WATER RESOURCES RESEARCH, VOL. 42, W07416, doi:10.1029/2005WR004661, 2006

A Bayesian perspective on input uncertainty in model calibration: Application to hydrological model "abc"

David Huard[1] and Alain Mailhot[1]

Received 12 October 2005; revised 20 March 2006; accepted 29 March 2006; published 22 July 2006.

[1] The impact of input errors in the calibration of watershed models is a recurrent theme in the water science literature. It is now acknowledged that hydrological models are sensitive to errors in the measures of precipitation and that those errors bias the model parameters estimated via the standard least squares (SLS) approach. This paper presents a Bayesian uncertainty framework allowing one to account for input, output, and structural (model) uncertainties in the calibration of a model. Using this framework, we study the impact of input uncertainty on the parameters of the hydrological model "abc." Mostly of academic interest, the "abc" model has a response linear to its input, allowing the closed form integration of nuisance variables under proper assumptions. Using those analytical solutions to compute the posterior density of the model parameters, some interesting observations can be made about their sensitivity to input errors. We provide an explanation for the bias identified in the SLS approach and show that in the input error context the prior on the input "true" value has a significant influence on the parameters' posterior density. Overall, the parameters obtained from the Bayesian method are more accurate, and the uncertainty over them is more realistic than with SLS. This method, however, is specific to linear models, while most hydrological models display strong nonlinearities. Further research is thus needed to demonstrate the applicability of the uncertainty framework to commonly used hydrological models.

Citation: 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.

1. Introduction

[2] *Gupta et al.* [2003] identify two important issues that need to be addressed in order to improve the calibration of hydrological models: accounting for all sources of uncertainty (input, state, structural, parameter and output uncertainties), and basing model calibration on multiple noncommensurable measures of model performance. This article tackles the first issue using Bayesian analysis. We propose a theoretical framework in which all sources of uncertainty are accounted for. Using this framework ensures that the calibration of the model and its predictions remain coherent despite the underlying uncertainties. This article does not, however, present a general algorithm to use this framework. Indeed, the resolution method we use is tailor-made for the chosen application and would be inadequate for the vast majority of hydrological models. It allows, however, a deeper look into the obstacles that any such algorithm will have to face.

[3] Section 2 introduces the backbone of the method, the Bayesian uncertainty framework. We discuss the different sources of uncertainties occurring in hydrological models and describe how they fit into this framework. The standard Bayesian approach to calibration is then presented, and we show how it can be modified to take various sources of uncertainties into account. Familiarity of the reader with Bayesian analysis is assumed. Note that although input, output, model and state uncertainties can be treated in the proposed framework, the paper focuses on input uncertainty.

[4] Section 3 applies the method to the seminal problem of fitting a straight line to a data set. This topic has been discussed extensively and provides a computationally simple benchmark for our method. Importantly, it allows a simple and intuitive interpretation of the issues related to input uncertainties.

[5] Section 4 details the application of the uncertainty framework to "abc," a pedagogical hydrological model. Chosen for its analytical properties, "abc" is linear with respect to its input, while exhibiting a behavior similar to that of more complex hydrological models. Section 5 proceeds with the calibration of the "abc" model in different settings using numerical simulations. The analysis of these results relies in part on those from the straight line model. Section 6 summarizes the most interesting observations about the treatment of input errors. Finally, section 7 discusses the simplifications made in the paper and the issues that will have to be tackled in the future.

2. Calibration, Uncertainties, and Bayesian Analysis

[6] In order to use any model, whether it describes physical, biological or hydrological processes, its parame-

[1]Centre Eau, Terre and Environnement, Institut National de la Recherche Scientifique, Quebec, Quebec, Canada.

0043-1397/06/2005WR004661$09.00

ters must be specified. Indeed, models almost always rely on some set of parameters, allowing the user to tune the model to a particular setting. Ultimately, the objective is to find the "best" set of parameters. What best means is quite subjective and relative to the situation, and we use it here meaning the set of parameters leading to the most accurate predictions possible given the data at hand. Uncertainties in the estimation of parameters arise from data errors and modeling errors. Note the distinction between an error, a difference between true values and measurements, and the uncertainty, the incomplete state of knowledge about the true values.

2.1. Challenge of Calibration

[7] The estimation of a model's parameters, the calibration, goes from deceptively simple to highly complex depending on the problem at hand and the user's requirements. A typical calibration involves two broad steps: defining a function measuring the agreement of model output with the data, and finding the parameters that maximize this function. In most applications, the agreement is described by e^{-S}, with $S \equiv \sum (y_i - \tilde{y}_i)^2$, the sum of the squared differences between model output and measurements. Since in the common usage we are only interested in the best fitting parameters, the extreme value of the function, there is no harm in taking the logarithm and minimizing S instead of maximizing the exponential. This approach is called the standard least squares (SLS) approach, and has been applied with success over the years. There are other measures of agreement, a review of some used in hydrology is given by *Gupta et al.* [1998]. The second step, finding the maximum of the function, is relatively easy in one dimension, but can become a daunting task for high-dimensional cases due to the presence of local maxima. Novel algorithms exploit stochastic methods to explore the parameter space while avoiding staying trapped into local maxima. While there is no guarantee that the global maximum is found, *Vrugt et al.* [2003] report reliable results using an efficient Markov chain Monte Carlo sampler.

[8] In general, the SLS approach provides sound estimates as long as the input data is precisely known. However, if significant input errors are present and if the model is sensitive to those errors, the parameter estimates are biased and the confidence intervals are much too optimistic [*Kavetski et al.*, 2002]. In the last thirty years or so, a number of studies have highlighted the sensitivity of hydrological models to input errors [*Troutman*, 1982; *Andreassian et al.*, 2001; *Oudin et al.*, 2005], as well as their sensitivity to model errors [*Engeland et al.*, 2005] (see *Mein and Brown* [1978] for a dated but interesting review). Yet, despite these warnings, few calibration methods directly address the issue of input uncertainty. Known exceptions are the generalized likelihood uncertainty estimation (GLUE) methodology [*Beven and Binley*, 1992], Bayesian total error analysis (BATEA) [*Kavetski et al.*, 2002], particle filters [*Moradkhani et al.*, 2005], simultaneous optimization and data assimilation (SODA) [*Vrugt et al.*, 2005], and Gaussian processes [*Kennedy and O'Hagan*, 2001]. While these methods provide seducing solutions to the treatment of input uncertainty (and model uncertainty for some), their concern for numerical efficiency and practicality overshadows important theoretical issues. Our aim in this paper is to temporarily lay aside practical considerations and address the fundamental issues related to uncertainties. To do this, the first step is to define an uncertainty framework linking the different sources of uncertainties to the data and model.

[9] The most promising avenue to include all types of uncertainties in the calibration process is certainly Bayesian analysis. Bayesian analysis consists in the manipulation of probability statements about hypotheses via two logical rules, the sum rule and the product rule [*Jaynes and Bretthorst*, 2003]. It is worthwhile mentioning that in this context, a probability has the usual common sense meaning of a degree of confidence. This contrasts with the usual statistical probability, defined as a frequency of occurrence. The thriving literature feeding the feud between Bayesians and "frequentists" [*Efron*, 1986; *Clark*, 2005] will no doubt be of interest to fans of passionate debates. On a more serious note, reference textbooks usually cited are those of *Bernardo and Smith* [1994] and *Gelman et al.* [1995].

2.2. Uncertainty Framework

[10] The first step to design a method able to account for uncertainties is to lay down an uncertainty framework. This framework describes how errors occur and propagate through the physical model. It is based on an idealization of the sampling and modeling processes. Hence it should be viewed as an approximation of how "real" errors influence data and modeling.

[11] Although not identical, the framework we propose is very similar to the one implicitly used by *Vrugt et al.* [2005] in the SODA method, based on ensemble Kalman filters. Their focus, however, was not the theoretical issues related to input errors but rather the implementation of a practical algorithm to consider different sources of errors.

2.2.1. Error Models

[12] The framework we propose (see Figure 1) assumes the existence of true variables and a true process. It further assumes that, with the knowledge of the true inputs and the true process, it is theoretically possible to determine exactly the true outputs. However, as P.-S. Laplace noted, such determinism is only theoretical. In practice, we only have access to a finite number of imperfect measurements, data, and a more or less naive understanding of nature's behavior: the model. Indeed, models typically work at scales very different from the natural scales, simulate only aggregated input and output variables, account for a handful of effects, neglect all exterior influences too difficult to measure or simulate and are limited by our understanding of the physical laws and our computing power. Models are nevertheless useful to validate new physical laws, understand phenomena, predict events and give decision makers realistic scenarios to compare projects or costs. The idea behind our strategy is to give those models a boost by coupling them to comprehensive error models.

[13] We will consider three different types of errors that affect the modeling process: input, output and structural errors. It is worthwhile to define the nature and origin of these errors since they are at the core of the method. Input errors are defined as the difference between the input data and the true inputs. They originate from the inherent imprecision of measurements, as well as from their imperfect representativeness. For example, a lumped hydrological model may take as true input the total amount of precipitation over the watershed during the last month. A pluviometer, however, only averages rainfall over a few square centimeters, and rainfall over the whole catchment must be

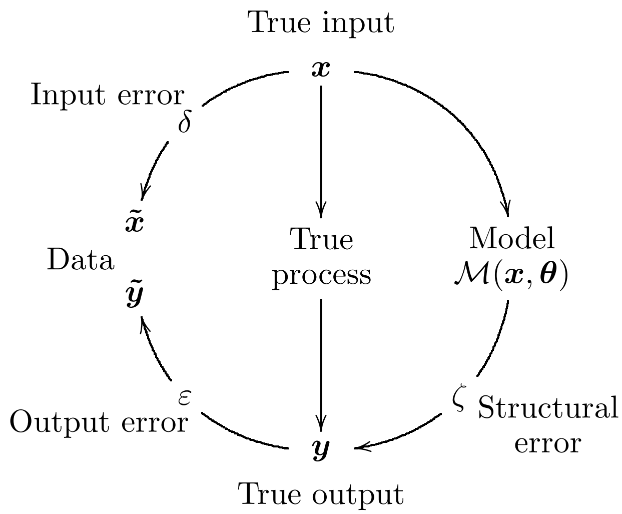

Figure 1. Idealized uncertainty framework.

extrapolated. Thus even with an infinitely precise pluviometer, there would still be a difference between the pluviometer readings and the theoretical input. This difference is what we call the input error.

