

Integrated uncertainty assessment of discharge predictions with a statistical error model

M. Honti,¹ C. Stamm,¹ and P. Reichert¹

Received 15 June 2012; revised 25 May 2013; accepted 17 June 2013; published 7 August 2013.

[1] A proper uncertainty assessment of rainfall-runoff predictions has always been an important objective for modelers. Several sources of uncertainty have been identified, but their representation was limited to complicated mechanistic error propagation frameworks only. The typical statistical error models used in the modeling practice still build on outdated and invalidated assumptions like the independence and homoscedasticity of model residuals and thus result in wrong uncertainty estimates. The primary reason for the popularity of the traditional faulty methods is the enormous computational requirement of full Bayesian error propagation frameworks. We introduce a statistical error model that can account for the effect of various uncertainty sources present in conceptual rainfall-runoff modeling studies and at the same time has limited computational demand. We split the model residuals into three different components: a random noise term and two bias processes with different response characteristics. The effects of the input uncertainty are simulated with a stochastic linearized rainfall-runoff model. While the description of model bias with Bayesian statistics cannot directly help to improve on the model's deficiencies, it is still beneficial to get realistic estimates on the overall predictive uncertainty and to rank the importance of different uncertainty sources. This feature is particularly important if the error sources cannot be addressed individually, but it is also relevant for the description of remaining bias when input and structural errors are considered explicitly.

Citation: 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.

1. Introduction

[2] Like other mathematical models, conceptual rainfall-runoff models (CRRMs) are simplified representations of heterogeneous and complex systems. Since simplification and aggregation (be it spatial and/or temporal) form the backbone of the modeling process, our predictions can produce reasonable results at best but will not be very accurate. This limitation enhances the need to estimate the confidence we can attribute to model-based predictions. One has to quantify the expected divergence between the model predictions and reality.

[3] Uncertainty can arise at any stage of the modeling process. When describing observed output, one can distinguish between the uncertainty of input, of model structure and parameters, and of the observations. The classical approach considers parameter and observation uncertainty only and describes the deviations between the deterministic model output and observations by a random noise term corresponding to an assumed measurement error. This corre-

sponds to a calibration procedure of minimizing the sum of the squared deviations between the deterministic model results and the measurements to determine parameter estimates. However, uncertainty affects CRRMs and other hydrological models in much more profound ways. Generally, there is substantial input uncertainty in the atmospheric drivers (i.e., precipitation and evapotranspiration) to the hydrological response of a catchment. Due to the scarcity of gauging stations compared to the spatial variability of precipitation, a perfect model would even fail to reproduce the real discharge due to the observation error. Additional uncertainty arises from necessary simplification and (spatial) aggregation. Such structural model errors affect the model predictions in a different way than purely random measurement errors. Similarly to input uncertainty, structural uncertainty produces model residuals that are auto-correlated in time, which leads to “wrong” representations of the internal state of the catchment (e.g., soil moisture status and groundwater levels). Because the hydrological response is generally state dependent, such errors also affect subsequent time steps of the CRRM predictions. A further general issue in hydrological modeling is the identifiability problem of model parameters. This problem stems from the fact that a CRRM calibration data set consisting of time series of atmospheric input and observed discharge (at a single or a few monitoring stations in most cases) is typically insufficient to identify all parameters. As a consequence, numerous parameter sets may

¹Eawag: Swiss Federal Institute of Aquatic Sciences and Technology, Dübendorf, Switzerland.

Corresponding author: M. Honti, Eawag: Swiss Federal Institute of Aquatic Sciences and Technology, Ueberlandstrasse 133, Dübendorf CH-8600, Switzerland. (mark.honti@eawag.ch)

yield simulations having similarly good agreement with the observed data. All these factors will typically lead to model residuals that have more complex statistical properties (i.e., autocorrelation, heteroscedasticity, heavy tails, and skewness) than the classical white noise error.

[4] Since discrepancies between the statistical assumptions and the true properties of model residuals result in biased parameters and unreliable prediction uncertainty intervals, several approaches have been developed to build more realistic statistical error models for rainfall-runoff simulations. The normality of residuals can be improved by the standard statistical procedure of transformation: power transforming both the model output and the measurements simultaneously reduces heteroscedasticity, skewness, and heavy tails [Abdulla *et al.*, 1999; Bates and Campbell, 2001; Demaria *et al.*, 2007; Duan *et al.*, 2007; Yang *et al.*, 2007; Frey *et al.*, 2011]. High autocorrelation can be treated with an autoregressive error model [Sorooshian and Dracup, 1980; Bates and Campbell, 2001; Yang *et al.*, 2007; Frey *et al.*, 2011]. Schoups and Vrugt [2010] combined a deterministic bias correction with a heteroscedastic autoregressive process and a versatile Skew Exponential Power (SEP) distribution to build a universal, yet entirely statistical, error model.

[5] While these techniques allow us to make less restrictive and thus more realistic statistical assumptions on total model error, they yield practically no insight into the origin and propagation of uncertainty. This is especially true for input uncertainty, which has the most complex propagation mechanism. Therefore, Bayesian uncertainty assessment frameworks have been developed that are able to propagate errors through the nonlinear deterministic model [Kuczera *et al.*, 2006; Ajami *et al.*, 2007]. The Bayesian foundation enables the analyst to treat model parameters as stochastic variables and incorporate existing knowledge about them via prior distributions. The Bayesian Total Error Analysis (BATEA) [Kavetski *et al.*, 2006] and Integrated Bayesian Uncertainty Estimator (IBUNE) [Ajami *et al.*, 2007] uncertainty assessment concepts and the study by Vrugt *et al.* [2008] using the Differential Evolution Adaptive Metropolis (DREAM) sampler (we refer to that whole study herein-after as DREAM) all provide methods to treat uncertainty in rainfall measurements. Considering input uncertainty adds complexity to these calibration methods. The BATEA and DREAM studies introduce storm-specific rainfall multipliers and infer them together with the model parameters. The technical difficulty lies in the fact that the number of estimated parameters becomes much larger due to the storm-specific parameters that make the sampling of the posterior more demanding. IBUNE applies a set of rainfall multipliers a priori drawn from a normal distribution, which are later shifted and scaled according to two additional input error parameters (the unknown mean and variance) for the estimation of the likelihood. Besides the estimation of parameter uncertainty, error propagation can support the detection of structural deficiencies of models by using time-variable parameters [Reichert and Mieleitner, 2009] or can quantify other sources of uncertainty to derive a more precise prediction of uncertainty intervals [Renard *et al.*, 2010]. These frameworks offer flexibility as one can account for almost any desired uncertainty compo-

nent, but this comes at the price of a high-computational burden and mathematical complexity.

[6] In such studies, there is still a “remnant error” that contains all uncertainty that has not been accounted for elsewhere in the error model. While the remnant error represents only a part of the total uncertainty, it can still show some statistical complexity, which obviously depends on the ability of the rest of the error model to describe all sources of uncertainty. In the worst case (a totally inappropriate error model), the remnant error can be identical to the model residuals. Despite this and likely due to the limited theoretical relevance of the remnant error, some error-propagation studies still assume that it is independent and normally distributed [Kuczera *et al.*, 2006; Renard *et al.*, 2010].

[7] The complexity of Bayesian uncertainty assessment frameworks prevents their widespread usage in cases when the exact description of error propagation is not absolutely required. Göttinger and Bárdossy [2008] have already attempted to provide a simple standalone error model that separates the effects of various sources of uncertainty. Nevertheless, they still assumed that the errors were independent and that the structural uncertainty was bound to the process sensitivities through a linear combination. This neglects that sensitivities derived from a potentially incorrect model structure are not guaranteed to reflect the true importance of the main hydrological processes.

[8] Therefore, the goal of this study is to develop a formal statistical error model that is able to account for the effects of all sources of uncertainty by emulating the key properties of error propagation through the CRRM. Such a method could bridge the gap between the fast yet typically unsatisfactory traditional statistical error models and the accurate yet computationally demanding mechanistic error propagating methods. In addition, the method could be also used alongside mechanistic error propagation to describe remnant errors. We inspect whether the new method can fulfill the requirements of reasonable speed and of statistical assumptions by comparing it to three existing Gaussian error models: (i) the traditional model of independent, normally distributed (measurement) errors (error model E), (ii) the first-order autoregressive models (error model B), and (iii) a recently introduced error model (B + E), which describes the residual series as the composite of systematic and independent error processes.

2. Methods

[9] Statistical error models are formulated by making statistical assumptions about the contributions of the different sources of uncertainty and their interaction with the model residual series. Based on these assumptions, the likelihood calculation algorithm infers the parameters of the different components from the composite residual series. This is a statistically demanding task. Kennedy and O'Hagan [2001] described how to distinguish between the effects of two error processes on model output given some prior knowledge about their statistical properties. This method was subsequently applied by Bayarri *et al.* [2007] for general purpose statistical modeling, but it has been overlooked by the environmental and hydrological modeling communities until the work of Reichert and Schuwirth

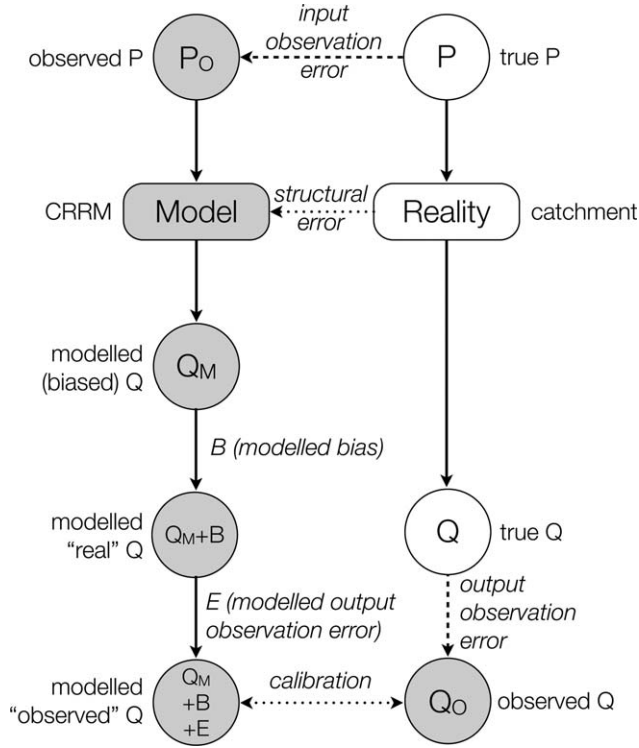


Figure 1. A schema for uncertainty description in conceptual rainfall-runoff modeling. P : precipitation, Q : discharge.

[2012], who used it for multicriteria calibration. We introduce some modifications to this method to construct a likelihood function that can simultaneously consider structural, input, and output uncertainty for CRRMs.

[10] We start with a generic setup following the example in *Reichert and Schuwirth* [2012]. They described the residual process with the sum of a Gaussian bias process B representing the effects of structural and input uncertainty on model results and an independent noise E (Figure 1). The E -term is often equated with observation uncertainty, which is not completely appropriate. The uncertainty of discharge measurements stems also from systematic deviations (e.g., due to errors in rating curves) that show up in the B -term with only random fluctuations being represented by E . Neglecting this difference due to the lack of information on the systematic errors in discharge, the observed, real, and modeled discharges (Q_O , Q , and Q_M , respectively) relate to each other according to

$$Q_O = Q + E = [Q_M + B] + E. \quad (1)$$

[11] The CRRM transforms an input error into a discharge error through the model's structure, state, and parameterization. Since our objective is to avoid mechanistic error propagation, our statistical error model should be able to emulate the a priori unknown model response. For this purpose, we introduce two separate precipitation-dependent structural error processes. The first can account for errors coming from the fast responding model mechanisms (like runoff formation), and the second for errors from the slower ones (like base flow). We assume that the

errors of the fast-responding mechanisms ($B_f(P)$) are memoryless and only active when it is raining, while errors of the slowly responding mechanisms ($B_s(P)$) have significant self-dependence. Then equation (1) changes to

$$Q_O = Q + E = [Q_M + B_f(P) + B_s(P)] + E. \quad (2)$$

[12] With the error model structure specified, the next task is to formulate $B_f(P)$, $B_s(P)$, and E . Since our objective is to describe the effects of input uncertainty without propagating the input errors through the nonlinear CRRM, we have to emulate the deterministic model's response to the stochastic rainfall uncertainty.

2.1. The Memoryless Error Term E

[13] We assume that the random measurement errors are independent and follow a normal distribution with the standard deviation of σ_E . Thus, the parameter set of E is $\Psi = [\sigma_E]$.

2.2. The Fast Bias Component $B_f(P)$

[14] The precipitation-discharge response of a CRRM can be locally linearized by assuming that a minor change in the precipitation will alter runoff formation proportionally. This is simply the century-old rational method (for a recent summary see *Butler and Davies* [2000]) for the calculation of discharge errors caused by input uncertainty. The local slope of the response function is the runoff coefficient that depends on the internal state of the model and the precipitation intensity. The linearization ensures that if the uncertainty of precipitation input is normally distributed then runoff will be too, with

$$\sigma_r = \sigma_{pm} P c_r(P, \Omega), \quad (3)$$

where $c_r()$ is a function specifying the local runoff coefficient depending on precipitation (P) and the internal model state (Ω), σ_r , and σ_{pm} are the standard deviations of the resulting runoff and of the precipitation multipliers, respectively. If we neglect the variation in c_r caused by the different internal model states, equation (3) becomes simply

$$\sigma_r = \kappa_f P. \quad (4)$$

[15] Since $B_f(P)$ stands for the error in runoff formation and we assume that runoff appears in stream discharge instantly, the variance of $B_f(P)$ is 0 when there is no precipitation and σ_r^2 when it rains. Using the simple formulation in equation (4), we get the following covariance matrix for $B_f(P)$:

$$\Sigma_{B_f}(i, j) = \begin{cases} (\kappa_f P(t_i))^2 & \text{if } i = j \\ 0 & \text{if } i \neq j \end{cases}, \quad (5)$$

where $P(t_i)$ is the precipitation intensity at t_i . Thus, the parameter set of $B_f(P)$ will be $\Xi_f = [\kappa_f]$.

