Bayesian spectral likelihood for hydrological parameter inference

Bettina Schaefli[1] and Dmitri Kavetski[2]

D. Kavetski, School of Civil, Environmental and Mining Engineering, University of Adelaide, Adelaide, SA, Australia

B. Schaefli, Institute of Earth Surface Dynamics, Faculty of Geosciences and Environment, University of Lausanne, 1015 Lausanne, Switzerland (bettina.schaefli@unil.ch)

[1]Institute of Earth Surface Dynamics,
Faculty of Geosciences and Environment,
University of Lausanne, Switzerland

[2]School of Civil, Environmental and
Mining Engineering, University of Adelaide,
Adelaide, SA, Australia

Abstract. This paper proposes a spectral-domain likelihood function for the Bayesian estimation of hydrological model parameters from a time series of model residuals. The spectral-domain error model is based on the Power-Density-Spectrum (PDS) of the stochastic process assumed to describe residual errors. The Bayesian Spectral Likelihood (BSL) is mathematically equivalent to the corresponding Bayesian Time-domain Likelihood (BTL) and yields the same inference when all residual error assumptions are satisfied (and all residual error parameters are inferred). However, the BSL likelihood function does not depend on the residual error distribution in the original time-domain, which offers a theoretical advantage in terms of robustness for hydrological parameter inference. The theoretical properties of BSL are demonstrated and compared to BTL and a previously proposed spectral likelihood by Montanari and Toth (2007), using a set of synthetic case studies and a real case study based on the Leaf River catchment in the US. The empirical analyses confirm the theoretical properties of BSL when applied to heteroscedastic and autocorrelated error models (where heteroscedasticity is represented using the log-transformation and autocorrelation is represented using an AR(1) process). Unlike MTL, the use of BSL did not introduce additional parametric uncertainty compared to BTL. Future work will explore the application of BSL to challenging modeling scenarios in arid catchments and "indirect" calibration with non-concomitant input/output time series.

1. Introduction

Bayesian and other likelihood-based inference methods have a strong tradition in hydrological modeling, with the overall goal of providing reliable hydrological predictions and uncertainty estimates [e.g., Kuczera, 1983; Beven and Binley, 1992; Kuczera and Parent, 1998; Bates and Campbell, 2001, and many others]. The key ingredient of likelihood-based inference is the likelihood function, which should provide a probabilistic description of the uncertainty in the model predictions [e.g., Box and Tiao, 1992]. In the simplest case, the likelihood function aims to describe the statistical properties of the model residual errors, i.e., the time series of differences between observed responses (e.g., streamflow) and corresponding model predictions [e.g., Box and Tiao, 1992; Kuczera and Parent, 1998].

A major concern is that, in hydrology, probabilistic inference methods have often been used with ostensibly wrong assumptions [e.g., as noted by Beven and Binley, 1992; Kavetski et al., 2006; Honti et al., 2013, and others]. For example, it is still common for hydrological calibration applications to assume independent and identically distributed Gaussian model residuals, and relatively few studies rigorously assess how well these assumptions are actually satisfied [e.g., Engeland et al., 2005; Schaefli et al., 2007]. Recent work is addressing these shortcomings, contributing more statistically reliable error models and likelihood functions [e.g., Kuczera, 1983; Kavetski et al., 2006; Schaefli et al., 2007; Thyer et al., 2009; Schoups and Vrugt, 2010; Smith et al., 2010; Pianosi and Raso, 2012; Evin et al., 2014; McInerney et al., 2017, and many others].

In this paper, we explore new perspectives for hydrological parameter inference by introducing a Bayesian Spectral Likelihood (BSL), based on a statistical description of

the model residuals in the spectral-domain, i.e., in the Fourier-transformed-domain rather than in the time-domain. The term "spectral-domain" rather than "frequency-domain" is used to emphasize that the proposed likelihood is based on the power-density spectrum.

This work is not the first attempt to use spectral methods in hydrological modeling. For example, Montanari and Toth [2007] applied the Whittle likelihood [Whittle, 1953] to calibrate hydrological model parameters. Other studies used spectral likelihoods or spectral signatures in more informal settings. Quets et al. [2010] used the sum of squared differences between the Fourier amplitudes, or between the Fourier amplitudes and phases of observed and simulated streamflow time series, to calibrate the SWAT model. A similar approach was followed by Pauwels and De Lannoy [2011] and De Vleeschouwer and Pauwels [2013]. Moussu et al. [2011] used the root-mean-squared difference between the estimated autocorrelation functions of the observed and simulated streamflow series to calibrate two conceptual rainfall-runoff models of a karst system. Winsemius et al. [2009] and Hartmann et al. [2013] used the streamflow autocorrelation function as a signature to assess the model performance in a multi-criteria model identification setting. Schaefli and Zehe [2009] proposed to assess hydrologic model performance in terms of the Kolmogorov-Smirnoff distance between the estimated wavelet power spectra of observed and simulated streamflow series. Several studies proposed to use spectral calibration for non-concomitant (or indirect) calibration, where input and output observations are not available over the same time period [Montanari and Toth, 2007; De Vleeschouwer and Pauwels, 2013].

The main motivation for this paper is to present the key theoretical aspects of spectral parameter inference, especially in light of recent interest in spectral model calibration and performance assessment. Our paper addresses the current research gap that the major-

ity of hydrological calibration approaches based on spectral techniques do not explicitly articulate the probabilistic assumptions underlying their choice of objective (likelihood) function; this limitation complicates the derivation of probability limits on the estimated model parameters and predictions. The properties of BSL are investigated using a series of synthetic and real data case studies, and are compared to the properties of the corresponding (standard) time-domain likelihood and of the spectral-domain likelihood used previously by Montanari and Toth [2007]. The method of Montanari and Toth [2007] is of particular relevance to this work, because to our knowledge it is the only spectral calibration method in the hydrological literature that uses a spectral likelihood function with an explicit probabilistic interpretation.

The BSL approach introduced in this work is obtained by expressing the probability density function (pdf) of residual errors of a hydrological model in terms of their Fourier power-density spectrum [e.g., Jenkins and Watts, 1968]. This spectral-domain probabilistic characterization is presented in considerable detail because, despite spectral analysis being widely used in time series analysis, it remains relatively rare in hydrology, and existing literature generally does not describe the pdf of the entire power-density spectrum.

The remainder of the paper is structured as follows. Section 2 presents all required definitions and the derivation of BSL for common stochastic error models. This section also briefly outlines the relationship of BSL to the likelihood presented by Montanari and Toth [2007]. Section 3 details the case studies and the analysis methodology. Section 4 presents and discusses the case study results. Section 6 summarizes the key conclusions of the paper, outlines some important open questions and suggests future research directions.

2. Theoretical development

2.1. Bayesian time-domain likelihood (BTL)

Consider a hydrological model $\boldsymbol{H}$

$$\widehat{\boldsymbol{Y}} = \boldsymbol{H}(\boldsymbol{\theta}, \boldsymbol{X}) \tag{1}$$

where $\boldsymbol{X} = (\boldsymbol{X}_t)_{t=1,..,N}$ are the system inputs (e.g., rainfall and potential evapotranspiration) at time steps $t = 1, .., N$, $\widehat{\boldsymbol{Y}} = (\widehat{\boldsymbol{Y}}_t)_{t=1,..,N}$ is the system output predicted by the model (e.g. streamflow), and $\boldsymbol{\theta}$ is a vector of model parameters. On overview of all used mathematical notations is given in Table 1.

In practice, the true system input $\boldsymbol{X}$ is unknown, and we only have observed inputs $\widetilde{\boldsymbol{X}}$, which are affected by sampling and measurements errors. In this paper, we represent total predictive uncertainty using residual errors, which are assumed to aggregate the effects of all sources of error including data uncertainty and model structural errors. We do not attempt error decomposition, i.e., to model individual sources of error using separate error models [e.g., Kavetski et al., 2006; Renard et al., 2011].

The simulated system output $\widehat{\boldsymbol{Y}} = (\widehat{\boldsymbol{Y}}_t)_{t=1,..,N}$ differs from the observed system output, $\widetilde{\boldsymbol{Y}} = (\widetilde{\boldsymbol{Y}}_t)_{t=1,..,N}$, for several reasons: i) errors in the observed system inputs, e.g., raingauge sampling errors [Renard et al., 2011; McMillan et al., 2011]; ii) errors in the observed system output, e.g., rating curve errors [Thyer et al., 2009; McMillan and Westerberg, 2015]; iii) structural errors in the model equations, e.g., due to the inability of lumped models to represent spatially distributed processes, incomplete representation of dominant hydrological processes, etc. [Beven and Binley, 1992; Kuczera et al., 2006]; (iv) parameter errors, including those due to the uncertainty arising from finite-length and uncertain calibration data, due to limitations of parameter optimization algorithms, etc.

Consider the vector of model residuals $\boldsymbol{\varepsilon}$,

$$\boldsymbol{\varepsilon} = q(\widetilde{\boldsymbol{Y}}) - q(\widehat{\boldsymbol{Y}}) \tag{2}$$

where we allow for a response transformation $q()$ [e.g., logarithmic, see McInerney et al., 2017].

By construction, Equation 2 lumps all sources of error in the residual error term. Therefore, in this case, likelihood-based inference requires the specification of a statistical model of the residual errors, i.e., a "residual error model",

$$\boldsymbol{\varepsilon} \sim \boldsymbol{\mathcal{E}}(\widehat{\boldsymbol{Y}}, \boldsymbol{\vartheta}) = \boldsymbol{\mathcal{E}}(\boldsymbol{X}, \boldsymbol{\theta}, \boldsymbol{\vartheta}), \tag{3}$$

where $\boldsymbol{\vartheta}$ denotes the error model parameters (which can be inferred or fixed a priori).

Hydrological model residuals are often well-described by Gaussian AR(1) processes ("red noise") [Schaefli et al., 2007; Evin et al., 2013; Li et al., 2013].

$$\varepsilon_t = \mu_{\boldsymbol{\varepsilon}} + \rho(\varepsilon_{t-1} - \mu_{\boldsymbol{\varepsilon}}) + \delta_t, \delta_t \sim \mathrm{NID}(\mu_\delta, \sigma_\delta^2) \tag{4}$$

where δ_t is the innovation at time step t, ρ is the (lag-1) autoregressive parameter, μ_δ is the innovation mean, σ_δ^2 the innovation variance, $\mu_{\boldsymbol{\varepsilon}}$ the residual mean and NID denotes the independent Gaussian distribution. To assist in the residual error analysis, it is convenient to define the innovation mean, $\mu_\delta = \mu_{\boldsymbol{\varepsilon}}(1 - \rho_{\boldsymbol{\varepsilon}})$.

The BTL corresponding to AR(1) residuals with Gaussian innovations is:

$$p(\widetilde{\boldsymbol{Y}} \mid \boldsymbol{X}, \boldsymbol{\theta}, \boldsymbol{\vartheta}) = \frac{\sqrt{1-\rho^2}}{(2\pi\sigma_\delta^2)^{N/2}} \exp\left[-\frac{1}{2\sigma_\delta^2}\left\{(1-\rho^2)(\varepsilon_1 - \mu_{\boldsymbol{\varepsilon}})^2 + \sum_{t=2}^{N}(\varepsilon_t - \mu_{\boldsymbol{\varepsilon}} - \rho(\varepsilon_{t-1} - \mu_{\boldsymbol{\varepsilon}}))^2\right\}\right]. \tag{5}$$

where ε_t are the raw residuals computed from Equation 2 [see e.g. Priestley, 1981, for the derivation of the probability density of Gaussian AR(1) processes].

In many cases, the assumption of Gaussian errors is not supported by residual analysis [e.g., Schoups and Vrugt, 2010]. An alternative assumption that we consider in this study is that the innovations follow a Laplace distribution, with pdf $f_{\text{Laplace}}(x|\mu, b) = 1/(2b)\exp\left(-|x-\mu|/b\right)$, where the standard deviation is $\sqrt{2}b$.

