Water Resources Research

Key Points:

- Stochastic, time-dependent parameters consider intrinsic model uncertainty and propagate this uncertainty to the model output
- Cross-validation sensitively identifies model structure deficits and the time-course of the identified parameters gives hints for model improvements
- Posterior uncertainty naturally gets autocorrelated and reflects the difference in knowledge between calibration and validation or extrapolation periods

Correspondence to:
P. Reichert,
peter.reichert@eawag.ch

Potential and Challenges of Investigating Intrinsic Uncertainty of Hydrological Models With Stochastic, Time-Dependent Parameters

Peter Reichert[1] , **Lorenz Ammann[1]** , and **Fabrizio Fenicia[1]**

[1]Eawag: Swiss Federal Institute of Aquatic Science and Technology, Dübendorf, Switzerland

Abstract Stochastic hydrological process models have two conceptual advantages over deterministic models. First, even though water flow in a well-defined environment is governed by deterministic differential equations, a hydrological system, at the level we can observe it, does not behave deterministically. Reasons for this behavior are unobserved spatial heterogeneity and fluctuations of input, unobserved influence factors, heterogeneity and variability in soil and aquifer properties, and an imprecisely known initial state. A stochastic model provides thus a more realistic description of the system than a deterministic model. Second, hydrological models simplify real processes. The resulting structural deficits can better be accounted for by stochastic than by deterministic models because they, even for given parameters and input, allow for a probability distribution of different system evolutions rather than a single trajectory. On the other hand, stochastic process models are more susceptible to identifiability problems and Bayesian inference is computationally much more demanding. In this paper, we review the use of stochastic, time-dependent parameters to make deterministic models stochastic, discuss options for the numerical implementation of Bayesian inference, and investigate the potential and challenges of this approach with a case study. We demonstrate how model deficits can be identified and reduced, and how the suggested approach leads to a more realistic description of the uncertainty of internal and output variables of the model compared to a deterministic model. In addition, multiple stochastic parameters with different correlation times could explain the variability in the time scale of output error fluctuations identified in an earlier study.

1. Introduction

Traditionally, the uncertainty in observed streamflow has mostly been described by combining a deterministic model of hydrological processes with a stochastic error model for the observations. Such error models are often formulated as additive terms to the output of the deterministic model, potentially at a transformed scale to account for heteroscedasticity in the errors while keeping the error term at the transformed scale homoscedastic (Bates & Campbell, 2001; Kuczera, 1983). This approach suffers from the conceptual limitation that it tends to underestimate the uncertainty in the hydrological processes (only considered by uncertain parameters of a quite rigid, deterministic model structure and, potentially, by uncertainty in input) and compensates for this by mapping the effect of all remaining uncertainties to a lumped output error model. The lumped output error results from the propagation of input errors, errors in the initial state (these first two contributions only if they are not addressed explicitly), intrinsic, apparent stochasticity, model structural error, and output observation errors. Such a model has to consider autocorrelation (resulting from memory-effects of errors in the hydrological states), heteroscedasticity (as larger values of outputs can also be expected to have larger uncertainty), positivity of the resulting uncertain discharge, and non-normality (as distributions of positive values are skewed; the larger the relative uncertainty, the larger the skewness) (Bates & Campbell, 2001; Kuczera, 1983; Sorooshian & Dracup, 1980). As the mechanisms leading to the lumped error are not described explicitly, this lumped output error model needs an empirical parameterization. A lot of empirical evidence demonstrates that with this approach it is difficult to get model calibration and quantification of prediction uncertainty that seem appropriate to hydrologists (Ammann et al., 2019; Evin et al., 2013, 2014; Reichert & Schuwirth, 2012; Schoups & Vrugt, 2010).

Uncertainty analysis for deterministic models can be improved by considering input and model structure uncertainty explicitly. In hydrology, input uncertainty has mostly been addressed by modifying

rain input with storm-dependent multipliers (Kavetski et al., 2003, 2006a, 2006b; Kuczera, 1990; Vrugt et al., 2008), but continuously varying descriptions of input uncertainty have also been used (Del Giudice et al., 2016; Reichert & Mieleitner, 2009). Model structure uncertainty has been considered by multi-model approaches, either in the form of ensemble predictions based on Bayesian model averaging (Duan et al., 2007; Liang et al., 2013), or by a careful model selection procedure (Fenicia et al., 2016). Conceptual hydrological models can ideally distinguish different time scales of hydrological processes and associated storages, but they cannot directly provide their physical interpretation for a heterogeneous catchment. The comparison of the performance of multiple model structures can support the understanding of the underlying system (Fenicia et al., 2016). Nevertheless, despite the usefulness of this approach for hydrological systems analysis, we should be aware of the problem that multimodel approaches with deterministic models still suffer from the problem of rigid model structures defined by the deterministic process models.

There are two main reasons, why a stochastic process model provides a better description of the hydrological system than a deterministic model. First, the same observed input affecting the same observed initial state of a hydrological system will not lead to the same output, because (i) observations are incomplete regarding influence factors and their spatial and temporal resolution, (ii) the knowledge about the initial state of the hydrological system is incomplete, and (iii) changes in catchment properties caused by vegetation growth, structural changes in soil structure due to drying and wetting processes, bioturbation, etc. affect the response of the catchment. Thus, despite the deterministic nature of water flow in a given physical environment, at the resolution we can observe it, the hydrological system is not deterministic (Blöschl & Sivaplalan, 1995; Kuczera et al., 2006). This can only partly be accounted for by explicitly considering uncertainty in input and in the initial state of a deterministic process model. Second, hydrological models need simplification of real processes. Such structural deficits can better be accounted for by stochastic than by deterministic process models because they, even for given parameters and input, allow for a probability distribution of different system evolutions rather than a single system trajectory. Still, multimodel approaches and the explicit consideration of uncertainty in input and in the initial state remain important, also for stochastic models.

To account for intrinsic stochasticity of the hydrological system (in the sense defined above) and for model structural errors, the hydrological process model has to be made stochastic. Stochastic hydrological modeling has a long history: Originally, stochastic hydrological models were primarily introduced to describe statistical features observed in discharge time series, such as extreme events or long periods of low or high flow (Koutsoyiannis, 2002; Mandelbrot & Wallis, 1968). More recently, stochastic hydrological models were primarily used, often even without explicitly mentioning the need for stochasticity, in the context of sequential "data assimilation," where stochasticity in the mass balance equations of hydrological reservoirs makes it possible to iteratively learn about reservoir levels from discharge observation time series (Clark et al., 2008; Y. Liu et al., 2012; Moradkhani et al., 2005; Vrugt et al., 2013, and many more). Multiple reasons have been mentioned to make parameters of hydrological models stochastic and thus time-dependent: The inadequacy of modeling runoff with deterministic models (Kuczera et al., 2006, "the notion of a deterministic conceptual rainfall-runoff model is indefensible"), the "effective" nature of hydrological model parameters that makes them more susceptible to changes than universal, physical constants (Y. Liu & Gupta, 2007, "one might wish more generally to permit the system characteristics represented as 'parameters' to vary slowly with time"), the presence of model structure deficits (Leisenring & Moradkhani, 2010, "to account for imperfect model representation of the physical processes"), and opportunities for the identification of model structure deficits (Beck & Young, 1976; Wagener et al., 2003). Making the parameters stochastic can be an additional element of stochasticity to stochastic mass balance equations (Y. Liu & Gupta, 2007; J. Liu & West, 2001; Suweis et al., 2010), or it can be made the exclusive way of introducing stochasticity (Reichert & Mieleitner, 2009). The latter approach is conceptually more satisfying as it preserves the mass balance equations by making the fluxes between the reservoirs stochastic rather than the water levels in the reservoirs (resulting changes in the water levels of reservoirs occur then due to changes in fluxes between reservoirs that preserve mass balances). In addition, this approach allows us to directly learn about potential improvements of the mechanistic part of the model, by analyzing the dependence of inferred time-dependent parameters on states or external influence factors (Reichert & Mieleitner, 2009).

The additional degrees of freedom resulting from stochastic, time-dependent parameters can lead to identifiability problems or even to "misuse" of the time-dependence by partly explaining observed responses to input by variations of parameters. This can make the results sensitive to model assumptions and priors (Renard et al., 2010). Depending on the involved correlation time scales and the observation time interval, there can also be an identifiability problem between remaining uncorrelated, random errors and stochastic processes to be identified. This requires very careful analysis of the results. Another disadvantage of stochastic compared to deterministic process models is the much higher algorithmic and computational requirements needed for Bayesian inference. These problems may be the reasons that still the majority of hydrological modeling studies are based on deterministic process models (plus a stochastic, lumped error term).

The increasing availability of statistical inference techniques for stochastic models and the increasing computational power becoming available make it possible to gain experience with such models. It is the goal of this paper to demonstrate the feasibility, the potential and also the challenges of this approach. In particular, we have the following objectives:

- Clarify the concept of stochastic process models based on stochastic, time-dependent parameters and review numerical techniques to do Bayesian inference for these models
- Develop a methodology to deal with the challenges of the approach, in particular with potential overparameterization and identifiability problems
- Demonstrate how time-dependent parameters can be used as a diagnostic approach to identify model structure or parameterization deficits and guide their resolution
- Illustrate how stochastic, time-dependent parameters lead to a realistic description of predictive uncertainty once severe model deficits have been substantially reduced
- Explore the potential of a simple multimodel approach for stochastic hydrological models
- Investigate whether multiple time-dependent parameters with different correlation times may provide an explanation for multiple time scales found in lumped error models (Ammann et al., 2019)

In the following section, we will establish a connection between stochastic and deterministic process models, introduce the concept of stochastic, time-dependent parameters, establish an analysis procedure to avoid "misuse" of the stochastic parameters to model functional relationships, discuss numerical implementation schemes for Bayesian inference with stochastic parameters, and conclude with discussing the potential and challenges of the suggested approach. In the subsequent sections, we will describe our case study and its results, discuss methodological aspects beyond the case study, and draw our conclusions.

2. Methods

2.1. Connection Between Stochastic and Deterministic Process Models

A stochastic model evaluated for a finite number of observed variables at discrete locations in space and time can be written as a vector of random functions:

$$\mathbf{Y}_{\mathrm{obs}}(\mathbf{x},\theta,\psi) \quad , \tag{1}$$

where, $\mathbf{Y}_{\mathrm{obs}}$ is a vector of random variables representing the observations, $\mathbf{x}$ are the model inputs, θ are the parameters of the model of the investigated system, and ψ are the parameters of the observation (sub) model. In many cases it makes sense to split the system and observation models explicitly:

$$\mathbf{Y}_{\mathrm{obs}}(\mathbf{x},\theta,\psi) = \mathbf{Y}_{\mathrm{out}}(\mathbf{x},\theta) + \mathbf{E}_{\mathrm{obs}}\big(\mathbf{Y}_{\mathrm{out}}(\mathbf{x},\theta),\psi\big) \quad , \tag{2}$$

where, $\mathbf{Y}_{\mathbf{out}}$ is the model of the true output and $\mathbf{E}_{\mathrm{obs}}$ is the observation error model. Note that Equation 2 describes the probability distribution of observations conditional on input and parameters. This does not mean that input and parameter uncertainty cannot be considered. Known uncertainty can be propagated through the model. If input and/or parameter uncertainty have to be estimated, we need a prior for their distribution and an observation model for input. In our notation, this changes "true" input into parameters and observed input into additional observations.
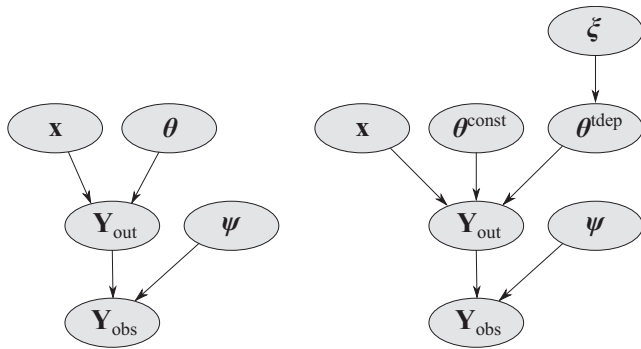
Figure 1. Illustration of the conceptual approach of making some of the model parameters stochastic. Left panel: model with constant parameters. Right panel: model with some of the parameters θ replaced by stochastic, time-dependent parameters with hyperparameters ξ (e.g., their means, standard deviations, and autocorrelation times).

The response of any stochastic model can be splitted into its expected value plus a total error term equal to the response minus the expected value

$$
\begin{aligned}
\mathbf{Y}_{\mathrm{obs}}(\mathbf{x},\theta,\psi) &= \underbrace{\mathrm{E}[\mathbf{Y}_{\mathrm{obs}}(\mathbf{x},\theta,\psi)]}_{\mathbf{y}_{\mathrm{det}}(\mathbf{x},\theta,\psi)} + \underbrace{\left(\mathbf{Y}_{\mathrm{obs}}(\mathbf{x},\theta,\psi) - \mathrm{E}[\mathbf{Y}_{\mathrm{obs}}(\mathbf{x},\theta,\psi)]\right)}_{\mathbf{E}_{\mathrm{tot}}(\mathbf{x},\theta,\psi)} \\
&= \underbrace{\mathrm{E}[\mathbf{Y}_{\mathrm{out}}(\mathbf{x},\theta)] + \mathrm{E}\left[\mathbf{E}_{\mathrm{obs}}\left(\mathbf{Y}_{\mathrm{out}}(\mathbf{x},\theta),\psi\right)\right]}_{\mathbf{y}_{\mathrm{det}}(\mathbf{x},\theta,\psi)} \\
&\quad + \underbrace{\left(\begin{array}{l}\mathbf{Y}_{\mathrm{out}}(\mathbf{x},\theta) + \mathbf{E}_{\mathrm{obs}}\left(\mathbf{Y}_{\mathrm{out}}(\mathbf{x},\theta),\psi\right) - \\ \mathrm{E}[\mathbf{Y}_{\mathrm{out}}(\mathbf{x},\theta)] - \mathrm{E}\left[\mathbf{E}_{\mathrm{obs}}\left(\mathbf{Y}_{\mathrm{out}}(\mathbf{x},\theta),\psi\right)\right]\end{array}\right)}_{\mathbf{E}_{\mathrm{tot}}(\mathbf{x},\theta,\psi)} \\
&= \mathbf{y}_{\mathrm{det}}(\mathbf{x},\theta,\psi) + \mathbf{E}_{\mathrm{tot}}(\mathbf{x},\theta,\psi).
\end{aligned}
\tag{3}
$$

In this equation, $\mathbf{y}_{\mathrm{det}}$ (x, θ, ψ) is the deterministic model defined as the expected value of the stochastic model including the observation model, and $\mathbf{E}_{tot}$ (x, θ, ψ) is the residual, lumped error model. While Equation 3 demonstrates that the use of a deterministic hydrological model plus a residual error term is not conceptually wrong, it also demonstrates why this approach is so difficult to implement: (i) The deterministic model is the expectation of a stochastic model. When formulating a deterministic model without deriving it from a stochastic model, it can only be an empirical parameterization of this expectation (note that the expectation of a nonlinear deterministic model that has been made stochastic by making parameters stochastic processes or by adding a noise term to the time evolution equations is not equal to the original deterministic model). As most models are anyway parameterized empirically, from a practical point of view, this may not be a very severe restriction. Still, the modeler should be aware that the parameter values reflect the empirical description of mean behavior and not of a single model trajectory. (ii) The derivation of the residual, lumped error in Equation 3 clarifies that the full output of the stochastic model is needed for its definition. The presence of apparent intrinsic stochasticity in a hydrological system, as outlined above, together with the propagation of this stochasticity through internal state variables to the output, implies that output errors will be autocorrelated due to the memory effect caused by the internal states. If this stochasticity and its propagation are not modeled explicitly by a stochastic model, the residual error term must be parameterized empirically. Experience has shown, that it is very difficult to formulate such an empirical term in a way that leads to satisfying results (Ammann et al., 2019; Evin et al., 2013, 2014; Reichert & Schuwirth, 2012; Schoups & Vrugt, 2010). The complex structure of the lumped error term, $\mathbf{E}_{tot}$ (x, θ, ψ), in Equation 3 clarifies why this is a very plausible outcome.

