Calibrating a hydrological model in stage space to account for rating curve uncertainties: general framework and key challenges

A.E. Sikorska, Benjamin Renard

Calibrating a hydrological model in stage space to account for rating curve uncertainties: general framework and key challenges

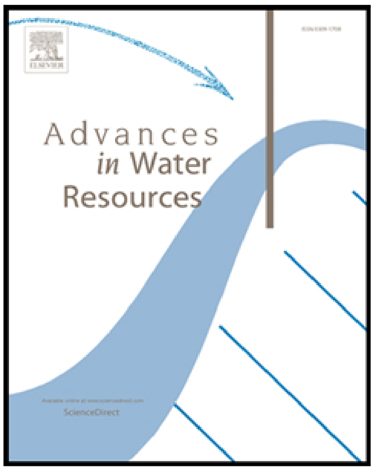
Sikorska, Anna E ; Renard, Benjamin

Abstract: Hydrological models are typically calibrated with discharge time series derived from a rating curve, which is subject to parametric and structural uncertainties that are usually neglected. In this work, we develop a Bayesian approach to probabilistically represent parametric and structural rating curve errors in the calibration of hydrological models. To achieve this, we couple the hydrological model with the inverse rating curve yielding the rainfall–stage model that is calibrated in stage space. Acknowledging uncertainties of the hydrological and the rating curve models allows assessing their contribution to total uncertainties of stages and discharges. Our results from a case study in France indicate that (a) ignoring rating curve uncertainty leads to changes in hydrological parameters, and (b) structural uncertainty of hydrological model dominates other uncertainty sources. The paper ends with discussing key challenges that remain to be addressed to achieve a meaningful quantification of various uncertainty sources that affect hydrological model, as including input errors.

DOI: https://doi.org/10.1016/j.advwatres.2017.04.011

Originally published at:
Sikorska, Anna E; Renard, Benjamin (2017). Calibrating a hydrological model in stage space to account for rating curve uncertainties: general framework and key challenges. Advances in Water Resources, 105:51-66.
DOI: https://doi.org/10.1016/j.advwatres.2017.04.011

Accepted Manuscript



Calibrating a hydrological model in stage space to account for rating curve uncertainties: general framework and key challenges

Anna E. Sikorska, Benjamin Renard

PII: S0309-1708(16)30391-8
DOI: 10.1016/j.advwatres.2017.04.011
Reference: ADWR 2826

To appear in: *Advances in Water Resources*

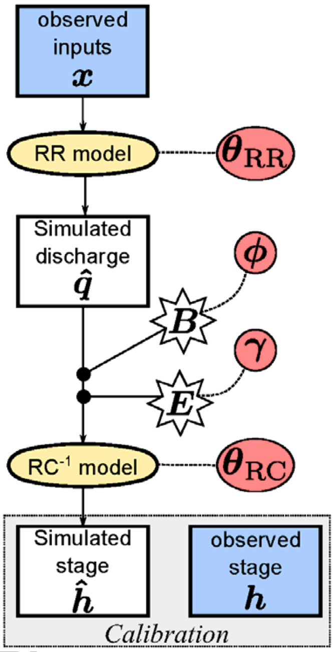
Received date: 31 August 2016
Revised date: 12 April 2017
Accepted date: 13 April 2017

Please cite this article as: Anna E. Sikorska, Benjamin Renard, Calibrating a hydrological model in stage space to account for rating curve uncertainties: general framework and key challenges, *Advances in Water Resources* (2017), doi: 10.1016/j.advwatres.2017.04.011

Highlights

- Method for quantifying rating curve uncertainties in discharge prediction is proposed
- A rainfall-stage model is developed and calibrated in stage space
- Such a rainfall-stage model couples a hydrological model with an inverse rating curve
- We consider both structural and parametric uncertainties of the rating curve
- Shares of these errors in the total uncertainty of stages and discharges are assessed
- Structural uncertainties of hydrological model dominates other uncertainty sources
- Ignoring rating curve errors affects the estimation of hydrological model parameters

Graphical Abstract



Calibrating a hydrological model in stage space to account for rating curve uncertainties: general framework and key challenges

Anna E. Sikorska[a,b,c,*], Benjamin Renard[a]

