: Pacific Northwest, Texas, and Mid-Atlantic (Fig. 1); and (3) at the set of 671 catchments to evaluate spatial consistencies in floods across large scales.

The evaluation at individual sites encompasses a comparison of observed and simulated distributional and temporal discharge characteristics. The distributional characteristics considered are the mean annual hydrograph showing variation of flow with season, three years of daily data illustrating the overall behavior of the series, the seasonal distributions (winter: Dec–Feb, spring: Mar–May, summer: June–Aug, fall: Sept–Nov), and monthly mean, maximum and minimum values. The temporal characteristics considered include the autocorrelation (acf) and partial autocorrelation functions (pacf) measuring the strength of temporal dependence for different time lags, the power spectrum indicating how power varies with frequency and showing high values at those frequencies that correspond to strong periodic components (Shumway and Stoffer, 2017), the normalized average power per scale over all time steps indicating oscillations, and the scale-averaged wavelet power (Erkyihun et al., 2016) for the three scales with the highest average power revealing non-stationarities in oscillations. To evaluate the capability of the approach to simulate extreme values, we compare observed and simulated low- and high-flows, i.e., flows below or above a threshold corresponding to the 0.05 and 0.95 percentiles, respectively.

The evaluation at multiple sites comprises both an assessment of how the general spatial dependence structure in the data is reproduced and an assessment of how the spatial dependence in high extremes is captured. The assessment of the general dependence structure encompasses a comparison of observed and simulated discharge time series for the catchments in the three example regions, a comparison of pairwise observed and simulated cross-correlations for the example stations in the Pacific Northwest region, and a comparison of variograms of the observed and simulated series across all stations (R-package SpatialExtremes; Ribatet, 2019) given by (Cressie, 1993):

$$2\gamma(\mathbf{s}_1, \mathbf{s}_2) = var(Z(\mathbf{s}_1) - Z(\mathbf{s}_2)) = E[(Z(\mathbf{s}_1) - Z(\mathbf{s}_2))^2], \tag{5}$$

where Z is a random variable (here, streamflow) measured at two locations s_1 and s_2. In order to be able to discern the shapes of the variograms, they are first smoothed using splines.

For assessing how spatial dependencies in extremes are reproduced, we first compare observed and simulated times of occurrences of flood events for the catchments in the three example regions. We then compare observed to simulated F-madograms for flood events across all stations. The F-madogram is a measure of spatial dependence that compares the ordering of extreme events between two time series of extreme events (Cooley et al., 2006) and is expressed as:

$$v^F(d) = \frac{1}{2} E |F[Z(\mathbf{s} + d)] - F[Z(\mathbf{s})]|, \tag{6}$$

where $Z(\mathbf{s})$ are transformed to have Fréchet margins so that $F(s) = \exp(-1/s)$, and d is the distance between a pair of stations (R-package SpatialExtremes; Ribatet, 2019). We finally compute the tail dependence coefficient χ across all stations defined as (Coles, 2001):

$$\chi(u) = Pr\{Y > G^{-1}(u) \mid X > F^{-1}(u)\} = Pr\{V > u \mid U > u\}, \tag{7}$$

where X and Y are uniformly distributed random variables with distribution functions F and G, and u is a threshold (R-package extRemes; Gilleland and Katz, 2016,). Tail dependence estimators depend heavily on the sample size and are subject to large uncertainty given the sample size of 38 years (Serinaldi et al., 2015).

4 Results

4.1 Single-site simulations

Both the distributional and temporal dependence characteristics of the time series at individual sites are well modeled as shown by comparing observed and stochastically simulated time series for the two stations on the Nehalem and Navidad rivers (Fig. 3 and 4). The hydrological regimes (3a), the overall behavior of the time series (3b–c), the seasonal (3d) and monthly distribution characteristics (3e–g) are well captured by the simulations. The simulations, however, are capable of simulating values going beyond the range of the observations as intended by using the kappa distribution. In addition to these distributional characteristics, the temporal correlation characteristics (4a–c), the oscillations in the data (4d), and the non-stationarities in these oscillations (4d) are well captured by the simulations as well.
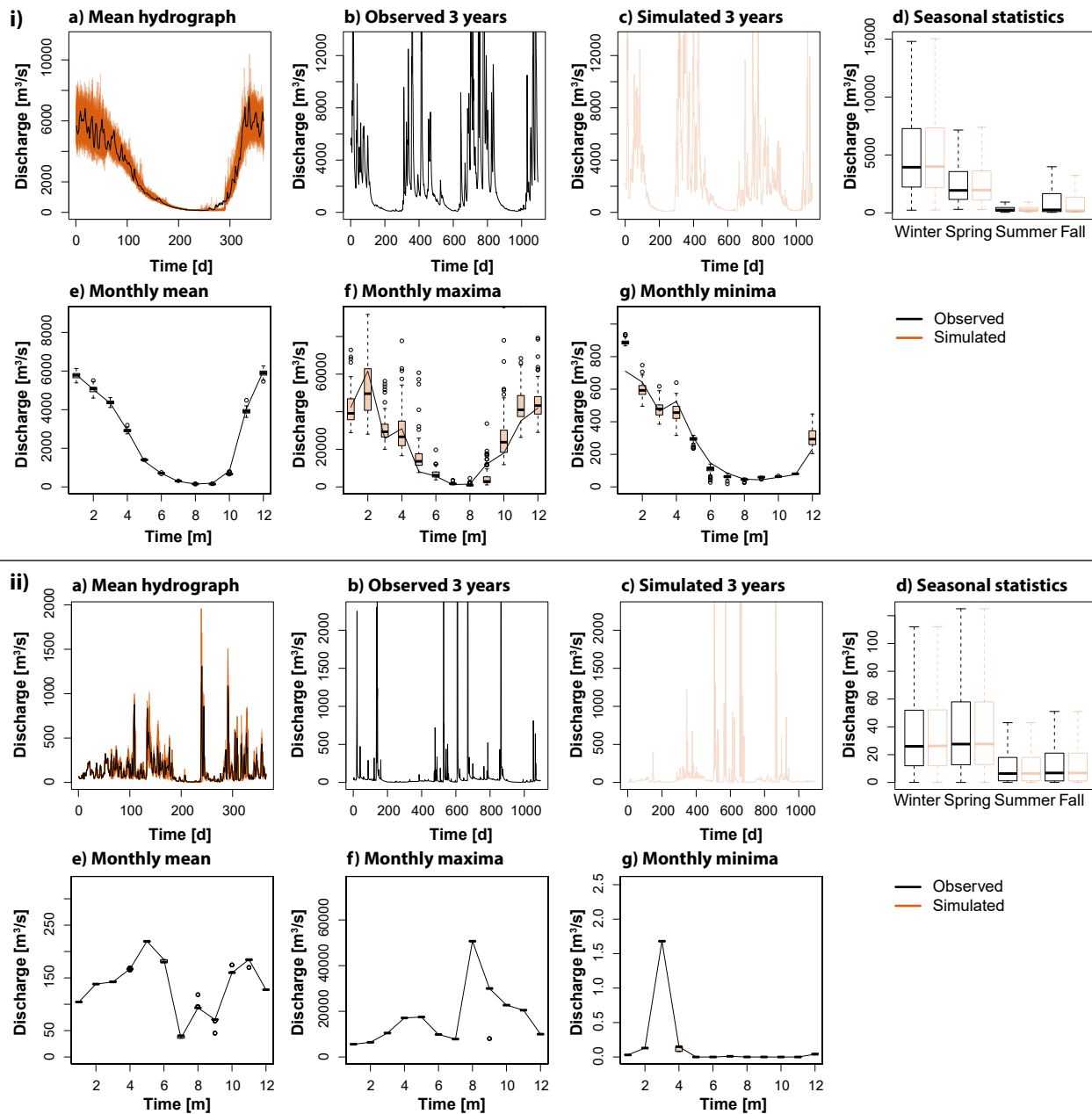
Figure 3. Comparison of observed (black) and simulated (orange) distributional discharge characteristics for (i) the station Nehalem River near Foss, OR (USGS 14301000, id 661) in the Pacific Northwest and (ii) the station Navidad Rv near Hallettsville, TX (USGS 08164300, id 464) in Texas: (a) mean annual hydrographs, (b and c) observed and simulated time series for three years, (d) seasonal discharge distribution characteristics, (e) monthly mean discharge, (f) monthly maximum discharge, and (g) monthly minimum discharge.