[14] Output errors are defined analogously as the difference between true outputs and output measurements. Again, both the imprecision of measurements and the representativeness of the measures are sources of errors. In hydrological modeling, the output variable is typically the streamflow at the basin outlet. Extrapolated from the water height and a rating curve, streamflow measurements tend to lose accuracy as the flow increases.

[15] Structural errors are defined here as the difference between the true outputs and the model output using true inputs and true parameters. Structural errors can arise from incorrect modeling hypotheses or unmodeled processes [*Sorooshian and Dracup*, 1980]. They are stochastic by nature, due to the intrinsic variability of natural processes, but may nonetheless display distinct biases. For example, some hydrological models underestimate peak flow or overestimate base flow. Hence, in those cases, the distribution of structural errors is a function of the input variable.

[16] The errors affecting the data and model will be described by an error model, containing three elements: the input error model, the output error model and the structural error model. These three error models relate, probabilistically, the data taken to the true values. Denoting the input, output and structural errors by δ, ε and ζ respectively, the error models describe $f_\delta(\delta)$, $f_\varepsilon(\varepsilon)$ and $f_\zeta(\zeta)$, the error probability density functions (pdf).

[17] The two most obvious ways to relate errors and data are additively and multiplicatively. To allow the use of Gaussian distributions, we will work with additive errors:

$$\begin{array}{ll} \text{Input} & \tilde{\mathbf{x}} = \mathbf{x} + \delta \\ \text{Structural} & \mathbf{y} = \mathcal{M}(\mathbf{x}, \theta) + \zeta \\ \text{Output} & \tilde{\mathbf{y}} = \mathbf{y} + \varepsilon, \end{array} \tag{1}$$

where $\mathbf{x}$, $\mathbf{y}$, θ stand for the vectors of true inputs, true outputs and true parameters respectively. Here and in the following, measurements are differentiated from true values by a tilde. Note that the structural error is defined by the difference between the true output and the model simulation using the true input $\mathbf{x}$ and true parameters θ.

2.2.2. Generalized Output Error Model

[18] Inspection of equation (1) suggests that the output error model can be made to include the structural error model

$$\tilde{\mathbf{y}} = \mathcal{M}(\mathbf{x}, \theta) + \varepsilon + \zeta.$$

Thus it is possible to replace the output error model and the structural error model by a "generalized" output error model, combining both errors. For the time being, however, we keep the two models apart since they are qualitatively and quantitatively different. Indeed, output errors are generally assumed to be unbiased, with homoskedastic or heteroskedastic variance. On the other hand, structural error are known to exhibit biases, due to incorrect or incomplete modeling of the underlying processes. Moreover, the magnitude and direction of this bias may be a function of the input variable.

2.2.3. Note on Parameter Uncertainty

[19] Our framework assumes the existence of input, output and structural uncertainties, but we have yet made no mention of the parameter uncertainty. The reason for this is that from our point of view, parameter errors are included in structural errors. Having a method that takes care of model errors, parameter errors lose their interest. We will, however, use the expression parameter uncertainty to describe our incomplete state of knowledge about the true parameters. Thus once the calibration is completed, there remains a parameter uncertainty owed to structural errors and limited supply of inexact data. By following the rules of probability, the resulting parameter uncertainty should incorporate the uncertainties about the model, inputs and outputs. Predictions can then be carried out by using not only the most probable parameter set, but the entire distribution. It is precisely the use of the whole parameter distribution that allows us to compute realistic confidence intervals about the predictions, incorporating the different sources of uncertainties.

2.3. Parameter Inference

2.3.1. Probability Inversion

[20] As stated earlier, the input, output and structural models define the probability of observing an error δ, ε and ζ. Yet, since the errors cannot be observed, it is more convenient to write the error models in terms of measured and true values:

$$\begin{array}{ll} \text{Input} & p(\delta) \Leftrightarrow p(\tilde{\mathbf{x}}|\mathbf{x}) \\ \text{Structural} & p(\zeta) \Leftrightarrow p(\mathbf{y}|\mathbf{x}, \theta, \mathcal{M}) \\ \text{Output} & p(\varepsilon) \Leftrightarrow p(\tilde{\mathbf{y}}|\mathbf{y}). \end{array} \tag{2}$$

The notation of probabilistic statements in this paper follows the general usage, in which p stands for the probability density. It should be remembered, however, that using the laws of probability on such expressions is a shortcut and that it can, in some cases, lead to paradoxes [*Jaynes and Bretthorst*, 2003].

[21] The next step is to combine the error models into a unified equation describing $p(\theta|\tilde{\mathbf{x}}, \tilde{\mathbf{y}}, \mathcal{M})$. To do so, we introduce $\mathbf{x}$ and $\mathbf{y}$ as nuisance variables and inverse the probability using Bayes' theorem:

$$\begin{aligned} p(\theta|\tilde{\mathbf{x}}, \tilde{\mathbf{y}}, \mathcal{M}) &= \iint p(\theta, \mathbf{x}, \mathbf{y}|\tilde{\mathbf{x}}, \tilde{\mathbf{y}}, \mathcal{M})\, \mathrm{d}\mathbf{x}\, \mathrm{d}\mathbf{y} \\ &= \iint p(\tilde{\mathbf{x}}, \tilde{\mathbf{y}}|\mathbf{x}, \mathbf{y}, \theta, \mathcal{M}) \\ &\quad \cdot p(\mathbf{x}, \mathbf{y}, \theta|\mathcal{M})\, \mathrm{d}\mathbf{x}\, \mathrm{d}\mathbf{y} \cdot \frac{1}{p(\tilde{\mathbf{x}}, \tilde{\mathbf{y}}|\mathcal{M})} \\ &= \iint p(\tilde{\mathbf{x}}, \tilde{\mathbf{y}}|\mathbf{x}, \mathbf{y}) p(\mathbf{y}|\mathbf{x}, \theta, \mathcal{M}) p(\mathbf{x}, \theta|\mathcal{M}) \\ &\quad \cdot \mathrm{d}\mathbf{x}\, \mathrm{d}\mathbf{y} \cdot \frac{1}{p(\tilde{\mathbf{x}}, \tilde{\mathbf{y}})} \\ &= \iint p(\tilde{\mathbf{x}}|\mathbf{x}) \cdot p(\tilde{\mathbf{y}}|\mathbf{y}) p(\mathbf{y}|\mathbf{x}, \theta, \mathcal{M}) p(\mathbf{x})\, \mathrm{d}\mathbf{x}\, \mathrm{d}\mathbf{y} \cdot \frac{p(\theta)}{p(\tilde{\mathbf{x}}, \tilde{\mathbf{y}})} \end{aligned} \quad (3)$$

Equation (3) encloses the three error models, as well as a prior on the parameters $p(\theta)$, and a prior on the true input variable $p(\mathbf{x})$. We stress that this prior appears from the inversion of $p(\mathbf{x}|\tilde{\mathbf{x}})$ to $p(\tilde{\mathbf{x}}|\mathbf{x})$, the input error model defined above. As we will soon see, failure to distinguish both expressions leads to serious shortcomings in the estimation of parameters. Note that we used the simplifying assumption $p(\tilde{\mathbf{x}}, \tilde{\mathbf{y}}|\mathbf{x}, \mathbf{y}) = p(\tilde{\mathbf{y}}|\mathbf{y})p(\tilde{\mathbf{x}}|\mathbf{x})$, that is, we supposed that the errors δ and ε were conditionally independent. While this is not mandatory, it simplifies the computations. Finally, the denominator $p(\tilde{\mathbf{x}}, \tilde{\mathbf{y}})$ is simply a normalization constant, provided the integral in the numerator converges.

[22] Since this approach is derived from that of *Kavetski et al.* [2002], it is worthwhile to highlight the main difference between the two. In the work of *Kavetski et al.* [2002], the structural errors are not defined. Also, the nuisance variables $\mathbf{x}$ are not integrated but rather estimated, on the same footing as the model parameters. This approach makes sense since in their application, the number of nuisance variables is considerably lower than the sample size. In the case where there is one nuisance variable for each measurement, however, their approach would suffer from overparametrization.

2.3.2. Standard Shortcuts

[23] In Bayesian analysis as in standard least squares schemes, we implicitly assume that the input variables are exact, that is $p(\tilde{\mathbf{x}}|\mathbf{x}) = \delta(\tilde{\mathbf{x}} - \mathbf{x})$. Integrating equation (3) over $\mathbf{x}$ under this assumption, we obtain:

$$\begin{aligned} p(\theta|\tilde{\mathbf{x}}, \tilde{\mathbf{y}}, \mathcal{M}) &= \int p(\tilde{\mathbf{y}}|\mathbf{y}) p(\mathbf{y}|\tilde{\mathbf{x}}, \theta, \mathcal{M})\, \mathrm{d}\mathbf{y} \cdot \frac{p(\theta)}{p(\tilde{\mathbf{y}}|\tilde{\mathbf{x}})} \\ &= \frac{p(\tilde{\mathbf{y}}|\tilde{\mathbf{x}}, \theta, \mathcal{M}) p(\theta)}{p(\tilde{\mathbf{y}}|\tilde{\mathbf{x}})}, \end{aligned}$$

the usual result of Bayes' theorem. Further assuming that the likelihood $p(\tilde{\mathbf{y}}|\tilde{\mathbf{x}}, \theta, \mathcal{M})$ is Gaussian and that the prior $p(\theta)$ is uniform, we obtain the SLS solution.

[24] It is worthwhile to stress that in this derivation of the usual Bayesian or SLS solution, there is no assumption about the form of the structural error model. Indeed, the likelihood $p(\tilde{\mathbf{y}}|\tilde{\mathbf{x}}, \theta, \mathcal{M})$ is rather a generalized likelihood, combining the output error model and the structural error model via the convolution of both distributions:

$$p(\tilde{\mathbf{y}}|\tilde{\mathbf{x}}, \theta, \mathcal{M}) = \int p(\tilde{\mathbf{y}}|\mathbf{y}) p(\mathbf{y}|\tilde{\mathbf{x}}, \theta, \mathcal{M})\, \mathrm{d}\mathbf{y}. \quad (4)$$

Hence once an output error model and a structural error model are specified, the minimization of equation (4) with respect to θ leads to an optimal solution, taking structural and output uncertainties into account. While apparently simple, this treatment of structural uncertainties hides a formidable challenge, namely, the adequate characterization of the structural error model.

