Technical note: Stochastic simulation of streamflow time series using phase randomization

Manuela I. Brunner[1], András Bárdossy[2], and Reinhard Furrer[3]

[1]Mountain Hydrology and Mass Movements, Swiss Federal Institute for Forest, Snow and Landscape Research WSL, Birmensdorf ZH, Switzerland
[2]Institute for Modelling Hydraulic and Environmental Systems, University of Stuttgart, Stuttgart, Germany
[3]Institute of Mathematics, University of Zurich, Zurich, Switzerland

Correspondence: Manuela Brunner (manuela.brunner@wsl.ch)

Abstract. Stochastically generated streamflow time series are widely used in water resource planning and management. Such series represent sets of plausible yet unobserved streamflow realizations which should reproduce the main characteristics of observed data. These characteristics include the distribution of daily streamflow values and their temporal correlation as expressed by short- and long-range dependence. Existing streamflow generation approaches have mainly focused on the time domain, even though simulation in the frequency domain provides good properties. These properties comprise the simulation of both short- and long-range dependence, as well as extension to multiple sites. Simulation in the frequency domain is based on the randomization of the phases of the Fourier transformation. We here combine phase randomization simulation with a flexible, four-parameter Kappa distribution, which allows for the extrapolation to yet unobserved low and high flows. The simulation approach consists of seven steps: 1) fitting the theoretical Kappa distribution, 2) normalization and deseasonalization, 3) Fourier transformation, 4) random phase generation, 5) inverse Fourier transformation, 6) back transformation, and 7) simulation. The simulation approach is applicable both to individual and multiple sites. It was applied to and validated on a set of four catchments in Switzerland. Our results show that the stochastic streamflow generator based on phase randomization produces realistic streamflow time series with respect to distributional properties and temporal correlation. However, cross-correlation among sites was in some cases found to be underestimated. The approach can be recommended as a flexible tool for various applications such as the dimensioning of reservoirs or the assessment of drought persistence.

Keywords: Fourier transformation, spectral analysis, generator, frequency domain, catchment, temporal dependence, correlation

Keypoints:

1. Stochastic simulation of streamflow time series for individual and multiple sites by combining phase randomization using the Kappa distribution

2. Simulated time series reproduce temporal correlation, seasonal distributions, and extremes of observed time series

3. Simulation procedure suitable for use in water resource planning and management

1 Introduction

Stochastically generated streamflow time series are used in various applications of water resource planning and management. These applications include water and reservoir management, the determination of the dimensions of hydraulic structures such as reservoirs, and the estimation of hydrological extremes such as droughts and floods. Stochastically generated time series
5 mimic the characteristics of observed data and represent sets of plausible but as yet unobserved streamflow sequences (Ilich, 2014; Borgomeo et al., 2015; Tsoukalas et al., 2018). They are essential for many uncertainty studies in hydrology because they can serve as input for deterministic water system models in which they allow for the propagation of natural variability and uncertainty (Tsoukalas et al., 2018).

    Stochastic models for the generation of synthetic streamflow time series need to fulfill certain requirements. They should
10 reproduce both the marginal distribution of observed streamflow time series as well as their temporal dependence structure (Sharma et al., 1997; Salas and Lee, 2010). Temporal dependence encompasses both short- and long-range dependence. While short-range dependence typically refers to the dependence of daily streamflow values measured within a few days, long-range dependence refers to dependencies across months or years. This temporal dependence has been found to depend on magnitude, in that low values have stronger dependence than high values (Lee and Salas, 2011). A proper representation of this long-range
15 dependence is of particular importance in studies where storage in reservoirs is of interest (Tsoukalas et al., 2018). If one is interested in extreme events, the model should allow for the generation of values that go beyond the magnitude of those observed (Herman et al., 2016). This requires the choice of a suitable theoretical marginal distribution. Streamflow typically exhibits a skewed distribution, which requires the use of a three or more-than-three-parameter distribution (Koutsoyiannis, 2000; Blum et al., 2017). Studies looking at individual hydrological events such as floods or droughts require a daily resolution.
20 Therefore, the stochastic model should allow for outputs at such a fine temporal resolution. Often, study regions encompass several sites whose streamflows are correlated. Consequently, the model ideally not only allows for the simulation of streamflow at individual sites, but also for the joint simulation of streamflow at multiple sites, taking into account their spatio-temporal dependence. From a practitioner's point of view, the model should not only reproduce the characteristics of the observed data, but it should also be simple (Sharma et al., 1997).

25   Many different approaches have been proposed for the stochastic simulation of streamflow time series, each able to fulfill some but usually not all of the desired properties listed above. One commonly used approach is the use of a synthetic weather generator in combination with a rainfall-runoff model (Pender et al., 2015). This approach is affected by uncertainties due to hydrological model selection and calibration, which can be avoided by using direct synthetic streamflow generation approaches (Herman et al., 2016). According to Stedinger and Taylor (1982), the development of such direct approaches consists of the
30 following steps: 1) selection of a model for seasonal marginal distributions, 2) selection of a model for spatial and temporal dependence, and 3) validation of the model. Different groups of direct approaches exist which are distinct in terms of their flexibility regarding marginal distributions and temporal dependence structures.