2.3. The Slow Bias Component $B_s(P)$

[16] As a basis for our $B_s(P)$ process, we use the Ornstein-Uhlenbeck (OU) process, which is the continuous equivalent of the first-order autoregressive (AR(1)) model. The OU process is a stationary Gauss-Markov process: its

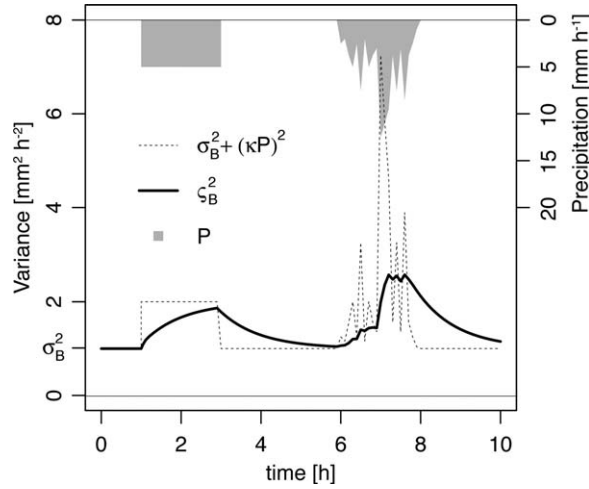


Figure 2. Time dynamics of the unconditional ($\sigma_B^2 + (\kappa P)^2$) and asymptotic (ζ_B^2) variance of a precipitation-dependent bias process under uniform and “natural” precipitation intensity (P). The total rainfall depth was the same (10 mm in 2 h) for both events.

distribution is Gaussian at any time with a constant variance of σ_B^2 and its future values depend only on the present state. This process has been used to describe model bias in hydrology in its original stationary form [Yang et al., 2007].

[17] Here, we assume that $B_s(P)$ is a stationary OU process, but our variant suffers from additional stochastic disturbances correlated to an external input. In this way, the stochastic process can follow the basic dynamics of the CRRM while being mathematically independent from it. Details about the statistical properties of the standard and disturbed OU processes are described in Appendix A; here, we summarize the most relevant information.

[18] When started from a Gaussian initial distribution, the asymptotic (conditional) variance ζ_B^2 of the stationary OU process follows simple kinetics:

$$\frac{d\zeta_B^2}{dt} = 2\beta(\sigma_B^2 - \zeta_B^2). \quad (6)$$

where t is time and β is the inverse correlation length.

[19] The stochastic disturbance is formulated in a similar form to the fast bias component $B_f(P)$ in equation (4). The stationary variance increases linearly with the precipitation with κ_s being the “runoff coefficient” (Figure 2). Then, the asymptotic variance kinetics becomes:

$$\frac{d\zeta_B^2}{dt} = 2\beta\left(\left[\sigma_B^2 + (\kappa_s P)^2\right] - \zeta_B^2\right). \quad (7)$$

[20] For equidistant Δt time steps and uniform precipitation distribution between adjacent time steps, we get

$$\zeta_{B_i}^2 = \sigma_B^2 + \left(\zeta_{B_{i-1}}^2 - \sigma_B^2\right) \exp\{-2\beta\Delta t\} + (\kappa_s P)^2 \quad (8)$$

with $\kappa_s = \kappa_s \sqrt{1 - \exp\{-2\beta\Delta t\}}$. For changing time steps κ_s will depend on the length of the time step $[\Delta t]$, neverthe-

less the same formula applies. The exponential covariance structure of the OU process is preserved despite the additional disturbance with the covariance matrix of $B_s(P)$ given by

$$\Sigma_{B_s}(i, j) = \begin{cases} \zeta_B^2(t_i) & \text{if } i = j \\ \zeta_B^2(t_i) \exp(-\beta|t_j - t_i|) & \text{if } i < j \\ \zeta_B^2(t_j) \exp(-\beta|t_i - t_j|) & \text{if } i > j \end{cases} \quad (9)$$

[21] Since the disturbed process is still Gaussian, the conditional probability of $B_s(P)$ given its value in the preceding step follows a normal distribution:

$$f(B_s(P_i, t_i) = b_i | B_s(P_{i-1}, t_{i-1}) = b_{i-1}) = \mathcal{N}(b_{i-1} \exp(-\beta\Delta t), \sigma_B^2(1 - \exp(-2\beta\Delta t) + (\kappa_s P_{i-1})^2)) \quad (10)$$

[22] The parameter set of $B_s(P)$ consists of three elements $\Xi_s = [\sigma_B, \beta, \kappa_s]$.

2.4. Transformation of Discharge Time Series

[23] We power transformed the measured and modeled discharge series according to Box and Cox [1964] before the application of the error models. The reasons were the following:

[24] 1. The compared additive statistical error models all have an unconstrained normal error distribution and so they are likely to give a significant probability to negative discharge values when the uncertainty is higher than actual discharge. This can be avoided with the application of a log-transform or a Box-Cox transform with $\lambda = 0$, which would turn the additive error model into a multiplicative one. Unfortunately, a strong transformation usually reduces the quality of fit for high discharge events to an unacceptable level. Less-skewed transformations ($1 > \lambda > 0$) cannot guarantee strictly positive discharges, but can reduce their occurrence and still maintain a healthier balance between the importance of high and low Q events.

[25] 2. Transformation is the only way to introduce some heteroscedasticity into the E, B, and E + B error models. They are known to fail without a transformation due to the strongly changing variance of discharge errors. To compare the B(P) + E error model to the simpler error models, we have to ensure that the latter also perform acceptably. Moreover, we can check whether the heteroscedasticity of B(P) + E is just equivalent to using a simpler error model with a power transformation or if there are additional benefits coupled to the more complicated structure.

[26] 3. The assumption on the linearity between precipitation and discharge errors with a uniform coefficient is a strong restriction in the B(P) + E error model. Some dependence on Q and thus the internal state of the system can be introduced by the application of transformation. However, this also affects the theoretically independent measurement noise besides the bias, so a strong transformation should be avoided.

[27] Based on the above points, we tested three different transformation parameter settings: $\lambda = 1, 0.5$, and 0.3 . The $\lambda = 1$ case corresponds to no transformation, while the

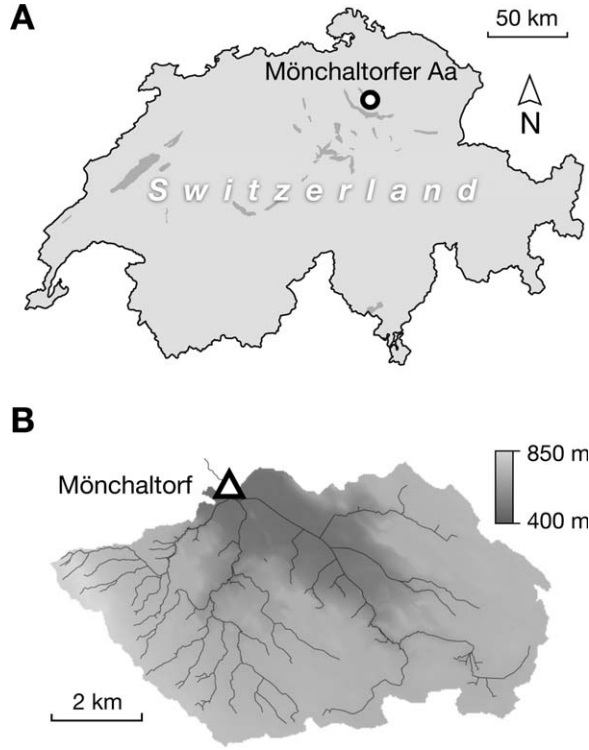


Figure 3. (a) Location of the study site within Switzerland and (b) the topographic map of the Mönchaltorfer Aa catchment with the gauge site (triangle).

latter values are typical for transforming discharge series [Summer et al., 1997; Thyer et al., 2002; Willems, 2009].

2.5. The Posterior Distribution

[28] As the transformation parameter is not calibrated, the entire parameter space consists of the parameters of the deterministic CRRM (Θ), the parameters of $B_s(P)$ (Ξ_s), of $B_f(P)$ (Ξ_f), and those of $E(P)$ (Ψ).

[29] The posterior probability of a specific parameter set $[\theta, \xi_f, \xi_s, \psi]$ given the actual transformed measurements (\mathbf{Q}_O^r) can be calculated with a multivariate normal distribution that builds on the transformed simulations of the deter-

ministic CRRM (\mathbf{Q}_M^r) and the prior distributions of the parameters [Reichert and Schuwirth, 2012]:

$$f(\theta, \psi, \xi_f, \xi_s | \mathbf{Q}_O^r) \propto \frac{f(\theta, \xi_f, \xi_s, \psi)}{\sqrt{\det(\Sigma_E + \Sigma_{B_f} + \Sigma_{B_s})}} \times \exp\left(-\frac{1}{2} [\mathbf{Q}_O^r \quad \mathbf{Q}_M^r]^T (\Sigma_E + \Sigma_{B_f} + \Sigma_{B_s})^{-1} [\mathbf{Q}_O^r \quad \mathbf{Q}_M^r]\right) \quad (11)$$

[30] The posterior distribution is taken in its original form from Reichert and Schuwirth [2012], because this general solution applies to all cases when $B = B_f + B_s$ and E are Gaussian.

2.6. Case Study

2.6.1. Study Site

[31] We tested the newly developed error model with the hydrological response of the small catchment of the Mönchaltorfer Aa, located on the Swiss Plateau (Figure 3). We used the daily Q data from the gauge at Mönchaltorf [Amt für Abfall, Wasser, Energie und Luft der Baudirektion des Kanton Zürich (AWEL), 2010]. The upstream catchment area is 46 km² with intensive agriculture (57%) and settlements (20%) being the dominant land use categories [SWISSTOPO, 2008]. The hydrology of the catchment is characterized by a relatively low base flow index (0.42–0.48, Siber et al. [2009]). Annual average precipitation was 1220 mm in the study period with a mean discharge of 0.98 m³ s⁻¹ (672 mm yr⁻¹) at the monitoring station. The dominant soil types are cambisols on hillsides and gleysols on flat areas (see the soil map in Wittmer et al. [2010]).

[32] The observed flow values were aggregated from 10 min observations to daily values, which practically eliminated all random scattering in the data. Consequently, we expected the random observation error (E) to be negligible after calibration. The study period lasted from 1 July 2000 to 31 December 2009 when there was continuous local precipitation data available from the WWTP of Mönchaltorf [AWEL, 2010]. The Q-P data set was divided into a calibration period covering 75% of the data set (from 2000 to December 2007) and verification periods corresponding to 25% of the data.

Table 1. Process Matrix of the CRRM^a

Process	Symbol	Rate	Affected Storage			
			h_{snow}	h_s	h_{gw}	h_q
Snowfall	P_{snow}	$\begin{cases} P & \text{if } T < T_{crit} \\ 0 & \text{otherwise} \end{cases}$	+			
Snowmelt	P_{melt}	$\begin{cases} k_{melt}(T - T_{melt}) & \text{if } T > T_{melt} \\ 0 & \text{otherwise} \end{cases}$		+		
Rainfall	P_{rain}	$\begin{cases} P & \text{if } T \geq T_{crit} \\ 0 & \text{otherwise} \end{cases}$		+		
Evapotranspiration	E	$E_{pot} f_{et}$				
Saturation excess runoff	Q_{runoff}	$f_{sat}(P_{rain} + P_{melt})$				+
Groundwater recharge	Q_{rge}	$k_{rge} f_{sat} q_{seep}^{max}$			+	
Subsurface flow	Q_{ssf}	$(1 - k_{rge}) f_{sat} q_{seep}^{max}$				+
Baseflow	Q_{bf}	$k_{bf} h_{gw}$				+
Stream discharge	Q_q	$k_q h_q$				+

^aAll storages are in (mm) while processes are in (mm d⁻¹). f_{sat} and f_{et} are defined in equations (12) and (13), respectively.

Table 2. Parameters of the CRRM

Symbol	Description	Unit	Prior ^a
T_{crit}	Critical temperature for snowfall	°C	N(1, 0.5)
T_{melt}	Threshold temperature for snowmelt	°C	N(0, 1)
k_{melt}	Temperature-specific snowmelt rate constant	mm °C ⁻¹ d ⁻¹	LN(3, 1.2)
h_{FS}	Soil moisture when almost all the catchment is saturated	mm	LN(350, 20)
h_{FC}	Soil moisture when almost no area is saturated	mm	LN(225, 15)
h_{WP}	Soil moisture when evapotranspiration almost ceases	mm	LN(100, 15)
k_{rge}	Proportion of groundwater recharge from seepage		B(14, 6)
q_{seep}^{max}	Maximal seepage rate	mm d ⁻¹	LN(100, 50)
k_{bf}	Baseflow constant	yr ⁻¹	LN(0.2, 0.2)
k_q	Stream constant ^b	d ⁻¹	

^aNotation for prior distributions: N(μ , σ): normal, LN(μ , σ): lognormal, and B(α , β): beta.