[25] Our prime interest in this paper, however, lies not with structural errors but with the input errors and their impact on the parameters. Our focus on this subject stems from two realizations. The first one is that falsely assuming exact input causes parameter estimates to be biased. This bias is observed for hydrological models [*Kavetski et al.*, 2002] as well as for basic linear regressions [*York*, 1966], and can have a significant impact on the reliability of hydrological predictions. The second one is that despite a growing number of publications on this subject, the processes by which input errors affect model parameters are still obscure. Therefore this article focuses on basic issues related to input errors using simple, academic models and error models. These simplifications allow us to concentrate on the essential problems, instead of diverting our attention to technical and numerical issues.

3. First Application: The Straight Line Model

[26] Despite its apparent simplicity, the straight line model "when both variables are subject to error" hides complex difficulties and has been the focus of a large number of publications from various research areas: statistics [*Lindley and El-Sayyad*, 1968; *Kendall and Stuart*, 1983; *Fuller*, 1987; *Cheng and Ness*, 1994], econometrics [*Zellner*, 1971; *Erickson*, 1989], physics [*York*, 1966; *Reed*, 1989; *Gull*, 1989], and image reconstruction [*Werman and Keren*, 2001]. In the following, we will review the standard least squares approach (also called ordinary least squares) and its inherent bias in presence of input errors. We will then apply the Bayesian uncertainty framework to better understand the origin of this bias on the slope. Finally, we will study the impact that priors have on the results.

3.1. Standard Least Squares

[27] The standard least squares (SLS) solution to an optimization problem dates back from Gauss, who invented the method to identify the orbit of Ceres. SLS works under the assumption that the independent variable is known exactly. The optimal parameters are then simply the ones that make the sum of the squared errors as small as possible. Stated otherwise, the SLS parameters are those maximizing the probability of "drawing" the errors from a Gaussian distribution with zero mean. The wide success enjoyed by SLS has since been linked to the exceptional mathematical properties of the Gaussian distribution: the product of two Gaussian distributions is an unnormalized Gaussian distribution, the convolution of two Gaussians is another Gaussian, it is the only distribution whose maximum likelihood estimate is also the arithmetic mean, it is the maximum

entropy distribution for fixed mean and variance and the limiting distribution of additive errors from any distribution. In contrast with its huge success, SLS fails when errors affect both dependent and independent variables [*York*, 1966].

[28] Let's assume that our model is a straight line passing through the origin $y = \theta x$. Different measurements are taken, denoted by $\tilde{x}_i$, $\tilde{y}_i$. To simplify the example, we will suppose that there is no output error, that is, $\tilde{y}_i = y_i$. Now, if we assume, correctly, that the output data is exact, and that errors are only on the input data, the SLS method provides the following estimate for θ:

$$\widehat{\theta} = \frac{\sum \tilde{y}_i^2}{\sum \tilde{x}_i \tilde{y}_i}.$$

This solution is correct, in the sense that it converges toward the true value as the sample size increases. However, if we falsely assume that the errors are on the output variable and the inputs are exact, we find instead

$$\widehat{\theta} = \frac{\sum \tilde{x}_i \tilde{y}_i}{\sum \tilde{x}_i^2} \sim \frac{\sum \tilde{x}_i \tilde{y}_i}{\sum x_i^2 + \sigma^2}. \tag{5}$$

Of course, if $\sigma^2 = 0$, there is no uncertainty and both solutions are equal. However, as the error variance σ^2 increases, the slope gets underestimated. Moreover, this bias does not decrease as the number of samples increases.

[29] In most real life situations, input and output data contain errors, and the slope will systematically be under or overestimated depending on which variable we choose to be the independent one. Some authors have chosen to compute both slopes, and then take the average. Far more ingenuous algorithms have since been devised [*York*, 1966], based on the individual weighting of each data. Our objective, however, is not to review the various solutions to the straight line model but to understand the origin of the bias.

3.2. Density Mapping Effect

[30] For the sake of simplicity and pedagogy, we will consider a simple situation.

[31] 1. The model $\mathcal{M}$ is a straight line with intercept at zero, i.e., $y = \theta x$, with $\theta = 3$.

[32] 2. There are no output nor structural error, $p(\tilde{y}|x, \theta, \mathcal{M}) = \delta(\tilde{y} - \theta x)$.

[33] 3. The input error is Gaussian $p(\tilde{x}|x) = \mathcal{N}(\tilde{x}|x, \sigma = 1)$.

[34] 4. The prior on θ is uniform on the interval $[\theta_a = 0, \theta_b = 5]$.

[35] 5. The prior on the true value $p(x)$ is uniform on the interval $[x_a = 0, x_b = 10]$.

[36] Let's imagine we have a single input measurement at $\tilde{x} = 2$. Knowing that the true value of x lies around the measurement, with a Gaussian probability

$$p(x|\tilde{x}) \propto p(\tilde{x}|x)p(x)$$
$$\propto \mathcal{N}(x|\tilde{x}, \sigma) \qquad x \in [x_a, x_b],$$

we want to compute the probability that $\tilde{y}$ is the true value knowing the slope θ. After a short analysis, we obtain

$$p(\tilde{y}|\tilde{x}, \theta, \mathcal{M}) \propto \frac{1}{\theta} \mathcal{N}(\tilde{y}/\theta|\tilde{x}, \sigma) \qquad \tilde{y} \in [\theta x_a, \theta x_b]. \tag{6}$$

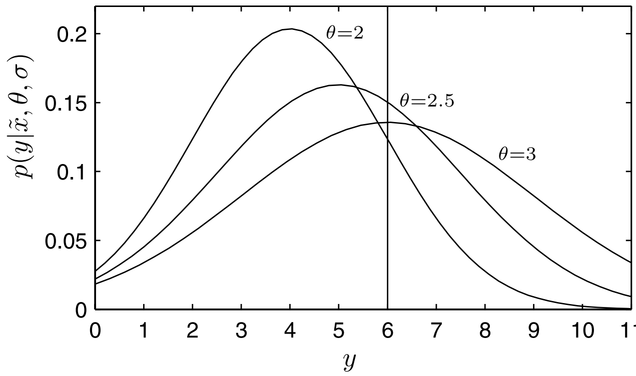

Figure 2. Probability density $p(y|\tilde{x}, \theta, \sigma)$ with $\tilde{x} = 2$, $\theta = 2$, 2.5, and 3, and $\sigma = 1$.

Figure 2 plots equation (6) for $\tilde{x} = 2$ and three values of θ: 2, 2.5 and 3. For small slopes, the y distribution is sharper than for bigger slopes. If we interpret function (6) as a mapping of the x elements onto the $\tilde{y}$ space, the values of $\tilde{y}$ are mapped more closely together with smaller slopes. Hence the pdf for $\tilde{y} = 6$ is higher with $\theta = 2.5$ than with $\theta = 3$, even though our intuitive guess, and the SLS estimate, would rather be $6/\tilde{x} = 3$.

[37] Suppose we now want to estimate the slope using measurements ($\tilde{x} = 2$, $\tilde{y} = 6$). Bayesian analysis tells us that

$$p(\theta|\tilde{x}, \tilde{y}, \mathcal{M}) \propto p(\tilde{y}|\tilde{x}, \theta, \mathcal{M})p(\theta), \tag{7}$$

but since $p(\theta) = \mathcal{U}(\theta_a, \theta_b)$, the probability for θ given $\tilde{y}$ is simply proportional to the probability of $\tilde{y}$ given θ. Thus, judging from Figure 2, $\theta = 2.5$ should be more probable than $\theta = 3$. Indeed, if we plot equation (7), we find that the most probable slope is around 2.5 (Figure 3).

[38] Figure 3 shows quite clearly how the uncertainty about x is asymmetrically mapped onto an uncertainty about θ. Of course, this asymmetry disappears as the uncertainty over x decreases. We will call the mapping of the input uncertainty onto the parameter uncertainty, the "density mapping" effect.

3.3. Straight Line Fitting

[39] A key question is whether or not this density mapping effect vanishes as the sample size increases. Unfortunately, it is difficult to find a simple answer based on analytical calculations. Hence we resort to numerical simulations to gain an insight into the problem. In this demonstration we use the first four assumptions of section 3.2. The only difference with the previous example is the number of samples and the prior on the true values, that will be modified to understand its impact on the parameter posterior pdf.

3.3.1. Synthetic Data

[40] We first generate a synthetic sample consisting of 500 couples $(\tilde{x}_i, \tilde{y}_i)$, shown in Figure 4. The true inputs x_i are randomly generated from a Gamma distribution, $x \sim \mathcal{G}(3, 1)$. The input measurements are synthesized by adding a Gaussian error to the true inputs: $\tilde{x}_i \sim \mathcal{N}(x_i, \sigma = 1)$. The output measurements are simply equal to the true outputs which are computed via the model and the true inputs: $\tilde{y}_i = y_i = \theta x_i$ with $\theta = 3$.

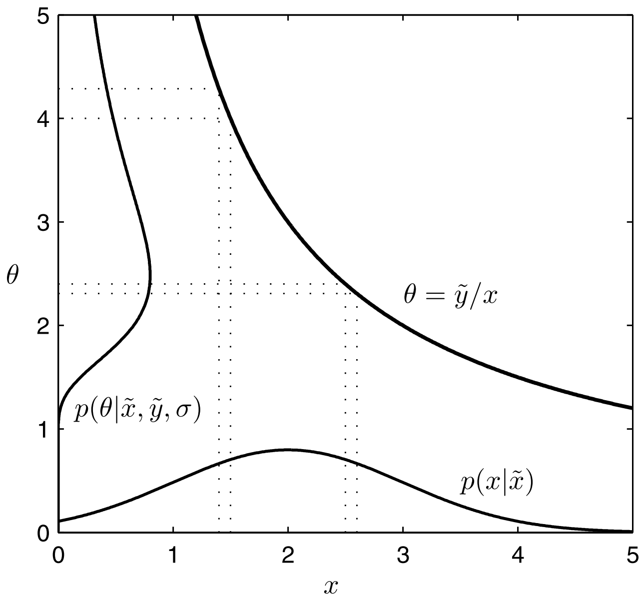

Figure 3. Probability densities $p(\theta|\tilde{x}, \tilde{y}, \sigma)$ and $p(x|\tilde{x})$ for $\tilde{x} = 2$, $\tilde{y} = 6$, and $\sigma = 1$. The hyperbola shows the relation between the slope and the true input for a given output $\tilde{y}$.