11
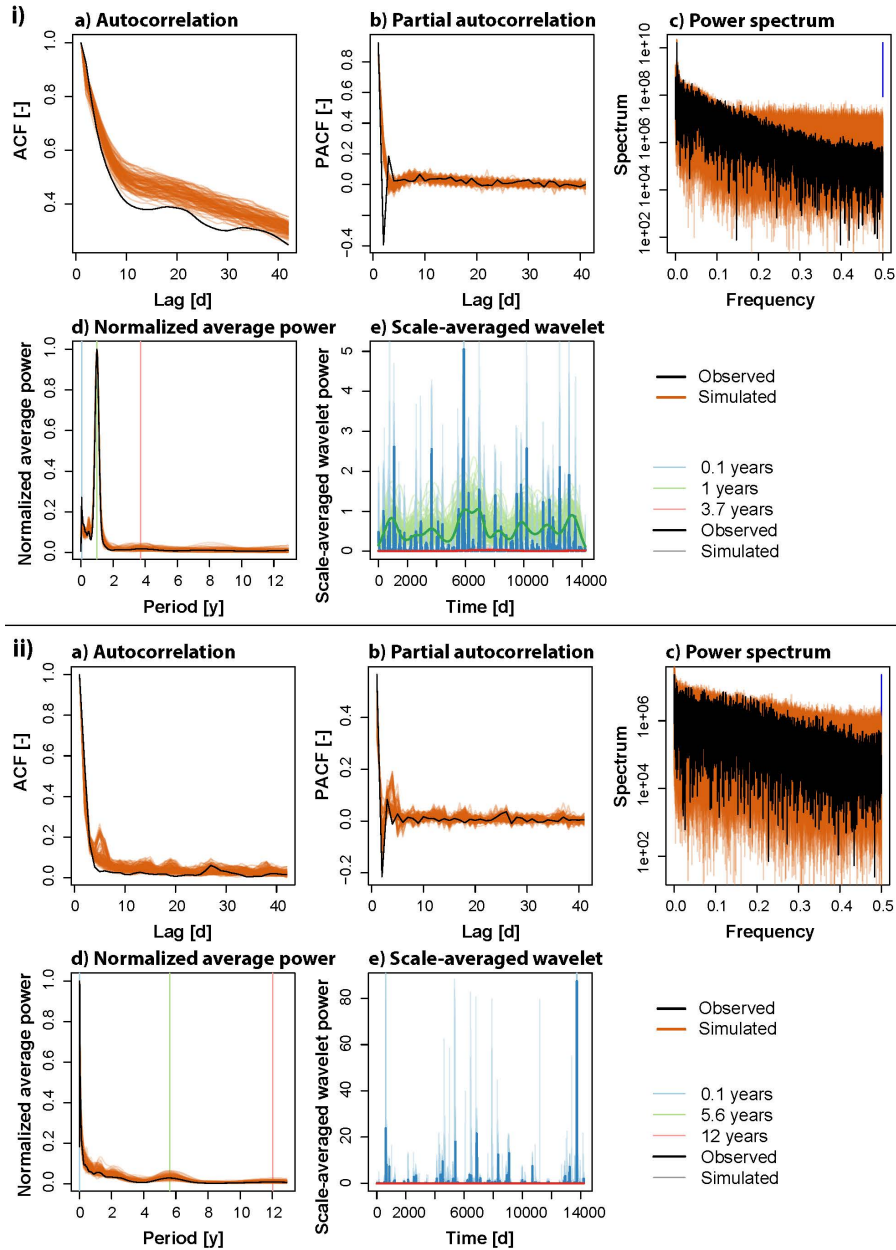
Figure 4. Comparison of observed (black) and simulated (orange) temporal discharge characteristics for (i) the station Nehalem River near Foss, OR (USGS 14301000, id 661) in the Pacific Northwest and (ii) the station Navidad Rv near Hallettsville, TX (USGS 08164300, id 464) in Texas: (a) acf, (b) pacf, (c) power spectrum, (d) normalized average power, and (e) scale averaged-wavelet power.
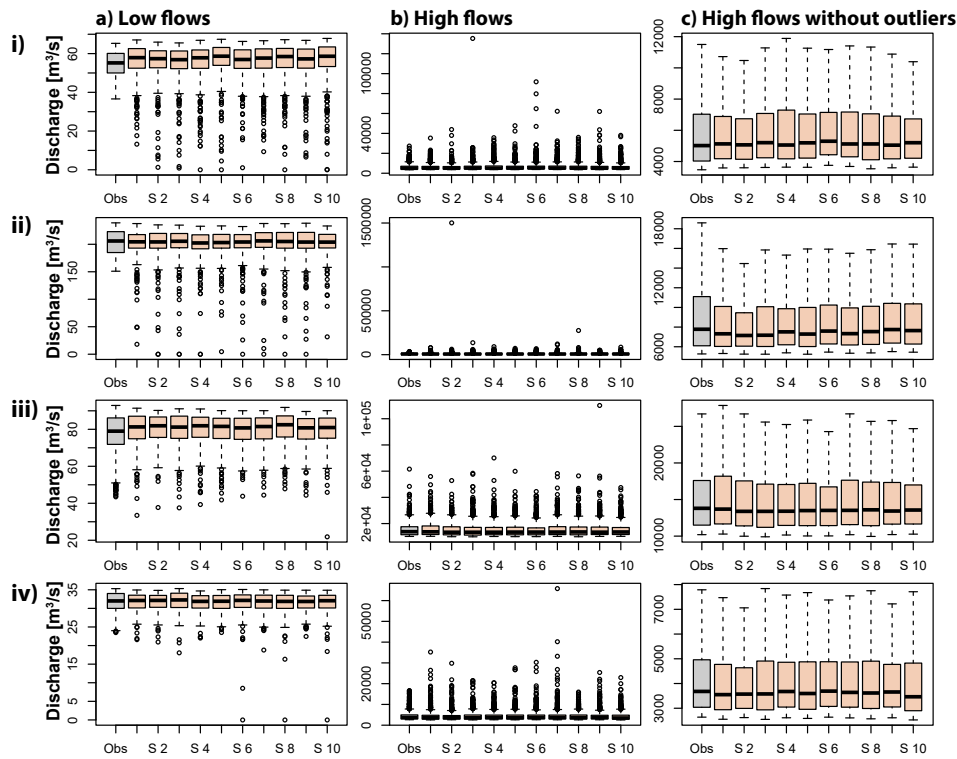
Figure 5. Comparison of observed (grey) (a) low- and (b–c) high-flow distributions (with and without outliers) with simulated distributions of 10 runs (orange) for the four catchments in the Pacific Northwest: (i) Calawah River near Forks, WA (USGS 12043000, id 590), (ii) Stillaguamish River near Arlington, WA (USGS 12167000, id 608), (iii) Nehalem River near Foss, OR (USGS 14301000, id 661), and (iv) Steamboat Creek near Glide, OR (USGS 14316700, id 668).