The BTL corresponding to AR(1) residual errors with Laplacian innovations is

$$p(\widetilde{\boldsymbol{Y}} \mid \boldsymbol{X}, \boldsymbol{\theta}, \boldsymbol{\vartheta}) = \sqrt{1-\rho^2}\left(\frac{\sqrt{2}}{2\sigma_\delta}\right)^N \exp\left[-\frac{\sqrt{2}}{\sigma_\delta}\left\{\sqrt{1-\rho^2}|\boldsymbol{\varepsilon}_1 - \mu| + \sum_{t=2}^{N}|\boldsymbol{\varepsilon}_t - \mu - \rho(\boldsymbol{\varepsilon}_{t-1}-\mu)|\right\}\right], \tag{6}$$

For details of more general AR(n) processes, see Box and Jenkins [1976, p. 274ff].

2.2. Spectral-domain: basic concepts

To derive the likelihood of the model residuals in the spectral-domain rather than in the time-domain, we need the same key ingredients as for BTL: (i) a residual error model in the spectral-domain (e.g., based on the power-density spectrum of the stochastic process assumed to describe the residuals), (ii) a parametric description of the probability distribution function associated with this error model, and (iii) spectral-domain realizations of the model residuals (either obtained directly in the spectral-domain or from a transform of time-domain realizations). The derivation of these ingredients is presented next.

In the derivations to follow, it is important to distinguish between the application of Fourier transforms to deterministic vs stochastic processes. In general, we use lower case symbols (e.g., $\boldsymbol{z}$) to denote deterministic processes and realizations (samples) from stochastic processes and upper case symbols (e.g., $\boldsymbol{Z}$) to denote stochastic processes (for example, the process $\boldsymbol{Z}$ that generated the realization $\boldsymbol{z}$).

Consider a deterministic process $\boldsymbol{z} = (z_t)_{t=-\infty,..,\infty}$, defined over discrete time steps t of length Δt. If the process is absolutely summable [Oppenheim and Schafer, 1989, p. 47],

its discrete time Fourier transform can be written as

$$\mathbb{F}_\omega[\boldsymbol{z}] = \Delta t \sum_{t=-\infty}^{\infty} z_t \exp(-it\omega\Delta t) \tag{7}$$

where $i = \sqrt{-1}$.

The discrete time Fourier transform $\mathbb{F}_\omega$ is a vector of complex numbers. Its components are indexed by the angular frequency ω [rad/T] [e.g., Oppenheim and Schafer, 1989, p. 698].

Next, consider a stochastic process $\boldsymbol{Z} = (Z_t)_{t=-\infty,..,\infty}$. In this case, the Fourier transform $\mathbb{F}_\omega[\boldsymbol{Z}]$ is itself a stochastic process, obtained as a derived distribution. In particular, the Fourier transform maps a time series of random variables $\boldsymbol{Z}$ into a set of frequency-ordered random variables $\mathbb{F}_\omega[\boldsymbol{Z}]$. This can be seen by considering the application of the Fourier transform to a set of individual time series $\boldsymbol{z}$ sampled from a stochastic process $\boldsymbol{Z}$, and then considering the distribution of the set of transformed time series $\mathbb{F}_\omega[\boldsymbol{z}]$.

Stochastic processes can be analyzed using the power-density spectrum (PDS), defined as the Fourier transform of the autocovariance function of $\boldsymbol{Z}$ [Oppenheim and Schafer, 1989, p. 843]:

$$\mathbb{P}_\omega[\boldsymbol{Z}] = \mathbb{F}_\omega\left[\upsilon_\ell[\boldsymbol{Z}]\right] \tag{8}$$

where $\upsilon_\ell[\boldsymbol{Z}] = \mathrm{E}\left[Z_t \ \mathrm{conj}(Z_{t+\ell})\right]$ is the autocovariance function of process $\boldsymbol{Z}$, E[] is the expectation operator, ℓ is the lag [T] and conj() denotes complex conjugation. Note that the autocovariance function is often referred to as the "autocorrelation sequence" in the signal processing literature [e.g., Oppenheim and Schafer, 1989, p. 743].

Assuming process ergodicity, the PDS of a stochastic process can be related to the expectation of the Fourier transform of the stochastic process. For a finite-domain process

$\boldsymbol{Z}^{(N)} = (Z_t)_{t=1,..,N}$, it holds that [Oppenheim and Verghese, 2015, chapter 11]

$$\mathbb{P}_\omega[\boldsymbol{Z}^{(N)}] = \frac{1}{N}\mathrm{E}\left[\left|\mathbb{F}_\omega[\boldsymbol{Z}^{(N)}]\right|^2\right], \tag{9}$$

where $|\cdot|$ denotes the absolute value (or, more generally, the complex modulus). The PDS of an infinite-domain process is obtained by taking the limit $N \to \infty$ in Equation 9.

In the signal processing literature, both formulations of the PDS are attributed to Wiener-Khinchin; in recent literature, the formulation in Equation 9 is referred to as the Einstein-Wiener-Khinchin theorem [Oppenheim and Verghese, 2015, chapter 11].

We stress that, unlike the Fourier transform of a stochastic process, the PDS of a stochastic process is a deterministic quantity: it is defined either in terms of the autocovariance function (Equation 8) or in terms of expectations (Equation 9).

2.3. The PDS variate, PDSV

Given a finite-length sample $\boldsymbol{z}^{(N)} = (z_t)_{t=1,..,N}$, the Fourier transform $\mathbb{F}_\omega[\boldsymbol{z}^{(N)}]$ provides a spectral-domain sample of the discrete-time Fourier transform of the entire infinite-domain process $\boldsymbol{Z}$ [Oppenheim and Schafer, 1989, p. 695].

Due to the finite length of the sample, the mapping between the time- and spectral-domain is possible only at a finite number of frequencies ω_j. These frequencies are given by integer multiples of the fundamental frequency ω_f:

$$\omega_j = j\omega_f; j = 0, .., N-1 \tag{10}$$

$$\omega_f = \frac{2\pi}{N\Delta t} \tag{11}$$

Accordingly, we will use the subscript j as the index of the discrete-time finite-sample Fourier transform. Since the exponential basis functions used by the Fourier transform are

orthogonal and complex-valued, $N/2$ frequencies are sufficient to describe the N elements of $\boldsymbol{z}^{(N)}$ and hence we have $j = 0, ..., N/2 - 1$.

We now define a transformation of a stochastic process such that the expected value of this transformation is the PDS of the original stochastic process. To this end, we define the "power-density spectrum variate" (PDSV) corresponding to a finite-length process $\boldsymbol{Z}^{(N)} = (Z_t)_{t=1,..,N}$, as:

$$\mathbb{Q}_j[\boldsymbol{Z}^{(N)}] = \frac{1}{N\Delta t}\left|\mathbb{F}_j\left[\boldsymbol{Z}^{(N)}\right]\right|^2. \tag{12}$$

The following observations can be made:

1. The probability distribution $p(\mathbb{Q}[\boldsymbol{Z}^{(N)}])$ is a derived distribution that depends on $p(\boldsymbol{Z}^{(N)})$;

2. The power-density spectrum $\mathbb{P}_j[\boldsymbol{Z}^{(N)}]$ is the expectation of $\mathbb{Q}_j[\boldsymbol{Z}^{(N)}]$, i.e. $\mathbb{P}_j[\boldsymbol{Z}^{(N)}] = \mathrm{E}\left[\mathbb{Q}_j[\boldsymbol{Z}^{(N)}]\right]$;

3. A sample from the distribution of $\mathbb{Q}_j[\boldsymbol{Z}^{(N)}]$ can be obtained by applying the transformation in Equation 12 to a time-domain sample $\boldsymbol{z}^{(N)}$ from the stochastic process $\boldsymbol{Z}^{(N)}$.

$$P_j[\boldsymbol{z}^{(N)}] = \frac{1}{N\Delta t}\left|\mathbb{F}_j\left[\boldsymbol{z}^{(N)}\right]\right|^2. \tag{13}$$

The quantity $P_j[\boldsymbol{z}^{(N)}]$ is often called the periodogram of $\boldsymbol{z}^{(N)}$ (usually defined with $\Delta t = 1$) [Oppenheim and Schafer, 1989]. The periodogram can be seen to represent a "single sample" estimator of the mean of the power-density spectrum.

2.4. Statistical properties of quantities in the spectral-domain

We now turn our attention to the probability distributions of quantities in the spectral-domain.

For $j \geq 1$, the quantity $\left|\mathbb{F}_j\left[\boldsymbol{Z}^{(N)}\right]\right|^2$ is known to have the χ^2-distribution with two degrees of freedom, i.e. an exponential distribution with pdf $f_{\exp}(x|\beta) = 1/\beta\exp(-x/\beta)$, where $\mathrm{E}[x] = \beta$ and $\mathrm{var}[x] = \beta^2$ [e.g. Bartlett, 1950]. The periodogram at $j = 0$ follows the χ^2 distribution with 1 degree of freedom (see below). These distributional properties of the PDS variate have important implications for parameter inference, and are elaborated in further detail in Section 2.10.

Since $P_j[\boldsymbol{z}^{(N)}]$ represents a sample from $\mathbb{Q}_j[\boldsymbol{Z}^{(N)}]$, and $\mathbb{P}_j[\boldsymbol{Z}^{(N)}]$ is by definition the expected value of $\mathbb{Q}_j[\boldsymbol{Z}]$, we can express the probability density of $P_j[\boldsymbol{z}^{(N)}]$ for $j > 0$ using the exponential pdf with mean $\mathbb{P}_j[\boldsymbol{Z}] = \mathrm{E}[\mathbb{Q}_j[\boldsymbol{Z}^{(N)}]]$:

$$p(P_j[z]|\mathbb{P}_j[\boldsymbol{Z}]) = f_{\exp}(P_j[z]|\mathbb{P}_j[\boldsymbol{Z}]); j = 1, .., N/2 - 1, \tag{14}$$

where, for simplification, we omitted the time-domain superscript (N) on Z.

The probability distribution of $\mathbb{Q}_j[\boldsymbol{Z}^{(N)})]$ at $j = 0$ is the scaled χ^2 distribution with 1 degree of freedom,

$$p(P_j[z]|\mathbb{P}_j[\boldsymbol{Z}]) = f_{\chi^2_1}(P_j[z]|\mathbb{P}_j[\boldsymbol{Z}]); j = 0, \tag{15}$$

where $f_{\chi^2_1}(x|\beta) = \frac{1}{\beta}\frac{1}{\sqrt{2\pi(x/\beta)}}\exp[-(x/\beta)/2]$.

Unlike the exponential distribution, for the χ^2 distribution with 1 degree of freedom, we have $\mathrm{E}[x] = \beta$ and $\mathrm{var}[x] = 2\beta^2$. Inspection of the χ^2 distribution with 1 degree of freedom indicates that it corresponds to the distribution of the square of a Gaussian variate (hence it can be referred to as a "squared-Gaussian" distribution).

Having derived the distribution of the PDSV for all frequencies ($p(P_j[z]|\mathbb{P}_j[\boldsymbol{Z}])$ in equations 14 and 15, and having a method to compute samples from the PDSV (the peri-

189 odogram in Equation 13), the remaining step is to obtain an expression for the PDS
190 $\mathbb{P}_j[\boldsymbol{Z}]$ of specific stochastic processes used to describe residual errors.

2.5. PDS of uncorrelated processes

For simple uncorrelated processes, the following relations hold:

$$\mathbb{P}_0[\boldsymbol{Z}] = \Delta t \sigma_Z^2 + N \Delta t \mu_Z^2 \tag{16}$$

$$\mathbb{P}_j[\boldsymbol{Z}] = \Delta t \sigma_Z^2; j = 1, .., N/2 - 1. \tag{17}$$

191 Equation 17, obtained from Parseval's theorem [Jenkins and Watts, 1968], is a well-
192 known result and generally referred to as the "mean value of the periodogram".

193 Equation 16 is less known. In fact, the PDS spike at zero frequency of any process
194 with non-zero constant mean is often discarded, e.g., in the Whittle estimator [Whittle,
195 1953] used by Montanari and Toth [2007], and in the mixed time-domain spectral-domain
196 calibration presented by Morlando et al. [2016]. An example where it is explicitly included
197 is the analysis of De Vleeschouwer and Pauwels [2013].