3.3.2. Error Models and Prior Hypotheses

[41] The parameter density is computed by recursively applying equation (3). That is, measurements are analyzed sequentially and update the posterior probability distribution $p(\theta_i|\tilde{x}_{1...i}, \tilde{y}_{1...i}, \mathcal{M})$. The input and generalized output error models are chosen correctly, that is, they correspond exactly to how the errors were applied to the true values:

$$p(\tilde{x}|x) = \mathcal{N}(\tilde{x}|x, \sigma)$$

for the input error model and

$$p(\tilde{y}|x, \theta) = \delta(\tilde{y} - \theta x)$$

for the generalized output error model. The only remaining hypothesis concerns the prior for the true input values $p(x)$. To understand its impact on the calibration of θ, we look at three distinct cases, (1) Gamma, $p(x) = \mathcal{G}(x|3, 1)$; (2) Gaussian, $p(x) = \mathcal{N}(3, \sqrt{3})$; and (3) uniform, $p(x) = \mathcal{U}(0, 10)$, and compute the posterior pdf for the slope θ for each of these cases. We also compute the slope and the 90% confidence intervals using SLS.

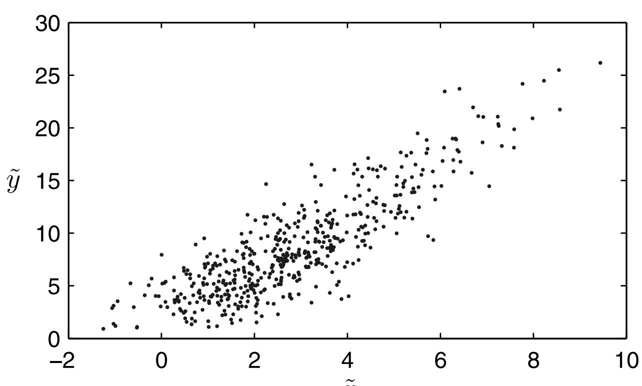

Figure 4. Synthetic data used to study the calibration of a straight line. The inputs $\tilde{\mathbf{x}}$ are computed using draws from a Gamma distribution to which a Gaussian noise is added. The outputs are given by $\tilde{y}_i = 3x_i$.

3.3.3. Results

[42] The results are shown in Figure 5. Slope estimates from the Gamma and Gaussian priors are very near to the true value, θ = 3, whereas those of SLS and the uniform prior are significantly underestimated. The similarity between SLS results and the uniform prior is not a coincidence. Indeed, the parameter maximizing the posterior density using a uniform prior over the real domain is identical to the "major axis," a close cousin of SLS. That is, the slope for which the perpendicular distance from the points to the line is a minimum. Note that at this sample size, 500, the results do not vary much for different simulations.

[43] The most surprising observation about these results is that contrary to the Bayesian motto "the effect of the prior weakens with increasing sample size," the prior plays here a significant role even with a sample of 500 observations. Indeed, there is a significant difference between the posterior pdf using a uniform prior and a Gamma or Gaussian prior. Also note that the choice of the distribution parameters is also significant. A Gamma distribution with crude parameters may lead to worst results than a Gaussian with sensible ones.

3.3.4. Analysis

[44] The drastic effect of the prior can be explained as follows. For each input measurement, there is an unknown true input, whose value is inferred from the data using the error model and the prior on the true inputs. If the error model assumes a very small error, the effect of the prior is weak, whereas when the error model has a large variance, the prior plays an important role in the inference of the true input. Thus the effect of the prior does not weaken since with each new sample, another prior is added to infer the input true value. Moreover, if the prior is incorrect, each inference about the true value is flawed and the estimation of the parameters will not "converge" to the parameter's true value, even with an infinitely large sample.

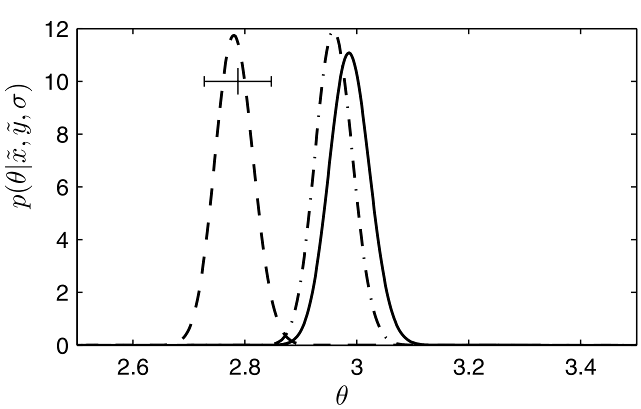

Figure 5. Probability density $p(\theta|\tilde{\mathbf{x}}, \tilde{\mathbf{y}}, \sigma)$ computed over a sample of 500 points using equation (3) with three different priors on the true values: Gamma (solid line), Gaussian (dash-dotted line), and uniform (dashed line). The cross indicates the SLS result as well as the 90% confidence interval. The true value of θ is 3.

[45] The importance of the prior on the results changes our perception of priors and how to choose them. There exists a wide variety of priors to choose from, an exhaustive review is given by *Kass and Wasserman* [1996]. In the light of the preceding results, we would recommend to avoid ignorance priors and rather look for additional evidence that would describe adequately the distribution of the true inputs. If a similar data set is available (other than the one under study), one solution is to fit a distribution to the data and use it as the prior. In the case where no such data sets exist, but where some descriptors such as the mean or variance are known (or can ge guessed), maximum entropy methods can be used to find the distribution with the highest entropy given those descriptors [*Jaynes*, 1983].

3.4. Highlights

[46] Before we go to the next section and begin our study of the hydrological model "abc," it is worthwhile to recall the highlights of this section: (1) The mapping of input uncertainties over the parameter space, the mapping density effect, is responsible for the "bias" observed in the parameter posterior density. (2) The influence of the mapping density effect decreases as data accumulate, but only as long as the prior on the true inputs allows reliable inference. That is, the estimated slope will only converge toward the true slope if the prior reflects the true input distribution.

4. Application to Hydrological Model "abc"

[47] We will now apply the Bayesian uncertainty framework to a simple hydrological model, the "abc" model. The "abc" model has been devised by Harold A. Thomas and introduced as a pedagogical tool by *Fiering* [1967]. Since then, it has been used as a benchmark for various computational methods [*Vogel and Sankarasubramanian*, 2003]. Its main advantages are that (1) it is linear with respect to the inputs, (2) it has only three parameters and one boundary condition, and (3) despite its simplicity, its calibration displays pathologies similar to more complex hydrological models.

[48] In the following, we first describe the "abc" model and how it can be written in a convenient form using linear algebra. Then, we describe the data set and justify the use of synthetic discharges in our numerical simulations. The next and last paragraphs discuss the error models chosen along with the prior distribution for the parameters and true values.

4.1. Description

[49] The "abc" model comprises two equations, one for the discharge Q_t and the other for water storage S_t, where t denotes time steps (annual or monthly). The model has three parameters (a, b, c) (hence the name) and a state parameter, S_t, the storage. The model is driven by the rainfall r_t:

$$\begin{aligned} Q_t &= (1-a-b)r_t + cS_t \\ S_{t+1} &= (1-c)S_t + ar_t \end{aligned} \tag{8}$$

Each parameter has a pseudophysical signification: a stands for the proportion of rainfall entering the storage, b is the proportion of rainfall lost to evapotranspiration and c is the percentage of water seeping from the storage to the basin outlet. All parameters take values in the interval [0, 1], and the conservation of water imposes an additional constraint on a and b, namely $a + b \leq 1$.

4.2. Matrix Formulation

[50] The main advantage of "abc" is that it can also be written elegantly in matrix form. To do so, we use induction to write the storage at the nth time step in terms of the initial storage and the rainfall history:

$$S_n = (1-c)^{n-1}S_1 + a\sum_{k=1}^{n-1}(1-c)^{n-k-1}r_k, \quad n > 1 \tag{9}$$

By defining the following vectors

$$\begin{aligned} \mathbf{x} &= [r_1, r_2, \ldots, r_n]^t \\ \mathbf{y} &= [Q_1, Q_2, \ldots, Q_n]^t \\ \mathbf{t} &= \left[c, c(1-c), c(1-c)^2, \ldots, c(1-c)^{n-1}\right]^t, \end{aligned}$$

and plugging equation (9) into (8), the model output $\mathbf{y}$ can be expressed in a convenient matrix formulation:

$$\mathbf{y} = \mathsf{A}\mathbf{x} + S_1\mathbf{t},$$

where matrix A is defined as

$$\mathsf{A} = \begin{bmatrix} (1-a-b) & 0 & 0 & \cdots & 0 \\ ac & (1-a-b) & 0 & & 0 \\ ac(1-c) & ac & (1-a-b) & & 0 \\ \vdots & & & \ddots & \vdots \\ ac(1-c)^{n-2} & ac(1-c)^{n-3} & ac(1-c)^{n-4} & \cdots & (1-a-b) \end{bmatrix}$$

4.3. Error Models Selection

[51] The selection of error models is restricted by a formidable requirement: closed form integrability. Indeed, equation (3) contains an integral over $2n$ variables. Standard numerical integration algorithms are not efficient with more than ten dimensions, and Markov chain Monte Carlo integration would take a huge amount of computing power to integrate over a mere 50 dimensions. Fortunately, Gaussian multivariate functions are readily integrable for any number of dimensions and are generally considered a valid depiction of a random error probability from an inferential point of view [*Jaynes and Bretthorst*, 2003]. For the particular context of "abc," we will thus limit our error models to multivariate Gaussian distributions in order to retain closed form integrability:

$$p(\tilde{\mathbf{x}}|\mathbf{x}) = \mathcal{N}(\tilde{\mathbf{x}}|\mathbf{x}, \Sigma_\delta) \tag{10}$$

$$p(\tilde{\mathbf{y}}|\mathbf{y}) = \mathcal{N}(\tilde{\mathbf{y}}|\mathbf{y}, \Sigma_\varepsilon) \tag{11}$$

$$p(\mathbf{y}|\mathbf{x}, \theta, \mathcal{M}) = \mathcal{N}\left(\mathbf{y}|A\mathbf{x} + S_1\mathbf{t}, \Sigma_\zeta\right), \tag{12}$$

where $\mathbf{x}$ and $\mathbf{y}$ are positively defined. Note that there are no constraints on the covariance matrices Σ_δ, Σ_ε and Σ_ζ. That is, correlated and heteroskedastic errors can be assumed by assigning nonzero off diagonal terms and variances dependent on the input data. See *Sorooshian and Dracup* [1980] for examples of such matrices.