240   Both high- and low- extremes are realistically modeled as illustrated by the boxplots depicting the distributions of the above and below threshold events of the four catchments in the Pacific Northwest (Fig. 5). While the median of the observed low and high-flow distributions is well met by the simulated medians, again the simulations allow for the generation of extreme low- and high flows going beyond the observed values because of the use of the theoretical kappa distribution.

13

4.2 Multi-site simulations
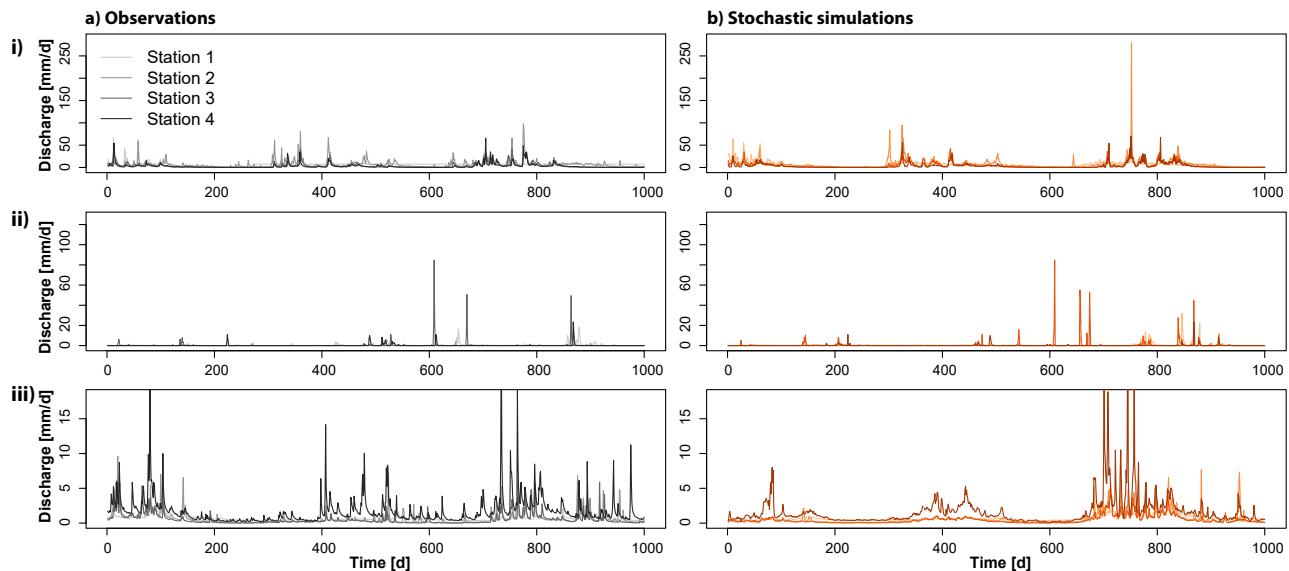
Figure 6. Comparison of three years of multi-site observations (a) with multi-site stochastic simulations (b) for the four catchments in the three example regions (i) Pacific Northwest, (ii) Texas, and (iii) Mid-Atlantic. Each region is displayed on its own scale.

245   The stochastic simulation approach *PRSim.wave* not only allows for the reproduction of the distributional and temporal characteristics of time series at single sites but also for simulating spatially coherent time series at multiple sites (Fig. 6). Independent of the region considered, the simulations realistically represent the observed behavior of the time series. This visual impression of a good performance with respect to the reproduction of spatial correlations in daily discharge data is confirmed by comparing observed and stochastically simulated cross-correlation functions for the catchments in the Pacific Northwest (Fig. 7). Both the

250   shape and magnitude of the cross-correlation functions are well simulated. The good performance in terms of reproducing the general spatial dependence structure in the data can be generalized to other regions as shown by a comparison of observed and simulated variograms (Fig. A2). Figures A1 and A2 in the appendix demonstrate that neither the observed cross-correlation nor variogram 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.