2.6. Bayesian spectral likelihood for uncorrelated processes

The Bayesian spectral likelihood (BSL) of time-domain observations $\widetilde{\boldsymbol{Y}}$ is obtained as the joint probability of the corresponding spectral-domain residual realizations $P[\boldsymbol{\varepsilon}]$ at all frequencies. When the residual errors are assumed to be uncorrelated, we obtain,

$$\begin{aligned} p(\widetilde{\boldsymbol{Y}} \mid \boldsymbol{X}, \boldsymbol{\theta}, \boldsymbol{\vartheta}) &= p(P_0[\boldsymbol{\varepsilon}]|\mathbb{P}_0[\boldsymbol{\mathcal{E}}]) \prod_{j=1}^{N/2-1} p(P_j[\boldsymbol{\varepsilon}]|\mathbb{P}_j[\boldsymbol{\mathcal{E}}]) \\ &= f_{\chi_1^2}(P_0[\boldsymbol{\varepsilon}]|\mathbb{P}_0[\boldsymbol{\mathcal{E}}]) \prod_{j=1}^{N/2-1} f_{\exp}(P_j[\boldsymbol{\varepsilon}]|\mathbb{P}_j[\boldsymbol{\mathcal{E}}]), \end{aligned} \tag{18}$$

198 where $\mathbb{P}_0[]$ is given in Equation 16 and $\mathbb{P}_j[]$ is given in Equation 17.

Note that the PDSV is obtained from the actual realization of residuals $\boldsymbol{\varepsilon}$ (Equation 2), whereas the PDS is a property of the underlying stochastic process $\boldsymbol{\mathcal{E}}$ (e.g., Gaussian).

Substituting the expressions for $\mathbb{P}_0[\boldsymbol{Z}]$ and $\mathbb{P}_j[\boldsymbol{Z}]$ from Equation 16 and Equation 17, and assuming $\Delta t = 1$ and $\mu^2_{\boldsymbol{\mathcal{E}}} = 0$, yields the BSL for zero-mean white noise with variance σ_δ:

$$p(\widetilde{\boldsymbol{Y}} \mid \boldsymbol{X}, \boldsymbol{\theta}, \sigma_\delta) = f_{\chi^2_1}(P_0[\boldsymbol{\varepsilon}]|\sigma^2_\delta) \prod_{j=1}^{N/2-1} f_{\exp}(P_j[\boldsymbol{\varepsilon}]|\sigma^2_\delta), \tag{19}$$

where the definitions of $f_{\exp}()$ and $f_{\chi^2_1}()$ are as given in Section 2.3.

2.7. PDS of autocorrelated processes

As the residuals of hydrological errors are typically highly autocorrelated, it is of interest to consider the PDS of autocorrelated processes.

Regardless of the autocorrelation structure, Equation 16 holds for frequency $j = 0$. However, for $j > 0$, the PDS of an autocorrelated process depends on j.

For Gaussian red noise, AR(1), it can be shown that [Brockwell and Davis, 1987]

$$\mathbb{P}_j[\boldsymbol{Z}] = \frac{\Delta t \sigma^2_Z}{\rho^2 \sin^2(\omega_j) + [1 - \rho\cos(\omega_j)]^2}. \tag{20}$$

More generally, the PDS of any stationary process $\boldsymbol{Z}$ with finite variance and linear autocorrelation structure (e.g., with moving-average components, non-Gaussian innovations, etc.) can be expressed using a "profile" function, $\zeta^Z_j(\boldsymbol{\vartheta}_\zeta)$ [Fox and Taqqu, 1986]:

$$\mathbb{P}_j[\boldsymbol{Z}^{(N)}] = \Delta t \sigma^2_Z \zeta^Z_j(\boldsymbol{\vartheta}_\zeta). \tag{21}$$

where $\boldsymbol{\vartheta}_\zeta$ are the parameters of the autocorrelation structure.

The profile function ζ_j^Z is a function of the frequency index j and depends solely on the autocorrelation structure of the process; it depend neither on the probability distribution e innovations, nor on the variance of the innovations.

Equation 20 can be derived from the general Equation 21, with parameters $\boldsymbol{\vartheta}_\zeta = \rho$ [Box et al., 1994]. The corresponding expressions for other autoregressive processes can be found in references such as Box et al. [1994] and others.

2.8. General Bayesian spectral likelihood

The general BSL formulation is obtained by expressing the joint probability distribution of the PDS variate (PDSV) at all frequencies $j = 0,..,N/2-1$:

$$p(\widetilde{\boldsymbol{Y}} \mid \boldsymbol{X}, \boldsymbol{\theta}, \boldsymbol{\vartheta}) = \prod_{j=0}^{N/2-1} f_j(P_j[\boldsymbol{\varepsilon}]|\mathbb{P}_j[\boldsymbol{\mathcal{E}}]). \tag{22}$$

Recalling the different form of the probability distribution for frequency $j = 0$ (Equation 15) than for frequencies $j > 0$ (Equation 14), the above equation becomes:

$$p(\widetilde{\boldsymbol{Y}} \mid \boldsymbol{X}, \boldsymbol{\theta}, \boldsymbol{\vartheta}) = f_{\chi_1^2}(P_0[\boldsymbol{\varepsilon}]|\mathbb{P}_0[\boldsymbol{\mathcal{E}}]) \prod_{j=1}^{N/2-1} f_{\exp}(P_j[\boldsymbol{\varepsilon}]|\mathbb{P}_j[\boldsymbol{\mathcal{E}}]), \tag{23}$$

$\mathbb{P}_0$ is given by Equation 16 and $\mathbb{P}_j$ by Equation 21:

$$\mathbb{P}_0[\boldsymbol{\mathcal{E}}] = \Delta t \sigma_\delta^2 + N \Delta t \mu_\varepsilon^2 \tag{24}$$

$$\mathbb{P}_j[\boldsymbol{\mathcal{E}}] = \Delta t \sigma_\delta^2 \zeta_j^{\boldsymbol{\mathcal{E}}}[\boldsymbol{\vartheta}_\zeta]. \tag{25}$$

The above equations hold for any homoscedastic (constant-variance) residual model, independent of the residual distribution. For hydrological model residuals, the main focus will be on AR(1) residual models, for which $\zeta_j^{\boldsymbol{\mathcal{E}}}$ is given by Equation 20. Note

223 that the homoscedasticity assumption can be addressed by using transformations such as
224 logarithmic or Box-Cox when calculating the residuals in Equation 2 (see below).

2.9. Incorporation of BSL into a full Bayesian framework

The preceding Section 2.6 and Section 2.8 derived the likelihood function $p(\widetilde{\boldsymbol{Y}}|\widetilde{\boldsymbol{X}}, \boldsymbol{\theta}, \boldsymbol{\vartheta})$ for Bayesian spectral-domain inference. The Bayesian posterior distribution $p(\boldsymbol{\theta}, \boldsymbol{\vartheta}|\widetilde{\boldsymbol{Y}}, \widetilde{\boldsymbol{X}})$ is then obtained by specifying a prior distribution for all inferred quantities, $p(\boldsymbol{\theta}, \boldsymbol{\vartheta})$,

$$p(\boldsymbol{\theta}, \boldsymbol{\vartheta}|\widetilde{\boldsymbol{Y}}, \widetilde{\boldsymbol{X}}) \propto p(\widetilde{\boldsymbol{Y}}|\widetilde{\boldsymbol{X}}, \boldsymbol{\theta}, \boldsymbol{\vartheta}) \times p(\boldsymbol{\theta}, \boldsymbol{\vartheta}) \tag{26}$$

225 The specification of the prior distribution allows incorporating existing (approximate)
226 knowledge of hydrological model parameters, eg, based on previous investigations [e.g.,
227 Viglione et al., 2013], theoretical constraints, as well as estimates of error model param-
228 eters from auxiliary studies such as rainfall and rating curve error analysis [e.g., Renard
229 et al., 2011]. In the simplest instance where such additional information is not avail-
230 able, such as in the case studies of this paper, a uniform prior distribution can be used,
231 $p(\boldsymbol{\theta}, \boldsymbol{\vartheta}) \propto const.$

In general, the likelihood function must account for any data transformations, such as the logarithmic or Box-Cox transformations often used to stabilize the error variance,

$$p(\widetilde{\boldsymbol{Y}}|\widetilde{\boldsymbol{X}}, \boldsymbol{\theta}, \boldsymbol{\vartheta}) = \det J_q(\widetilde{\boldsymbol{Y}}) \times p(\boldsymbol{\mathcal{E}}|\widetilde{\boldsymbol{X}}, \boldsymbol{\theta}, \boldsymbol{\vartheta}) \tag{27}$$

232 where $\det J_q$ denotes the Jacobian determinant of transformation q, e.g., in the case
233 of the logarithmic transformation $q(y) = \log y$ used in Section 3.4, we have $\det J_q(\tilde{y}) =$
234 $\prod_{t=1}^{N} 1/\tilde{y}$ [e.g., see McInerney et al., 2017]. Unless the data transformation includes fitted
235 parameters (e.g., the Box-Cox transformation applied with a fitted rather than fixed value

of the power parameter λ), the Jacobian term is constant with respect to the inferred quantities $(\boldsymbol{\theta}, \boldsymbol{\vartheta})$, and can be treated as part of the proportionality constant in Equation 26.

2.10. Theoretical advantages of BSL

An interesting and useful feature of BSL is that its likelihood function, given in Equation 23, is "almost" (asymptotically for large N) independent from the distribution of residuals in the original (time) domain. This behavior arises from the χ^2 form of the probability distribution of the PDS variate (PDSV), used to derive BSL (see equations 14 and 15).

Intuitively, the asymptotic properties of the PDSV can be related to the mathematical form of the Fourier transform, which is defined as a sum of a series of variables (see Equation 7). When these variables are random, as is the case when the Fourier transform is applied to a stochastic process, the Central Limit Theorem results in an asymptotic convergence to a Gaussian distribution almost irrespective of the distribution of the individual terms in the sum [see Brillinger, 1981; Cohen, 1998, for details]. Next, when we consider the definition of the power-density spectrum variate (PDSV, $\mathbb{Q}$) in Equation 12, we see that, for $j > 0$, the complex modulus operation results in the sum of squares of two Gaussian terms, which by definition yields the χ^2 distribution with 2 degrees of freedom, i.e., the exponential distribution. A slightly different result holds for $j = 0$ (see Equation 14).

A more formal derivation of the (asymptotic) distributional properties of spectral quantities is provided by Brillinger [1981, theorem 4.4.2] and [Cohen, 1998]. This behavior and associated advantages are illustrated and discussed in Section 4.

It is emphasized that, strictly speaking, the Power-density spectrum variate (PDSV) follows the exact χ^2 distribution only for white noise (uncorrelated homoscedastic Gaussian processes). For strongly non-Gaussian, heteroscedastic and/or auto-correlated processes, the distribution of the PDSV converges to the χ^2 distribution asymptotically as $N \to \infty$ [e.g., Duchon and Robert Hale, 2012, chapter 1]. The greater the departure from white noise, the longer data period (larger value of N) is needed before the χ^2 distribution becomes a reasonable approximation (similar to Central Limit Theorem converging slower when summing highly non-Gaussian, heteroscedastic and/or autocorrelated random variables). Given the long times series used in this paper (e.g., $N = 2^{15}$ in case study 2), convergence of the PDSV to the χ^2 distribution is not a limiting factor.

Another useful theoretical feature of BSL is that it can quite readily accommodate virtually any residual autocorrelation structure, as long as the autocorrelation decays to zero. This can be achieved by substituting the appropriate parameterization for $\zeta_j^{\boldsymbol{\varepsilon}}$ in Equation 25. Examples of estimating the parameters of the well-known "$1/f$" noise [e.g. West and Shlesinger, 1990; Ward and Greenwood, 2007a] and of a process with an exponentially decaying PDS are provided in Section 3. In contrast, it might be difficult to derive the corresponding autocorrelation functions in the time-domain, and indeed impossible in the case of "$1/f$" noise [Ward and Greenwood, 2007b]. The practical advantages afforded by this flexibility of BSL are discussed further in Section 5.