^bThis parameter was set to a fixed value of 48 in all simulations.

2.6.2. Rainfall-Runoff Model

[33] The CRRM is a modified version of the simple logistic saturated path model logSPM of *Kavetski et al.* [2006]. The model was extended with a snow module to have all together four storages. The process formulae and parameters are specified in Tables 1 and 2, respectively.

[34] Precipitation (P) was filtered by a snow module constructed following *Martinez and Rango* [1981]. Precipitation that falls below a critical temperature (T_{crit}) was defined as snow and it accumulates in the snow storage h_{snow} . If air temperature exceeds a melt threshold (T_{melt}), snow starts to melt with the common degree-day method [*Martinez and Rango*, 1981]. Precipitation falling at an air temperature exceeding T_{crit} directly reaches the underlying soil storage.

[35] The rest of the model routes flow between the soil moisture (h_s), groundwater (h_{gw}), and stream (h_q) storages. Soil moisture is modeled according to the original logSPM model, but some reformulation was applied (see below) to improve the identifiability of parameters. The logSPM model [*Kuczera et al.*, 2006] belongs to the saturated path family of semidistributed hydrological models, which implements the variable contributing area concept by an event-invariant saturation function that maps between average soil moisture and the runoff generating area [*Kavetski et al.*, 2003; *Lazzarotto et al.*, 2006]. The soil moisture profile is simplified to a homogeneous depth distribution. The core of the model is the sigmoid saturation function f_{sat} :

$$f_{sat}(h_s) = \frac{1}{1 + \exp\left(4 \frac{h_{FS} + h_{FC} - 2h_s}{h_{FS} - h_{FC}}\right)} - \frac{1}{1 + \exp\left(4 \frac{h_{FS} + h_{FC}}{h_{FS} - h_{FC}}\right)}, \quad (12)$$

where h_{FS} and h_{FC} are the catchment-scale storage level equivalents of full saturation and field capacity, respectively. The parameterization of f_{sat} differs from that of *Kuczera et al.* [2006]. This ensures that the prior knowledge on the characteristic moisture content could be directly introduced, but the function is mathematically equivalent to the original. Runoff is formed on the saturated proportion of the catchment for any rain event [*Reichert and Mieleitner*, 2009]. Underground flow components are active only in the saturated area: groundwater recharge and fast groundwater flow are generated proportionally to f_{sat} . Evapotranspiration from the soil moisture storage is controlled in a similar manner to f_{sat} :

$$f_{et}(h_s) = \frac{1}{1 + \exp\left(4 \frac{h_s - h_{WP}}{h_{FC} - h_{WP}}\right)} - \frac{1}{1 + \exp\left(4 \frac{h_{WP}}{h_{FC} - h_{WP}}\right)}, \quad (13)$$

where h_{WP} is the catchment-scale moisture level equivalent of the wilting point.

[36] The groundwater and stream storages are simple linear reservoirs without size constraints.

2.7. Numerical Implementation

2.7.1. Solving the CRRM Equations

[37] The snow module in the CRRM exhibits a strong threshold behavior (Table 1), which causes problems for the solver and optimization routines [*Kavetski and Kuczera*, 2007]. To prevent this, the threshold function is computed with a numerically well-behaved reformulation of the soft maximum function (J. D. Cook, 2010, How to compute the soft maximum, <http://www.johndcook.com/blog/2010/01/20/how-to-compute-the-soft-maximum>):

$$\text{softmax}(x, y) = \frac{\ln(1 + \exp(k \cdot \min(x, y) - k \cdot \max(x, y)))}{k} + \max(x, y) \quad (14)$$

where k is an arbitrary factor specifying the scale of smoothness around the threshold. The soft maximum function converges to the original maximum function as k approaches infinity.

[38] The differential equations of the CRRM are solved with the LSODA (a variant of LSODE - Livermore Solver for Ordinary Differential Equations) solver [*Hindmarsh*, 1983; *Petzold*, 1983], which automatically switches between stiff and nonstiff solution methods according to the behavior of the system to be solved. Together with the smoothing technique described above, this ensures that the objective function is free from roughness and virtual optima generated by numerical artifacts [*Clark and Kavetski*, 2009].

2.7.2. Likelihood Calculation

[39] The direct evaluation of the likelihood function equation (11) on long input series can pose serious numerical problems when Σ_E , Σ_{B_r} , and Σ_{B_s} are full covariance matrices [*Reichert and Schuwirth*, 2012]. Since hydrologic observation series often span decades, these matrices can grow to enormous sizes. The storage requirements can be solved if we store them as sparse matrices, but the matrix

operations are still prone to numerical problems (over or underflow) and require a long-computational time.

[40] However, using a descendant of the OU process for $B_s(P)$ simplifies the solution due to the Markov property:

[41] 1. *Rybicki and Press* [1995] showed that the inverse covariance matrix $\Sigma_{B_s}^{-1}$ of a Gauss-Markov process is tri-diagonal and symmetric and they also provided a construction algorithm for the inverse matrix. Since Σ_E^{-1} and $\Sigma_{B_f}^{-1}$ are diagonal, the sum $(\Sigma_E + \Sigma_{B_f})^{-1} + \Sigma_{B_s}^{-1}$ is then tri-diagonal. Symmetric tri-diagonal matrices can be inverted directly with the algorithm of *Usmani* [1994]. The inverse of the covariance matrix of the likelihood function can be expressed as $\Sigma_{E+B}^{-1} = (\Sigma_E + \Sigma_{B_f} + \Sigma_{B_s})^{-1} = (\Sigma_E + \Sigma_{B_f})^{-1} - (\Sigma_E + \Sigma_{B_f})^{-1} (\Sigma_E + \Sigma_{B_f})^{-1} + \Sigma_{B_s}^{-1} (\Sigma_E + \Sigma_{B_f})^{-1}$, so there is no need to invert any generic full matrix to compute equation (11) (note: Σ_{B_s} is special, it is a full, but Markovian covariance matrix).

[42] 2. The realization of $B_s(P)$ in the prediction phase depends only on the last observation step. Thus, prediction samples of $B(P)$ can be simply generated by subsequent draws from the conditional distribution equation (10) starting from the last value in the observation period.

[43] Due to the limited memory of the $B_s(P)$ process, we can apply a simple and robust numerical solution technique. Although $\Sigma_{E+B}^{-1} = (\Sigma_E + \Sigma_{B_f} + \Sigma_{B_s})^{-1}$ is a full matrix due to $E(P) + B_f(P) + B_s(P)$ it does not retain the handy Markov property of $B_s(P)$ anymore. Nevertheless, its elements quickly decay as we get farther from the diagonal. This means that after a certain lag time the likelihoods of specific sections in the residual series ($r = Q_O^r - Q_M^r$) are quasi independent.

[44] Taking a sufficient memory length m (we applied 10 daily steps), we can estimate the likelihood of the residual series in smaller parts. The algorithm requires two instances of the inverted covariance matrix Σ_{E+B}^{-1} , one with the size of m and the other with $m+1$ (called here the inner and outer covariance kernels, respectively). First, the likelihood of the first m residual elements is calculated with the inner kernel according to equation (11). Then, for each following element from the index of $i = m+1$, we estimate the conditional likelihood based on the previous m values:

$$f(r_i | r_{i-1}, \dots, r_{i-m}) = \frac{f(r_i, r_{i-1}, \dots, r_{i-m})}{f(r_{i-1}, \dots, r_{i-m})}. \quad (15)$$

[45] The approximate likelihood of the entire residual series is the product of the likelihood of the first part and the estimated conditional likelihoods of the subsequent residuals:

$$f(r_n, \dots, r_1) \approx f(r_m, \dots, r_1) \prod_{i=m+1}^n f(r_i | r_i, \dots, r_{i-m}). \quad (16)$$

[46] Since the two covariance kernels have different sizes, it is necessary to include the $1/\sqrt{2\pi k}$ term of the normal density function (f) in equations (15) and (16) with $k = m$ and $m+1$ according to the kernel in question.

2.7.3. Markov Chain Monte Carlo Sampling

[47] Posterior parameter distributions are sampled with the traditional Metropolis algorithm [*Gamerman*, 1997]. The width of the jump distribution is tuned during the burn-in period so that the average acceptance rate is between 15 and 40% afterwards [*Gelman et al.*, 1996]. Realizations of the

bias and error processes are generated according to the method of *Reichert and Schuwirth* [2012] for the observation period and with their respective conditional distributions (see equation (10) for the bias) for the prediction phase.

2.8. Comparison of Error Models

[48] To evaluate the performance of the newly developed error model, we tested three other Gaussian error models with the same data and CRRM. Due to its high degree of freedom, the newly developed error model was considered as a generalization of several existing models. To simplify the calculations, the competitors were selected so that they could be simulated by disabling selected parts in the new error model.

[49] The simplest competitor was the traditional normal noise model (E), which—despite its inadequacy—is still used for the description of remnant errors. This was imitated with the new model by switching off all error parameters except σ_E^2 ($\beta = \infty$, everything else = 0). Another traditional selection was the first-order autoregressive bias model (B), which we adapted by setting all parameters except σ_E^2 and β to 0. The most complex alternative was the bias-noise composite (B + E) from *Reichert and Schuwirth* [2012]. The difference in $E = Q_O - Q$ was assumed to be white noise, while $B = Q - Q_M$ was an autoregressive bias process. This was achieved by setting κ_s and κ_f to 0.

[50] Since the selection of an inappropriate error model would introduce a bias in the parameters and result in unreliable uncertainty intervals, it is an important objective during any uncertainty analysis to assess the suitability of the applied error model a posteriori [*Thyer et al.*, 2009]. For entirely frequentist error models (in this case: E), this is indeed feasible in the form of, e.g., a statistical test that checks the probability that the model residuals were realizations of the hypothesized error distributions. However, formal Bayesian error models (in this case: B from B + E and B(P) from B(P) + E) usually combined elements of epistemic uncertainty (due to lack of knowledge) with aleatory uncertainty (due to random behavior of the system described by the model). The epistemic part of uncertainty can only be assessed by its conceptual foundation and the elicitation from knowledgeable experts, and thus it is the purely aleatory part that can be explicitly examined with frequentist tests. Along these lines, we examined the independence and normality of E in the E, B + E, and B(P) + E error models.

3. Results

3.1. Transformation Parameter

[51] The effect of the transformation parameter on the normality of E is illustrated in Figure 4. As λ decreases, the normality of the maximum likelihood realization of E improves for the E and B + E error models, but at the same time the lag 1 autocorrelation increases causing a conflict in the assumptions. The B(P) + E error model did not seem to be influenced significantly as it can simulate heteroscedasticity on its own. Based on this outcome, we fixed λ to 0.5 as a reasonable compromise for the remaining part of the analysis.

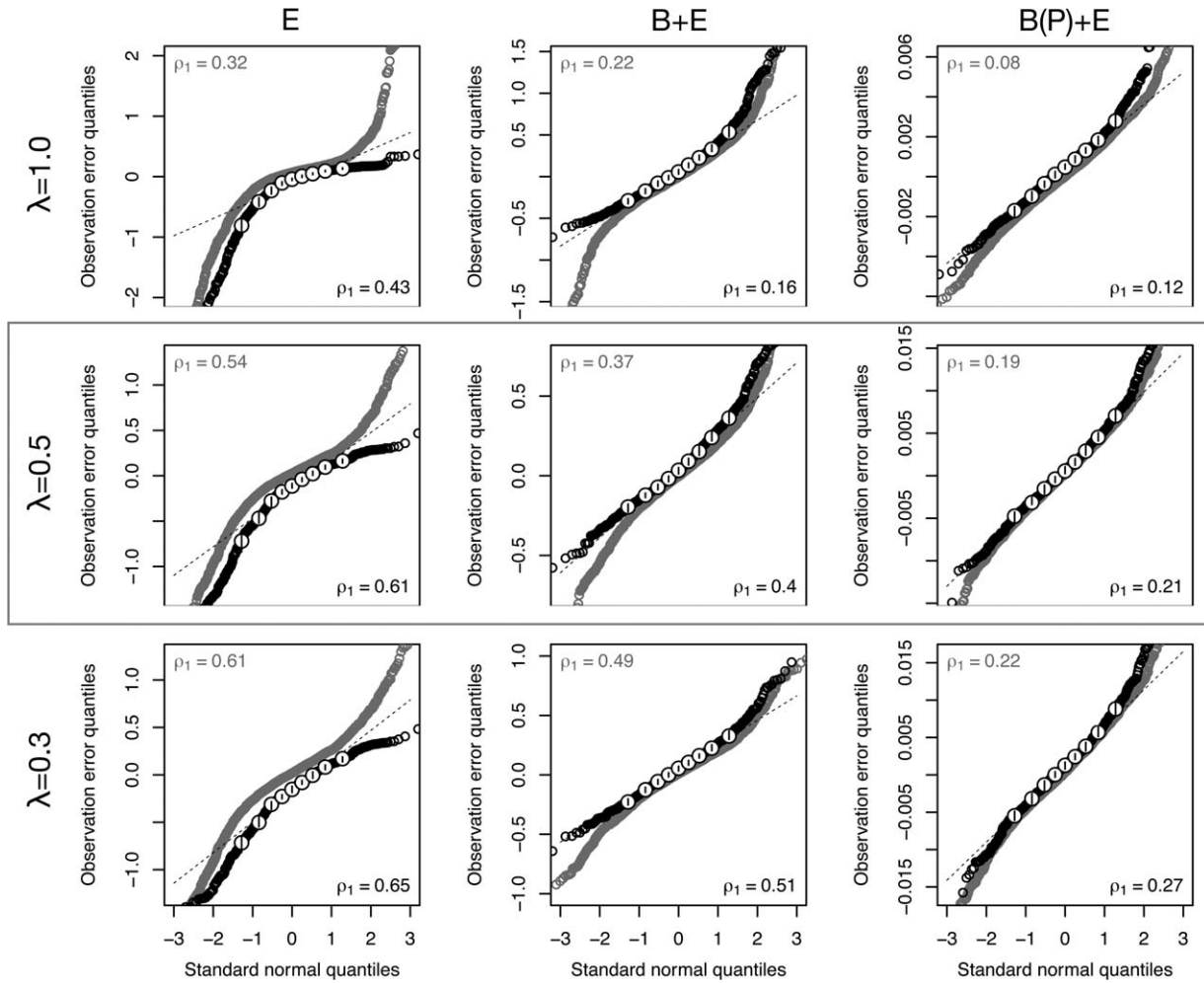


Figure 4. Quantile-quantile plots for the posterior maximum likelihood realizations of E in the calibration (black) and validation (gray) periods. Open white marks indicate the calibration deciles with the small bars showing their 95% confidence interval. ρ_1 is the 1 day lag autocorrelation coefficient.

