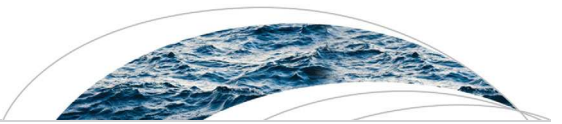
AGU PUBLICATIONS

Water Resources Research



RESEARCH ARTICLE

10.1002/2015WR017871

Describing the catchment-averaged precipitation as a stochastic process improves parameter and input estimation

Dario Del Giudice[1,2,3], Carlo Albert[1], Jörg Rieckermann[1], and Peter Reichert[1,4]

[1]Eawag, Swiss Federal Institute of Aquatic Science and Technology, Dübendorf, Switzerland, [2]Institute of Environmental Engineering, ETH Zurich, Swiss Federal Institute of Technology, Zurich, Switzerland, [3]Now at Department of Global Ecology, Carnegie Institution for Science, Stanford, California, USA, [4]Institute of Biogeochemistry and Pollutant Dynamics, ETH Zurich, Swiss Federal Institute of Technology, Zurich, Switzerland

Key Points:
- We develop a novel error model to infer the whole-catchment precipitation
- Parameter and input estimates are more realistic than with rainfall multipliers
- This way to assess and reduce input data errors supports uncertainty separation

Supporting Information:
- Supporting Information S1

Correspondence to:
D. Del Giudice,
ddelgiudice@carnegiescience.edu

Citation:
Del Giudice, D., C. Albert, J. Rieckermann, and P. Reichert (2016), Describing the catchment-averaged precipitation as a stochastic process improves parameter and input estimation, *Water Resour. Res.*, *52*, 3162–3186, doi:10.1002/2015WR017871.

Received 18 JUL 2015
Accepted 25 FEB 2016
Accepted article online 1 MAR 2016
Published online 29 APR 2016

Abstract Rainfall input uncertainty is one of the major concerns in hydrological modeling. Unfortunately, during inference, input errors are usually neglected, which can lead to biased parameters and implausible predictions. Rainfall multipliers can reduce this problem but still fail when the observed input (precipitation) has a different temporal pattern from the true one or if the true nonzero input is not detected. In this study, we propose an improved input error model which is able to overcome these challenges and to assess and reduce input uncertainty. We formulate the average precipitation over the watershed as a stochastic input process (SIP) and, together with a model of the hydrosystem, include it in the likelihood function. During statistical inference, we use "noisy" input (rainfall) and output (runoff) data to learn about the "true" rainfall, model parameters, and runoff. We test the methodology with the rainfall-discharge dynamics of a small urban catchment. To assess its advantages, we compare SIP with simpler methods of describing uncertainty within statistical inference: (i) standard least squares (LS), (ii) bias description (BD), and (iii) rainfall multipliers (RM). We also compare two scenarios: accurate versus inaccurate forcing data. Results show that when inferring the input with SIP and using inaccurate forcing data, the whole-catchment precipitation can still be realistically estimated and thus physical parameters can be "protected" from the corrupting impact of input errors. While correcting the output rather than the input, BD inferred similarly unbiased parameters. This is not the case with LS and RM. During validation, SIP also delivers realistic uncertainty intervals for both rainfall and runoff. Thus, the technique presented is a significant step toward better quantifying input uncertainty in hydrological inference. As a next step, SIP will have to be combined with a technique addressing model structure uncertainty.

1. Introduction

One of the main sources of uncertainty in hydrological modeling are input errors. These are predominantly associated with errors in the estimation of the true precipitation (here used interchangeably with rainfall) over a watershed [*Kavetski et al.*, 2006; *Vrugt et al.*, 2008]. Hydrological systems are indeed heavily input-driven and inaccuracies in rainfall characterization can dramatically impair the quality of calibration results and model output [*Bardossy and Das*, 2008].

Rainfall input errors affecting model calibration arise for a variety of reasons: inadequate areal coverage of point-scale pluviometers, inexact spatial interpolation, mechanical limitation of the gauge, wind effects, etc. [*McMillan et al.*, 2011; *Renard et al.*, 2011]. Furthermore, precipitation provided at an insufficient temporal resolution can substantially impair the model's ability to represent runoff, especially for small and fast-reacting catchments [*Ochoa-Rodriguez et al.*, 2015].

In studies focusing on precipitation simulation, several statistical models have been proposed to determine the time evolution of precipitation intensity (or rain rate) and its uncertainty [e.g., *Rodriguez-Iturbe et al.*, 1987; *Cowpertwait et al.*, 1996; *Deidda et al.*, 1999; *Paschalis et al.*, 2013; *Langousis and Kaleris*, 2014]. Results of those stochastic weather generators can provide (uncertain) inputs to rainfall-runoff models, thus helping to assess the influence of rainfall errors on runoff predictive uncertainty. However, as discussed e.g., by *Sikorska et al.* [2012], in studies focusing on hydrological model calibration and probabilistic prediction, input uncertainty has largely been neglected in the statistical inference process. This is probably due to the

computational complexity of including it in a likelihood function [*Honti et al*., 2013]. The likelihood (function) is the probability density of observations given the values of model parameters and inputs. The assumption is often that the observations are generated by the underlying model, consisting of a deterministic part describing the system, and a stochastic part describing the errors. This function is needed to extract information about model parameters and input from observed data. To make correct inference, the likelihood should consider all relevant mechanisms and error contributions. However, as discussed, e.g., by *Yang et al*. [2008], *Sikorska et al*. [2012], and *Reichert and Schuwirth* [2012], likelihood functions formulated as uncorrelated normal distributions and centered at the outputs of a deterministic model are still frequently used. Due to input errors and/or structural deficits of the model, this assumption is usually unrealistic [*Yang et al*., 2007b]. Independent and identically distributed (iid) normal likelihoods have repeatedly been shown to produce biased estimates of model parameters and unreliable predictions [*Renard et al*., 2011; *Honti et al*., 2013; *Del Giudice et al*., 2015b].

One alternative to the iid uncertainty description is the use of autoregressive error models [*Kuczera*, 1983; *Yang et al*., 2007a]. Although these likelihoods are still simple, they implicitly acknowledge the existence of errors besides the random output measurement noise (including inaccuracies in the input estimation). The effects of these errors on model output have been described by autocorrelated stochastic processes added to the model output [*Frey et al*., 2011; *Evin et al*., 2013]. In recent studies focusing on reliable runoff predictions, (iid) observation errors have been explicitly considered in addition to the "bias process," describing correlated deviations [*Reichert and Schuwirth*, 2012; *Del Giudice et al*., 2013; *Dietzel and Reichert*, 2014]. While likelihoods describing bias are more plausible than those assuming iid errors, they still have some limitations: (i) they can only provide limited information about the causes of model bias and, therefore, do not help much to disentangle input from structural errors; (ii) they can only partially buffer the corruption of model parameter estimates; (iii) they do not contribute to quantifying the uncertainty of unobserved variables (such as water level in an arbitrary point of the drainage network) [*Reichert and Mieleitner*, 2009; *Del Giudice et al*., 2015a].

A more satisfying approach for considering input errors is to make the input uncertain and to propagate it through the model [*Honti et al*., 2013; *McMillan et al*., 2011]. A simple way of doing so, which has become popular in hydrology, is the use of so-called rainfall multipliers [*Kavetski et al*., 2006; *Sun and Bertrand-Krajewski*, 2013]. These are event-specific random variables multiplied with the observed rain to provide the input to the model. These multipliers and their uncertainty are then estimated jointly with the other model parameters to correct for possible rainfall input errors during the calibration period. Most of these investigations in statistical model calibration have focused on the temporal dynamics of precipitation rather than its spatial variability, thus using the whole-catchment precipitation as input for a lumped runoff model. For specific applications in distributed modeling, one multiplier per grid point or subcatchment (and event) has been considered [*Salamon and Feyen*, 2010; *Li et al*., 2012].

While using rainfall multipliers is relatively straightforward, they have important drawbacks: multipliers do not provide a realistic assessment of input uncertainty if, for example, the temporal dynamics, i.e., the "shape," of a recorded storm event is significantly different from the true precipitation dynamics, or if a storm bypasses the pluviometric stations so that they do not record any precipitation although the catchment shows a runoff response [*Kavetski et al*., 2006; *Renard et al*., 2011]. While the first disadvantage can be reduced with multipliers varying within the storm event [*Reichert and Mieleitner*, 2009], the second one cannot be solved within this framework and requires a fresh approach.

In this study, we therefore suggest a novel input uncertainty model that describes the input of a hydrosystem as a continuous stochastic process. This makes it possible to formulate a more realistic likelihood function than those discussed above. This also allows us to learn about and reduce input as well as output uncertainties. By means of Bayesian inference, we show how to update our prior beliefs about parameters and rainfall patterns from the simultaneous use of input data (here: from pluviometers), output data (from a flowmeter at the outlet of the catchment), the runoff model (a lumped linear reservoir), a rainfall model (a transformed Gauss-Markov process), and models of the input and output observation errors (both normal distributions). We name this method Stochastic Input Process, SIP.

SIP, compared to the previous methods, has the following benefits:

1. It can probabilistically estimate the true input to a system in cases of sparse, inaccurate, or imprecise input measurements, if the output measurements and the underlying model are comparably accurate. This can be valuable to reconstruct past precipitation records from flow data or to spatially upscale point measurements. Reliable precipitation estimates can also be very useful to test hydrological theories and benchmark recordings from other sensors like radars [*Kirchner*, 2009].
2. It can reduce the bias in inferred parameters of hydrological models and therefore in runoff predictions. This can substantially support regionalization studies, which try to establish relations between hydrological model parameters calibrated in gauged catchments and properties of these catchments [*Kavetski et al*., 2006].
3. It can produce not only a reliable assessment of total output uncertainty, but also quantify the contributions due to parameter and input uncertainty. Supporting uncertainty separation, SIP can help assess to what extent prediction uncertainty can be reduced by providing better rainfall data and can therefore guide our efforts to minimize the uncertainty sources [*Sikorska et al*., 2012].

We examine the ability of our approach to produce realistic posterior parameter estimates and reliable predictions, the two pivotal features of "Bayesian data assimilation" [*Rougier*, 2013]. We then compare SIP with the three methods mentioned above: the simple least squares (LS) formulation assuming iid errors, an autoregressive bias description (BD), and the event-dependent rainfall multiplier (RM) error model. As an illustrative example, we perform inference and prediction for a monitored urbanized watershed which we model with a parsimonious hydrological model of combined waste and rainwater discharge.

2. Method

We briefly describe the LS, BD, and RM approaches, three commonly used techniques to calibrate and predict with environmental and, specifically, rainfall-runoff models. All techniques are implemented in a Bayesian framework, meaning that the likelihood function is combined with a prior distribution of the parameters to obtain posterior parameter estimates from observations. This allows us to make use of our existing knowledge of physical and error model parameters. Prior knowledge can reduce the identifiability problem between the process-based model and the statistical error model [*Bayarri et al*., 2007]. After this review, we explain more in depth the concepts and numerics of the method developed in this paper that is based on describing rainfall as a Stochastic Input Process (SIP). Finally, we discuss the case study we selected to demonstrate the usefulness of the SIP calibration scheme in the presence of important input errors. A graphical comparison of the methods is provided in Figure 1.

2.1. Alternative Methods Used for Comparison

2.1.1. Standard Least Squares: LS Method

The standard nonlinear least squares methodology is the simplest of the four approaches. This regression method describes the residual errors, i.e., the differences between the output of a deterministic model, $\mathbf{y}_M$, and observations, $\mathbf{y}_o$, as normally and identically distributed and independent (Figure 1). The implicit assumption here is that model results deviate from data only because of random observation errors, **E**. The other error sources, like input and structural errors, are neglected or somehow considered to have the same effect as white observation noise [*Vrugt et al*., 2008; *McMillan et al*., 2011]. Relaxing the assumptions of constant variance and normality of the residual errors via output transformation can make the LS approach slightly more appropriate for hydrological applications [*Wang et al*., 2012]. The probabilistic model used for inference and prediction in this approach is:

$$f(\mathbf{y}_o\,|\,\boldsymbol{\theta},\boldsymbol{\psi}_y,\mathbf{x}) = \frac{(2\pi)^{-q/2}}{\sqrt{\det\left(\boldsymbol{\Sigma}(\boldsymbol{\psi}_y)\right)}} \cdot \exp\left(-\frac{1}{2}[g(\mathbf{y}_o)-g(\mathbf{y}_M(\boldsymbol{\theta},\mathbf{x}))]^{\mathrm{T}}\boldsymbol{\Sigma}(\boldsymbol{\psi}_y)^{-1}[g(\mathbf{y}_o)-g(\mathbf{y}_M(\boldsymbol{\theta},\mathbf{x}))]\right)\prod_{i=1}^{q}\frac{\mathrm{d}g}{\mathrm{d}y}(y_{o,i})\,, \tag{1}$$

where $\boldsymbol{\theta}$ is the vector of hydrological model parameters, $\boldsymbol{\psi}_y$ is the vector of parameters of the error term, $\mathbf{x}$ is the time-varying model input (in our case precipitation), q is the number of elements of $\mathbf{y}_o$ and $\mathbf{y}_M$, and g is the scalar transformation function, which, when applied to a vector, returns the vector of the function applied to all components of its argument. For the LS method, the input used in equation (1) is given by the observed input (ignoring input uncertainty),

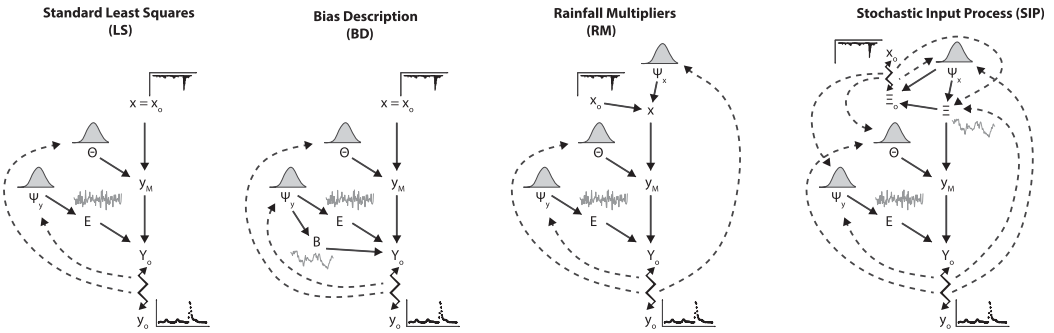


Figure 1. Representation of the four methods we use to describe the uncertainties during inference and predictions (section 2). Greek letters denote (random) calibration parameters of the hydrological model, $\boldsymbol{\Theta}$, input, $\boldsymbol{\Psi}_x$, and output error model, $\boldsymbol{\Psi}_y$ (more details in Table 1). The random processes **E**, **B**, and **Ξ** have different illustrations depending on their consideration of autocorrelation. The dashed lines exemplify the learning process of inference while the solid lines illustrate the information flow during predictions. The zigzags typify the comparison between modeled and measured time series during inference. While the RM implicitly accounts for input uncertainty by inferring additional parameters, SIP explicitly considers it in the likelihood function.

$$\mathbf{x}=\mathbf{x}_o \tag{2}$$

and the covariance matrix $\boldsymbol{\Sigma}(\boldsymbol{\psi}_y)$ is given by the diagonal matrix

$$\boldsymbol{\Sigma}_{ij}(\boldsymbol{\psi}_y)=\delta_{ij}\sigma_E^2\,, \tag{3}$$

where i and j are subscripts running from 1 to q, and δ represents the Kronecker delta. The results of the deterministic model are represented by $\mathbf{y}_M(\boldsymbol{\theta},\mathbf{x})$, whereas the corresponding observed values are denoted by $\mathbf{y}_o$. We here consider the heteroscedasticity of the errors (i.e., the dependence of the error variance on the corresponding model output) via an output transformation, g, whose functional form is given in Appendix A. The only parameter $\boldsymbol{\psi}_y$ of this error model is σ_E, the (constant) standard deviation of the output measurement noise in transformed units. Several studies have shown that this error description is too simple for complex environmental systems where input and structural errors play an important role [e.g., *Yang et al*., 2007a,2007b; *Vrugt et al*., 2008; *Renard et al*., 2011; *Reichert and Schuwirth*, 2012; *Honti et al*., 2013]. However, we chose the LS method as a benchmark for comparison with conceptually more satisfactory methods because it is still widely applied in environmental modeling.