[a]*Irstea, UR HHLY, Hydrology-Hydraulics, 5 rue de la Doua CS70077, 69626 Villeurbanne Cedex, France*
[b]*University of Zurich, Department of Geography, Winterthurerstr. 190, 8057 Zürich, Switzerland*
[c]*Warsaw University of Life Sciences – SGGW, Department of Hydraulic Engineering, Nowoursynowska 166, 02-787 Warsaw, Poland*

Abstract

Hydrological models are typically calibrated with discharge time series derived from a rating curve, which is subject to parametric and structural uncertainties that are usually neglected. In this work, we develop a Bayesian approach to probabilistically represent parametric and structural rating curve errors in the calibration of hydrological models. To achieve this, we couple the hydrological model with the inverse rating curve yielding the rainfall-stage model that is calibrated in stage space. Acknowledging uncertainties of the hydrological and the rating curve models allows assessing their contribution to total uncertainties of stages and discharges. Our results from a case study in France indicate that a) ignoring rating curve uncertainty leads to changes in hydrological parameters, and b) structural uncertainty of hydrological model dominates other uncertainty sources. The paper ends with discussing key challenges that remain to be addressed to achieve a meaningful quantification of various uncertainty sources that affect hydrological model, as including input errors.

Keywords: rating curve, rainfall-stage model, structural uncertainty, parametric uncertainty, Bayesian inference, hydrological modeling

*Corresponding author at: University of Zurich, Department of Geography, Winterthurerstr. 190, 8057 Zürich, Switzerland
Email address: as@annasikorska.eu (Anna E. Sikorska)

1. Introduction

1.1. The importance of rating curve uncertainty in hydrological modeling

Flood risk analysis relies on estimates of hydrological models and associated uncertainty [1, 2, 3]. This uncertainty results mainly from four components: (i) parametric uncertainty of the hydrological model, (ii) its limited approximation of the catchment hydrological processes (model structural error), (iii) uncertainty in external model inputs (typically rainfall, temperature or evapotranspiration), and in (iv) output calibration data (typically discharge series) [4, 5, 6, 7, 3, 8, 9].

Among these four uncertainty contributors, input errors are considered to be one of the major uncertainty sources in hydrological models [10, 11, 12, 3] and thus more research has been devoted to investigate their effect on hydrological predictions than the effect of output uncertainty. Hence, different techniques have been proposed to represent input uncertainty which include a rainfall multiplier approach [10, 11, 13, 3], an addition to the bias [14, 8], or a more advanced stochastic description [15]. All of these studies, however, indicated that the inclusion of input errors raises several challenges. First, the computational cost is much higher than with traditional calibration. But even more importantly, substantial difficulties arise from the interaction between input errors and other uncertainty components. For instance, Renard et al. [6] discussed the challenge of identifying both input and structural errors; similarly, Del Giudice et al. [15] reported difficulties in distinguishing between different observational errors (input and output) if they have similar properties, i.e., are systematic. Hence, in this study we do not describe input errors explicitly, to be able to focus entirely on the effect of output uncertainty (due to the rating curve) on calibration and prediction of a hydrological model. Input errors will be implicitly encompassed in the structural error of the hydrological model.

As opposed to input errors, less attention has been given to the output uncertainty which is often assumed to be relatively small in comparison to the other three parts and thus has been evenly neglected in uncertainty analysis frameworks [16, 17]. Such a strong assumption might be justified for a direct measurement of discharge, for which measurement errors of 5% on average could be assumed [18, 19]. For practical applications, however, measuring discharge continuously becomes impossible [20, 21]. Instead, a measure of discharge is obtained from an observed stage using a stage - discharge relationship (called rating curve) [22, 23]. This relationship needs to be established at a hydrometric station with few direct (discrete) measurements of gauging pairs (stage and discharge) [18]. Using pre-established rating curves to compute discharges therefore allows deriving continuous quasi-observed discharge series [22, 23], which next serve for calibration of hydrological models [24, 25].