    A first group of models consists of parametric models such as autoregressive moving average (ARMA) models and their modifications. While these models are commonly used in stochastic hydrology, they only allow for modeling short-range

dependence because their autocorrelation decreases strongly with increasing lag time (Sharma et al., 1997). This means they guarantee neither the reproduction of observed persistence of annual flows nor the correlation structure among flows in different months (Stedinger and Taylor, 1982). This makes them unsuitable for applications where long-range dependence is important (Koutsoyiannis, 2000). A second group of parametric models is based on the temporal disaggregation of annual series and

5 enables the representation of long-range dependence (Stedinger and Taylor, 1982; Salas and Lee, 2010). These models include 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). Disaggregation models can be extended to multi-site applications (Grygier and Stedinger, 1988). However, this group of models has been shown to exhibit parameter estimation problems, and only allows for the representation of a narrow range of autocorrelation functions

10 (Koutsoyiannis, 2000). A third group of models is nonparametric in its approach and includes kernel density estimation (Lall and Sharma, 1996; Sharma et al., 1997) and various bootstrap approaches. The latter include simple bootstrap, which is only useful if data are uncorrelated, moving block-bootstrap, nearest-neighbor bootstrap (Salas and Lee, 2010; Herman et al., 2016), matched-block bootstrap (Srinivas and Srinivasan, 2006), and maximum-entropy bootstrap (Srivastav and Simonovic, 2014), which also take lagged correlations into account. These nonparametric techniques resample from the data with perturbations

15 and directly reproduce the characteristics of the original data (Sharma et al., 1997). Consequently, they allow for extrapolation to unobserved values (Salas and Lee, 2010). Moreover, the reproduction of long-range dependence is difficult and variance can be under- or overestimated (Salas and Lee, 2010). To allow for values that go beyond the observed distribution, Salas and Lee (2010) proposed a model employing k-nearest neighbor resampling with a gamma kernel perturbation. A further group of models consists of models that employ Markov chains and their variations. These models account for transition probabilities

20 between different hydrological states (Stagge and Moglen, 2013; Bracken et al., 2014; Pender et al., 2015) and can be combined with nonparametric approaches such as k-nearest neighbors (Prairie et al., 2008).

Several alternatives to these well-established simulation procedures have been proposed, which allow for a flexible choice of marginal distributions. These include models where the temporal dependence structure is modelled with copula functions, which are, however, difficult to apply for higher orders of autocorrelation (Lee and Salas, 2011). An example of a new sim-

25 ulation procedure is the SMARTA model by Tsoukalas et al. (2018), which employs Nataf's joint distribution model for the simulation of stochastic time series, representing both short- and long-range dependence. In addition, simulation schemes based on wavelet decomposition, which avoid assumptions about the temporal dependence structure, have been proposed by Kwon et al. (2007); Wang et al. (2010) and Erkyihun et al. (2016). Borgomeo et al. (2015) have shown how simulated annealing can be used to generate synthetic streamflow time sequences that represent possible climate-induced changes in user-specified

30 streamflow properties.

All these previously mentioned models are based on the time domain. An alternative to time-domain models is frequency-domain models (Shumway and Stoffer, 2017), which allow for the simulation of surrogate data with the same Fourier spectra as the raw data (Theiler et al., 1992). Such methods are based on the randomization of the phases of the Fourier transformation and have been commonly applied in hypothesis testing, when identifying nonlinearity in time series (Venema et al., 2006;

35 Maiwald et al., 2008), and in trend detection (Radziejewski et al., 2000). We hereafter refer to such methods, which are

also known as amplitude-adjusted Fourier transformations (Lancaster et al., 2018), as *phase randomization* simulations. In hydrology, phase randomization simulation has rarely been applied for purposes other than hypothesis testing (Fleming et al., 2002) even though it has desirable properties which make it suitable for a wider range of applications. Ineed, its implementation is relatively simple, it can simulate time series with both short- and long-range dependence, and it can be extended to multiple

5   sites. However, its application is limited to Gaussian data. We here propose the use of phase randomization simulation for the stochastic generation of streamflow time series at individual and multiple sites. To allow for non-Gaussian distributions, as commonly observed for daily streamflow values, we combine the data simulated by phase randomization with the Kappa distribution, which is a flexible four-parameter marginal distribution. The simulated streamflow series can be used for various water resource management applications.

10   We now turn to some theoretical background on Fourier transformation and phase randomization. For a more detailed introduction to the Fourier transformation, the reader is referred to textbooks by Morrison (1994) or Shumway and Stoffer (2017). We then discuss the use of phase randomization for the stochastic generation of streamflow time series. For illustration purposes, we apply and validate the approach on a set of four catchments in Switzerland. Finally, we discuss potential applications of the simulation approach.

2   Theoretical background

The basic idea behind all surrogate methods is to randomize the Fourier phases of the underlying (hydrological) process. The Fourier transformation converts a time-domain signal into a frequency-domain signal, which is complex-valued. This transformation may be depicted as a decomposition of the time series into sine and cosine waves of different amplitude, phase, and period (Fleming et al., 2002; Shumway and Stoffer, 2017). In the frequency domain, the power spectral density (power

20   spectrum) expresses the same information in cycles as the autocovariance function expresses in lags in the time domain. The periodogram, the empirical counterpart of the power spectrum, shows high values at those frequencies which correspond to strong periodic components (Shumway and Stoffer, 2017).

    The surrogate approach utilizes the property that realizations of linear Gaussian processes differ only in their Fourier phases and not their power spectrum. It preserves the autocorrelation structure of the raw series by conserving its power spectrum

25   through phase randomization. The procedure consists of three main steps (Radziejewski et al., 2000; Maiwald et al., 2008; Kim et al., 2010). In the first step, the streamflow series is converted from the time domain to the frequency domain by Fourier transformation (Morrison, 1994). In this frequency domain, the data are represented by the phase angle and the power spectrum, as represented by the periodogram. The phase angle θ of the power spectrum is uniformly distributed over the range of $-\pi$ to π. In the second step, the phases in the phase spectrum are randomized while the power spectrum is preserved. In the third step,

30   the inverse Fourier transformation is applied to transform the data from the frequency domain back to the temporal domain (Maiwald et al., 2008).