3.2. Predictive Uncertainty With Different Error Models

[52] Thanks to the local input data, the simple modified logSPM CRRM achieved good performance. In a test calibration with the Nash-Sutcliffe efficiency as the objective function, the highest scores reached $NS = 0.85$. With the involvement of the Box-Cox transformation ($\lambda = 0.5$), the highest NS scores reached 0.94 for the transformed discharge series, but this meant a decrease to 0.81 considering Q without transformation. Interestingly, neither the maximum likelihood solutions nor their NS scores differed significantly for the different error models (Figure 5). The deterministic CRRM overestimated the amount of base flow in all cases (Figure 6), but this was compensated by the error models to a varying degree.

[53] However, the uncertainty intervals differed between the error models. The difference mostly came from the varying rigour of the applied error models (Figure 7). Generally speaking, the independent identically distributed normal error model (E) showed the least tolerance against deviations. According to the independence hypothesis behind it, any exceptionally large deviation between the

model results and the measurements would be exclusively attributed to measurement errors and can be expected to diminish in a single time step. Since this typically does not happen due to the memory of the CRRM, the E error model tended to underestimate the likelihood of parameter sets causing temporary but systematic deviations. The consequences were twofold. First, uncertainty caused by parameter variability was supposed to be very small, which—according to the hypotheses—means that the overall uncertainty of Q and Q_M (considered to be identical) was severely underestimated (Figure 7). The incorrect assumptions on the model residuals were indicated by the statistical properties of the residuals in the calibration period. The distribution of residuals had too heavy tails (Figure 4). There was significant autocorrelation ($\rho_1 = 0.54$) in the residual series, and the estimated measurement errors were orders of magnitude higher (sometimes reaching $\pm 50\%$ of the average discharge depending on the transformation) than the usual noise in automatic discharge measurements (Figure 8). The latter was obvious when one compared predictions of Q_O to observations: the magnitude of independent random fluctuations was too high compared to the

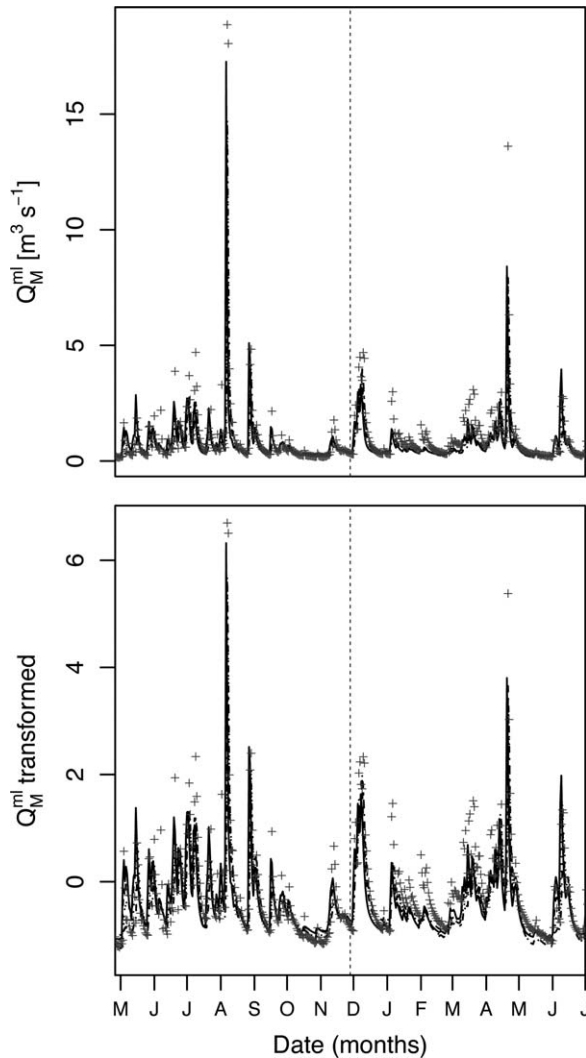


Figure 5. Maximum likelihood solutions (Q_M^{ml}) with different error models in (top) normal and (bottom) transformed space in a 400 day period between 2007 and 2008. Crosses show Q_O , the dashed line separates the calibration and prediction periods. The results from the four different error models cannot be distinguished.

discharge itself, which ruined the slowly changing parts of the hydrograph (Figure 9).

[54] The B error model eliminated some of the problems of the E model. Although it was similarly homoscedastic (Figure 8), the strong autocorrelation of residuals could be properly accounted for. This practically means that it was the innovations of the error process that need to be random, which then resulted in much smaller likelihood penalties for systematic deviations. Contrary to the E error model, the B error model considered the structural problems of the CRRM as the main cause for the deviations between Q_M and Q_O , which resulted in $Q_M = Q$. However, the uncertainty ranges seemed to be too optimistic for high discharges and additionally too pessimistic for low flows based on the validation measurement points (Figure 7).

[55] While the E and B error models possessed a single deviation term specifying $Q_O - Q_M$, the composite structure of the B + E error model, potentially coupled with prior in-

formation on the statistical properties of B and E enabled us to distinguish the systematic errors caused by structural or parameter uncertainty from output errors. The bounds of the predictive uncertainty intervals for the B + E error model closely resembled those of the B model (Figure 7), which was a surprise considering the additional complexity of this error model compared to B. However, there was an important difference. The uncertainty bounds showed only the unconditional variance of the error process, but not how individual predictions evolved in time. The maximum likelihood estimation for ρ_1 of the bias process was 0.61 and 0.88 with the B and the B + E model, respectively. This meant that in a short-term operational prediction the two error models would deliver significantly different results despite their similar unconditional variance (Figure 9). The fulfillment of statistical assumptions was almost satisfactory for the B + E model. The observation error E was approximately normally distributed (Figure 4), but there were issues with its independence in the calibration period: it showed signs of a long, but not too vivid, memory ($\rho_1 = 0.37$). In addition, the amplitude of E (SD $\approx 20\%$ of \bar{Q}) still exceeded the (negligible) scattering of our discharge data set (Figure 8), which then introduced a noticeable zig-zag on the predicted trajectories of Q_O (Figure 9).

[56] The newly developed B(P) + E error model finally produced different uncertainty bounds compared to B and B + E (Figure 8). Due to the input-dependent components, the highest uncertainty was concentrated in the vicinity of intense precipitation events (August 2007 in Figure 10). This potential heteroscedasticity allowed the error model to decrease the error variance in low-flow periods (August 2003 and April 2007 in Figure 10), which indicated that the uncertainty for low flow could have been overestimated by the other error models. The good performance in the verification period showed that the algorithm could reasonably estimate the statistical properties of the actual residual process. This resulted in significantly less outlying observation points during flood events (Figure 7). The fulfillment of statistical assumptions had further improved: the standard deviation of the observation error was ultimately calibrated to a reasonably low level (Figure 8 and Table 3) and the values of E followed the prescribed normal distribution (Figure 4) with better independence ($\rho_1 = 0.20$). These improvements were reflected in the proper smooth shape of predicted recession paths of Q_O (Figure 9). If we had used the original discharge measurements with 10 min resolution, E could have been calibrated to have higher standard deviation as the averaging from 10 min to daily steps would have reduced the variance of any white noise process 144-fold.

[57] Contrary to the high predictive uncertainty of individual events, the predictive uncertainty for an aggregated hydrological indicator like the flow duration curve was much lower for all error models with values below 20% with 95% confidence (Figure 6). The distribution of flow duration curves was generated in predictive mode for the entire period covered by data. The individual flow duration curves belonging to different parameter sets in the MCMC sample were used to estimate the distribution of Q belonging to a specific exceedance probability. Similarly to the individual events, the estimated width of predictive uncertainty was the smallest for the simplest E error model. However, the widest interval was now produced by the

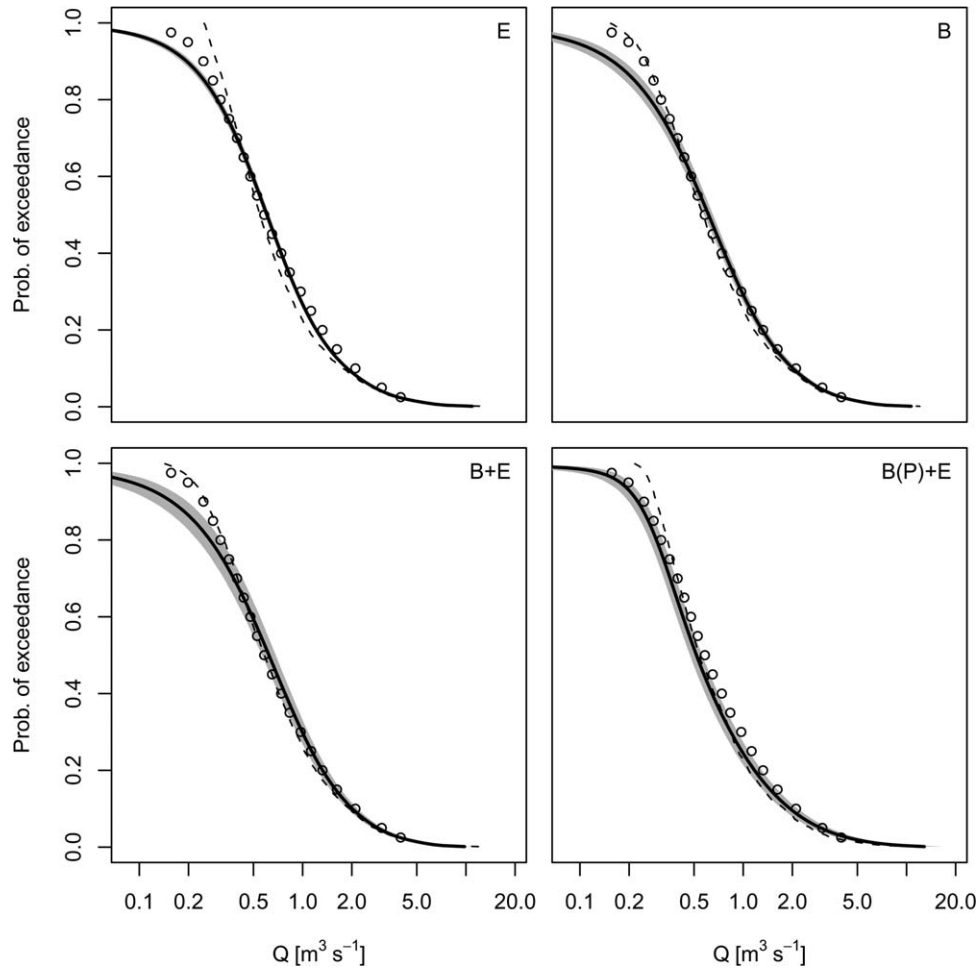


Figure 6. Flow duration curves of the prediction period with different error models. The dashed line stands for Q_M^{ml} , the maximal likelihood solution of the deterministic CRRM, open circles for the quantiles of Q_O (observations), Q_M^{ml} (solid line), and Q (shaded area) are the maximum likelihood solution and the 95% uncertainty interval for the modeled “real” discharge ($Q = Q_M + B_f(P) + B_s(P)$), respectively.

B + E error model and not by the most complex B(P) + E. Although the maximum likelihood solution for Q_M was quite inaccurate for low flow in all cases, the bias process in the B(P) + E error model could almost perfectly compensate for this and produced the closest agreement with the observations.

3.3. Posterior CRRM Parameters

[58] The choice of an error model and the corresponding likelihood function influences the posterior parameter distribution in Bayesian model calibration, because the error models in rainfall-runoff modeling are all statistically imperfect to some degree. The error models taking part in our comparative study feature a wide range of complexity (E has 1 parameter, B(P) + E has 5). The posterior marginal distributions for different CRRM parameters are similarly diverse (Figure 11).