2.11. Relationship to the estimator of Montanari and Toth (2007)

Montanari and Toth [2007] have previously investigated the calibration of hydrological models using the maximum likelihood estimator introduced by Whittle [1953]. The

likelihood proposed by Montanari and Toth [2007] is

$$p(\widetilde{\boldsymbol{Y}} \mid \boldsymbol{X}, \boldsymbol{\theta}, \boldsymbol{\vartheta}) = \prod_{j=1}^{N/2} f_{\exp}(P_j[\widetilde{\boldsymbol{Y}}] | P_j[\boldsymbol{H}(\boldsymbol{\theta}, \widetilde{\boldsymbol{X}})] + \mathbb{P}_j[\boldsymbol{\mathcal{E}}|\boldsymbol{\vartheta}]) \tag{28}$$

where $P_j[\boldsymbol{H}(\boldsymbol{\theta}, \widetilde{\boldsymbol{X}})]$ is the periodogram of the model simulation and is used as an estimate of the PDS of the hydrological model $\mathbb{P}_j[\boldsymbol{H}(\boldsymbol{\theta})]$. The term $\mathbb{P}_j[\boldsymbol{\mathcal{E}}|\boldsymbol{\vartheta}]$ is the PDS of the residual model as in Equation 23, and $P_j[\widetilde{\boldsymbol{Y}}]$ is the periodogram of the observed output. Note that, as discussed in Section 2.11, Equation 28 does not include the PDS value for $j = 0$. Furthermore, both periodograms (of the model simulation and of the observed output) are obtained via the Fast Fourier transform without windowing [Montanari and Toth, 2007, and the R-code available from the authors].

The approach suggested by Montanari and Toth [2007], which will be referred to as the "Montanari-Toth likelihood" (MTL), requires two important assumptions:

1. The PDS of the observations $\widetilde{\boldsymbol{Y}} = \boldsymbol{H}(\boldsymbol{\theta}, \widetilde{\boldsymbol{X}}) + \boldsymbol{\mathcal{E}}$ can be approximated as $\mathbb{P}[\widetilde{\boldsymbol{Y}}] = \mathbb{P}[\boldsymbol{H}(\boldsymbol{\theta}, \widetilde{\boldsymbol{X}})] + \mathbb{P}[\boldsymbol{\mathcal{E}}]$ or, more generally, $\mathbb{P}[q(\widetilde{\boldsymbol{Y}})] = \mathbb{P}[q(\boldsymbol{H}(\boldsymbol{\theta}, \widetilde{\boldsymbol{X}}))] + \mathbb{P}[\boldsymbol{\mathcal{E}}^*]$ when the residuals $\boldsymbol{\mathcal{E}}^*$ are defined in transformed space. This approximation holds only if the hydrological model and the residual model are independent, i.e., if their cross-spectrum is zero [e.g., see Brockwell and Davis, 1987].

2. The generally unknown PDS of the hydrological model $\mathbb{P}[\boldsymbol{H}(\boldsymbol{\theta}, \widetilde{\boldsymbol{X}})]$ can be approximated by the periodogram of a model simulation, $P[\boldsymbol{H}(\boldsymbol{\theta}, \widetilde{\boldsymbol{X}})]$. As the periodogram is a "single-sample" estimator of the PDS, this approximation introduces additional noise into the estimation procedure. More stable (less noisy) PDS estimators exist [e.g., Welch, 1967] but have not been investigated in the context of the Whittle estimator for hydrological model calibration.

As shown by Montanari and Toth [2007], assumptions 1-2 above can often provide useful practical results. However, the additional approximation errors introduced by these assumptions can be expected to inflate posterior parameter uncertainty compared to BSL, which does not make these assumptions. These theoretical considerations are investigated empirically in Section 4 and discussed in Section 5.

Note also that MTL is formulated in terms of the observed streamflows given a model simulation and a residual error model, whereas BSL is formulated in terms of the observed residuals given a residual error model. In this respect, the convenience of the BSL formulation presented in this work is that it does not require dealing with spectral analysis of the hydrological model equations.

3. Case studies

Four case studies are presented: (i) illustration of the properties of the periodogram (used in the BSL inference), using pure random processes; (ii) inference of parameters of autocorrelation functions of pure random processes; (iii) synthetic hydrological calibration, where we investigate the inference of hydrological and error model parameters under controlled conditions; and (iv) real hydrological calibration, where we investigate parameter inference when model assumptions are not fully met.

A summary of the case studies is given in Table 3 and a summary of notations in Table 4.

3.1. Case study 1 (synthetic): Properties of the periodogram / BSL

One of the interesting properties of the periodogram of a random process is that its elements follow (approximately) the χ^2 probability distribution regardless of the probability distribution of the original random process (see Section 2.8). This property is illustrated

for uncorrelated non-Gaussian processes, with innovations from the following four probability distributions: i) uniform distribution in $[-1, 1]$, ii) Laplace distribution with $\mu = 0$, $\sigma = 1$, iii) bimodal Gaussian distribution with $\mu_1 = 1$, $\mu_2 = -1$, $\sigma_1 = \sigma_2 = 0.5$ and weight 0.2 of the first component; (iv) an AR(1) process with the parameters of the Gaussian error model of Table 4, selected based on hydrological experience.

The methodology employed to empirically confirm the probability distribution properties of the periodogram is given in Appendix A1.

3.2. Case study 2 (synthetic): Inference of pure random processes

The ability of BSL to retrieve the parameters of the process that generated the "observed" data are first illustrated using a synthetic case study based on pure random processes (i.e., without a deterministic component).

The following stochastic processes are investigated:

$$f_{\mathbb{P}1}(\omega; A_1, B_1) = A_1 \exp(-B_1 \omega) \tag{29}$$

$$f_{\mathbb{P}2}(\omega; A_2, B_2) = A_2 / \omega^{B_2} \tag{30}$$

with reference parameter values $(A_1, B_1) = (100, 0.001)$ and $(A_2, B_2) = (10.5, 3)$, respectively.

The methodology for generating the synthetic data for this case study is detailed in Appendix A2. This analysis allows establishing the theoretical properties of BSL under idealized conditions. As we do not carry out a Markov Chain Monte Carlo (MCMC) analysis of the posterior distribution, we are limited to examining the properties of the optimal BSL estimate, rather than of the entire BSL distribution. Note that optimization

of the likelihood corresponds to optimization of the Bayesian posterior under uniform prior assumptions.

3.3. Case study 3 (synthetic): simple hydrological model

Following the basic verification of the BSL using pure random processes, we investigate its properties when applied to hydrological models with synthetic rainfall-runoff data. The synthetic data is generated using a simple rainfall generator and hydrological model, in order for the synthetic streamflow data to generally resemble real observations; see Appendix A3 for a detailed description.

The rainfall generator used is a Poisson rectangular pulse model with an exponential distribution for both the rain cell intensity and the duration [e.g. Bierkens and Puente, 1990]. The model has three parameters: the arrival rate λ, the mean intensity $\bar{i}_r$ and the mean duration $\bar{t}_r$.

The hydrological model used is a simple model with two linear reservoirs in series, described by three parameters. The reservoir outflow is $q = ks$, where q is the outflow, s is the storage and k^{-1} is the residence time. The residence time of the first reservoir, k_1^{-1}, is shorter than the residence time of the second reservoir, k_2^{-1} . The leaching from reservoir 1 to reservoir 2 is assumed to be constant and equal to l_g. We refer to this model as the *simple linear HM* ("linearity" here refers to the flux formulation).

The following residual error models are used: (i) Gaussian AR(1) process (Section 2.1); (ii) Laplacian AR(1) process (Section 2.1); and (iii) a process with the autocorrelation structure $\upsilon_\ell = f(\ell; \varrho_1, \varrho_2) = \exp(-\varrho_1 \ell - \varrho_2 \ell^{0.5})$, which corresponds to the autocorrelation structure of the model residuals of Schaefli et al. [2007]. We consider cases where the error model is specified correctly and cases where it is misspecified. The analyses of misspecified

error models include erroneous distributional assumptions and erroneous autocorrelation structure assumptions.

The model parameters and the selected reference values for the synthetic case studies are summarized in Table 4. The maximum likelihood parameter set is estimated using the Nelder-Mead simplex algorithm [e.g. Press et al., 2007] in Matlab Version 2010b.

The statistical reliability of the predictions in this case study is assessed against multiple realizations of synthetic data using a predictive quantile-quantile plot, constructed as a generalization of the predictive qq-plot plot proposed by Thyer et al. [2009] for the case of a single reference realization (the observed data) (see Appendix A3).

3.4. Case study 4 (real data): Leaf River modeling

The behavior of BSL under real data conditions is investigated by calibrating the hydrological model HYMOD [Boyle, 2000] to the well-known Leaf River basin near Collins, Mississippi [e.g. Sorooshian et al., 1993; Vrugt et al., 2005; Smith et al., 2008]. The catchment has an area of about 1950 km^2. Daily area-average precipitation, evapotranspiration and streamflow estimates are available from the Hydrologic Research Laboratory of the National Weather Service. The calibration period ranges from October 1948 to September 1951. The validation period ranges from January 1951 to December 1969.

To stabilize the variance of the model residuals, we apply a log-transformation to the observed and simulated streamflow

$$\varepsilon_t = \log(\widetilde{y}_t + A) - \log(\widehat{y}_t + A), \tag{31}$$

where A is a small fixed offset to avoid numerical problems when applying the transformation to zero and near-zero flows. Here, we use $A = 10^{-4}$ (mm/d).

We assume that the residuals of log-transformed responses can be described by a random vector $\boldsymbol{\mathcal{E}}$ that (approximately) follows an AR(1) process with Gaussian innovations. In this case, BTL and BSL can be applied without further modification to the log-transformed residuals (see Section 2.9). The posterior parameter distribution with BSL or BTL are sampled using the Metropolis algorithm described in Schaefli et al. [2007], which was used to produce 1000 samples from a stable chain (no update of the sampling distribution).

All error model parameters are sampled jointly with the hydrological model parameters. We use uniform priors for all parameters except the error model innovation variance σ_δ^2, for which Jeffreys prior is used ($p(\sigma_\delta^2) = 1/\sigma_\delta^2$) [see Schaefli et al., 2007]. Note that the mean of the innovations μ_δ, which effectively acts as a mass balance parameter, is inferred jointly with all other parameters. In principle this estimation approach can lead to non-robust predictions, as shown empirically by Evin et al. [2014]. Although such non-robustness was not seen in the current case study, we note that joint inference of mass balance parameters, error variance and error autocorrelation should be undertaken with care to avoid poor inference and predictions.

4. Results

4.1. Case study 1: Theoretical properties

An important property of BSL, arising from its use of the periodogram, is that its formulation does not depend on the process distribution in the original (time-)domain (Section 2.10). This is illustrated in Figure **??**, which shows Gaussian qq-plots of the realizations from three different non-Gaussian random processes, and the χ^2 qq-plots of the corresponding periodograms. Figure **??**a-b show three theoretical process examples, while Figure **??**c-d applies this analysis to the (highly non-Gaussian) residuals time series

obtained from the inferred maximum likelihood parameter set of the Leaf River case study (Section 4.4). In all cases, irrespective of the process distribution in the original domain, the periodograms follow a χ^2 distribution with two degrees of freedom for all frequencies $j > 0$. This findings provides empirical confirmation of the theoretical considerations given in Section 2.10.

4.2. Case study 2: Inference of PDS parameters

Another important property of BSL is that it can be readily used to infer the parameters of processes with virtually any PDS. This is illustrated in Figure **??**, which shows, for each parameter of the two pure random processes given in Table 3, the distributions of optimal estimates obtained by maximizing the likelihood function over multiple process realizations with the same underlying true parameters (see Section 3.2).

Figure **??** show that the distributions are (correctly) centered on the true parameter values used to generate the original process realizations. The variability of the optimal estimates is indicative of the parametric uncertainty associated with fitting the model to finite-length realizations (in this particular example, parametric uncertainty is quite small, less than 5% in both cases, due to the relatively long realization used (see Appendix A2).

4.3. Case study 3: Synthetic hydrological calibration

4.3.1. Correct versus misspecified error models

The mathematical equivalence of BSL and BTL is illustrated in Figure **??**, which shows the parameter distributions obtained for the simple linear HM with a Gaussian AR(1) residual model where all model assumptions are respected. The inferred parameter distributions are almost indistinguishable for the BSL and BTL likelihoods.

Figure ?? shows the same experiment, except that the residuals are generated from a Laplace AR(1) distribution. We consider three likelihoods: Gaussian BTL, Laplace BTL and BSL (which remains unchanged because it does not depend on the assumed distribution of residual errors).