[52] At this point, we are ready to integrate equation (3) over **y**, effectively convoluting the structural and output error model into a generalized output error model:

$$\int_{\mathbb{R}} p(\tilde{\mathbf{y}}|\mathbf{y})p(\mathbf{y}|\mathbf{x},\theta,\mathcal{M})\,d\mathbf{y} = p(\tilde{\mathbf{y}}|\mathbf{x},\theta,\mathcal{M}) \tag{13}$$

Substituting functions (11) and (12) in (13) and integrating over the real domain, we find

$$p(\tilde{\mathbf{y}}|\mathbf{x},\theta,\mathcal{M}) = \mathcal{N}(\tilde{\mathbf{y}}|A\mathbf{x}+S_1\mathbf{t},\Sigma_\epsilon), \tag{14}$$

where $\Sigma_\epsilon = \Sigma_\varepsilon + \Sigma_\zeta$. Note that since discharge is a purely positive quantity, the integral in equation (13) should only span the positive orthant. The integration over the real domain is an approximation, acceptable only if the means of $\mathsf{A}\mathbf{x} + S_1\mathbf{t}$ are three or four standard deviations away from 0. The same comment applies to the integration over **x**. Before we can proceed with the integration over the input variable, the prior $p(\mathbf{x})$ must first be defined.

4.4. Prior on the True Rainfall

[53] We choose to define the prior on the true rainfall as a historical prior, that is, $p(\mathbf{x})$ will reflect the distribution of monthly rainfall in the region. Details about how this is done are given in section 5.2. For now, it suffices to say that the prior is described by a sum of three Gaussian distributions with different means and variances. It is therefore possible to describe distributions very different in shape from a unique Gaussian, while respecting the requirement that all error models are to be Gaussian.

$$p(\mathbf{x}) = \sum_{j=1}^{3} f_j\mathcal{N}\left(\mathbf{x}|\mathbf{R}_j,\Sigma_{Rj}\right), \tag{15}$$

where $\sum f_j = 1$, $\mathbf{R}_j = R_j\,[1, 1, \ldots, 1]^t$ and $\Sigma_{Rj} = \sigma_{Rj}^2 I$, I standing for the $n \times n$ identity matrix.

4.5. Integration of the Input Nuisance Variables

[54] The integral

$$p(\tilde{\mathbf{y}}|\tilde{\mathbf{x}},\theta,\mathcal{M}) = \int_{\mathbb{R}} p(\tilde{\mathbf{y}}|\mathbf{x},\theta,\mathcal{M})p(\tilde{\mathbf{x}}|\mathbf{x})p(\mathbf{x})\,d\mathbf{x} \tag{16}$$

can now be solved using the identity:

$$K = \int_{\mathbb{R}} \prod_{i=1}^{n} \mathcal{N}(\mathbf{x}|\mu_{\mathbf{i}},\Sigma_i)\,d\mathbf{x} = \frac{(2\pi)^{\frac{n}{2}}|\widetilde{\Sigma}|^{\frac{1}{2}}}{\prod_i (2\pi)^{\frac{n}{2}}|\Sigma_i|^{\frac{1}{2}}}\exp\left\{\frac{1}{2}\left(\widetilde{\mu}^t\widetilde{\Sigma}^{-1}\widetilde{\mu} - \sum_{i=1}^{n}\mu_i^t\Sigma_i^{-1}\mu_i\right)\right\}. \tag{17}$$

where

$$\widetilde{\Sigma} = \left(\sum_{i=1}^{n}\Sigma_i^{-1}\right)^{-1}$$
$$\widetilde{\mu} = \widetilde{\Sigma}\left(\sum_{i=1}^{n}\Sigma_i^{-1}\mu_i\right).$$

[55] Putting equations (10), (14) and (15) into (16), we find using the identity (17):

$$p(\tilde{\mathbf{y}}|\tilde{\mathbf{x}},\theta,\mathcal{M}) = \sum_{j=1}^{3} f_jK_j, \tag{18}$$

using covariance and means

$$\widetilde{\Sigma}_j = \mathsf{A}^t\Sigma_\epsilon^{-1}\mathsf{A} + \Sigma_\delta^{-1} + \Sigma_{Rj}^{-1}$$
$$\widetilde{\mu}_j = \widetilde{\Sigma}_j\left[\mathsf{A}^t\Sigma_\epsilon^{-1}(\tilde{y} - S_1\mathbf{t}) + \Sigma_\delta^{-1}\tilde{x} + \Sigma_{Rj}^{-1}\mathbf{R}_j\right].$$

The calculations are detailed explicitly in Appendix A.

4.6. Prior on the Parameters

[56] The prior on the parameters a, b and c remains to be defined. The prior density should reflect our degree of confidence in the value of the parameters. Since the parameters take values between 0 and 1, we will use the Beta distribution for the sake of generality

$$\mathcal{B}(x|r,s) = \frac{1}{\beta(r,s)}x^{r-1}(1-x)^{s-1} \qquad 0 \le x \le 1,$$

where $(r, s) > 0$. The prior on c will thus be $p(c) = \mathcal{B}(x|r_c, s_c)$. The prior on a and b is a little bit more involved due to the constraint $a + b \le 1$. Indeed, this constraint suggests a bivariate distribution. To define $p(a, b)$, we impose a dummy prior $\mathcal{B}(q|r_{ab}, s_{ab})$ on the direct runoff $q = 1 - a - b$, and using variable substitution, we compute the corresponding prior on (a, b):

$$p(a,b) = \frac{\mathcal{B}(1-a-b|r_{ab},s_{ab})}{(a+b)}.$$

The prior for the parameters is then completely specified by

$$p(\theta) = p(a,b)p(c).$$

[57] Now that all the ingredients are assembled, we are ready for numerical computations. The following section presents various results that highlight the impact of input errors on the parameters and the workings of the Bayesian uncertainty framework.

5. Results From "abc"

[58] The results from "abc" share traits similar to those of the straight line: influence of the density mapping and sensitivity to the prior $p(\mathbf{x})$. Other properties are also observed, such as model filtering and parameter averaging. Each one of these observations will be discussed in the following paragraphs. As we will see, the "abc" model is more complex than the straight line, due to the fact that it has more parameters and a memory of past events, via water storage. Hence past uncertainties propagate to blur the actual state of the system, regardless of the initial state uncertainty.

[59] The effective display of results is a troublesome issue; there seems to be no perfect way to visualize a 3-D parameter space at a glance. The solution adopted was to cut

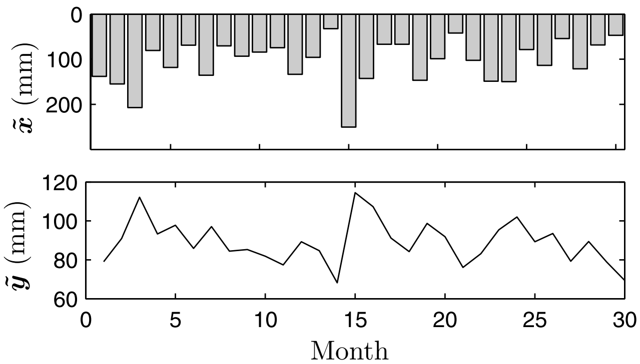

Figure 6. Inputs ($\tilde{\mathbf{x}}$) and outputs ($\tilde{\mathbf{y}}$) used in this study. The inputs are rainfall readings from a meteorological station on the Famine River. The outputs (discharges) were generated using the "abc" model with parameters $a = 0.6$, $b = 0.15$, and $c = 0.2$.

orthogonal slices in the parameter mesh at the posterior density's maximum ($a_{\max}$, $b_{\max}$, $c_{\max}$). Figures will thus contain three plots, one for each slice. The color bar uses white to indicate a null probability and black for the maximum probability reached by the pdf.

5.1. Density Mapping Effect

[60] To demonstrate the effect of the density mapping on the parameters, we isolate this effect from other influences. We eliminate the effect of random noise by using exact data, specifying an input error model with a large variance, an output error model with a very small variance, a uniform prior on the true values and vague priors on the model parameters.

5.1.1. Data

[61] The true inputs **x** consist of monthly rainfall measurements from a station on the Famine River in the Beauce region, located south of Québec city. The true outputs **y** are the discharges simulated using the "true" rainfall and the "abc" model with parameters $a = 0.6$, $b = 0.15$ and $c = 0.2$ (the same parameters used by *Kavetski et al.* [2002]). The use of synthetic discharges allows us to check the consistency of the method and analyze the results more easily, that is, compare estimates to the "true parameters."

[62] In the following simulation, the sample ($\tilde{\mathbf{x}}$, $\tilde{\mathbf{y}}$) consists of 30 months of rainfall and stream flow. The data contain no errors ($\sigma_x = \sigma_y = 0$), thus ($\tilde{\mathbf{x}}$, $\tilde{\mathbf{y}}$) = ($\mathbf{x}$, $\mathbf{y}$) and the data fit perfectly the "abc" model with parameters $\theta = (0.6, 0.15, 0.2)$. Figure 6 shows a plot of the rainfall and stream flow used in this simulation.