14
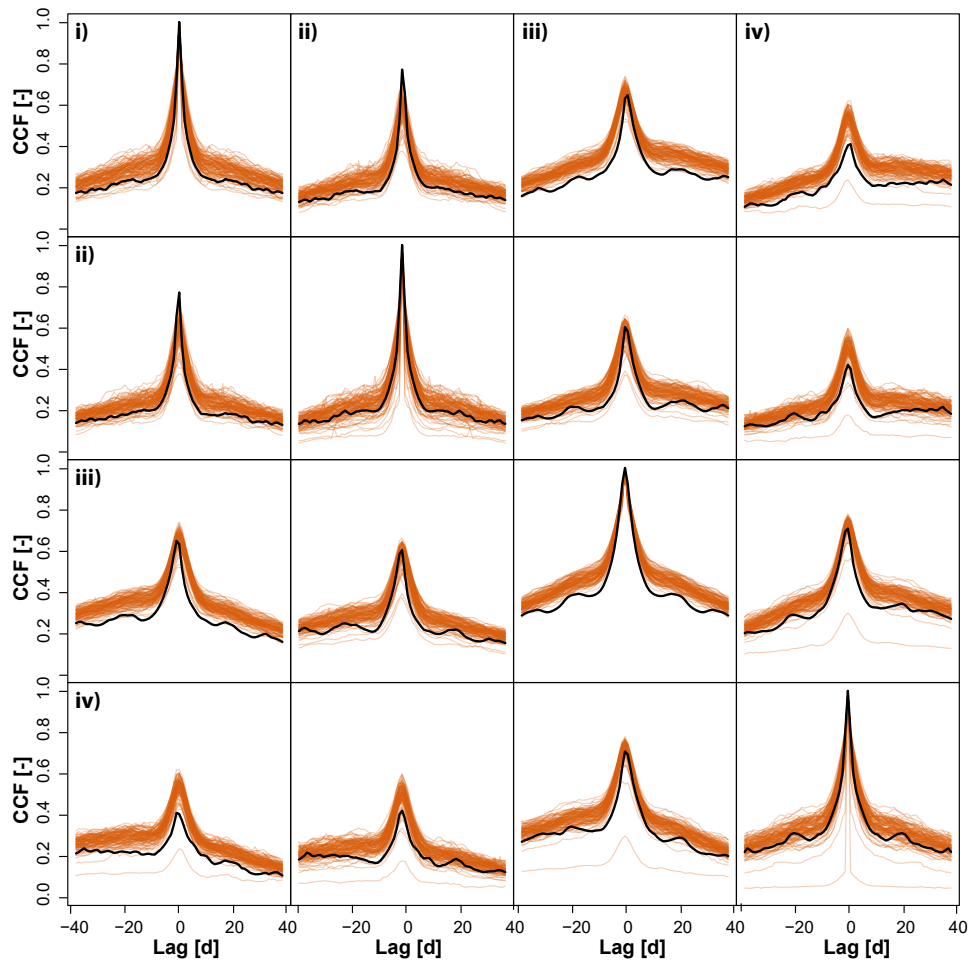
Figure 7. Comparison of observed (black) and simulated (orange) cross-correlation functions (ccfs) for the daily discharge values for pairs of stations in the set of the four catchments (i–iv) in the Pacific Northwest.
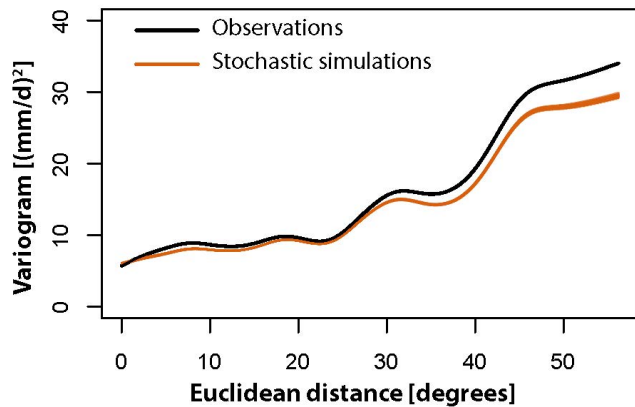
Figure 8. Comparison of observed (black) and simulated (orange) variograms.

255   Spatial dependencies are not only maintained for the bulk of the distribution, by which we mean the part of the distribution excluding extremes or outliers, but also for extreme values as illustrated by the peak-over-threshold (POT) values for the different stations in the three illustration regions (Fig. 9). These results show that besides regional flood co-occurrences, the temporal clustering behavior of events is also reproduced.
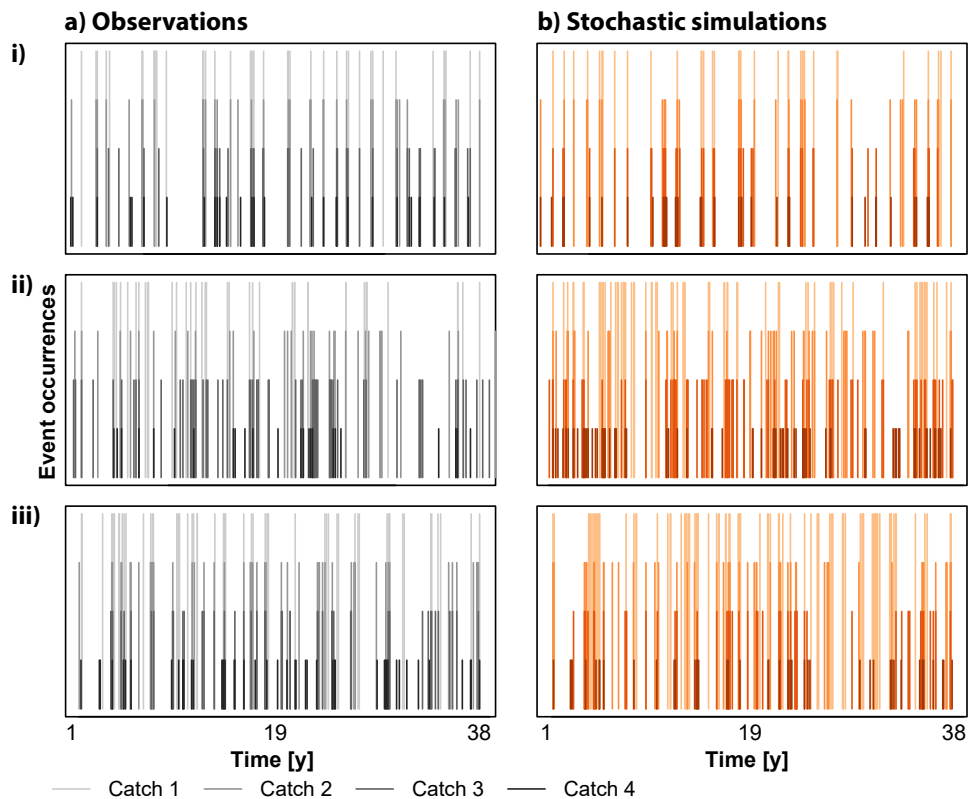
Figure 9. Observed (a; black) vs. stochastically simulated (b; orange) POT events for the four stations (color shadings) in the three regions (i) Pacific Northeast, (ii) Texas, and (iii) Mid-Atlantic. Vertical bars indicate event occurrences over the 38 years.
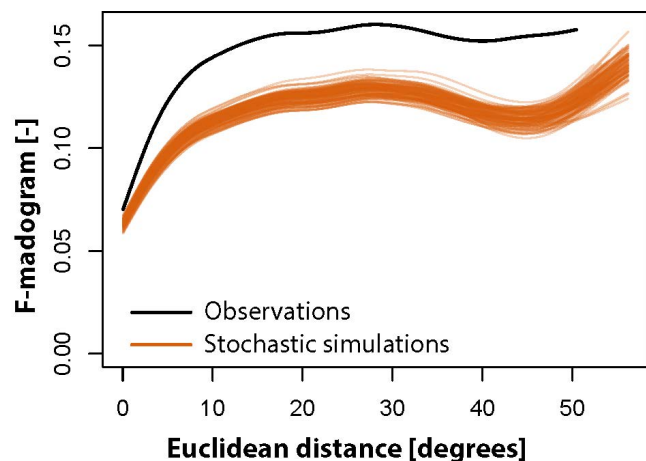
Figure 10. Observed (black) vs. simulated (orange) F-madograms , a measure for the strength of spatial dependence, plotted against Euclidean distance. The lower the value, the higher the dependence between a pair of stations.