The Fourier transformation of a given time series $x = (x_1, \ldots, x_t, \ldots, x_n)$ of length n is

$$f(\omega) = \frac{1}{\sqrt{2\pi n}} \sum_{t=1}^{n} e^{-i\omega t} x_t, -\pi \leq \omega \leq \pi, \tag{1}$$

where $i = \sqrt{-1}$ is the imaginary unit. The original time series can be recovered by the back transformation

$$x_t = \sqrt{\frac{2\pi}{n}} \sum_{j=1}^{n} e^{j\omega_j t} f(\omega_j), t = 1, 2, \ldots, n, \tag{2}$$

5   if the transformation is calculated for discrete frequencies $\omega_j = 2\pi/n, j = 1, 2, \ldots, n$. The Fourier transformation surrogate method constructs a new time series y_t with the same periodogram as the observations. Apart from this, the new series are statistically independent of x_t. This can be achieved by fixing the Fourier amplitudes $|f(\omega_j)|$ and replacing the Fourier phases $\phi(\omega_j) = \arg(f(\omega_j))$ by uniformly distributed random numbers $\phi_{\text{rand}}(\omega_j) \in [-\pi, \pi]$. A new realization is given by

$$y_t = \sqrt{\frac{2\pi}{n}} \sum_{j=1}^{n} e^{j\omega_j t} |f(\omega_j)| e^{i\phi} \text{rand}^{\omega_i}. \tag{3}$$

10   The surrogate data consist of the same values as the original data in another temporal order but with the same time dependence structure as the original data (Schreiber and Schmitz, 2000). The approach can be extended to multiple sites by multiplying the phases of each site by the same set of random phases. This is possible because the cross-spectrum, which describes the cross-correlation of the data in the frequency domain, reflects relative phases only (Prichard and Theiler, 1994; Schreiber and Schmitz, 2000).

3   Methods

3.1   Stochastic streamflow simulation

We use phase randomization here to simulate stochastic streamflow time series to be used in various water resource management studies. The stochastic series generated using phase randomization are combined with a theoretical distribution to allow extrapolation to unobserved values, which still realistically represent daily streamflow values. The observed streamflow time series require pre-treatment before phase randomization can be applied. First, they need to be normalized because phase randomization assumes Gaussianity (Maiwald et al., 2008). Second, they need to be deseasonalized in order to remove monthly/daily fluctuations (Pender et al., 2015). The stochastic simulation procedure consists of the following seven steps.

1. **Fitting of theoretical Kappa distribution:** The four-parameter Kappa distribution (Hosking, 1994) is fit to the daily values of the observed input time series. This distribution will be used for the back transformation in Step 7, and permits extreme values going beyond the empirical distribution to be obtained. It has four parameters and its cumulative distribution function is expressed as

$$F(x) = \left\{ 1 - h[1 - k(x - \xi)\alpha]^{1/k} \right\}^{1/h}, \tag{4}$$

where ξ is the location parameter, α is the scale parameter which must be positive, and k and h are the shape parameters.

The Kappa distribution was found suitable for fitting observed streamflow data in U.S. catchments (Blum et al., 2017). A suitable fit was also found for our data as confirmed by a Kolmogorov-Smirnov test which did not reject the null hypothesis at $\alpha = 0.05$ for most catchments. 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 used 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 were removed to maintain constant sample sizes across years as in Blum et al. (2017).

2. **Normalization and deseasonalization:** The input time series are normalized using the normal transform, i.e., values corresponding to a certain rank are replaced with respective values from a standard normal distribution. The normal transform is applied to each day of the year separately, which results in the deseasonalization of the data.

3. **Fourier transformation:** The normalized and deseasonalized data are transferred to the frequency domain using the Fourier transformation (Equation 1). The Fourier phases (i.e., the arguments of the Fourier transformation) are computed.

4. **Random phase generation:** Random phase series are generated by sampling from the uniform phase distribution. The observed spectrum (i.e., the modulus of the Fourier transformation) is preserved.

5. **Inverse Fourier transformation:** The random phases are combined with the observed spectrum and inverse Fourier transformation is applied (Equation 2) to transform the data back to the time domain.

6. **Back transformation:** The data are back transformed from the normal to the Kappa domain using the fitted daily Kappa distributions (Equation 4), which achieves reseasonalization. This is done by generating a sample of length n (length of observed time series) and reassigning values according to the ranks in the simulated series. Negative simulated values are replaced by 0, which corresponds to the lower boundary of the Kappa distribution. Using the empirical distribution instead of the Kappa distribution would prevent us from obtaining values that go beyond the range of observed data (Srinivas and Srinivasan, 2006).

7. **Simulation:** Steps 4-7 are repeated m times to generate m time series of the same length as the observed time series.

The method is extended to the simulation of stochastic streamflow time series at multiple sites. To model the cross-correlation between sites, the phase randomization performed in Step 4 of the procedure is performed in the same way for all the stations in the dataset (Prichard and Theiler, 1994). In contrast, the parameters of the monthly Kappa distributions and the power spectrum are calculated for each individual site separately.

3.2 Model validation

The simulation was validated on the observed streamflow time series of a set of four catchments in Switzerland (Figure 1), namely, Plessur-Chur, Birse-Moutier, Thur-Jonschwil, and Cassarate-Pregassona. The catchments are characterized by diverse

catchment characteristics and flow regimes (Table 1). Their catchment areas range between 74 and 493 km^2 and their mean elevations between 930 and 1,850 m a.s.l. Plessur represents a catchment with a melt-dominated flow regime with high flows in summer but low flows in winter. In contrast, the flow regimes of Birse and Thur are dominated by precipitation with high flows in winter and low flows in summer. The regime of Cassarate shows two peaks, one in spring due to melt processes, and

5   one in fall due to precipitation.