Awkwardly, these computed discharge series are often communicated to modellers or practitioners without any uncertainty statement [22, 26]. It is however clear that such estimated discharge series contain several errors. It has been reported in literature that although these errors are on average about $3-6\%$ of an estimated value, they may increase to about 20% under poor measurement conditions [27], and to more than 25% outside the range of measured stage-discharge pairs [16, 28]. However, the level of these errors is case-specific [29] and results from many sources: measurement errors of gauging pairs (instrumental errors, measurement technique), temporal shifts in the rating curve (unstable stream channel due to vegetation, bank erosion, sediment deposition, ice jams, etc.), transient hydrological conditions during measurement of gauging pairs, hysteresis effect, and rating curve parametric and structural uncertainties [12, 30, 31, 26, 32].

All these errors affect calibration of the hydrological model and have serious implications for discharge simulations [12, 26, 23], flood frequency analysis [33, 34, 35], and for regionalization of model parameters [36]. As these errors are often not explicitly considered in uncertainty estimation, their effect on discharge uncertainty cannot be quantified. Moreover, when fully neglected, the uncertainty caused by rating curve errors may be wrongly attributed to other uncertainty source(s), leading to biased estimates that might be misunderstood by practitioners [35]. Given the above considerations and the number of studies dealing with calibration of hydrological models based on such quasi-observed discharge series, an accurate assessment of the rating curve uncertainties and their impact on the hydrological model becomes essential for flood risk assessment and management.

1.2. Existing approaches to describe rating curve uncertainty

Although a number of recent studies have investigated different aspects of rating curve uncertainties [20, 22, 37, 24, 38, 30, 39], the contribution of the rating curve to the uncertainty in hydrological simulations has not been assessed systematically so far. In many uncertainty frameworks, rating curve errors are either not explicitly represented or are combined with other error sources. For instance, a common practice in uncertainty analysis is to pool all uncertainties (apart from parametric uncertainty but including rating curve uncertainty) into a lumped error term, which properties need to be mathematically described [40]. We call the latter approach when only parametric and structural errors of hydrological model are represented and the model is calibrated against discharges computed from rating curves as *a standard uncertainty estimation approach*. Another possible solution is mapping all uncertainty sources (including rating curve errors) to parameter uncertainty as in the original GLUE (generalized likelihood uncertainty estimation) methodology [41]. Further developments allowed to relate "limits of acceptability" with the rating curve uncertainty, although the need to extend these limits to account for other error sources (input errors in particular) was recognized [42]. Other approaches allow distinguishing input and structural errors [43, 11]. However, they don't explicitly represent rating curve errors, which are hence implicitly merged with structural errors. Finally, a recently introduced bias addition approach [14] gives the possibility to distinguish, aside from the parametric uncertainty, two different structural error types of the hydrological model, i.e., systematic and random errors. These errors are interpreted as structural and observational errors respectively. The bias approach pools however all observational errors (i.e., input and output) together and thus the uncertainty linked to the rating curve cannot be assessed. Hence, the major drawback of all these different approaches available to assess uncertainty of hydrological models is their inability to quantify the uncertainty contribution of the rating curve in total uncertainty estimates of hydrological models.

One possibility to indirectly tackle rating curve uncertainty is to propagate rating curve errors to discharge series which are then represented as spaghetti lines or uncertainty bands [44]. Such multiple realizations of discharge series yield however a practical question of how to calibrate a hydrological model with hundreds of "observed" discharges.

As an alternative, Sikorska et al. [26] and Thyer et al. [45] have recently proposed to avoid the issue of multiple "observed" discharges by simulating directly stages instead of discharges. Thus, they proposed to couple the hydrological model with the inverse rating curve yielding a so-called *rainfall-stage* model, for which uncertainty was evaluated in the stage space. In this way, rating curve uncertainty could be directly incorporated into simulations of the hydrological model and the contribution of the rating curve uncertainty could be assessed. Yet, this method was mostly suitable to estimate stages while it was lacking the possibility to provide discharge predictions along with their uncertainty estimates (as discharge was only an intermediate step and was not directly modelled). Moreover, only the assessment of the parametric rating curve uncertainty was possible, while the structural errors of the rating curve could not be separated from those of the hydrological model.

Finally, other authors proposed specific error models to describe rating curve errors, based on an analysis of the rating curve itself [24]. Thyer et al. [37] and Renard et al. [12] proposed a specific error model within the Bayesian total error analysis methodology (BATEA) of Kavetski et al. [43, 10], to represent structural errors of rating curves in discharge data along other uncertainty components (input and structural errors of hydrological model). In this way, contributions of those three main uncertainty components could be evaluated. Yet, they did not make an explicit distinction between parametric and structural uncertainties of rating curves, pooling all rating curve errors into a lumped structural error.