2.2. Concept of Stochastic, Time-Dependent Parameters

The idea of using stochastic, time-dependent parameters for considering intrinsic stochasticity of a system and making the model less susceptible to structural errors is very old. Originally, inference was implemented for discrete-time models by applying an Extended Kalman Filter (EKF) algorithm (Beck, 1987; Beck & Young, 1976). The concept was later on also occasionally mentioned in the hydrological literature on "data assimilation," mostly in addition to, not instead of, stochasticity in the mass balance equations (Kuczera et al., 2006; Leisenring & Moradkhani, 2010; Y. Liu & Gupta, 2007; J. Liu & West, 2001; Moradkhani et al., 2005). The idea of using stochastic parameters was later on generalized to make it applicable to nonlinear, continuous-time models without the need for linearization (Buser, 2003; Tomassini et al., 2009) and further extended to a systematic procedure of model identification and uncertainty analysis (Reichert & Mieleitner, 2009). Figure 1 illustrates the approach of making parameters stochastic processes. The parameter vector, θ, is split into constant and time-dependent parameters, $\theta = (\theta^{const}, \theta^{tdep})$. The probability distribution of the constant parameters, θ^{const}, and of the hyperparameters, ξ, of the stochastic process that underlies the time-dependent parameters describes the uncertain knowledge about these parameter values, whereas the stochastic processes of the time-dependent parameters, θ^{tdep}, describe intrinsic randomness of the system described by the model. By choosing parameters to be stochastic and time-dependent, we can make the model a more realistic description of the real system as we can consider effects of unobserved spatial variation and fluctuations in observed and unobserved influence factors.

The switch from the model shown in the left panel of Figure 1 to that in the right panel is associated with the switch from the joint probability

$$
p(\mathbf{y}_{\mathrm{obs}},\mathbf{y}_{\mathrm{out}},\boldsymbol{\psi},\boldsymbol{\theta} \mid \mathbf{x}) = \underbrace{p(\mathbf{y}_{\mathrm{obs}} \mid \mathbf{y}_{\mathrm{out}},\boldsymbol{\psi})}_{\text{obs. likelihood}} \cdot \underbrace{p(\mathbf{y}_{\mathrm{out}} \mid \mathbf{x},\boldsymbol{\theta})}_{\text{process model}} \cdot \underbrace{p(\boldsymbol{\psi},\boldsymbol{\theta})}_{\text{prior of par.}}
\tag{4}
$$

to the joint probability with an additional hierarchical level for the time-dependent parameters

$$
\begin{aligned}
& p(\mathbf{y}_{\mathrm{obs}},\mathbf{y}_{\mathrm{out}},\boldsymbol{\psi},\boldsymbol{\theta}^{\mathrm{const}},\boldsymbol{\theta}^{\mathrm{tdep}},\boldsymbol{\xi} \mid \mathbf{x}) \\
& = \underbrace{p(\mathbf{y}_{\mathrm{obs}} \mid \mathbf{y}_{\mathrm{out}},\boldsymbol{\psi})}_{\text{obs. likelihood}} \cdot \underbrace{p\Big(\mathbf{y}_{\mathrm{out}} \mid \mathbf{x},(\boldsymbol{\theta}^{\mathrm{const}},\boldsymbol{\theta}^{\mathrm{tdep}})\Big)}_{\text{process model}} \cdot \underbrace{p(\boldsymbol{\theta}^{\mathrm{tdep}} \mid \boldsymbol{\xi})}_{\text{stochastic par.}} \cdot \underbrace{p(\boldsymbol{\psi},\boldsymbol{\theta}^{\mathrm{const}},\boldsymbol{\xi})}_{\text{prior of const. par.}}.
\end{aligned}
\tag{5}
$$

In these equations, the observational likelihood does not change formally. However, in Equation 4, depending on the structure of the process model, it may have to consider autocorrelation as it may represent a lumped error that includes the effect of intrinsic stochasticity and model structure errors propagated to the output. On the other hand, the process model does not change except that it has to consider the time-dependence of some of its parameters. Note that in the Equations 4 and 5 the process model, $p(\mathbf{y}_{out}|\mathbf{x}, \theta)$, is assumed to be stochastic (we get a probability distribution of outputs even for given inputs and parameters). For a deterministic process model, $\mathbf{Y}_{\mathrm{out}}(\mathbf{x}, \theta)$, these equations simplify to

$$
p(\mathbf{y}_{\mathrm{obs}},\boldsymbol{\psi},\boldsymbol{\theta} \mid \mathbf{x}) = \underbrace{p\Big(\mathbf{y}_{\mathrm{obs}} \mid \underbrace{\mathbf{y}_{\mathrm{out}}(\mathbf{x},\boldsymbol{\theta})}_{\text{proc. mod.}},\boldsymbol{\psi}\Big) \cdot \underbrace{p(\boldsymbol{\psi},\boldsymbol{\theta})}_{\text{prior of par.}}}_{\text{obs. likelihood}\,(\text{incl. proc. mod.})}
\tag{6}
$$

and

$$
\begin{aligned}
& p(\mathbf{y}_{\mathrm{obs}},\boldsymbol{\psi},\boldsymbol{\theta}^{\mathrm{const}},\boldsymbol{\theta}^{\mathrm{tdep}},\boldsymbol{\xi} \mid \mathbf{x}) \\
& = \underbrace{p\Big(\mathbf{y}_{\mathrm{obs}} \mid \underbrace{\mathbf{y}_{\mathrm{out}}\Big(\mathbf{x},(\boldsymbol{\theta}^{\mathrm{const}},\boldsymbol{\theta}^{\mathrm{tdep}})\Big)}_{\text{process model}},\boldsymbol{\psi}\Big)}_{\text{obs. likelihood}\,(\text{incl. process model})} \cdot \underbrace{p(\boldsymbol{\theta}^{\mathrm{tdep}} \mid \boldsymbol{\xi})}_{\text{stochastic par.}} \cdot \underbrace{p(\boldsymbol{\psi},\boldsymbol{\theta}^{\mathrm{const}},\boldsymbol{\xi})}_{\text{prior of const. par.}}.
\end{aligned}
\tag{7}
$$

Again, the observational likelihood can still be the same across all models (Equations 4–7), but the need for an autocorrelated, lumped error model is even higher in Equation 6 than in Equation 4, as a deterministic process model often has serious structural errors.

Note that, although the function $\mathbf{y}_{\mathrm{out}}$ remains deterministic in Equation 7, the stochastic, time-dependent parameters, θ^{tdep}, make the deterministic model stochastic when seen as a function of the hyperparameters $\boldsymbol{\xi}$ rather than of realized time series of time-dependent parameters θ^{tdep} (the dependence on inputs, $\mathbf{x}$, and constant parameters, θ^{const}, remains the same). The probability density of $(\mathbf{Y}_{\mathrm{out}}|\mathbf{x}, \theta^{\mathrm{const}}, \boldsymbol{\xi})$ is obtained by propagating the distribution $p(\theta^{tdep} \mid \boldsymbol{\xi})$ through the deterministic model $\mathbf{y}_{\mathrm{out}}\Big(\mathbf{x},(\boldsymbol{\theta}^{\mathrm{const}},\boldsymbol{\theta}^{\mathrm{tdep}})\Big)$. This leads to the formal expression

$$
f_{\mathbf{Y}_{\mathrm{out}}|\mathbf{x},\boldsymbol{\theta}^{\mathrm{const}},\boldsymbol{\xi}}(\mathbf{y}) = \int_{\left\{\boldsymbol{\theta}^{\mathrm{tdep}} \mid \mathbf{y}_{\mathrm{out}}\Big(\mathbf{x},(\boldsymbol{\theta}^{\mathrm{const}},\boldsymbol{\theta}^{\mathrm{tdep}})\Big)=\mathbf{y}, \det(JJ^{\mathrm{T}})\neq 0\right\}} \frac{f(\boldsymbol{\theta}^{\mathrm{tdep}} \mid \boldsymbol{\xi})}{\sqrt{\det(JJ^{\mathrm{T}})}} \, \mathrm{d}\sigma_{\mathbf{y}}(\boldsymbol{\theta}^{\mathrm{tdep}}),
\tag{8}
$$

where

$$
J_{\mathbf{y}_{\mathrm{out}}}\Big(\mathbf{x},(\boldsymbol{\theta}^{\mathrm{const}},\boldsymbol{\theta}^{\mathrm{tdep}})\Big) = \frac{\partial \mathbf{y}_{\mathrm{out}}}{\partial(\boldsymbol{\theta}^{\mathrm{tdep}})^{\mathrm{T}}}\Big(\mathbf{x},(\boldsymbol{\theta}^{\mathrm{const}},\boldsymbol{\theta}^{\mathrm{tdep}})\Big)
\tag{9}
$$

is the Jacobian of $\mathbf{Y}_{\mathrm{out}}$ with respect to θ^{tdep},

$$
JJ^{\mathrm{T}} = J_{\mathbf{y}_{\mathrm{out}}}\Big(\mathbf{x},(\boldsymbol{\theta}^{\mathrm{const}},\boldsymbol{\theta}^{\mathrm{tdep}})\Big) \cdot J_{\mathbf{y}_{\mathrm{out}}}^{\mathrm{T}}\Big(\mathbf{x},(\boldsymbol{\theta}^{\mathrm{const}},\boldsymbol{\theta}^{\mathrm{tdep}})\Big)
\tag{10}
$$

and $\sigma_y(\theta^{tdep})$ is the Hausdorff measure on the surface $\{\theta^{\text{tdep}} \mid \mathbf{y}_{\text{out}}\left(\mathbf{x},\left(\theta^{\text{const}},\theta^{\text{tdep}}\right)\right) = \mathbf{y}\}$ (Stroock, 1999, Equation 5.3.28). Note that we will approximate the stochastic parameters, θ^{tdep}, by their values on a fine grid in time, and use linear interpolation between these values in the function $\mathbf{y}_{\text{out}}$, so that the integral (Equation 8) becomes a well-defined, although very high dimensional integral. Equation 8 is very complicated because analytically propagating a probability distribution through a deterministic model is difficult as different combinations of input may lead to the same result (see the complicated domain of integration below the integral sign in Equation 8) and the transformation of volume further modifies the probability density (see the denominator in the integrand of Equation 8). Note, however, that whenever distributions are represented numerically by samples, propagation just reduces to propagating all sample points through the model to get the sample that approximates the distribution of the result of Equation 8. This means that we will hardly ever have to apply Equation 8 explicitly.

Equation 8 is the basis for prediction of our knowledge of true output, $\mathbf{y}_{\text{out}}$, with prior or posterior distributions of the parameters. In case of prediction for a time period immediately following calibration and if the stochastic processes θ^{tdep} are Markov processes, conditioning the predictive distribution of θ^{tdep} on the posterior distribution consists only of using the posterior distribution of θ^{tdep} at the final time point of the calibration period as the distribution of θ^{tdep} to start prediction of the stochastic parameters with the posterior distribution of the process parameters ξ. Prediction of model outputs then consists of propagating this posterior distribution of θ^{tdep} together with the posterior distribution of the constant parameters θ^{const} through the model. The observational error, $\mathbf{E}_{\text{obs}}$, can still be added to get predicted observations, $\mathbf{Y}_{\text{obs}}$.

In the following, we choose the time-dependent parameters or adequately transformed parameters (e.g., their logarithms for parameters bounded to positive values) to follow Ornstein-Uhlenbeck processes (Uhlenbeck & Ornstein, 1930). The Ornstein-Uhlenbeck process is the simplest continuous stochastic process with finite variance and can be characterized by its mean, μ, its asymptotic standard deviation, σ, and its rate, γ, or correlation time, $\tau = 1/\gamma$. It is defined by the combination of a random walk and a drift back to the mean with rate γ. This concept is realized by the stochastic differential equation

$$\mathrm{d}\theta(t) = -\gamma\left(\theta(t) - \mu\right) + \sqrt{2\gamma}\,\sigma\,\mathrm{d}W(t) \quad , \tag{11}$$

where $W(t)$ is a Wiener process (continuous-time random walk). The solution to this differential equation at time t, given the solution at time $s \leq t$, is given by

$$\theta(t) \mid \theta(s) \sim \mathrm{N}\left(\mu + \left(\theta(s) - \mu\right)\exp\left(-\gamma(t - s)\right), \sigma\sqrt{1 - \exp\left(-2\gamma(t - s)\right)}\right) \quad , \tag{12}$$

where N is the Normal distribution with mean and standard deviation as its arguments. Equation 12 shows the drift from the solution at time s towards the mean with an increasing standard deviation that asymptotically approaches σ. This process is useful for characterizing fluctuating parameters as its standard deviation is bounded.

When formulating discretized Ornstein-Uhlenbeck processes on a fine grid of time values, $\{t_j\}_{j=1}^n$, we get the following probability density of the stochastic parameter i:

$$\begin{aligned}
f_{\theta_i^{\text{tdep}}\mid\xi}\left(\theta_i^{\text{tdep}}(t_1,\ldots,t_n) \mid \xi\right) &= f_{\mathrm{N}(\mu_i,\sigma_i)}\left(\theta_i^{\text{tdep}}(t_1)\right) \cdot \prod_{j=2}^n f_{\mathrm{N}(\mu_{i,j},\sigma_{i,j})}\left(\theta_i^{\text{tdep}}(t_j)\right) \\
&\text{with } \mu_{i,j} = \mu_i + \left(\theta_i^{\text{tdep}}(t_{j-1}) - \mu_i\right)\exp\left(-\gamma_i(t_j - t_{j-1})\right) , \\
&\quad\quad \sigma_{i,j} = \sigma_i\sqrt{1 - \exp\left(-2\gamma_i(t_j - t_{j-1})\right)} .
\end{aligned} \tag{13}$$

here, $\xi = (\mu_1,\sigma_1,\gamma_1,\ldots,\mu_{n_{\text{tdep}}},\sigma_{n_{\text{tdep}}},\gamma_{n_{\text{tdep}}})$, contains the means, asymptotic standard deviations, and rates of the Ornstein-Uhlenbeck processes of all stochastic parameters and n_{tdep} is the number of time-dependent parameters. Finally in the Equations 5, 7, and 8, we use the joint density of all stochastic parameters

$$p(\boldsymbol{\theta}^{\text{tdep}} \mid \boldsymbol{\xi}) = \prod_{i=1}^{n_{\text{tdep}}} f_{\boldsymbol{\theta}_i^{\text{tdep}} \mid \boldsymbol{\xi}}\left(\theta_i^{\text{tdep}}(t_1, \ldots, t_n) \mid \boldsymbol{\xi}\right) \quad . \tag{14}$$

2.3. Bayesian Inference and Recommended Analysis Procedure

2.3.1. Bayesian Inference

Bayesian inference consists of jointly inferring the constant parameters of the process model, θ^{const}, and of the observation model, $\boldsymbol{\psi}$, the time courses of the stochastic, time-dependent parameters, θ^{tdep}, the hyper-parameters, $\boldsymbol{\xi}$, of the stochastic processes of the time-dependent parameters, and, in case of a stochastic process model, the distribution of the outputs, $\mathbf{Y_{out}}$, from given observational data for $\mathbf{Y_{obs}}$. The joint posterior density of all these variables is proportional to the joint density that includes the observations given by Equation 5 or 7, respectively, with the actual observations substituted for $\mathbf{y_{obs}}$. Options for the numerical implementation of Bayesian inference for these equations will be discussed in Section 2.4.