The parameter distributions inferred with BSL, Gaussian BTL and Laplace BTL are almost identical, and are centered on the true parameter values. This finding demonstrates the general robustness of the Gaussian BTL with respect to the underlying distribution of model residuals. This robustness is confirmed by repeating the same experiment (results not shown) with different residual error parameters (namely $\rho_\varepsilon \in 0, 0.5, 0.8$, $\mu_\delta \in 0, 0.23, 1$) and $\sigma_\delta \in 0.06, 0.13, 0.5$), and for log-normal residual distributions. This robustness expresses the fact that in any of these experiments, BTL-Gauss is maximized for almost the same parameter set as BTL-Laplace or BTL-lognormal (differences in optimal parameter values of a few percent). The robustness of least squares parameter estimates to moderate departures from Gaussian distribution assumptions is fairly well established in the statistical literature [e.g., White, 1981]. Note that here were are concerned with departures from the overall shape of the error distribution rather than to the presence of strong outliers, as in the latter case least squares estimates can indeed deteriorate very rapidly [Press et al., 2007].

The robustness of BTL does start to break down if wrong assumptions are made about the residual autocorrelation structure (rather than about the residual distribution). This is illustrated in Figure ??, which shows the parameter distributions obtained for BTL, BSL and MTL for the synthetic case when the residuals have an exponential rather than AR(1) autocorrelation structure (see Section 3.2). The parameter distributions show

that both BTL-Gauss-AR1 and BSL-AR1 yield unbiased parameter estimates. However, compared to the distributions obtained under the correct likelihood (called BSL-nonAR1), BTL-Gauss-AR1 yields too wide hydrological parameter distributions, in particular for parameter k_2 (compare Figure **??** top row and bottom row). The distributions of the parameters k_1, k_2 and l_g under BSL-AR1 are similar to the distributions obtained under the correct likelihood BSL-nonAR1 (compare Figure **??** 2nd row and bottom row). This finding suggests that BSL is slightly more robust than BTL to violations of assumptions describing the residual error autocorrelation.

The differences between the distributions become more visible when comparing their reliability using the predictive qq-plots shown in Figure **??**. These predictive qq-plots show the probability distribution of the underlying true reference simulations within the model simulations. The predictive qq-plot for BSL-nonAR1 is clearly closer to the 1:1 line (corresponding to a perfectly reliable probabilistic model) than for the other likelihoods. Finally, Figure **??** shows the spectral-domain differences between the PDS inferred under the assumptions of BTL-Gauss-AR1, BSL-AR1 and BSL-nonAR1 versus the true PDS. It can be seen that there is a relatively pronounced difference between the PDS inferred with the correct likelihood BSL-nonAR1 and the PDS obtained with the likelihoods assuming (wrongly) a AR1 process.

4.3.2. Comparison of MTL versus BSL and BTL

Figure **??** compares the parameter distributions for the simple linear HM obtained using the MTL likelihood to the parameter distributions obtained using BTL and BSL.

The MTL inference of all model parameters (hydrological and error model) is unbiased despite the fact that the error innovation mean cannot be inferred (Section 2.11). However,

the posterior distributions (of all parameters) have a larger variance under MTL than under BTL and BSL, in particular for the hydrological model parameter k_2 and the error model autocorrelation ρ.

In terms of sensitivity to residual error assumptions, MTL does not, in theory, depend on the residual distribution in the original domain (this property is similar to BSL). For the previously discussed case of Laplace distributed residuals, MTL indeed results in unbiased parameter distributions (Figure **??**), despite the fact that the mean of the innovations, μ_δ, cannot be inferred with MTL (the zero frequency does not enter the computation of the Whittle likelihood). This example shows that non-zero-mean residuals do not necessarily lead to biased MTL estimates.

In contrast, MTL is highly sensitive to wrong autocorrelation assumptions as demonstrated with the experiment with non-AR(1) residual realizations: the resulting parameter distributions are biased (Figure **??**), the prediction range does not correspond to the range of reference simulations (Figure **??**) and, compared to the periodogram of the residuals, the PDS does not show enough power for high frequencies (Figure **??**).

4.4. Case study 4: Leaf river case study

The posterior distributions of HYMOD and residual error model parameters inferred in the Leaf River case study using BTL-Gauss, BSL and MTL are shown in Figure **??**. The corresponding maximum likelihood parameter sets are listed in Table 5.

The distributions produced using BTL-Gauss and BSL are very similar. Notable differences arise between the inferred residual innovation means μ_δ, the HYMOD parameter b_{H} (which parameterizes the degree of spatial variability of the soil moisture) and the HYMOD parameter α (which parameterizes the distribution of flow between the slow

and quick reservoirs). MTL gives a very different range of posterior parameter values for the two reservoir parameters ν_s and ν_q resulting in lower base flow, slower recessions and a slower response to rainfall events. Overall, this leads to more strongly autocorrelated residuals, as reflected in the distribution of ρ for MTL.

Figure **??** shows the streamflow simulation during the validation period corresponding to the parameter set with the highest BSL value. Included are plots of the log-transformed streamflow to show the model performance during low flows, a plot of the corresponding residual time series, a plot of the residuals against the rank of the simulated streamflow, a predictive qq-plot and a plot of the partial autocorrelation of the residuals.

The diagnostic plots in Figure **??** show that, in the case of BSL, the logarithmic transformation stabilizes the variance of residual errors and the assumption of constant-variance Gaussian residuals holds at least approximately. The autocorrelation of the residual errors is reasonably approximated by the AR(1) process (Figure **??**f).

In contrast, the corresponding diagnostic plots for MTL clearly show that the results obtained with this likelihood do not comply with the underlying assumptions. In particular, the residuals are strongly non-symmetric (Figure **??**d), do not have an AR(1) autocorrelation structure (Figure **??**c,f) and are non-Gaussian (Figure **??**e).

The results for the maximum likelihood simulation with BTL are very similar to BSL Figure **??**. However, the residual time series computed in the BSL and BTL inferences are not identical. In particular, the distributions of residual model parameters are slightly different (especially for the innovation mean μ_δ, see Figure **??**), which translates into a different mean and total variance in the AR(1) residual model.

Accordingly, the total prediction limits for BSL and BTL are also different (Figure **??**). These limits are obtained from 500 random model realizations (hydrologic model simulation plus residual error realization) corresponding to random draws from the posterior parameter distributions. As can be seen in Figure **??**, the parametric uncertainty is relatively small for BSL, BTL and MTL (because of the length of the calibration data), and the majority of the predictive uncertainty is due to residual errors.

Overall, the total 90% prediction limits obtained with BSL and BTL from the 1000 samples span, respectively, around 93% and 90% of observed values in the calibration period and around 94% and 87% of observed values in the entire simulation period (calibration and validation). The predictive qq-plots for the validation period (Figure **??**a and d) show that both likelihoods lead to very similar statistical reliability, with minor deviations from the uniform distribution. Considering high flow and low flow separately, (Figure **??**b-c and e-f) suggests that BSL gives more reliable results than BTL for low flow simulations (Figure **??**c and f).

5. Discussion

Given the theoretical aspects presented in Section 2 and the results of the empirical case studies reported in Section 3, we are now in a position to discuss the advantages and limitations of the BSL approach, relate it to the existing techniques for parameter inference in the time- and spectral-domains, and outline directions for further investigations.

We begin by comparing BSL to other spectral calibration methods proposed in the hydrological literature, with a particular focus on the MTL approach [Montanari and Toth, 2007], and then make a broader comparison to traditional time-domain calibration (BTL).

The majority of spectral-domain calibration methods in the hydrological literature are heuristic, in the sense that they do not explicitly articulate a probabilistic model of the system of interest (here, the catchment and observation systems) when constructing the objective function. For example, consider the case of parameter calibration that searches for the hydrological parameter set to match the autocorrelation function of the simulated discharge and the autocorrelation function of the observed discharge, using the root-mean-squared-error as a distance metric [Moussu et al., 2011]. This approach is useful from the point of view of maximizing particular model fit features, and establishing the sensitivity of the fit to parameter values, but cannot provide *probabilistic* estimates of uncertainty in the estimated parameters and predictions.

In addition, heuristic approaches, such as matching the autocorrelation function, hide a number of assumptions, such as the distributional properties of the errors. For example, using the sum-of-squared differences between the autocorrelation functions of observed versus simulated streamflow implies an assumption that these differences follow an independent Gaussian distribution. Unless these assumptions are stated and tested explicitly, the ability of the inference to provide meaningful probabilistic estimates is questionable.

For these reasons, our interest in this study is on likelihood functions explicitly derived from probabilistic models of the hydrological system and observations systems.

To the best of our knowledge, MTL is the only approach in the hydrological literature where a likelihood function is formally articulated from a probabilistic description of the data. The Whittle likelihood employed in the MTL approach is used outside of hydrology, in particular to infer the parameters of time series models [Ives et al.] or to estimate the power-density spectrum of time series [Choudhuri et al., 2004]. However, it is usually used

in contexts where the PDS of the model can be computed directly [e.g., Montanari et al., 1997], whereas in the MTL approach, the PDS of the (error) model is approximated (with potentially large errors) by subtracting the periodogram of model simulations from the periodogram of observed times series (Section 2.11).

In this work, we use a different strategy when deriving the BSL approach - we articulate the spectral-domain error model by computing the distribution of residual errors in the spectral-domain from the time series of residuals in the time-domain.

What are the advantages and limitations of the BSL approach versus the MTL approach? By explicitly computing the residual error time series and then transforming to the spectral-domain, BSL avoids the approximations and ensuing noise incurred by MTL. This behavior can be seen in the empirical case studies 3.1 and 3.2 (Table 3, Figure **??**, Figure **??**) where parametric uncertainty in the BSL approach was similar to the BTL inference. Especially under synthetic conditions with the correct error model (case study 3.1, Figure **??**), we can take BTL as the reference solution because it works directly with the raw residuals without any spectral-domain transformations.

In contrast, MTL inference yields parameter distributions that are clearly wider than those of BTL and BSL, as can be seen in case studies 3.1, 3.2, 3.3 (Figure **??**, Figure **??**, Figure **??**). We can attribute the additional noise in MTL to at least two potential reasons: (i) MTL uses the periodogram to estimate the PDS of the hydrological model. As the periodogram is a single-sample estimator of the PDS, it increases the variance of the resulting parameter estimates; (ii) MTL excludes the 0th frequency of the PDS from the likelihood function and hence looses information about the mean of the stochastic process assumed to describe residual errors.

This paper has not investigated the individual impact of these approximations on the MTL inference. It is possible that the use of more robust PDS estimators, for example, Welch's method [Welch, 1967] could reduce the impact of some of the limitations. That said, testing the theoretical properties of approximations to spectral properties of hydrological models is difficult because, except for special cases, the true PDS of hydrological models is unknown. For example, Bierkens and Puente [1990] proposed analytical expressions for the autocorrelation of the outputs from a simple hydrological model forced with stochastic inputs. However, their derivations hold only for the specific model used and only for small lags; these results are hence of limited value with respect to more general analyses of the MTL inference scheme.

It is worth adding that, in BSL, the periodogram of residual errors is used *not* as an estimator of their power-density spectrum (PDS), but to compute samples of the power-density spectrum variate (PDSV). Hence, replacing the periodogram $P(\boldsymbol{\varepsilon})$ used in BSL by a more stable estimator of the PDS [e.g., the method of Welch, 1967] would be detrimental to the probabilistic properties of BSL, because it would (by construction) under-estimate the variability of residual errors in the spectral domain.

The advantage of the BSL approach in avoiding the approximations of MTL does not come free. In particular, the MTL approach appears simpler to apply to indirect calibration problems such as non-concomitant calibration, because the residual errors do not have to be computed explicitly. BSL will require further development to be applicable to this problem, including an approximation of the PDS of the residual errors that does not ignore the cross-spectrum (Section 2.11).

These extensions to the MTL and BSL approaches lie beyond the scope of current work and will be explored in follow-up studies.

