Technical University of Denmark

DTU

Comparison of two stochastic techniques for reliable urban runoff prediction by modeling systematic errors

Del Giudice, Dario; Löwe, Roland; Madsen, Henrik; Mikkelsen, Peter Steen; Rieckermann, Jörg

Link to publication

RESEARCH ARTICLE

Key Points:
- We describe systematic errors of drainage models with two statistical approaches
- Both methods reliably assess the uncertainty of urban hydrologic simulations
- One is optimal for long-term predictions, the other for real-time forecasts

Comparison of two stochastic techniques for reliable urban runoff prediction by modeling systematic errors

Dario Del Giudice[1,2], Roland Löwe[3], Henrik Madsen[3], Peter Steen Mikkelsen[4], and Jörg Rieckermann[1]

[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]Department of Applied Mathematics and Computer Science, Technical University of Denmark, Lyngby, Denmark, [4]Department of Environmental Engineering, Technical University of Denmark, Lyngby, Denmark

Abstract In urban rainfall-runoff, commonly applied statistical techniques for uncertainty quantification mostly ignore systematic output errors originating from simplified models and erroneous inputs. Consequently, the resulting predictive uncertainty is often unreliable. Our objective is to present two approaches which use stochastic processes to describe systematic deviations and to discuss their advantages and drawbacks for urban drainage modeling. The two methodologies are an external bias description (EBD) and an internal noise description (IND, also known as stochastic gray-box modeling). They emerge from different fields and have not yet been compared in environmental modeling. To compare the two approaches, we develop a unifying terminology, evaluate them theoretically, and apply them to conceptual rainfall-runoff modeling in the same drainage system. Our results show that both approaches can provide probabilistic predictions of wastewater discharge in a similarly reliable way, both for periods ranging from a few hours up to more than 1 week ahead of time. The EBD produces more accurate predictions on long horizons but relies on computationally heavy MCMC routines for parameter inferences. These properties make it more suitable for off-line applications. The IND can help in diagnosing the causes of output errors and is computationally inexpensive. It produces best results on short forecast horizons that are typical for online applications.

1. Introduction

Any model in urban hydrology usually delivers results that substantially differ from observations of water level, flow, or water quality [*Dotto et al.*, 2012]. These mismatches between modeled and observed output are caused by errors in the input estimation and by simplifications of the system description [*Del Giudice et al.*, 2013]. These systematic output deviations can affect the operation of urban drainage and wastewater systems as well as design decisions, which are usually based on model predictions [*Vezzaro and Grum*, 2014]. Consequently, an appropriate description of these systematic deviations can meliorate forecasting and control [*Löwe et al.*, 2014]. Significant efforts have therefore been made in past and recent hydrological research to quantify the uncertainties of model results [*Jonsdottir et al.*, 2007; *Yang et al.*, 2007; *Salamon and Feyen*, 2010; *Breinholt et al.*, 2012; *Freni and Mannina*, 2012; *Sikorska et al.*, 2012; *Evin et al.*, 2013; *Honti et al.*, 2013].

Runoff modeling in urban hydrology distinguishes itself from its counterpart in natural catchment hydrology by the usually smaller temporal and spatial scales involved in peak discharge generation. Typical time steps for peak discharge simulations are 6 [*Kleidorfer et al.*, 2009] to 15 min [*Breinholt et al.*, 2012], but seconds [*Freni et al.*, 2009] to days [*Mejía et al.*, 2014] have been reported. Typical study areas of sewer watersheds range from dozens [*Del Giudice et al.*, 2015] to more than 1000 ha [*Breinholt et al.*, 2011]. Furthermore, the majority of sewer peak flow comes from sealed surfaces which dominate urban landscapes [*Coutu et al.*, 2012]. As a result, concentration times of 1 h or less are common, which makes model predictions highly sensitive to variations of rainfall input on small scales. This sensitivity to input uncertainty was underlined by previous investigations which suggested that forecasting errors are mainly due to discrepancies in the rainfall input, in particular an insufficient quantification of the spatial rainfall distribution on a scale of a few kilometers or less [*Schilling and Fuchs*, 1986; *Sikorska et al.*, 2012; *Borup et al.*, 2013].

The systematic rainfall errors, their routing through a possibly nonlinear model, and deficits in the model structure usually lead to an autocorrelated and heteroscedastic behavior of the residuals of runoff simulations [see *Reichert and Mieleitner*, 2009; *Evin et al.*, 2013]. Most of the techniques applied for uncertainty quantification in urban hydrology do not explicitly account for this dynamic nature of model errors. Typically, only parametric uncertainty and output measurement noise are considered. This usually leads to biased parameter estimates and to suboptimal forecasting [*Thyer et al.*, 2009; *Schoups and Vrugt*, 2010; *Willems*, 2012; *Del Giudice et al.*, 2013].

Recent developments have focused on the attempt to account for systematic behavior of runoff model residuals (by some authors referred to as model bias or discrepancy). The present work aims at comparing two such approaches that have recently been applied in urban hydrology [*Bechmann et al.*, 2000; *Breinholt et al.*, 2012; *Del Giudice et al.*, 2013; *Löwe et al.*, 2014]. In the following, we will denote them as "external bias description" (EBD) and "internal noise description" (IND). Both approaches aim at describing and compensating for the dynamic variations of model residuals. However, they are implemented in different mathematical frameworks, originate from different scientific fields, utilize a distinct terminology, and to date focus on dissimilar applications.

The EBD, on the one hand, was developed against the background of statistical inference in a regression-type framework [see *Craig et al.*, 2001; *Kennedy and O'Hagan*, 2001; *Higdon et al.*, 2005; *Bayarri et al.*, 2007; *Reichert and Schuwirth*, 2012, for example] and has a strong focus on the estimation of parameters and system output, as well as their related uncertainties. The IND, on the other hand, originated from research related to stochastic processes and time series analysis and was originally applied to forecasting and control of engineered systems such as chemical reactors or heating systems [see *Bechmann et al.*, 2000; *Kristensen et al.*, 2004, 2005; *Friling et al.*, 2009, for example].

Based on the existing literature, it is difficult to identify the relative advantages and disadvantages of the approaches and to make recommendations on their overall applicability which depends on forecasting horizon and model type. Therefore, the main objectives and innovations of this work are to

1. Present in commensurate terms two advanced approaches for probabilistic model calibration and predictions. Because of their different origins, the EBD and IND have been presented with dissimilar "idioms," which has hindered the collaboration between their respective communities.

2. Explore new aspects of the two approaches. For the EBD, this implies testing its performances in short-term predictions, in combined sewer flow modeling, and in the presence of substantial and nonstationary model deficiencies. For the IND, this means testing its performances in discrete short-term and long-term predictions, observing the uncertainty expansion from the last observation point, and discussing its likelihood function in more detail.

3. Discuss the lessons learned from the two approaches and their respective strengths and weaknesses. To do so, we consider both theoretical aspects and the performances of the EBD and IND when applied to a common and complex system and an oversimplified model.

The discussions and results of this investigation will help the modeler to make a more conscious choice about which method to adopt. This choice will depend on the study resources (e.g., black-box/modifiable model, sufficient/limited computational power) and goals (e.g., predicting over long/short horizons). Furthermore, the reciprocal understanding of the EBD and IND ensuing from this study will help direct future developments of both approaches.

2. Brief Review of Methods Applied for Uncertainty Quantification in Conceptual Rainfall-Runoff Modeling

This section provides a brief overview of the techniques applied for quantifying uncertainties, with a focus on conceptual rainfall-runoff modeling. We classify the techniques as shown in Table 1 according to their main characteristics: model formulation (rows) and representation of the errors (columns).

A natural distinction of the different approaches derives from the way the model is formulated [*Renard et al.*, 2010]. In hydrology, we traditionally model the output of a system by using a deterministic model (or simulator). The model output can then be combined with one or more probabilistic error terms. This

Table 1. Probabilistic Approaches for Runoff Predictions[a]

	Errors iid	Systematic Deviations Described	Error Sources Represented
Output error modeling (deterministic model + stochastic errors)	*Dotto et al.* [2012]* *Freni and Mannina* [2012]* *Kleidorfer et al.* [2009]* *Vezzaro et al.* [2013]*	*Del Giudice et al.* [2013]* *Kuczera* [1983] *Schoups and Vrugt* [2010] *Wilkinson et al.* [2011]	*Kavetski et al.* [2006] *Renard et al.* [2010] *Sikorska et al.* [2012]* *Sun and Bertrand-Krajewski* [2013]*
Internal error modeling (stochastic model + stochastic errors)		*Breinholt et al.* [2011, 2012]* *Löwe et al.* [2014]* *Moradkhani et al.* [2012]	*Beck and Young* [1976] *Vrugt et al.* [2005] *Bulygina and Gupta* [2009] *Reichert and Mieleitner* [2009] *Salamon and Feyen* [2010]

[a]We included examples from the urban drainage (marked with an asterisk) and natural hydrology literature. Note that it is not possible to assume the residual errors to be independent and identically distributed (iid) when the system equations contain a noise term.

approach is shown in the first row of Table 1 and we denote it as "output error modeling." Alternatively, the model itself can be stochastic. This is usually done by considering the model states [e.g., in *Vrugt et al.*, 2005; *Breinholt et al.*, 2012; *Moradkhani et al.*, 2012] or parameters [e.g., in *Beck and Young*, 1976; *Reichert and Mieleitner*, 2009] as time-varying, random variables. Such approaches are usually implemented in a state space form, which is common in system theory and statistical filtering [*Lin and Beck*, 2007; *Bulygina and Gupta*, 2009; *Quinn and Abarbanel*, 2010]. The model output, a function of these stochastic states, is additionally affected by an observation error term, and the approach is usually combined with data assimilation methods. We denote these approaches as "internal error modeling" and summarize them in the second row of Table 1.

Complementary to how they formulate the model, methods for uncertainty analysis of runoff predictions can be classified by how they characterize modeling errors (columns in Table 1). We suggest distinguishing between three cases:

1. Approaches that do not explicitly account for dynamic model discrepancies. These may be Bayesian approaches which assume uncorrelated model residuals or pseudo-Bayesian approaches (such as GLUE) [*Beven*, 1993]. Common to these frameworks is that input and structural uncertainties are assigned to the (constant) model parameters (see discussions in *Yang et al.* [2008] and *Reichert and Mieleitner* [2009]). As a result, parameter estimates can become difficult to interpret and the resulting output prediction intervals may be unreliable [*Renard et al.*, 2010; *Reichert and Schuwirth*, 2012].

2. Approaches that explicitly account for dynamic model discrepancies. In the case of output error modeling, this can be done by adding a time-varying error term to the model output (for example ARMA models as already suggested by *Kuczera* [1983] or stochastic differential equations (SDEs) as in *Yang et al.* [2008] and *Del Giudice et al.* [2013]). In the case of internal error modeling, a random noise is added to the states to reflect that they can rarely be predicted exactly [see *Breinholt et al.*, 2012, for example]. The state noise provides a quantification of forecast uncertainties. In both methods, structural and input uncertainties are aggregated into one term.