As outlined in the introduction, inference of stochastic, time-dependent parameters can lead to the "misuse" of the large number of degrees of freedom to compensate for model deficits by inferring a time series that does not fluctuate randomly, but partly models a missing relationship of the time-dependent parameter as a function of inputs or model state variables. This violates our statistical assumptions and has to be avoided. For this reason, inference has to be followed by a careful analysis for such problems and strong model deficits have to be removed before using posterior distributions for prediction. In the next two subsections, we briefly discuss how this can be done by exploratory analysis of results and by cross-validation, respectively.

2.3.2. Exploratory Analysis of Results

The most obvious way of identifying the "misuse" of the large number of degrees of freedom of the inference process to compensate for model deficits is to search for relationships of the values of the time-dependent parameters on inputs, model state variables, and outputs. Any kind of exploratory statistical analysis of the values of the time-dependent parameters, inputs, internal model states and outputs at all points in time can be useful for this purpose. For this analysis we are not primarily interested in uncertainty, but in finding potential dependences of the parameter time series realization that led to the best model fit on model states, inputs or outputs. For models with low-dimensional state and parameter spaces, the simplest technique is to use scatter plots and smoothing algorithms to visualize and quantify potential relationships between the values of the time-dependent parameter and model states or inputs. In higher dimensions, multivariate exploratory statistical analyses can be done or machine learning algorithms can be applied to find potential relationships. The identified relationships can either be used directly to improve the model or they can inspire parameterized model extensions or even structural changes to the model. Inference of stochastic model parameters should then be redone with the improved model until no strong relationships can be identified which would indicate that a stochastic process may be a realistic description of system behavior.

2.3.3. Cross-Validation

For models with high-dimensional parameter or state spaces, it may be difficult to get confidence in having found all relevant relationships between the posterior time series of the parameter and states, inputs or outputs by the techniques described in the previous subsection. For this reason, it may be useful to learn from different approaches whether it makes sense to invest more effort in searching for such relationships. Predictive cross-validation is a simple technique to do so. If a time-dependent parameter was "misused" to compensate for model deficits, it will not be able to correct for this deficit in "prediction mode," when model results are based on the posterior of model and stochastic process parameters, but not on the posterior time-course of the stochastic parameter. In this case, this will thus lead to poor predictions. In case of short time-series that do not allow for predictive cross-validation, this can even be done by running the model in "prediction mode" for the calibration period, and still a potential compensation of model deficits by the posterior parameter time courses would be indicated by poor predictive performance.

2.4. Numerical Implementation

If the stochastic process used to describe the time-dependent parameters has the Markov property (such as the Ornstein-Uhlenbeck process), hydrological models that are based on differential equations become state-space or hidden Markov models. The following subsections briefly describe options for the numerical implementation of Bayesian inference for this class of models.

2.4.1. Particle Markov Chain Monte Carlo

States of state-space or hidden Markov Models are often inferred using the Ensemble Kalman Filter (EKF) (Evensen, 2009). This approach has the advantage of being very efficient, but the disadvantage of relying on linear approximations the accuracy of which is difficult to assess. Alternative approaches are particle filters or particle smoothers, which approximate the distribution of the states at each time point through a sample of values, called "particles." These particles are propagated through the time-series combining probabilistic propagation with the model (sampling from the distribution that describes the time evolution of the "particles" to the next output time point) with weighting based on the likelihood of the observations at the next time step (Fearnhead & Künsch, 2018; Godsill et al., 2004; Künsch, 2001; Van Leeuwen et al., 2019). The difference between particle filters and smoothers is that the former condition each state on current and past observations, while the latter condition on the full time series including future observations. Particle Markov Chain Monte Carlo (PMCMC) techniques combine particle filtering or smoothing for the states with Markov Chain Monte Carlo (MCMC) for the constant parameters either based on an approximation to the marginal likelihood calculated from the particle sample at each step of the Markov chain or by Gibbs sampling between states and parameters (Andrieu et al., 2010; Andrieu & Roberts, 2009; Kantas et al., 2015; Kattwinkel & Reichert, 2017; Sukys & Kattwinkel, 2018).

2.4.2. Hamiltonian Monte Carlo

The concept of Hamiltonian Monte Carlo (HMC) is to construct a Hamiltonian (energy as a function of locations and momentums of particles) with a potential energy equal to minus the log of the posterior density (interpreting the parameters as locations) and with a kinetic energy as a sum of quadratic terms of corresponding momenta. Drawing momenta randomly and integrating the Hamiltonian equations of motion over time leads to a proposal with an acceptance rate close to unity (Duane et al., 1987; Neal, 2011). This makes it possible to proceed with large steps without compromising the acceptance probability and thus makes this algorithm very efficient. The particle masses needed to formulate the kinetic energy and the time integration interval are tuning parameters of the algorithm.

Note that for performing the large sampling steps the equations of Hamiltonian dynamics have to be integrated numerically. This does not only require numerical integration, but it also requires multiple evolutions of the derivatives of the Hamiltonian, or the posterior probability density. Whenever this can be done analytically or by automatic differentiation, HMC is likely to considerably outperform Metropolis or Gibbs sampling. An interesting variant of HMC is to automatically adjust the integration time as it is done by the No-U-Turn Sampler, NUTS (Hoffman & Gelman, 2014). The idea of HMC can even be applied to numerically solve inference problems for stochastic differential equations (Albert et al., 2016).

2.4.3. Approximate Bayesian Computation

For many models it is difficult to evaluate the likelihood function analytically, but it may be (relatively) easy to sample from the model. Specific methods have been developed for this kind of problems (Albert et al., 2015; Beaumont, 2010; Beaumont et al., 2002; Marjoram et al., 2003). The simplest algorithm just consists of sampling from the prior and accepting the sample if the model prediction is closer to the data than some predefined tolerance (Marjoram et al., 2003). This and similar algorithms become increasingly inefficient with increasing dimension of the observation space. In this case, to increase the numerical efficiency of the algorithm, a small number of summary statistics of observations are used rather than the original observations. Such summary statistics can either be generated (Fearnhead & Prangle, 2012) or constructed based on characteristic properties of the solution. In hydrology, the latter choice is called "signature-based calibration" (Fenicia et al., 2018; Kavetski et al., 2018). Because of the use of summary statistics and the need of using a tolerance level as described above, we are sampling from an approximate posterior. In the context of this paper, this technique could be applied to infer the constant and stochastic
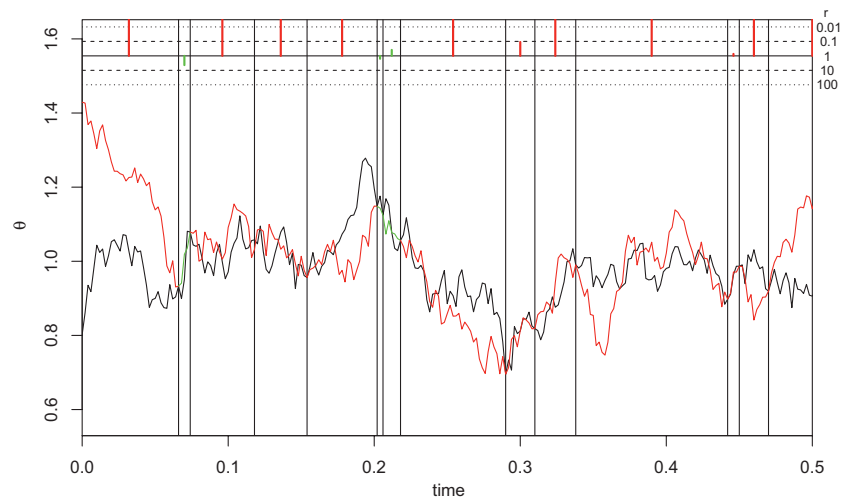
Figure 2. Illustration of conditional sampling from an Ornstein-Uhlenbeck process in subsequent subintervals. Black line: time course of the time-dependent parameter from the previous iteration step. Red and green lines: proposals within the subintervals separated by the vertical lines, green proposals were accepted, red proposals were rejected. On the top, the ratio of probability densities of proposal to previous time course are indicated on a logarithmic scale. Values larger than one lead to acceptance (they have to be green), values smaller than one are accepted with the probability equal to this ratio (they can thus be green or red).

process parameters, but, due to the high dimensionality of the inference problem, hardly for inferring the time-series of the stochastic parameters.

2.4.4. Conditional Ornstein-Uhlenbeck Sampling

The final technique discussed in this paper is to combine Metropolis or Metropolis-Hastings sampling of the constant parameters with conditional sampling from Ornstein-Uhlenbeck processes for the time-dependent parameter(s) in an overarching Gibbs sampling framework (Buser, 2003; Reichert & Mieleitner, 2009; Tomassini et al., 2009). The key idea is to increase the acceptance rate of sampling from an Ornstein-Uhlenbeck process by dividing the time range into subintervals and iteratively re-sample the Ornstein-Uhlenbeck process within the subintervals conditional on the values at the end points of the intervals to guarantee a continuous process (Buser, 2003; Tomassini et al., 2009).

The conditional Ornstein-Uhlenbeck sampling (COUS) process is illustrated in Figure 2. At each iteration step for the time-dependent parameters a random division of the time range into subintervals is chosen. Figure 2 illustrates then conditional proposals in these subintervals and their acceptance or rejection. When all subintervals were sampled, Metropolis or Metropolis-Hastings MCMC steps are taken for the constant parameters and for the process parameters of the Ornstein-Uhlenbeck process(es). For more details, see the original publications (Buser, 2003; Tomassini et al., 2009).

2.4.5. Comparison of the Approaches

The numerical efficiency of the algorithms can be expected to increase from Approximate Bayesian Computation (ABC) to COUS and PMCMC and from these to HMC. On the other hand, the implementation effort on the algorithmic side and regarding the need for adapting the simulation program is likely to increase in the same order. Unless user friendly packages become available for PMCMC or HMC for stochastic differential equations (there are packages available for HMC with ordinary differential equation models, see e.g., stan, https://mc-stan.org), it may be a useful strategy to do preliminary trials with relatively short time series with COUS and switch only to better techniques if the results are promising. This paper is based on the application of COUS (Buser, 2003; Tomassini et al., 2009) as implemented in the R package timedeppar, https://cran.r-project.org/package=timedeppar developed for this paper.

2.5. Metrics for Comparing Results

As the lumped error model combines the effects of input, model structure and observation error on modeled observed output, we do not have a complete probabilistic description of our knowledge of model states and output (in this approach, states and model output do not contain the effect of model structure error, and of input only if input uncertainty is considered explicitly; on the other hand, output with the lumped error includes also observation error). This is in contrast to the approach with stochastic parameters, where the probability distributions of states and output reflect our knowledge about states and true output that we can formulate without the random part of the observation error. (Note that systematic observation errors, e.g., could be considered as well (Sikorska & Renard, 2017; Sikorska et al., 2013; Thyer et al., 2011). We can still add the observation error to get predictions for observations. This difficulty implies that a fair comparison of model output between the two approaches is only possible at the observation level. Nevertheless, we will calculate the metrics discussed below for the probability distributions of $\mathbf{Y_{obs}}$ and $\mathbf{Y_{out}}$, even though, as discussed above, $\mathbf{Y_{out}}$ has a different meaning for the two approaches. In the following, we denote the probabilistic model outcomes as $\mathbf{Y}$ and later on substitute $\mathbf{Y_{obs}}$ or $\mathbf{Y_{out}}$ for $\mathbf{Y}$ for both of the modeling approaches. This leads to four different results for each of the metrics.

The most widely used metric to quantify model performance in hydrology is the Nash-Sutcliffe Efficiency (NSE) (Nash & Sutcliffe, 1970) which, when taking the expectation over a posterior distribution of model output or modeled observations, $\mathbf{Y}$, is given by

$$NSE(\mathbf{Y}, \mathbf{y}_{\text{obs}}) = \mathrm{E}\left[1 - \frac{\sum_{i=1}^{n}\left(Y(t_i) - y_{\text{obs}}(t_i)\right)^2}{\sum_{i=1}^{n}\left(y_{\text{obs}}(t_i) - \text{mean}\left(\left\{y_{\text{obs}}(t_j)\right\}_{j=1}^{n}\right)\right)^2} \right], \tag{15}$$

where mean(.) is the mean of the set provided as its argument, and the expected value, E, can be ignored if a sequence of real values is provided for $\mathbf{Y}$ rather than a sequence of random variables. NSE takes values between $-\infty$ and 1, where 1 indicate a perfect fit and values above 0.5 are often denoted as indicating an acceptable agreement.

To compare the distribution of data with predictive distributions, if the latter vary from observation to observation, it makes sense to evaluate the corresponding cumulative distribution functions of the predictions at the data points: $\left\{F_{Y(t_j)}\left(y_{\text{obs}}(t_j)\right)\right\}_{j=1}^{n}$. If the observations represent a sample from the predictive distributions, these transformed points should be a sample of a uniform distribution. This can be visualized by a QQ plot of these values (Renard et al., 2010, is an example with explanations in a hydrological context), or, as a numeric indicator, the "reliability," that is defined as the integral of the absolute deviation of the empirical cumulative distribution from the 1:1 line in this plot (McInerney et al., 2017; Renard et al., 2010). Unfortunately, this absolute deviation does not indicate whether the model prediction is too wide or too narrow; it just indicates a deviation of the observations from this distribution. As our knowledge is incomplete, we expect the predictive distribution to be wider than the empirical distribution of the data. For this reason, we are interested to distinguish excessive uncertainty (which we may accept) from overconfidence (which are indications of structural problems of the model). For this reason, we modify the "reliability" by omitting the absolute value in the deviation from the uniform cumulative distribution, splitting the integral at 0.5, and reverting the sign for the second part. If we denote the empirical cumulative distribution function of the observations relative to a predictive distribution transformed to a uniform distribution as F, we thus define the Data Coverage Deviation (DCD) as follows:

$$\begin{aligned} DCD(F) &= 4 \cdot \left(\int_{0}^{0.5}\left(x - F(x)\right)\mathrm{d}x + \int_{0}^{0.5}\left(F(1-x) - (1-x)\right)\mathrm{d}x \right) \\ &= 4 \cdot \left(\int_{0}^{0.5}\left(x - F(x)\right)\mathrm{d}x + \int_{0.5}^{1}\left(F(x) - x\right)\mathrm{d}x \right) . \end{aligned} \tag{16}$$

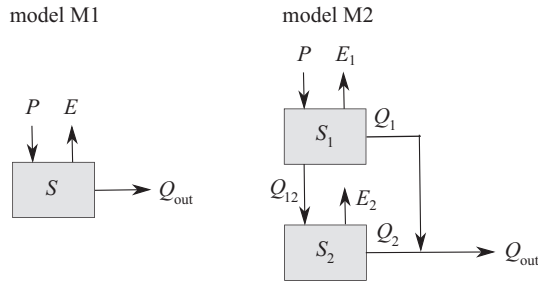model M1   model M2

Figure 3. Illustration of the hydrological models M1 and M2. See text and equations for an explanation of the symbols.

This measure is motivated as follows: $F(x)$ denotes the probability mass in the left tail up to the value x, which indicates the probability mass (= fraction of data points) we would expect for the uniform distribution. $x - F(x)$ is thus a measure of excess predictive uncertainty (or lack of data points) in the left tail of the predictive distribution up to a predicted probability x. The same is true for $F(1 - x) - (1 - x)$ for the right tail. Adding the two terms provides the excess uncertainty (or overconfidence if negative) in both tails together. To become independent of the threshold x, we integrate over all values of x and normalize accordingly to get a measure in the interval $[-1, 1]$ (this justifies the factor of 4 in Equation 16). The interpolation points of the distribution function F can easily be obtained by combining the sorted points of the sample $\left\{ F_{Y(t_j)}\left(y_{\text{obs}}(t_j)\right) \right\}_{j=1}^{n}$ for x values with equidistantly increasing points in the interval $[0, 1]$ for the y values. Integration in Equation 16 can then be done, for example, by using the trapezoidal rule. Note that the empirical distribution function, F, corresponds to the function in the plot of Figure 3 in Renard et al. (2010) with interchanged x and y axes. With this definition, we get positive values for excess uncertainty and negative values for overconfidence or strong over- or underprediction. A small degree of over- or underprediction combined with a narrow distribution of the data can still lead to positive values of DCD, as the data points may then still be sufficiently covered by the predictive distribution. Perfect agreement of data with the predictive distribution results in a deviation DCD of zero.