We now shift our attention to a broad comparison of BSL and time-domain estimation (BTL). In view of the close correspondence of BSL and BTL, does BSL represents a new error model, or a new solution approach for the same error model as BTL? In principle, BSL is obtained via Fourier transform of the time-domain realization and, therefore, could be seen to rely on the same initial assumptions as BTL (here, that residuals follow a Gaussian AR(1) process). However, due to the properties of the Fourier transform and the PDS definition listed in Section 2, the influence of assumptions such as that the errors are Gaussian is greatly diminished. In this respect, the BSL approach could be viewed to represent a more robust error model, especially with respect to distributional assumptions.

The advantages / attractive features of BSL versus BTL can be summarizes as follows.

a) Theoretical robustness against violation of distributional assumptions. As noted in the theory section (Section 2.3), the power-density spectrum variate central to the BSL approach has a χ^2 distribution almost independently from the (residual error) time series it is computed from. This is an attractive theoretical property because it can be expected to reduce the impact of violating distributional assumptions such as Gaussian errors, etc.

We note that least squares methods are often robust against departures from Gaussian assumptions [e.g., White, 1981, , and case study 3.2]. For this reason, a comparison of BSL and BTL in catchments with strongly non-Gaussian errors is of interest and recommended for future work. The impact of strong outliers is of particular interest given the susceptibility of least squares estimation to this particular departure from Gaussianity [Press et al., 2007].

b) Flexibility in representing the autocorrelation structure of the model residuals. This flexibility arises due to the structure of BSL where the autocorrelation profile function ((21)) is formulated directly in the spectral-domain. In case study 2.2, we considered processes with autocorrelation structures that cannot be readily formulated in the time-domain. For example, the correlation structure of "1/f" (pink) noise has no simple representation in the time-domain. Using BTL for such problems would require approximations, e.g., by AR(n) processes, and does not appear robust (e.g., case study 3.3).

The theoretical flexibility of BSL over BTL in representing error autocorrelation becomes particularly attractive when modeling environmental processes with strong cyclic behavior. For example, water temperature time series typically exhibit a pronounced diurnal cycle [e.g., Comola et al., 2015]. A water temperature model that does not capture this diurnal cycle will generally yield residuals with a cyclic autocorrelation structure peaking every 24 hours. This type of autocorrelation structure is difficult to represent in the time-domain, but is relatively easier to represent in the spectral-domain (e.g., with the profile function, equation (21)). This example illustrates a case where BSL offers useful practical advantages for modeling environmental systems where the autocorrelation of the model and/or observation errors is (much) easier in the spectral-domain than in the time-domain.

c) Opportunities to extend the method to indirect calibration problems, by taking advantage of working in the spectral-domain (see discussion above).

The practicality of a calibration scheme such as BSL depends not only on its inferential properties, but also on its computational cost. A major computational feature of spectral-domain methods is their use of Fourier transformed quantities. In practical work, the

Fourier transform is invariably implemented using the Fast Fourier Transform (FFT) algorithm, which requires of the order of $N \log N$ operations [Rao et al., 2010]. BTL does not require any FFT operations (it operates exclusively in the time-domain), MTL requires a single FFT operation per likelihood evaluation (FFT of the hydrological model simulations, assuming the FFT of observed data is pre-computed once) and BSL also requires a single FFT operation per likelihood evaluation (FFT of the residual error time series). While the cost of FFT can be appreciable for very long time series, in most cases we expect it to be dominated by the cost of running the hydrological model, which in general requires the solution of differential and algebraic equations at each time step. Consequently, it is unlikely that the computational cost of FFT within the BSL (and MTL) approaches could be a major limiting factor in practical work.

Finally, in terms of future work, we note that many aspects of hydrological calibration in the spectral-domain remain poorly understood. Based on the findings reported in this paper, the following specific directions deserve focused investigation:

(a) Investigate the robustness of BSL versus MTL and BTL under conditions of strongly non-Gaussian errors. In principle, BSL and MTL should provide practical robustness, which should be established using both "realistically constructed" synthetic data and real data case studies, in particular in arid/semi-arid catchments;

(b) Apply BSL to environmental modeling problems where the error time series exhibit nontrivial persistence patterns (for example, the diurnal water temperature models mentioned earlier);

(c) Indirect calibration, including calibration using non-concomitant input-output data time series, and calibration in ungauged catchments;

(d) Using BSL in studies exploring error decomposition. In this paper, we focused exclusively on aggregated treatment of errors using a single residual error model. The alternative paradigm of error decomposition is of tremendous interest, as it allows estimating dominant sources of uncertainty and devising strategies for reducing these errors. In principle, BSL can be incorporated directly into the likelihood terms of hierarchical Bayesian approaches such as BATEA [Kavetski et al., 2006; Renard et al., 2011], but the advantages, limitations and practicalities of doing so remain to be established empirically.

6. Conclusions

This paper presents the derivation of the spectral-domain counterpart of the widely used time-domain likelihood for Bayesian inference of environmental models. The theoretical and empirical properties of the proposed Bayesian spectral likelihood (BSL) are compared to the properties of the Bayesian time-domain likelihood (BTL), and to the Whittle-type spectral-domain likelihood (MTL) previously proposed by Montanari and Toth [2007].

The key conclusions of this paper are as follows:

1. The Bayesian spectral-domain likelihood derived in this work is mathematically equivalent to its time-domain counterpart in the case when the residual errors are assumed to be Gaussian (and autocorrelated). However, the spectral formulation offers two theoretical benefits: (i) its likelihood function is (asymptotically) independent from the probability distribution of residual errors, and (ii) it can accommodate residual errors with more complicated autocorrelation structure (for which time-domain representations are difficult or impossible).

2. At least under synthetic conditions, the time-domain likelihood is relatively robust to departures from the assumption of Gaussian residuals, but this robustness breaks down

for departures from the assumed autocorrelation structure. This is an aspect in which the spectral-domain inference might offer practical benefits over time-domain inference.

3. For the Whittle-type spectral-domain likelihood proposed by Montanari and Toth [2007], the synthetic and real data studies suggest that the simplifying assumptions made in this likelihood tend to produce parameter distributions that are too wide compared to inference in the time-domain, and potentially biased when autocorrelation assumptions are violated. The Bayesian spectral-domain likelihood introduced in this work does not appear to suffer from these limitations and does not incur a loss of information compared to the corresponding time-domain inference.

4. The real data case study based on the Leaf River and the hydrological model HYMOD reinforces the (relative) robustness of the time-domain and spectral-domain inference for a typical hydrological setting. Both inferences produced similar results, despite some moderate departures from the residual error model assumptions.

The theoretical derivations and analyses presented in this paper represent the first step towards formal Bayesian inference in the spectral-domain. Further work is required to better understand the properties of spectral-domain inference and its potential advantages in environmental model calibration. Future studies will include: (i) investigation of the robustness of BTL and BSL in cases where the model residuals are strongly non-Gaussian, e.g., as common in models of arid and semi-arid catchments; (ii) a wider range of hydrological case studies to gain more general insights into the practical performance of spectral-domain inference, including for models with non-trivial / cyclic autocorrelation structures; (iii) extensions of BSL to parameter inference with non-concomitant input-

output time series; and (iv) extensions of BSL to more comprehensive inference setups with individual treatment or sources of uncertainty (error decomposition).

A Matlab implementation of BSL for AR(1) error models is available in the Supporting Information.

7. Acknowledgements

The research of the first author was supported by research grants of the Swiss National Science Foundation (SNF, PZ00P2_147366, PP00P2_157611). A Matlab implementation of the HYMOD model, including the Leaf river data set, was provided by the model identification toolbox of Hoshin Gupta (University of Arizona). No other data sets were used for this theoretical study. An R implementation of the MTL likelihood was provided by Alberto Montanari (University of Bologna). We thank two anonymous reviewers, Anna Sikorska and Chief Editor Alberto Montanari for their constructive comments during the review process.

Appendix A: Details of case study methodology

A1. Analysis of periodogram properties in case study 1

This section details the methodology employed in case study 1 to empirically confirm the probability distribution properties of the periodogram:

1. Select a time-domain process (e.g. from the list of case studies summarized in Table 3);

2. Generate a sample of multiple independent realizations from the stochastic process. Here we generate a sample $\boldsymbol{z}$ of length $N = 2^{13}$.

3. Compute the periodogram $\mathbb{P}_j(\boldsymbol{z})$ using the FFT operation [Welch, 1967];

4. Produce a Gaussian qq-plot of the sample z from in Step 2. This plot is used to illustrate that the distribution of the process $\boldsymbol{Z}$ in the time-domain is clearly non-Gaussian.

5. Produce a χ^2 qq-plot of the values of the periodogram $\mathbb{P}_j(\boldsymbol{z})$ from Step 3. This plot is used to demonstrate that distribution of the process in the spectral-domain follows the χ^2 distribution. Note that for an uncorrelated process, the distribution of the PDS element $\mathbb{P}_j[\boldsymbol{Z}]$ does not depend on the element index j.

A2. Generation of synthetic data and inference verification in case study 2

This section describes the methodology employed in case study 2 to generate the synthetic data and use it to verify the results of the inference und all tested likelihoods.

1. Select a stochastic process, defined by its PDS function $f_{\mathbb{P}}(\omega;\boldsymbol{\gamma})$, where $\boldsymbol{\gamma}$ is a set of parameters. We also select a reference set of parameters, $\boldsymbol{\gamma}_r$.

2. Generate a random process realization $\boldsymbol{z}^{(\imath)}_{\mathbb{P}(\gamma_r)}$ of length N with PDS $f_{\mathbb{P}}(\boldsymbol{\gamma}_r)$, as follows

$$\boldsymbol{z}^{(\imath)}_{\omega} \leftarrow N(0, 1^2) \tag{A1}$$

$$\boldsymbol{z}^{(\imath)}_{\mathbb{P}(\boldsymbol{\gamma}_r)} = f^{-1}_{\mathrm{FFT}}\Big[f_{\mathbb{P}}(\omega;\boldsymbol{\gamma}_r)^{0.5} f_{\mathrm{FFT}}(\boldsymbol{z}^{(\imath)}_{\omega})\Big], \tag{A2}$$

where f_{FFT} denotes the fast Fourier transform (FFT) operation, f^{-1}_{FFT} is the inverse FFT operation, and $\boldsymbol{z}^{(\imath)}_{\omega}$ is a standard Gaussian white noise realization (but any other probability distribution could be used here). $f_{\mathbb{P}}(\omega;\boldsymbol{\gamma}_r)^{0.5} f_{\mathrm{FFT}}(\boldsymbol{z}^{(\imath)}_{\omega})$ corresponds to an element by element (i.e. frequency by frequency) multiplication of the square-root of the PDS $f_{\mathbb{P}}(\boldsymbol{\gamma}_r)$ with $f_{\mathrm{FFT}}(\boldsymbol{z}^{(\imath)}_{\omega})$.

3. Infer the maximum BSL estimate $\widehat{\boldsymbol{\gamma}}^{(\imath)}$ by maximizing the log-BSL, $\log p(\boldsymbol{z}^{(\imath)}_{\mathbb{P}(\boldsymbol{\gamma}_r)} \mid \boldsymbol{\gamma})$ with respect simplex direct search algorithm [Lagarias et al., 1998] for this maximization. direct search algorithm [Lagarias et al., 1998] for this maximization.

4. Repeat steps 2-3 for $\imath = 1, .., m$ to obtain an empirical distribution of optimized parameter estimates $\widehat{\boldsymbol{\gamma}}^{(\imath)}$. All presented results use $m = 400$ replicates each of length $N = 2^{15}$.

5. Compare the empirical distribution of $\widehat{\boldsymbol{\gamma}}^{(\imath)}$ to the reference value $\boldsymbol{\gamma}_r$.

This analysis allows establishing the theoretical properties of BSL under idealized conditions. As we do not carry out a Markov Chain Monte Carlo analysis of the posterior distribution, we are limited to examining the properties of the optimal BSL estimate, rather than of the entire BSL distribution. Note that optimization of the likelihood corresponds to optimization of the Bayesian posterior under uniform prior assumptions.

A3. Generation of synthetic data in the hydrological case study 3

This section describes the methodology employed in case study 3 to investigate the BSL inference under synthetic hydrological conditions.

1. Select a rainfall generator $\boldsymbol{G}$, and a reference parameter set $\boldsymbol{\varphi}_r$.

2. Select a rainfall-runoff model $\boldsymbol{H}$, and a reference parameter set $\boldsymbol{\theta}_r$.

3. Select a stochastic process to represent residual errors $\boldsymbol{\mathcal{E}}$, and a reference parameter set $\boldsymbol{\vartheta}_r$.