1.3. Objectives

Therefore, within this work, we further advance uncertainty quantification of rating curves by developing a Bayesian approach to probabilistically represent rating curve errors in the estimation of the hydrological model. In contrast to previous works, for the first time, we explicitly represent the parametric and the structural uncertainties of both the hydrological and the rating curve models. To achieve this, we couple the hydrological model with the inverse rating curve yielding the rainfall-stage model that can be calibrated in stage space, as previously proposed by Sikorska et al. [26]. Specifically, we describe structural errors of the hydrological model as an Ornstein-Uhlenbeck process [46] in the form implemented by Sikorska et al. [47], and the structural errors of the rating curve as Gaussian errors with a zero mean and a standard deviation proportional to the discharge value following the BaRatin method [23]. Because of such an explicit consideration of different uncertainty components of the rating curve and the hydrological model, the coupled total error can be decomposed into its constitutive sources. Hence, the approach is suitable for providing both stage and discharge simulations along with their associated uncertainties.

Specifically, we formulate the following objectives for this study:

1. Propose a generic framework for quantifying parametric and structural uncertainties of rating curves in hydrological models, and derive the corresponding inference equations;
2. Examine the effects of ignoring a specific source of rating curve uncertainty (parametric or structural) in the inference of model parameters and in model simulations;
3. Discuss pros and contras of using an advanced calibration approach (representing both structural and parametric rating curve errors explicitly) over a "standard" uncertainty estimation approach (when uncertainty is attributed only to parametric and structural errors of the hydrological model and uncertainties of rating curve are neglected).

Our approach is developed and tested on a medium-size study catchment in France. This study restricts its attention solely to investigate uncertainties in output (discharge) of hydrological models, while uncertainty in input data (typically rainfall), although non negligible, is not explicitly acknowledged and is implicitly represented in structural errors of the hydrological model. We debate possible consequences of this assumption in the discussion part. Moreover, we recognize that an explicit and reliable treatment of all error sources remains a key challenge for hydrologic modeling: while not the objective of this paper, we also discuss this long-term objective in section 5.5.

2. Uncertainty representation

2.1. Rating curve

2.1.1. Rating curve model

We describe an instantaneous discharge at time t predicted with the rating curve (RC), $\breve{q}_t$, as

$$\breve{q}_t = f_{\mathrm{RC}}\left(h_t, \boldsymbol{\theta}_{\mathrm{RC}}\right) \tag{1}$$

where $f_{\mathrm{RC}}\left(h_t, \boldsymbol{\theta}_{\mathrm{RC}}\right)$ is the deterministic RC equation, h_t is the instantaneous stage at time t and $\boldsymbol{\theta}_{\mathrm{RC}} = (\theta_{\mathrm{RC}_1}, ..., \theta_{\mathrm{RC}_w})$ are parameters of the RC. Because parameters of the RC are unknown, they must be calibrated and thus they will introduce parametric uncertainty to the RC (see Sect. 3.1 for description of model calibration).

2.1.2. Structural error

The rating curve equation is a simplified mathematical representation of the true stage-discharge relationship prevailing at the gauging station. We therefore introduce a structural error E_t to describe the difference between the RC-predicted discharge $\breve{q}_t$ and the (unknown) true discharge q_t:

$$\breve{q}_t = q_t + E_t(\boldsymbol{\gamma}) \tag{2}$$

The structural error E_t is assumed to be a realization from a Gaussian distribution with mean zero and standard deviation varying with the RC-predicted discharge as parameterized below:

$$E_t \overset{indep}{\sim} N\left(0, g(\breve{q}_t, \boldsymbol{\gamma})^2\right); \quad g(\breve{q}_t, \boldsymbol{\gamma}) = \gamma_1 + \gamma_2 \cdot \breve{q}_t \tag{3}$$

where $\boldsymbol{\gamma} = (\gamma_1, \gamma_2)$ are the unknown parameters of the RC structural error model. This equation calls for the following comments:

1. The assumption that the standard deviation of structural errors is an affine function of the RC-predicted discharge is made to account for heteroscedasticity, which is often observed in practice (see e.g. [22, 23]). A homoscedastic model can easily be obtained by fixing $\gamma_2 = 0$. Conversely, more complex heteroscedasticity models can in principle be derived by replacing the affine function g by another function (e.g. an higher-order polynomial), at the cost of introducing more unknown parameters;
2. Since the true discharge q_t is unknown, we assume that the standard deviation of structural errors is a function of the RC-predicted discharge $\breve{q}_t$;
3. Eq. 3 also makes the strong assumption that structural errors are independent in time. This will be further discussed in section 5.2.