3. Approaches that, instead of just describing the output errors, focus on identifying the causes of model inadequacies. To quantify input uncertainty, rainfall multipliers have been proposed [*Kuczera et al.*, 2006; *Sun and Bertrand-Krajewski*, 2013]. Structural uncertainty, instead, has been dealt with by inferring the model equations [*Bulygina and Gupta*, 2009], the behavior of dynamic parameters [*Reichert and Mieleitner*, 2009], or the value of model parameters and states [*Vrugt et al.*, 2005].

From the literature [e.g., *Dotto et al.*, 2012; *Sikorska et al.*, 2012; *Del Giudice et al.*, 2013], it is clear that the majority of uncertainty studies in urban hydrology does not account for time-dependent systematic model errors. In contrast, the two approaches considered in this article explicitly account for systematic dynamic output errors (second column of Table 1). However, they are generally less conceptually complex and computationally demanding than those presented in the third column of Table 1. The EBD is an output error modeling approach (first row of Table 1), while the IND is an internal error modeling approach (second row of Table 1). The works of *Breinholt et al.* [2012] and *Del Giudice et al.* [2013] in urban hydrology and multiple works in natural catchment hydrology (see Table 1) have demonstrated that such approaches are generally capable of producing reliable predictions in conceptual rainfall-runoff modeling.

3. Methods

3.1. Terminology

We here provide a brief unifying nomenclature to describe our analyses with the two methodologies. We also mention alternative terminology used in hydrology, statistics, and control theory. An illustrative description of this terminology is given in Figure 1.

Parameter estimation consists in identifying parameter values by comparing the model and the output observations. This learning process is also known as parameter inference [*Reichert and Schuwirth*, 2012], calibration [*O'Hagan*, 2006], or inverse modeling.

Smoothing refers to identifying system states and/or outputs in a past time, e.g., the calibration period, using the available data before and after that point [*Bulygina and Gupta*, 2009; *Law and Stuart*, 2012].

Forecasting denotes the generation of model outputs (and states) starting from the last observation up to an arbitrary number of time steps in the future. This process is also loosely described as making predictions (in the validation period) [*Dietzel and Reichert*, 2012; *Renard et al.*, 2010; *Law and Stuart*, 2012; *Einicke*, 2012], simulating [*Platen and Bruti-Liberati*, 2010] or, more precisely, ex-post hindcasting (when the input is assumed to be known) [*Beven and Young*, 2013].

Filtering consists in characterizing the system state at the current time given inputs and observations up to the current point [*Bulygina and Gupta*, 2009; *Platen and Bruti-Liberati*, 2010; *Law and Stuart*, 2012]. Data assimilation is also used to define this process of learning about the current state [*O'Hagan*, 2006].

3.2. Two Approaches to Explicitly Account for Dynamic Systematic Errors in Rainfall-Runoff Modeling

We here explain the external bias description (EBD) and the internal noise description (IND). While the first adds a stochastic process to the system output, the second adds a stochastic process to the states *and* to the output.

3.2.1. Output Error Modeling and External Bias Description (EBD)

In deterministic conceptual modeling, differential equations are applied to describe the variation d**s** of a set of model states **s** (e.g., water level in an unobserved combined sewer overflow tank, hydraulic heads in specific points of an aquifer, or soil moisture content in a catchment) depending on a vector of driving forces (e.g., a rainfall time series) **x** and parameters θ in a function f_M (equation (1)). Bold minuscules denote deterministic vectors while bold majuscules denote stochastic vectors.

$$\frac{d\mathbf{s}}{dt} = f_M(\mathbf{s}, \mathbf{x}, t, \theta). \tag{1}$$

The model output $\mathbf{y}_M$ relates to the model states, input, and parameters through a function h:

$$\mathbf{y}_M = h(\mathbf{s}, \mathbf{x}, t, \theta). \tag{2}$$

So far, no modeling error has been considered. In order to account for the fact that no system description is perfect and that output observations are affected by errors, two strategies are possible: external or internal error modeling. In external (or output) error modeling, the observed system output $\mathbf{Y}_o$ (e.g., measured discharge just before the entrance of a sewage treatment plant) can be represented as the sum of $\mathbf{y}_M$ plus a stochastic error term. This term aggregates modeling and observation errors and can be independently and identically distributed (iid) [e.g., in *Kleidorfer et al.*, 2009; *Freni and Mannina*, 2012], or autocorrelated in time [e.g., in *Kuczera*, 1983; *Bates and Campbell*, 2001; *Frey et al.*, 2011; *Evin et al.*, 2013]. Several studies [e.g., *Yang et al.*, 2007; *Sikorska et al.*, 2012; *Honti et al.*, 2013] have demonstrated that describing the autocorrelated behavior of the errors produces more reliable predictions. Instead of adding only one autocorrelated error term, recent statistical literature has suggested considering observation noise in addition to input, structural, and parameter uncertainty (equation (3)) [*Craig et al.*, 2001; *Kennedy and O'Hagan*, 2001; *Higdon et al.*, 2005; *Bayarri et al.*, 2007]. Following the notation of *Reichert and Schuwirth* [2012], who transferred this approach to environmental modeling, we model the observable system output as

$$\mathbf{Y}_o = \mathbf{y}_M(\mathbf{s}, \mathbf{x}, t, \theta) + \mathbf{B}_M(\mathbf{x}, t, \psi) + \mathbf{E}(\psi), \tag{3}$$

where $\mathbf{B}_M$ is a random process that mimics systematic deviation of model results from the true system output, $\mathbf{E}$ represents uncorrelated observation errors, and (θ, ψ) are the parameters of the simulator and error model. Simplified iid approaches only consider $\mathbf{E}$ while neglecting $\mathbf{B}_M$.
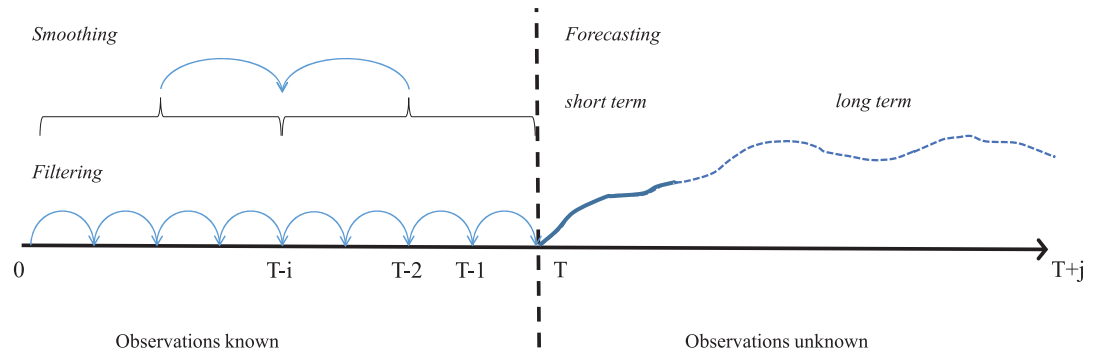
Figure 1. Illustration of the different types of predictions according to the conditioning on output observations.

To further improve the error description, modeled and observed outputs could be transformed by a function. This can be useful in hydrology, where the error variance increases during peak discharge. This effect can, however, also be reproduced by a heteroschedastic error model [*Evin et al.*, 2013; *Del Giudice et al.*, 2013]. In this study, we achieve satisfactory results with an input-dependent bias description. The specific formulation we use assumes that the bias follows an Ornstein-Uhlenbeck process with input-dependent variance [see *Honti et al.*, 2013, for derivation]. In other words, $\mathbf{B}_M$ is modeled as a continuous version of a first-order autoregressive process with normal independent noise whose variance grows with the rain rate, x, shifted in time by a lag d. The evolution of $\mathbf{B}_M$ and $\mathbf{E}$ for the scalar case is described by equations (4) and (5):

$$dB_M(t) = -\frac{B_M(t)}{\tau}dt + \sqrt{\frac{2}{\tau}\left(\sigma_{B_{ct}}^2 + (\kappa x(t-d))^2\right)}dW(t), \qquad (4)$$

$$E(t) = \sigma_E E_{\mathcal{N}}, \qquad (5)$$

where κ is a scaling factor, d denotes the response time of the system to rainfall, τ is the correlation time of the error process, and $\sigma_{B_{ct}}$ is the asymptotic standard deviation of the random fluctuations around the equilibrium. $dW(t)$ represents increments of a standard Wiener process and therefore has a normal distribution [*Kloeden and Platen*, 1992; *Iacus*, 2008], while $E_{\mathcal{N}}$ is a standard normal random variable.

3.2.2. Internal Error Modeling and Internal Noise Description (IND)

An alternative way to account for uncertainties when modeling the behavior of a hydrosystem with equations (1) and (2) is via internal error modeling, which is usually applied in combination with state updating [*Kristensen et al.*, 2004; *Moradkhani et al.*, 2012]. Instead of adding stochasticity only to the system output, this approach (also known as state space modeling or stochastic gray-box modeling) describes the internal evolution of the system as

$$d\mathbf{S} = f_M(\mathbf{S}, \mathbf{x}, t, \boldsymbol{\theta})dt + \boldsymbol{\sigma}(\mathbf{S}, \mathbf{x}, t, \boldsymbol{\psi})d\mathbf{W}(t). \qquad (6)$$

This so-called "state" (or "transition," or "system") equation describes the continuous evolution of some "hidden" (or "latent") states $\mathbf{S}$ which, being now stochastic, directly account for modeling errors. This vector of usually unmeasurable variables can be estimated from the measured outputs [*Einicke*, 2012]. $\boldsymbol{\sigma}$ is called "diffusion term," "state noise," or "level disturbance" and accounts for modeling errors by making the states uncertain or random. $f_M(\cdot)$ is called "drift term" and corresponds to the functions constituting the deterministic (part of the) model M. Adding noise to the state equations reflects that the states cannot be predicted exactly. Hence, the model itself is stochastic. This is an important distinction from the EBD, where randomness is only added to the model output. The IND is instead more similar to approaches making model parameters stochastic and time varying [*Reichert and Mieleitner*, 2009].

The dynamics of the observed output $\mathbf{Y}_o$ are related to the state equations via an observation equation:

$$\mathbf{Y}_o = h(\mathbf{S}, \mathbf{x}, \boldsymbol{\theta}, \boldsymbol{\psi}, t) + \mathbf{E}(\boldsymbol{\psi}), \qquad (7)$$

which is a potentially nonlinear function of states **S** and parameters (θ, ψ). The modeled observation process $\mathbf{Y}_o$ is assumed to be subject to independent random normal observation errors **E**. Similarly to the EBD, transformations can be applied to the observed and modeled output [*Breinholt et al.*, 2012].

We here parametrized the diffusion term as linearly increasing with the model states

$$\boldsymbol{\sigma}(\mathbf{S}, \mathbf{x}, t, \psi) = diag(\boldsymbol{\sigma}_s \circ \mathbf{S}), \tag{8}$$

where $\circ$ is the Hadamard (entrywise) product between the vector of diffusion parameters $\boldsymbol{\sigma}_s$ and the vector of states **S**, and *diag* indicates that the matrix is diagonal. This formulation produced satisfactory results in previous urban hydrological studies [*Breinholt et al.*, 2011; *Löwe et al.*, 2014]. The assumption of state-dependent noise in the IND is another relevant distinction from the EBD, where the additive noise terms can depend on the input or output, but not on a (hidden) state variable.