The F-madograms shown in Fig. 10 indicate that there is generally good agreement between observed and simulated spatial dependence despite a slight overestimation of the spatial dependence of floods by the stochastic simulations. This overestimation means that a certain pair of stations may experience more joint floods according to the simulations than seen in the observations. The good agreement between observed and simulated spatial dependence and the weak overestimation in spatial dependence is also visible if we look at the tail dependence coefficient χ for the two thresholds 0.8 and 0.95 (Fig. 11). For most pairs of stations, both the simulations and observations indicate no tail dependence (grey dots). If the observations indicate tail dependence, the simulations mostly also simulate upper tail dependence (green dots). Only in very few cases, the simulations do not capture the tail dependence in the observations (black dots). There are, however, quite a few cases where the simulations indicate tail dependence despite its absence in the observations (orange dots). Overall, the model shows good performance in the reproduction of observed spatial (in)dependencies.
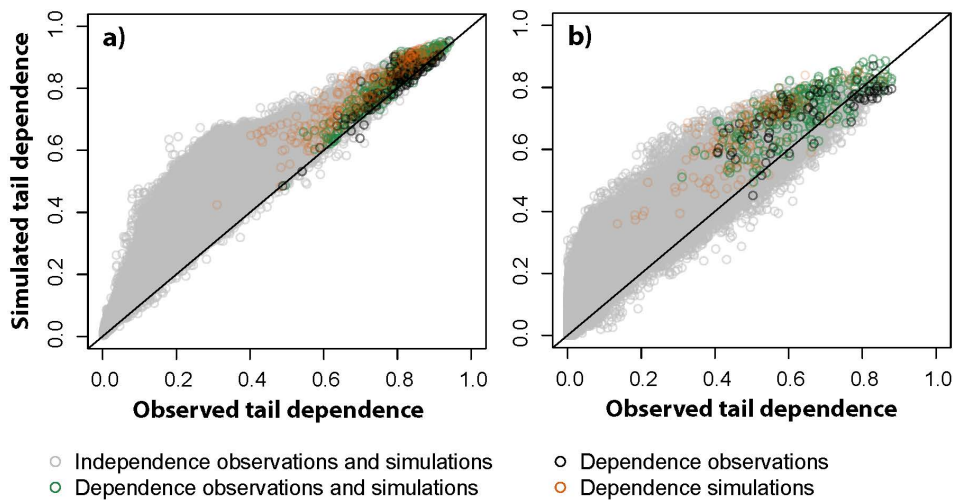
Figure 11. Observed vs. simulated tail dependence coefficient χ for the two thresholds 0.8 (a) and 0.95 (b). Grey dots indicate pairs of stations with no upper tail dependence, green dots pairs where both observations and simulations indicate upper tail dependence, black dots pairs of stations where only the observations indicate tail dependence, and orange dots pairs of stations where only the simulations indicate tail dependence.

5   Discussion

Similar to the Fourier transform based simulation approach *PRSim* by Brunner et al. (2019), the wavelet-based simulation approach *PRSim.wave* presented here allows for the generation of many time series of the same length as the observed series. This means that the representation of temporal dependence is limited to ranges within the length of the observed series. However, the modeled range of dependence is also limited to the one in the observed series if one very long time series is generated. In addition to the representation of temporal dependence, *PRSim.wave* allows for the reproduction of realistic

spatial dependencies both in the general distribution and in extreme events. This representation of spatial dependence is not possible if using the Fourier transform as in *PRSim*. This difference between methods may be related to the fact that the wavelet transform compared to the Fourier transform does not necessitate a transformation to the normal domain, and a back transformation to the domain of the skewed distribution, which has been shown to weaken spatial correlations (Embrechts et al., 2010; Papalexiou, 2018). The use of the kappa distribution in combination with the spatio-temporal model allows for the generation of extremes beyond the range of the observed values. However, it requires the fitting of many parameters, which make the model non-parsimonious (Koutsoyiannis, 2016). Depending on the application, the approach can therefore be used with a distribution with fewer parameters, a distribution fitted to a monthly instead of daily scale, or the empirical distribution.

The application of the approach is not limited to observed streamflow time series. It is applicable to other variables such as precipitation if combined with a suitable distribution as well as to modeled time series. The use of streamflow time series generated with a hydrological model extends the application of *PRSim.wave* to climate impact studies where a hydrological model is driven by meteorological time series generated with global and/or regional climate models.

6 Conclusions

Our results show that the continuous, wavelet-based stochastic simulation approach *PRSim.wave* reliably simulates discharge and extremes at multiple sites. Thanks to a spatio-temporal model based on phase randomization, temporal short- and long range dependencies, non-stationarities, and spatial dependencies are reproduced. In combination with the parametric kappa distribution, spatial extremes at multiple sites can be reliably simulated as well. The stochastic approach of *PRSim.wave* is very flexible and easy to use because of its availability in the R-package *PRSim*. Its versatility and advantageous properties make it generally useful for various water management applications where spatio-temporal patterns are of interest and, in particular, valuable for hazard assessments requiring information on spatial extremes.