2.1.3. Gauging measurement error

The RC is typically calibrated using gaugings, i.e., pairs of stage-discharge values measured at different stage levels and flow conditions [48, 49, 50]. The measurement error on stage is assumed to be negligible. Conversely, the measurement error on the gauged discharge can be considerable. Hence, we represent the gauged discharge observed at time t, $\tilde{q}_t$, as the sum of the true discharge q_t and a measurement error W_t:

$$\tilde{q}_t = q_t + W_t \tag{4}$$

The measurement error W_t is further assumed to be a realization from a Gaussian distribution with mean zero and known standard deviation δ_t:

$$W_t \overset{indep}{\sim} N\left(0, \delta_t^2\right), \tag{5}$$

This equation calls for the following comments:

1. We assume that δ_t is known because the uncertainty of the gauged discharge can be quantified before RC estimation by analyzing the measurement process (see e.g. [51, 52, 23]). Note that each gauging has its specific uncertainty;
2. As for structural errors, eq. 5 also makes the assumption that measurement errors are independent in time. However this assumption is probably much more realistic here.

2.2. Hydrological model

2.2.1. Rainfall-runoff model

For simplicity sake, we prefer to substitute the hydrological model with a rainfall-runoff model which abbreviates to RR since h notation is restricted for stage and thus could be confused with the abbreviation of a hydrological model. We represent a RR-predicted discharge at time t, $\hat{q}_t$, as:

$$\hat{q}_t = f_{\mathrm{RR}}\left(\boldsymbol{x}_{1:t}, \boldsymbol{\theta}_{\mathrm{RR}}\right) \tag{6}$$

where $f_{\mathrm{RR}}\left(\boldsymbol{x}_{1:t}, \boldsymbol{\theta}_{\mathrm{RR}}\right)$ represent the deterministic RR equations, $\boldsymbol{x}_{1:t}$ are inputs time series up to time t and $\boldsymbol{\theta}_{\mathrm{RR}} = (\theta_{\mathrm{RR}_1}, ..., \theta_{\mathrm{RR}_z})$ are the parameters. Note that for simplicity this notation makes initial conditions implicit. Similarly to the parameters of RC, parameters of the RR are unknown and they must be estimated from observations. Hence they will introduce parametric uncertainty to the RR model (see further Sect. 3.2 describing model calibration).

2.2.2. Structural error

To account for the imperfect nature of the RR model, a structural error B_t is introduced to describe the mismatch between the RR-predicted discharge and the (unknown) true discharge q_t:

$$\psi(\hat{q}_t) = \psi(q_t) + B_t\left(\boldsymbol{\phi}\right) \tag{7}$$

where $\psi(\cdot)$ is a transformation function applied to the true and the RR-predicted discharges (typically, a Box-Cox transformation, see appendix section Appendix A). The aim of this transformation is to make the probabilistic model used to describe B_t (described next) more realistic.
In order to explicitly describe the autocorrelated nature of structural errors,

B_t is represented as an Ornstein-Uhlenbeck (OU) process [46] with parameters $\boldsymbol{\phi} = (\phi_1, \phi_2)$.

$$B_t \sim OU\left(\phi_1, \phi_2\right) \tag{8}$$

The OU process is a continuous-time equivalent of more standard time series models such as the autoregressive (AR) error model, which are only defined for data sampled at regular discrete times. Such a continuous-time model allows dealing with unequally spaced data, which are commonly used for routine monitoring of instantaneous water stage or discharge (typically, more frequent records during floods than during low flows). We choose the correlation structure of B_t in such a way that it becomes similar to the AR(1) model [14, 47] with the variance at time t_i conditioned on a previous time step t_j being equal to:

$$Var(B_{t_{i|j}}) = \phi_1^2 \cdot \left(1 - exp\left(-\frac{2 \cdot |t_i - t_j|}{\phi_2}\right)\right) \tag{9}$$

ϕ_1 can be interpreted as the asymptotic standard deviation (for infinitely-spaced time points) and ϕ_2 is a characteristic correlation time.