The linear state-dependent diffusion imposes a lognormal distribution on the model outputs. We thus use the logarithmic transformation of the modeled and observed outputs for parameter inference. We then back-transform $h(\mathbf{S}, \mathbf{x}, \theta, \psi, t) + \mathbf{E}(\psi)$ into the real space for forecasting.

As the numerical solution of equation (6) with stochastic state-dependent diffusion can be challenging, a Lamperti transformation is commonly applied [*Kloeden and Platen*, 1992; *Iacus*, 2008; *Møller*, 2010; *Breinholt et al.*, 2011].

3.3. Inference and Generation of Model Outputs

To describe how the EBD and IND differ regarding parameter estimation and forecasting, we first discuss the approaches on a conceptual level before addressing their numerical implementation.

3.3.1. Parameter Estimation

In a probabilistic framework, the inverse problem of parameter estimation requires assumptions about the error distribution. These assumptions are formalized by a likelihood function $\mathcal{L}_M(\mathbf{y}_o | \theta, \psi, \mathbf{x})$ that describes the conditional probability density of producing the observed output data given a certain model structure M, inputs **x**, and parameters (θ, ψ). Calibration parameters of the hydrological model (θ) and of the error description (ψ) are presented in Table 2.

3.3.1.1. Parameter Estimation in the EBD Approach

In the current state of the EBD approach, we assume that the data-generating process follows a multivariate normal distribution with mean $\mathbf{y}_M$ and covariance $\boldsymbol{\Sigma}$:

$$\mathcal{L}_M(\mathbf{y}_o | \theta, \psi, \mathbf{x}) = \frac{(2\pi)^{-\frac{n}{2}}}{\sqrt{\det(\boldsymbol{\Sigma}(\psi, \mathbf{x}))}} \exp\left(-\frac{1}{2} \left[\mathbf{y}_o - \mathbf{y}_M(\theta, \mathbf{x}) \right]^{\mathsf{T}} \boldsymbol{\Sigma}(\psi, \mathbf{x})^{-1} \left[\mathbf{y}_o - \mathbf{y}_M(\theta, \mathbf{x}) \right] \right), \tag{9}$$

where n is the number of observations, i.e., the length of the vector $\mathbf{y}_o$ (e.g., a measured discharge time series at the outlet of a catchment). $\boldsymbol{\Sigma} = \boldsymbol{\Sigma}_{\mathbf{B}_M} + \boldsymbol{\Sigma}_{\mathbf{E}}$ is the total error-covariance matrix accounting for the auto-correlated and heteroskedastic bias process arising from input and structural errors and for iid observation errors.

Since equation (3) has three terms to identify given one observation vector, a Bayesian approach involving the use of prior information is necessary [*Craig et al.*, 2001; *Bayarri et al.*, 2007; *Reichert and Schuwirth*, 2012]. For statistical inference, the likelihood function is combined with the prior information on parameters to infer their posterior distribution according to Bayes' law:

$$f_{post}(\theta, \psi | \mathbf{y}_o, \mathbf{x}) = \frac{f(\theta, \psi) \mathcal{L}_M(\mathbf{y}_o | \theta, \psi, \mathbf{x})}{\int \int f(\theta, \psi) \mathcal{L}_M(\mathbf{y}_o | \theta, \psi, \mathbf{x}) \mathrm{d}\theta \mathrm{d}\psi}. \tag{10}$$

Numerically, we approximated this distribution by a Markov chain Monte Carlo (MCMC) algorithm [*Honti et al.*, 2013; *Del Giudice et al.*, 2013].

3.3.1.2. Parameter Estimation in the IND Approach

Considering the focus of the IND on online (i.e., real time) applications, computationally efficient routines for parameter inference are important. For time series data, the likelihood function is given as a product of one-step-ahead conditional densities [*Box et al.*, 2008; *Madsen*, 2008]. This approach is more efficient and easier to implement than sampling from the multivariate likelihood function when accounting for all the

Table 2. Conceptual Model and Error Model Calibration Parameters (θ, ψ)[a]

Name	Description and Alternative Name	Units	Prior (for EBD)	Prior (for IND)
Deterministic model parameters (θ)				
$ln(A_{imp})$	$\log_e$ of the impervious catchment (A)	ln (ha)	N (4.31, 0.86)	N (4.31, 0.86)
K	Mean reservoir residence time	h	TN (4.5, 0.9, 0, ∞)	TN (4.5, 0.9, 0, ∞)
$s_{1,0}$	Initial condition of reservoir 1 (*s1_ini*)	m^3	LN (675, 135)	
$s_{2,0}$	Initial condition of reservoir 2 (*s2_ini*)	m^3	LN (675, 135)	
$ln(s_{1,0})$	Initial condition of reservoir 1	ln (m^3)		N (6.5, 0.19)
$ln(s_{2,0})$	Initial condition of reservoir 2	ln (m^3)		N (6.5, 0.19)
Error model parameters (ψ)				
τ	Correlation length of B (*corrlen*)	h	LN (10, 3)	
$\sigma_{B_{ct}}$	Standard deviation of B (*sd.B_Q*)	m^3/h	TN (0, 40, 0, ∞)	
κ	Proportionality constant between input and uncertainty increase (*ks_Q*)	m^2	TN (0, 57,965, 0, ∞)	
d	Lag (in time steps) between input and uncertainty increase (*Delta*)	10 min	Exp (6)	
σ_{s_1}	Diffusion scaling for $ln(s_1)$			N (-10, 1000)
σ_{s_2}	Diffusion scaling for $ln(s_2)$			N (-10, 1000)
σ_E	Standard deviation of E (*sd.Eps_Q*)	m^3/h	LN (20, 2)	
$ln(\sigma_E)$	Standard deviation of E	ln (m^3/h)		N (-2.55, 0.255)

[a]The notation for prior distributions is LN(μ, σ): lognormal, N(μ, σ): normal, TN(μ, σ, a_1, a_2): truncated normal, and Exp(λ^{-1}): exponential. The symbols are μ: expected value, σ: standard deviation, a_1: lower limit, a_2: upper limit, and λ: rate.

observations at a time. This likelihood would be a path integral, i.e., an infinite-dimensional integral over all possible realizations of the model states [e.g., *Restrepo*, 2008; *Balaji*, 2009; *Quinn and Abarbanel*, 2010]. We define

$$
\mathcal{L}_M(\mathbf{y}_o|\boldsymbol{\theta}, \boldsymbol{\psi}, \mathbf{x}) = \left(\prod_{i=2}^{n} p(\mathbf{y}_{o_i}|\mathbf{y}_{o_{i-1}}, \boldsymbol{\theta}, \boldsymbol{\psi}, \mathbf{x}) \right) p(\mathbf{y}_{o_1}|\boldsymbol{\theta}, \boldsymbol{\psi}, \mathbf{x}) = \frac{(2\pi)^{-\frac{n}{2}}}{\sqrt{\det\left(\Sigma(\mathbf{y}_{o_i}|\mathbf{y}_{o_{i-1}}, \boldsymbol{\theta}, \boldsymbol{\psi}, \mathbf{x})\right)}} \cdot
$$

$$
\exp\left(\sum_{i=2}^{n} \left(-\frac{1}{2} \left[\mathbf{y}_{o_i} - \mathrm{E}(\mathbf{y}_{o_i}|\mathbf{y}_{o_{i-1}}, \boldsymbol{\theta}, \boldsymbol{\psi}, \mathbf{x})\right]^{\mathsf{T}} \Sigma(\mathbf{y}_{o_i}|\mathbf{y}_{o_{i-1}}, \boldsymbol{\theta}, \boldsymbol{\psi}, \mathbf{x})^{-1} \left[\mathbf{y}_{o_i} - \mathrm{E}(\mathbf{y}_{o_i}|\mathbf{y}_{o_{i-1}}, \boldsymbol{\theta}, \boldsymbol{\psi}, \mathbf{x})\right] \right) \right) \tag{11}
$$

$$
\cdot p(\mathbf{y}_{o_1}|\boldsymbol{\theta}, \boldsymbol{\psi}, \mathbf{x}),
$$

where $\mathrm{E}(\mathbf{y}_{o_i}|\mathbf{y}_{o_{i-1}}, \boldsymbol{\theta}, \boldsymbol{\psi})$ is the mean and $\Sigma(\mathbf{y}_{o_i}|\mathbf{y}_{o_{i-1}}, \boldsymbol{\theta}, \boldsymbol{\psi})$ the covariance of the one-step-ahead predictions generated using an extended Kalman filter. This product of conditional densities assumes independence and normality of the one-step-ahead forecast errors ("innovations") at each time step given the observations up to time $i - 1$. These innovations are the results of input and structural errors. It is implicitly assumed that the transformed states given all observations up to $i - 1$ are also normally distributed [*Law and Stuart*, 2012] and that they follow a Markov process [*Bulygina and Gupta*, 2009; *Moradkhani et al.*, 2012]. To gain insight into whether the conditional densities of the states can be considered Gaussian, we can analyze the empirical distribution of the one-step-ahead errors.

In the IND, inference is usually performed on a frequentist basis [*Breinholt et al.*, 2012], but a Bayesian framework has also been adopted [*Melgaard*, 1994; *Sadegh et al.*, 1994]. For comparability with the EBD, we will use a Bayesian calibration and therefore also make use of equation (10). Traditionally, in the IND, Bayesian estimation has consisted in maximizing the posterior $f(\boldsymbol{\theta}, \boldsymbol{\psi}|\mathbf{y}_o, \mathbf{x})$ rather than characterizing its full distribution [*Melgaard*, 1994; *Sadegh et al.*, 1994; *Walter and Pronzato*, 1997]. Numerically, the so-called maximum a posteriori (MAP) estimation is here performed with an extended Kalman filter (EKF) [*Law and Stuart*, 2012]. The EKF provides a consistent first-order approximation to the estimate of a nonlinear model at the observation time, as well as the errors of this estimate [*Kao et al.*, 2004]. Details on the EKF equations can be found in Appendix A. *Quinn and Abarbanel* [2010], *Balaji* [2009], and *Law and Stuart* [2012] provide further discussions on the assumptions behind approximate Gaussian filters (as the EKF).

3.3.2. Smoothing

It can be useful to predict system output and/or states for points in time where flow data have been employed for parameter inference, in the so-called "calibration period" (or calibration layout). This retrospective analysis, called smoothing, consists in identifying system states (or output) from all available (noisy) output data [*Einicke*, 2012; *Bulygina and Gupta*, 2009].

3.3.2.1. Smoothing With EBD

Here we condition the Gaussian bias process on the observations and updated parameters, and propagate the parametric uncertainty of the simulator and the error models via Monte Carlo simulations [*Reichert and Schuwirth*, 2012; *Del Giudice et al.*, 2013]. To predict the observed system response in the calibration layout, we approximate the distributions of $\mathbf{y}_M + \mathbf{B}_M + \mathbf{E}$ for every temporal point i of the data set, i.e., for $i = 1, \ldots, n$.

3.3.2.2. Smoothing With IND

Commonly, the IND is applied in combination with extended Kalman filtering to update the model states considering one data point at a time [*Kristensen and Madsen*, 2003]. For comparability with the EBD, we here generate smoothened estimates of the model states and outputs in the calibration period. In this setting, conditioning on data can be performed by combining a filter moving forward in time with one going backward (i.e., from the future to the present) [*Einicke*, 2012]. The smoothed model states are assumed to be normally distributed and related to the output through equation (7).