[59] The B and B + E error models produced the most similar posterior marginals thanks to their common stationary AR(1) process kernel. This indicated that the distinct measurement error (the difference between B and B + E) could not fundamentally change the posteriors in

case of our CRRM and data. The E and B(P) + E error models differed frequently from the previous two models. The memoryless E error model was exceptionally selective for sensitive parameters. This was why the magnitude of parameter uncertainty became so low for this error model.

[60] There was a basic difference between the new B(P) + E error model and its counterparts. The new model was the only one that changed its unconditional variance even in transformed space depending on the rainfall situation and history. This was not a completely new property as Yang *et al.* [2007] had already introduced different parameter sets for their B error model in the wet and dry seasons, but it was unique in this selection of error models. While the spectrum from the E through the B to the B + E error models could be regarded as a gradual relaxation of the unrealistic 0 autocorrelation constraint for an otherwise stationary process, the introduction of heteroscedasticity meant a more fundamental change. The B(P) + E error model frequently produced totally different posteriors than the others from the same prior knowledge and measurement data (Figure 11). This demonstrated the conditionality

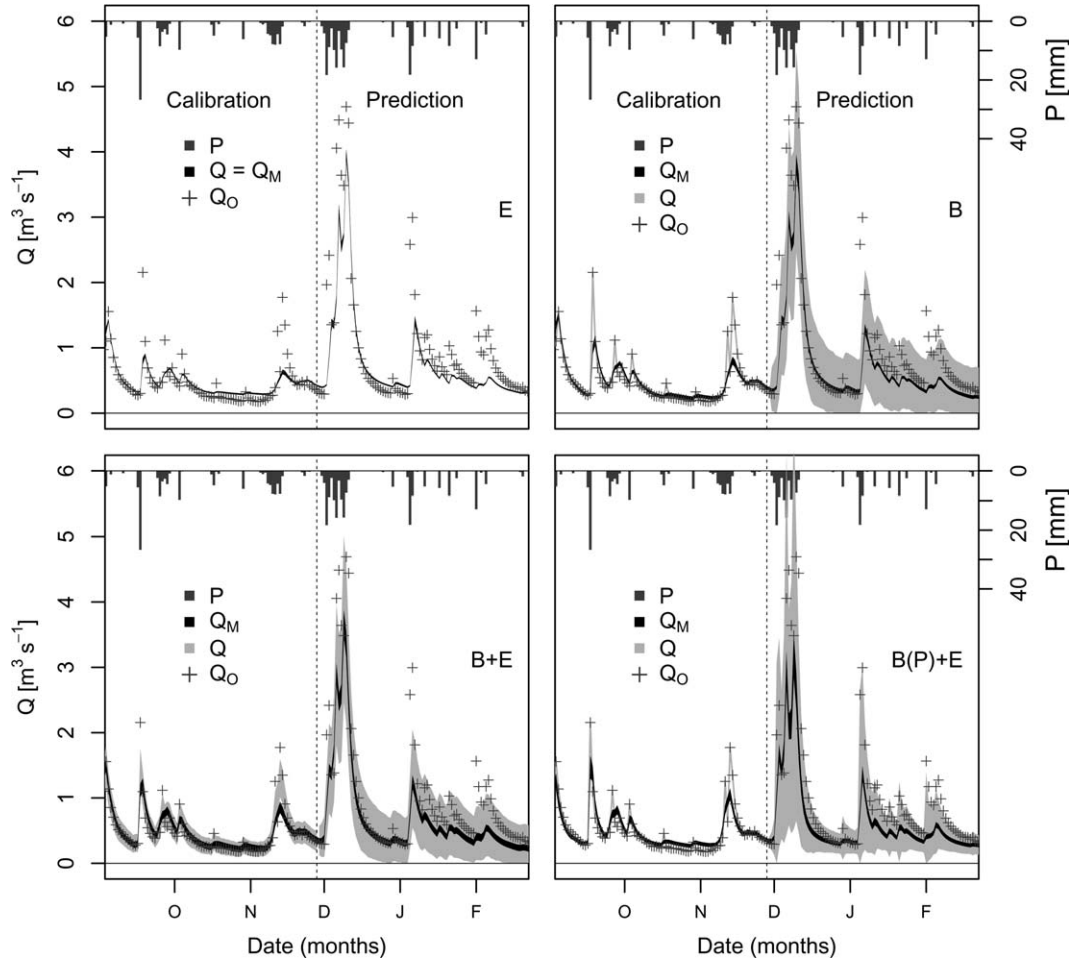


Figure 7. Differences in calibration and predictive uncertainty of the modeled (Q_M) and the real (Q) discharge with different error models in a 160 day interval around the end of the calibration period. Q_O : observations, P : precipitation.

of the posterior parameter distribution on the assumptions of the error model.

[61] The similar maximum likelihood solutions for Q could occur besides different parameter posteriors because the internal storages of the CRRM also differed by the error model (Figure 12). While there was an almost perfect consensus on the size of the snow storage (h_{snow}) for all error models, the mean value of the soil moisture storage (h_s) scattered between 200 and 250 mm. The difference was most expressed for the groundwater storage (h_{gw}) where the E, B, and B+E error models all settled below 5 m, while the B(P)+E model resulted in a stunning storage size of 15 m.

3.4. Posterior Error Parameters

[62] We considered noninformative prior distributions for Ξ_f , Ξ_s , and Ψ . Similarly to the CRRM parameters, the maximum likelihood values and the posterior distributions varied by the specific error model (Table 3). The decrease of σ_E and σ_B from B+E to B(P)+E showed that the introduction of precipitation-dependence created a valuable opportunity for the B(P)+E model to decrease the stationary uncertainty when the water fluxes between the CRRM compartments were less intense. The stationary

variance of B(P)+E in dry periods approximately halved in comparison to the other error models. The practical extinction of σ_E was especially a big improvement, since it was finally in accordance with the truly negligible stationary scattering of our Q measurements. This resulted in realistically predicted hydrographs for Q_O in each single realization (Figure 9).

[63] The values of κ_s and κ_f represented input uncertainty and the structural uncertainty in runoff formation. Their values suggested that these uncertainties affected both the fast and slow reacting components of the CRRM in a roughly equal way. This hypothesis was consistent with the posterior marginal of the k_{rge} parameter, which assigned similar amounts to interflow and recharge. The magnitudes of κ_s and κ_f indicated that a precipitation event of a mere 2 mm could practically double the stationary uncertainty on the very day of the event and keep it 35% higher for the next day.

[64] A rough and naïve comparison of κ_s and κ_f with the actual storm-specific runoff coefficients demonstrated the amount of input-related uncertainty in this error model. The total input-related uncertainty came from the interplay between true input uncertainty and uncertainty of runoff formation (see equation (3)), so we had to estimate the

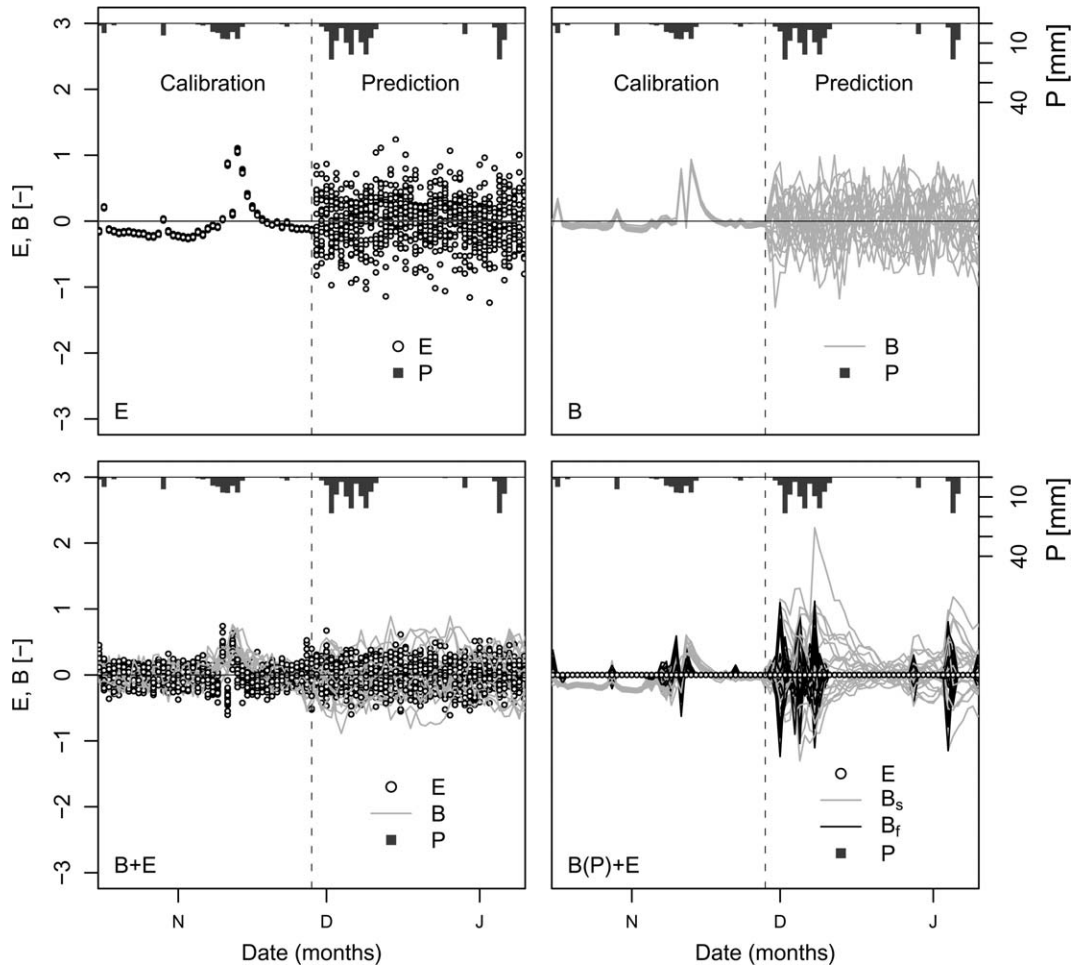


Figure 8. Sample realizations of error components (20 realizations in transformed space) during calibration and prediction with different error models. The figure covers a 80 day interval around the end of the calibration period.

runoff coefficient in order to get the input uncertainty. According to the maximum-likelihood parameter values of κ_f and κ_s , the total input-related uncertainty of Q (mm d^{-1}) reached 13% of the actual daily precipitation (mm d^{-1}) on average. The observed runoff yield was 0.53 for the biggest flood event in August 2007. This meant that the standard deviation of the rainfall multiplier σ_{rm} should have been around 22% in this case. This figure seemed to be in the realistic range for input uncertainty, but rather it was a minimal limit of input uncertainty as most smaller flood events possessed a much lower runoff yield.

4. Discussion

4.1. Linearization of CRRM Response to Input Uncertainty

[65] Our assumptions about the linear propagation of input errors through the CRRM corresponded to the structure of the “abc” rainfall-runoff model [Fiering, 1967]. This extremely simple linear CRRM was primarily important for education, but its simplicity allows the direct inspection of some otherwise complicated details. This aspect made it popular for system diagnostic purposes [Kuc-

zera, 1982; Vogel and Sankarasubramanian, 2003; Huard and Mailhot, 2006]. The whole catchment was simulated by a single storage S . The rainfall P was routed according to fixed proportions. The parameter a determined how much of P can enter the storage, while b described the proportion of P immediately lost to evapotranspiration. The remaining c specified the rate of base flow from S . Thus,

$$\frac{dS}{dt} = a \cdot P - c \cdot S \quad \text{and} \quad Q = (1 - a - b) \cdot P + c \cdot S. \quad (17)$$

[66] This was similar to our assumption that a difference in P caused an immediate and a lasting effect in Q . Considering a precipitation multiplier P/P_M with a mean of 1 and variance of σ_{pm} , we could analytically express the true values for κ_s and κ_f for the B(P) + E error model based on equation (4):

$$\kappa_f = \sigma_{pm} \cdot (1 - a - b) \quad \text{and} \quad \kappa_s = \sigma_{pm} \cdot a \cdot c \quad \text{and} \quad \beta = \log(1 - c). \quad (18)$$

[67] In this sense, the B(P) + E error model emulated the input error propagation through the nonlinear CRRM using