The model outlined in the previous section was fit to the observed time series over 50 years (1960-2009) for each individual catchment. The application of this approach is only recommended for records longer than 30 years to reduce uncertainty in the estimation of the parameters of the Kappa distribution. The model was then run, on the one hand, for each individual catchment and, on the other hand, for the four sites jointly. In both cases, 100 sets of stochastic streamflow time series of the same length

10   as the observed series were generated as in Salas and Lee (2010) and Pender et al. (2015).
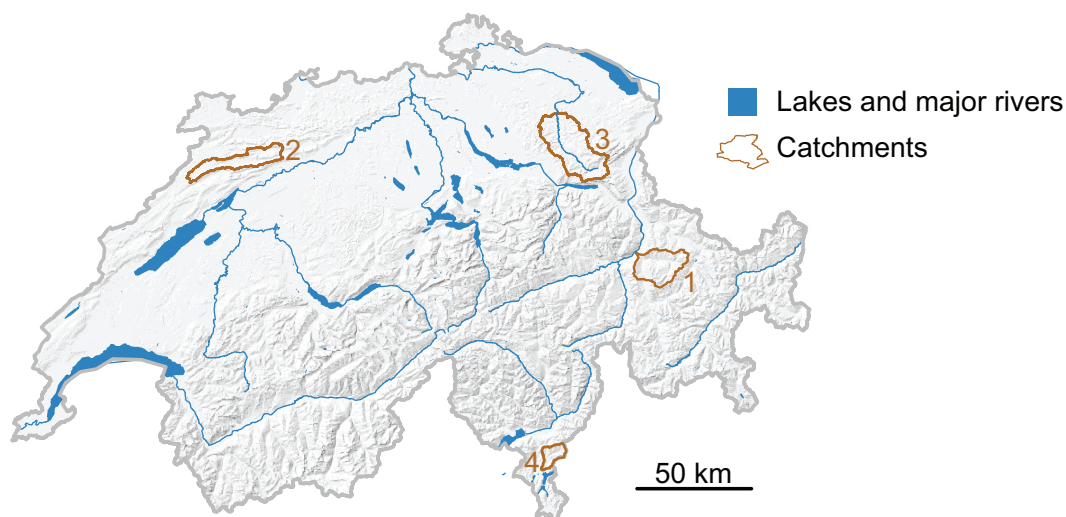
Figure 1. Map showing the four Swiss catchments.

Table 1. List of catchments and catchment summary including ID, river name, gauging station, catchment area, station elevation, mean elevation, and flow regime.

ID	River	Gauging station	Area (km²)	Station elevation (m a.s.l.)	Mean elevation (m a.s.l.)	Flow regime
1	Plessur	Chur	263	573	1,850	Melt-dominated
2	Birse	Moutier	183	519	930	Rainfall-dominated
3	Thur	Jonschwil	493	534	1,030	Rainfall-dominated
4	Cassarate	Pregassona	74	291	990	Mixed

Both the temporal correlation structure and seasonal streamflow statistics were used to compare observed and simulated streamflow time series in order to assess the validity of the stochastic streamflow generation model. As in Kim et al. (2010),

we used the autocorrelation function on daily values to represent the short-range temporal correlation. Further, we also used the partial autocorrelation function (Stedinger and Taylor, 1982). In addition to short-range dependence, long-range dependence was assessed by looking at the autocorrelation function of annual discharge sums. The seasonal statistics were validated with respect to the seasonal distributions (winter: Dec–Feb, spring: Mar–May, summer: Jun–Aug, and fall: Sep–Nov) and the monthly means, maxima, minima, and standard deviations. In addition to general distribution characteristics, the approach was validated for low and high flows because these characteristics are often of interest in hydrological simulation studies (Borgomeo et al., 2015). High and low flows were defined as above or below threshold values, respectively. For high flows, the 95th percentile was used as a threshold, while the 5th percentile was used for low flows.