3.3.3. Forecast of Future Output
3.3.3.1. Forecast With EBD

The posterior predictive distribution of runoff in the extrapolation layout (also called validation period) is computed via Monte Carlo simulations. To approximate the distribution of $\mathbf{y}_M + \mathbf{B}_M + \mathbf{E}$, we first obtain realizations of $\mathbf{y}_M$ by propagating a sample of θ_{post} through f_M. Second, we compute trajectories of $\mathbf{B}(\psi_{post})$ and $\mathbf{E}(\psi_{post})$ and add them to the results of the simulator [*Reichert and Schuwirth*, 2012; *Del Giudice et al.*, 2013]. In this procedure, the bias-corrected model is not conditioned on data and therefore its predictive uncertainty becomes larger than in the calibration period. However, as the autocorrelated bias has a "memory," observed output still influences these predictions if the analyzed time is close to the last calibration point. An explanation on how to produce EDB forecasts in this (initial) extrapolation phase is given in Appendix C.

3.3.3.2. Forecast With IND

Unconditional output can be generated from stochastic gray-box models by performing "scenario (or ensemble) simulations" [*Platen and Bruti-Liberati*, 2010] from equation (6). To compute trajectories from the stochastic differential equations describing the state space model, we use discrete-time approximations. For each solution of equation (6), the predictions for $\mathbf{Y}_o$ are derived by inserting the simulated paths of the states into equation (7). In this setting, normality is assumed for the model states at the forecast starting point j, conditional on the previous time steps observations $\mathbf{Y}_{o,j-1}$.

3.4. Design of Computer Experiments

To compare the performances of the two approaches, we performed three numerical experiments. First, we analyzed the parameter estimates we obtained after calibration. Second, we compared the quality of long-term predictions over 14 days (5328 time steps) and, third, short-term forecasts over 200 min (20 time steps). Although this is longer than the usual one to five time steps of online applications, we selected this forecasting horizon for illustrative purposes. Since future rainfall was assumed known, both types of predictions were, strictly speaking, ex-post hindcasts.

3.5. Performance Metrics

To evaluate the performances of the EBD and IND, we used four performance metrics, together with a visual inspection of model predictions and quantile-quantile plots. To assess the quality of the underlying deterministic model, we considered the median of the probabilistic simulations. We used (i) the Nash-Sutcliffe efficiency index (NS, optimally approaching 1 from below) and (ii) the normalized (or relative) bias (NB, optimally approaching 0). Both statistics are commonly used in hydrology to assess the accuracy in fitting the peaks of the hydrographs and preserve water balance, respectively [*Bennett et al.*, 2013; *Bulygina and Gupta*, 2009; *Coutu et al.*, 2012].

To assess the quality of ex-post forecasts, we focused on 95% prediction intervals, while also analyzing the other quantiles via QQ plots (supporting information). Specifically, we evaluated the (iii) "coverage," which measures the percentage of validation measurements falling into the 95% prediction intervals, and (iv) the interval (skill) score ($S_{0.05}^{int}$, optimally approaching 0 from above), which provides a simultaneous assessment of the precision and reliability of the prediction intervals [*Gneiting and Raftery*, 2007]:

$$S_\alpha^{int} = (u - l) + \frac{2}{\alpha}(l - y_{o_j})H\{l - y_{o_j}\} + \frac{2}{\alpha}(y_{o_j} - u)H\{y_{o_j} - u\}, \tag{12}$$
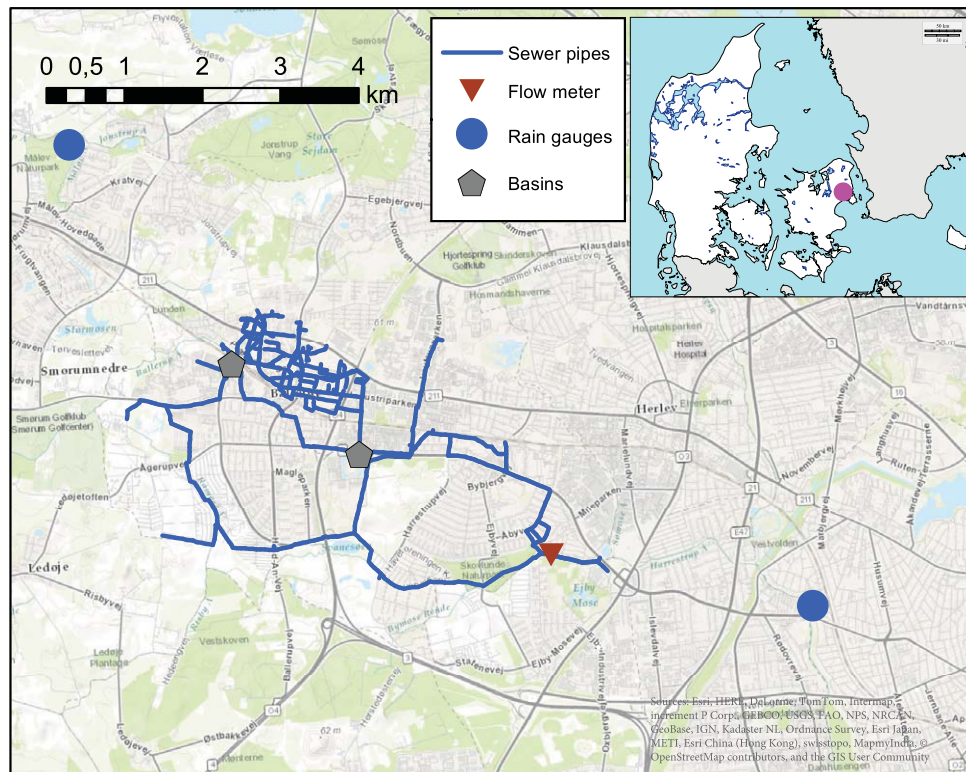
Figure 2. The studied Ballerup sewer network with the rain gauges used for deriving the model input and the flow meter used for measuring the system output.

where $\alpha = 0.05$ corresponds to the confidence level, u and l to the 97.5 and 2.5 quantiles of the predictive distribution of $\mathbf{Y}_o$ at the time point j, and y_{o_j} to the data in the extrapolation layout. H denotes the unit step function, which takes the value of 1 if its argument is greater than 0 and 0 otherwise. We averaged $S_{0.05}^{int}$ over all time steps considered.

4. Hydrological Application

In the following, we describe the analyzed watershed, the deterministic model used, the available hydrological measurements, the chosen priors, and the computer implementation of our study.

4.1. Case Study

For our application, we chose a sewer system located in the Ballerup area close to Copenhagen (Denmark) (Figure 2). The catchment has a total surface area of approximately 1300 ha and is mainly laid out as a separate system, although it does have a small combined section. The runoff in this area is strongly influenced by rainfall-dependent infiltration, and the catchment contains several basins and pumping stations. Several previous modeling studies were undertaken using this catchment [*Breinholt et al.*, 2011, 2012; *Löwe et al.*, 2014]. Tipping bucket rain gauge measurements were available from the Danish Water Pollution Committee's (SVK) network [*Jørgensen et al.*, 1998]. One minute observations from the two pluviometers located near the catchment were averaged and used as input for the runoff model. Flow measurements were available with a temporal resolution of 5 min. The time of concentration of the catchment is approximately 60 min. As *Schilling* [1991] recommends a temporal resolution of rainfall measurements of at least 0.2–0.33 times the concentration time of an urban watershed, we adopted a modeling time step of 10 min and averaged flow and pluviometric data to this time discretization.

4.2. A Parsimonious Hydrological Model

The sewer flow at the monitoring point, $\mathbf{y}_M$, is modeled as a superposition of wastewater flow and rainfall-runoff. While the storm water runoff (equations (13) and (14)) is described by a cascade of two virtual

reservoirs, the wastewater hydrograph (equation (15)) is represented as a superposition of four harmonic functions (Figure 3). The model dynamics is defined by the following deterministic equations:

$$f_M(\mathbf{s}, \mathbf{x}, t, \boldsymbol{\theta})dt = d\begin{bmatrix} s_1(t) \\ s_2(t) \end{bmatrix} = \begin{bmatrix} A_{imp} \cdot x(t) + a_0 - \dfrac{1}{k}s_1(t) \\ \dfrac{1}{k}s_1(t) - \dfrac{1}{k}s_2(t) \end{bmatrix} dt,$$ (13)

with output

$$y_M(\mathbf{x}, t, \boldsymbol{\theta}) = \frac{1}{k}s_2(t) + w_{dw}(t),$$ (14)

where $w_{dw}(t)$ describes the diurnal variation of dry-weather wastewater flow

$$w_{dw}(t, \boldsymbol{\theta}) = \sum_{i}^{2}\left(\varsigma_i sin\frac{i2\pi t}{24} + \chi_i cos\frac{i2\pi t}{24}\right).$$ (15)

s_1 and s_2 correspond to the states of the system, i.e., the levels in the virtual storage tanks, and vary as a function of time (in hours). The vector $\boldsymbol{\theta}$ of physical model parameters includes the impervious catchment area A_{imp}, the mean dry-weather flow at the catchment outlet a_0, the mean travel time (or reservoir residence time) k, and parameters ς_1, ς_2, χ_1, and χ_2. These last four variables describe the dry-weather variation of the catchment outflow as a harmonic function. The vector $\mathbf{x}$ of model inputs includes the rainfall measurements averaged from the two pluviometers.

This simplified model disregards infiltration and does not include losses from sewer overflows. However, as a so-called "gray-box model," it captures the major processes with components that have a physical meaning. As such, its major advantage is that its equations are suitable to be incorporated into the IND framework (Appendix B) and it is computationally fast enough to be applied in a forecast setting with data assimilation [*Breinholt et al.*, 2012; *Löwe et al.*, 2014]. Simple models have often proven useful and sufficient in offline and online applications [*Coutu et al.*, 2012; *Wolfs et al.*, 2013; *Mejía et al.*, 2014] and when modeling the integrated urban drainage system [*Freni et al.*, 2009].

4.3. Prior Knowledge of Model Parameters

We selected prior distributions for the EBD based on the experience gained during previous studies in the same and similar catchments [*Breinholt et al.*, 2012; *Löwe et al.*, 2014]. Prior knowledge on simulator parameters was described by lognormal or normal distributions with a coefficient of variation of 0.2.

For the bias, we defined a probability density decreasing with increasing values of $\sigma_{B_{ct}}$ and κ (here a truncated normal distribution) [*Reichert and Schuwirth*, 2012; *Del Giudice et al.*, 2013]. This helps to reduce the identifiability problem between the deterministic model and the bias term and avoids model bias as much as possible. Regarding the correlation time of the bias, τ, we chose a prior value of 10 h, close to 1/3 of the recession time of a consequential flood event not used for calibration.