4. Generate a realization of the rainfall, $\boldsymbol{x}^{(\imath)} \leftarrow \boldsymbol{G}(\boldsymbol{\varphi}_r)$ of length N.

5. Compute the synthetic "true" streamflow realization, $\boldsymbol{y}^{(\imath)} = \boldsymbol{H}(\boldsymbol{\theta}_r, \boldsymbol{x}^{(\imath)})$ of length N.

6. Generate a realization of residuals, $\boldsymbol{\varepsilon}^{(\imath)} \leftarrow \boldsymbol{\mathcal{E}}(\boldsymbol{\vartheta}_r)$ of length N.

7. Compute the synthetic "observed" streamflow realization, $\widetilde{\boldsymbol{y}}^{(\imath)} = \boldsymbol{y}^{(\imath)} + \boldsymbol{\varepsilon}^{(\imath)}$, of length N.

8. Select a likelihood formulation (BTL, BSL or MTL) and infer the maximum likelihood estimates $(\widehat{\boldsymbol{\theta}}^{(\imath)}, \widehat{\boldsymbol{\vartheta}}^{(\imath)})$ by maximizing the log-likelihood, $\log p(\widetilde{\boldsymbol{y}}^{(\imath)} \mid \boldsymbol{x}^{(\imath)}, \boldsymbol{\theta}, \boldsymbol{\vartheta})$ with respect to $(\boldsymbol{\theta}, \boldsymbol{\vartheta})$. Note that, similar to case study 1, this procedure corresponds to maximizing the Bayesian posterior under uniform prior assumptions.

9. Repeat steps 4-8 for $\imath = 1, ..., m$ to obtain an empirical distribution of the optimized rainfall-runoff model parameters $\widehat{\boldsymbol{\theta}}^{(\imath)}$ and residual model parameters $\widehat{\boldsymbol{\vartheta}}^{(\imath)}$. All presented results use $m = 400$ replicates.

The statistical reliability of the predictions obtained in this study is assessed using a predictive quantile-quantile (qq) plot constructed as follows: (i) compute the median of the reference simulations at time step t; (ii) estimate the quantiles of this median simulation within the predictions at time step t; (iii) estimate the frequency of these quantiles for selected bins; (iv) repeat steps (i) - (iii) but swapping the reference simulations and the predictions; (v) plot the frequencies against each other; this should plot on a 1:1 line. Note that the predictive qq-plot constructed using the approach above is a generalization

of the predictive qq-plot proposed by [Thyer et al., 2009], where only a single reference realization (the observed data) was used.

References

M. S. Bartlett. Periodogram analysis and continuous spectra. *Biometrika*, 37:1–16, 1950.

B. C. Bates and E. P. Campbell. A markov chain monte carlo scheme for parameter estimation and inference in conceptual rainfall-runoff modeling. *Water Resour. Res.*, 37(4):937–947, 2001.

K. J. Beven and A. Binley. The future of distributed models: model calibration and uncertainty prediction. *Hydrol. Proc.*, 6(3):279–298, 1992. doi: 10.1002/hyp.3360060305.

M. F. P. Bierkens and C. E. Puente. Analytically derived runoff models based on rainfall point processes. *Water Resour. Res.*, 26(11):2653–2659, 1990.

G. Box, G. M. Jenkins, and G. C. Reinsel. *Time series analysis : forecasting and control*. Prentice Hall, Englewood Clifs, NJ, 1994.

G. E. Box and G. Jenkins. *Time series analysis : forecasting and control*. Holden-Day, San Francisco, 1976.

G. E. P. Box and G. Tiao. *Bayesian inference in statistical analysis*. Wiley, 1992.

D. P. Boyle. *Multicriteria calibration of hydrological models*. PhD thesis, University of Arizona, 2000.

P. Brockwell and R. Davis. *Time Series: Theory and Methods*. Springer series in statistics. Springer, New York, 1987.

N. Choudhuri, S. Ghosal, and A. Roy. Bayesian estimation of the spectral density of a time series. *Journal of the American Statistical Association*, 99(468):1050–1059, 2004.

L. Cohen. Generalization of the wiener-khinchin theorem. *IEEE Signal Processing Letters*, 5(11):292–294, 1998.

F. Comola, B. Schaefli, A. Rinaldo, and M. Lehning. Thermodynamics in the hydrologic response: Travel time formulation and application to alpine catchments. *Water Resources Research*, 51(3):1671–1687, 2015. doi: 10.1002/2014wr016228.

N. De Vleeschouwer and V. R. N. Pauwels. Assessment of the indirect calibration of a rainfall-runoff model for ungauged catchments in flanders. *Hydrol. Earth Syst. Sci.*, 17 (5):2001–2016, 2013. doi: 10.5194/hess-17-2001-2013.

C. Duchon and R. Robert Hale. *Time Series Analysis in Meteorology and Climatology: An Introduction*. Wiley-Blackwell, Oxford, 2012.

K. Engeland, C. Y. Xu, and L. Gottschalk. Assessing uncertainties in a conceptual water balance model using bayesian methodology. *Hydrol. Sci. J.*, 50(1):45–63, 2005.

G. Evin, D. Kavetski, M. Thyer, and G. Kuczera. Pitfalls and improvements in the joint inference of heteroscedasticity and autocorrelation in hydrological model calibration. *Water Resour. Res.*, 49:4518–4524, 2013. doi: 10.1002/wrcr.20284.

G. Evin, M. Thyer, D. Kavetski, D. McInerney, and G. Kuczera. Comparison of joint versus postprocessor approaches for hydrological uncertainty estimation accounting for error autocorrelation and heteroscedasticity. *Water Resour. Res.*, 50:2350–2375, 2014. doi: 10.1002/2013wr014185.

R. Fox and M. S. Taqqu. Large-sample properties of parameter etimates for strongly dependent stationary gaussian time-series. *Ann. Stat.*, 14(2):517–532, 1986.

A. Hartmann, M. Weiler, T. Wagener, J. Lange, M. Kralik, F. Humer, N. Mizyed, A. Rimmer, J. A. Barber, B. Andreo, C. Butscher, and P. Huggenberger. Process-based karst

824 modelling to relate hydrodynamic and hydrochemical characteristics to system proper-
825 ties. *Hydrol. Earth Syst. Sci.*, 17(8):3305–3321, 2013. doi: 10.5194/hess-17-3305-2013.
826 M. Honti, C. Stamm, and P. Reichert. Integrated uncertainty assessment of discharge
827 predictions with a statistical error model. *Water Resour. Res.*, 49:4866–4884, 2013.
828 doi: 10.1002/wrcr.20374.
829 A. R. Ives, K. C. Abbott, and N. L. Ziebarth. Analysis of ecological time series with
830 arma(p,q) models. *Ecology*, 91(3):858–871. doi: 10.1890/09-0442.1.
831 G. Jenkins and D. G. Watts. *Spectral analysis and its applications*. Holden-Day Series in
832 Time Series Analysis. Holden-Day, San Francisco, 1968.
833 D. Kavetski, G. Kuczera, and S. W. Franks. Bayesian analysis of input uncertainty
834 in hydrological modeling: 1. theory. *Water Resour. Res.*, 42(3):W03407, 2006. doi:
835 10.1029/2005WR004368.
836 G. Kuczera. Improved parameter inference in catchment models. 1. evaluating parameter
837 uncertainty. *Water Resour. Res.*, 19(5):1151–1162, 1983.
838 G. Kuczera and E. Parent. Monte carlo assessment of parameter uncertainty in conceptual
839 catchment models: The metropolis algorithm. *J. Hydrol.*, 211(1-4):69–85, 1998. doi:
840 10.1016/S0022-1694(98)00198-X.
841 G. Kuczera, D. Kavetski, S. Franks, and M. Thyer. Towards a bayesian total er-
842 ror analysis of conceptual rainfall-runoff models: Characterising model error us-
843 ing storm-dependent parameters. *Journal of Hydrology*, 331(5):61–177, 2006. doi:
844 10.1029/WR019i005p01163.
845 J. Lagarias, J. Reeds, M. Wright, and P. Wright. Convergence properties of the nelder-
846 mead simplex method in low dimensions. *SIAM Journal of Optimization*, 9(1):112–147,

1998.

L. Li, C. Y. Xu, and K. Engeland. Development and comparison in uncertainty assessment based bayesian modularization method in hydrological modeling. *Journal of Hydrology*, 486:384–394, 2013. doi: 10.1016/j.jhydrol.2013.02.002.

D. McInerney, M. Thyer, D. Kavetski, J. Lerat, and G. Kuczera. Improving probabilistic prediction of daily streamflow by identifying pareto optimal approaches for modeling heteroscedastic residual errors. *Water Resources Research*, pages n/a–n/a, 2017. doi: 10.1002/2016WR019168.

H. McMillan and I. K. Westerberg. Rating curve estimation under epistemic uncertainty. *Hydrol. Process.*, 29:1873–1882, 2015. doi: 10.1002/hyp.10419.

H. McMillan, B. Jackson, M. Clark, D. Kavetski, and R. Woods. Rainfall uncertainty in hydrologic modelling: An evaluation of multiplicative error models. *Journal of Hydrology*, 400:83–94, 2011.

A. Montanari, R. Rosso,M. S. Taqqu. Fractionally differenced ARIMA models applied to hydrologic time series: Identification, estimation, and simulation. *Water Resour. Res.*, 33(5):1035–1044, 1997. doi: 10.1029/97WR00043.

A. Montanari and E. Toth. Calibration of hydrological models in the spectral domain: an opportunity for ungauged basins? *Water Resour. Res.*, 43(5):W05434, 2007. doi: 10.1029/2006WR005184.

F. Morlando, L. Cimorelli, L. Cozzolino, G. Mancini, D. Pianese, and F. Garofalo. Shot noise modeling of daily streamflows: A hybrid spectral- and time-domain calibration approach. *Water Resources Research*, page Early View, 2016. doi: 10.1002/2015wr017613.

F. Moussu, L. Oudin, V. Plagnes, A. Mangin, and H. Bendjoudi. A multi-objective calibration framework for rainfall-discharge models applied to karst systems. *J. Hydrol.*, 400(3-4):364–376, 2011. doi: DOI: 10.1016/j.jhydrol.2011.01.047.

A. Oppenheim and Schafer. *Discrete-Time Signal Processing, 1st ed.* Prentice-Hall, Upper Saddle River, NJ, 1989.

A. Oppenheim and G. Verghese. *Signals, Systems and Inference.* Prentice Hall,, 2015.

V. R. N. Pauwels and G. J. M. De Lannoy. Multivariate calibration of a water and energy balance model in the spectral domain. *Water Resour. Res.*, 47(7):W07523, 2011. doi: 10.1029/2010wr010292.

F. Pianosi and L. Raso. Dynamic modeling of predictive uncertainty by regression on absolute errors. *Water Resour. Res.*, 48(3):W03516, 2012. doi: 10.1029/2011wr010603.

W. Press, S. Teukolsky, W. Vetterling, and B. Flannery. Numerical recipes: The art of scientific computing (3rd ed.). Cambridge University Press, New York, 2007.

M. Priestley. *Spectral Analysis and Time Series.* Academic Press, London, 1981.

J. J. Quets, G. J. M. De Lannoy, and V. R. N. Pauwels. Comparison of spectral and time domain calibration methods for precipitation-discharge processes. *Hydrol. Proc.*, 24(8): 1048–1062, 2010. doi: 10.1002/hyp.7546.

K. R. Rao, D. N. Kim, J. J. Hwang. Fast Fourier transform: algorithms and applications. Springer, Dordrecht, 423 p., 2010.

B. Renard, D. Kavetski, M. Leblois, Etienneand Thyer, G. Kuczera, and S. W. Franks. Towards a reliable decomposition of predictive uncertainty in hydrologic modeling: Characterizing rainfall errors using conditional simulation. *Water Resour. Res.*, 47:W11516, 2011. doi: 10.1029/2011WR010643.

B. Schaefli and E. Zehe. Hydrological model performance and parameter estimation in the wavelet-domain. *Hydrology and Earth System Sciences*, 13:1921–1936, 2009. doi: 10.5194/hess-13-1921-2009.