2.1.2. Statistical Bias Description: BD Method

A way to consider the effects of input errors and structural deficits on model output is to mimic the systematic deviations of model results from data with an autocorrelated stochastic process. This approach was introduced in hydrology decades ago [*Sorooshian and Dracup*, 1980; *Kuczera*, 1983]. The version we use, however, besides modeling the autocorrelated process **B**, which represents the effects of input and structural errors, additionally includes an uncorrelated process **E**, representing the observation noise. [*Kennedy and O'Hagan*, 2001; *Bayarri et al*., 2007; *Reichert and Schuwirth*, 2012; *Brynjarsdóttir and O'Hagan*, 2014], As adequate information on measurement precision is usually available, and **B** and **E** have different properties, their identifiability is typically high. However, there is an identifiability problem between **B** and the parameters of the deterministic model which can be resolved by defining appropriate prior distributions [*Reichert and Schuwirth*, 2012]. The bias correction **B** here follows an Ornstein-Uhlenbeck (OU) dynamics [e.g., *Platen and Bruti-Liberati*, 2010; *Kroese et al*., 2011, and references therein]

$$\mathrm{d}B(t)=-\frac{B(t)}{\tau}\mathrm{d}t+\sqrt{\frac{2}{\tau}}\sigma_B\mathrm{d}W(t), \tag{4}$$

where τ is the correlation time and σ_B is the asymptotic standard deviation of the statistical fluctuations around the average value of **B**, here 0. $W(t)$ is a Wiener process, also called standard Brownian motion, or random walk with independent Gaussian increments. The first part of the (Langevin) equation (4) describes a deterministic dampening, central-restoring force, or pull towards the long-run mean of zero. The second term counterbalances this tendency by adding stochastic white noise. This leads to random oscillations of realizations of this process around the equilibrium state with standard deviation σ_B and correlation time τ. We chose an OU process, because it is time-continuous, linear, Markovian, has a finite stationary variance,

and can be integrated analytically [*Ibe*, 2009; *Paul and Baschnagel*, 2013]. Using Gaussian error models in combination with an appropriate transformation not only keeps the inference relatively simple due to their analytical properties, but has additionally been proven to appropriately describe errors in runoff modeling [*Honti et al*., 2013; *Del Giudice et al*., 2013]. Similar output error models might be appropriate as well, although care has to be taken to avoid overparameterization [*Evin et al*., 2013].

In the BD approach, the likelihood function has the same basic form as in equation (1), but the covariance matrix is nondiagonal:

$$\Sigma_{ij}(\psi_y)=\sigma_B^2 e^{-\tau^{-1}|t_i-t_j|}+\delta_{ij}\sigma_E^2. \tag{5}$$

In this equation, i and j are subscripts spanning over the time domain, and τ and σ_B are the (hyper)parameters, ψ_y, of the Gaussian bias process. As the effect of input errors is corrected at the output, this technique is based on using the observed input (2) when applying the likelihood function (1) with (5).

2.1.3. Multiplicative Rainfall Error Model: RM Method

The RM (rainfall multiplier) approach explicitly considers input (in our case rainfall) uncertainty by perturbing the observed precipitation time series with independent random factors for all storm events [*Kavetski et al*., 2006; *McMillan et al*., 2011]. To make the inference tractable, these (latent) factors are kept constant during an event [*Vrugt et al*., 2008]. We then apply the likelihood function (1) with (3) and with the perturbed input

$$x_i=\beta_{j(i)}x_{o,i}, \tag{6}$$

where the index i runs through all elements of the rainfall time series, whereas the index j remains constant for all values of i within any given storm event. The parameters $\boldsymbol{\beta}=(\beta_1,\ldots,\beta_{n_s})$ represent the rainfall bias corrections for all (n_s) storm events. The priors for the elements of $\boldsymbol{\beta}$ are formulated as lognormal distributions centered at 1 and with a joint standard deviation σ^β. Centering these distributions at 1 implies assuming the observed input to be correct. Inferring σ^β leads to a hierarchical parameter estimation problem. While some applications of RM kept σ^β fixed [*Sun and Bertrand-Krajewski*, 2013], making the error model nonhierarchical, we prefer to infer this hyperparameter to learn about the overall input variance detected during calibration [*Li et al*., 2012; *Sikorska et al*., 2012]. Despite using the same likelihood function as for the LS method, the replacement of the input description (2) by (6) leads to the consideration of input uncertainty at the level of whole storm events and augments the parameter vector with the parameters $\boldsymbol{\Psi}_x=\{\boldsymbol{\beta},\sigma^\beta\}$ of the rainfall error model. The basic difference between LS and RM is that, while the former assumes the observed rainfall to be the true input, the latter considers sections of the true input to be unknown multiples of the recorded precipitation during that time period (Figure 1). While the RM technique provides a simple approximation for the uncertainty of the rainfall volumes, its limited ability to deal with strongly dynamic input errors has been widely acknowledged [*Kavetski et al*., 2006; *Vrugt et al*., 2008; *Sikorska et al*., 2012]. Consequently, a more realistic statistical representation of the catchment-averaged precipitation is needed [*Salamon and Feyen*, 2010; *Renard et al*., 2011].

2.2. Joint Inference of Input, Hydrological Model, and Output Error Parameters: SIP Method

2.2.1. Overall Concept

The framework we propose to quantify and propagate input uncertainty is based on the inference of a latent Gauss-Markov stochastic process, the "rainfall potential," ξ [*Sigrist et al*., 2012]. This time series can be transformed to the areal average watershed precipitation. As in previous studies developing techniques to consider input uncertainty in hydrological inference [e.g., *Kavetski et al*., 2006; *Vrugt et al*., 2008; *Reichert and Mieleitner*, 2009; *Kirchner*, 2009; *Sikorska et al*., 2012], we concentrate on the time evolution of precipitation rather than on its spatial variability. We thus develop a stochastic rainfall model which satisfies two requirements: (a) its realizations should have some statistical properties of the real rainfall and (b) the conditional probabilities (needed e.g., in equations (19) and (20)) are available in analytical form. Modeling the rainfall potential ξ as a normal, linear, and Markovian process simplifies the calculation of conditional probabilities, thus making inference easier. Together with ξ, the parameters of the hydrological model, $\boldsymbol{\Theta}$, those of the input, $\boldsymbol{\Psi}_x$, and those of the output error model, $\boldsymbol{\Psi}_y$, are also inferred. Furthermore, similar to *Sigrist et al*. [2012], we simultaneously estimate ξ_o, the "rainfall potential" at the pluviometric station. The rainfall potential over the catchment, ξ, is only inferred indirectly through both

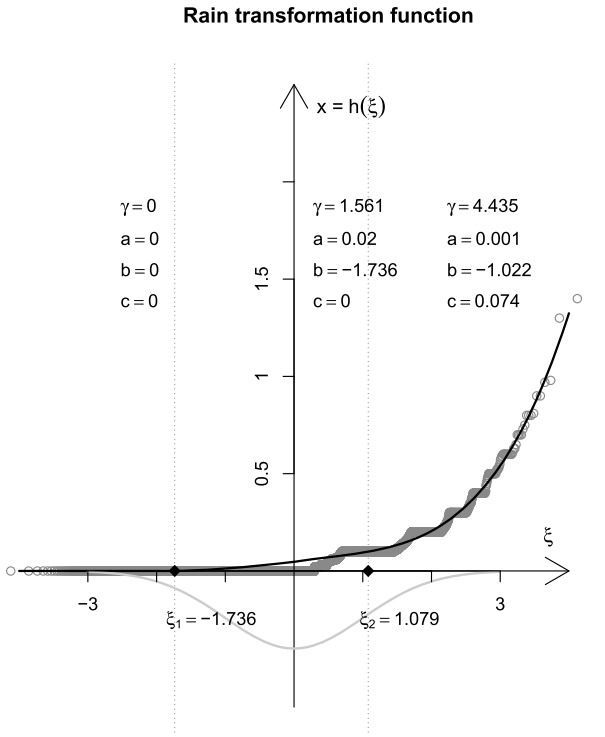


Figure 2. Empirical function to transform a normally distributed value into a rain value (see equation (11)). The data used to parameterize the function are sorted according to their value. The values of b and c for $\xi > \xi_2$ were derived from continuity constraints for h and $\frac{dh}{d\xi}$ at ξ_2. The dotted vertical lines define three domains of the transformation whose parameters are indicated above. These regime switches have been empirically determined. Note that x, a, and c have units of mm/min.

measurements of rainfall at the observation site, $\mathbf{x}_o$, and of runoff, $\mathbf{y}_o$. Thus, for inference we need two models, one for the observed rainfall (section 2.2.4) and the other for observed runoff (section 2.2.4). The "rainfall potential" describes the rainfall in a given catchment or at a given site and is not meant as a potential in a physical sense. These "rainfall potentials" can be transformed to the associated rainfall by means of the scalar function h:

$$\mathbf{x}=h(\xi), \quad \mathbf{x}_o=h(\xi_o). \tag{7}$$

(again, the application of h to ξ is to be understood element-wise). As discussed by *Sigrist et al.* [2012] and *Ailliot et al.* [2015], transforming a latent Gaussian process is a parsimonious way to model occurrence and amount of precipitation. As multiple values of the "rainfall potentials" are mapped to zero precipitation, the function h is not invertible (Figure 2). Therefore, we avoid denoting the elements of ξ and ξ_o the "transformed rainfall." However, whenever the rainfall intensities are not zero, and thus h^{-1} exists, they are transformed rainfall intensities.

The two main differences between the suggested technique and RM are:

1. The SIP technique does not assume (pieces of) the true precipitation to be proportional to the observed time series. Instead, our knowledge of true input is inferred from prior knowledge, input observations, and output observations. This makes it possible to deal with time-varying observation errors of the rain rate and with unrecorded storms that bypassed the observation site but led to a runoff increase, a situation intractable with rainfall multipliers. These features make the suggested technique conceptually more satisfying than the techniques described in section 2.1.
2. The joint input and output likelihood function of SIP does not have a simple explicit form as for the RM (equation (1)), but is instead given in a discretized form of a high-dimensional integral over all possible realizations of ξ and ξ_o (equation (8)). Unfortunately, this makes the suggested technique computationally more demanding than all three techniques used to compare with.

The SIP likelihood function can be written as:

$$f(\mathbf{y}_o, \mathbf{x}_o|\boldsymbol{\theta}, \boldsymbol{\psi}_y, \boldsymbol{\psi}_x) = \int f(\mathbf{y}_o|\boldsymbol{\theta}, \boldsymbol{\psi}_y, \mathbf{x}=h(\xi)) f(\mathbf{x}_o|\xi_o) f(\xi_o|\xi, \boldsymbol{\psi}_x) f(\xi|\boldsymbol{\psi}_x) \mathrm{d}\xi \mathrm{d}\xi_o \,, \tag{8}$$

where integration is over all possible discretized time series of ξ and ξ_o. This formulation is the discretized version of what in physics is called path integral (for an application in the environmental sciences see, e.g., *Quinn and Abarbanel* [2010]).

In the following, we describe the factors of the integrand above. $f(\mathbf{y}_o|\boldsymbol{\theta}, \boldsymbol{\psi}_y, \mathbf{x})$ is the likelihood of observed output given the parameters of the hydrological model, $\boldsymbol{\theta}$, the parameters of the output error model, $\boldsymbol{\psi}_y$, and the rainfall input, $\mathbf{x}$; $f(\mathbf{x}_o|\xi_o)$ is the model of observed input, $\mathbf{x}_o$, given the input potential at the observation site, ξ_o; $f(\xi_o|\xi, \boldsymbol{\psi}_x)$ is the model for the rainfall potential at the observation site, ξ_o, given the rainfall potential for the whole catchment, ξ, and the input model parameters, $\boldsymbol{\psi}_x$; and $f(\xi|\boldsymbol{\psi}_x)$

is the a priori model for the rainfall potential for the whole catchment, $\boldsymbol{\xi}$, given the input model parameters, $\boldsymbol{\psi}_x$. Subsequently, we describe these models in more detail, the numerical method to implement inference of parameters and time series of rainfall potential, and, finally, we elucidate how to make predictions with SIP.

2.2.2. Output Probabilistic Model

The term $f(\mathbf{y}_o|\boldsymbol{\theta},\boldsymbol{\psi}_y,\mathbf{x})$ represents the probabilistic hydrological model for the observed discharge, $\mathbf{y}_o$, as a function of the rainfall, $\mathbf{x}$, the parameters of the hydrological model, $\boldsymbol{\theta}$, and those of the output error model, $\boldsymbol{\psi}_y$. This probability density function is assumed to be given by equation (1) with the covariance matrix given by equation (3) as for the LS and RM approaches. The difference is, again, the representation of the input. The LS approach assumes the observed input to be error-free (equation (2)). The RM approach assumes the true input to be a piecewise-scaled version of the observed one (equation (6)). In our new approach, we infer the rainfall potential, $\boldsymbol{\xi}$, and the corresponding input, $\mathbf{x}$, jointly with the parameters of the hydrological model and the error models using rainfall (input) and runoff (output) data. In some sense, we use the output of the catchment as an additional rain gauge, to gain spatially integrated information about the catchment-averaged precipitation.

Similarly to the RM approach, the use of the probability density function (1) with the covariance matrix given by equation (3) assumes that model structural deficits are negligible and can be "absorbed" by parameter uncertainty. This means that we assume that the systematic deviations of model output from observations are dominated by problems in acquiring the input with sufficient accuracy. For the simple hydrosystem under study, this assumption is very plausible (see section 4). However, it would be possible to use an output model that accounts for the effect of structural errors (see section 5.4).