For the IND approach, all parameters, except k, are defined in a logarithmic space to avoid
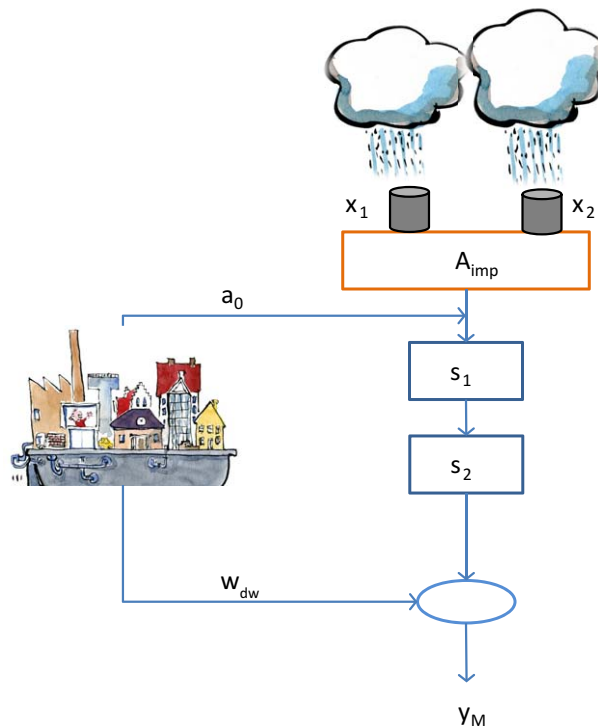

Figure 3. The linear reservoir cascade model considered for hydrological modeling. (left) The wastewater generation and (right) the rainfall-runoff process are illustrated. Symbols' description is given in section 4.2. Drawings by F. Ahlefeldt.

negative values for the parameters. With respect to the standard deviation of the observation error σ_ϵ, we specified a prior as consistent as possible with the one of the bias description. Regarding the initial model states, we analytically calculated the filling of the reservoirs for no rain condition. The results obtained were similar to the system states in dry weather calculated in previous studies [*Breinholt et al.*, 2011].

The parameters of the dry-weather-flow compartment were not inferred simultaneously with the other parameters due to numerical difficulties encountered in the IND routine. Instead, we independently estimated them with a least squares method. For that, we selected data (not shown) from a period with no rain ranging from 18 July 2010 until 28 July 2010. The resulting dry-weather parameters were $a_0 = 281.5 \frac{m^3}{h}$, $\chi_1 = -47.4$, $\chi_2 = 21.3$, $s_1 = -43.4$, and $s_2 = -84.2$. The prior distributions of simulator and error model parameters are summarized in Table 2.

4.4. Computer Implementation
The conceptual hydrological model and the EBD routine for uncertainty analysis were implemented in `R` [*R Core Team*, 2014]. During inference (equation (10)), we first obtained an optimal jump distribution and chain starting point by sequentially using the stochastic techniques described by *Haario et al.* [2001] and *Vihola* [2012], and then sampled from the target distribution by using a Metropolis-Hastings algorithm [*Hastings*, 1970]. Finally, we approximated the predictive distribution of $\mathbf{Y}_o$ by propagating a posterior parameter sample through the simulator and the error model.

The IND routine was implemented in the open source software `CTSM` [*Juhl et al.*, 2013], which is available as a package for `R`. Posterior maximization was performed using the PORT algorithm through the `R` function `nlminb` [*Gay*, 1990]. To generate forecasts with the SDEs, we applied an Euler-Maruyama scheme [see e.g., *Kloeden and Platen*, 1992; *Iacus*, 2008], which involved 5000 realizations of the process **S**.

5. Results
Predicting sewage flow with the EBD and IND approaches, we found that (i) both methodologies provided forecast coverage of the validation data close to the nominal 95% and (ii) reproducing the observations during heavy storm events (where the model has high discrepancies from data) was challenging for both methods. Even so, the uncertainty estimates of the two approaches dramatically outperformed those of a simplified approach using an iid error model (see supporting information).

5.1. Experiment 1: Parameter Estimation
The data used for inference include two separate periods, as presented in supporting information Figure S1. The parameters inferred for the different modeling approaches are shown in Figure 4. The calibration with the IND was approximately 2 orders of magnitude faster than with the EBD. In the EBD, the inference produced approximately bell-shaped marginals. The only distribution with a complex shape is that of d, which represents the time steps after which the rainfall influences runoff uncertainty. The posterior initial model states s_1 and s_2 remained close to their prior estimates and were similar for the EBD and IND. For the effective area A_{imp}, we observed bigger values of approximately 39 ha for the EBD approach, while the IND estimated an optimum of 33 ha. For the time constant k, approximately the same value was obtained with both frameworks (2.5 h). In both approaches, the inferred observation noise was considerably smaller than the bias or diffusion term (Figure 4). Due to the different ways of considering errors in the two methods, the other stochastic process parameters ψ cannot be compared directly.

5.2. Experiment 2: Long-Term Forecasting
Long-term predictions for the two approaches were similar in terms of interquantile width and reliability (Figure 5). Credible intervals for IND predictions, however, were slightly wider than those for the EBD and therefore covered the validation data better. Higher data coverage also resulted in a $\approx$50% better average interval score $S_{0.05}^{int}$ than for the EBD. The median of the probabilistic predictions was closer to the observations for the EBD than for the IND approach. The model calibrated with the EBD fitted validation peak discharge data better and obtained a better NS than the IND (32% higher). In general, with both error descriptions, the model consistently underestimated wet weather flows. This underprediction is confirmed by the QQ plot analysis (supporting information Figure S2). Here the EBD-calibrated simulator performed slightly better than the IND. The latter had a NB $\approx$40% larger and quantiles more distant from the 1:1 line.
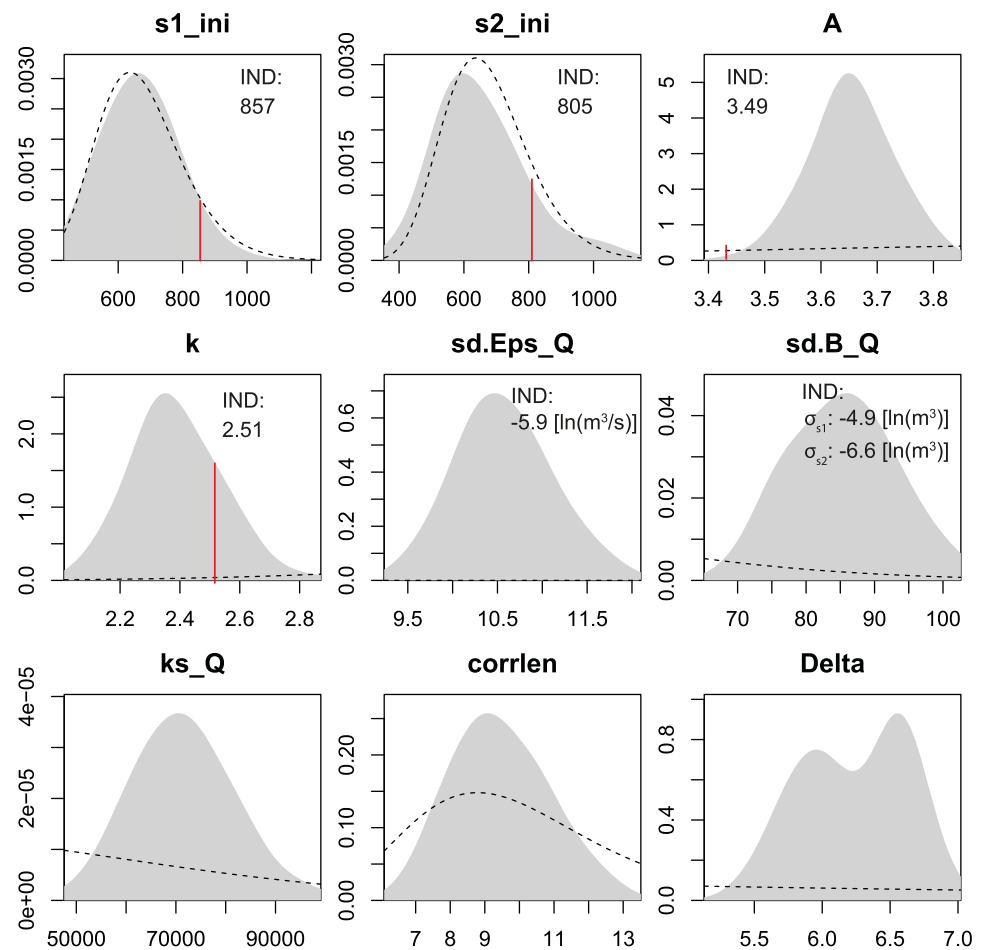
Figure 4. Prior (black, dashed) and posterior (gray area) marginal distributions from Bayesian inference in the EBD framework. The corresponding maximum a posteriori estimates from the IND framework are also displayed. Meaning and units of the parameters are given in Table 2.

As expected, the EDB and IND outperformed the forecasts where model bias was neglected, both in terms of data coverage (i.e., reliability) and interval scores (supporting information Figure S4).

5.3. Experiment 3: Short-Term Forecasting

As shown in Figure 6, the percentage of data points covered by the 95% credible interval of the short-term predictions was close to the nominal coverage. This means that the predictions were approximately reliable, although the underlying simulator appears to systematically deviate from reality. This is particularly interesting during the flood event on the right side of Figure 6, where the underlying model heavily underestimated the receding section of the hydrograph, yet the probabilistic predictions, after data assimilation, still encompassed most of the validation data. Indeed, with the simplified analysis that uses an iid error model, we obtained much poorer prediction intervals than with the two proposed methodologies (supporting information Figure S5).

During storm events, interval scores $S_{0.05}^{int}$, which penalize too wide and unreliable uncertainty bands, were moderately higher (i.e., worse) for the EBD, especially in the decreasing limb of the flood hydrograph. Visual inspection shows that this is related to the slightly overconfident predictions of the EBD in this period. In contrast, during dry weather, the EBD and IND produced similar predictions.