2.3. Rainfall-stage model

The basic idea behind the construction of the rainfall-stage (RS) model is to apply the inverse of the RC to the discharge simulated by the RR model [26]. The advantage of such a RS model is that its parameters encompass both the RR and the RC parameters, which allows explictly accounting for RC uncertainty in the calibration of the RR parameters. However, the structural errors affecting both the RR and the RC models propagate to the RS model and therefore need to be accounted for, as described next.

2.3.1. Structural error

Let h_t denote the true stage value at time t. From the RC model eqs. 1 and 2 we get:

$$f_{\mathrm{RC}}\left(h_t, \boldsymbol{\theta}_{\mathrm{RC}}\right) = q_t + E_t(\boldsymbol{\gamma}) \tag{10}$$

Inverting the RC therefore yields the following relation:

$$h_t = f_{\mathrm{RC}}^{-1}\left(q_t + E_t(\boldsymbol{\gamma}), \boldsymbol{\theta}_{\mathrm{RC}}\right) \tag{11}$$

Moreover from the RR structural error model eq. 7 we get:

$$q_t = \psi^{-1}\left(\psi(\hat{q}_t) - B_t\left(\boldsymbol{\phi}\right)\right) \tag{12}$$

where $\psi(\cdot)$ and $\psi^{-1}(\cdot)$ are the forward and the backward transformation. Combining eqs. 11 and 12, the true instantaneous stage at time t can be written as:

$$h_t = f_{\mathrm{RC}}^{-1}\left(\psi^{-1}\left[\psi\left(\underbrace{f_{\mathrm{RR}}\left(\boldsymbol{x}_{1:t}, \boldsymbol{\theta}_{\mathrm{RR}}\right)}_{\text{RR model}}\right) - \underbrace{B_t\left(\boldsymbol{\phi}\right)}_{\text{RR structural error}}\right] + \underbrace{E_t\left(\boldsymbol{\gamma}\right)}_{\text{RC structural error}}, \boldsymbol{\theta}_{\mathrm{RC}}\right) \tag{13}$$

We stress that the structural error model described in eq. 13 is a pure consequence of the individual error models used for the RR and the RC models: no new assumption has been made to derive eq. 13.

2.3.2. Input/output measurement errors

The RS model needs to be calibrated using observations of its input/output variables. The input variables typically comprise precipitation and potential evapotranspiration, while the output variable is stage.

In this paper, we make the strong assumption that measurement errors in all input/output variables are negligible. We acknowledge that this assumption is unrealistic in most studies. For instance, errors in estimating areal precipitation may be large when the raingauge density is small (see e.g. [53, 12]). Similarly, continuously-measured stage values may be affected by non-negligible errors, of both random and systematic nature. Typically, the inherent uncertainty of the stage sensor corresponds to a random error, while the periodic recalibration of the stage sensor with respect to the staff gauge produces an unknown systematic error between two successive recalibrations (for more details, see e.g. [32]).

Making this restrictive assumption allows focusing entirely on the uncertainty induced by the rating curve while minimizing possible interactions between input and output errors. In practice, unaccounted input/output errors will be implicitly absorbed by the structural error terms (B_t and E_t). One should therefore keep in mind that while these terms are intended to represent structural errors, they may also encompass the effect of ignored input/output errors.

3. Calibration

In this paper, we apply Bayesian estimation to estimate all unknown parameters. The posterior distributions are explored by means of an adaptive Markov Chain Monte Carlo sampler described in Haario et al. [54]. The convergence of the chains is assessed visually by plotting the simulated chains and verifying their stationarity.

The general calibration strategy is made of two successive steps. We first estimate the RC using available gauging pairs (these gaugings are not used afterwards). In a second stage, we estimate the RS model combining the RC and the RR submodels (thus the RC model is re-calibrated). Since the RS model comprises parameters related to the RC (namely, $\boldsymbol{\theta}_{\mathrm{RC}}$ and $\boldsymbol{\gamma}$, see section 2.1), the posterior distribution of these parameters obtained after stage 1 becomes their prior distribution in stage 2. Note that this informative prior for the RC model, based on an analysis of rating curve data, strongly constrains the inference. This allows avoiding non-identifiability and equifinality problems in the estimation of all parameters during stage 2.