2.2.3. Prior Rainfall Model

We base the description of the prior rainfall model on a "rainfall potential," $\boldsymbol{\xi}$, that follows an Ornstein-Uhlenbeck process with mean zero, asymptotic standard deviation unity, and correlation time τ_ξ, from which we get the distribution of rainfall intensity by a transformation, h: $\mathbf{x}=h(\boldsymbol{\xi})$. In continuous-time formulation, the rainfall potential then follows the stochastic differential equation

$$\mathrm{d}\xi(t)=-\frac{\xi(t)}{\tau_\xi}\mathrm{d}t+\sqrt{\frac{2}{\tau_\xi}}\mathrm{d}W(t), \tag{9}$$

which is solved by a Gaussian process with conditional expectation and variance given by

$$\mathrm{E}[\xi(t)|\xi(t_0)]=\xi(t_0)\exp\left(\frac{t_0-t}{\tau_\xi}\right),\quad \mathrm{Var}[\xi(t)|\xi(t_0)]=1-\exp\left(-2\frac{t-t_0}{\tau_\xi}\right). \tag{10}$$

The discrete time series, $\boldsymbol{\xi}$, used for our model, consists of an evaluation of this process for a discrete set of time points. The resulting probability density, $f(\boldsymbol{\xi}|\boldsymbol{\psi}_x)$, describes our prior knowledge of the rainfall potential time series at catchment scale during rainy periods. Time modeling of precipitation as a simple hidden (i.e., latent or potential) and censored (where an observation only becomes available when a threshold is exceeded) stochastic process to be transformed into precipitation and updated using data is well known in statistical meteorology [*Ailliot et al.*, 2015].

In our case study, we estimated the parameterization of the transformation h and the choice of the parameter values from a long precipitation time series with few zeros and including the most intense rain events recorded in 2013 in the area (Figure 2). Similar to *Sigrist et al.* [2012], we chose a power function, but we differentiated its coefficients for three rain intensity intervals (or regimes): no rain, light rain, and heavy rain. The coefficients were constrained to guarantee the differentiability of h over the full range of its argument. This led to:

$$\begin{aligned} x&=h(\xi)=a(\xi-b)^\gamma+c,\\ \xi&=h^{-1}(x)=b+\left(\tfrac{x-c}{a}\right)^{1/\gamma}\quad \text{if } a\neq 0,\\ \frac{\mathrm{d}h}{\mathrm{d}\xi}&=a\gamma(\xi-b)^{\gamma-1}, \end{aligned} \tag{11}$$

with three sets of parameters for the intervals] $-\infty$, ξ_1], [ξ_1, ξ_2], and [ξ_2, ∞ [(see Figure 2). Note that for $\xi\leq\xi_1$, $a=c=0$, so that h is only invertible for $\xi>\xi_1$. The correlation time τ_ξ and the parameters of the

transformation h are used to formulate our prior knowledge about the process ξ. Thus, they are not included in Bayesian inference. However, the actual time course of ξ is inferred.

As will be shown later, the properties of this Ornstein-Uhlenbeck process are convenient to efficiently sample from the posterior (section 2.2.5). Furthermore, these types of censored power-transformed Gaussian models have few parameters, are analytically tractable, and have shown satisfactory performances in precipitation generation [*Sigrist et al.*, 2012; *Ailliot et al.*, 2015]. The parameterization chosen makes this prior stochastic model particularly suitable for typical rainy periods. Thanks to the Bayesian framework, however, this model can be updated to correctly reproduce a wide range of calibration events [*Renard et al.*, 2011]. Alternative stochastic models for precipitation [e.g., *Paschalis et al.*, 2013] could also be used. For instance, if the analysis of extreme events is of particular interest, more complicated multifractal models randomly simulating the cascade of rainfall across scales could be envisaged [e.g., *Deidda et al.*, 1999]. Incorporating those models in hydrological inference, however, would make the calculation of conditional probabilities difficult and thus would require a different numerical approach.

2.2.4. Model for Rainfall Observations

As mentioned in the introduction, the true input to the catchment, $\mathbf{x}$, differs from the observed input, $\mathbf{x}_o$, mainly due to sampling errors caused by insufficient gauge coverage and/or the imperfect spatial interpolation scheme between gauges [*McMillan et al.*, 2011]. Similar to *Sigrist et al.* [2012], our error model for input observation errors related to the distance and intrinsic inaccuracy of pluviometers is multivariate normal in the space of the "rainfall potential":

$$f(\xi_o|\xi,\psi_x)=\frac{1}{(2\pi)^{n_{\mathbf{x}_o}/2}}\frac{1}{\sqrt{\det(\Sigma_{\xi_o}(\psi_x))}}\exp\left(-\frac{1}{2}(\xi_o-\xi)^{\mathrm{T}}\Sigma_{\xi_o}(\psi_x)^{-1}(\xi_o-\xi)\right). \tag{12}$$

Due to insignificant correlations found in long time series of reference data sets (supporting information Figure S1), we parametrized the covariance matrix in equation (12) as:

$$\Sigma_{\xi_o,ij}(\psi_x)=\delta_{ij}\sigma_\xi^2, \tag{13}$$

where σ_ξ^2 is the variance characterizing the deviation in rainfall potential between the observation site and the input to the catchment (supporting information Figure S2). The larger σ_ξ, the less accurately the observed rainfall represents the true rainfall at catchment scale. While we expect an increase in σ_ξ^2 with increasing distance from the catchment, other factors, such as topography and wind direction, can influence the value of σ_ξ^2 as well. Additionally, while we use rain gauge data to estimate the areal average catchment precipitation, unlike in *Bardossy and Das* [2008], this does not involve spatial interpolation but rather a Bayesian updating of the prior input process for the whole catchment $\boldsymbol{\xi}$ via assimilation of rainfall $\mathbf{x}_o$ and discharge data $\mathbf{y}_o$. Our strategy is similar to using hidden Markov models where the a priori parameterization of the weather state (here: precipitation) is optimally fitted/conditioned to the data [*Ailliot et al.*, 2015].

Finally, in our model for rainfall observations, the probability distribution of rainfall given the rainfall potential is given by

$$f(\mathbf{x}_o|\xi_o)=\delta(\mathbf{x}_o-h(\xi_o)). \tag{14}$$

This Dirac function represents the transformation from rainfall potential into actual rain rate at the measurement station.

2.2.5. Numerical Implementation of Inference With SIP

Bayesian updating of the prior process $\boldsymbol{\Xi}$ and distributions $\boldsymbol{\Theta}$, $\boldsymbol{\Psi}_y$, and $\boldsymbol{\Psi}_x$ are based on the Markov chain Monte Carlo (MCMC) scheme proposed by *Tomassini et al.* [2009]. In particular, we adopt a Metropolis-within-Gibbs algorithm, which sequentially samples different conditional distributions while keeping the other parameters or process realizations constant (supporting information Figure S3). Using the index k for the elements of these Markov chains, we sequentially generate the elements $k+1$ of the different chains as outlined below. The starting point for the iterations can be obtained by drawing a vector of parameters and a realization of the input processes from the prior distribution. The pseudocode of this algorithm is as follows:

1. **Sample** a new point of the Markov chain of the hydrological model and **output** error **parameters**, $(\boldsymbol{\theta}^{k+1}, \boldsymbol{\psi}_y^{k+1})$, for the conditional distribution

$$f(\theta^{k+1}, \psi_y^{k+1}|\mathbf{y}_o, \mathbf{x}_o, \psi_x^k, \xi^k, \xi_o^k) = f(\theta^{k+1}, \psi_y^{k+1}|\mathbf{y}_o, \xi^k) \propto f(\mathbf{y}_o|\theta^{k+1}, \psi_y^{k+1}, h(\xi^k)) \cdot f(\theta^{k+1}, \psi_y^{k+1}) \quad (15)$$

using Metropolis sampling: draw a candidate point for $(\theta^{k+1}, \psi_y^{k+1})$ from $\mathcal{N}((\theta^k, \psi_y^k), \Sigma_y)$ as the proposal (or jump) distribution with covariance matrix Σ_y and accept or reject this candidate by the Metropolis rule using the density (15). This step requires running the deterministic model.

2. **Sample** a new point of the Markov chain of the **input** error model **parameters**, ψ_x^{k+1}, for the conditional distribution

$$f(\psi_x^{k+1}|\mathbf{y}_o, \mathbf{x}_o, \theta^{k+1}, \psi_y^{k+1}, \xi^k, \xi_o^k) = f(\psi_x^{k+1}|\xi^k, \xi_o^k) \propto f(\xi_o^k|\xi^k, \psi_x^{k+1}) \cdot f(\xi^k|\psi_x^{k+1}) \cdot f(\psi_x^{k+1}) \quad (16)$$

using Metropolis sampling: draw a candidate point for ψ_x^{k+1} from $\mathcal{N}(\psi_x^k, \Sigma_x)$ as the proposal distribution with covariance matrix Σ_x and accept or reject this candidate using the Metropolis rule using the density (16). The parameters of the Ornstein-Uhlenbeck process for ξ are mean zero and standard deviation unity. Furthermore, the correlation time (here: $\tau_\xi = 10.6$ min) was estimated from a long precipitation time series not used during inference. Therefore, in the current application, the density $f(\xi|\psi_x)$ does not depend on ψ_x and cancels for the rejection rate calculation. This step does not require any hydrological model run.

3. **Sample** a new element of the Markov chain of the **rainfall** potential time series **at the input observation site**, ξ_o^{k+1}, for the conditional distribution

$$f(\xi_o^{k+1}|\mathbf{y}_o, \mathbf{x}_o, \theta^{k+1}, \psi_y^{k+1}, \psi_x^{k+1}, \xi^k) = f(\xi_o^{k+1}|\mathbf{x}_o, \psi_x^{k+1}, \xi^k) \propto f(\mathbf{x}_o|\xi_o^{k+1}) f(\xi_o^{k+1}|\xi^k, \psi_x^{k+1}) . \quad (17)$$

For time indices, I, at which $x_{o,i} > 0$, we can directly calculate $\xi_{o,i}^{k+1} = h^{-1}(x_{o,i})$ since h^{-1} exists for arguments that are larger than zero. For time indices, I_0, at which $x_{o,i} = 0$, we sample $\xi_{o,i}^{k+1}$ from a normal distribution with mean $\xi_{I_0}^k + \Sigma_{\xi_o,I_0,I}\Sigma_{\xi_o,I,I}^{-1}(h^{-1}(\mathbf{x}_{o,I}) - \xi_I^k)$ and covariance matrix $\Sigma_{\xi_o,I_0,I_0} + \Sigma_{\xi_o,I_0,I}\Sigma_{\xi_o,I,I}^{-1}\Sigma_{\xi_o,I_0,I}^T$ that is truncated to values $\xi < \xi_1$.

4. **Sample** a new element of the Markov chain of the **rainfall** potential time series **for the whole catchment**, ξ^{k+1}, for the conditional distribution

$$f(\xi^{k+1}|\mathbf{y}_o, \mathbf{x}_o, \theta^{k+1}, \psi_y^{k+1}, \psi_x^{k+1}, \xi_o^{k+1}) = f(\xi^{k+1}|\mathbf{y}_o, \theta^{k+1}, \psi_y^{k+1}, \psi_x^{k+1}, \xi_o^{k+1}) \propto f(\mathbf{y}_o|\theta^{k+1}, \psi_y^{k+1}, h(\xi^{k+1})) f(\xi_o^{k+1}|\xi^{k+1}, \psi_x^{k+1}) f(\xi^{k+1}|\psi_x^{k+1}) . \quad (18)$$

This is similar to drawing a new element of the Markov chain of a time-dependent parameter [*Tomassini et al.*, 2009; *Reichert and Mieleitner*, 2009], with the difference that we condition not only on the parameters, (θ, ψ_y, ψ_x), and on the observed output, $\mathbf{y}_o$, but, in addition, on the rainfall potential of the observed input, ξ_o. In principle, we could draw a candidate realization from the Ornstein-Uhlenbeck process $f(\xi^{k+1}|\psi_x^{k+1})$, and use the two other factors in (18), $f(\mathbf{y}_o|\theta^{k+1}, \psi_y^{k+1}, h(\xi^{k+1}))$ and $f(\xi_o^{k+1}|\xi^{k+1}, \psi_x^{k+1})$, for the calculation of the rejection ratio of Metropolis sampling. However, in that case we would not profit from what we have learned in the past (up to step k) about the posterior of ξ. This would lead to a very low acceptance rate. To improve efficiency, we draw a realization only for pieces of the time series of ξ^{k+1}, keeping the remainder of the time series at their previous values, ξ^k, for the intervals that were not yet updated, or their new values, ξ^{k+1}, for the intervals that were already updated (supporting information Figure S4). This leads to the following algorithm for the construction of ξ^{k+1} from ξ^k (adapted from *Tomassini et al.* [2009]):

4.1 Divide the calibration period and the previous realization of the rainfall potential, ξ^k, into m subintervals of similar length (using random disturbances to prevent the interval boundaries from being the same in successive steps). Denote with $\xi_l^k = \xi^k|_{I_l}$ the restrictions of ξ^k to the subinterval I_l.

4.2 Repeat the following three substeps $\forall\, l = 1,\ldots, m$ subintervals I_l in order to draw ξ^{k+1}, a sample of the updated Ξ. The more substeps are considered, the more hydrological model runs will be required for one iteration.

4.2.1. Draw a candidate sample $\xi_l^{k+1\prime}$ over the subinterval I_l from an Ornstein-Uhlenbeck process conditional on the values of ξ^k at its start time point, s, and its end point, u. Updating the current sample over the subinterval I_l guarantees continuity of the process over the full time domain. Conditional mean and variance of this process are given by

$$\begin{aligned}\mathrm{E}[\Xi(t)|\xi(s),\xi(u)] = &\frac{\exp\left(-(t-s)/\tau_\xi\right)\left[1-\exp\left(-2(u-t)/\tau_\xi\right)\right]}{1-\exp\left(-2(u-s)/\tau_\xi\right)}\,\xi(s)\\ &+\frac{\exp\left(-(u-t)/\tau_\xi\right)\left[1-\exp\left(-2(t-s)/\tau_\xi\right)\right]}{1-\exp\left(-2(u-s)/\tau_\xi\right)}\,\xi(u),\end{aligned} \tag{19}$$

$$\mathrm{Var}[\Xi(t)|\xi(s),\xi(u)] = \frac{\left[1-\exp\left(-2(u-t)/\tau_\xi\right)\right]\left[1-\exp\left(-2(t-s)/\tau_\xi\right)\right]}{1-\exp\left(-2(u-s)/\tau_\xi\right)}\,. \tag{20}$$

Then, replace the current sample (that may have already been modified on previous intervals) by the candidate, $\xi_l^{k+1\prime}$, in the interval I_l. We denote this candidate sample over the full time domain by $\xi_{<l}^{k+1} \cup \xi_l^{k+1\prime} \cup \xi_{>l}^{k}$.

4.2.2. Compute the acceptance probability, r, of this candidate sample according to:

$$r = \min\left[1, \frac{f(\xi_o|\xi_{<l}^{k+1} \cup \xi_l^{k+1\prime} \cup \xi_{>l}^{k}, \psi_x^{k+1}) f(\mathbf{y_o}|\theta^{k+1}, \psi_y^{k+1}, h(\xi_{<l}^{k+1} \cup \xi_l^{k+1\prime} \cup \xi_{>l}^{k}))}{f(\xi_o|\xi_{<l}^{k+1} \cup \xi_l^{k} \cup \xi_{>l}^{k}, \psi_x^{k+1}) f(\mathbf{y}_o|\theta^{k+1}, \psi_y^{k+1}, h(\xi_{<l}^{k+1} \cup \xi_l^{k} \cup \xi_{>l}^{k}))}\right]. \tag{21}$$

This part requires running the hydrological model, which might be time consuming. In contrast to *Tomassini et al*. [2009], in this acceptance ratio, the rainfall observations are considered in the form of a probability density of the rainfall potential at the observation site, $f(\xi_o|\xi, \psi_x)$, in addition to the probability density for the observed output, $f(\mathbf{y}_o|\theta, \psi_y, h(\xi))$.