4 Results

4.1 Simulation at individual sites

The stochastic streamflow generator was found to produce realistic annual hydrograph realizations as illustrated in Figure 2 for the Plessur catchment (Figure 1). This is confirmed visually by observing the temporal correlation structure, as well as the seasonal statistics (see Figure 3 for Thur and Plessur).
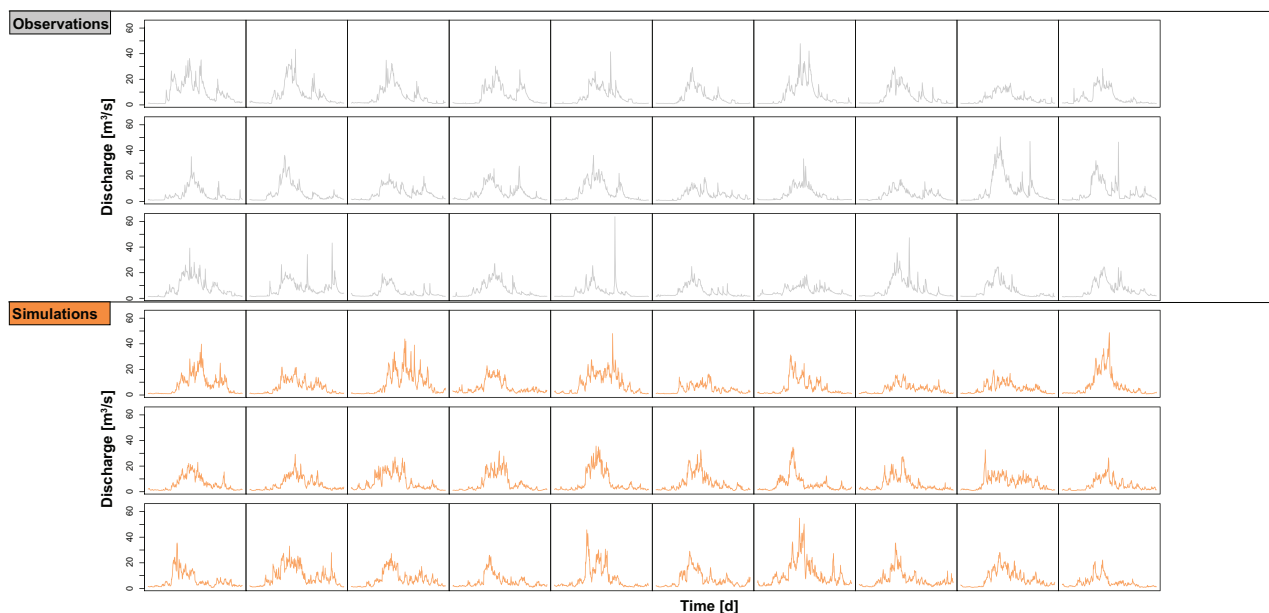
Figure 2. Observed (grey) and stochastically generated (orange) annual hydrographs at daily resolution over 30 years for the Plessur catchment.
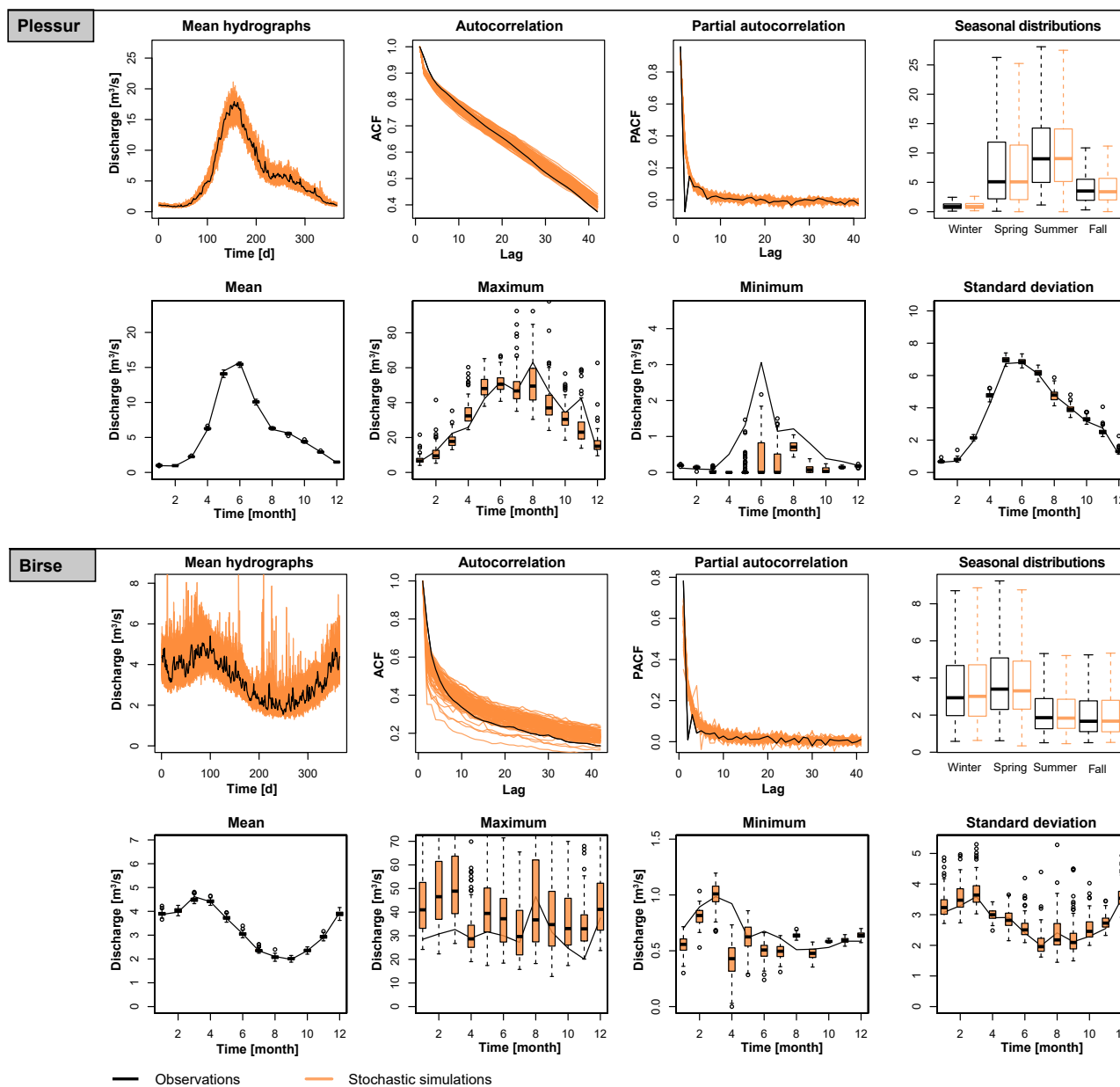
Figure 3. Comparison of observed and stochastically generated time series for the melt-dominated Plessur catchment (upper panels) and the rainfall-dominated Birse catchment (lower panels) for the following characteristics: Mean hydrograph over 50 years, autocorrelation function, partial autocorrelation function, seasonal distributions, monthly means, monthly maxima, monthly minima, and monthly standard deviations. Black lines represent observations while orange lines represent simulations.

The stochastic generator produces time series with mean regimes similar to the observed mean regime, and reproduces both the autocorrelation (ACF) and partial autocorrelation functions (PACF). Seasonal distributions match well thanks to the good fit

of the Kappa distribution to the data. Monthly means and standard deviations match particularly well, while monthly maxima and minima show some deviations from the observed maxima and minima, as was intended by using a theoretical instead of an empirical distribution. The suitability of the Kappa distribution to produce realistic high and low flows is confirmed in Figure 4. The distribution produces low flows similar to observed low flows but with different outliers. In two catchments (Thur and