3.1. Stage 1: rating curve calibration

From the assumptions described in section 2.1, the gauged discharge at time t can be written as follows (combining equations 2 and 4):

$$\tilde{q}_t = f_{\mathrm{RC}}\left(\tilde{h}_t, \boldsymbol{\theta}_{\mathrm{RC}}\right) - E_t(\boldsymbol{\gamma}) + W_t \tag{14}$$

Conditional on unknown parameters, the gauged discharge $\tilde{q}_t$ is therefore a realization from a Gaussian distribution with mean $\breve{q}_t = f_{\mathrm{RC}}\left(\tilde{h}_t, \boldsymbol{\theta}_{\mathrm{RC}}\right)$ and variance $(\gamma_1 + \gamma_2 \cdot \breve{q}_t)^2 + \delta_t^2$. The likelihood function can therefore be written:

$$p\left(\tilde{\boldsymbol{q}}|\boldsymbol{\theta}_{\mathrm{RC}}, \boldsymbol{\gamma}, \tilde{\boldsymbol{h}}\right) = \prod_{k=1}^{N_{gauging}} f_G\left(\tilde{q}_{t_k}; \breve{q}_{t_k}, (\gamma_1 + \gamma_2 \cdot \breve{q}_{t_k})^2 + \delta_{t_k}^2\right) \tag{15}$$

where $f_G(u; m, v)$ is the Gaussian pdf with mean m and variance v, evaluated at u.

The posterior distribution is then computed up to a constant of proportionality using Bayes' theorem:

$$p\left(\boldsymbol{\theta}_{\mathrm{RC}}, \boldsymbol{\gamma}|\tilde{\boldsymbol{q}}, \tilde{\boldsymbol{h}}\right) \propto p\left(\tilde{\boldsymbol{q}}|\boldsymbol{\theta}_{\mathrm{RC}}, \boldsymbol{\gamma}, \tilde{\boldsymbol{h}}\right) \cdot p\left(\boldsymbol{\theta}_{\mathrm{RC}}, \boldsymbol{\gamma}\right) \tag{16}$$

The prior distribution for RC parameters $\boldsymbol{\theta}_{\mathrm{RC}}$ is derived from an analysis of the hydraulic configuration of the gauging station, as will be described in the case study (for more general considerations, see Le Coz et al. [23]). For the parameters $\boldsymbol{\gamma}$ governing the standard deviation of structural errors, wide non-informative priors are used.

3.2. Stage 2: rainfall-stage model calibration

Let $\boldsymbol{h} = (h_{t_k})_{k=1:N}$ denote the observed time series of stage values used to calibrate the RS model. Computing the likelihood requires deriving the distribution of $\boldsymbol{h}$ conditional on all inferred quantities. Unfortunately, this cannot be done directly on the basis of eq. 13. Indeed, this conditional distribution is not Gaussian, because the Gaussian error terms E_t and B_t transit through nonlinear models (the backward transformation ψ^{-1} and the inverse rating curve f_{RC}^{-1}). Moreover, this non-Gaussian pdf cannot be derived analytically. Indeed, eq. 13 involves the sum of two independent random variables. The pdf of this sum can be obtained by convolution, but this convolution has no analytical solution because one of the random variables is not Gaussian.

In order to circumvent this issue, we partly linearize eq. 13 as described next. We introduce the following shorthand notation for this section:

$$\begin{aligned} \hat{q}_t(\boldsymbol{\theta}_{\mathrm{RR}}) &= f_{\mathrm{RR}}\left(\boldsymbol{x}_{1:t}, \boldsymbol{\theta}_{\mathrm{RR}}\right) \\ d_t^{(\psi)}(\boldsymbol{\theta}_{\mathrm{RR}}) &= \psi'\left(\hat{q}_t\right) \end{aligned} \tag{17}$$