4.2.3. Set $\xi_l^{k+1} = \xi_l^{k+1\prime}$, i.e., accept $\xi_l^{k+1\prime}$, with probability r, otherwise set $\xi_l^{k+1} = \xi_l^{k}$, i.e., reject $\xi_l^{k+1\prime}$.

4.3. After having completed these m substeps, set $\xi^{k+1} = \xi_1^{k+1} \cup \ldots \cup \xi_m^{k+1}$ and move to the next iteration (step 1 above).

After having repeated these steps 1–4 of the MCMC algorithm to convergence, we obtain a sample of the joint posterior distribution and input processes $f(\theta, \psi_y, \psi_x, \xi, \xi_o|\mathbf{y}_o, \mathbf{x}_o)$. The posterior of the parameters only can be gained through marginalization: $f(\theta, \psi_y, \psi_x|\mathbf{y}_o, \mathbf{x}_o) = \int f(\theta, \psi_y, \psi_x, \xi, \xi_o|\mathbf{y}_o, \mathbf{x}_o)\mathrm{d}\xi\mathrm{d}\xi_o$. A sample from this distribution is obtained from the sample of $f(\theta, \psi_y, \psi_x, \xi, \xi_o|\mathbf{y}_o, \mathbf{x}_o)$ by disregarding the information on ξ and ξ_o.

2.3. Predictions in the Calibration and Validation Periods

Once having a statistically calibrated model, we are usually interested in quantifying our knowledge of the true system output, $\mathbf{y}$. This is done by calculating

$$\begin{aligned}&f(\mathbf{y}^{L_2}|\mathbf{y}_o^{L_1}, \mathbf{x}_o^{L_1 \cup L_2})\\ &= \int f(\mathbf{y}^{L_2}|\theta, \psi_y, \psi_x, \mathbf{y}_o^{L_1}, \mathbf{x}_o^{L_1 \cup L_2}) f(\theta, \psi_y, \psi_x|\mathbf{y}_o^{L_1}, \mathbf{x}_o^{L_1 \cup L_2})\mathrm{d}\theta\mathrm{d}\psi_y\mathrm{d}\psi_x\,,\end{aligned} \tag{22}$$

where the superscripts L_1 and L_2 indicate that we may be interested in predictions for another time period (here: "layout," L), L_2, than we have observations for, L_1. As in most studies on inference and uncertainty analysis, we still assume input data to be available also in L_2 and thus operate in "prediction" or "hindcasting" mode [*Renard et al*., 2011; *Del Giudice et al*., 2015b].

When evaluating the quality of the models, we want to compare observations of the system with predicted observations. This requires us to predict our knowledge of observations, $\mathbf{y}_o^{L_2}$, rather than our knowledge of the true output, $\mathbf{y}^{L_2}$:

$$f(\mathbf{y}_o^{L_2}|\mathbf{y}_o^{L_1},\mathbf{x}_o^{L_1\cup L_2}) = \int f(\mathbf{y}_o^{L_2}|\boldsymbol{\theta},\psi_y,\psi_x,\mathbf{y}_o^{L_1},\mathbf{x}_o^{L_1\cup L_2}) f(\boldsymbol{\theta},\psi_y,\psi_x|\mathbf{y}_o^{L_1},\mathbf{x}_o^{L_1\cup L_2})\mathrm{d}\boldsymbol{\theta}\mathrm{d}\psi_y\mathrm{d}\psi_x\,. \tag{23}$$

The essential difference between equations (22) and (23) is that the latter also considers output observation errors, usually in the form of iid Gaussian noise.

In the following subsections, we describe the specifics of predicting $\mathbf{y}^{L_2}$ and $\mathbf{y}_o^{L_2}$ with the four different methods described here. In addition, for the techniques RM and SIP, which also infer the rainfall input, we will discuss the formulation of our posterior knowledge of the rainfall.

2.3.1. Predictions With Alternative Methods

2.3.1.1. Predictions With LS

As the traditional least squares approach assumes that the uncertainty in the system output predictions only arises from incomplete knowledge about model parameters, its predictive distribution can thus be obtained by propagating the posterior of the model parameters:

$$\mathbf{Y}^{L_2} = \mathbf{y}_M^{L_2}(\boldsymbol{\Theta}_{\text{post}}^{L_1},\mathbf{x}). \tag{24}$$

For the prediction of observations, the observation error, $\mathbf{E}$, must be added in the g-transformed space:

$$\mathbf{Y}_o^{L_2} = g^{-1}(g(\mathbf{y}_M^{L_2}(\boldsymbol{\Theta}_{\text{post}}^{L_1},\mathbf{x})) + \mathbf{E}^{L_2}(\boldsymbol{\Psi}_{y,\text{post}}^{L_1})) \tag{25}$$

for time points in L_2 that are not identical to time points in L_1, where we know the observed output. Note that according to the model assumption (2), the observed input is used for $\mathbf{x}$ in these equations.

Numerically, a sample of $\mathbf{Y}^{L_2}$ is generated by propagating the parameter sample through the deterministic model, $\mathbf{y}_M^{L_2}$. To generate a sample of $\mathbf{Y}_o^{L_2}$, sample points of the normal distribution of $\mathbf{E}^{L_2}$ with the corresponding sample points of $\boldsymbol{\Psi}_{y,\text{post}}^{L_1}$ must be added on the transformed scale and the sum transformed back to the original scale as indicated in equation (25).

2.3.1.2. Predictions With BD

The bias description approach assumes that the uncertainty in the system output predictions arises from incomplete knowledge about model parameters and from input and structural errors. Thus, our best knowledge of the true system output requires consideration of the bias (and the transformation g):

$$\mathbf{Y}^{L_2} = g^{-1}(g(\mathbf{y}_M^{L_2}(\boldsymbol{\Theta}_{\text{post}}^{L_1},\mathbf{x})) + \mathbf{B}_{\text{post}}^{L_2}(\boldsymbol{\Psi}_{y,\text{post}}^{L_1})). \tag{26}$$

For the prediction of observations, the observation error, $\mathbf{E}$, must be added to the bias (on the transformed scale also):

$$\mathbf{Y}_o^{L_2} = g^{-1}(g(\mathbf{y}_M^{L_2}(\boldsymbol{\Theta}_{\text{post}}^{L_1},\mathbf{x})) + \mathbf{B}_{\text{post}}^{L_2}(\boldsymbol{\Psi}_{y,\text{post}}^{L_1}) + \mathbf{E}^{L_2}(\boldsymbol{\Psi}_{y,\text{post}}^{L_1})). \tag{27}$$

Again, the observed input (2) is used in these equations, as the effect of input errors to the output is corrected in the output by the additive term $\mathbf{B}$. As the distributions of $\mathbf{B}_{\text{post}}$ and $\mathbf{E}$ conditional on their parameters are normal (with expectation and variance given by equations (35)–(38) in *Reichert and Schuwirth* [2012]), we can again propagate the posterior sample of the model parameters through these equations and sample from the corresponding normal distributions to get a posterior sample of (26) and (27), respectively.

2.3.1.3. Predictions With RM

The rainfall multipliers approach assumes that the uncertainty in the system output predictions arises from incomplete knowledge about model parameters and from input imprecision. The output of the system is assumed to be equal to the model output forced with uncertain input $\mathbf{y}_M(\boldsymbol{\Theta},\mathbf{X})$, as for each storm event, j, the uncertain input is equal to the observed input times a factor β_j. Compared to the LS approach, this leads to the expansion of the parameter vector. The predictions are thus still given by the equations (24) and (25) with the exception that the use of the observed input (2) is replaced by the input given by equation (6). When predicting beyond calibration, our knowledge of the rainfall multiplier is described by a hierarchical model based on a conditional lognormal distribution with mean unity, the standard deviation of which is distributed according to the posterior of the parameter $\sigma_{\text{post}}^{\beta}$.

In addition to the posterior of the model output, the RM technique provides a posterior estimate of the rainfall given by (see equation (6))

$$X_i = \beta_{j(i)} x_{o,i}. \tag{28}$$

The numerical implementation is again similar to the LS approach. For storm events included in the calibration phase, the rainfall multiplier β_j is part of the parameter sample. For other events, β_j is drawn from a log-normal distribution with mean unity and standard deviation $\sigma^{\beta}_{\text{post}}$.

2.3.2. Predictions With SIP

Our approach of using a stochastic input process assumes that the uncertainty in the system output predictions arises from incomplete knowledge about model parameters and from input imprecision. The distributions representing our knowledge of true and observed output are given by considering an additional integration over the rainfall potentials, ξ and ξ_o, in the equations (22) and (23), and eliminating arguments that are not relevant. This leads to

$$\begin{aligned} f(\mathbf{y}^{L_2}|\mathbf{y}_o^{L_1}, \mathbf{x}_o^{L_1\cup L_2}) = \int f(\mathbf{y}^{L_2}|\boldsymbol{\theta}, \psi_y, h(\xi^{L_1\cup L_2})) \\ \cdot f(\boldsymbol{\theta}, \psi_y, \psi_x, \xi^{L_1\cup L_2}, \xi_o^{L_1\cup L_2}|\mathbf{y}_o^{L_1}, \mathbf{x}_o^{L_1\cup L_2}) \mathrm{d}\boldsymbol{\theta}\mathrm{d}\psi_y\mathrm{d}\psi_x\mathrm{d}\xi^{L_1\cup L_2}\mathrm{d}\xi_o^{L_1\cup L_2} \end{aligned} \tag{29}$$

and

$$\begin{aligned} f(\mathbf{y}_o^{L_2}|\mathbf{y}_o^{L_1}, \mathbf{x}_o^{L_1\cup L_2}) = \int f(\mathbf{y}_o^{L_2}|\boldsymbol{\theta}, \psi_y, h(\xi^{L_1\cup L_2})) \\ \cdot f(\boldsymbol{\theta}, \psi_y, \psi_x, \xi^{L_1\cup L_2}, \xi_o^{L_1\cup L_2}|\mathbf{y}_o^{L_1}, \mathbf{x}_o^{L_1\cup L_2}) \mathrm{d}\boldsymbol{\theta}\mathrm{d}\psi_y\mathrm{d}\psi_x\mathrm{d}\xi^{L_1\cup L_2}\mathrm{d}\xi_o^{L_1\cup L_2}\,. \end{aligned} \tag{30}$$

The posterior of our knowledge of the true catchment-mean precipitation is given by

$$\begin{aligned} f(\mathbf{x}^{L_1\cup L_2}|\mathbf{y}_o^{L_1}, \mathbf{x}_o^{L_1\cup L_2}) = \int f(\mathbf{x}^{L_1\cup L_2}|\xi^{L_1\cup L_2}) \\ \cdot f(\boldsymbol{\theta}, \psi_y, \psi_x, \xi^{L_1\cup L_2}, \xi_o^{L_1\cup L_2}|\mathbf{y}_o^{L_1}, \mathbf{x}_o^{L_1\cup L_2}) \mathrm{d}\boldsymbol{\theta}\mathrm{d}\psi_y\mathrm{d}\psi_x\mathrm{d}\xi^{L_1\cup L_2}\mathrm{d}\xi_o^{L_1\cup L_2}\,. \end{aligned} \tag{31}$$

Numerically, this involves assimilating input data for both L_1 and L_2 and, additionally, using output data from L_1 to estimate the input. For time points which are several correlation lengths τ_ξ ahead of L_1, the discharge calibration data $\mathbf{y}_o^{L_1}$ will have a negligible (direct) influence on the estimation of ξ^{L_2}. Indirectly, however, $\mathbf{y}_o^{L_1}$ influences the estimation of ξ^{L_2} by affecting the posterior distribution of σ^2_ξ. A sample for the true output (29) is then obtained by propagating the posterior sample components corresponding to $\boldsymbol{\theta}$, ψ_y, and ξ through the model $\mathbf{y}_M(\boldsymbol{\theta}, \psi_y, h(\xi))$. For the sample for (30), we have to add sample points of the normal distribution of observation errors with the corresponding parameters ψ_x as well as consider output transformation, as in equation (25). Finally, to draw realizations of rainfall intensities, e.g., to numerically quantify input uncertainty, we simply need to generate random paths of ξ and transform those into time series of $\mathbf{x}$ via h. This Monte Carlo method is similar to drawing samples from a desired distribution by drawing from a standard uniform distribution and then transforming those samples via an inverse distribution function [*Platen and Bruti-Liberati*, 2010; *Kroese et al*., 2011]. Our method, however, additionally aims to capture the autocorrelation structure of the rainfall.

3. Materials

To demonstrate the relevance of our method in hydrology, we tested it in an urban catchment with real observations.

3.1. Rainfall Scenarios

To test the performance of the SIP method compared to the other three error descriptions, we considered two typical scenarios of rainfall data availability. Scenario Sc1 uses as input the rainfall recorded by two of our own pluviometers located in the direct vicinity of the catchment (section 3.4). Averaging the data from two pluviometers close to the catchment centroid has appropriately characterized the input of systems with areas ranging from dozens [*Del Giudice et al*., 2015a] to more than 1000 ha [*Del Giudice et al*., 2015b]. For larger catchments (>10000 ha), more complex methods, such as angular distance-weighting [*Li et al*., 2012], conditional simulation [*Renard et al*., 2011], or external drift kriging [*Bardossy and Das*, 2008], could be more appropriate for spatial interpolation of rainfall data. Using the highly-representative (located only

400 m away from the catchment centroid) high-resolution (recorded every minute) data from these gauges is an illustrative example of the best case scenario of input data availability. We thus define Sc1 as using "accurate input." Scenario Sc2 uses as input the rainfall recorded by a pluviometer managed by the Swiss meteorological office (section 3.4). Using data from this less-representative gauge is a typical example of suboptimal input data availability. We define Sc2 as using "inaccurate input." In this study, we focus on point-scale pluviometers, since they still are the most common source of rainfall measurements [*McMillan et al.*, 2011].

3.2. System

The test case hydrosystem is a small partially combined sewer network located in Adliswil in the proximity of Zurich, Switzerland (Figure 3). The watershed surface is about 28.6 ha, only a fraction of which contributes to the stormwater outflow. The area is characterized by medium density of housing and a slope of about 8.7%.

3.3. Hydrological Model

The hydrological model of the hydrosystem consists of two components, one for the stormwater runoff and the other for the wastewater produced. This concept is akin to the one adopted by *Del Giudice et al.* [2015b]. Parsimonious linear models, using as input spatially aggregate rainfall, are effective tools to reproduce the discharge dynamics at the catchment outlet during storm events [*Coutu et al.*, 2012; *Sun and Bertrand-Krajewski*, 2013].

The stormwater runoff is modeled by a linear reservoir which is alimented by a time-varying precipitation input, $\mathbf{x}$, and a constant base flow partly coming from groundwater, x_{gw}. The dynamics of this compartment is described by the following ODE which can be solved analytically:

$$\frac{ds(t)}{dt} = A \cdot x(t) + x_{gw} - \frac{s(t)}{k}, \tag{32}$$

where s is the water volume within the reservoir, A is the area contributing to the rainfall-runoff, and k is the mean residence time in a virtual reservoir.