To check whether data coverage was achieved by very wide distributions, it makes sense, also to quantify the width of the predictive distribution. We do this by the average of the quotient of the standard deviation of the predictive distribution at each observation point and the observation:

$$\bar{\sigma}_{\text{rel}}(\mathbf{Y}, \mathbf{y}_{\text{obs}}) = \frac{1}{n} \sum_{i=1}^{n} \frac{\text{SD}\left[Y(t_i)\right]}{y_{\text{obs}}(t_i)} \quad , \tag{17}$$

where SD is the standard deviation of the random variable provided as its argument.

Another property of hydrological observations or model prediction is their strength of fluctuations, which can be quantified by the flashiness index, FI (Baker et al., 2004; Fenicia et al., 2018),

$$FI(\mathbf{Y}) = \text{E}\left[\frac{\dfrac{1}{n-1} \sum_{i=2}^{n} \mid Y(t_i) - Y(t_{i-1}) \mid}{\dfrac{1}{n} \sum_{i=1}^{n} Y(t_i)} \right] \quad , \tag{18}$$

where again the expected value, E, can be ignored, if the argument is a sequence of numbers rather than random variables.

2.6. Potential and Challenges of the Suggested Approach

Summarizing the arguments outlined in the introduction and methods sections, we expect the following potential of the suggested approach:

- At the initial stage of the analysis, stochastic, time-dependent parameters may be "misused" to partly represent missing deterministic relationships in the model. While these results cannot be used for uncertainty analysis, by the analysis of potential dependences of the time-dependent parameter(s) on model states and inputs, they provide information about model deficits in a constructive and nonparametric way and stimulate model structure improvement
- After removing severe structural deficits, stochastic parameters may lead to a more realistic description of the hydrological system by considering apparent intrinsic stochasticity and model structure deficits

while still maintaining mass balances exactly. In particular, propagating intrinsic errors to the output should contribute substantially to the observed autocorrelation in output errors that is difficult to parameterize empirically in lumped output error models

The suggested approach can be seen as an approach to combine mechanistic modeling with data science in the sense of "theory-guided data science" (Jiang et al., 2020; Karpatne et al., 2017) or "interpretable data science" (Molnar, 2019). It combines the advantages of a mechanism-based approach regarding interpretability and uncertainty quantification by being based on simple, hydrological models and a statistically rigorous description of uncertainty with the advantages of a data-based approach of profiting from a higher flexibility in process rates for better describing the data.

On the other hand, we see the following challenges of the approach:

- Identifiability problems between processes at close time scales can hamper the separate identification of different internal sources of stochasticity as well as the separation of internal stochasticity and the random component of observation errors
- Overlooked "misuses" of the time-dependence of parameters to describe nonstochastic relationships may lead to poor predictions
- High algorithmic and computational demand

These challenges need careful consideration and we have to learn from applications of the suggested technique to better being able to assess under which circumstances it is worth to follow this path. In the following case study, we will investigate the power of the suggested approach as well the challenges mentioned above.

3. Case Study

3.1. Catchment and Data

To demonstrate the methodology, we chose a small catchment that can be described by relatively simple, conceptual hydrological models and for which input uncertainty is small. The Maimai catchments on the South Island of New Zealand are a set of small headwater catchments with a long history of hydrological research and are thus well-suited for this purpose (Brammer & McDonnell, 1996; Seibert & McDonnell, 2002). We here use hourly precipitation (P), discharge (Q) and potential evaporation (E) data from the 3.8 ha M8 Maimai catchment (Freer et al., 2004; McDonnell et al., 2020). We use a calibration period of 6 months starting on June 5, 1985 after a long dry period that allows us to use priors favoring small initial water levels of the reservoirs. The 6 months calibration period is followed by a 2 months validation period until February 4, 1986 that contains multiple high discharge peaks as well as low flow periods.

3.2. Hydrological Models

Figure 3 illustrates the two very simple hydrological models used in this case study. They are intentionally kept simple to illustrate how the analysis of time-dependent parameters can support model development. The model M1 is defined by the following equations:

$$
\begin{aligned}
\frac{\mathrm{d}S}{\mathrm{d}t}(t) &= P(t) - Q_{\mathrm{out}}(t) - E(t), \\
Q_{\mathrm{out}}(t) &= \frac{f_k\, k}{S_{\mathrm{typ}}^{\alpha-1}} S^{\alpha}(t), \quad E(t) = c_{\mathrm{e}} E_{\mathrm{pot}}(t)\left[1 - \exp\left(-\frac{S(t)}{m}\right)\right].
\end{aligned}
\tag{19}
$$

The model M2 extends the model M1 by an additional "groundwater reservoir" and is defined by the following equations:

$$\frac{\mathrm{d}S_1}{\mathrm{d}t}(t) = P(t) - Q_1(t) - Q_{12}(t) - E_1(t),$$

$$\frac{\mathrm{d}S_2}{\mathrm{d}t}(t) = Q_{12}(t) - Q_2(t) - E_2(t),$$

$$Q_1(t) = \frac{f_{k_1} k_1}{S_{\mathrm{typ}}^{\alpha_1 - 1}} S_1(t)^{\alpha_1}, \quad Q_2(t) = \frac{f_{k_2} k_2}{S_{\mathrm{typ}}^{\alpha_2 - 1}} S_2(t)^{\alpha_2}, \quad Q_{12}(t) = k_{12} S_1(t),$$

$$E_1(t) = c_e E_{\mathrm{pot}}(t) \left[1 - \exp\left(-\frac{S_1(t)}{m_1} \right) \right],$$

$$E_2(t) = c_e E_{\mathrm{pot}}(t) \exp\left(-\frac{S_1(t)}{m_1} \right) \left[1 - \exp\left(-\frac{S_2(t)}{m_2} \right) \right],$$

$$Q_{\mathrm{out}}(t) = Q_1(t) + Q_2(t) \quad . \tag{20}$$

In these equations and in Figure 3, P is precipitation (mm/h); E_{pot} is the potential evapotranspiration rate (mm/h); E, E_1, and E_2 are actual evapotranspiration rates from different reservoirs (mm/h), Q_{out} is the calculated outflow of the catchment (mm/h); S, S_1, and S_2 are reservoir levels (mm); S_{typ} is a typical reservoir level, introduced to make the reservoir outflow parameters k, k_1, k_2, and k_{12} less dependent on the catchment; k, k_1, k_2, and k_{12} are reservoir outflow parameters (1/h); α, α_1, and α_2 are parameters to characterize nonlinearity in reservoir outflow as a function of reservoir level (-); c_e is a correction factor of potential evapotranspiration (-); m, m_1, and m_2 are parameters to describe limitation of evapotranspiration with decreasing reservoir levels (mm); and f_k, f_{k_1} and f_{k_2} are multipliers that are usually set to 1.

We distinguish the following submodels of the models M1 and M2: Model M1a is based on a linear reservoir ($\alpha = 1$) whereas in model M1b the exponent, α, of the relationship between discharge and outflow is estimated. Similarly, model M2a is based on linear reservoirs ($\alpha_1 = \alpha_2 = 1$), in model M2b the outflow of reservoir 1 is made nonlinear by estimating the exponent α_1 whereas α_2 is still kept at 1, and in model M2c both exponents, α_1 and α_2, are estimated.

3.3. Output Error Model

We assume the observed discharge at time t to be distributed according to the distribution $D_Q\big(Q_{\mathrm{out}}(t), \psi\big)$ (Ammann et al., 2019). The dependence of this distribution on $Q_{out}(t)$ allows us to use different distributions (of the same class) for different outflow rates. In particular, this allows us to describe heteroscedasticity.

In the likelihood function, we will have to evaluate this distribution for the observed discharge at the observation times t_i. To make the residuals at the observation times comparable, we transform them to normally distributed "residuals," η_i, according to the following transformation (Ammann et al., 2019):

$$\eta_i = F_{\mathrm{N}(0,1)}^{-1} \left(F_{D_Q(Q_{\mathrm{out}}(t_i), \psi)}\big(Q_{\mathrm{obs}}(t_i)\big) \right) \quad . \tag{21}$$

In this equation, F is the cumulative distribution function of the distribution given as its index and N represents the Normal distribution with mean and standard deviation given by its arguments.

In our specific example, we assume $D_Q\big(Q_{\mathrm{out}}, \psi\big)$ to be a shifted Normal or Student-t distribution with its mean and standard deviation given by:

$$\mu_{D_Q}(Q_{\mathrm{out}}) = Q_{\mathrm{out}} \quad , \qquad \sigma_{D_Q}(Q_{\mathrm{out}}, \psi) = \left[a \left(\frac{Q_{\mathrm{out}}}{Q_{\mathrm{typ}}} \right)^c + b \right] Q_{\mathrm{typ}} \tag{22}$$

with observation model parameters $\psi = (a, b, c)$. This distribution is truncated at zero and the integral of the tail in the negative domain is assigned to a point probability mass at $Q = 0$. The constant Q_{typ} in Equation 22 is a typical discharge and makes the parameters a and b nondimensional. In our application, Q_{typ} is chosen to be the mean of the observed discharge time series.
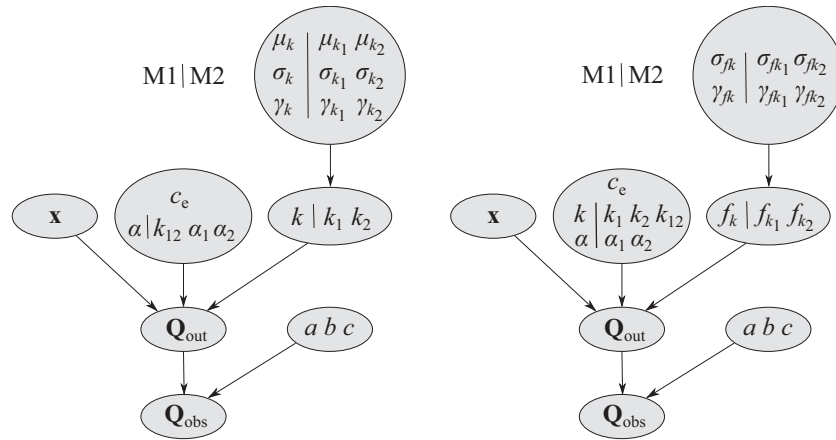
Figure 4. Illustration of the two slightly different implementations of the stochastic parameter approach shown in the right panel of Figure 1 to the hydrological models given by the Equations 19 and 20 (vertical lines separate parameters of model M1 from those of M2): Left panel: the parameters k (M1) or k_1 and k_2 (M2) become stochastic (here $f_k = f_{k_1} = f_{k_2} = 1$). Right panel: the parameters k (M1) or k_1 and k_2 (M2) remain constant parameters, but their multipliers, f_k (M1) or f_{k_1} and f_{k_2} (M2), become stochastic (here $\mu_{f_k} = \mu_{f_{k_1}} = \mu_{f_{k_2}} = 1$).

In our application we assume the observational likelihood to be the product of the densities of $D_Q\left(Q_{\mathrm{out}}(t_i), \psi\right)$ at all-time points, t_i, with data. This independence assumption is justified, as in the presence of stochastic, time-dependent parameters, the output error model only describes the small random part of the observation error and input and structural errors are described by propagating the uncertainty of the time-dependent parameters to the output. Note that systematic observation errors, for example, resulting from inaccurate gauging curves, are not considered by this approach but could be considered as well (Sikorska & Renard, 2017; Sikorska et al., 2013; Thyer et al., 2011). For better comparability, we use the same output error model for inference with constant parameters where the output error model describes the lumped errors resulting from input, model structure, and observation error. This makes the output errors much larger and violates the independence assumption. This will be discussed when interpreting the results.

3.4. Implementation of the Time-Dependent Parameter Concept for Hydrological Models

Due to the small catchment in our case study, we can assume the precipitation to be quite precisely known. The key uncertainty of hydrological dynamics in the models defined by the Equations 19 and 20 and illustrated in Figure 3 is then caused by the dependence of the outflows Q (M1) or Q_1 and Q_2 (M2) on the corresponding reservoir levels S (M1) or S_1 and S_2 (M2). These outflow rates are determined by the outflow rate parameters k (M1) or k_1 and k_2 (M2) and in case of nonlinear reservoirs by the exponents α (M1b), α_1 (M2b, M2c), and α_2 (M2c).

As the outflow rate parameters k (M1) or k_1 and k_2 (M2) most directly affect the modeled discharge and as reservoir outflow - water level relationships are very crucial for conceptual hydrological models, it seems natural to make the parameters k (M1) or k_1 and k_2 (M2) stochastic. Making instead the exponents, α or α_1 and α_2 stochastic would modify the same release rates and would thus not lead to fundamentally different results. Note that this choice of the stochastic parameters does not allow us to profit from the advantage of getting (additional) autocorrelation of fluctuations by the retention of water in internal reservoirs as they affect outflows that are directly linked to the model output. This would be different when making k_{12} stochastic. Nevertheless, due to the smaller sensitivity of the results to k_{12} than to k_2 and to avoid having two stochastic parameters in series, we decided to keep k_{12} constant. The left panel of Figure 4 shows the most straightforward way of applying the concept shown in the right panel of Figure 1 to our models. The parameters k (M1) or k_1 and k_2 (M2) become stochastic and time-dependent with the corresponding means, μ, asymptotic standard deviations, σ, and rates, γ as hyperparameters. An alternative way of making the parameters stochastic is to keep the parameters k (M1) or k_1 and k_2 (M2) constant and make the factors, f_k (M1) or f_{k_1} and f_{k_2} (M2), stochastic parameters with the mean kept fixed at 1. With this latter parameterization, a

Table 1
Prior Distributions of Different Variables (TruncNormal Refers to a Normal Distribution Truncated From Negative Values at Zero; in This Case, Mean and Std. Dev. Refer to the Distribution Before Truncation)

Variables	Value if not est.	Distribution	Mean	Std. dev.
		Parameters of the hydrological model		
k, k_1, k_2, k_{12}		Lognormal	0.02, 0.04, 0.005, 0.01 h^{-1}	0.25·mean
c_e		Lognormal	1	0.05
m, m_1, m_2	0.5, 0.5, 2 mm			
$\alpha, \alpha_1, \alpha_2$	1, 1, 1	Lognormal	2, 2, 1	0.25·mean
α_{12}	1			
$S_{ini}, S_{1,ini}, S_{2,ini}$		TruncNormal	0, 0, 0	5, 5, 5 mm
		Parameters of the output error model		
a, b	0.05, 0.02	Lognormal	0.1, 0.05	0.25·mean
c	0.7			
		Parameters of Ornstein-Uhlenbeck processes for factors f_k, f_{k_1}, f_{k_2}		
$\mu_{f_k}, \mu_{f_{k_1}}, \mu_{f_{k_2}}$	1			
$\sigma_{f_k}, \sigma_{f_{k_1}}, \sigma_{f_{k_2}}$	0.1	TruncNormal	0	0.05
$\gamma_{f_k}, \gamma_{f_{k_1}}, \gamma_{f_{k_2}}$	0.1 h^{-1}	Lognormal	0.1 h^{-1}	0.1·mean

change in the parameter k_i leads to a shift of the whole time series of $f_{k_i} \cdot k_i$. This improves convergence of our Gibbs sampling algorithm compared to keeping f_{k_i} fixed to unity and inferring k_i as a time-dependent parameter with inferred mean, as in the latter case, the mean can only shift as a consequences of shifts in many subintervals of our sampling scheme. For this reason we applied the parameterization shown in the right panel of Figure 4 to analyze stochastic, time-dependent parameters.