5.1.2. Error Models and Prior Hypotheses

[63] Following are the modeling hypotheses, designed to bring out the density mapping effect and reduce other potential influences on the parameters posterior distribution: (1) homoskedastic input error model with a large standard error $\Sigma_\delta = 25^2 I$ mm^2 (almost 30% of the mean precipitation), (2) homoskedastic output error model with a negligible variance ($\Sigma_\epsilon = 1I$ mm^2), (3) uniform prior over the true rainfall ($p(\mathbf{x}) \propto 1$), and (4) vague prior for θ, that is, the prior described in section 4.5 with parameters $r_{ab} = s_{ab} = 1$ and $r_c = s_c = 1$.

5.1.3. Results and Analysis

[64] Computation of the parameters posterior density reveals a displacement of the cloud from the true parameters indicated by a cross (see Figure 7). This displacement amplifies as the assumed input uncertainty increases. Since all other influences have been curbed, we relate the pdf displacement from the true values to the density mapping effect. In "abc", its consequence is to displace the pdf density toward higher values of evapotranspiration b and infiltration a. In other words, the calibrated model has a lower percentage of direct runoff ($1 - a - b$) than the "real" model.

[65] The underestimation of the direct runoff can be explained using an analogy with the straight line model. We have seen in Figure 2 how smaller slopes map the outputs more densely than larger ones, with the consequence that the slope probability distribution has its maximum below the intuitive slope $\tilde{y}/\tilde{x}$. In the "abc" model, the direct runoff is akin to the slope: it specifies the instantaneous response of the model to the input. Models with high direct runoff display a great variability of discharges, and the output distribution is stretched over a wider range. Models with a low direct runoff, where storage plays the dominant part, exhibit more steady discharges, and the output distribution is more concentrated. Therefore, with a given discharge and an input uncertainty, we expect the most probable model to display a lower direct runoff than the "true" model.

[66] Now, one must realize that the density mapping effect is not an artifact of the Bayesian uncertainty framework but a very common phenomenon. When the input error is not accounted for, as in SLS for example, the effect is still present, biasing the results in an apparently uncontrollable way. The Bayesian framework merely provides the means to identify the effect, and as data accumulate, to reduce its effect on the parameters' pdf. The fact that the method accounts for the input uncertainties makes sure that the pdf is consistent with the data and reflects the input uncertainties' impact on calibration.

5.2. Prior $p(\mathbf{x})$ Impact

[67] As shown in section 3, the prior on the input true values plays a significant role in the estimation of the

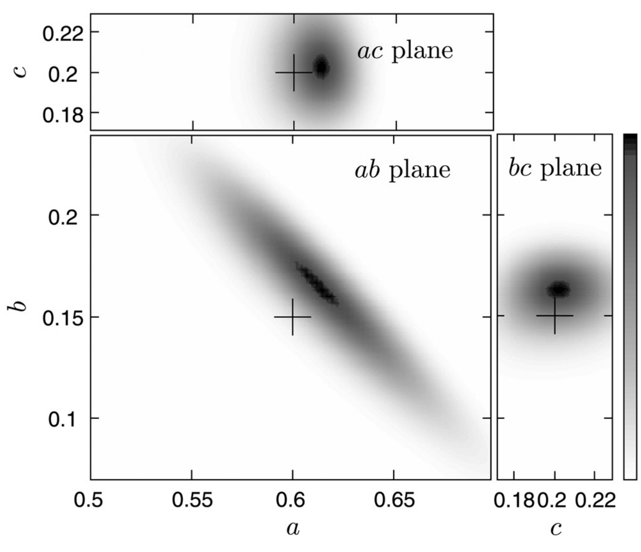

Figure 7. Density mapping effect on the parameters' posterior distribution. The slices are taken at the parameters most likely value [a, b, c] = [0.61, 0.16, 0.2]. True parameter values are indicated by crosses.

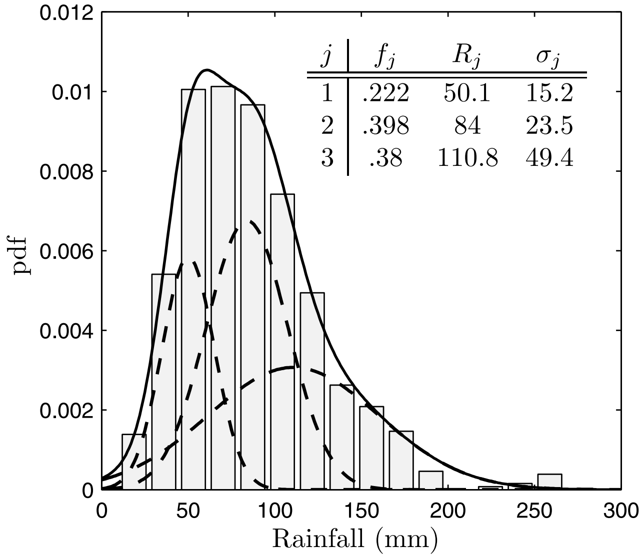

Figure 8. Histogram of St-Ephrem monthly rainfall based on data collected from 1929 to 2003. Superimposed (dashed lines) are the three Gaussian distributions whose sum (solid line) fits the histogram best and constitutes the historical prior $p(\mathbf{x})$ over the true rainfall.

parameters. In the last example, we used a uniform prior over the true rainfall to isolate the density mapping effect. We now use the same data and assumptions, except for a more realistic prior on the true inputs, an historical prior.

[68] To build this historical prior, we select a meteorological station (St-Ephrem) located near the watershed under study and plot an histogram of the monthly rainfall from 1929 to 2003. The best fitting distributions are Gamma, Weibull, and lognormal. However, if we choose to use any one of these distributions, the integral over $\mathbf{x}$ in equation (3) has no closed form solution. The trick is to fit the rainfall distribution by a sum of three Gaussian distributions (equation (15)), whose parameters are given in Figure 8.

[69] Although crude, this prior is adequate for our didactic purposes. In a real case study, however, it would be best to take seasonality into account. That is, specify a distinct prior for each month of the year. In the case where there is only one station on the watershed, the prior could also try to include the effect of spatially averaging the point rainfall input. The more relevant information that can be added to the prior, the better the calibration.

[70] The probability distribution of the parameters estimated using the historical prior is shown in Figure 9. Comparison with Figure 7 allows us to measure the difference made by the prior. The main change is in the storage parameter a, displaced to the left and achieving better agreement with the true parameters.

5.3. Model Filtering and Rainfall Smoothing

[71] An interesting observation from "abc" is that the model exhibits a form of filtering. By channeling rainfall in a storage compartment, the model is able to filter some of the noise imposed to the input data. That is, due to the time spent in storage, input errors are absorbed and averaged. This filtering makes the model relatively robust to non-biased input errors. Indeed, the noise imposed to the precipitation must be considerable before its effect becomes apparent. Following this line of thought, a watershed with a very fast response (small size, flash floods) should be more sensitive to input errors than a large watershed with a long response time.

[72] Our last observation concerns the averaging of parameters due to the historical prior. When a historical prior is chosen for the true values of precipitation, and the input error model has a large variance, the prior plays an important role in determining the distribution of probable true rainfall. Indeed, as the variance of the input error model increases, the distribution of the true rainfall approaches the prior distribution. In other words, the intrinsic variability of rainfall embodied by measurements is discarded in favor of the historical knowledge about its distribution. By this process, the data are "smoothed" toward the prior's mode, reflecting more and more the average rainfall, and thus the average behavior of the model. Since the parameters a and c have, on average, no influence (they affect only the timing of discharges), they become increasingly difficult to estimate.

5.4. Comparison With SLS

[73] As can be seen, there are some pitfalls to avoid when using the Bayesian uncertainty framework, and the question is whether or not the quality of the calibration is worth the extra effort. To answer that question, we compare results obtained by our method assuming input and output uncertainties, with those obtained assuming only output uncertainties (this is effectively SLS grafted with priors on parameters).

5.4.1. Data

[74] The data is a series of 100 monthly rainfall measurements from the Famine River and the synthetic discharge computed using the "abc" model. In this case, a Gaussian noise is added both on the input and output data. The noise is homoskedastic and devoid of autocorrelation, with standard variations of $\sigma_x = 25$ mm and $\sigma_y = 15$ mm for the input and output variables respectively.

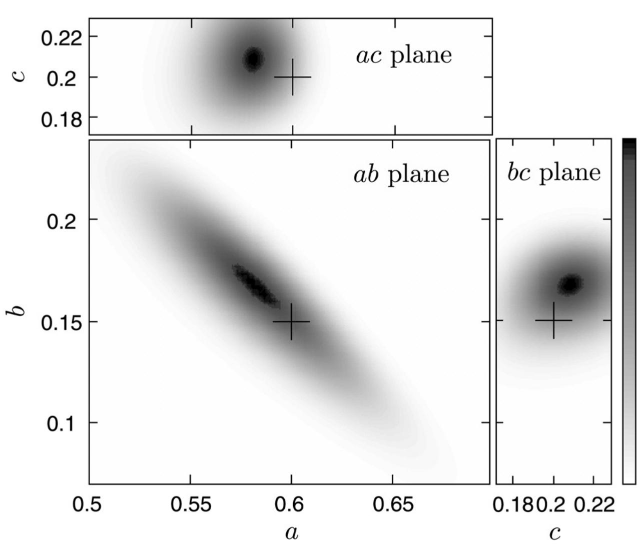

Figure 9. Impact of the prior for the true inputs. The slices are taken at the parameters most likely value [a, b, c] = [0.59, 0.16, 0.2]. The only difference with Figure 7 is the prior $p(x)$, now a historical prior instead of a uniform prior.