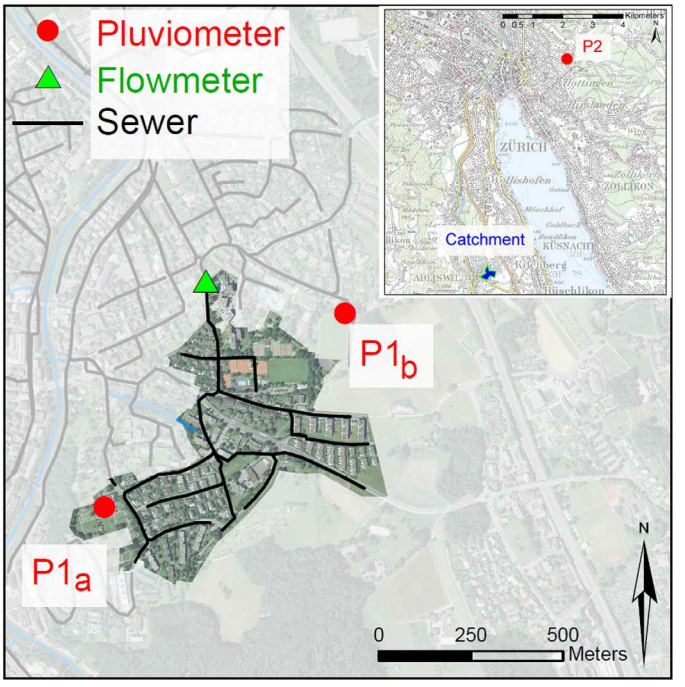


Figure 3. Map of the study area. The urban catchment and monitoring sites for input and output measurements are represented. Data from $P1_a$ and $P1_b$ are combined to provide the input of Sc1, while data from P2 are used as input in Sc2.

The daily variations in discharge due to wastewater input are described by the harmonic function:

$$w(t) = \sum_{i=1}^{2} \left(\varsigma_i \sin \frac{2\pi i t}{24} + \chi_i \cos \frac{2\pi i t}{24} \right), \tag{33}$$

with σ_1, σ_2, χ_1, and χ_2 representing the coefficients of the trigonometrical series. The base flow parameter x_{gw} makes sure that the outlet discharge be nonnegative. For the system studied, the stormwater component during high precipitation can be more than 20 times larger than the average wastewater flow. The combined discharge at the outlet of the system is modeled by the superposition of the storm and wastewater:

$$y_M(t) = \frac{s(t)}{k} + w(t). \tag{34}$$

The model, as well as the analyses, has been coded in the statistical programming language R [*R Core Team*, 2014].

3.4. Data Set

The measurements of precipitation and discharge were performed from June to November 2013. We recorded the rainfall data with two weighing gauges (Figure 3, $P1_a$ and $P1_b$). Averaging the recordings ($\Delta t = 1$ min) from those gauges (OTT Pluvio^2), we derived the input to the system in scenario Sc1 (section 3.1). As input in scenario Sc2 we instead used rainfall observed ($\Delta t = 10$ min) at the pluviometric station of Zurich-Fluntern (Figure 3, P2), which belongs to the network of the Swiss meteorological office (www.hw.zh.ch/hochwasser/foto/DB\%20SMA.pdf). Precipitation data from four stations located around the catchment (not shown) were analyzed to parameterize the function transforming the standard OU-process into precipitation (see Figure 2) and the prior of the "rainfall potential" (supporting information Figures S1 and S2). These pluviometers located outside the catchment provided valuable information about the dynamics of precipitation in the region, but were not selected for hydrological modeling due to their distance from the catchment.

Wastewater flow was measured ($\Delta t = 4$ min) at the outlet of the catchment by a radar-based contact-free sensor (Flo-Dar 4000 SR). From the recorded data, we selected three events for calibration and two for validation (see section 4). These storms were characterized by a significant response of the system (i.e., maximal flowrate one order of magnitude larger than during dry-weather) and negligible infiltration from groundwater. Being separated by several days in between, the individual precipitation events can be regarded as independent. During these storms, the volume of precipitation ranged from 5 to 50 mm, the duration from 1 to 15 h, and the maximal intensity from 0.3 to 0.8 mm measured in 1 min. The average precipitation intensity observed both during calibration and validation was 2 mm/h. Although not extremes, those storms where among the largest for the recorded period and location. The use of typical storm events is generally regarded as appropriate in hydrological studies on rainfall uncertainty (section 1).

3.5. Prior Distributions

The marginal prior distributions of the hydrological model and error model parameters are given in Table 1. We estimated the prior marginal distributions of A, k, x_{gw}, σ_1, σ_2, χ_1, χ_2, and σ_E based on a least squares calibration employing measurements not used in the final analysis (data not shown). Having an interpretation related to the hydrological system (A is connected to the volume of effective precipitation, k to the rapidity of the system response, x_{gw} to the base flow, and σ_1, σ_2, χ_1, χ_2 to the harmonic dynamics of wastewater) and measurement device (σ_E represents the imprecision of runoff observations), we refer to these constants as "physical parameters." Regarding the priors of the bias error model (BD), we followed the guidelines provided in previous works [*Reichert and Schuwirth*, 2012; *Brynjarsdóttir and O'Hagan*, 2014; *Del Giudice et al*., 2015b]. For σ_B, the magnitude of the bias, we determined its prior standard deviation by analyzing the model discrepancy between the model forced with accurate rainfall and the measured discharge. The prior expected value of τ, the bias autocorrelation time scale, was set approximately equal to 1/3 of the duration of the falling limb of a storm hydrograph. As for the other parameters, the priors of the bias were based on analyses of events independent from those used for calibration and validation. In the multiplicative error

Table 1. Hydrological Model and Error Model Calibration Parameters ($\boldsymbol{\theta}, \psi_y, \psi_x$)[a]

Symbol	Description	Units	Prior
A	Area contributing to outflow	m^2	LN(11815.8, 1181.6)
k	Water residence time	hr	LN(0.079, 0.016)
x_{gw}	Groundwater infiltration and sewage base flow	L/s	LN(2.05, 0.013)
$-\sigma_1$	Trigonometric coefficient of the sewage flow	L/s	LN(0.25, 0.094)
$-\sigma_2$	Trigonometric coefficient of the sewage flow	L/s	LN(0.84, 0.019)
$-\chi_1$	Trigonometric coefficient of the sewage flow	L/s	LN(0.68, 0.019)
χ_2	Trigonometric coefficient of the sewage flow	L/s	LN(0.077, 0.01)
σ_E	Standard deviation of **E**	g(L/s)	$\mathrm{LN}(4.1\,\frac{dg}{dy}\vert_{50},\ 0.41\,\frac{dg}{dy}\vert_{50})$
σ_B	Standard deviation of **B** (BD method)	g(L/s)	$\mathrm{TN}(0,\ 3.77\frac{dg}{dy}\vert_{50},\ 0,\ \infty)$
τ	Correlation length of **B** (BD method)	hr	LN(0.47, 0.047)
β_j	Rainfall multiplier for the event j (RM method)		$\mathrm{LN}(1, \sigma^{\beta})$
σ^{β}	Standard deviation of the multipliers (RM method)		LN(0.1, 0.02)
σ^2_{ξ}	Variance between rainfall potentials (SIP method)		LN(0.4, 0.2)

[a]The notation for prior distributions is: LN(μ, σ): lognormal, TN(μ, σ, a_1, a_2): truncated normal. The symbol meaning is: μ: expected value, σ: standard deviation, a_1: lower limit, and a_2: upper limit.

model (RM), similar to *Sikorska et al*. [2012], we assumed a priori no bias in the rainfall measurements and estimated the mean standard deviation of the input uncertainty, σ^{β}, to be 10%.

3.6. Performance Assessment

An optimal error description should produce posterior model parameters which are highly representative of the average conditions of the physical system [*Vrugt et al*., 2008; *Brynjarsdóttir and O'Hagan*, 2014]. Furthermore, it should ensure reliable (i.e., with high coverage of data), accurate (i.e., on average close to the data or unbiased), and precise (i.e., sharp or with low dispersion) predictions, especially in the extrapolation domain. For this reason, we inspected the following factors:

1. **Consistency of the updated parameters** $\boldsymbol{\theta}, \psi_{\boldsymbol{y}}, \psi_{\boldsymbol{x}}$. To assess the corruption of the estimated parameters due to input errors, we compare the posterior marginals obtained with accurate and inaccurate rainfall data and with the different likelihood functions.
2. **Prediction accuracy**. As a measure of model adequacy, we calculate the Nash-Sutcliffe efficiency at the maximum of the posterior, *NS*. The closer this coefficient is to 1, the better the model fits the data, especially during high-flow periods [*Reichert and Mieleitner*, 2009; *Coutu et al*., 2012].
3. **Prediction reliability**. We analyze the data coverage of the 95% interquantile intervals. If the percentage of data points falling into these total uncertainty bands is larger than or equal to 95, we consider the predictions to be reliable [*Del Giudice et al*., 2013; *Li et al*., 2012].
4. **Prediction precision**. We compute the average band width (ABW) of the 95% interquantile intervals. The lower this value, the lower prediction uncertainty is.
5. **Integrated predictive performance**. We adopt two metrics to quantify how reliable, accurate, and precise predictions are. The first is the interval (skill) score [*Gneiting and Raftery*, 2007]. The better the quality of the predictions, the closer to 0 this statistic is. Second, we also create predictive quantile-quantile plots, which analyze the probability of the observations being distributed as the model output (including all uncertainties). The more reliable and precise the predictive distribution is, the closer to the identity line the observed *p* values are [*Renard et al*., 2011].

4. Results

In the following, we present the outcomes of the case study application in terms of "calibration distributions," representing the posterior parameters, and "smoothing distributions," representing the posterior input and output in the calibration phase (where both rainfall and runoff data are assimilated) and in the extrapolation phase (where only rainfall data are assimilated).

4.1. Estimated Parameters During Calibration

Violin plots of the parameters are illustrated in Figure 4 and are based on the MCMC samples presented in supporting information. The convergence to the target posterior distributions was typically reached within 10^4-10^5 iterations and was corroborated by running multiple chains with different initial conditions and by observing their consistency, time-independence, and mixing. Compared to the other inference schemes, calibration with SIP involved a computational expense 10–100 times higher (supporting information Figure S5). In the scenario with accurate rainfall data (Sc1), inferred hydrological model parameters had a similar distribution for all error representations (Figure 4, top row). *A* and σ_E, however, showed some dependencies on the error model. The first parameter, connected to the fraction of precipitation converted into discharge, decreased slightly more during inference with LS and BD, both of which do not describe input uncertainty explicitly. The second parameter, connected to output measurement uncertainty, reached the lowest value with BD (50% less than the average of the other error models), which partitions output uncertainty into two terms. Very little model bias was identified (posterior σ_B similar to posterior σ_E and ~70% lower than for Sc2), a condition confirmed by the RM error model which did not display an increase in σ^{β}.

Scenario Sc2, with lower input data quality, induced different performances of the error models (Figure 4, second row). With BD and SIP, the posterior median of most physical parameters was very similar and also differed minimally from Sc1 (apart from a 75% increase in *k* with BD and a 30% decrease in σ_E with SIP). Interestingly, in Sc2 the spread of the distributions of the parameters directly related to rainfall, *A* and *k*, increased for the BD (standard deviation on average 260% higher) while remaining almost the same for SIP.

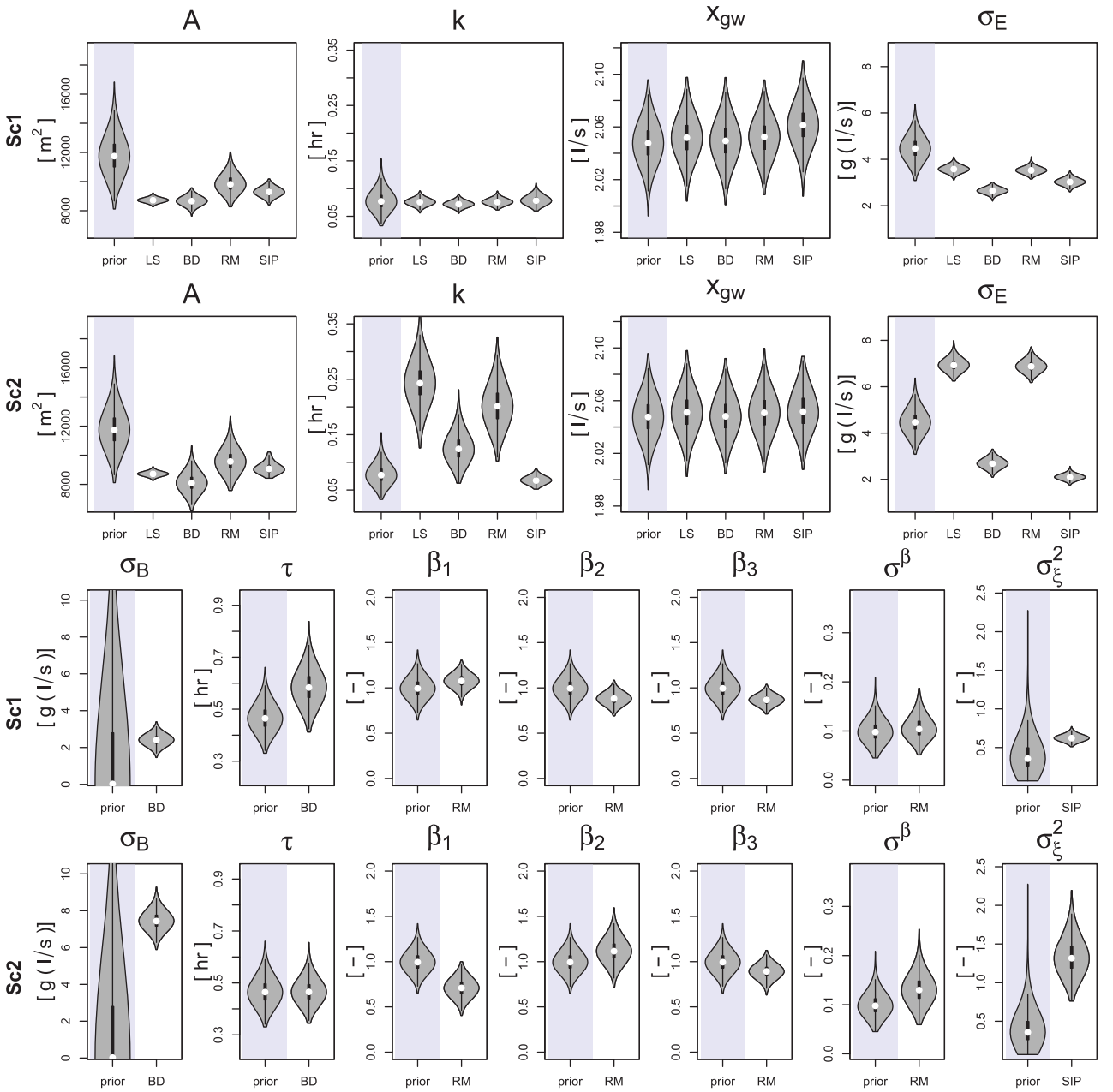


Figure 4. Marginal prior and posterior distributions of physical parameters common to all error models (first two rows) and of parameters typical of each error model (bottom two rows). As in a Box plot, the median (white dot), 50% interquantile range (thick line), and the range of typical values (thin line) are displayed; additionally, the Kernel density estimates are also shown. Comparing results between Sc1 (accurate rain) and Sc2 (faulty rain) shows the corrupting effects of input uncertainties on the physical parameters and should be detected by the error model parameters. See Table 1 for an explanation of the symbols.