6. Discussion

6.1. Prediction Analysis

As shown in the case study application, both methodologies were able to provide both short-term and long-term reliable predictions. This is remarkable for two reasons. First, the underlying lumped reservoir
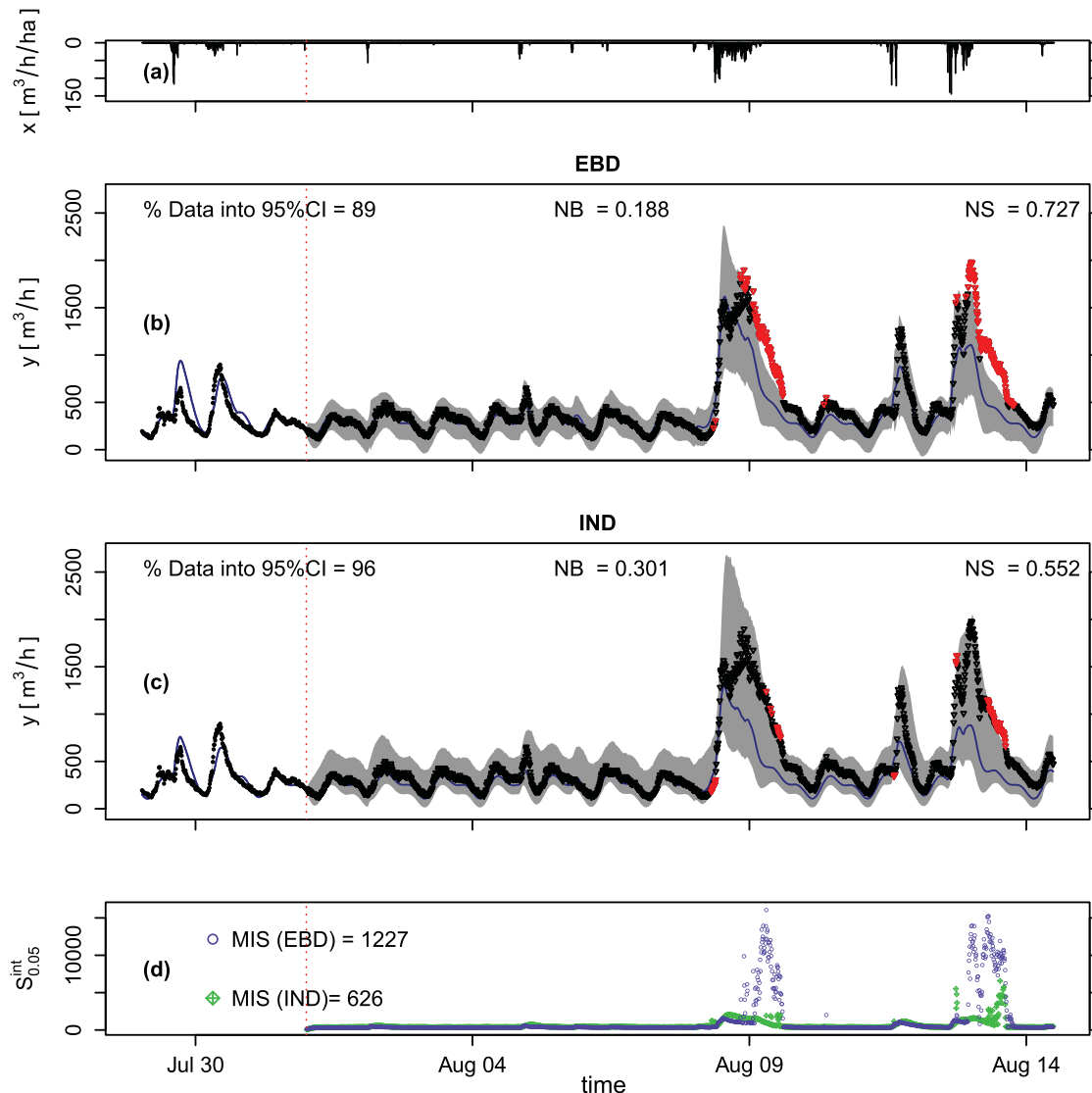
Figure 5. (left) Smoothing and (right) long-term forecasting results with the two methods. (a) Average rain intensity over the catchment (input data); (b) 95% credible intervals (gray) using the EBD approach, output data (dots, red when outside the intervals), median of the deterministic model (blue line); (c) 95% credible interval using the IND approach; (d) interval skill scores $S_{0.05}^{int}$ for the validation period together with its mean value MIS. The Nash-Sutcliffe coefficient (NS), the Normalized Bias index (NB), and the other performance indicators refer to the extrapolation period on the right of the dotted line.

model was a simplified representation of reality and therefore unable to consider all mechanisms occurring in the catchment (e.g., spatially varying soil water content, infiltration). Second, the validation conditions were consistently different from the calibration circumstances (more substantial peak discharges and infiltration-inflow). These considerations suggest that the methods are relatively robust against nonstationary inputs and boundary conditions, and structural errors of the model. Furthermore, both the EBD and IND could account for increased uncertainty during more dynamic wet periods, the first thanks to the input dependence of the bias and the second due to the state dependency of the noise. This is consistent with the conclusions of previous studies [*Breinholt et al.*, 2012; *Dietzel and Reichert*, 2012; *Honti et al.*, 2013; *Del Giudice et al.*, 2015]. Furthermore, for both methods, conditioning on data generated generally reliable and precise short-term forecasts in all flow conditions, even when the calibrated simulator heavily deviated from the measurements (Figure 6).

Large deviations between model predictions and observations on long forecast horizons are mostly caused by the very simple model structure and system nonstationarities but are also influenced by the error description. As discussed in *Bayarri et al.* [2007] and *Del Giudice et al.* [2013], the bias description might produce
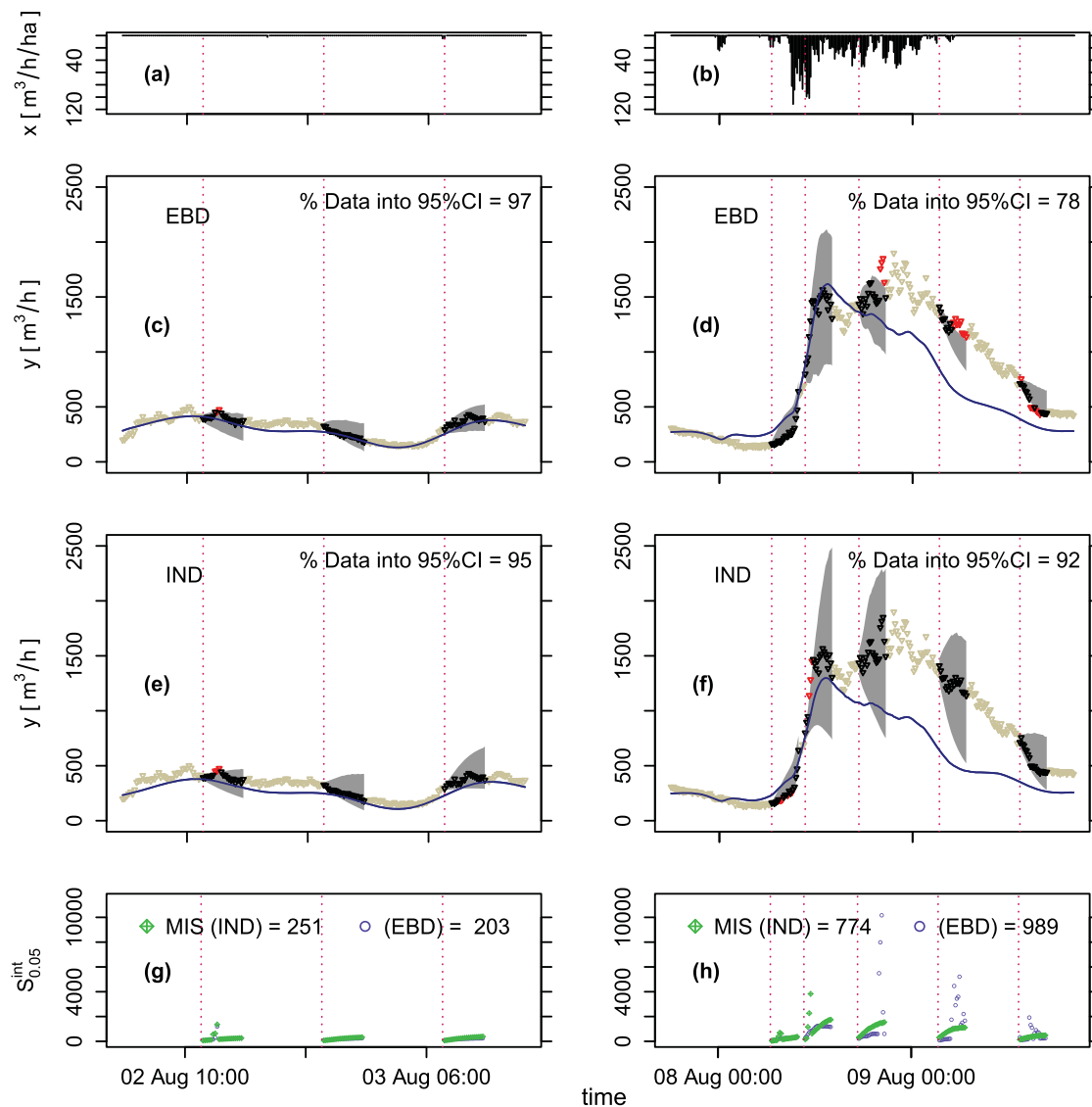
Figure 6. Short-term forecasts for illustrative points during (a, c, e, and g) a dry and (b, d, f, and h) a wet period. (Figures 6a and 6b) Rain intensity (input data); (Figures 6c and 6d) 20 step flow forecasts for the EBD approach (95% credible intervals) (gray), output data (dots), median for the previously calibrated deterministic model (blue line); (Figures 6e and 6f) 20 step flow forecasts for the IND approach; (Figures 6g and 6h) interval skill scores $S_{0.05}^{int}$ for the different forecast horizons together with their mean value MIS.

model performances which are slightly inferior to simplified approaches based on an iid error assumption. This can be explained by the fact that the inference with the EBD does not force the simulator to reproduce the observations with biased (i.e., over-tuned) parameters. The reverse, however, can also be true, and in this experiment the model fitted the data better with the bias than without it. Reduced model fit can be even more pronounced in the IND where parameter inference is performed in a one-step-ahead prediction setting. *Breinholt et al.* [2012] demonstrated a very satisfactory forecast performance of the approach on short horizons, which diminishes on longer horizons until becoming inferior to simplified approaches. In the present study, we also observe the highest forecast accuracy on the shortest horizons (Figure 6).

Parameter estimation in the IND relies on the assumption of normality and independence of the one-step-ahead prediction errors (innovations) and of Gaussianity of the transformed system states. By inspecting the innovations (supporting information Figures S13–S15), this assumption appears to be valid in our study.

In agreement with previous studies [*Honti et al.*, 2013; *Del Giudice et al.*, 2013; *Breinholt et al.*, 2012], we generally found that both the EBD and the IND (Figures 5 and 6) produced much less overconfident and therefore more reliable uncertainty bands than simplified approaches (supporting information Figures S4 and S5).

6.2. Commonalities and Differences of the Methods
6.2.1. Theoretical Considerations

The main difference between the two approaches considered is that the IND describes model inadequacies as part of the model states, while the EBD adds them to the model output. In other words, the IND propagates the input and structural errors identified during calibration through the model, while the EBD treats the model as a "perfect" black box to which these errors are added. In addition, the EBD was developed with a focus on statistical inference and long-term prediction, while online applications were the focus for the IND.

Flexible model structures for describing the time-dependent behavior of systematic errors can be implemented in both approaches. Input and output dependence of systematic errors can be considered in the EBD [*Del Giudice et al.*, 2013]. In the IND, state-dependent and input-dependent diffusion terms can be implemented, but only the former were documented in previous applications [*Breinholt et al.*, 2012; *Löwe et al.*, 2014], while the latter are the subject of ongoing research.

6.2.2. Practical Aspects

The most suitable error characterization needs to be identified depending on the specific case study. Adding linear state-dependent noise in the model equations, as in the IND, has the advantage that it guarantees positive values of the model output. When modeling the errors in the output equations, as in the EBD, output transformation might be required to ensure nonnegative predictions [e.g., in *Frey et al.*, 2011; *Sikorska et al.*, 2012].

On the one hand, the implementation of the noise term as part of the model states in the IND seems intuitively more appropriate, because the systematic error description becomes a part of the model and the noise is routed through the model. In combination with data assimilation routines, the IND also allows for the identification of hidden states from data, which is a useful feature in process monitoring and system control, for example. On the other hand, the solution of stochastic differential equations is more complex than that of ordinary differential equations and this limits how complex the model can be.

The IND, being an "intrusive" method, cannot easily be applied to existing hydrological software packages such as SWMM (Rossman and Supply, 2010). Instead, this is easily done with the "nonintrusive" EBD, on condition that the model is fast enough to be applied in MCMC.