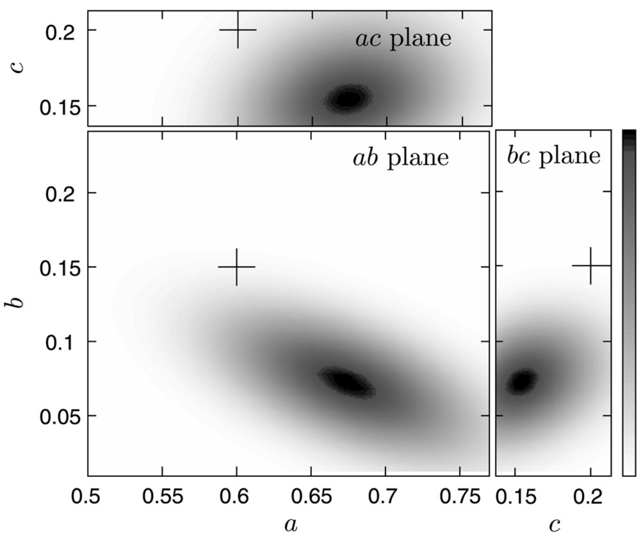

Figure 10. Parameters' posterior density, computed using only an output error model, thereby simulating a modified SLS algorithm. The slices are taken at the parameters most likely value [a, b, c] = [0.68, 0.07, 0.16].

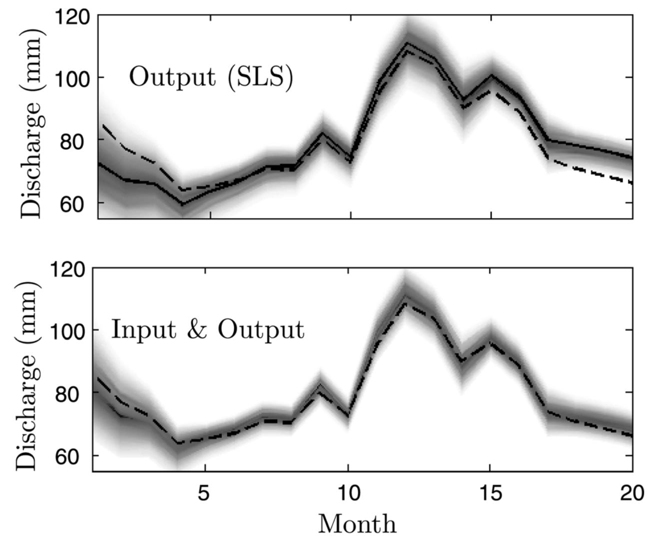

Figure 12. Comparison of predictions from two cases: (top) assuming only output errors and (bottom) assuming both input and output errors. The dashed line indicates the true synthetic discharge, and the solid line (SLS) indicates the discharge computed from the likeliest parameters.

5.4.2. Error Models and Prior Hypotheses

[75] The Bayesian method uses Gaussian error models with the correct covariance matrices ($\Sigma_\delta = 25^2 I$ mm^2 and $\Sigma_\epsilon = 15^2 I$ mm^2). Although the "SLS" method treats only the output error model, it is possible to include, partially, the effect of input errors on the parameters by defining an effective variance [*Orear*, 1982; *Lybanon*, 1984]. This effective variance is the sum of the output variance with the input variance multiplied by the model response: $\sigma_{SLS}^2 = \sigma_\delta^2(1 - b)^2 + \sigma_\epsilon$. Here $1 - b$ is the average response of the model, so that using $b = 0.15$, we obtain $\Sigma_{SLS} = 26^2 I$ mm^2.

[76] The prior on the true inputs is the historical prior defined in section 5.2. The prior on the parameters (defined in section 4.5) is identical for the Bayesian and SLS calibrations.

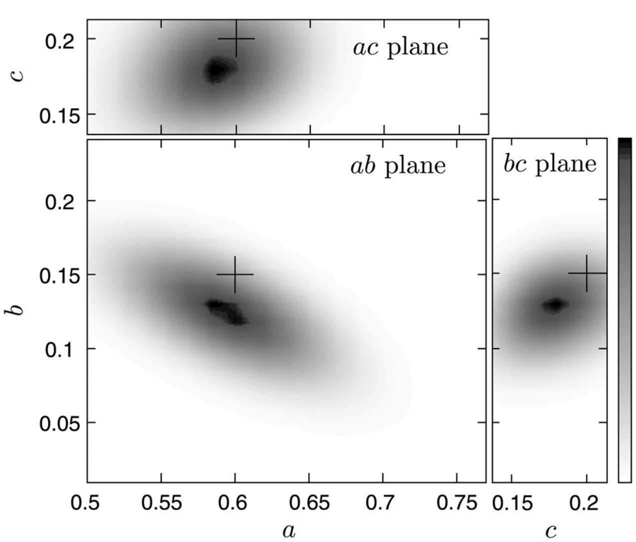

Figure 11. Parameters' posterior density, computed using an input and an output error model. The slices are taken at the parameters most likely value [a, b, c] = [0.59, 0.12, 0.18].

5.4.3. Results and Analysis

[77] The results are displayed in Figures 10 and 11. For this particular realization, the results obtained by the Bayesian method seem more accurate than those of SLS. Other simulations (not shown for lack of space) offer different pictures of the situation, but overall, the Bayesian density almost always presents superior results. What really matters, however, is not the parameters themselves but their prediction capability.

[78] To make predictions, we use as input 20 monthly rainfall values following the series used to calibrate the model. We uniformly sample the posterior density around 3000 times and simulate the discharge. Each discharge series is weighted by the probability of the parameter set, and a histogram of the discharge is computed. The results are linearly interpolated and displayed in Figure 12, where they can be compared to the true synthetic discharge (dashed line). The Bayesian predictions are closer to the synthetic discharge than those of SLS.

[79] The reader must be aware that these results constitute only one realization of a random Gaussian noise. The authors have conducted similar simulations where results are not so clear cut. In fact, when the noise is smaller, or when the data set is around n = 30, the predictions of SLS and of the Bayesian method are often very similar. Overall, it seems that Bayesian results are almost always better than those of SLS, the difference, however, may be negligible.

[80] Hence the decision of using the Bayesian method to account for input uncertainty depends upon two factors: (1) the impact of estimated input errors on model outputs and (2) the requirements of the end user. Since the difference between SLS and Bayesian estimates is directly related to the sensitivity of the model to input errors, the first step to assess the pertinence of a full Bayesian analysis is to evaluate the magnitude of input errors and their impact

on model predictions. If the impact is not significant with respect to output errors, then there is no need to include input uncertainties in the calibration process. If the impact is significant, the decision to use a Bayesian method relies on the needs of the end user. If the user only needs crude parameter estimates, there is no need to get fancy and a simple SLS will do. However, if the issue is sensitive and the reliable assessment of uncertainties plays a crucial role in the decision making process, using a method that considers input uncertainties would be preferable.

5.5. Comparison With *Kavetski et al.* [2002]

[81] Our study of the impact of input uncertainty was motivated in part by the article of *Kavetski et al.* [2002], on which we want to comment. The article presents a Bayesian method, BATEA, Bayesian Total Error Analysis, to account for input uncertainties. BATEA is applied to "abc", and the performances of BATEA and SLS are compared. The conclusion is that in the presence of input errors, SLS provides precisely wrong parameter estimates while BATEA gives probably right estimates. Indeed, the posterior pdf computed using BATEA are dead centered on the parameter's true values, a performance we could not replicate. We think this can be explained by the differences between the error models. In their case, the "abc" model is used with hourly time steps, where errors are likely to be highly correlated. Hence, instead of corrupting individual measurements, *Kavetski et al.* [2002] divide the rainfall history into storms, and multiply all the precipitations in a given storm by the same random factor. For example, for a time series of 1000 hourly precipitation, there may be five storm multipliers. These storm multipliers are then treated as parameters and estimated, rather than integrated, via a Bayesian analysis using MCMC algorithms.

[82] The method of *Kavetski et al.* [2002] is effective with large data sets and a relatively low number of nuisance variables. The method presented in this article deals with small data sets where each measurement has its own nuisance variable (the true value). In a sense, the two methods apply to different situations and cannot be easily compared in terms of performance or accuracy.

6. Summary

[83] The article describes a Bayesian framework to account for input, output and structural uncertainties in the calibration of models. The method is applied to two models: a simple straight line passing through the origin and the hydrological model "abc." In both cases, the main object of study is the impact of input uncertainties on the estimation of the model parameters. The analysis of those models has led us to some interesting observations concerning the calibration of models in the presence of input uncertainty.

[84] 1. The prior on the true inputs plays a major role. Contrary to the usual belief that the influence of the prior vanishes as data accumulate, the prior on the true inputs has a significative influence on the parameters estimation even for large samples. This is due to the fact that every datum added to the set comes with an additional prior, contributing to the shape of the posterior distribution.

[85] 2. Error models need to be carefully specified. The error models should be as realistic as possible, since underestimating the input error can bias the results, while overestimating the input error leads to the smoothing of inputs, averaging of the model behavior and difficulties in parameter estimation.

[86] 3. Biases are meaningful. The uncertainty on the parameter combines the output uncertainty and the input uncertainty via the model structure. Hence a bias on the parameter posterior density does not necessarily mean that the calibration is flawed, but rather that the "true" parameters are not those for which the measured outputs are the most probable. This displacement of the parameter posterior from the true values is related to how the model maps uncertain inputs onto outputs. In classical statistics, such displacements of the parameter's density maximum are called biases and regarded with suspicion. In the Bayesian framework, the origin of this bias is understandable, and we see no motivation to make corrections in order to obtain an "unbiased" estimator.

[87] 4. The entire posterior is significant. The fact that biases are commonplace in input error context underlines the importance of using the whole posterior distribution to make predictions, and not only the maximum of the distribution. By sampling the parameter's posterior to make predictions, we insure that the prediction uncertainty is faithful to the parameter uncertainty, and reflects the data and structural uncertainty.

7. Conclusion

[88] The uncertainty framework presented in the paper is general and theoretically, can be applied to any problem. The application to the "abc" model, however, made use of the fact that the model is linear in its input and that equation (3) could be integrated analytically. In general, hydrological models are nonlinear in their inputs, and such direct integration is impossible. The resolution method used here is thus inadequate for virtually all commonly used hydrological models. The usefulness of the uncertainty framework in applied hydrology is hence questionable until an algorithm is developed to apply it to nonlinear models. This is the subject of the authors' current research.

[89] The article bypasses the entire question of the hyperparameters. Indeed, in all simulations, we posed as known the error model parameters (noise variance, mean and correlation). In real cases, those hyperparameters must be estimated, or integrated as nuisance variable. The treatment of hyperparameters, computationally difficult, would have overshadowed the focus of the paper, treatment of input uncertainty. It is clear, however, that realistic applications of the method will have to tackle this issue.