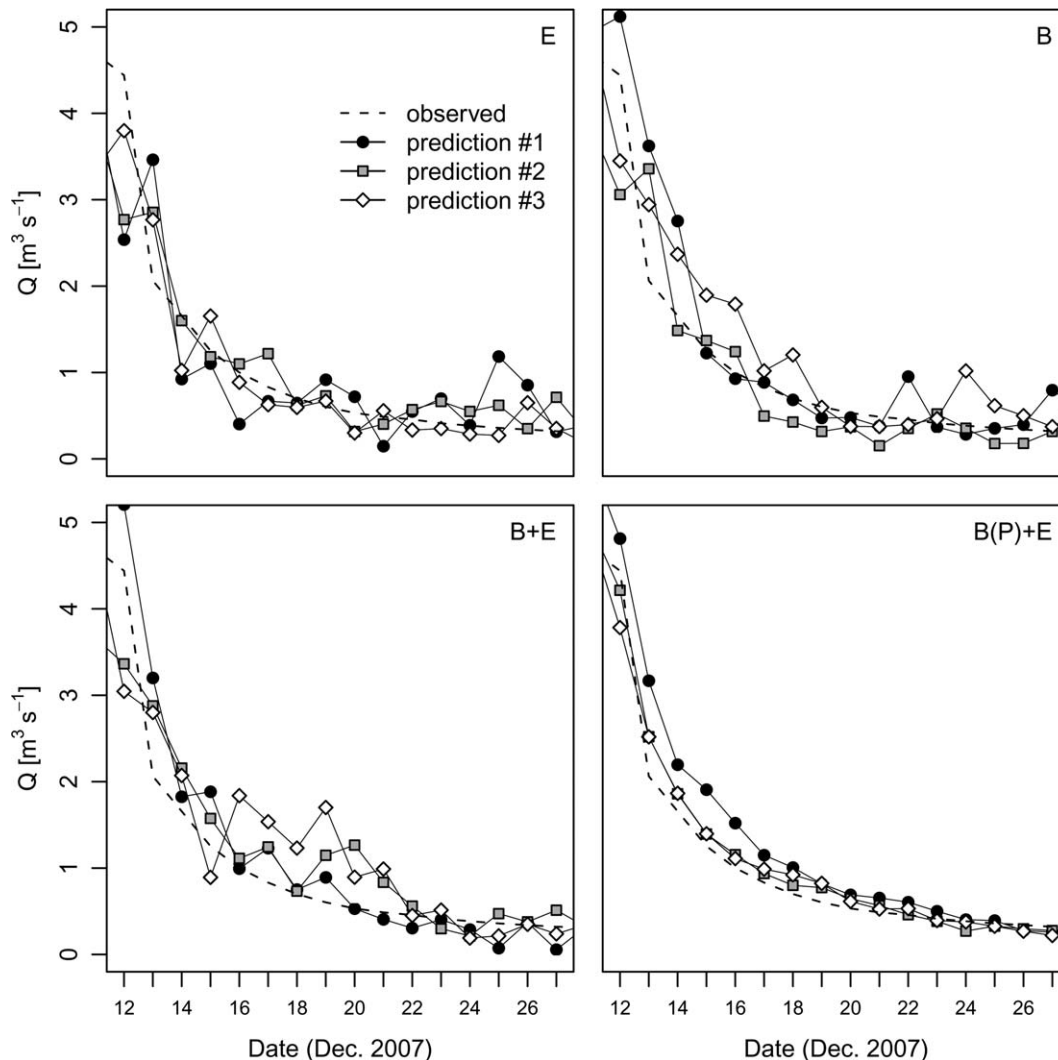


Figure 9. Comparison of observed flow during a single recession period with selected realizations of predicted Q_O for different error models.

the blueprint of the “abc” hydrological model. This implied that the error would be overestimated during low flows and underestimated during high flows. The linear error propagation resulted in heteroscedastic predictive uncertainty bands, where the unconditional variance was mainly controlled by the rainfall series. While it was a simplification compared to uncertainty assessment methods relying on true error propagation, it was a step forward from purely statistical approaches. This feature was similar to the results by Götzinger and Bárdossy [2008], although our study differed in some details. The trajectories of discharge predictions in our study were smooth in the recession phases thanks to the autoregressive bias process (B(P)+E in Figure 9), while their error varied independently around the deterministic model predictions with the given variance (similarly to E in Figure 9).

4.2. Quantification of Input Uncertainty

[68] Several studies attempted to examine different sources of uncertainty by accounting for them in the Bayesian inference procedure. From the posterior parameter distributions of the error model or the manifestation of stochastic time-

variable parameters, conclusions could be drawn about the relative importance of different sources of uncertainty [see, e.g., Kuczera et al., 2006; Vrugt et al., 2008; Reichert and Mieleitner, 2009]. A common finding in these studies was that precipitation had the strongest impact on the model output and consequently input uncertainty was declared to dominate.

[69] Mantovan and Todini [2006] warned that in Bayesian calibration theory the posterior parameter distribution is conditional on the deterministic model structure and the error model. Thus, in an unconditional sense, parameters are simple mathematical utilities to adjust the output of the imperfect deterministic model, regardless of their physical or other meaning in the model. Consequently, Bayesian calibration does not guarantee that we get generally valid parameter distributions by inferring the parameters from an observation series. The exclusive objective of the inference procedure is to find a posterior distribution for the parameter set so that the conditional dependence of the predictions on the parameters can be marginalized out. Doherty and Christensen [2011] demonstrated that the inevitable difference between the model structure and reality introduces a certain bias to the parameter estimates. Strictly speaking,

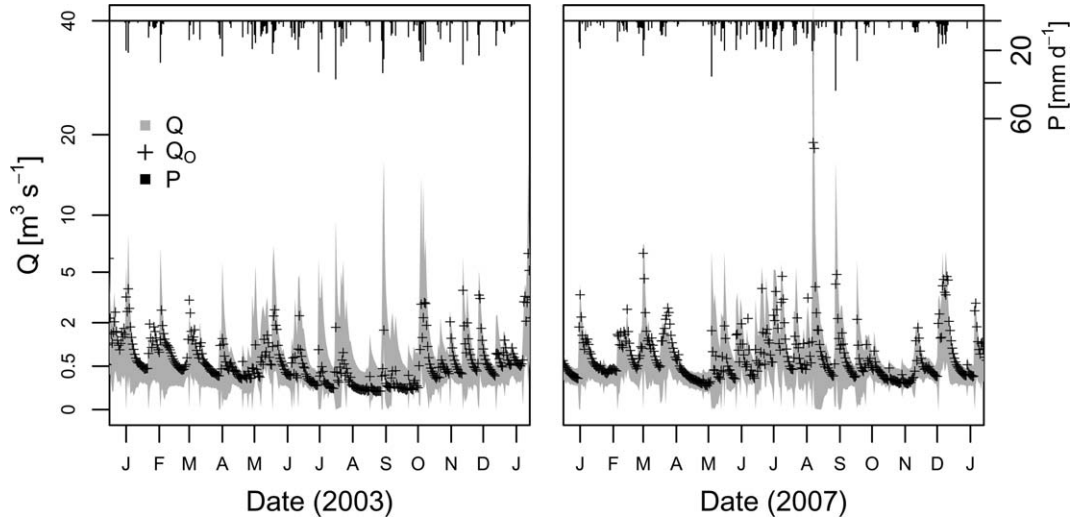


Figure 10. Modeled and observed discharge in 2003 and 2007 (exceptionally dry and wet years, respectively) with the B(P) + E error model. The uncertainty intervals were generated in prediction mode.

the only unconditional products of the inference procedure are the distributions of model predictions.

[70] In line with the observation of *Beven* [1989] about the information content of discharge series, *Kirchner* [2009] showed that quite simple models can be fed with a synthetic rainfall series so that they produce a close match to the observed discharge. The radically simple storage-discharge function applied in the backward inference of precipitation did not even attempt to describe the common hydrological processes featured in most CRRMs [*Kirchner*, 2009] and thus it could be declared to be a warehouse of structural errors. However, at the same time most of the uncertainty could be assigned to the input in a hypothetical experiment as the adjustment of rainfall would have been able to create a perfect match between the model output and the observations. It seems that the importance of precipitation is so huge for CRRMs that it might compensate for other sources of uncertainty.

[71] It was difficult to figure out the importance of input uncertainty on the present results due to several reasons. Even if we limited ourselves to statements conditional on the applied CRRM and error model and neglect the power transformation of Q , the inherent interaction between input and model response happening inside equation (4) would still have created an identification problem. Simply looking at the units, κP could have given the standard deviation of precipitation specific runoff error (mm mm^{-1}), but actually it was the product of the input uncertainty and the sensitivity of the discharge response to it equation (3). Consequently, we did not meaningfully compare κ_s and κ_f to traditional rainfall multipliers, which could have been directly interpreted as relative measurement inaccuracy.

4.3. Absolute Levels of Uncertainty

[72] The discovery that traditional error models seriously underestimated the predictive uncertainty in real-world applications was a main motivation for the development of more sophisticated uncertainty assessment techniques [*Kennedy and O'Hagan*, 2001]. This phenomenon was typical

in environmental modeling, but it was especially expressed in hydrology due to usual rainfall-runoff model residuals exhibiting a complex statistical behavior. With the development of a more realistic error model, one could have expected to get a more accurate view on the actual level of predictive uncertainty, but it was a surprising recognition what levels this uncertainty could have actually reached.

[73] The test case utilized a locally recorded input data set coupled with a simple yet efficient CRRM. The maximal likelihood model results for discharge were very close to the measurements, but even in this relatively simple and well-monitored catchment the 95% relative predictive uncertainty interval could reach ± 50 to 150% for low flows with any error model. For flood peaks, the picture was more diverse. The E, B, and B + E error models predicted smaller relative uncertainty, which logically came from their assumptions on stationary errors.

[74] These surprising levels of uncertainty for individual events seemed very high compared to the fact that the model achieved a very good fit in the calibration period. The general cause was that the occasional low likelihood events trained the error model to believe that the CRRM could regularly miss the target discharge with over 100% error. This widened the bounds of unconditional uncertainty. The specific cause could have been the fact that the statistically more realistic error models relied on less strict yet invalid assumptions (e.g., that relative errors are the same for floods and base flow), so they were more tolerant to heavy-tailed error distributions [*Vrugt et al.*, 2008].

Table 3. Maximum Likelihood Values for Error Parameters

Error Model	Error Parameter				
	σ_E	σ_B	β	κ_f	κ_s
E	0.368				
B		0.385	0.491		
B + E	0.211	0.337	0.127		
B(P) + E	0.004	0.140	0.167	0.038	0.030

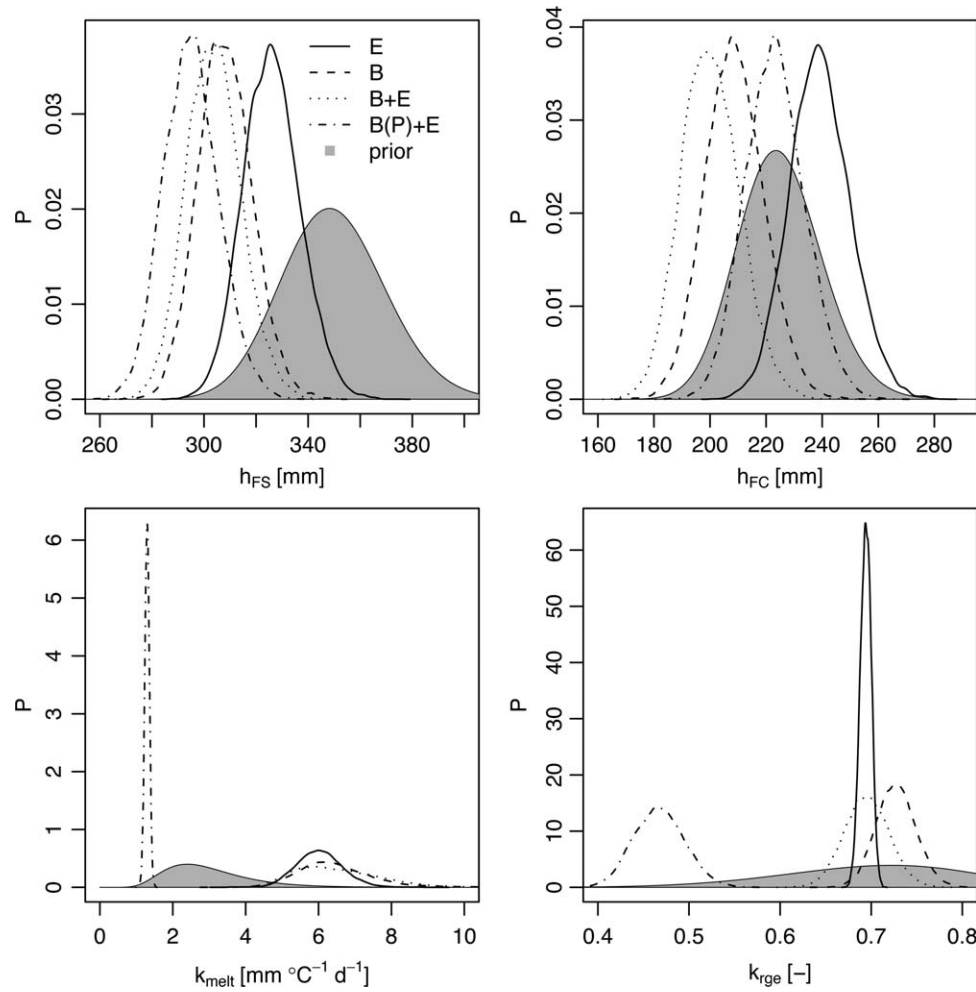


Figure 11. Prior and posterior marginals for some selected CRRM parameters with different error models. For a description of parameters, please refer to Table 2.

[75] The predictive uncertainty for flow duration curves was typically around $\pm 10\%$. This suggested that the CRRM was indeed a good representation of catchment behavior in the case study. The prevalence of input uncertainty prevented the CRRM from making precise predictions on individual hydrological events like a flood peak or a low-flow period, but the more robust performance indicator proved its predictive capability.