With LS and RM, some posterior parameters were substantially affected by the increased input errors of Sc2. In particular, for the hydrologic response time of the catchment (k) the median increased 220% with LS and 170% with RM, while the standard deviation increased 505% with LS and 570% with RM. For the output measurement errors (σ_E), both the median and standard deviation increased ~95% with both error models. The other two parameters common to all error models were less affected by the inaccurate rain data. In particular, x_{gw}, representing low-water flow, was not affected by inference scheme or rainfall data quality. The effective impervious area, A, was also only mildly altered by the inaccurate input data. This appears to be connected to the rainfall characteristics of both scenarios (Figure 5) which, despite showing different

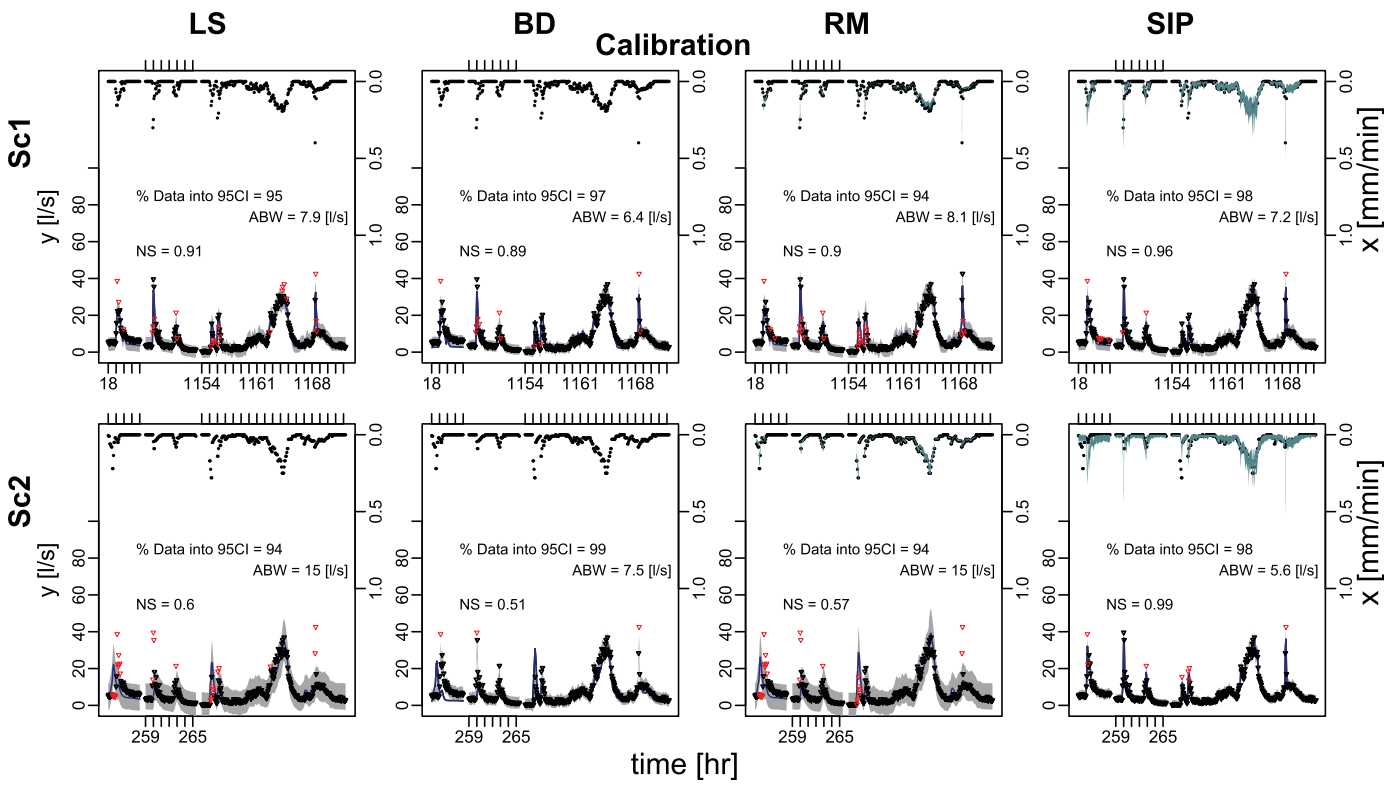


Figure 5. Total output (bottom of each frame) and input uncertainty (top of each frame) in the calibration phase with all error models and rainfall scenarios. Each tick indicates 1 h and, for each event, marks are either plotted on the top or on the bottom of the figure. The points represent measured data. Red triangular dots indicate runoff data not included within the 95% output credible intervals. Predictions are considered reliable if red dots are $\leq$ 5% of the total. ABW is the average band width. NS expresses the accuracy of the model median (blue line). While LS and BD do not assess input uncertainty, RM and SIP do.

temporal behavior, have similar volumes. This is also in agreement with the only slight deviations of the multipliers β_1, β_2, β_3 from unity, even with the worst rain (Sc2).

An analysis of the error-model-specific parameters (Figure 4, bottom two rows) shows that, as expected, all parameters related to the amount of input uncertainty increase when comparing Sc1 with Sc2. In particular, with less accurate rain, more output bias is detected (210% higher σ_B) and more uncertainty is identified with SIP (110% higher σ^2_ξ). With RM, instead, the increased input uncertainty is barely recognized (σ^β hardly increases).

4.2. Estimated Input and Output During Calibration

Forced with accurate rainfall data (Sc1), the chosen hydrological model fitted the calibration data accurately (Nash-Sutcliffe efficiency around 0.9 with all error models, Figure 5, first row). Predictions were also reliable for all error descriptions (data coverage around 95%) and sharp (ABW $\ll$ maximum discharge). Estimated input uncertainty with RM and SIP was also very low. In the more realistic scenario Sc2, however, a substantial distinction among error models becomes evident (Figure 5, second row, and Figure 6). Both LS and RM significantly increase output uncertainty, producing unrealistically wide (i.e., imprecise) prediction intervals while still missing the misrecorded rainfall peaks. Instead, BD was able to effectively assimilate the deviating flow data and in this way correct model output in a reliable and precise way. The most interesting result, however, was produced by SIP. Not only were the output predictions the most accurate and precise among the four cases but, compared to the other methods, total uncertainty intervals were mostly above zero. This desirable feature can be difficult to achieve in ephemeral catchments when modeling uncertainty at the output level, even with heteroschedastic error models [*Evin et al.*, 2013]. Instead, propagating uncertainty through the model, as SIP does, can ensure nonnegative discharge predictions in a way conceptually more satisfying than when constructing heteroschedastic output error models [*Honti et al.*, 2013; *Del Giudice et al.*, 2015b]. Additionally, SIP also generated the most realistic input estimates. As shown in Figure 6 and

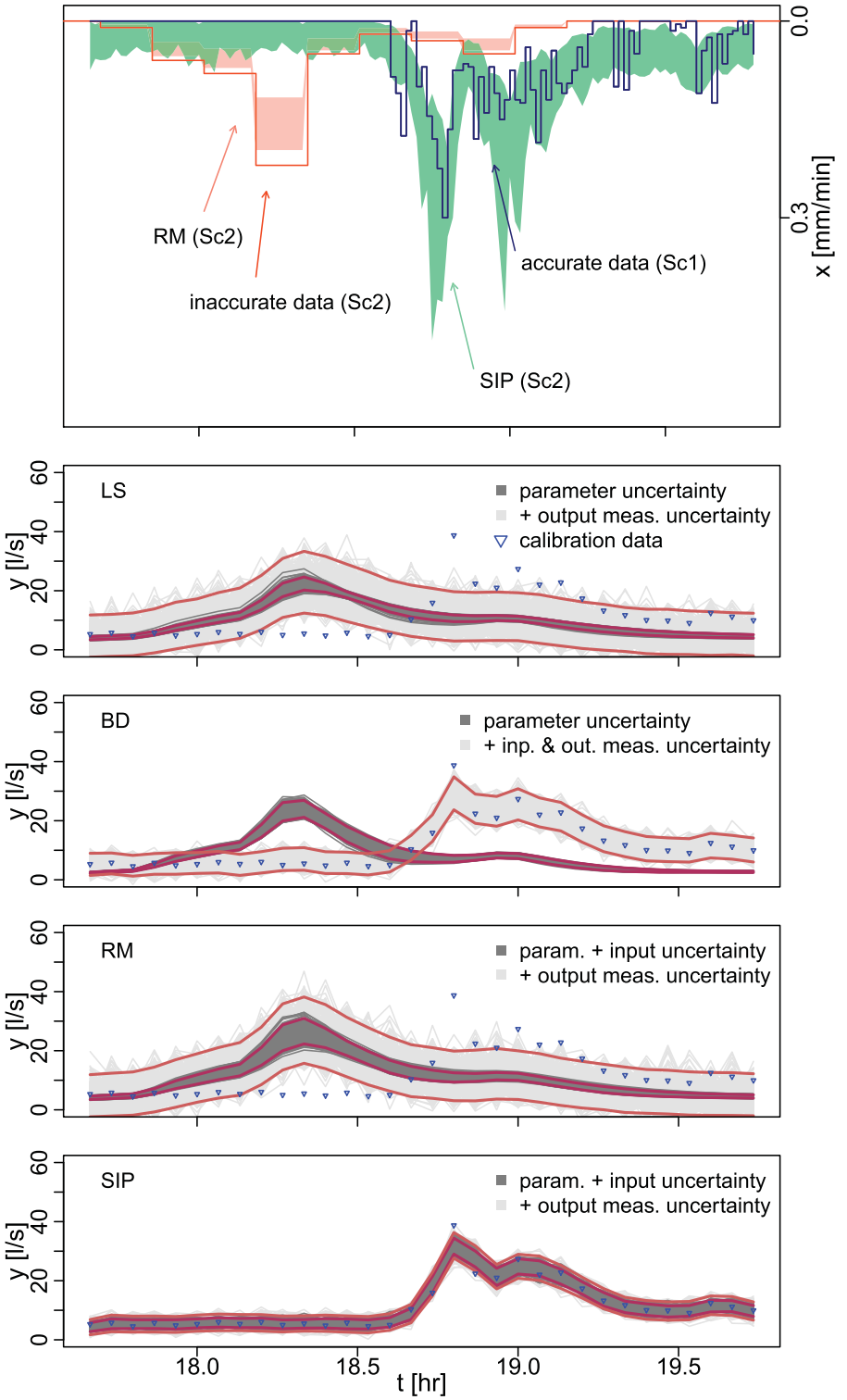


Figure 6. Zoom of the estimated input uncertainty with RM and SIP using inaccurate rainfall data for a storm event in the calibration period (Figure 5, second row, fourth column). The accurate data from Sc1 are only used for a posteriori validation. Whole-catchment precipitation inferred with SIP is substantially more realistic than the one with RM. SIP, as a continuous-time stochastic process, is better able to learn from the flow data.

supporting information Figure S6, even when using highly biased input data (Sc2), SIP produced estimates very close to the optimal data (plotted for comparison). The rainfall multipliers, instead, were unable to deal with these dynamic input biases.

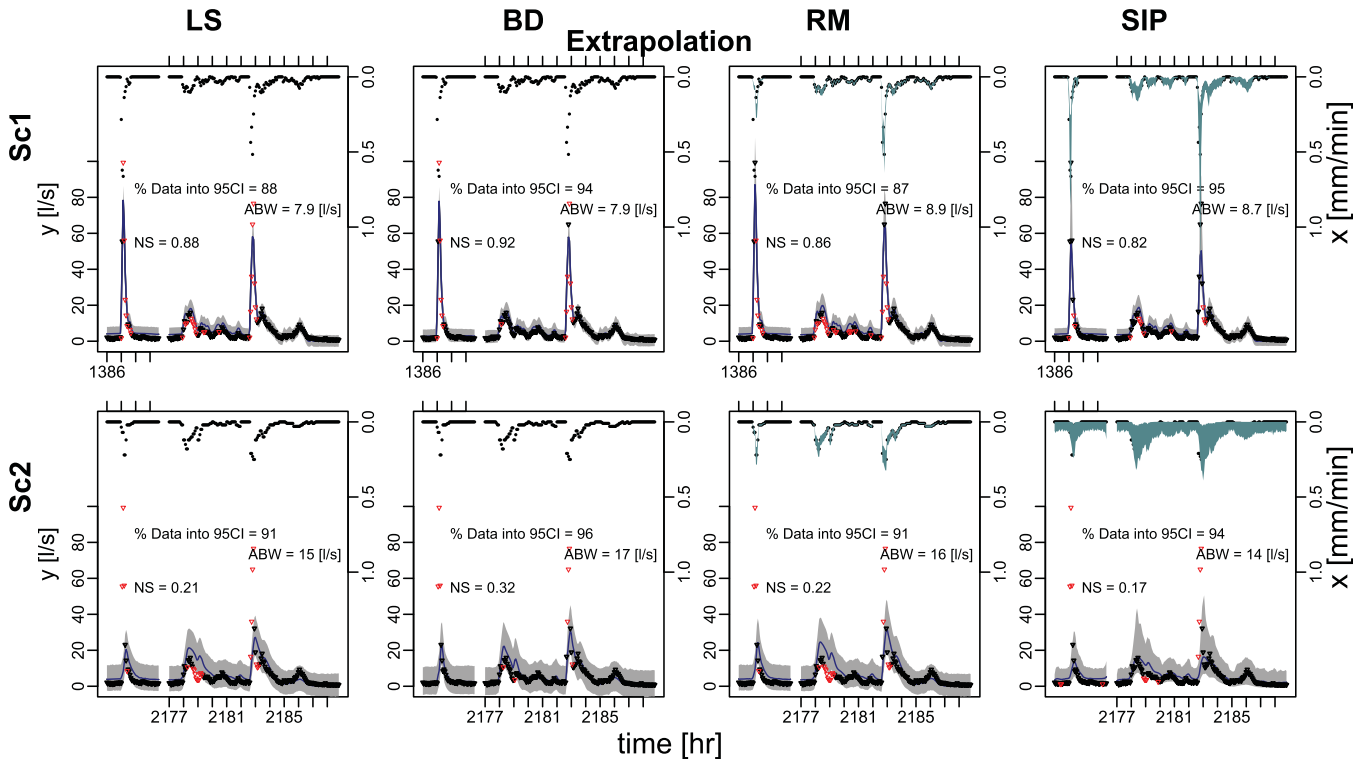


Figure 7. Total output (bottom of each frame) and input uncertainty (top of each frame) in the extrapolation phase with all error models and rainfall scenarios. Each tick indicates 1 h and, for each event, marks are either plotted on the top or on the bottom of the figure. The points represent measured data. Red triangular dots indicate runoff data not included within the 95% output credible intervals. Predictions are considered reliable if red dots are $\leq$ 5% of the total. ABW is the average band width. NS expresses the accuracy of the model median (blue line). While LS and BD do not assess input uncertainty, RM and SIP do. Runoff data here are only used for a posteriori validation.