3.5. Parameter Values and Priors

The values of the fixed parameters as well as the priors of the estimated parameters are defined in Table 1. The priors for the water release coefficients could well be estimated from recession curves within a part of the hydrograph that was not used for calibration. For the other parameters of the hydrological model we used values based on our experience with other modeling studies. The priors for the initial reservoir levels have a preference for low values as the simulation starts after a long dry period. The prior for the asymptotic standard deviation of the Ornstein-Uhlenbeck process was chosen to be very narrow with a maximum at zero to induce a preference for constant parameters to avoid unnecessary time variations. This reflects our goal of getting the key pattern of the discharge time series from the hydrological model and not producing it with a "misuse" of the stochastic water release parameter. The joint prior density was chosen to be equal to the product of the individual prior densities of the estimated parameters. In addition to these parameters, we set the typical reservoir level, $S_{typ} = 10$ mm, and the typical discharge, Q_{typ} to the mean of the observed discharge. These settings just scale some of the other parameters to make them better transferable to other applications.

3.6. Statistical Inference and Computational Considerations

We applied the method COUS as described in Section 2.4.4 (Buser, 2003; Reichert & Mieleitner, 2009; Tomassini et al., 2009) to jointly infer constant and stochastic parameters. The hydrological model was implemented in R using the package deSolve, https://cran.r-project.org/package=deSolve with a C implementation of the right-hand side of the differential Equations 19 and 20 to profit from fast numerical integration. Inference was done using the R package timedeppar, https://cran.r-project.org/package=timedeppar. Inference with multiple chains for multiple models was run on Intel servers operated under Linux.

3.7. Results and Discussion

As outlined in Section 2.6, we have to identify and reduce model deficits before interpreting the results and their uncertainty. After the convergence analysis of the Markov chains in Section 3.7.1, we therefore analyze model deficits in Section 3.7.2. For the model with reduced deficits, we then interpret the inferred parameters in Section 3.7.3 and posterior output uncertainty during calibration and validation time intervals in Section 3.7.4.

3.7.1. Convergence and Computational Requirements

Convergence was easily achieved for all models with constant parameters when inferring hydrological parameters and the parameters a and b of the lumped error model (see Figures S1–S5 for 1d projections of the Markov chains and 1d marginals of all model parameters and Figures S6–S10 for 2d projections of the same Markov chain samples). Despite the good convergence, we simulated Markov chains of the posterior of length 100'000 (with a thinning factor of 100) after having established reasonable starting points and proposal distributions in earlier runs (see the supporting information figures mentioned above).

When estimating the parameters a and b of the observation error model (Equation 22) jointly with the parameters of the hydrological model and the time-dependent parameter f_k, and after choosing a large number of subintervalues (200–400, depending on the parameter) and an adaptive weighting for the sampling process of the interval end points, we achieved convergence of the algorithm for model M1a, with residual error model parameters in the order of $a = 0.07$ and $b = 0.0075$. However, for the other models that allow a better fit already with constant parameters, when estimating time-dependent parameters, the observation error model parameters decreased so much that the acceptance rate of the time-dependent parameters became very small and convergence could not be achieved (multiple chains got stuck at different local maxima of the posterior). This reflects the problem, that very small, random observation errors cannot be separated from internal model stochasticity. To resolve this problem, for the remaining analysis, we fixed the parameters $a = 0.05$ and $b = 0.02$. This leads to a random part of the observation error that is much smaller than the effect of parameter and intrinsic uncertainty on the output.

We then achieved acceptable convergence for all investigated model structures when running very long Markov chains of length 600'000 (with a thinning factor of 100) (see Figures S26–S30 for a complete overview of the Markov chains and 1d marginals of all constant and Ornstein-Uhlenbeck process parameters of all models, Figures S31–S35 for the corresponding 2d marginals, and Figures S36–S43 for the Markov chains and 1d marginals of the time-dependent parameters of all models at selected points in time).

The slow convergence process that requires many subintervals to sample the time-dependent parameters and long Markov chains for the constant and Ornstein-Uhlenbeck parameters leads to a very high computational burden. For the more complex model structure M2 and with 400 subintervals for the parameter f_{k_1} and 200 for f_{k_2} in our case study, it takes about a day for completing 10,000 Markov chain steps on a single core of modern Intel hardware. In addition, our 6 min time-resolution of the Ornstein-Uhlenbeck process (10 times higher resolution than the hourly data) leads to high memory requirements for storing the Markov chains for a half year calibration period, which we handled by choosing a thinning factor of 100 (storing only each 100th point of the Markov chain). Still, memory requirements for storing a sample of a single time-dependent parameter are then around 183 (days) · 24 (hours) · 10 (steps per hour) · 8 (bytes per double) · 600'000/100 (stored steps of the Markov chain) ≈ 2.1 Gbytes. Running multiple chains and multiple model structures in parallel on different cores, can thus lead to memory allocation problems. This could be avoided by not keeping the whole chain in memory which would, however, make postprocessing more difficult.

3.7.2. Deficit Analysis and Model Improvement

3.7.2.1. Scatter Plots

As discussed in Section 2.3.2, for our models with low-dimensional state and parameter spaces, scatter plots and smoothing of the solution with the highest observational likelihood can be used to identify potential relationships of the time-dependent parameters on model states or inputs.

The top left panel of Figure 5 shows the scatter plot of the variation factor, f_k, of the reservoir release rate, k, as a function of the reservoir level, S, for the model M1a in which the exponent, α, of the reservoir level is
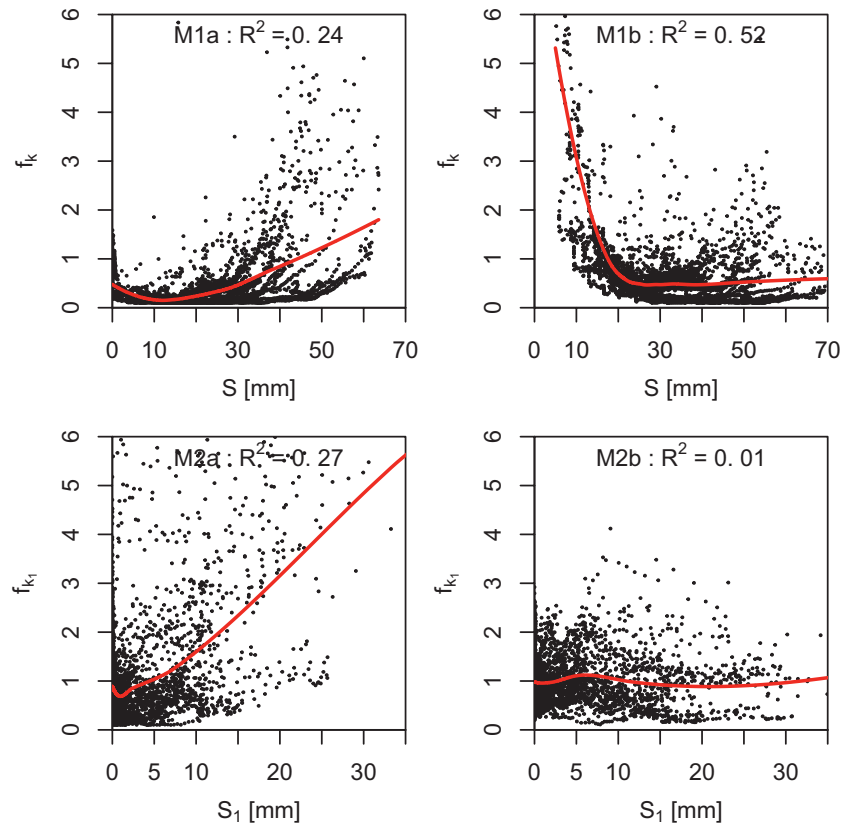
Figure 5. Scatter plot of the time-dependent parameters (multipliers), f_k versus the reservoir level S for the models M1a and M1b, and f_{k_1} versus the reservoir level S_1 for the models M2a and M2b. The red line indicates a potential relationship between the variables produced with a smoothing algorithm implemented in the R function loess. The value of R^2 at the top of the plots shows the degree of variance reduction achieved by the smoothing model.

set to unity (see Figure 3 and Equation 19). This plot shows that there is not a unique relationship between f_k and S but that there is very strong evidence that a stronger increase of the outflow as a function of the reservoir level than provided by the linear model should improve the model. This was the motivation for the model M1b in which the exponent, α, of the reservoir level was fitted in addition to k (see Figure 3 and Equation 19). This led to a considerable increase in the value of α that will be discussed in Section 3.7.3 (see in particular Figure 7). The resulting scatter plot for the model M1b is shown in the top right panel of Figure 5. This figure shows that this parameterization, while not being perfect, mostly resolves the problem of a systematic increase of f_k with S at large values of S. However, there is now a very strong increase in f_k needed at small reservoir levels, S (to a much smaller degree this was already the case for the model M1a). This indicates that the new parameterization of outflow as a function of reservoir level leads to problems with modeling base flow (or increases the problem already present for model M1a). Although a one box model may be able to describe a hydrological system sufficiently well (Kirchner, 2009) it often makes more sense to resolve fast response to rain events and the provision of relatively constant base flow by a two box model (Jakeman & Hornberger, 1993). This is thus the strategy for the next model extension (see Figure 3 and Equation 20).

To check, whether also in a two box model nonlinearity in outflow as a function of reservoir water level is needed, we start with the linear model M2a (see Figure 3 and Equation 20 with $\alpha_1 = \alpha_2 = 1$). The bottom left plot in Figure 5 shows the resulting scatter plots of the release rate modification factor f_{k_1} as a function of the reservoir level S_1. The plot clearly indicates the need for a nonlinear dependence of the outflow of reservoir 1 on S_1 analogously to the one box model. This is implemented in model M2b in which the exponent α_1 is fitted in addition to the other model parameters. This led to a significant increase in the parameter α_1 similarly to the increase in α for the model M1b (see Section 3.7.3). The bottom right plot in Figure 5

demonstrates that now scatter dominates potential remaining relationships. The smoothing relationship can only reduce 1% of the variance in f_{k_1}. The scatter plot of f_{k_1} versus S_1 for the model M2c, in which also potential nonlinearity in the outflow of the slow reservoir by an exponent α_2 is considered, looks quite similar to the one for the model M2b as the marginal posterior of α_2 in model M2c did not strongly deviate from its value of 1 (see Figure S59 and the black density of α_2 in Figure 7). To the degree assessable from the analyzed potential relationships, we conclude that the model M2b sufficiently reduced the biases to allow the quantification of its uncertainty with stochastic, time-dependent parameters. The model M2c does only slightly better.

3.7.2.2. Predictive Cross-Validation

In this simple example with a low-dimensional parameter space, it was easy to identify the "misuse" of stochastic, time-dependent parameters during calibration to compensate for model deficits by scatter plots of the time-dependent parameter versus model state variables (see Figure 5 and its discussion in the text). This may be more difficult in higher-dimensional settings where multivariate analyses or machine learning approaches may be needed. As discussed in Section 2.3.3, predictive cross-validation is an excellent tool to assess the presence of such "misuse," as the compensation of model deficits by the time-dependence during calibration will lead to a poor predictive performance of the model when operated in "predictive mode" where only information about the parameters of the stochastic processes but not about the realized time-courses are available. This is demonstrated in Figure 6 which clearly illustrates the breakdown of the prediction of the model M1a and the much better predictive power of the model M2c (see Figures S49–S53 for the time-series plots of all investigated model structures over this time interval, Figures S44–S48 for the results over the full calibration and validation periods, and Figures S54–S58 for the results within a short part of the validation period). The lack of predictive power of the model M1a with stochastic parameters seen in Figure 6 is supported by a strong decrease of the NSE from the calibration to the prediction range. This is shown in the top left panel of Figure 10 and will be discussed in Section 3.7.4.

3.7.2.3. Model Improvement

The analysis in this section, in particular the dependence analysis in Figure 5 and the cross-validation discussed in the previous paragraph and illustrated in Figure 6 show that the Model M2b mostly resolves the structural deficits of the models M1a, M1b and M2a and thus this model could be used for further analysis. Nevertheless, we will proceed with the model M2c that has the potential to perform even slightly better without adding too much additional complexity that could deteriorate its identifiability.

3.7.3. Posterior Distribution of Model Parameters

3.7.3.1. Marginal Posterior Parameter Distributions

Figure 7 shows the marginal posterior densities of the constant and stochastic process parameters of the model M2c with constant and time-dependent parameters. The simulations were started with a precipitation event after a long dry period. This leads to a rapid initial increase in the water level, S_1, as can be seen in the full time series in the supporting information (Figure S15 for constant parameters and Figure S48 for stochastic parameters). This makes the importance of the differences in initial values $S_{1,ini}$ and $S_{2,ini}$ less relevant than it seems from the two top left plots in Figure 7. The calibrations with and without time-dependent parameters lead to values of α_1 around 2.5 to 3 indicating a strong nonlinear increase of the outflow of reservoir 1 with increasing water level. On the other hand, we have a stronger nonlinearity with α_2 around 0.5 with constant parameters whereas it is close to 1 with time-dependent parameters. As will be seen in the next subsection, this leads to a poorer prediction of base flow for the model with time-dependent parameters. The values of the parameters k_1 and k_2 are not easy to compare because of the different exponents. However, the values of the (linear) interflow rate parameter k_{12} shows a strong difference between calibration with constant and time-dependent parameters. This difference indicates a different separation of the underground water body into the two compartments of the conceptual model. Its much smaller value for calibration with constant parameters implies that the slow reservoir reacts much more slowly to changes in the fast reservoir than for calibration with time-dependent parameters. This will be clearly seen in the slower dynamics of S_2 for constant than for time-dependent parameters in the next subsection (Figures 8
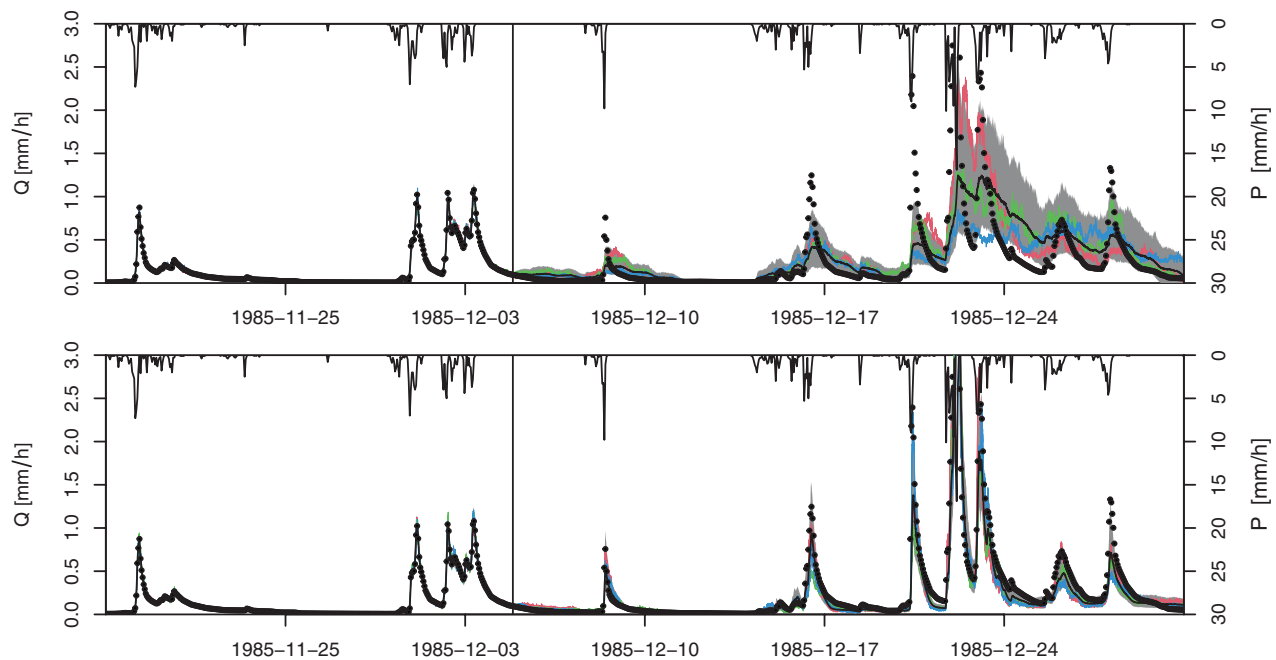
Figure 6. Final part of the calibration time range and initial range of the prediction time range of discharge, Q, for the models M1a (top) and M2c (bottom). Points represent data, lines and gray areas median, and 95% uncertainty bands; the vertical line separates the calibration from the prediction period. The colored lines represent three realizations of the model.

and 9). The asymptotic standard deviation, $\sigma_{f_{k_1}}$ in the order of 0.38 compared to $\sigma_{f_{k_2}}$ in the order of 0.25 indicates the need for a higher amplitude of the stochasticity of the fast reservoir.