Code and data availability. The wavelet-based stochastic simulation procedure for multiple sites *PRSim.wave* using the empirical, kappa, or any other distribution and some of the functions used to generate the validation plots are provided in the R-package *PRSim*. The stable version can be found in the CRAN repository https://cran.r-project.org/web/packages/PRSim/index.html, and the current development version is available at https://git.math.uzh.ch/reinhard.furrer/PRSim-devel. The observational discharge data was provided by the USGS and can be downloaded via: https://waterdata.usgs.gov/nwis.

Appendix A: Simulations randomizing phases for each catchment individually

19
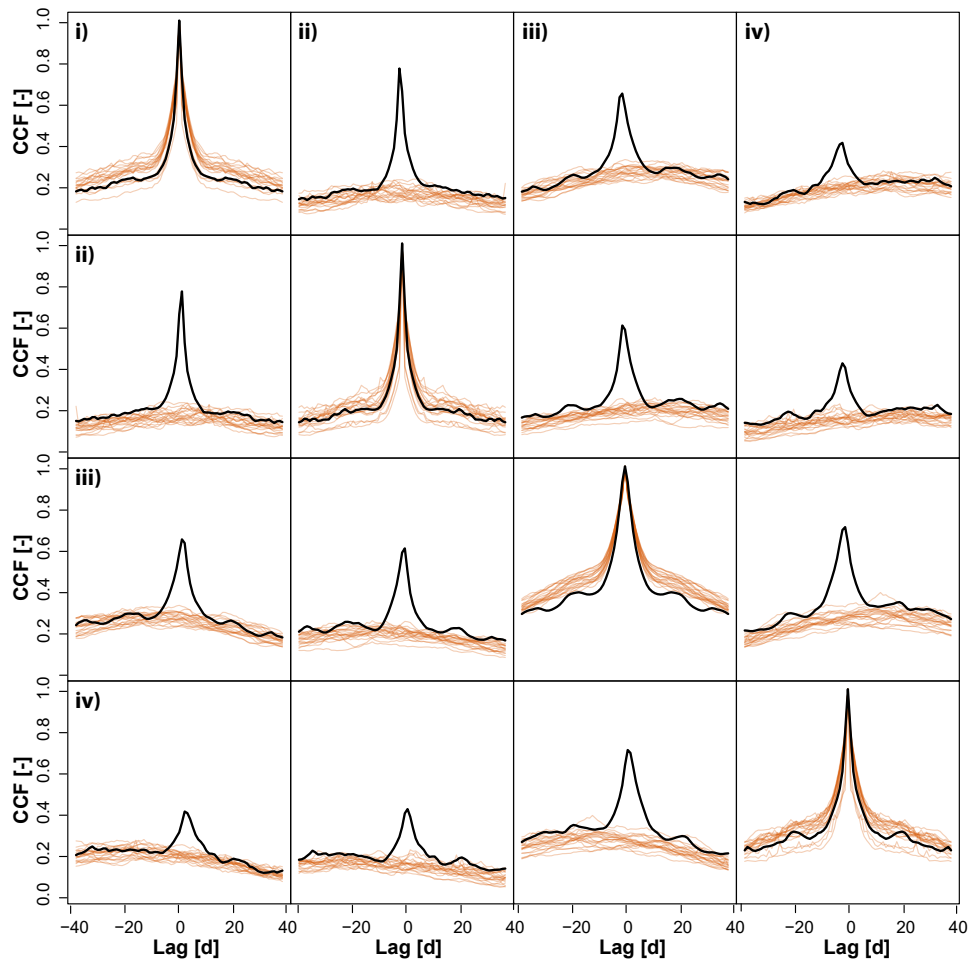
Figure A1. Comparison of observed (black) and simulated (orange) cross-correlation functions (ccfs) for the daily discharge values for pairs of stations in the set of

the four catchments (i–iv) in the Pacific Northwest. 20 simulations were generated for each site individually, neglecting spatial dependence.
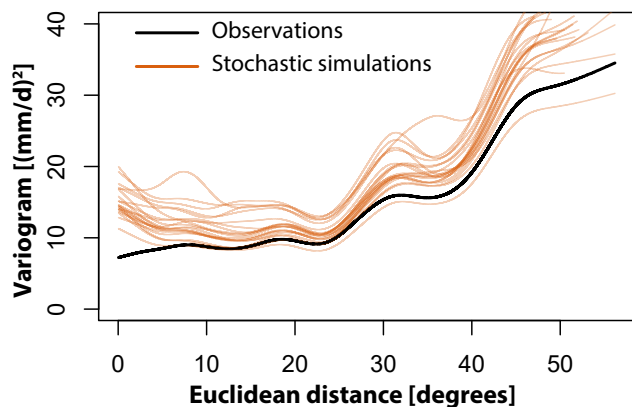
Figure A2. Comparison of observed (black) and simulated (orange) variograms for 20 simulation runs where the phases were randomized for each station individually.

305 *Competing interests.* The authors declare that they have no conflict of interest.

21

References

Addor, N., Newman, A. J., Mizukami, N., and Clark, M. P.: The CAMELS data set: Catchment attributes and meteorology for large-sample studies, Hydrology and Earth System Sciences, 21, 5293–5313, https://doi.org/10.5194/hess-21-5293-2017, 2017.

Blum, A. G., Archfield, S. A., and Vogel, R. M.: The probability distribution of daily streamflow in the conterminous United States, Hydrology and Earth System Sciences, 21, 3093–3103, https://doi.org/10.5194/hess-2016-460, 2017.

Bracken, C., Rajagopalan, B., Cheng, L., Kleiber, W., and Gangopadhyay, S.: Spatial Bayesian hierarchical modeling of precipitation extremes over a large domain, Water Resources Research, 52, 6643–6655, https://doi.org/10.1111/j.1752-1688.1969.tb04897.x, 2016.

Breakspear, M., Brammer, M., and Robinson, P. A.: Construction of multivariate surrogate sets from nonlinear data using the wavelet transform, Physica D: Nonlinear Phenomena, 182, 1–22, https://doi.org/10.1016/S0167-2789(03)00136-2, 2003.

Brunner, M. I. and Furrer, R.: PRSim: Stochastic Simulation of Streamflow Time Series using Phase Randomization, https://cran.r-project.org/web/packages/PRSim/index.html, 2019.