4.4. Comparison to a Non-Gaussian Model of Total Error

[76] To assess the performance of $B(P) + E$ in describing total predictive uncertainty, we did not only compare it to its subsets but also to an independent statistical model of total error. The frequentist generalized likelihood function (GL) model of *Schoups and Vrugt* [2010] was developed to

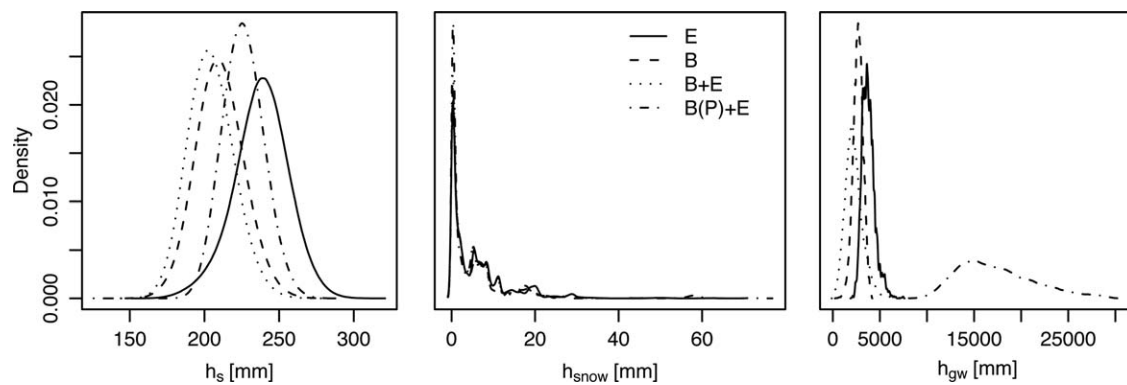


Figure 12. Storage size densities with different error models. h_s is the soil moisture storage, h_{snow} is the snow height in water equivalent, and h_{gw} is the groundwater storage.

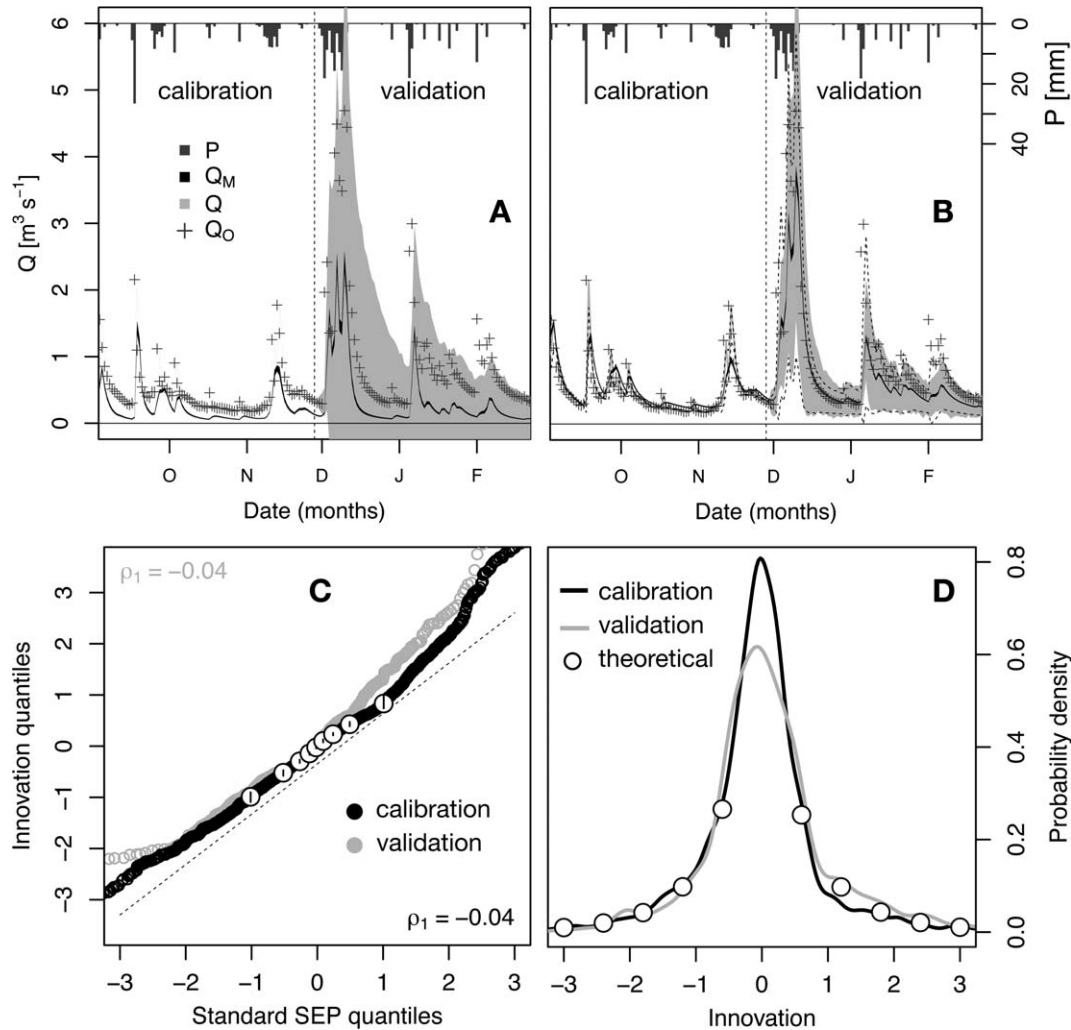


Figure 13. Calibration and predictive uncertainty with two versions of the generalized likelihood function (GL) of Schoups and Vrugt [2010]. (a) Calibrating all error parameters; (b) fixing ϕ_1 at 0.7, dashed lines indicate the uncertainty interval of $B(P) + E$. (bottom) Posterior maximum likelihood realizations of error innovations for the version shown in Figure 13b. (c) Q-Q plot and (d) empirical probability density functions.

handle all kinds of statistical problems that usually result in conflicts between the modeler's typical assumptions and the true properties of the residuals (autoregression, heteroscedasticity, heavy tails).

[77] The GL error model is essentially an autoregressive error process with non-Gaussian innovations (innovations have a skew exponential power (SEP) distribution). It has five parameters plus a bias correction factor.

[78] Despite its apparent conceptual versatility the successful application of the GL error model appears to depend on the suitability of the CRRM to the specific catchment. Schoups and Vrugt [2010] reported test applications for a wet and an arid catchment with mixed results. The wet catchment was well described by their CRRM and the error parameters were inferred without difficulty. However, the CRRM did not perform so well for the arid catchment and this implied several problems in calibrating the parameters of the GL error model. It was found that the maximum likelihood solution for Y was significantly bi-

ased and predictive uncertainty was extremely and unreasonably wide. To overcome these problems the ultimate remedy was to fix the lag 1 autocorrelation coefficient of the error process arbitrarily to a moderate value ($\phi_1 = 0.4$).

[79] Interestingly, we got similar experiences for our test catchment. The application of the GL error model without constraints on its parameters resulted in extremely wide predictive uncertainty intervals (Figure 13a) and biased Y values. Just like in the arid case study of Schoups and Vrugt [2010], the reason was the very high value of ϕ_1 (maximum likelihood value: 0.93). The unrealistic amount of uncertainty could only be reduced by fixing ϕ_1 arbitrarily to 0.7 based on the autocorrelation range specified by the other error models. This helped to narrow down the uncertainty intervals to approximately the level of the $B(P) + E$ error model (Figure 13b) and eliminated the bias of Y .

[80] While the GL error model seems to be advantageous compared to the $B(P) + E$ model because it is based on more general statistical assumptions, this versatility comes

at a price. The wide range of possible shapes of the SEP distribution appears to be the reason why the inference procedure may result in maximum likelihood for biased results and extremely wide predictive uncertainty. This behavior would require more in-depth analysis that is beyond the scope of this paper.

4.5. Perspectives in Statistical Error Modeling

[81] The purely statistical modeling of model errors provided a relatively simple way of uncertainty assessment. However, this also meant taking some compromises. Although the information needed to make a statistical error model can be sometimes much less compared to mechanistic error propagation frameworks such as BATEA [Kavetski *et al.*, 2006], DREAM [Vrugt *et al.*, 2008], or IBUNE [Ajami *et al.*, 2007], this also meant that the statistical error description might miss some crucial features of the error process.

[82] It is a common critique against formal error models that the traditional (“dumb”) statistical error models (E and B) eventually produced discharge uncertainty bands that do not make any physical sense by dipping below 0 or that the actual discharge may follow a path that actually cannot be described by the deterministic hydrological model [Beven *et al.*, 2008].

[83] While this is indeed a drawback, it is not a proper justification for abandoning the sound theoretical foundations of inference. Traditional methods often build on indefensible assumptions (like the E error model) or neglect axioms of inference (like GLUE). As a result, they may hide the majority of existing uncertainty from the analyst. We think that statistical error models are still worth further research to overcome most of their existing limitations to a point where they can provide simpler alternatives to the Bayesian mechanistic error propagation frameworks.

5. Conclusions

[84] In this study, we developed a formal statistical error model that can represent the effects of all important uncertainty sources (including input uncertainty) on model output that occur in conceptual rainfall-runoff modeling. The two main objectives were (i) to narrow the gap between the fast yet unsatisfactory traditional Gaussian error models and the accurate yet computationally demanding mechanistic error propagating methods and (ii) to account for the remaining bias of these methods. The composite bias + noise statistical error model from Kennedy and O’Hagan [2001], Bayarri *et al.* [2007], and Reichert and Schuwirth [2012] was extended with the linearized propagation of input uncertainty equivalent to the “abc” rainfall-runoff model. The development introduced intrinsic heteroscedasticity into the error model dependent on the amplitude of the most important driver: precipitation. The new error model was tested on data from the Mönchaltorfer Aa catchment (Switzerland) and compared with altogether four other statistical error models. Based on the results we concluded that:

[85] 1. The involvement of input uncertainty significantly improved the agreement between the statistical properties of posterior residuals and the assumptions of the error model. The maximum likelihood predictions of discharge were similar for all error models, but the underlying deterministic model parameters and the predictive uncertainty intervals were rather different.

[86] 2. Besides the statistical improvements, the newly developed error model showed additional refinements: it properly assigned a lower importance to the observation noise, which allowed for smooth recession patterns in predicted discharge observations. Furthermore, it delivered what appeared to be a realistic estimate of predictive uncertainty with different bandwidths for flood and recession periods.

[87] 3. Despite the good performance of the deterministic CRRM on the test data set, the predictive uncertainty of individual flood events in the validation period was very high reaching 100% relative error. The input and runoff uncertainty seemed to be a major cause for this low confidence, but its quantified contribution is certainly conditional on the assumptions about the structure and propagation of errors.

[88] 4. In contrast to the single events, the overall flow regime was simulated with high confidence (about 10% relative error in the flow duration curve), which indicated that the CRRM managed to capture the most important aspects of the local hydrology.

[89] 5. While complex statistical error models could not provide insight into the reasons of errors or possible structural improvements to reduce the model bias, they still remained to be computationally cheap alternatives to full Bayesian error propagation frameworks in the theoretically sound assessment of total predictive uncertainty. Additionally, they could be used in full error propagation frameworks to provide a statistical description of the remaining bias.

Appendix A: Statistical Properties of the Disturbed Ornstein-Uhlenbeck Process

A1. A Standard Ornstein-Uhlenbeck Process With Zero Mean

[90] The Ornstein-Uhlenbeck (OU) or Gauss-Markov process is a mean-reverting Gaussian process. For 0 expected value, the process B is defined with its unconditional variance σ_B^2 and the inverse correlation length β :

$$dB(t) = -\beta B(t) + \sqrt{2\beta\sigma_B^2} dW(t), \quad (A1)$$

where $W(t)$ is the Wiener process at time t .

[91] This equation has an analytical solution for the conditional distribution of B . The conditional mean is

$$E[B(t_i)|B(t_{i-1}) = b] = b \exp\{-\beta|t_i - t_{i-1}|\}. \quad (A2)$$

[92] According to the definition of the process in equation (A1), the random component is independent of the process itself. This means that the variance can be described separately from the actual process value:

$$\text{Var}[B(t_i)|B(t_{i-1}) = b] = \sigma_B^2(1 - \exp\{-2\beta|t_i - t_{i-1}|\}). \quad (A3)$$

[93] Similarly, the covariance between earlier and later values of the process depends only on the time difference (if the variance in $t_i - 1$ is σ_B^2):

$$\text{Cov}[B(t_i), B(t_{i-1})] = \sigma_B^2 \exp\{-\beta|t_i - t_{i-1}|\} \quad (A4)$$

and in this case the covariance matrix for discrete observations becomes

$$\Sigma_B(i, j) = \sigma_B^2 \exp \{ -\beta |t_i - t_j| \}. \quad (\text{A5})$$

A2. Asymptotic Variance Kinetics of the Standard Ornstein-Uhlenbeck Process

[94] The stationary variance of the process is the unconditional variance σ_B^2 . The process variance will converge to the stationary variance regardless of the initial state.

[95] If the distribution in t_{i-1} is Gaussian with $\varsigma_B^2(t_{i-1})$ then in t_i its variance will be

$$\varsigma_B^2(t_i) = \sigma_B^2 + (\varsigma_B^2(t_{i-1}) - \sigma_B^2) \exp \{ -2\beta(t_i - t_{i-1}) \}. \quad (\text{A6})$$

[96] To get the conditional variance, we can simply substitute $\varsigma_B^2(t_{i-1}) = 0$ to get the solution equivalent to equation (A3):

$$\varsigma_B^2(t_i) = \sigma_B^2 (1 - \exp \{ -2\beta(t_i - t_{i-1}) \}). \quad (\text{A7})$$

[97] The kinetics of the asymptotic variance can be described by taking the derivative of equation (A6):

$$\frac{d\varsigma_B^2}{dt} = 2\beta(\sigma_B^2 - \varsigma_B^2). \quad (\text{A8})$$

[98] The solution of this differential equation indeed satisfies the conditional variance equation (A6) if we use $\varsigma_B^2 = \sigma_B^2$ if $t_i - t_{i-1} = \infty$ as the boundary condition during integration.