5   Cassarate) however, observed low flows were rather overestimated. High flow distributions match well in all catchments, and values exceeding observed values are generated. The four-parameter Kappa distribution (Houghton, 1978; Griffiths, 1989) was found to be more suitable for representing daily streamflow values compared to distributions with even more parameters, which are rather prone to over-fitting.
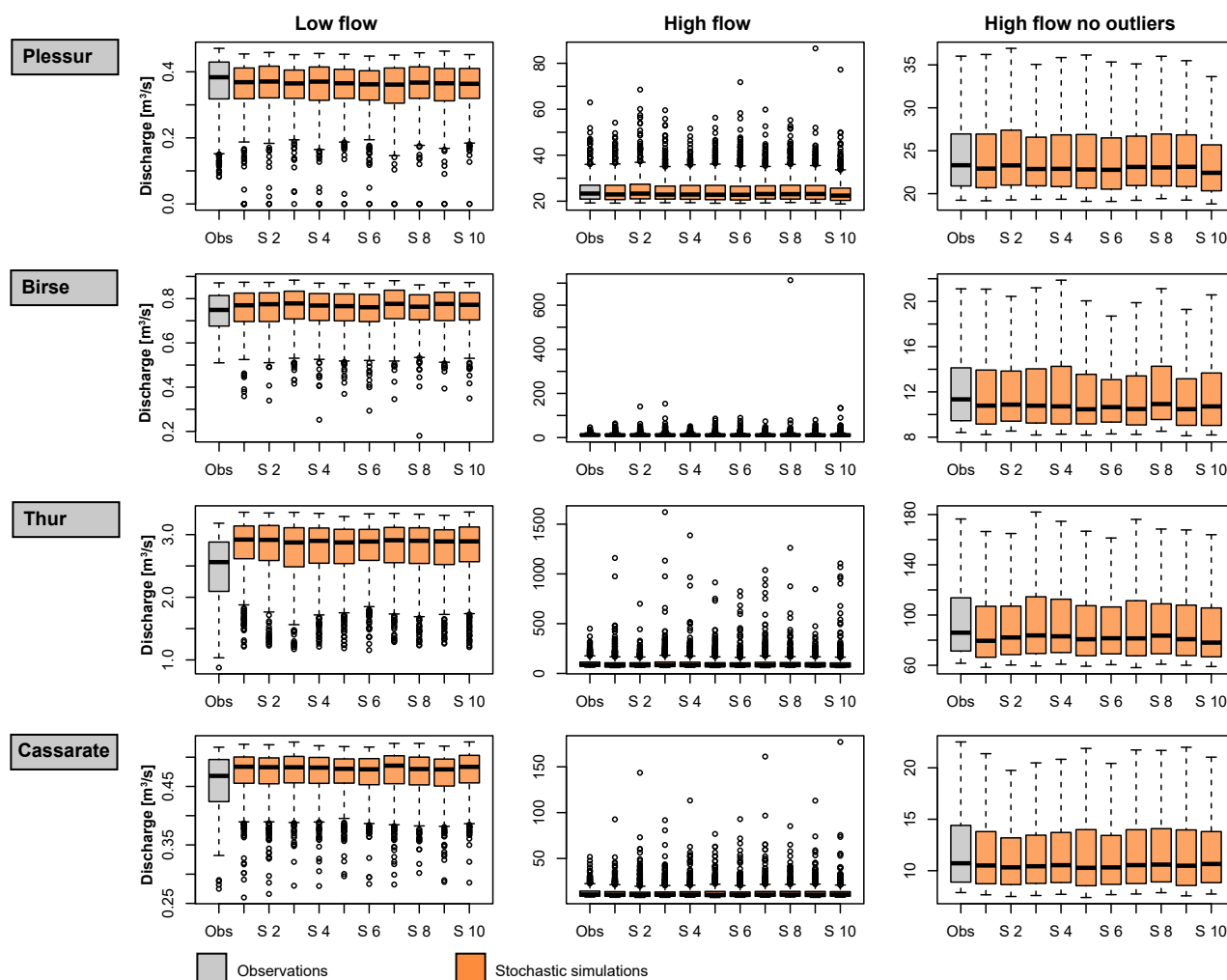
Figure 4. Low and high flows for observed (grey) and simulated (orange) time series for the four catchments Plessur, Birse, Thur, and Cassarate. The results are given for ten simulation runs (S1-S10), and high flows are plotted with (middle panel) and without outliers (right panel). Whiskers extend to the lowest/highest data point which is still within 1.5 times the interquartile range.

The stochastic streamflow generator is not only able to reproduce the streamflow distribution and the short-range dependence in the data, but also the long-range dependence over several years (Figure 5). Both the rapid decrease in the ACF at short lags (up to five years) and the cyclical behavior at lags longer than five years are reproduced as well.
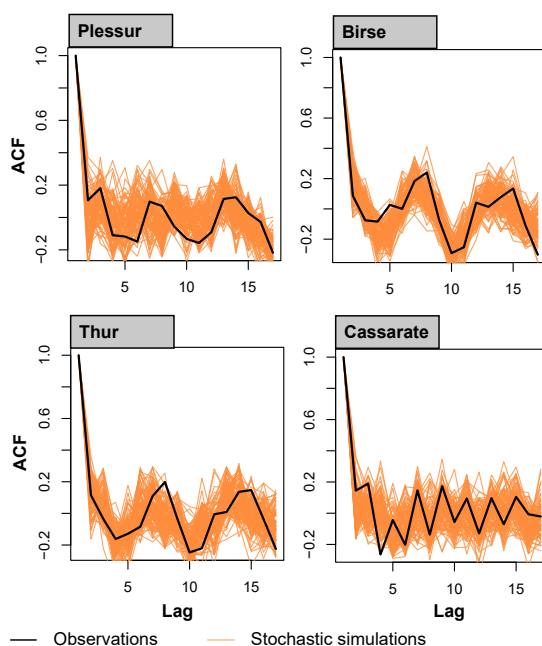
Figure 5. Autocorrelation (ACF) of annual streamflow sums of the observed and simulated streamflow time series for the catchments Plessur, Birse, Thur, and Cassarate.

The good performance of the stochastic streamflow generator with respect to streamflow distribution and temporal correlation - both short and long range - is not limited to these four example catchments but generalizes to other data sets used as input.