It is a very remarkable result of this study that, despite the identical priors, we get considerably different posterior values for the time correlation parameters of the fast and the slow reservoirs, $\gamma_{f_{k_1}}$ and $\gamma_{f_{k_2}}$. The release rate parameter of the fast reservoir has a fluctuation rate in the order of $\gamma \approx 0.058\ \text{h}^{-1}$ which corresponds to a characteristic time $\tau \approx 17$ h whereas the slow (groundwater) reservoir has a fluctuation rate in the order of $\gamma \approx 0.019\ \text{h}^{-1}$ which corresponds to a characteristic time $\tau \approx 53$ h. This result matches the expectation that the fluctuation rates of the release rate parameter are larger for fast than for slow reservoirs. In addition, this result provides a nice explanation of a recent observation that an empirical, autocorrelated error model needs less autocorrelation during rain events when the discharge is driven by fast reservoirs than during dry weather periods during which the discharge is driven by slow reservoirs (Ammann et al., 2019). Our approach with stochastic, time-dependent parameters thus replaces the need for an explicit dependence of error model autocorrelation by stochastic parameters that can keep their characteristic time scales. The variation in output fluctuation time scale results then from the dominance of the outflow from one or the other reservoir (from reservoir 1 during and shortly after rain events and from reservoir 2 during longer dry periods).

3.7.4. Posterior Uncertainty of Model Output During Calibration and Prediction Time Intervals

The Figures 8 and 9 show time series of discharge and water levels of the model M2c with constant parameters and of the model M2c with time-dependent parameters f_{k_1} and f_{k_2}, respectively, over the final part of the calibration time range and the first part of the validation time range (see Figures S11–S25 for time series results of all models with constant parameters over the full calibration an validation ranges, over part of these ranges and over a short part of the validation range, and Figures S44–S58 for the corresponding plots with stochastic parameters). These plots and the known properties of the stochastic processes lead to the following observations:

•   The suggested approach with stochastic parameters leads to a description of uncertainty in all model state variables that is propagated to the output and is much larger than the random observation error
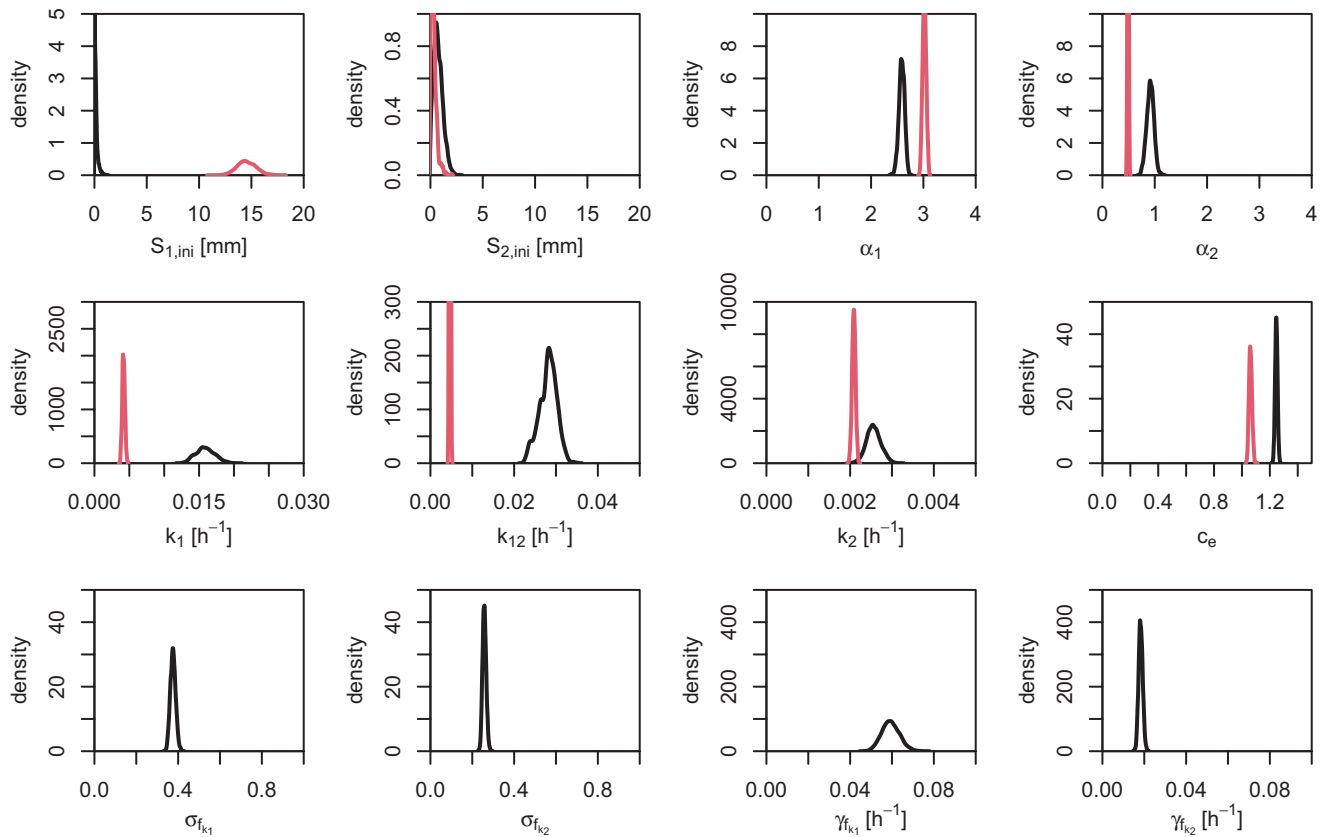
Figure 7. Posterior marginals of the constant and stochastic process parameters of model M2c (black) and of the same model with constant parameters (red).

- The approach leads to very narrow uncertainty bands during calibration, to an intermediate phase where the knowledge about observed discharge fades out, and a final phase with considerably higher prediction uncertainty in model states and output
- The approach naturally leads to autocorrelation in model output, in the presence of multiple reservoirs even to different output error time scales during phases in which different reservoirs have a dominating effect on the output
- The approach leads to a distinction of intrinsic uncertainty of a system from the random component of the observation error

Figure 10 quantifies the observations discussed in the previous paragraph and the previous subsections with the diagnostics measures introduced in Section 2.5.

The top left panel of Figure 10 shows the results for the NSE (Equation 15). The Nash-Sutcliffe Efficiency is very stable during calibration and prediction with constant parameters, although at much lower levels for the model structures M1a and M2a that do not consider nonlinearity in reservoir outflow - water level relationships. Due to the large number of degrees of freedom with stochastic parameters, the Nash-Sutcliffe Efficiency is very close to unity during the calibration phase with stochastic parameters. However, it drops very strongly for prediction with the model structures M1a and M2a which clearly show the model structure deficits of these models that were discussed in Section 3.7.2. It also drops quite strongly for the model M1b that considers nonlinearity but still relies on a single reservoir. As already discussed in Section 3.7.2, the models M2b and M2c perform similarly well as nonlinearity seems to be much more important for the fast than for the slow reservoir.

The top right panel of Figure 10 shows the results for the Flashiness Index, *FI* (Equation 18). In contrast to the other measures shown in Figure 10, the Flashiness Index is not a comparative measure between simulations and data but it reflects a property of a single time series. The horizontal lines at the bottom of the top
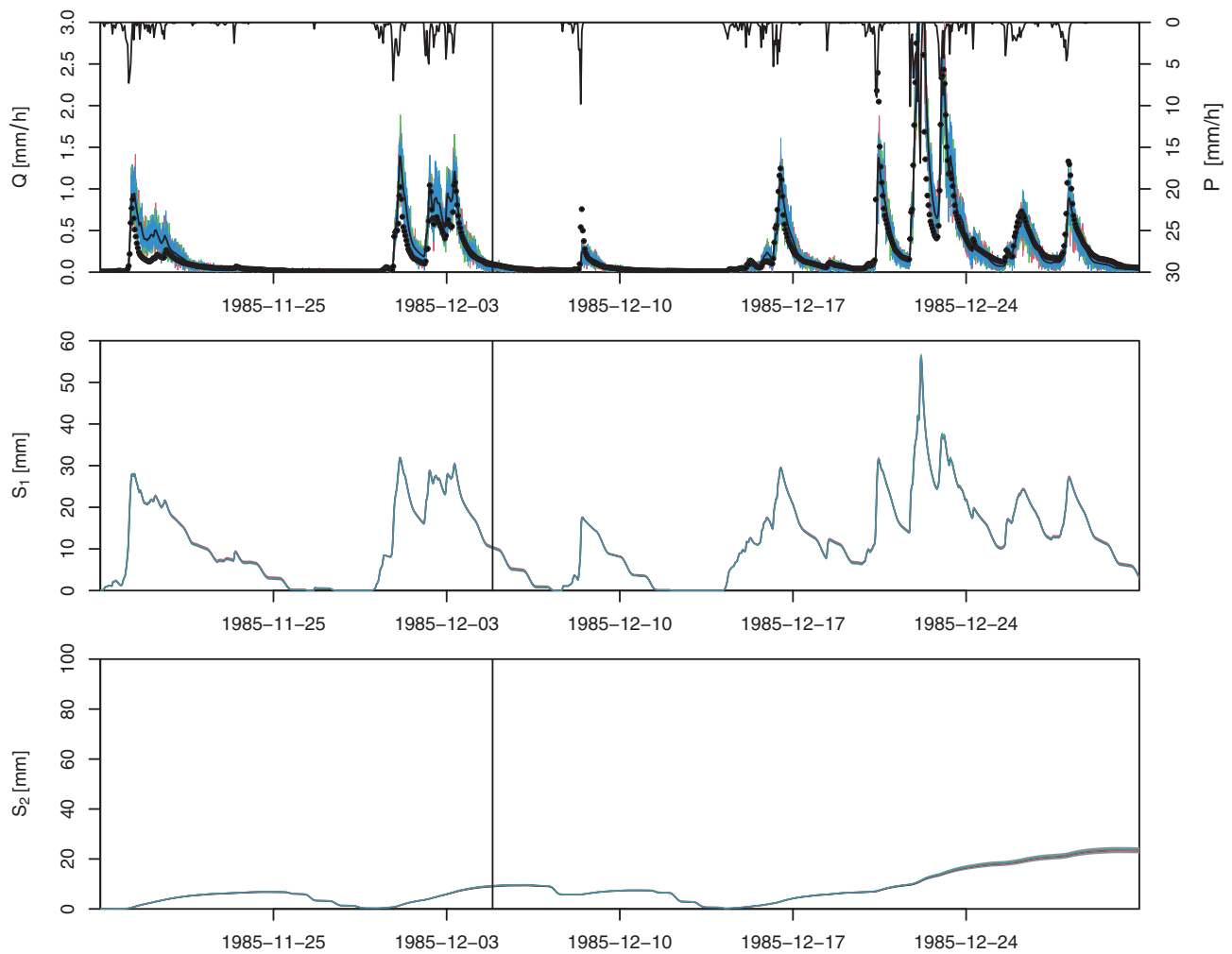
Figure 8. Time series of discharge, Q, and reservoir levels, S_1 and S_2 for the model M2c with constant parameters over the final part of the calibration time range and the first part of the validation time range. Points represent data, lines and gray areas median, and 95% uncertainty bands; the vertical line separates the calibration from the prediction period. The colored lines represent three individual samples of time evolution.

right panel of Figure 10 show the Flashiness Index of the observations. It is evident that the lumped error model without autocorrelation greatly overestimates the flashiness of the observations whereas the output of this model without the lumped error term, which only considers parametric uncertainty, has a flashiness very close to the data. The model with stochastic parameters overestimates the flashiness, in particular when considering the random component of the observation error, which still seems to have to be chosen to be too large, although it is a minor contribution to the overall prediction uncertainty that is dominated by the intrinsic stochasticity in this model. In all model versions, the flashiness is quite stable for prediction compared to calibration.

The bottom left panel of Figure 10 shows the results for spread, the mean relative standard deviation, $\overline{\sigma}_{\mathrm{rel}}$ (Equation 17) of the posterior distributions. We expect this quantity to be stable between calibration and prediction for the models with constant parameters. The decrease for prediction for these models is probably caused by a different fraction of rain and dry weather periods in these two time intervals. In contrast, there is a large increase in prediction uncertainty compared to calibration for the model with stochastic parameter. This reflects our much better knowledge of state and output during calibration than for prediction. This is a very important feature missing in the simple lumped error model.