[90] Another simplification concerns the state uncertainties. In the case of "abc", the model state is embodied by S_1, the initial storage. In our simulations, we took the initial storage as known, in order to simplify an already lengthy equation. A brief inspection, however, will convince the reader that it is relatively easy to integrate equation (18) with respect to S_1. Thus, for a uniform prior on the initial storage, an analytical formula for the parameters posterior density can be derived. The effect of our ignorance of the initial storage is then included into the uncertainty on the parameters.

[91] If the uncertainty framework can be successfully applied to nonlinear models, two issues will have to be addressed to make full use of its potential. The first one is

the specification of a rigorous structural error model. This is already a recognized open problem in hydrology. The second issue is the description of a detailed prior on the true values, taking scaling effects, seasonality and correlation into account. While these effects have been studied separately, their integration into a unique prior remains to be done.

[92] We hope by this article to promote the realistic assessment of uncertainties and their integration into model calibration methods. It is important to acknowledge the complexity of the calibration process and its importance for reliable predictions. The credibility of science is not built upon its vast knowledge, but upon its honesty in front of ignorance and uncertainty.

Appendix A: Calculations

[93] Here we detail the derivation of the posterior pdf for the parameters θ of the ''abc'' model. Starting with equation (16) and plugging the input and output error models along with the historical prior, we find

$$p(\tilde{\mathbf{y}}|\tilde{\mathbf{x}},\theta,\mathcal{M}) = \int_{\mathbb{R}} \mathcal{N}(\tilde{\mathbf{y}}|\mathsf{A}\mathbf{x}+S_1\mathbf{t},\Sigma_\epsilon)\mathcal{N}(\tilde{\mathbf{x}}|\mathbf{x},\Sigma_\delta) \cdot \left[\sum_{j=1}^{3} f_j \mathcal{N}\left(\mathbf{x}|\mathbf{R}_j,\Sigma_{R_j}\right)\right] \mathrm{d}\mathbf{x} \quad \text{(A1)}$$

In order to apply the identity (17), we need to express the Gaussians as functions of **x**:

$$\mathcal{N}(\tilde{\mathbf{y}}|\mathsf{A}\mathbf{x}+S_1\mathbf{t},\Sigma_\epsilon) = \mathcal{N}\left(\mathbf{x}|\mathsf{A}^{-1}(\tilde{\mathbf{y}}-S_1\mathbf{t}),\mathsf{A}^{-1}\Sigma_\epsilon\mathsf{A}^{-1'}\right)$$
$$\mathcal{N}(\tilde{\mathbf{x}}|\mathbf{x},\Sigma_\delta) = \mathcal{N}(\mathbf{x}|\tilde{\mathbf{x}},\Sigma_\delta)$$

Substituting those last expressions into equation (A1) and writing it as the sum of a product of Gaussian, we can apply the identity for each element of the sum

$$p(\tilde{\mathbf{y}}|\tilde{\mathbf{x}},\theta,\mathcal{M}) = \sum_{j=1}^{3} f_j \int_{\mathbb{R}} \mathcal{N}(\mathbf{x}|\mu_1,\Sigma_1) \cdot \mathcal{N}(\mathbf{x}|\mu_2,\Sigma_2)\mathcal{N}\left(\mathbf{x}|\mu_{3j},\Sigma_{3j}\right) \mathrm{d}x,$$

where

$$\mu_1 = \mathsf{A}^{-1}(\tilde{\mathbf{y}}-S_1\mathbf{t}) \quad \Sigma_1 = \mathsf{A}^{-1}\Sigma_\epsilon\mathsf{A}^{-1'}$$
$$\mu_2 = \tilde{\mathbf{x}} \quad \Sigma_2 = \Sigma_\delta$$
$$\mu_{3j} = \boldsymbol{R}_j \quad \Sigma_{3j} = \Sigma_{Rj}.$$

[94] **Acknowledgments.** The authors would like to thank the Associate Editor and reviewers for their insightful comments. David Huard is grateful for the financial support of the Natural Sciences and Engineering Research Council of Canada (NSERC Postgraduate Scholarships).

References

Andreassian, V., C. Perrin, C. Michel, I. Usart-Sanchez, and J. Lavabre (2001), Impact of imperfect rainfall knowledge on the efficiency and the parameters of watershed models, *J. Hydrol.*, *250*(1–4), 206–223.

Bernardo, J., and A. Smith (1994), *Bayesian Theory*, John Wiley, Hoboken, N. J.

Beven, K., and A. Binley (1992), The future of distributed models: Model calibration and uncertainty prediction, *Hydrol. Processes*, *6*, 279–298.

Cheng, C.-L., and J. W. V. Ness (1994), On estimating linear relationships when both variables are subject to errors, *J. R. Stat. Soc., Ser. B*, *56*(1), 167–183.

Clark, J. (2005), Why environmental scientists are becoming Bayesians, *Ecol. Lett.*, *8*(1), 2–14.

Efron, B. (1986), Why isn't everyone a Bayesian?, *Am. Stat.*, *40*(1), 1–5.

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

Erickson, T. (1989), Proper posteriors from improper priors for an unidentified errors-in-variables model, *Econometrica*, *57*(6), 1299–1316.

Fiering, M. B. (1967), *Streamflow Synthesis*, Harvard Univ. Press, Cambridge, Mass.

Fuller, W. A. (1987), *Measurement Error Models*, John Wiley, Hoboken, N. J.

Gelman, A., J. B. Carlin, H. S. Stern, and D. B. Rubin (1995), *Bayesian Data Analysis*, 526 pp., CRC Press, Boca Raton, Fla.

Gull, S. F. (1989), Bayesian data anaysis: Straight-line fitting, in *Maximum Entropy and Bayesian Methods*, edited by J. Skilling, pp. 511–518, Springer, New York.

Gupta, H. V., S. Sorooshian, and P. Yapo (1998), Towards improved calibration of hydrologic models: Multiple and non-commensurable measures of information, *Water Resour. Res.*, *34*(4), 751–764.

Gupta, H. V., S. Sorooshian, T. S. Hogue, and D. P. Boyle (2003), Advances in automatic calibration of watershed models, in *Calibration of Watershed Models*, *Water Sci. Appl. Ser.*, vol. 6, edited by Q. Duan et al., pp. 9–28, AGU, Washington, D. C.

Jaynes, E. T. (1983), *E.T. Jaynes: Papers on Probability, Statistics, and Statistical Physics*, edited by R. D. Rosenkrantz, Hingham, Boston, Mass.

Jaynes, E. T., and G. L. Bretthorst (2003), *Probability Theory: The Logic of Science*, Cambridge Univ. Press, New York.

Kass, R. E., and L. Wasserman (1996), The selection of prior distributions by formal rules, *J. Am. Stat. Assoc.*, *91*(435), 1343–1370.

Kavetski, D., S. W. Franks, and G. Kuczera (2002), Confronting input uncertainty in environmental modeling, in *Calibration of Watershed Models*, *Water Sci. Appl. Ser.*, vol. 6, edited by Q. Duan et al., pp. 49–68, AGU, Washington, D. C.

Kendall, M., and A. Stuart (1983), Functional and structural relationships, in *The Advanced Theory of Statistics*, vol. 2, 4th ed., chap. 29, Oxford Univ. Press, New York.

Kennedy, M. C., and A. O'Hagan (2001), Bayesian calibration of computer models, *J. R. Stat. Soc., Ser. B*, *63*, 425–450.

Lindley, D. V., and G. M. El-Sayyad (1968), The Bayesian estimation of a linear functional relationship, *J. R. Stat. Soc., Ser. B*, *30*(1), 190–202.

Lybanon, M. (1984), A better least-squares method when both variables have uncertainties, *Am. J. Phys.*, *52*(1), 22–26.

Mein, R., and B. Brown (1978), Sensitivity of optimized parameters in watershed models, *Water Resour. Res.*, *14*(2), 299–303.

Moradkhani, H., K.-L. Hsu, H. Gupta, and S. Sorooshian (2005), Uncertainty assessment of hydrologic model states and parameters: Sequential data assimilation using the particle filter, *Water Resour. Res.*, *41*, W05012, doi:10.1029/2004WR003604.

Orear, J. (1982), Least squares when both variables have uncertainties, *Am. J. Phys.*, *50*(10), 912–916.

Oudin, L., F. Hervieu, C. Michel, C. Perrin, V. Andreassian, F. Anctil, and C. Loumagne (2005), Which potential evapotranspiration input for a lumped rainfall-runoff model?: Part 2—Towards a simple and efficient potential evapotranspiration model for rainfall-runoff modelling, *J. Hydrol.*, *303*(1–4), 290–306.

Reed, B. C. (1989), Linear least-squares fits with errors in both coordinates, *Am. J. Phys.*, *57*(7), 642–646.

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*(2), 430–442.

Troutman, B. M. (1982), An analysis of input errors in precipitation-runoff models using regression with errors in independent variables, *Water Resour. Res.*, *18*(4), 947–964.

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., H. Gupta, W. Bouten, and S. Sorooshian (2003), A shuffled complex evolution Metropolis algorithm for optimization and uncertainty assessment of hydrologic model parameters, *Water Resour. Res.*, *39*(8), 1201, doi:10.1029/2002WR001642.

Vrugt, J. A., C. G. H. Diks, H. V. Gupta, W. Bouten, and J. 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.

Werman, M., and D. Keren (2001), A Bayesian method for fitting parametric and nonparametric models to noisy data, *IEEE Trans. Pattern Anal. Mach. Intell.*, *23*(5), 528–534.

York, D. (1966), Least-squares fitting of a straight line, *Can. J. Phys.*, *44*, 1079–1086.

Zellner, A. (1971), *An Introduction to Bayesian Inference in Econometrics*, John Wiley, Hoboken, N. J.

D. Huard and A. Mailhot, Centre Eau, Terre et Environnement, Institut National de la Recherche Scientifique, Québec, QC, Canada G1K 9A9. (david.huard@ete.inrs.ca; alain.mailhot@ete.inrs.ca)