Brunner, M. I. and Tallaksen, L. M.: Proneness of European catchments to multiyear streamflow droughts, Water Resources Research, 55, 8881– 8894, https://doi.org/10.1029/2019WR025903, 2019.

Brunner, M. I., Bárdossy, A., and Furrer, R.: Technical note : Stochastic simulation of streamflow time series using phase randomization, Hydrology and Earth System Sciences, 23, 3175–3187, https://doi.org/10.5194/hess-23-3175-2019, 2019.

Caraway, N. M., McCreight, J. L., and Rajagopalan, B.: Multisite stochastic weather generation using cluster analysis and k-nearest neighbor time series resampling, Journal of Hydrology, 508, 197–213, https://doi.org/10.1016/j.jhydrol.2013.10.054, 2014.

Chavez, M. and Cazelles, B.: Detecting dynamic spatial correlation patterns with generalized wavelet coherence and non-stationary surrogate data, Scientific Reports, 9, 1–9, https://doi.org/10.1038/s41598-019-43571-2, 2019.

Chernobai, A., Rachev, S. T., and Fabozzi, F. J.: Composite goodness-of-fit tests for left-truncated loss samples, in: Handbook of financial econometrics and statistics, edited by Lee, C.-F. and Lee, J., chap. 20, pp. 575–596, Springer Science+Business Media, New York, 2015.

Coles, S.: An introduction to statistical modeling of extreme values, Springer, London, 2001.

Cooley, D., Naveau, P., and Poncet, P.: Variograms for spatial max-stable random fields, in: Lecture notes in statistics. Dependence in probability and statistics, pp. 373–390, Springer, New York, 2006.

Cressie, N. A. C.: Statistics for spatial data, Wiley series in probability and mathematical statistics, John Wiley & Sons, Inc, Iowa State University, 1993.

Daubechies, I.: Ten lectures on wavelets, in: CBMS-NSF regional conference series in applied mathematics, p. 357, Society for Industrial and Applied Mathematics, Philadelphia, 1992.

De Cicco, L. A., Lorenz, D., Hirsch, R. M., and Watkins, W.: dataRetrieval: R packages for discovering and retrieving water data available from U.S. federal hydrologic web services, https://doi.org/10.5066/P9X4L3GE, https://code.usgs.gov/water/dataRetrieval, 2018.

Diederen, D., Liu, Y., Gouldby, B., Diermanse, F., and Vorogushyn, S.: Stochastic generation of spatially coherent river discharge peaks for continental event-based flood risk assessment, Natural Hazards and Earth System Sciences, 19, 1041–1053, https://doi.org/10.5194/nhess-19-1041-2019, 2019.

Embrechts, P., McNeil, A. J., and Straumann, D.: Correlation and dependence in risk management: Properties and pitfalls, in: Risk Management, edited by Dempster, M. A. H., chap. 7, pp. 176–223, Cambridge University Press, Cambridge, https://doi.org/10.1017/cbo9780511615337.008, 2010.

Erkyihun, S. T., Rajagopalan, B., Zagona, E., Lall, U., and Nowak, K.: Wavelet-based time series bootstrap model for multidecadal streamflow simulation using climate indicators, Water Resources Research, 52, 4061–4077, https://doi.org/10.1002/2016WR018696, 2016.

Gilleland, E. and Katz, R. W.: ExtRemes 2.0: An extreme value analysis package in R, Journal of Statistical Software, 72, 1–39, https://doi.org/10.18637/jss.v072.i08, 2016.

350 Gordon, A.: Classification, Chapman & Hall/CRC, Boca Raton, 2nd edn., 1999.

Herman, J. D., Reed, P. M., Zeff, H. B., Characklis, G. W., and Lamontagne, J.: Synthetic drought scenario generation to support bottom-up water supply vulnerability assessments, Journal of Water Resources Planning and Management, 142, 1–13, https://doi.org/10.1061/(ASCE)WR.1943-5452.0000701, 2016.

Hosking, J.: The four-parameter kappa distribution, IBM Journal of Research and Development, 38, 251–258, 1994.

355 Hosking, J. R. M.: Modeling persistence in hydrological time series using fractional differencing, Water Resources Research, 20, 1898–1908, https://doi.org/10.1029/WR020i012p01898, 1984.

Keylock, C. J.: A wavelet-based method for surrogate data generation, Physica D: Nonlinear Phenomena, 225, 219–228, https://doi.org/10.1016/j.physd.2006.10.012, 2007.

Koutsoyiannis, D.: Generic and parsimonious stochastic modelling for hydrology and beyond, Hydrological Sciences Journal, 61, 225–244, 360 https://doi.org/10.1080/02626667.2015.1016950, 2016.

Kugiumtzis, D.: Test your surrogate data before you test for nonlinearity, Phydrical Review E, 60, 2808–2816, https://doi.org/10.1103/PhysRevE.60.2808, 1999.

Kwon, H. H., Lall, U., and Khalil, A. F.: Stochastic simulation model for nonstationary time series using an autoregressive wavelet decomposition: Applications to rainfall and temperature, Water Resources Research, 43, 1–15, https://doi.org/10.1029/2006WR005258, 2007.

365 Labat, D., Ronchail, J., and Guyot, J. L.: Recent advances in wavelet analyses: Part 2 - Amazon, Parana, Orinoco and Congo discharges time scale variability, Journal of Hydrology, 314, 289–311, https://doi.org/10.1016/j.jhydrol.2005.04.004, 2005.

Lafrenière, M. and Sharp, M.: Wavelet analysis of inter-annual variability in the runoff regimes of glacial and nival stream catchments, Bow Lake, Alberta, Hydrological Processes, 17, 1093–1118, https://doi.org/10.1002/hyp.1187, 2003.

Lall, U. and Sharma, A.: A nearest neighbor bootstrap for resampling hydrologic time series, Water Resources Research, 32, 679–693, 1996.

370 Lancaster, G., Iatsenko, D., Pidde, A., Ticcinelli, V., and Stefanovska, A.: Surrogate data for hypothesis testing of physical systems, Physics Reports, 748, 1–60, https://doi.org/10.1016/j.physrep.2018.06.001, 2018.

Maiwald, T., Mammen, E., Nandi, S., and Timmer, J.: Surrogate data - A qualitative and quantitative analysis, in: Mathematical methods in time series analysis and digital image processing, edited by Dahlhaus, R., Kurths, J., Maass, P., and Timmer, J., chap. 2, pp. 41–74, Springer, Berlin Heidelberg, 2008.