Finally, the bottom right panel of Figure 10 shows the results for the DCD (Equation 16). The value close to zero for the model with constant parameters including the lumped error model indicates a similar shape of
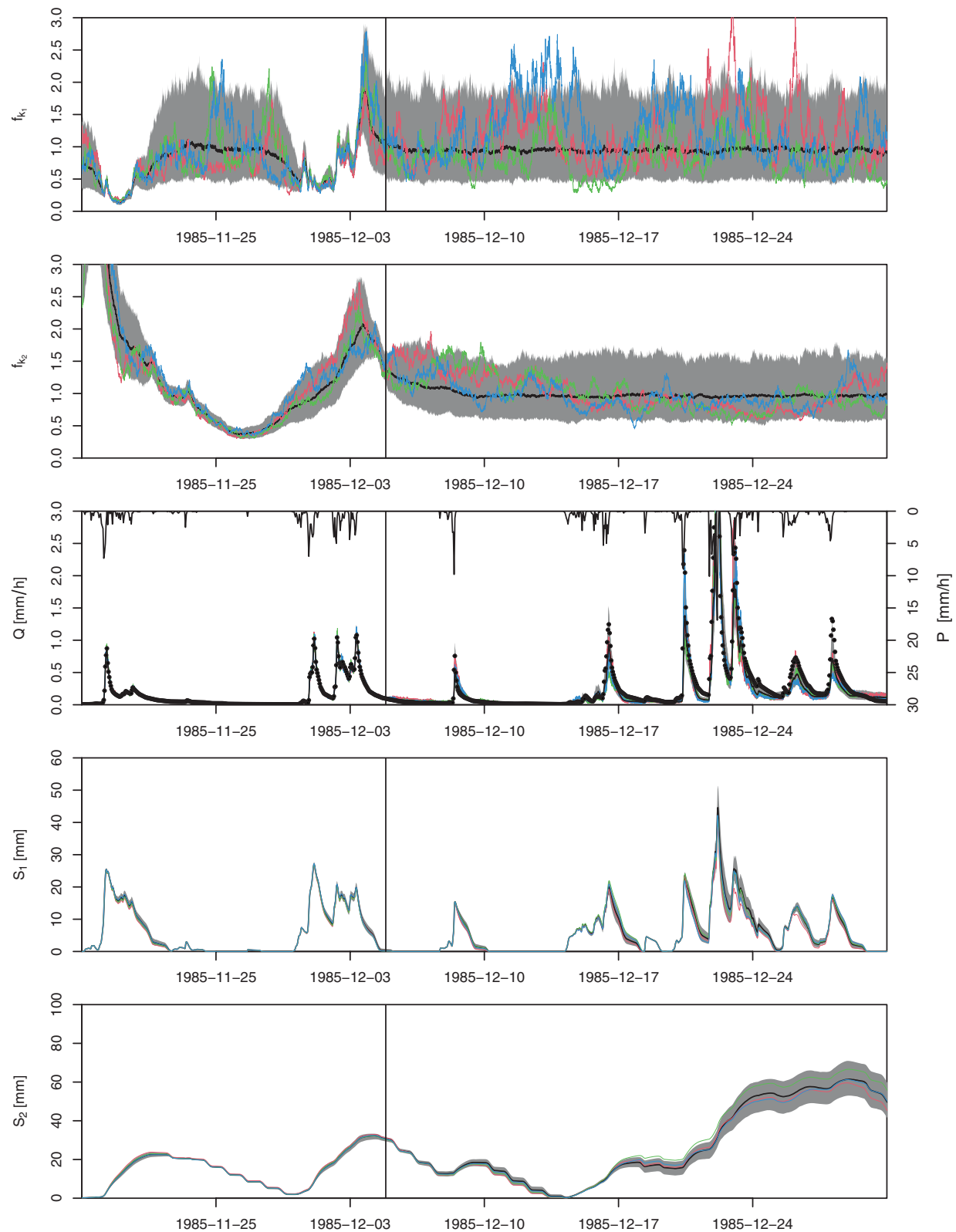
Figure 9. Time series of the time-dependent parameters (multiplier), f_{k_1} and f_{k_2}, discharge, Q, and reservoir levels, S_1 and S_2 for the model M2c over the final part of the calibration time range and the first part of the validation time range. Points represent data, lines and gray areas median and 95% uncertainty bands; the vertical line separates the calibration from the prediction period. The colored lines represent three individual samples of time evolution.
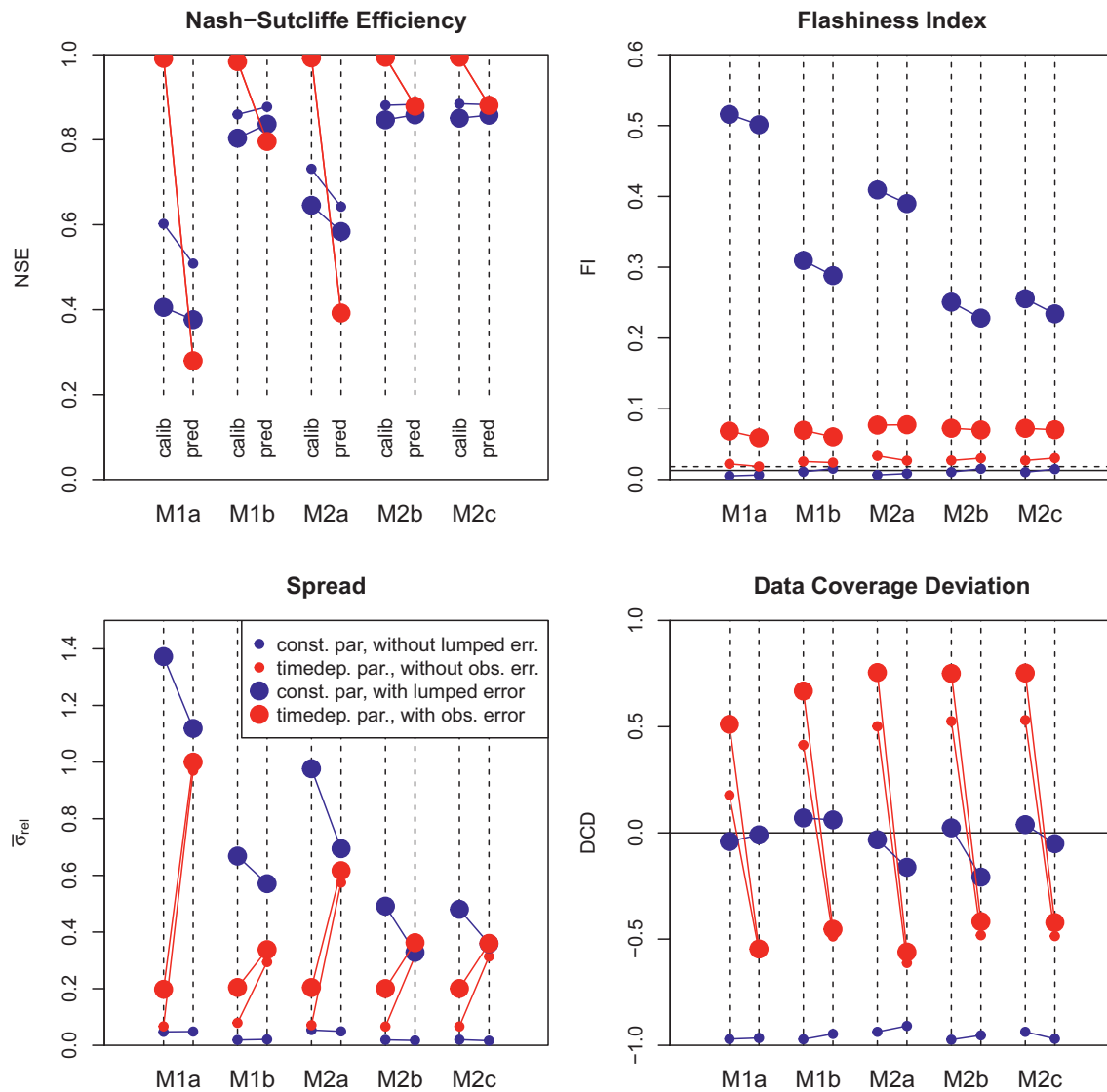
Figure 10. Values of diagnostics measures introduced in Section 2.5 for the calibration and validation time domains and for model output with and without the lumped/observation error model for all model structures. Small circles indicate results for the model output without the lumped error model (in case of constant parameters) or without the observation error model (in case of time-dependent parameters), blue circles indicated results for constant model parameters and red circles results with stochastic model parameters. The results for the calibration range of each model are plotted left of those for the validation time range and the two corresponding values are connected with a line to clarify which pairs belong together.

the posterior distributions and the data. Omitting the lumped error model clearly leads to a very strong underestimation of uncertainty which is a natural result (ignoring major parts of uncertainty cannot lead to a good uncertainty estimate). The models with stochastic parameters show a strong decrease in data coverage from calibration to validation which is in contrast to the strong increase in spread. It seems that the increase in spread is still insufficient to lead to a good data coverage for prediction. Figure 11 identifies the cause of this problem. Insufficient coverage of the data for the model with stochastic parameters occurs primarily during low flows where even a quite large relative error is insufficient to cover the data if the predicted discharge is too low. The model with constant parameters performs better at low discharge first, because its lower value of α_2 allows more discharge and second, because the offset parameter, b, of the lumped error model increases the relative error strongly for small values of discharge. A similar mechanism is harder to get with stochastic parameters. First, in the current model formulation, uncertainty in the parameter k_i is independent of the discharge and cannot easily model such an offset. Second, an increase in the offset of the
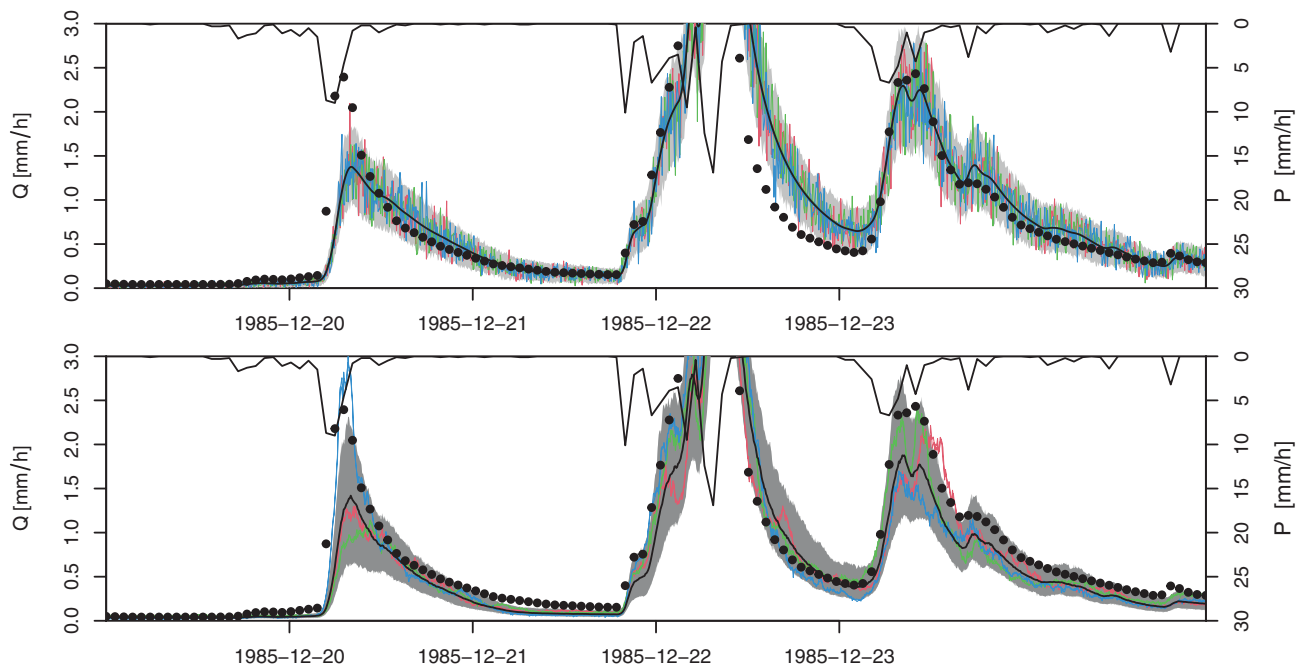
Figure 11. Time series of discharge, Q, for the model M2c with constant parameters over a small part of the validation time range for calibration with constant model parameters (top) and with stochastic model parameters (bottom). Points represent data, lines and gray areas median, and 95% uncertainty bands. The colored lines represent three individual samples of time evolution.

random observation error, that could easily increase the uncertainty at low discharges, is in clear disagreement with the observations as it would even more increase the flashiness of the time series.

3.8. Conclusions of the Case Study

Our results clearly demonstrated some of the advantages of stochastic model parameters, both as a diagnostic approach to understand model deficits, and as an instrument to improve the characterization of predictive uncertainty. However, even this simple application also showed some of the challenges that are associated with having to solve a much more complex inference problem than when using time constant model parameters.

Failures in predictive cross-validation of models with stochastic parameters (see Figure 6 and the top left panel of Figure 10) clearly identified deficient model structures, and the (simple) exploratory analyses of the time-course of the parameters led to constructive suggestions for model improvement (see Figure 5). After reducing model deficits, stochastic process parameters lead to much more realistic uncertainty estimates of internal model states than with a lumped error model that adds uncertainty only to the output (compare uncertainty bands of the states S_1 and S_2 in the Figures 8 and 9). In addition, the stochastic model structure allows us to describe our knowledge about the true output and add (the random part of) the observation error if we want to describe observations. In contrary, a lumped error model leads to a difficult to interpret model output without the error term, as it considers only part of the uncertainty and when adding the lumped error term, we have to include observation error also (as it cannot be not separated from the other errors). This does not allow us to make predictions about the true output, only of observations. The stochastic model reflects the difference in posterior knowledge we have between calibration and prediction time ranges, by having much narrower uncertainty bands in the calibration period (where we use the observations) than in the prediction period (where we do not use the observations). Simple lumped error models do not correctly describe this important property of the results. Note however, that this feature can also be reproduced with more sophisticated output error models (Reichert & Schuwirth, 2012). One of the key advantages of the stochastic process parameters found in our case study was the identification of different

fluctuation time scales in different reservoir outflow coefficients that explain varying time scales found in a lumped error model in an earlier study (Ammann et al., 2019).

On the other hand, stochastic model parameters result in a greatly higher computational demand and more severe identifiability and convergence problems than constant model parameters. In addition, the resulting prediction uncertainty bands were too narrow at low predicted discharge (see Figure 11). This lead to an insufficient coverage of the data by the prediction bands (see the bottom right panel of Figure 10). The cause of this problem is that an uncertain water release parameter, k_i, leads to a relative error in the resulting discharge (see Equations 19 and 20) which is known to be insufficient for describing uncertainty for discharge predictions that are much smaller than the observations. In a lumped error model, this problem is easily solved by adding an offset to the width of the empirical error distribution, but this simple recipe cannot directly be transferred to the setting in the case study as the uncertainty in the parameters k_i is independent of the reservoir level, S_i.

4. Conclusions

Considering the theoretical development and extrapolating from the case study, we can draw the following conclusions:

The potential of the suggested approach can be summarized as follows:

- Stochastic, time-dependent parameters are a very convincing concept for considering intrinsic model uncertainty (apparent stochasticity and model structure uncertainty) while still exactly keeping mass balances
- Cross-validation of models with stochastic parameters very sensitively identifies model structure deficits and exploratory analysis of potential dependences of the stochastic parameter on model states and external influence factors provide constructive hints to model improvement
- After having eliminated strong model structure deficits, models with stochastic, time-dependent parameters lead to a much better description of (apparent) nondeterministic behavior and uncertainty in internal states and output than approaches based on deterministic models. In particular: (i) The propagation of (apparent) intrinsic stochasticity to model output naturally leads to autocorrelated output errors (this was not fully exploited in our case study). (ii) Stochastic parameters with different characteristic time scales may explain different dominating fluctuation time scales in model output during times in which different hydrological processes dominate the output. (iii) The characterization of posterior knowledge of stochastic models naturally leads to much narrower posterior distributions of model output during the calibration phase, where the knowledge is conditioned to the observations, than during the prediction phase where no observations of output are used

On the other hand, there are remaining challenges that need more research:

- Bayesian inference with stochastic, time-dependent parameters is computationally very demanding which is a particular problem for inference from long time-series. This problem may become less severe with improved numerical algorithms, with the availability of easy-to-use numerical implementations, and with increasing power of computational resources. The R package timedeppar, https://cran.r-project.org/package=timedeppar, developed for this paper, is a first step in this direction
- In the presence of processes with similar fluctuation time scales or small amplitudes, unique identification of different sources of stochasticity may be difficult (in our case study, we could not identify the random component of the observation error)
- As in hydrological modeling the intrinsic error will typically dominate the observation error, it may be more difficult to precisely characterize output uncertainty than with an empirical parameterization of a large, lumped error term (in our case study, uncertainty for low values of predicted discharge was underestimated due to the lack of an empirical parameter for an offset in the error variance at low discharge)

As we improve predictions by combining a mechanistic model with learning from adding many degrees of freedom to the inference problem, the suggested approach can be seen as an implementation of "theory-guided data science" or "interpretable data science." More applications of this approach to other case

studies and the consideration of some of the above-mentioned research needs could strongly contribute to a better assessment of its usefulness for improving dynamic environmental models and their uncertainty quantification.

Data Availability Statement

Current versions of the R package timedeppar developed for this paper and the code of the hydrological simulation program are available at: https://cran.r-project.org/package=timedeppar, https://gitlab.com/p.reichert/timedeppar, https://gitlab.com/p.reichert/conhydmod. In addition, the frozen versions of software and data exactly as used for this paper are available in our institutional repository https://opendata.eawag.ch at https://doi.org/10.25678/0002TG.

References

Albert, C., Künsch, H.-R., & Scheidegger, A. (2015). A simulated annealing approach to Approximate Bayes Computations. *Statistics and Computing*, *25*(6), 1217–1232.