A3. The Disturbed Ornstein-Uhlenbeck Process

[99] We assume that precipitation increases the unconditional variance to $[\sigma_B^2 + (\kappa P)^2]$. Since β remains unaffected, the asymptotic variance kinetics from equation (A8) still applies:

$$\frac{d\varsigma_B^2}{dt} = 2\beta([\sigma_B^2 + (\kappa P)^2] - \varsigma_B^2). \quad (\text{A9})$$

[100] The conditional distribution of $B(t_i)$ given the initial value $B(t_0)$ is then

$$B(t_i) = B(t_0) \exp(-\beta |t_i - t_0|) + \int_{t_0}^{t_i} \varsigma_B(s) \exp(-\beta |t_i - s|) dW(s). \quad (\text{A10})$$

[101] When the P precipitation rate is uniform within a time step Δt , then the actual process variance at the end of the time step becomes

$$\varsigma_B^2(t + \Delta t) = \sigma_B^2 + (\kappa P)^2 + (\varsigma_B^2(t) - \sigma_B^2 - (\kappa P)^2) \exp \{ -2\beta \Delta t \}. \quad (\text{A11})$$

[102] This can be rearranged for the equidistant discrete case into a form that resembles equation (A6):

$$\varsigma_{B_i}^2 = \sigma_B^2 + (\varsigma_{B_{i-1}}^2 - \sigma_B^2) \exp \{ -2\beta \Delta t \} + (\kappa P)^2 \quad (\text{A12})$$

with

$$\kappa = \kappa \sqrt{1 - \exp \{ -2\beta \Delta t \}}. \quad (\text{A13})$$

[103] This way we can separate the effects of precipitation on the actual process variance from the mean-reverting mechanism that is also present in a standard OU process.

[104] If we consider the asymptotic process variance of a standard OU process in t_i and t_{i-1} , the increase caused by the introduction of the uniform P precipitation rate between the observation points can be regarded as if an independent standard normal random number $Z(t_i)$ multiplied with κP has been added to the standard OU process. Due to its independence, Z does not change the covariance between the two subsequent observation points:

$$\begin{aligned} \text{Cov}[B(t_{i-1}), B(t_i) + \kappa P Z(t_i)] &= \text{Cov}[B(t_{i-1}), B(t_i)] \\ &+ \kappa P \text{Cov}[B(t_{i-1}), Z(t_i)] \\ &= \text{Cov}[B(t_{i-1}), B(t_i)]. \end{aligned} \quad (\text{A14})$$

[105] For nonadjacent observation points that encompass several precipitation events, the overall effect is similar. The disturbance caused by the past precipitation appears as an integral of independent normal noise terms with weights that decay with the time lag (see equation (A10)). Due to the independence of dW from B , we find that

$$\begin{aligned} \text{Cov}[B(t_i), B(t_0)] &= \text{Cov}[B(t_0) \exp \{ -\beta |t_i - t_0| \}, B(t_0)] \\ &= \text{Var}[B(t_0)] \exp \{ -\beta |t_i - t_0| \} \\ &= \varsigma_B^2(t_0) \exp \{ -\beta |t_i - t_0| \} \end{aligned} \quad (\text{A15})$$

[106] If we generalize this for any discrete observation points, we get the following covariance matrix:

$$\Sigma_B(i, j) = \begin{cases} \varsigma_B^2(t_i) & \text{if } i = j \\ \varsigma_B^2(t_i) \exp(-\beta |t_j - t_i|) & \text{if } i < j. \\ \varsigma_B^2(t_j) \exp(-\beta |t_i - t_j|) & \text{if } i > j \end{cases} \quad (\text{A16})$$

[107] This highlights that the disturbed OU process is still a Gauss-Markov process since it has the same decay pattern in covariance as its standard version.

[108] **Acknowledgments.** This study was part of the iWaQa project financed by the Swiss National Science Foundation (National Research Program 61 on Sustainable Water Management, grant 406140-125866) and the Swiss Federal Office for the Environment. We thank two anonymous reviewers for their valuable suggestions and Michael Exner-Kittridge for his support to improve the language.

References

- Abdulla, F., D. Lettenmaier, and X. Liang (1999), Estimation of the ARNO model baseflow parameters using daily streamflow data, *J. Hydrol.*, 222(1–4), 37–54, doi:10.1016/S0022-1694(99)00096-7.
- Ajami, N. K., Q. Duan, and S. Sorooshian (2007), An integrated hydrologic Bayesian multimodel combination framework: Confronting input, parameter and model structural uncertainty in hydrologic prediction, *Water Resour. Res.*, 43, W01403, doi:10.1029/2005WR004745.
- Amt für Abfall, Wasser, Energie und Luft der Baudirektion des Kanton Zurich (AWEL) (2010), Water Quantity and Quality Monitoring Database of Canton Zurich, Zurich, Switzerland.
- Bates, B. C., and E. P. Campbell (2001), A Markov chain Monte Carlo scheme for parameter estimation and inference in conceptual rainfall-runoff modeling, *Water Resour. Res.*, 37(4), 937–947, doi:10.1029/2000WR900363.
- Bayarri, M. J., J. O. Berger, R. Paulo, J. Sacks, J. A. Cafeo, J. Cavendish, C.-H. Lin, and J. Tu (2007), A framework for validation of computer models, *Technometrics*, 49(2), 138–154.

- Beven, K. J. (1989), Changing ideas in hydrology – The case of physically-based models, *J. Hydrol.*, 105, 157–172.
- Beven, K. J., P. J. Smith, and J. E. Freer (2008), So just why would a modeler choose to be incoherent?, *J. Hydrol.*, 354(1–4), 15–32.
- Box, G. E. P., and D. R. Cox (1964), An analysis of transformations, *J. R. Stat. Soc., Ser. B*, 26(2), 211–252.
- Butler, D., and J. W. Davies (2000), *Urban Drainage*, Taylor and Francis, London.
- Clark, M. P., and D. Kavetski (2009), Ancient numerical demons of conceptual hydrological modeling: 1. Fidelity and efficiency of time stepping schemes, *Water Resour. Res.*, 46, W10510, doi:10.1029/2009WR008894.
- Demaria, E. M., B. Nijssen, and T. Wagener (2007), Monte Carlo sensitivity analysis of land surface parameters using the variable infiltration capacity model, *J. Geophys. Res.*, 112, D11113, doi:10.1029/2006JD007534.
- Doherty, J., and S. Christensen (2011), Use of paired simple and complex models to reduce predictive bias and quantify uncertainty, *Water Resour. Res.*, 47, W12534, doi:10.1029/2011WR010763.
- Duan, Q., N. K. Ajami, X. Gao, and S. Sorooshian (2007), Multi-model ensemble hydrologic prediction using Bayesian model averaging, *Adv. Water Resour.*, 30(5), 1371–1386, doi:10.1016/j.advwatres.2006.11.014.
- Fiering, M. B. (1967), *Streamflow Synthesis*, Harvard Univ. Press, Cambridge, Mass.
- Frey, M. P., C. Stamm, M. K. Schneider, and P. Reichert (2011), Using discharge data to reduce structural deficits in a hydrological model with Bayesian inference approach and the implications for the prediction of critical source areas, *Water Resour. Res.*, 47, W12529, doi:10.1029/2011WR009993.
- Gamerman, D. (1997), *Markov Chain Monte Carlo – Statistical Simulation for Bayesian Inference*, Chapman and Hall, London.
- Gelman, A., G. O. Roberts, and W. R. Gilks (1996), Efficient Metropolis jumping rules, *Bayesian Stat.*, 5, 599–607.
- Götzinger, J., and A. Bárdossy (2008), Generic error model for calibration and uncertainty estimation of hydrological models, *Water Resour. Res.*, 44, W00B07, doi:10.1029/2007WR006691.
- Hindmarsh, A. C. (1983), ODEPACK, a systematized collection of ODE solvers, *IMACS Trans. Sci. Comput.*, 1, 55–64.
- Huard, D., and A. Mailhot (2006), A Bayesian perspective on input uncertainty in model calibration: Application to hydrological model “abc,” *Water Resour. Res.*, 42, W07416, doi:10.1029/2005WR004661.
- Kavetski, D., and G. Kuczera (2007), Model smoothing strategies to remove microscale discontinuities and spurious secondary optima in objective functions in hydrological calibration, *Water Resour. Res.*, 43, W03411, doi:10.1029/2006WR005195.
- Kavetski, D., G. Kuczera, and S. W. Franks (2003), Semidistributed hydrological modeling: A ‘saturation path’ perspective on TOPMODEL and VIC, *Water Resour. Res.*, 39(9), 1246, doi:10.1029/2003WR002122.
- Kavetski, D., G. Kuczera, and S. Franks (2006), Bayesian analysis of input uncertainty in hydrological modeling: 1. Theory, *Water Resour. Res.*, 42, W03407, doi:10.1029/2005WR004368.
- Kennedy, M. C., and A. O’Hagan (2001), Bayesian calibration of computer models, *J. R. Stat. Soc. B*, 63(3), 425–464.
- Kirchner, J. W. (2009), Catchments as simple dynamical systems: Catchment characterization, rainfall-runoff modeling, and doing hydrology backward, *Water Resour. Res.*, 45, W02429, doi:10.1029/2008WR006912.
- Kuczera, G. (1982), On the relationship between the reliability of parameter estimates and hydrologic time series data used in calibration, *Water Resour. Res.*, 18, 146–154.
- 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(1–2), 161–177.
- Lazzarotto, P., C. Stamm, V. Prasuhn, and H. Fluhler (2006), A parsimonious soil-type based rainfall-runoff model simultaneously tested in four small agricultural catchments, *J. Hydrol.*, 321(1–4), 21–38, doi:10.1016/j.jhydrol.2005.07.038.
- Mantovan, P., and E. Todini (2006), Hydrological forecasting uncertainty assessment: Incoherence of the GLUE methodology, *J. Hydrol.*, 330, 368–381.
- Martinez, J., and A. Rango (1981), Areal distribution of snow water equivalent evaluated by snow cover monitoring, *Water Resour. Res.*, 17, 1480–1488.
- Petzold, L. R. (1983), Automatic selection of methods for solving stiff and nonstiff systems of ordinary differential equations, *SIAM J. Sci. Stat. Comput.*, 4, 136–148.
- 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 description of bias to multi-objective model calibration, *Water Resour. Res.*, 48, W09543, doi:10.1029/2011WR011391.
- Renard, B., D. Kavetski, G. Kuczera, M. Thyer, and S. 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.
- Ryabicki, G. B., and W. H. Press (1995), A class of fast methods for processing irregularly sampled or otherwise inhomogeneous one-dimensional data, *Phys. Rev. Lett.*, 74, 1060–1063.
- 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.
- Siber, R., C. Stamm, and P. Reichert (2009), Modeling potential herbicide loss to surface waters on the Swiss plateau, *J. Environ. Manage.*, 91, 290–302.
- Sorooshian, S., and J. A. Dracup (1980), Stochastic parameter estimation procedures for hydrologic rainfall-runoff models: Correlated and heteroscedastic error cases, *Water Resour. Res.*, 16, 430–442.
- Summer, N. R., P. M. Fleming, and B. C. Bates (1997), Calibration of a modified SFB model for twenty-five Australian catchments using simulated annealing, *J. Hydrol.*, 197(1–4), 166–188, doi:10.1016/S0022-1694(96)03277-5.
- SWISSTOPO, Bundesamt für Landestopographie (2008), Digital Land-Use Map of Switzerland, Wabern, Switzerland.
- Thyer, M., G. Kuczera, and Q. J. Wang (2002), Quantifying parameter uncertainty in stochastic models using the Box-Cox transformation, *J. Hydrol.*, 265(1–4), 246–257, doi:10.1016/S0022-1694(02)00113-0.
- Thyer, M., B. Renard, D. Kavetski, G. Kuczera, S. W. Franks, and S. Srikanthan (2009), Critical evaluation of parameter consistency and predictive uncertainty in hydrologic modeling: A case study using Bayesian total error analysis, *Water Resour. Res.*, 45, W00B14, doi:10.1029/2008WR006825.
- Usmani, R. (1994), Inversion of Jacobi’s tridiagonal matrix, *Comput. Math. Appl.*, 27, 59–66.
- Vogel, R. M., and A. Sankarasubramanian (2003), Validation of a watershed model without calibration, *Water Resour. Res.*, 39(10), 1292, doi:10.1029/2002WR001940.
- 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.
- Willems, P. (2009), A time series tool to support the multi-criteria performance evaluation of rainfall-runoff models, *Environ. Modell. Software*, 24(3), 311–321, doi:10.1016/j.envsoft.2008.09.005.
- Wittmer, I. K., H.-P. Bader, R. Scheidegger, H. Singer, A. Luck, I. Hanke, C. Carlsson, and C. Stamm (2010), Significance of urban and agricultural land use for biocide and pesticide dynamics in surface waters, *Water Res.*, 44(9), 2850–2862, doi:10.1016/j.watres.2010.01.030.
- Yang, J., P. Reichert, K. C. Abbaspour, and H. Yang (2007), Hydrological modelling of the Chaohe Basin in China: Statistical model formulation and Bayesian inference, *J. Hydrol.*, 340(3–4), 167–182.