Parameter inference in the two approaches is largely driven by their focus areas, and that applies to both conceptual formulation and the numerical techniques. The EBD applies a Bayesian approach using MCMC which is slow but allows for the identification of the whole distribution of the parameters. The IND commonly applies Maximum a Posteriori (or Likelihood) estimation for parameter inference. Currently, only the mode of the parameter distribution is considered and parametric uncertainty is neglected during forecasting. This approach is computationally very efficient and identifies model parameters which are optimal for online predictions.

An updating of the model states is readily implemented in the IND framework, but it leads to a violation of the water balance [see e.g., *Salamon and Feyen*, 2010; *Reichert and Mieleitner*, 2009]. It is therefore not particularly suitable for design studies, while it can be very useful in online applications where only the correspondence between forecasted and observed output is of interest.

7. Conclusions

In this study we, for the first time, compared and discussed two probabilistic techniques to reliably quantify predictive uncertainty in rainfall-runoff modeling in urban catchments. The first approach was an external bias description (EBD), representing model discrepancies in the output space. The second was an internal noise description (IND), considering model inadequacies in the system equations. Based on theoretical considerations and the results of the case study, we conclude that

1. Both approaches describe systematic model errors in a way suitable for hydrological modeling. Both can produce reliable forecasts in the short term, which is useful, e.g., for real-time model predictive control of sewer networks and wastewater treatment plants, as well as for long-term analyses. As demonstrated in our case study, this seems to be the case even for very simple rainfall-runoff models applied to a complex sewer system with nonstationary behavior.

2. Both methods also have some limitations. First, although they explicitly account for the effects of model inadequacies, neither of them provides comprehensive information on underlying causes of bias. The IND, through an analysis of model states, can, however, give some hints on which model compartment is most uncertain. The EBD can be rather demanding on a computational level during parameter inference because it requires tens of thousands of MCMC simulations. Furthermore, it does not provide a data assimilation routine in its current implementation. In contrast to the IND, the EBD can readily be applied to any existing engineering software. Additionally, in its current implementation, the IND makes simplifying assumptions on the distribution of the states and outputs. These guarantee a very high computational efficiency but need to be tested via residual analysis.

3. Although both techniques generally outperform those that do not account for systematic model errors, especially in quantifying predictive uncertainties, each has its optimal field of application. The EBD is usually able to provide accurate and precise long-term forecasts with various kinds of models, provided that the model reasonably describes the system studied. The IND, on the other hand, is especially suitable for short-term forecasts where new output measurements are continuously available for updating. Additionally, it appears able to provide reliable predictions even in cases where the underlying model is highly simplified. Finally, it allows for the identification of hidden model states, which is useful to identify the behavior of a variable when only indirect measurements are available.

4. Expected developments of the EBD involve the investigation of the reasons for bias. Current research in the IND is focusing on reducing the likelihood approximations and producing an ensemble-based version that would make it applicable to existing models.

Appendix A: Equations for State Updating With the IND Using the EKF

Posterior maximization with the IND likelihood (equation (11)) adopts an extended Kalman filter (EKF). The filtering procedure is briefly synthesized from *Kristensen et al.* [2004] and *Kao et al.* [2004]. For each candidate parameter set (θ_c, ψ_c) generated during optimization, the innovations $y_{o_i} - E(y_{o_i}|y_{o_{i-1}}, \theta_c, \psi_c)$ and their covariances $\Sigma(y_{o_i}|y_{o_{i-1}}, \theta_c, \psi_c)$ are continuously updated following this assimilation scheme:

Step i: Project the state ahead for the next time step i solving the state prediction equation representing the deterministic model:

$$\frac{d\mathbf{S}}{dt} = f_M(\mathbf{S}, \mathbf{x}, t, \theta), \tag{A1}$$

for time interval $[t_{i-1}, t_i[$. The so-obtained state $\mathbf{S}_{i|i-1}$ is used to predict the (a priori) output at time i:

$$E(y_{o_i}|y_{o_{i-1}}) = h(\mathbf{S}_{i|i-1}, \mathbf{x}_i, \theta_c, \psi_c, t_i). \tag{A2}$$

Step ii: Project the (a priori) error-covariance matrix ahead:

$$\frac{d\mathbf{P}}{dt} = \mathbf{MP} + \mathbf{PM}^T + \Sigma_\sigma, \tag{A3}$$

where the resulting covariance matrix is defined as $\mathbf{P}_{i|i-1} \equiv E[(\mathbf{S}_{i|i-1} - \mathbf{S}_*)(\mathbf{S}_{i|i-1} - \mathbf{S}_{*,i})^T]$ with $\mathbf{S}_{*,i}$ representing the true state. In equation (A3), $\mathbf{M}$ is the Jacobian matrix of the deterministic model f_M, and Σ_σ is the estimated system noise covariance for the prediction of $\mathbf{P}$.

Step iii: When the next output measurement y_{o_i} becomes available (or assimilable) the states are updated (or corrected):

$$\mathbf{S}_{i|i} = \mathbf{S}_{i|i-1} + \mathbf{K}_i(y_{o_i} - E(y_{o_i}|y_{o_{i-1}})), \tag{A4}$$

where $\mathbf{K}_i$ is the Kalman gain defined as $\mathbf{K}_i \equiv \mathbf{P}_{i|i-1}\mathbf{H}^T\Sigma^{-1}(y_{o_i}|y_{o_{i-1}})$, with $\mathbf{H}$ being the Jacobian matrix of the stochastic model h, and $\Sigma(y_{o_i}|y_{o_{i-1}}) \equiv \mathbf{HP}_{i|i-1}\mathbf{H}^T + \Sigma_E$ being the innovation covariance matrix.

Step iv: Finally, the updated (a posteriori) error-covariance matrix is computed as

$$\mathbf{P}_{i|i} = \mathbf{P}_{i|i-1} - \mathbf{K}_i\Sigma(y_{o_i}|y_{o_{i-1}})\mathbf{K}_i^T. \tag{A5}$$

This procedure of sequential state update is repeated for every time step i of the calibration period.

Appendix B: Specific Model Equations With the IND

Combining the simulator equations (equations (13–15)) with the state noise (equation (8)) and the Lamperti transformation, we obtain the following state space description of the system studied:

$$
d\begin{bmatrix} ln(s_1(t)) \\ ln(s_2(t)) \end{bmatrix} = \begin{bmatrix} \left(exp(ln(A_{imp})) \cdot x(t) + a_0\right) \cdot exp(-ln(s_1(t))) - \frac{1}{k} - \frac{1}{2}\sigma_{s_1}^2 \\ \left(\frac{1}{k}exp(ln(s_1(t)))\right) \cdot exp(-ln(s_2(t))) - \frac{1}{k} - \frac{1}{2}\sigma_{s_1}^2 \end{bmatrix} dt + \begin{bmatrix} \sigma_{s_1} & 0 \\ 0 & \sigma_{s_2} \end{bmatrix} dW_t, \quad (B1)
$$

$$
ln(Y_0) = ln(\frac{1}{k}s_2(t) + df(t)) + E. \quad (B2)
$$

Appendix C: Short-Term Forecasts With the EBD

In its current implementation, the online predictions with the bias correction are calculated by following these steps:

Step i: Select the current time point j (e.g., the last element of a data time series) and its corresponding output observation y_{o_j}.

Step ii: Condition the Gauss-Markov process on y_{o_j}. This involves computing the mean and variance of the bias according to *Reichert and Schuwirth* [2012, equations (27) and (28)], which in turn requires calculating $\Sigma_E(\psi_{post})$ and $\Sigma_{B_M}(\psi_{post})$ according to *Del Giudice et al.* [2013, equations (3) and (10)].

Step iii: Draw $\sim 10^3$ samples of the bias process in this past period.

Step iv: Use each last element (i.e., the one at time j) of the bias sample as starting point for simulating trajectories of B_M over the desired number of time steps in the future. These realizations are based on *Del Giudice et al.* [2013, equations (21) and (22)].

Step iv: As in equation (3), add to the bias realizations sample paths of the white noise (see equation (5)) and an equal number of runs of the model $y_M(\Theta_{post})$.

Step vi: Finally, produce the desired sample quantiles $y_M + B_M + E$ to plot the total uncertainty bands, usually corresponding to the region between the 95% credible intervals.

Step vii: Repeat for each time j of interest.

References

Balaji, B. (2009), Continuous-discrete path integral filtering, *Entropy*, *11*, 402–430, doi:10.3390/e110300402.

Bates, B., and E. Campbell (2001), A Markov chain Monte Carlo scheme for parameter estimation and inference in conceptual rainfall-runoff modeling, *Water Resour. Res.*, *37*, 937–947.

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.

Bechmann, H., H. Madsen, N. K. Poulsen, and M. K. Nielsen (2000), Grey box modeling of first flush and incoming wastewater at a wastewater treatment plant, *Environmetrics*, *11*(1), 1–12.

Beck, M. B., and P. Young (1976), Systematic identification of DO-BOD model structure, *J. Environ. Eng. Div. Am. Soc. Civ. Eng.*, *103*(5), 902–927.

Bennett, N. D., et al. (2013), Characterising performance of environmental models, *Environ. Modell. Software*, *40*, 1–20, doi:10.1016/j.envsoft.2012.09.011.

Beven, K., and P. Young (2013), A guide to good practice in modeling semantics for authors and referees, *Water Resour. Res.*, *49*, 5092–5098, doi:10.1002/wrcr.20393.

Beven, K. J. (1993), Prophecy, reality and uncertainty in distributed hydrological modelling, *Adv. Water Resour.*, *16*(1), 41–51.

Borup, M., M. Grum, and P. S. Mikkelsen (2013), Comparing the impact of time displaced and biased precipitation estimates for online updated urban runoff models, *Water Sci. Technol.*, *68*(1), 109–116, doi:10.2166/wst.2013.221.

Box, G. E. P., G. M. Jenkins, and G. C. Reinsel (2008), *Time Series Analysis: Forecasting and Control*, 4th edition, 748 pp., John Wiley & Sons, Hoboken, N. J.

Breinholt, A., F. O. Thordarson, J. K. Møller, M. Grum, P. S. Mikkelsen, and H. Madsen (2011), Grey-box modelling of flow in sewer systems with state-dependent diffusion, *Environmetrics*, *22*(8), 946–961, doi:10.1002/env.1135.

Breinholt, A., J. Møller, H. Madsen, and P. Mikkelsen (2012), A formal statistical approach to representing uncertainty in rainfall-runoff modelling with focus on residual analysis and probabilistic output evaluation-distinguishing simulation and prediction, *J. Hydrol.*, *472–473*, 36–52, doi:10.1016/j.jhydrol.2012.09.014, 2012.

Bulygina, N., and H. Gupta (2009), Estimating the uncertain mathematical structure of a water balance model via Bayesian data assimilation, *Water Resour. Res.*, *45*, W00B13, doi:10.1029/2007WR006749.

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.

Craig, P., M. Goldstein, J. Rougier, and A. Seheult (2001), Bayesian forecasting for complex systems using computer simulators, *J. Am. Stat. Assoc.*, *96*, 717–729.

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 (2015), 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.