Albert, C., Ulzega, S., & Stoop, R. (2016). Boosting Bayesian parameter inference of nonlinear stochastic differential equation models by Hamiltonian scale separation. *Physical Review E*, *93*, 043313.

Ammann, L., Fenicia, F., & Reichert, P. (2019). A likelihood framework for deterministic hydrological models and the importance of non-stationary autocorrelation. *Hydrology and Earth System Sciences*, *23*, 2147–2172.

Andrieu, C., Doucet, A., & Holenstein, R. (2010). Particle Markov chain Monte Carlo. *Journal of the Royal Statistical Society: Series B*, *72*, 269–342.

Andrieu, C., & Roberts, G. O. (2009). The pseudo-marginal approach for efficient Monte Carlo computations. *Annals of Statistics*, *37*, 697–725.

Baker, D. B., Richards, R. P., Loftus, T. T., & Kramer, J. W. (2004). A new flashiness index: characteristics and applications to mid-western rivers and streams. *Journal of the American Water Resources Association*, *40*, 503–522.

Bates, B. C., & Campbell, E. P. (2001). A Markov chain Monte Carlo scheme for parameter estimation and inference in conceptual rainfall-runoff models. *Water Resources Research*, *37*(4), 937–947.

Beaumont, M. A. (2010). Approximate Bayesian Computation in evolution and ecology. *Annual Review of Ecology, Evolution and Systematics*, *41*, 379–406.

Beaumont, M. A., Zhang, W., & Balding, D. J. (2002). Approximate Bayesian Computation in population genetics. *Genetics*, *23*, 2025–2035.

Beck, M. B. (1987). Water quality modeling: A review of the analysis of uncertainty. *Water Resources Research*, *23*(8), 1393–1442.

Beck, M. B., & Young, P. C. (1976). Systematic identification of DO-BOD model structure. *Journal of the Environmental Engineering Division, ASCE*, *103*(EE5), 902–927.

Blöschl, G., & Sivaplalan, M. (1995). Scale issues in hydrological modeling: a review. *Hydrological Processes*, *9*, 251–290.

Brammer, D. D., & McDonnell, J. J. (1996). En evolving perceptual model of hillslope flow at the Maimai catchment. *Advances in Hillslope Processes*, *1*, 35–60.

Buser, C. M. (2003). *Differentialgleichungen mit zufälligen zeitvariierenden Parametern*. Zürich: ETH Zürich.

Clark, M. P., Rupp, D. E., Woods, R. A., Zheng, X., Ibbitt, R. P., Slater, A. G., & Uddstrom, M. J. (2008). Hydrological data assimilation with the ensemble kalman filter: Use of streamflow observations to update states in a distributed hydrological model. *Advances in Water Resources*, *31*, 1309–1324.

Del Giudice, D., Albert, C., Rieckermann, J., & Reichert, P. (2016). Describing catchment-averaged precipitation as a stochastic process improves parameter and input estimation. *Water Resources Research*, *52*, 3162–3186.

Duan, Q., Ajami, N. K., Gao, X., & Sorooshian, S. (2007). Multi-model ensemble hydrologic prediction using Bayesian model averaging. *Advances in Water Resources*, *30*, 1371–1386.

Duane, S., Kennedy, A. D., Pendleton, B. J., & Roweth, D. (1987). Hybrid Monte Carlo. *Physics Letters B*, *195*(2), 216–222.

Evensen, G. (2009). *Data assimilation - the ensemble kalman filter* (2nd ed.) Springer.

Evin, G., Kavetski, D., Thyer, M., & Kuczera, G. (2013). Pitfalls and improvements in the joint inference of heteroscedasticity and autocorrelation in hydrological model calibration. *Water Resources Research*, *49*, 4518–4524.

Evin, G., Thyer, M., Kavetski, D., McInemey, D., & Kuczera, G. (2014). Comparison of joint versus postprocessor approaches for hydrological uncertainty estimation accounting for error autocorrelation and heteroscedasticity. *Water Resources Research*, *50*, 2350–2375.

Fearnhead, P., & Künsch, H.-R. (2018). Particle filters and data assimilation. *Annual Review of Statistics and its Application*, *5*, 11.1–11.29.

Fearnhead, P., & Prangle, D. (2012). Constructing summary statistics for approximate Bayesian computation: semi-automatic approximate Bayesian computation. *Journal of the Royal Statistical Society B*, *74*, 419–474.

Fenicia, F., Kavetski, D., Reichert, P., & Albert, C. (2018). Signature-domain calibration of hydrological models using approximate Bayesian computation: Empirical analysis of fundamental properties. *Water Resources Research*, *54*, 3958–3987.

Fenicia, F., Kavetski, D., Savenije, H. H. G., & Pfister, L. (2016). From spatially variable streamflow to distributed hydrological models: Analysis of key modeling decisions. *Water Resources Research*, *52*, 954–989.

Freer, J. E., McMillan, H., McDonnell, J. J., & Beven, K. J. (2004). Constraining dynamic TOPMODEL responses for imprecise water table information using fuzzy rule based performance measures. *Journal of Hydrology*, *291*, 254–277.

Godsill, S. J., Doucet, A., & West, M. (2004). Monte Carlo smoothing for nonlinear time series. *Journal of the American Statistical Association*, *99*(465), 156–168.

Hoffman, M. D., & Gelman, A. (2014). The No-U-Turn sampler: Adaptively setting path lengths in Hamiltonian Monte Carlo. *Journal of Machine Learning Research*, *15*, 1351–1381.

Jakeman, A. J., & Hornberger, G. M. (1993). How much complexity is warranted in a rainfall-runoff model? *Water Resources Research*, *29*(8), 2637–2649.

Jiang, S., Zheng, Y., & Solomatine, D. (2020). Improving ai system awareness of geoscience knowledge: Symbiotic integration of physical approaches and deep learning. *Geophysical Research Letters*, *46*, e2020GL088229.

Kantas, N., Doucet, A., Sing, S. S., Maciejowski, J., & Chopin, N. (2015). On particle methods for parameter estimation in state-space models. *Statistical Science*, *30*, 328–351.

Karpatne, A., Atluri, G., Faghmous, J. H., Steinbach, M., Banerjee, A., Ganguly, A., & Kumar, V. (2017). Theory-guided data science: a new paradigm for scientific discovery. *IEEE Transactions on Knowledge and Data Engineering*, *29*(10), 2318–2331.

Kattwinkel, M., & Reichert, P. (2017). Bayesian parameter inference for individual-based models using a Particle Markov Chain Monte Carlo method. *Environmental Modelling and Software*, *87*, 110–119

Kavetski, D., Fenicia, F., Reichert, P., & Albert, C. (2018). Signature-domain calibration of hydrological models using approximate Bayesian computation: Theory and comparison to existing applications. *Water Resources Research*, *54*, 3958–3987..

Kavetski, D., Franks, S. W., & Kuczera, G. (2003). Confronting input uncertainty in environmental modeling. In Q. Duan, H. V. Gupta, S. Sorooshian, A. N. Rousseau, & R. Turcotte (Eds.), *Calibration of watershed models* (pp. 49–68). Washington, DC: American Geophysical Union.

Kavetski, D., Kuczera, G., & Franks, S. W. (2006a). Bayesian analysis of input uncertainty in hydrological modeling: 1. Theory. *Water Resources Research*, *42*, W03407.

Kavetski, D., Kuczera, G., & Franks, S. W. (2006b). Bayesian analysis of input uncertainty in hydrological modeling: 2. Application. *Water Resources Research*, *42*, W03408.

Kirchner, J. W. (2009). Catchments as simple dynamical systems: Catchment characterization, rainfall-runoff modeling, and doing hydrology backward. *Water Resources Research*, *45*, W02429.

Koutsoyiannis, D. (2002). The hurst phenomenon and fractional gaussian noise made easy. *Hydrological Sciences Journal*, *47*(4), 573–595.

Kuczera, G. (1983). Improved parameter inference in catchment models. 1. evaluating parameter uncertainty. *Water Resources Research*, *19*(5), 1151–1162.

Kuczera, G. (1990). Estimation of runoff-routing parameters using incompatible storm data. *Journal of Hydrology*, *114*, 47–60.

Kuczera, G., Kavetski, D., Franks, S., & Thyer, M. (2006). Towards a Bayesian total error analysis of conceptual rainfall-runoff models: Characterizing model error using storm-dependent parameters. *Journal of Hydrology*, *331*(1-2), 161–177.

Künsch, H.-R. (2001). State Space and Hidden Markov Models. In O. E. Barndorff-Nielsen, D. R. Cox, & C. Klüppelberg (Eds.), *Complex stochastic systems* (pp. 109–173). Boca Raton: Chapman & Hall/CRC.

Leisenring, M., & Moradkhani, H. (2010). Snow water equivalent prediction using Bayesian data assimilation methods. *Stochastic Environmental Research and Risk Assessment*, *25*, 253–270.

Liang, Z., Wang, D., Guo, Y., Zhang, Y., & Dai, R. (2013). Application of bayesian model averaging approach to multimodel ensemble hydrologic forecasting. *Journal of Hydrologic Engineering*, *18*(11), 1426–1436.

Liu, J., & West, M. (2001). Combined parameter and state estimation in simulation-based filtering. In A. Doucet, N. de Freitas, & N. Gordon (Eds.), *Sequential Monte Carlo methods in practice* (pp. 197–223). New York: Springer.

Liu, Y., & Gupta, H. V. (2007). Uncertainty in hydrologic modeling: Toward an integrated data assimilation framework. *Water Resources Research*, *43*, W07401.

Liu, Y., Weerts, A. H., Clark, M., Hendricks Franssen, H.-J., Kumar, S., Moradkhani, H., & Restrepo, P. (2012). Advancing data assimilation in operational hydrologic forecasting: progresses, challenges, and emerging opportunities. *Hydrology and Earth System Sciences*, *16*, 3863–3887.

Mandelbrot, B. B., & Wallis, J. R. (1968). Noah, joseph, and operational hydrology. *Water Resources Research*, *4*(5), 909–918.

Marjoram, P., Molitorm, J., Plagnol, V., & Tavare, S. (2003). Markov chain Monte Carlo without likelihoods. *Proceedings of the National Academy of Sciences of the United States of America*, *100*(26), 15324–15328.

McDonnell, J., Gabrielli, C., Ameli, A., Ekanayake, J., Fenicia, F., Freer, J., & Woods, R. (2020). *Maimai experimental watershed. HydroShare*. http://www.hydroshare.org/resource/a292cb65a5d24a31a60978b2ab390266

McInerney, D., Thyer, M., Kavetsky, D., Lerat, J., & Kuczera, G. (2017). Improving probabilistic prediction of daily streamflow by identifying Pareto optimal approaches for modeling heteroscedastic residual errors. *Water Resources Research*, *53*, 2199–2239.

Molnar, C. (2019). *Interpretable machine learning - a guide for making black box models explainable*. https://christophm.github.io/interpretable-ml-book/

Moradkhani, H., Hsu, K.-L., Gupta, H., & Sorooshian, S. (2005). Uncertainty assessment of hydrologic model states and parameters: Sequential data assimilation using the particle filter. *Water Resources Research*, *41*, W05012. https://doi.org/10.1029/2005WR003604

Nash, J. E., & Sutcliffe, J. V. (1970). River flow forecasting through conceptual models. Part i - A discussion of principles. *Journal of Hydrology*, *10*, 282–290.

Neal, R. M. (2011). Mcmc using hamiltonian dynamics. In S. Brooks, A. Gelman, G. L. Jones, & X.-L. Meng (Eds.), *Handbook of Markov chain Monte Carlo* (pp. 113–162). Boca Raton: CRC Press.

Reichert, P., & Mieleitner, J. (2009). Analyzing input and structural uncertainty of nonlinear dynamic models with stochastic, time-dependent parameters. *Water Resources Research*, *45*, W10402. https://doi.org/10.1029/2009WR007814

Reichert, P., & Schuwirth, N. (2012). Linking statistical description of bias to multi-objective model calibration. *Water Resources Research*, *48*, W09543. https://doi.org/10.1029/2011WR011391

Renard, B., Kavetski, D., Kuczera, G., Thyer, M., & Franks, S. W. (2010). Understanding predictive uncertainty in hydrologic modeling: The challenge of identifying input and structural errors. *Water Resources Research*, *46*, W05521. https://doi.org/10.1029/2009WR008328

Schoups, G., & Vrugt, J. A. (2010). A formal likelihood function for parameter and predictive inference of hydrologic models with correlated, heterscedastic, and non-Gaussian errors. *Water Resources Research*, *46*, W10531. https://doi.org/10.1029/2009WR008933

Seibert, J., & McDonnell, J. J. (2002). On the dialog between experimentalist and modeler in catchment hydrology: Use of soft data for multicriteria model calibration. *Water Resources Research*, *38*(11), 1241. https://doi.org/10.1029/2001WR000978

Sikorska, A. E., & Renard, B. (2017). Calibrating a hydrological model in stage space to account for rating curve uncertainties: general framework and key challenges. *Advances in Water Resources*, *105*, 51–66.

Sikorska, A. E., Scheidegger, A., Banasik, K., & Rieckermann, J. (2013). Considering rating curve uncertainty in water level predictions. *Hydrology and Earth System Sciences*, *17*, 4415–4427.

Sorooshian, S., & Dracup, J. A. (1980). Stochastic parameter estimation procedures for hydrologic rainfall-runoff models: Correlated and heteroscedastic error cases. *Water Resources Research*, *16*(2), 430–442.

Stroock, D. W. (1999). *A concise introduction to the theory of integration* (3rd ed.). Basel, Boston: Birkhäuser.

Sukys, J., & Kattwinkel, M. (2018). SPUX: Scalable Particle Markov Chain Monte Carlo for uncertainty quantification in stochastic models. In S. E. A. Bsssini (Ed.), *Parallel computing is everywhere* (pp. 159–168). Amsterdam, The Netherlands: IOS Press.

Suweis, S., Bertuzzo, E., Botter, G., Porporato, A., Rodriguez-Iturbe, I., & Rinaldo, A. (2010). Impact of stochastic fluctuations in storage-discharge relations on streamflow distributions. *Water Resources Research*, *46*, W03517.

Thyer, M., Renard, B., Kavetski, D., Kuczera, G., & Clark, M. (2011). Improving hydrological model predictions by incorporating rating curve uncertainty. In *Proceedings of the 34th iahr world congress*.

Tomassini, L., Reichert, P., Künsch, H. R., Buser, C., Knutti, R., & Borsuk, M. E. (2009). A smoothing algorithm for estimating stochastic, continuous-time model parameters and its application to a simple climate model. *Journal of the Royal Statistical Society C: Applied Statistics*, *58*, 679–704.

Uhlenbeck, G. E., & Ornstein, L. S. (1930). On the theory of Brownian motion. *Physical Review*, *36*, 823–841.

Van Leeuwen, P. J., Künsch, H.-R., Nerger, L., Potthast, R., & Reich, S. (2019). Particle filters for high-dimensional geoscience applications: A review. *Quarterly Journal of the Royal Meteorological Society*, *145*(723), 2335–2365.

Vrugt, J. A., ter Braak, C. J. F., Clark, M. P., Hyman, J. M., & Robinson, B. A. (2008). Treatment of input uncertainty in hydrologic modeling: Doing hydrology backward with Markov chain Monte Carlo simulation. *Water Resources Research*, *44*, W00B09.

Vrugt, J. A., ter Braak, C. J. F., Diks, C. G. H., & Schoups, G. (2013). Hydrologic data assimilation using particle Markov chain Monte Carlo simulation: Theory, concepts and applications. *Advances in Water Resources*, *51*, 457–478.

Wagener, T., McIntyre, N., Lees, M. J., Wheater, H. S., & Gupta, H. V. (2003). Towards reduced uncertainty in conceptual rainfall-runoff modeling: Dynamic identifiability analysis. *Hydrological Processes*, *17*, 455–476.