375 Mandelbrot, B. B.: Une classe de processus stochastiques homothetiques a soi: Application a la loi climatologique de H. E. Hurst, Comptes rendus de l'Académie des sciences, 260, 3274–3276, 1965.

Mandelbrot, B. B.: A fast fractional Gaussian noise generator, Water Resources Research, 7, 543–553, 1971.

Mejia, J. M., Rodriguez-Iturbe, I., and Dawdy, D. R.: Streamflow simulation: 2. The broken line process as a potential model for hydrologic simulation, Water Resources Research, 8, 931–941, https://doi.org/10.1029/WR008i004p00931, 1972.

380 Newman, A. J., Clark, M. P., Sampson, K., Wood, A., Hay, L. E., Bock, A., Viger, R. J., Blodgett, D., Brekke, L., Arnold, J. R., Hopson, T., and Duan, Q.: Development of a large-sample watershed-scale hydrometeorological data set for the contiguous USA: Data set characteristics and assessment of regional variability in hydrologic model performance, Hydrology and Earth System Sciences, 19, 209–223, https://doi.org/10.5194/hess-19-209-2015, 2015.

Nowak, K., Prairie, J., Rajagopalan, B., and Lall, U.: A nonparametric stochastic approach for multisite disaggregation of annual to daily streamflow, Water Resources Research, 46, W08 529, https://doi.org/10.1029/2009WR008530, 2010.

Nowak, K. C., Rajagopalan, B., and Zagona, E.: Wavelet Auto-Regressive Method (WARM) for multi-site streamflow simulation of data with non-stationary spectra, Journal of Hydrology, 410, 1–12, https://doi.org/10.1016/j.jhydrol.2011.08.051, 2011.

Papalexiou, S. M.: Unified theory for stochastic modelling of hydroclimatic processes: Preserving marginal distributions, correlation structures, and intermittency, Advances in Water Resources, 115, 234–252, https://doi.org/10.1016/j.advwatres.2018.02.013, 2018.

Prichard, D. and Theiler, J.: Generating surrogate data for time series with several simultaneously measured variables, Physical Review Letters, 73, 951–954, 1994.

Quinn, N., Bates, P. D., Neal, J., Smith, A., Wing, O., Sampson, C., Smith, J., and Heffernan, J.: The spatial dependence of flood hazard and risk in the United States, Water Resources Research, 55, 1890–1911, https://doi.org/10.1029/2018WR024205, 2019.

Radziejewski, M., Bardossy, A., and Kundzewicz, Z.: Detection of change in river flow using phase randomization, Hydrological Sciences Journal, 45, 547–558, https://doi.org/10.1080/02626660009492356, 2000.

Rajagopalan, B., Salas, J. D., and Lall, U.: Stochastic methods for modeling precipitation and streamflow, in: Advances in data-based approaches for hydrologic modeling and forecasting, edited by Sivakumar, B. and Berndtsson, R., chap. 2, pp. 17–52, World Scientific, https://doi.org/doi.org/10.1142/7783, 2010.

Ribatet, M.: SpatialExtremes: Modelling spatial extremes, https://doi.org/10.1214/11-STS376, https://cran.r-project.org/package=SpatialExtremes, 2019.

Salas, J. D. and Lee, T.: Nonparametric simulation of single-site seasonal streamflows, Journal of Hydrologic Engineering, 15, 284–296, https://doi.org/10.1061/(ASCE)HE.1943-5584.0000189, 2010.

Sang, Y. F.: A practical guide to discrete wavelet decomposition of hydrologic time series, Water Resources Management, 26, 3345–3365, https://doi.org/10.1007/s11269-012-0075-4, 2012.

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

Schmitz, A. and Schreiber, T.: Improved surrogate data for nonlinearity tests, Physical Review Letters, 77, 635–638, https://doi.org/10.1103/PhysRevLett.77.635, 1996.

Schreiber, T. and Schmitz, A.: Surrogate time series, Physica D: Nonlinear Phenomena, 142, 346–382, https://doi.org/10.1016/S0167-2789(00)00043-9, 2000.

Serinaldi, F. and Lombardo, F.: General simulation algorithm for autocorrelated binary processes, Physical Review E, 95, 1–9, https://doi.org/10.1103/PhysRevE.95.023312, 2017.

Serinaldi, F., Bardossy, A., and Kilsby, C. G.: Upper tail dependence in rainfall extremes: would we know it if we saw it?, Stochastic Environmental Research and Risk Assessment, 29, 1211–1233, 2015.

Sharma, A., Tarboton, D. G., and Lall, U.: Streamflow simulation: a nonparametric approach, Water Resources Research, 33, 291–308, 1997.

Shumway, R. H. and Stoffer, D. S.: Time series analysis and its applications. With R examples, Springer International Publishing AG, Cham, 4 edn., https://doi.org/10.1007/978-1-4419-7865-3, 2017.

Srinivas, V. V. and Srinivasan, K.: Hybrid matched-block bootstrap for stochastic simulation of multiseason streamflows, Journal of Hydrology, 329, 1–15, https://doi.org/10.1016/j.jhydrol.2006.01.023, 2006.

Srivastav, R. K. and Simonovic, S. P.: An analytical procedure for multi-site, multi-season streamflow generation using maximum entropy bootstrapping, Environmental Modelling and Software, 59, 59–75, https://doi.org/10.1016/j.envsoft.2014.05.005, 2014.

Stedinger, J. R. and Taylor, M. R.: Synthetic streamflow generation. 1. Model verification and validation, Water Resources Research, 18, 909–918, https://doi.org/0733-9429, 1982.

425 Theiler, J., Eubank, S., Longtin, A., Galdrikian, B., and Farmer, J. D.: Testing for nonlinearity in time series: the method of surrogate data, Physica D, 58, 77–94, https://doi.org/10.1016/0167-2789(92)90102-S, 1992.

Torrence, C. and Compo, G. P.: A practical guide to wavelet analysis, Bulletin of the American Meteorological Society, 79, 61–78, 1998.

USGS: USGS Water Data for the Nation, https://waterdata.usgs.gov/nwis, 2019.

Venema, V., Bachner, S., Rust, H., and Simmer, C.: Statistical characteristics of surrogate data based on geophysical measurements, Nonlinear 430   Processes in Geophysics, 13, 449–466, 2006.

Viglione, A.: homtest: Homogeneity tests for regional frequency analysis, https://cran.r-project.org/package=homtest, 2009.

Wang, W., Hu, S., and Li, Y.: Wavelet transform method for synthetic generation of daily streamflow, Water Resources Management, 25, 41–57, https://doi.org/10.1007/s11269-010-9686-9, 2010.