Dietzel, A., and P. Reichert (2012), Calibration of computationally demanding and structurally uncertain models with an application to a lake water quality model, *Environ. Modell. Software*, *38*, 129–146.

Dotto, C., G. Mannina, M. Kleidorfer, L. Vezzaro, M. Henrichs, D. McCarthy, G. Freni, W. Rauch, and A. Deletic (2012), Comparison of different uncertainty techniques in urban stormwater quantity and quality modelling, *Water Res.*, *46*(8), 2545–58, doi:10.1016/j.watres.2012.02.009.

Einicke, G. A. (2012), *Smoothing, Filtering and Prediction: Estimating the Past, Present and Future*, 286 pp., InTech, doi:10.5772/2706.

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.

Freni, G., and G. Mannina (2012), Uncertainty estimation of a complex water quality model: The influence of Box-Cox transformation on Bayesian approaches and comparison with a non-Bayesian method, *Phys. Chem. Earth Parts A/B/C*, *42*, 31–41.

Freni, G., G. Mannina, and G. Viviani (2009), Uncertainty assessment of an integrated urban drainage model, *J. Hydrol.*, *373*, 392–404, doi:10.1016/j.jhydrol.2009.04.037.

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.

Friling, N., M. J. Jiménez, H. Bloem, and H. Madsen (2009), Modelling the heat dynamics of building integrated and ventilated photovoltaic modules, *Energy Buildings*, *41*(10), 1051–1057, doi:10.1016/j.enbuild.2009.05.018.

Gay, M. (1990), Usage summary for selected optimization routines, *Comput. Sci. Tech. Rep. 153*, AT&T Bell Lab., Murray Hill, N. J. [Available at http://netlib.bell-labs.com/cm/cs/cstr/153.pdf.]

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.

Haario, H., E. Saksman, and J. Tamminen (2001), An adaptive Metropolis algorithm, *Bernoulli*, *7*, 223–242.

Hastings, W. (1970), Monte Carlo sampling methods using Markov chains and their applications, *Biometrika*, *57*, 97–109.

Higdon, D., M. Kennedy, J. C. Cavendish, J. A. Cafeo, and R. D. Ryne (2005), Combining field data and computer simulations for calibration and prediction, *SIAM J. Sci. Comput.*, *26*, 448–466, doi:10.1137/S1064827503426693.

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.

Iacus, S. M. (2008), *Simulation and Inference for Stochastic Differential Equations: With R examples*, 286 pp., Springer-Verlag, N. Y., doi:10.1007/978-0-387-75839-8.

Jonsdottir, H., H. A. Nielsen, H. Madsen, J. Eliasson, O. P. Palsson, and M. K. Nielsen (2007), Conditional parametric models for storm sewer runoff, *Water Resour. Res.*, *43*, W05443, doi:10.1029/2005WR004500.

Jørgensen, H. K., S. Rosenørn, H. Madsen, and P. S. Mikkelsen (1998), Quality control of rain data used for urban runoff systems, *Water Sci. Technol.*, *37*(11), 113–120.

Juhl, R., N. R. Kristensen, P. Bacher, J. K. Kloppenborg, and H. Madsen (2013), *CTSM: Continuous Time Stochastic Modeling*, Version 0.6.7., Dept. Appl. Math. and Comput. Sci., Tech. Univ. of Den., Kongens Lyngby, Denmark. [Available at http://ctsm.info/.]

Kao, J., D. Flicker, R. Henninger, S. Frey, M. Ghil, and K. Ide (2004), Data assimilation with an extended Kalman filter for impact-produced shock-wave dynamics, *J. Comput. Phys.*, *196*, 705–723, doi:10.1016/j.jcp.2003.11.028.

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.

Kennedy, M. and A. O'Hagan (2001), Bayesian calibration of computer models, *J. R. Stat. Soc., Ser. B*, *63*, 425–464.

Kleidorfer, M., A. Deletic, T. D. Fletcher, and W. Rauch (2009), Impact of input data uncertainties on urban stormwater model parameters, *Water Sci. Technol.*, *60*, 1545–1554, doi:10.2166/wst.2009.493.

Kloeden, P. E., and E. Platen (1992), *Numerical Solution of Stochastic Differential Equations*, vol. 3, Springer-Verlag, Berlin Heidelberg, doi:10.1007/978-3-662-12616-5.

Kristensen, N. R., and H. Madsen (2003), *Continuous Time Stochastic Modelling CTSM 2.3—Mathematics Guide*. [Available at http://www.webcitation.org/6PI9H6pBR.]

Kristensen, N. R., H. Madsen, and S. B. Jørgensen (2004), Parameter estimation in stochastic grey-box models, *Automatica*, *40*(2), 225–237, doi:10.1016/j.automatica.2003.10.001.

Kristensen, N. R., H. Madsen, and S. H. Ingwersen (2005), Using stochastic differential equations for PK/PD model development, *J. Pharmacokinetics Pharmacodynamics*, *32*(1), 109–41, doi:10.1007/s10928-005-2105-9.

Kuczera, G. (1983), Improved parameter inference in catchment models: 1. Evaluating parameter uncertainty, *Water Resour. Res.*, *19*(5), 1151–1162, doi:10.1029/WR019i005p01151.

Kuczera, G., D. Kavetski, S. Franks, and M. Thyer (2006), Towards a Bayesian total error analysis of conceptual rainfall-runoff models: Characterising model error using storm-dependent parameters, *J. Hydrol.*, *331*(1), 161–177, doi:10.1016/j.jhydrol.2006.05.010.

Law, K. J. H., and A. M. Stuart (2012), Evaluating data assimilation algorithms, *Mon. Weather Rev.*, *140*, 3757–3782, doi:10.1175/MWR-D-11-00257.1.

Lin, Z., and M. Beck (2007), On the identification of model structure in hydrological and environmental systems, *Water Resour. Res.*, *43*, W02402, doi:10.1029/2005WR004796.

Löwe, R., P. S. Mikkelsen, and H. Madsen (2014), Stochastic rainfall-runoff forecasting: Parameter estimation, multi-step prediction, and evaluation of overflow risk, *Stochastic Environ. Res. Risk Assess.*, *28*, 505–516, doi:10.1007/s00477-013-0768-0.

Madsen, H. (2008), *Time Series Analysis*, 380 pp., Chapman and Hall.

Mejía, A., E. Daly, F. Rossel, T. Jovanovic, and J. Gironas (2014), A stochastic model of streamflow for urbanized basins, *Water Resour. Res.*, *50*, 1984–2001, doi:10.1002/2013WR014834.

Melgaard, H. (1994), Identification of physical models, PhD thesis, Tech. Univ. of Den., Institute of Mathematical Modelling. [Available at http://orbit.dtu.dk/en/publications/identification-of-physical-models%28b323400c-213b-48e2-afb1-22090ba1e530%29.html.]

Møller, J. K. (2010), Stochastic state space modelling of nonlinear systems, PhD thesis, Tech. Univ. of Den., Kgs. Lyngby, Denmark. [Available at http://orbit.dtu.dk/en/publications/stochastic-state-space-modelling-of-nonlinear-systems--with-application-to-marine-ecosystems%28a7077954-1b01-4c88-9f48-8010973f1e6e%29.html.]

Moradkhani, H., C. M. DeChant, and S. Sorooshian (2012), Evolution of ensemble data assimilation for uncertainty quantification using the particle filter-Markov chain Monte Carlo method, *Water Resour. Res.*, *48*, W12520, doi:10.1029/2012WR012144.

O'Hagan, A. (2006), Bayesian analysis of computer code outputs: A tutorial, *Reliab. Eng. Syst. Safety*, *91*, 1290–1300.

Platen, E., and Bruti-Liberati, N. (2010), *Numerical Solution of Stochastic Differential Equations With Jumps in Finance*, vol. 64, 856, pp., Springer-Verlag, Berlin Heidelberg, doi:10.1007/978-3-642-13694-8.

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. [Available at http://www.webcitation.org/6P7vrJNzs.]

Reichert, P. (2012), Conceptual and practical aspects of quantifying uncertainty in environmental modelling and decision support, paper presented at 2012 International Congress on Environmental Modelling and Software, Int. Environ. Modell. and Software Soc., edited by R. Seppelt et al., Leipzig, Germany. [Available at http://www.webcitation.org/6P7w8shw9.]

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.

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

Restrepo, J. M. (2008), A path integral method for data assimilation, *Atmos. Res.*, *237*(1), 14–27, doi:10.1016/j.physd.2007.07.020.

Rossman, L., and W. Supply (2010), Storm Water Management Model User's Manual, Version 5.0., National Risk Management Research Laboratory, Office of Research and Development, US Environmental Protection Agency.

Sadegh, P., H. Melgaard, H. Madsen, and J. Holst (1994), Optimal experiment design for identification of grey-box models, in *Proceedings of 1994 American Control Conference—ACC '94*, vol. 1, pp. 132–137, IEEE, doi:10.1109/ACC.1994.751709. [Available at http://orbit.dtu.dk/en/publications/optimal-experiment-design-for-identification-of-greybox-models%2812de45ce-ea56-4963-9f98-07ded8026d39%29.html.]

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.

Schilling, W. (1991), Rainfall data for urban hydrology: What do we need?, *Atmos. Res.*, *27*(1), 5–21.

Schilling, W., and L. Fuchs (1986), Errors in stormwater modeling—A quantitative assessment, *J. Hydraul. Eng.*, *112*(2), 111–123.

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

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.

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.

Thyer, M., B. Renard, D. Kavetski, G. Kuczera, S. Franks, and S. Srikanthan (2009), Critical evaluation of parameter consistency and predictive uncertainty in hydrological modeling: A case study using Bayesian total error analysis, *Water Resour. Res.*, *45*, W00B14, doi:10.1029/2008WR006825.

Vezzaro, L., and M. Grum (2014), A generalized dynamic overflow risk assessment (DORA) for urban drainage real time control, *J. Hydrol.*, *515*, 292–303, doi:10.1016/j.jhydrol.2014.05.019.

Vezzaro, L., P. S. Mikkelsen, A. Deletic, and D. McCarthy (2013), Urban drainage models—Simplifying uncertainty analysis for practitioners, *Water Sci. Technol.*, *68*(10), 2136–2143, doi:10.2166/wst.2013.460.

Vihola, M. (2012), Robust adaptive Metropolis algorithm with coerced acceptance rate, *Stat. Comput.*, *22*, 997–1008.

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

Walter, E., and L. Pronzato (1997), *Identification of Parametric Models From Experimental Data*, 413 pp., Springer-Verlag, London, U. K.

Wilkinson, R. D., M. Vrettas, D. Cornford, and J. E. Oakley (2011), Quantifying simulator discrepancy in discrete-time dynamical simulators, *J. Agric. Biol. Environ. Stat.*, *16*, 554–570, doi:10.1007/s13253-011-0077-3.

Willems, P. (2012), Model uncertainty analysis by variance decomposition, *Phys. Chem. Earth Parts A/B/C*, *42–44*, 21–30, doi:10.1016/j.pce.2011.07.003.

Wolfs, V., Villazon, M. and Willems, P. (2013), Development of a semi-automated model identification and calibration tool for conceptual modelling of sewer systems, *Water Sci. Technol.*, *68*, 167–175, doi:10.2166/wst.2013.237.

Yang, J., P. Reichert, and K. Abbaspour (2007), 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.