4.3. Estimated Input and Output During Extrapolation

In the validation period, when using only input data but not output data, BD and SIP produced the most reliable flow predictions (coverage close to 95%), regardless of data quality (Figure 7, bottom two rows, and supporting information Figures S17–S20). When using the SIP technique, however, the hydrological model produced ~13% less accurate results than with the other methods. Contrary to the calibration phase, differences among the error models are visible in both rainfall scenarios. With the most accurate and precise rainfall (Sc1), the bias description performed best, especially in terms of accuracy and precision during low flows. Regarding rainfall, SIP allowed for slightly more uncertainty than RM, especially during the moments of maximum intensity. Using rainfall data from the less representative pluviometer (Sc2) helped to further differentiate the input estimates of the two error models. While neither was able to account for missing peaks without considering output information, SIP was substantially less overconfident than RM (supporting information Figure S6). Concerning the flow predictions with inaccurate rainfall, SIP appears to have generated realistic uncertainty bands during high flows. It was also more precise than the other methods during low flows. All-in-all, the BD method dominated in accuracy and reliability, although it slightly overestimated predictive uncertainty (supporting information Figure S7).

5. Discussion

5.1. Interpreting Posterior Parameters, Input, and Output

As expected theoretically, and as confirmed by the results of the case study, describing input uncertainty in a more realistic way helps to protect model parameters from shifting to biased values, provided that the hydrological model is sufficiently accurate [*Kavetski et al.*, 2006; *Vrugt et al.*, 2008; *Li et al.*, 2012]. Parameters estimated with SIP using erroneous input data were very close to the value obtained using the best rain data (Figure 4). This means that SIP avoided the compensation of input errors by shifts in model parameters. Instead, with simpler error models like LS and RM, some parameters were forced to less meaningful posteriors to help the model fit flow data, notwithstanding the erroneous forcing. This occurs frequently in hydrological modeling where

input data errors can corrupt model parameters away from their original representation of average catchment or measurement characteristics [*Renard et al*., 2011; *Bardossy and Das*, 2008].

Interestingly, the bias description had a similar parameter preserving effect as SIP. This robustness of BD, only speculated about in previous studies [*Bayarri et al*., 2007; *Del Giudice et al*., 2015b], was confirmed by our study thanks to the comparison of the "unbiased model" (Sc1) with a "biased model" (Sc2). In other words, both SIP and BD helped alleviate the (distorting or overtuning) impact of errors in the regressor, i.e., the areal precipitation. The ability of BD to infer parameters close to their physical value even in the presence of input errors is very promising. This is probably due to a combination of two reasons. First, the a priori parameterization of the bias process and its parameters was plausible since it was based on a combination of system understanding and analyses of independent data sets (see section 3.5 and *Brynjarsdóttir and O'Hagan* [2014]). Second, our runoff model appeared to represent the hydrological processes sufficiently well (see Figure 5). SIP and BD also provided similarly reliable and precise predictions in the calibration phase for the output, even in Sc2. The mechanisms behind the two methods, however, are different. SIP avoided parameter overtuning by flexibly adapting the input process and therefore adjusting the biased model at its source. BD, in contrast, avoided parameter compensation by not forcing the model to fit the data but rather allowing the autocorrelated discrepancy term to bridge the gap between model output and data.

In contrast to our expectations, considering storm-dependent rainfall multipliers did not improve the calibration of physical parameters compared to the simplest least squares approach. This is probably linked to the fact that we are analyzing a rapidly reacting catchment with more detailed data than what is normally available for natural systems [*Ochoa-Rodriguez et al*., 2015]. This can explain why our results with RM differ from those of studies conducted at coarser time resolutions [*Vrugt et al*., 2008; *Li et al*., 2012] or with small and nonsystematic input errors [*Sun and Bertrand-Krajewski*, 2013].

As observed in Figures 5 and 6, model input is estimated much more realistically by the SIP method than by RM. Both methods learn from the output about the input dynamics. Discharge data integrate rainfall-runoff processes over the entire catchment [*Frey et al*., 2011] and, by using the hydrological model "backward," they can be used to reliably learn about precipitation [*Kirchner*, 2009]. Estimating the rainfall not only from pluviometric data but also from runoff data is one of the advantages of describing input uncertainty in hydrological inference with SIP or RM. This allows us not only to quantify but also to reduce input uncertainty, unlike the approaches that estimate true rainfall from pluviometric data alone [e.g., *Rodriguez-Iturbe et al*., 1987; *Ailliot et al*., 2015]. Contrary to RM, however, SIP, however, is not limited to following the temporal dynamics of the measured rain and can, therefore, more effectively learn from runoff dynamics. This flexibility is even larger than what was obtained by *Reichert and Mieleitner* [2009], who let the multiplicative factors vary within the event. Here, indeed, we can additionally handle time periods where rainfall was completely unobserved.

In the case of contradicting input and output measurements, SIP makes a compromise between matching the input to the rainfall data and the output to the discharge data. As expected, the direction of this compromise is mainly dictated by our a priori assumptions on the errors of the input measurements (σ_{ξ}) and output measurements (σ_E) and by the relative number of input and output data (defined by data resolution). Here, in the best case scenario (Sc1), we had 4 times more input than output data, whereas in the worst case scenario (Sc2) the size of the output data set was 2.5 times the size of the input data set. This last point probably explains why SIP considered the information content of the output time series relatively strongly (Figure 6). While the amount of rainfall data influences the precision with which we can estimate ξ_o, SIP is still able to distinguish which input time series is more informative. When comparing σ_{ξ}^2 estimated with the two scenarios (Figure 4), we see that SIP can correctly assess how good the rainfall data are, independently of the data size. This ability to "solve" the inconsistencies between input and output data by recognizing disinformative time series can be very valuable in hydrological inference [*Beven and Westerberg*, 2011].

Interestingly, although SIP was mainly developed as a method to improve hydrological inference, it also appears to reliably estimate input uncertainty during extrapolation. In this phase, where only rainfall data are assimilated, obviously no method can correct an erroneous input. Compared to RM, however, SIP more realistically reflected our increased lack of knowledge by estimating wider input uncertainty bands (supporting information Figure S6). Such a comparison of the inferred rainfall against

independent pluviometric data is a strong test of the robustness of the input estimates [*Kirchner*, 2009; *Vrugt et al*., 2008].

Runoff predictions during extrapolation had, in all cases, a similarly reasonable coverage, even in the worst scenario and with the simplest error model (Figure 5, last row, first column). This is probably because rainfall biases, although important, only last a few time steps. These quickly vanishing errors thus produced a considerable mismatch in the output but barely influenced the coverage (or reliability), which is a measure integrating all time points. Among all error models, however, BD still provided slightly more accurate and reliable uncertainty intervals than the other methods. This confirms the advantages of the bias description for reliable flow prediction, as discussed in previous studies [*Honti et al*., 2013; *Del Giudice et al*., 2015b]. Compared to the other methods, SIP produced extrapolative predictions which were similarly reliable, yet slightly less accurate. Rather than having to do with the estimated parameters, this is probably connected to the estimated input. Indeed, the methodology's focus is to better quantify input uncertainty and only to reduce it during the calibration phase. Therefore, as in this case study, it is possible to obtain a predictive median which fits validation data slightly less well than when using input data directly.

5.2. Advantages and Limitations of SIP

Based on theoretical reflections and experiences from this case study, this novel formulation of input uncertainty as a stochastic process has the following advantages over previous methods:

1. Compared to LS and BD, SIP provides a more accurate assessment of model input during the calibration phase and a realistic characterization of input uncertainty, also when extrapolating to the validation period. When input errors are the main contributor to predictive uncertainty, SIP helps to infer more realistic physical parameters than those obtained with LS. Furthermore, by stochastically describing and propagating the input, SIP can support the decomposition of output uncertainty into its sources, a highly desirable feature [*Vrugt et al*., 2008; *Renard et al*., 2011]. This is not possible with LS, which erroneously partitions the identified calibration errors into (only) parameter and output measurement uncertainty. Characterization of total output uncertainty with SIP is similar to that of BD (Figure 6 and supporting information Figure S18 versus S20). However, by describing the uncertainties where they arise instead of "at the end of the pipe," uncertainty separation with SIP is conceptually sounder than with BD.
2. Compared to RM, SIP represents a more appropriate model for forcing errors arising from rainfall measurements of suboptimal quality, e.g., collected by a low-resolution rain gauge or one which is located away from the catchment. SIP provides a better rainfall description, especially in two cases. First, when the temporal pattern of whole-catchment precipitation during the storm event is different from the observed dynamics. In fact, contrary to RM, SIP does not assume the storm to have a certain "shape" dictated by the input measurements. Therefore, SIP can compensate for time-varying input errors and generate a reasonable rainfall dynamics. Describing the input as a continuous stochastic process allows the rain rate fluctuations to be estimated at very fine scales [*Sigrist et al*., 2012]. This can be very useful in hydrology [*Paschalis et al*., 2013], for instance by helping to downscale coarse rainfall measurements to a temporal resolution appropriate for urban hydrology [*Ochoa-Rodriguez et al*., 2015]. Second, SIP can even handle storms not recorded at all, which cannot be tackled with RM. Because of this superior input characterization, especially when the number of events is not very high, SIP can estimate more meaningful physical parameters than RS.

Notwithstanding its several improvements with respect to existing methods, our approach still has some limitations:

1. The main disadvantage of the suggested technique is its computational requirements. Inferring input requires the propagation of a large number of suggested inputs through the model, which makes inference computationally one or two orders of magnitude more demanding. Describing the input stochastic process as an Ornstein-Uhlenbeck process and then sampling from it with an advanced MCMC strategy made the inference tractable. While this is more practical than estimating one multiplier per time point, which is virtually unfeasible, it is still computationally much more expensive than any of the other methods tested here. A harder and slower inference is a well-recognized problem when attempting to describe the sources of errors in environmental modeling [*Yang et al*., 2008; *Renard et al*., 2011; *Rougier*, 2013].

2. Inference of the dynamics of rainfall at an arbitrary temporal resolution from output is obviously limited by the retention time of the hydrological system. While this was possible for our urban catchment, it may be more limited in natural catchments. Rather than "doing hydrology backward," i.e., reconstructing the precipitation only from discharge data [*Kirchner*, 2009], we thus suggest adding a "backward component" ($f(\mathbf{y}_o|\theta,\psi_y,h(\xi))$ as used in equation (21)) to increase the information available to estimate precipitation.
3. A reasonable parameterization of the prior input process requires rainfall measurements additional to those used during calibration and validation. The collection of a relatively large rainfall data set necessary for SIP might be expensive. This, however, can also be seen as an advantage over the other methods, which cannot use this prior information as effectively. Furthermore, the need for additional data is usually not problematic as routine rainfall measurements, e.g., provided by meteorological offices, could be used to parameterize the prior input process.
4. Finally, the current implementation of SIP assumes that the main reason for model bias is input errors and therefore uses a LS likelihood (equations (1) and (3)) as an output error model. As recognized for multipliers [*Li et al*., 2012; *Sun and Bertrand-Krajewski*, 2013], this has the potential of producing rainfall estimates which unrealistically compensate for structural inadequacies. While this effect might be useful in some situations, e.g., to detect unexpected or difficult-to-measure inputs such as groundwater infiltration, we generally prefer having input estimates as independent as possible from the hydrological model [*Kirchner*, 2009]. This can be accomplished by using a model with minimal structural errors, as done here. However, for more general situations, in section 5.4 we discuss possible strategies to cope with model structural deficits.

5.3. Recommendations

Depending on the available resources and the specific objectives of hydrological modeling in a given study, we provide our perspective on which of the four techniques discussed in this paper to preferably apply (but see also section 5.4 for the development of even better alternatives):

1. If a realistic model is available but rainfall data are limited and of insufficient quality and the study's focus is to estimate the physical properties of the catchment, the dynamics of the catchment-averaged rain rate, or the contribution of the error sources to output (runoff) uncertainty, then we suggest using SIP whenever this is computationally feasible.
2. Under the same conditions as above, but if the model running time and the length of the time series make it impossible to perform millions of MCMC simulations, we suggest using RM.
3. If a realistic model is available and input and output data are of high quality, and the study focus is to estimate the physical properties of the catchment and to predict its output, then we suggest using LS. Note that these conditions are rarely met.
4. If the available model is structurally deficient and the study focus is to reliably predict runoff, then we suggest using BD. BD will not substantially help, however, to understand, disentangle, or minimize uncertainty.
5. If the model is structurally deficient (or, equivalently, if the output observations are inaccurate), and the input is poorly observed, we recommend to combine BD with either SIP or RM, depending on the computational possibilities. This is particularly relevant when, besides output predictions, input uncertainty estimation is also of interest. Although this is an extension of the four alternatives discussed in this paper, it should be straightforward to apply. In the next section, we will provide an outlook to promising developments toward even better techniques.

5.4. Outlook

In this first application of SIP, as done in several studies on input uncertainty [*Kavetski et al*., 2006; *McMillan et al*., 2011; *Sun and Bertrand-Krajewski*, 2013], for the sake of simplicity, we deliberately adopted a simple output error model similar to LS. As demonstrated by the very low bias identified ($\sigma_B < \sigma_E$) and by the high NS obtained with Sc1, for our system-model combination, it was plausible to assume minimal structural errors. However, since the long-term goal is to also target more realistic situations, we suggest some future directions of research to deal with input and structural errors in hydrology.

1. Further developing the numerics of SIP by advancing sampling techniques for inference based on likelihoods that are formulated as infinite dimensional integrals, is an important focus for research. One

option could be to use so-called "Hamiltonian Monte Carlo" algorithms, a promising class of methods profiting from concepts of molecular dynamics to increase the efficiency of MCMC schemes [*Brooks et al*., 2011; see also C. Albert and S. Ulzega, Bayesian parameter inference for 1D nonlinear stochastic differential equation models, submitted to New Journal of Physics, 2015]. Note that addressing errors where they are generated, although computationally demanding, is conceptually preferable to correcting the output for the effects of these errors.

2. Making more realistic distributional assumptions for SIP by not relying on a transformed Ornstein-Uhlenbeck process could lead to a better description of our prior knowledge on the rain rate. For an overview of recent promising approaches to model precipitation using e.g., copulas or multifractals, the reader can refer to *Paschalis et al*. [2013] or *Ailliot et al*. [2015]. These more sophisticated precipitation models, however, would require a different numerical approach for hydrological inference. Research under point 1 above could contribute to making this feasible.
3. Combining SIP with techniques describing model structural deficits to make it possible to infer the input jointly with model structural deficits. This is a very important research direction as we often have both sources of error and are interested in disentangling their contributions to the overall output error [*Salamon and Feyen*, 2010; *Renard et al*., 2011]. A promising way of doing this is to combine SIP with stochastic, time-dependent parameters as outlined in *Reichert and Mieleitner* [2009]. This approach, although computationally demanding, would enable us to directly capture the sources of uncertainty. Combining SIP with BD or similar autoregressive output error models (as done for RM by *Sikorska et al*. [2012] and *Li et al*. [2012]) could be also a pragmatic alternative. In both cases, however, to minimize the identifiability problem between model parameters and stochastic processes, the use of realistic priors for input errors [*Renard et al*., 2011] or model discrepancy [*Brynjarsdóttir and O'Hagan*, 2014] will be decisive.
4. Extending the input error model to combine different types of input data, such as those from radars or microwave links in addition to rain gauges, could further reduce input uncertainty. Indeed, these alternatives provide better information on spatially integrated rain rates than rain gauges.