4.2 Simulation at multiple sites

The stochastic streamflow generator can be extended from the simulation at individual sites to the joint simulation at multiple sites. In addition to reproducing distribution and temporal correlation at individual sites, it should then be able to reproduce the cross-correlation among sites, which describes the similarity of time series at two sites. Figure 6 shows the cross-correlation function (CCF) for pairs of stations among the example catchments for the observed time series and the 100 simulation runs. Cross-correlation is already generally low for observations because the selected sample catchments are characterized by diverse discharge regimes and seasonality. The shape of the cross-correlation is reproduced for all pairs of stations. However, the magnitude of cross-correlation is underestimated for certain pairs of stations in the simulated time series compared to the observed series independently of the simulation run considered. For the catchment pair Birse-Thur, whose discharge behavior is rainfall-dominated, the simulated cross-correlation is much lower than the observed one. In the observations, spatially

consistent rainfall events lead to a joint rise in discharge at both stations. This behavior is not captured by the stochastic discharge generator. The underestimation of cross-correlation is also visible when looking at the cross-correlation of below- or above-threshold events (not shown).
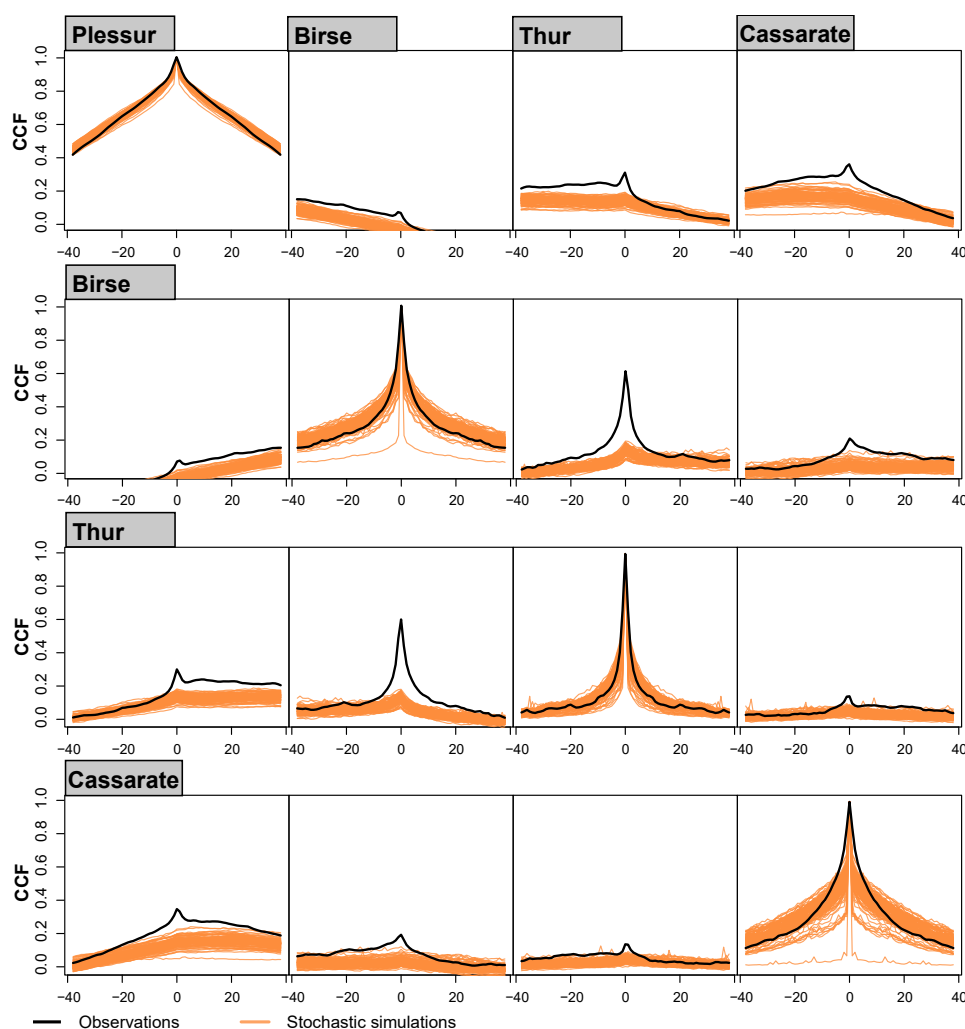
Figure 6. Cross-correlation function (CCF) of observed (black line) and simulated (orange lines) daily streamflow for pairs of stations at Plessur, Birse, Thur, and Cassarate.

5   Discussion and Conclusions

5   The stochastic streamflow generator based on phase randomization has been shown to produce realistic streamflow time series with respect to both distributional properties and temporal correlation. Compared to models commonly used for the stochastic generation of streamflow time series, such as autoregressive moving average models, the simulation approach presented here

not only reproduces short-range but also long-range dependence. The generator can, on the one hand, be used to simulate streamflow at individual sites, and, on the other, to simulate jointly at multiple sites, which is not necessarily the case for other existing models. Its application to the example catchments, however, has resulted in somewhat underestimated cross-correlations between stations. This underestimation can be explained by the fact that phase randomization preserves the cross-

5 correlation in the normal domain but not necessarily in the domain of the original distribution. Phase randomization was here combined with the flexible four-parameter Kappa distribution, which was found to effectively represent daily streamflow values. The distribution of daily flows was found to be modelled well in all seasons. The use of the Kappa distribution allows us to generate values that go beyond the range of observed values, which would not be the case if the empirical distribution was used. This ability of the generator to extrapolate extremes makes it suitable for applications where extreme events such

10 as floods and droughts are of interest. Nevertheless, it is important to note that the cross-correlation of high and low flows at multiple stations is underestimated even if the generator is jointly applied to multiple sites. This cannot be overcome even if the simulation run which best reproduces these cross-correlations is extracted from a large set of simulations. However, Stedinger and Taylor (1982) showed that estimators of the autocorrelation and cross-correlation of flows which do not match the historical sample estimates often provide more accurate estimates of the true but unknown correlations. Still, one possibility to improve

15 the reproduction of cross-correlations would be the use of phase annealing (Hörning and Bárdossy, 2018). Phase annealing modifies the Fourier phases in an iterative way in order to optimize certain statistics, such as the cross-correlation function, and makes it possible to take covariates into consideration for the generation of time series. However, using phase annealing increases the computational effort.