B. Schaefli, D. Balin Talamba, and A. Musy. Quantifying hydrological modeling errors through a mixture of normal distributions. *J. Hydrol.*, 332(3-4):303–315, 2007. doi: 10.1016/j.jhydrol.2006.07.005.

G. Schoups and J. A. Vrugt. A formal likelihood function for parameter and predictive inference of hydrologic models with correlated, heteroscedastic, and non-gaussian errors. *Water Resour. Res.*, 46(10):W10531, 2010. doi: 10.1029/2009WR008933.

P. J. Smith, K. J. Beven, and J. A. Tawn. Detection of structural inadequacy in process-based hydrological models: A particle-filtering approach. *Water Resour. Res.*, 44(1): W01410, 2008.

T. Smith, A. Sharma, L. Marshall, R. Mehrotra, and S. Sisson. Development of a formal likelihood function for improved bayesian inference of ephemeral catchments. *Water Resour. Res.*, 46:W12551, 2010. doi: doi:10.1029/2010WR009514.

S. Sorooshian, Q. Duan, and V. K. Gupta. Calibration of rainfall-runoff models: Application of global optimization to the sacramento soil moisture accounting model. *Water Resour. Res.*, 29(4):1185–1194, 1993.

D. Brillinger. *Time series - Data analysis and theory, expanded edition.* Holden-Day Inc., San Francisco, 1981.

M. Thyer, B. Renard, D. Kavetski, G. Kuczera, S. Franks, and S. S. Srikanthan. Critical evaluation of parameter consistency and predictive uncertainty in hydrological modeling: A case study using bayesian total error analysis. *Water Resour. Res.*, 45:W00B14, 2009.

915 doi: 10.1029/2008WR006825.

916 A. Viglione, R. Merz, J. L. Salinas, and G. Blschl. Flood frequency hydrology:
917 3. a bayesian analysis. *Water Resources Research*, 49(2):675–692, 2013. doi:
918 10.1029/2011wr010782.

919 J. A. Vrugt, C. G. H. Diks, H. V. Gupta, W. Bouten, and J. M. Verstraten. Im-
920 proved treatment of uncertainty in hydrologic modeling: Combining the strengths of
921 global optimization and data assimilation. *Water Resour. Res.*, 41:W01017, 2005. doi:
922 10.1029/2004WR003059.

923 L. M. Ward and P. E. Greenwood. 1/f noise. *Scholarpedia*, 2(12):1537, 2007a.

924 L. M. Ward and P. E. Greenwood. 1/f noise, scholarpe-
925 dia, 2(12):1537. doi:10.4249/scholarpedia.1537, revision 90924,
926 *www.scholarpedia.org/article/1/f_noise* (accessed on 29 march 2017). 2007b.

927 P. D. Welch. Use of fast fourier transform for estimation of power spectra - a method
928 based on time averaging over short modified periodograms. *Ieee Transactions on Audio*
929 *and Electroacoustics*, AU15(2):70–73, 1967. doi: 10.1109/tau.1967.1161901.

930 B. J. West and M. Shlesinger. The noise in natural phenonema. *American Scientist*, 78
931 (1):40–45, 1990.

932 H. White. Consequences and detection of misspecified nonlinear regression models. *J.*
933 *Am. Stat. Assoc.*, 76:419433, 1981.

934 P. Whittle. Estimation and information in stationary time series. *Ark. Mat.*, 2:423434,
935 1953.

936 H. Winsemius, B. Schaefli, A. Montanari, and H. Savenije. On the calibration of hydrolog-
937 ical models in ungauged basins: a framework for integrating hard and soft hydrological

938 information. *Water Resour. Res.*, 45:W12422, 2009. doi: 10.1029/2009WR007706.

939 I. Zhurbenko. Periodogram, http://www.encyclopediaofmath.org/index.php/periodogram.

940 In M. Hazewinkel, editor, *This text originally appeared in: Encyclopedia of mathematics.*

941 Kluwer, Dordrecht, 1991.

Table 1. Mathematical notations used for the methods (for case studies see Table 1) We use capital letters for stochastic processes, lower case letters for process realizations (samples) or deterministic processes. Bounded domain continuous-time stochastic processes are written as $(Z_t)_{t=1,..,N}$, bounded domain stochastic discrete-time processes are written as $(Z_k)_{k=1,..,N}$.

Notation	Type	Meaning
$p()$	Function	Probability density
E[]	Function	Expectation
t	Variable	Time step of a discrete-time process
Δt	Parameter	Length of the discrete time step
ℓ	Variable	Time lag (autocorrelation function)
N	Parameter	Number of time steps
ω	Variable	Angular frequency index of continuous Fourier transform
ω_f	Parameter	Fundamental frequency, $= \frac{2\pi}{N\Delta t}$
ω_j	Variable	Frequency discrete-time Fourier transform ($j-th$ multiple of ω_f)
j	Variable	Frequency index of discrete Fourier transform
$\boldsymbol{H}$	Process	Hydrological process model
$\boldsymbol{X}$	Process	True hydrol. system input
$\widetilde{\boldsymbol{X}}$	Process	Observed system input
$\boldsymbol{Y}$	Process	True system output
$\widetilde{\boldsymbol{Y}}$	Process	Observed system output
$\boldsymbol{\mathcal{E}}$	Stoch. Process	Model residual process
$\boldsymbol{Z}$	Stoch. Process	General stochastic process
$\boldsymbol{z}$	Det. Process	General deterministic process
$\boldsymbol{Z}^{(N)}$	Stoch. Process	Discrete process Z in the bounded domain N
$\boldsymbol{\varepsilon}$	Det. Process	Error process realisation
δ	Det. Process	Realisation of error process innovations
ρ	Parameter	AR(1) autoregressive parameter
μ	Parameter	Error process mean
σ	Parameter	Error process variance
μ_δ	Parameter	Innovation process mean
σ_δ	Parameter	Innovation process variance
$\boldsymbol{y}$	Det. Process	Realisation of hydrol. system output
$\widetilde{\boldsymbol{Y}}$	Det. Process	Realisation of observed system output
$\boldsymbol{x}$	Det. Process	Realisation of rainfall input
$\widehat{\boldsymbol{y}}$	Det. Process	Realisation of hydrol. process model (simulation)
$\mathbb{F}$	Variable	Fourier transform (discrete- and continuous-time)
$\mathbb{P}$	Variable	Power density spectrum (PDS)
$\mathbb{Q}$	Variable	Stochastic process having PDS as expected value (PDSV)
P	Variable	Sample of the PDSV, Periodogram
$\mathbb{P}_j$	Variable	Mean of the periodogram

Table 2. Table 1 (continued)

Notation	Type	Meaning
$\zeta_j^{\boldsymbol{Z}}$	Variable	Profile function (Equation 21)
$\zeta_j^{\boldsymbol{\varepsilon}}$	Variable	Spectral profile function of residuals
f_{FFT}	Operation	Fast Fourier transform (FFT) operation
$\boldsymbol{\theta}$	Parameter	Parameter vector of hydrol. process model
$\boldsymbol{\vartheta}$	Parameter	Error model parameters
$\boldsymbol{\vartheta}_\zeta$	Parameter	Autocorrelation parameters of error model
$\mu_{\boldsymbol{\varepsilon}}$	Parameter	Mean of error process
f_j	Function	Probability density function at frequency j
$f_{\chi^2_1}$	Function	χ^2 probability density function
f_{exp}	Function	Exponential probability density function
q	Function	Transformation function, e.g. log
J_q	Function	Jacobian of the transformation q
υ_ℓ	Function	Autocovariance function at lag ℓ

Table 3. Summary of all the used case studies.

Name	Type	Purpose
Case study 1.1	synthetic	uncorrelated process with innovations from a uniform distribution in $[-1, 1]$
Case study 1.2	synthetic	as 1.1 but Laplace distribution with $\mu = 0$, $\sigma = 1$
Case study 1.3	synthetic	as 1.1 but bimodal Gaussian distrib.: $\mu_1 = 1$, $\mu_2 = -1$, $\sigma_1 = \sigma_2 = 0.5$, weight of 1st component: 0.2
Case study 1.4	synthetic	AR(1) process with parameters of Gaussian error model of Table 4
Case study 2.1	synthetic	pure random process with $f_{\mathbb{P}1}(\omega; (A_1, B_1)) = A_1 \exp(-B_1\omega)$, $A_1 = 100, B_1 = 0.001$
Case study 2.2	synthetic	pure random process with $f_{\mathbb{P}2}(\omega; (A_2, B_2)) = A_2/\omega^{B_2}$, $A_2 = 10.5, B_2 = 3$
Case study 3.1	synthetic	simple hydrologic model + Gaussian AR(1) error process (Table 4)
Case study 3.2	synthetic	as 3.1 but Laplacian AR(1) error process
Case study 3.3	synthetic	as 3.1 but error process with autocorrel. structure $v_\ell = f(\ell; \varrho_1, \varrho_2) = \exp(-\varrho_1\ell - \varrho_2\ell^{0.5})$
Case study 4	real data	HYMOD model for Leave river + AR(1) Gaussian error process in log-transformed space

Table 4. Mathematical notations used for the case studies (for methods see Table 1), including the reference values for the model parameters.

Notation	Type	Meaning	Reference value
m	Parameter	Number of experiment repetitions	
$\boldsymbol{G}$	Process	Rainfall generator model	
$\boldsymbol{\varphi}$	Parameter	Parameter vector of rainfall generator	
λ	Parameter	Arrival rate (rainfall generator)	0.5 d^{-1}
$\bar{i}_r$	Parameter	Rainfall event mean intensity (rainfall generator) [L/T]	3.3 mm/d
$\bar{t}_r$	Parameter	Rainfall event mean duration (rainfall generator) [T]	0.8 d
ϱ	Parameter	Autocorrelation parameters (synthetic rainfall-runoff)	
ℓ	Variable	Lag (synthetic rainfall-runoff)	
k_1	Parameter	Linear reservoir coefficient 1 [1/T] (synthetic rainfall-runoff)	0.1 d^{-1}
k_2	Parameter	Linear reservoir coefficient 2 [1/T] (synthetic rainfall-runoff)	0.05 d^{-1}
l_g	Parameter	Leaching parameter [L/T] (synthetic rainfall-runoff)	$\lambda\bar{i}_r\bar{t}_r$ mm/d
$\rho_{\boldsymbol{\varepsilon}}$	Parameter	AR1 parameter (error model)	0.8
μ_δ	Parameter	Mean of innovations (error model)	0.25
σ_δ	Parameter	Standard deviation of innov. (error model)	0.015
$s_{\max}$	Parameter	Max. storage (HYMOD)[L]	
b_{H}	Parameter	Spatial variability (HYMOD)[-]	
α	Parameter	Flow splitting (HYMOD)[-]	
ν_s	Parameter	Residence time slow reservoir (HYMOD) [T]	
ν_q	Parameter	Residence time fast reservoirs (HYMOD) [T]	
m_H	Parameter	Number of fast reservoirs (HYMOD) [-]	
A	Parameter	Numerical offset for log computation	

Table 5. Leaf River case study: limits of the uniform priors, parameter values inferred using BTL and BSL (columns denoted with *inf*), and empirical quantities estimated from the computed residuals (columns denoted with *emp*). The number of fast reservoirs, m_H, is fixed to 2 after initial optimization. The prior for σ_δ is obtained according to the method used in [Schaefli et al., 2007]. NSE stands for the Nash-Sutcliffe efficiency.

	$s_{\max}$ mm	b_{H} -	α -	ν_s -	m_H -	ρ inf -	μ_δ inf mm/d	σ_δ inf mm/d	ρ emp -	μ_δ emp mm/d	σ_δ emp mm/d	NSE -
Prior min	50	0.05	0.01	0.001	0.001	0	-0.25	0	-	-		-
Prior max	800	1.95	1	0.20	0.95	0.99	0.25	-	-	-		-
BTL-Gauss	149	0.51	0.11	0.16	0.01	0.90	-0.02	0.19	0.71	-0.05	2.08	0.70
BSL	174	0.38	0.10	0.15	0.01	0.89	0.01	0.20	0.73	-0.02	2.10	0.68
MTL	173	0.34	0.08	0.08	0.84	0.97	-0.01	0.21	0.80	-0.02	2.13	0.59