6. Conclusions

In this study, we aimed at improving parameter inference, better estimating areal precipitation, and contributing to uncertainty separation in hydrological modeling. The main novelty of this work is to perform hydrological inference with a more realistic input error model. In particular, we suggest describing the catchment-averaged precipitation as a stochastic input process (SIP). This appropriately parameterized and transformed Ornstein-Uhlenbeck process is updated in a Bayesian framework jointly with model parameters by combining rainfall data (the input), system understanding (the hydrological model), and runoff data (the output). We applied SIP to a parsimonious urban rainfall-runoff model and compared the effects of optimal versus mediocre rainfall data. For a better understanding of SIP performance, we compared its results with those obtained with simpler methods, namely the standard least squares (LS), the rainfall multipliers (RM), and the bias description (BD). By combining conceptual arguments with the results of our case study, we conclude that:

1. SIP can effectively deal with severe input errors such as unrecorded or temporally shifted rainfall peaks. In such situations, simpler methods assuming multiplicative forcing errors provide inaccurate rainfall estimates and biased model parameter values. As shown in our real-world application, given an accurate hydrological model, high-quality discharge data, and inaccurate precipitation data, SIP was the only method capable of accurately reconstructing the whole-catchment precipitation and reliably quantifying input uncertainty.
2. In our case study, when forcing the model with inaccurate input data, similar to BD, SIP was able to produce physically consistent parameters. Simpler methods such as LS and RM instead produced biased parameter estimates. Furthermore, SIP estimated input uncertainty more reliably than RM, also in prediction mode.
3. Despite those advantages over previous methods, the increased computational requirements of SIP can be limiting for practical applications. Furthermore, as RM, SIP can unintentionally compensate for model structural deficits by incorrectly adjusting the input.
4. We recommend SIP to reduce the corrupting effects of input uncertainty on hydrological model parameters and to estimate the input to a catchment in an accurate probabilistic way. Producing rainfall realizations at every desired temporal resolution, SIP can also be useful to reconstruct past precipitation and support temporal downscaling of precipitation records. Further developments will aim at improving its

numerical efficiency and extending its applicability to the consideration of structurally inadequate models.

Appendix A: Log-sinh Transformation

The log-sinh transformation has recently shown very promising results for hydrological applications [*Wang et al.*, 2012; *Del Giudice et al.*, 2013]. In contrast to the original notation, we suggest a reparameterized notation with parameters that have a more intuitive meaning:

$$g(y) = \beta \log\left(\sinh\left(\frac{\alpha + y}{\beta}\right)\right), \tag{A1}$$

$$g^{-1}(z) = \left(\operatorname{arcsinh}\left(\exp\left(\frac{z}{\beta}\right)\right) - \frac{\alpha}{\beta}\right)\beta, \tag{A2}$$

$$\frac{dg}{dy} = \coth\left(\frac{\alpha + y}{\beta}\right), \tag{A3}$$

where α and β are "low" and "high" outputs, relative to observations. α and β control the degree of heteroscedasticity of the predictions (higher when $\alpha \ll \beta$). As in the aforementioned studies, we chose $\beta = 50$ L/s to be an intermediately high discharge above which uncertainty was assumed not to significantly increase. To ensure a mild degree of transformation, we set $\alpha = 25$ L/s (Figure S9). This provided a plausible representation of the output-dependent uncertainties with the best rainfall scenario and all error models and enabled predictive intervals to properly encompass high and low flow data.

Acknowledgments
The data and codes used are available upon request from the first author. The authors are very grateful to Tobias Doppler for data collection and compilation, and Hans Rudolf Künsch for stimulating discussions. We additionally would like to thank Anneli Schöniger, Fabrizio Fenicia, Anna Sikorska, Simone Ulzega, and two anonymous reviewers for their helpful suggestions that improved the manuscript. This work was supported by the Swiss National Science Foundation (grant CR2212_135551).

References

Ailliot, P., D. Allard, V. Monbet, and P. Naveau (2015), Stochastic weather generators: An overview of weather type models, *J. Soc. Fr. Stat.*, *156*, 101–113.
Bayarri, M., J. Berger, R. Paulo, J. Sacks, J. Cafeo, J. Cavendish, C. Lin, and J. Tu (2007), A framework for validation of computer models, *Technometrics*, *49*, 138–154.
Bardossy, A., and T. Das (2008), Influence of rainfall observation network on model calibration and application, *Hydrol. Earth Syst. Sci.*, *12*(1), 77–89, doi:10.5194/hess-12-77-2008.
Beven, K. J., and I. Westerberg (2011), On red herrings and real herrings: Disinformation and information in hydrological inference, *Hydrol. Processes*, *25*, 1676–1680, doi:10.1002/hyp.7963.
Brooks, S., A. Gelman, G. Jones, and X. Meng (2011), *Handbook of Markov Chain Monte Carlo, (Handbooks of Modern Statistical Methods)*, Chapman and Hall, Boca Raton, Fla.
Brynjarsdóttir, J., and A. O'Hagan (2014), Learning about physical parameters: The importance of model discrepancy, *Inverse Problems*, *30*, 114007.
Coutu, S., D. Del Giudice, L. Rossi, and D. Barry (2012), Parsimonious hydrological modeling of urban sewer and river catchments, *J. Hydrol.*, *464–465*, 477–484, doi:10.1016/j.jhydrol.2012.07.039.
Cowpertwait, P. S. P., P. E. O'Connell, A. V. Metcalfe, and J. A. Mawdsley (1996), Stochastic point process modelling of rainfall. I. Single-site fitting and validation, *J. Hydrol.*, *175*, 17–46.
Deidda, R., R. Benzi, and F. Siccardi (1999), Multifractal modeling of anomalous scaling laws in rainfall, *Water Resour. Res.*, *35*(6), 1853–1867.
Del Giudice, D., M. Honti, A. Scheidegger, C. Albert, P. Reichert, and J. Rieckermann (2013), Improving uncertainty estimation in urban hydrological modeling by statistically describing bias, *Hydrol. Earth Syst. Sci.*, *17*, 4209–4225, doi:10.5194/hess-17-4209-2013.
Del Giudice, D., P. Reichert, V. Bareš, C. Albert, and J. Rieckermann (2015a), Model bias and complexity - understanding the effects of structural deficits and input errors on runoff predictions, *Environ. Modell. Software*, *64*, 205–214, doi:10.1016/j.envsoft.2014.11.006.
Del Giudice, D., R. Löwe, H. Madsen, P. S. Mikkelsen, and J. Rieckermann (2015b), Comparison of two stochastic techniques for reliable urban runoff predictions by modeling systematic errors, *Water Resour. Res.*, *51*, 5004–5022, doi:10.1002/2014WR016678.
Dietzel, A., and P. Reichert (2014), Bayesian inference of a lake water quality model by emulating its posterior density, *Water Resour. Res.*, *50*, 7626–7647, doi:10.1002/2012WR013086.
Evin, G., D. Kavetski, M. Thyer, and G. Kuczera (2013), Pitfalls and improvements in the joint inference of heteroscedasticity and autocorrelation in hydrological model calibration, *Water Resour. Res.*, *49*, 4518–4524, doi:10.1002/wrcr.20284.
Frey, M. P., C. Stamm, M. K. Schneider, and P. Reichert (2011), Using discharge data to reduce structural deficits in a hydrological model with a Bayesian inference approach and the implications for the prediction of critical source areas, *Water Resour. Res.*, *47*, W12529, doi:10.1029/2010WR009993.
Gneiting, T., and A. E. Raftery (2007), Strictly proper scoring rules, prediction, and estimation, *J. Am. Stat. Assoc.*, *102*(477), 359–378, doi:10.1198/016214506000001437.
Honti, M., C. Stamm, and P. Reichert (2013), Integrated uncertainty assessment of discharge predictions with a statistical error model, *Water Resour. Res.*, *49*, 4866–4884, doi:10.1002/wrcr.20374.
Ibe, O. C. (2009), *Markov Processes for Stochastic Modeling*, Elsevier Sci. [Available at http://store.elsevier.com/product.jsp?isbn=9780080922454&pagename=search.]
Kennedy, M., and A. O'Hagan (2001), Bayesian calibration of computer models, *J. R. Stat. Soc. Ser. B*, *63*, 425–464.
Kirchner, J. W. (2009), Catchments as simple dynamical systems: Catchment characterization, rainfall-runoff modeling, and doing hydrology backward, *Water Resources Research*, *45*, *W02429*, doi:10.1029/2008WR006912.

Kroese, D., T. Taimre, and Z. Botev (2011), *Handbook of Monte Carlo Methods*, John Wiley, Hoboken, N. J.
Kuczera, G. (1983), Improved parameter inference in catchment models: 1. Evaluating parameter uncertainty, *Water Resour. Res.*, *19*(5), 1151–1162.
Langousis, A., and V. Kaleris (2014), Statistical framework to simulate daily rainfall series conditional on upper-air predictor variables, *Water Resour. Res.*, *50*, 3907–3932, doi:10.1002/2013WR014936.
Li, M., D. Yang, J. Chen, and S. Hubbard (2012), Calibration of a distributed flood forecasting model with input uncertainty using a Bayesian framework, *Water Resour. Res.*, *48*, W08510, doi:10.1029/2010WR010062.
Kavetski, D., G. Kuczera, and S. W. Franks (2006), Bayesian analysis of input uncertainty in hydrological modeling:1. Theory, *Water Resour. Res.*, *42*, W03407, doi:10.1029/2005WR004368.
McMillan, H., B. Jackson, M. Clark, D. Kavetski, and R. Woods (2011), Rainfall uncertainty in hydrological modelling: An evaluation of multiplicative error models, *J. Hydrol.*, *400*, 83–94.
Ochoa-Rodriguez, S., et al. (2015), Impact of spatial and temporal resolution of rainfall inputs on urban hydrodynamic modelling outputs: A multi-catchment investigation, *J. Hydrol.*, *531*, 389–407, doi:10.1016/j.jhydrol.2015.05.035.
Paschalis, A., P. Molnar, S. Fatichi, and P. Burlando (2013), A stochastic model for high-resolution space-time precipitation simulation, *Water Resour. Res.*, *49*, 8400–8417, doi:10.1002/2013WR014437.
Paul, W., and J. Baschnagel (2013), *Stochastic Processes: From Physics to Finance*, Springer. [Available at http://www.springer.com/us/book/9783319003269.]
Platen, E., and Bruti-Liberati, N. (2010), *Numerical Solution of Stochastic Differential Equations with Jumps in Finance*, vol. 64, Springer. [Available at http://www.springer.com/us/book/9783642120572.]
Quinn, J. C., and H. D. Abarbanel (2010), State and parameter estimation using Monte Carlo evaluation of path integrals, *Q. J. R. Meteorol. Soc.*, *136*, 1855–1867, doi:10.1002/qj.690.
R Core Team (2014), *R: A Language and Environment for Statistical Computing*, R Found. for Stat. Comput., Vienna, Austria. [Available at http://www.R-project.org/.]
Renard, B., D. Kavetski, E. Leblois, M. Thyer, G. Kuczera, and S. W. Franks (2011), Toward a reliable decomposition of predictive uncertainty in hydrological modeling: Characterizing rainfall errors using conditional simulation, *Water Resour. Res.*, *47*, W11516, doi:10.1029/2011WR010643.
Reichert, P. and J. Mieleitner (2009), Analyzing input and structural uncertainty of nonlinear dynamic models with stochastic, time-dependent parameters, *Water Resour. Res.*, *45*, W10402, doi:10.1029/2009WR007814.
Reichert, P., and N. Schuwirth (2012), Linking statistical bias description to multiobjective model calibration, *Water Resour. Res.*, *48*, W09543, doi:10.1029/2011WR011391.
Rodriguez-Iturbe, I., D. R. Cox, and V. Isham (1987), Some models for rainfall based on stochastic point processes, *Proc. R. Soc. London, Ser. A*, *410*, 269–298.
Rougier, J. (2013), 'Intractable and unsolved': Some thoughts on statistical data assimilation with uncertain static parameters, *Philos. Trans. R. Soc. A*, *371*, 20120297, doi:10.1098/rsta.2012.0297.
Salamon, P., and L. Feyen (2010), Disentangling uncertainties in distributed hydrological modeling using multiplicative error models and sequential data assimilation, *Water Resour. Res.*, *46*, W12501, doi:10.1029/2009WR009022.
Sigrist, F., H. Künsch, and W. Stahel (2012), A dynamic nonstationary spatiotemporal model for short term prediction of precipitation, *Ann. Appl. Stat.*, *6*(4), 1452–1477, doi:10.1214/12-AOAS564.
Sikorska, A., A. Scheidegger, K. Banasik, and J. Rieckermann (2012), Bayesian uncertainty assessment of flood predictions in ungauged urban basins for conceptual rainfall-runoff models, *Hydrol. Earth Syst. Sci.*, *16*, 1221–1236, doi:10.5194/hess-16-1221-2012.
Sorooshian, S., and J. A. Dracup (1980), Stochastic parameter estimation procedures for hydrologiy rainfall-runoff models: Correlated and heteroscedastic error cases, *Water Resour. Res.*, *16*(2), 430–442.
Sun, S., and J. Bertrand-Krajewski (2013), Separately accounting for uncertainties in rainfall and runoff: Calibration of event-based conceptual hydrological models in small urban catchments using Bayesian method, *Water Resour. Res.*, *49*, 5381–5394, doi:10.1002/wrcr.20444.
Tomassini, L., P. Reichert, H. R. Künsch, C. Buser, R. Knutti, and M. E. Borsuk (2009), A smoothing algorithm for estimating stochastic, continuous- time parameters and its application to a simple climate model, *J. R. Stat. Soc., Ser. C*, *85*, 679–704.
Vrugt, J. A., C. J. F. ter Braak, M. P. Clark, J. M. Hyman, and B. A. Robinson (2008), Treatment of input uncertainty in hydrologic modeling: Doing hydrology backward with Markov chain Monte Carlo simulation, *Water Resour. Res.*, *44*, W00B09, doi:10.1029/2007WR006720.
Wang, Q. J., D. L. Shrestha, D. E. Robertson, and P. Pokhrel (2012), A log-sinh transformation for data normalization and variance stabilization, *Water Resour. Res.*, *48*, W05514, doi:10.1029/2011WR010973.
Yang, J., P. Reichert, K. C. Abbaspour and H. Yang (2007a), Hydrological modelling of the Chaohe Basin in China: Statistical model formulation and Bayesian inference, *J. Hydrol.*, *340*, 167–182, doi:10.1016/j.jhydrol.2007.04.006.
Yang, J., P. Reichert, and K. C. Abbaspour (2007b), Bayesian uncertainty analysis in distributed hydrologic modeling: A case study in the Thur River basin (Switzerland), *Water Resour. Res.*, *43*, W10401, doi:10.1029/2006WR005497.
Yang, J., P. Reichert, K. C. Abbaspour, J. Xia, and H. Yang (2008), Comparing uncertainty analysis techniques for a SWAT application to the Chaohe Basin in China, *J. Hydrol.*, *358*, 1–23.