The streamflow generator was here used on observed streamflow time series. The input time series, however, do not nec-

20 essarily need to consist of observed values. One could also use the generator on streamflow simulated with a hydrological model. This extends its application to climate impact studies where a hydrological model is driven by meteorological time series generated with global and/or regional climate models. In addition, phase randomization simulation can potentially be applied to streamflow time series of anthropogenically-influenced catchments, and to other hydro-meteorological variables such as precipitation, evapotranspiration, or snowmelt. Spatio-temporal modelling of precipitation fields, for example, may be

25 performed using a technique based on phase randomization. However, it must be noted that due to the large number of zero observations (specifically with fine temporal resolution) the normal score transformation can become non-unique. In this case, additional efforts are needed to preserve the spatial structure of precipitation.

The stochastic streamflow generator presented here represents a flexible tool for streamflow simulation at individual or multiple sites. It can be used for various applications such as the design of hydropower reservoirs, the assessment of flood risk,

30 or the assessment of drought persistence and the estimation of the risk of multi-year droughts.

Code availability. The stochastic simulation procedure for a single site 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.

Data availability. The observational discharge data was provided by the Federal Office for the Environment (FOEN), and can be ordered from http://www.bafu.admin.ch/wasser/13462/13494/15076/index.

Author contributions. AB and MB jointly developed the concept and methodology of the study. MB and RF set up the simulation approach. MB did the data analysis, produced the figures, and wrote the first draft of the manuscript. The manuscript was revised by RF and AB and
5 edited by MB.

Competing interests. The authors do not have any competing interests

Acknowledgements. We acknowledge financial support provided by the FOEN (granted to MB via contract 15.0003.PJ/Q292-5096), the German Research Foundation (Deutsche Forschungsgemeinschaft, DFG) (granted to AB via project number Ba-1150/13-1 within the Space-Time Dynamics of Extreme Floods (SPATE) project), and the Swiss National Science Foundation (SNF) (granted to RF via project 175529).

References

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.

Borgomeo, E., Farmer, C. L., and Hall, J. W.: Numerical rivers: A synthetic streamflow generator for water resources vulnerability assessments, Water Resources Research, 51, 5382–5405, https://doi.org/10.1002/2014WR016259, 2015.

Bracken, C., Rajagopalan, B., and Zagona, E.: A hidden Markov model combined with climate indices for multidecadal streamflow simulation, Water Resources Research, 50, 7836–7846, https://doi.org/10.1002/2013WR014910.Received, 2014.

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.

Fleming, S. W., Marsh Lavenue, A., Aly, A. H., and Adams, A.: Practical applications of spectral analysis of hydrologic time series, Hydrological Processes, 16, 565–574, https://doi.org/10.1002/hyp.523, 2002.

Griffiths, G. A.: A theoretically based Wakeby distribution for annual flood series, Hydrological Sciences Journal, 34, 231–248, https://doi.org/10.1080/02626668909491332, 1989.

Grygier, J. A. N. C. and Stedinger, J. R.: Condensed disaggregation procedures and conservation corrections for stochastic hydrology, Water Resources Management, 24, 1574–1584, 1988.

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.

Hörning, S. and Bárdossy, A.: Phase annealing for the conditional simulation of spatial random fields, Computers and Geosciences, 112, 101–111, https://doi.org/10.1016/j.cageo.2017.12.008, 2018.

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

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.

Houghton, J. C.: Birth of a parent: The Wakeby Distribution for modeling flood flows, Water Resources Research, 14, 1105–1109, https://doi.org/10.1029/WR014i006p01105, 1978.

Ilich, N.: An effective three-step algorithm for multi-site generation of stochastic weekly hydrological time series, Hydrological Sciences Journal, 59, 85–98, https://doi.org/10.1080/02626667.2013.822643, 2014.

Kim, J. S., Huh, Y., and Suh, M. W.: A method to generate autocorrelated stochastic signals based on the random phase spectrum, Journal of the Textile Institute, 101, 471–479, https://doi.org/10.1080/14685240802528443, 2010.

Koutsoyiannis, D.: A generalized mathematical framework for stochastic simulation and forecast of hydrologic time series, Water Resources Research, 36, 1519–1533, 2000.

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.

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

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, https://doi.org/10.1016/j.physrep.2018.06.001, 2018.

Lee, T. and Salas, J. D.: Copula-based stochastic simulation of hydrological data applied to Nile River flows, Hydrology Research, 42, 318–330, https://doi.org/10.2166/nh.2011.085, 2011.

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.

Mandelbrot, 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.

Morrison, N.: Introduction to Fourier analysis, John Wiley & Sons, Inc, New York, 3 edn., 1994.

Pender, D., Patidar, S., Pender, G., and Haynes, H.: Stochastic simulation of daily streamflow sequences using a hidden Markov model, Hydrology Research, 47, 75–88, https://doi.org/10.2166/nh.2015.114, 2015.

Prairie, J., Nowak, K., Rajagopalan, B., Lall, U., and Fulp, T.: A stochastic nonparametric approach for streamflow generation combining observational and paleoreconstructed data, Water Resources Research, 44, 1–11, https://doi.org/10.1029/2007WR006684, 2008.

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

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.

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.

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.

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.

Stagge, J. H. and Moglen, G. E.: A nonparametric stochastic method for generating daily climate-adjusted streamflows, Water Resources Research, 49, 6179–6193, https://doi.org/10.1002/wrcr.20448, 2013.

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.

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.

Tsoukalas, I., Makropoulos, C., and Koutsoyiannis, D.: Simulation of stochastic processes exhibiting any-range dependence and arbitrary marginal distributions, Water Resources Research, 54, 9484–9513, https://doi.org/10.1029/2017WR022462, 2018.

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

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.