Using this notation and linearizing the backward transformation ψ^{-1}, eq. 13 can be approximated as follows (see Appendix B for details):

$$h_t \approx f_{\mathrm{RC}}^{-1}\left(\underbrace{\hat{q}_t(\boldsymbol{\theta}_{\mathrm{RR}}) - \frac{B_t(\boldsymbol{\phi})}{d_t^{(\psi)}(\boldsymbol{\theta}_{\mathrm{RR}})} + E_t(\boldsymbol{\gamma})}_{Z_t}, \boldsymbol{\theta}_{\mathrm{RC}}\right) \tag{18}$$

The term Z_t is now the sum of a constant plus two Gaussian terms, and is therefore itself Gaussian. More precisely, the vector $\boldsymbol{Z} = (Z_{t_1}, ..., Z_{t_N})$ follows a multivariate Gaussian distribution, with mean vector $\boldsymbol{\mu}$ (size N) and covariance matrix $\boldsymbol{\Sigma}$ (size $N \times N$) defined as follows:

$$\boldsymbol{\mu}(\boldsymbol{\theta}_{\mathrm{RR}}) = (\hat{q}_{t_1}(\boldsymbol{\theta}_{\mathrm{RR}}), ..., \hat{q}_{t_N}(\boldsymbol{\theta}_{\mathrm{RR}})) \tag{19}$$

$$\mathbf{\Sigma}(\boldsymbol{\theta}_{\mathrm{RR}}, \boldsymbol{\phi}, \boldsymbol{\gamma}) = \boldsymbol{D}^{(\psi)}\mathbf{\Sigma}^{(\mathrm{RR})}\boldsymbol{D}^{(\psi)} + \mathbf{\Sigma}^{(\mathrm{RC})} \tag{20}$$

In the latter equation, $\boldsymbol{D}^{(\psi)}$ denotes the square $N \times N$ diagonal matrix whose diagonal terms are equal to $1/d_t^{(\psi)}$, while $\mathbf{\Sigma}^{(\mathrm{RR})}$ and $\mathbf{\Sigma}^{(\mathrm{RC})}$ are the $N \times N$ covariance matrices of RR and RC structural errors:

$$\boldsymbol{D}^{(\psi)}(i,i) = \frac{1}{d_{t_i}^{(\psi)}(\boldsymbol{\theta}_{\mathrm{RR}})}; \quad \boldsymbol{D}^{(\psi)}(i,j) = 0 \ if \ i \neq j \tag{21}$$

$$\mathbf{\Sigma}^{(\mathrm{RR})}(i,j) = \phi_1^2 \cdot exp\left(-\frac{|t_i - t_j|}{\phi_2}\right) \tag{22}$$

$$\mathbf{\Sigma}^{(\mathrm{RC})}(i,i) = \left(\gamma_1 + \gamma_2 . \hat{q}_{t_i}\right)^2; \quad \mathbf{\Sigma}^{(\mathrm{RC})}(i,j) = 0 \ if \ i \neq j \tag{23}$$

Having derived the pdf of $\boldsymbol{Z}$, the pdf of $\boldsymbol{h} \approx f_{\mathrm{RC}}^{-1}(\boldsymbol{Z})$ (eq. 18) can be obtained by applying the change-of-variables formula. After some computation (see Appendix B for details), this yields the following likelihood:

$$\begin{aligned} p\left(\boldsymbol{h}|\boldsymbol{\theta}_{\mathrm{RR}}, \boldsymbol{\theta}_{\mathrm{RC}}, \boldsymbol{\phi}, \boldsymbol{\gamma}, \boldsymbol{x}\right) = \\ f_{MG}\left(f_{\mathrm{RC}}(\boldsymbol{h}, \boldsymbol{\theta}_{\mathrm{RC}}); \boldsymbol{\mu}(\boldsymbol{\theta}_{\mathrm{RR}}), \mathbf{\Sigma}(\boldsymbol{\theta}_{\mathrm{RR}}, \boldsymbol{\phi}, \boldsymbol{\gamma})\right) \times \prod_{k=1}^{N} |f'_{RC}(h_{t_k}, \boldsymbol{\theta}_{\mathrm{RC}})| \end{aligned} \tag{24}$$

where $f_{MG}\left(\boldsymbol{u}; \boldsymbol{m}, \boldsymbol{v}\right)$ is the multivariate Gaussian pdf with mean vector $\boldsymbol{m}$ (size N) and covariance matrix $\boldsymbol{v}$ (size $N \times N$), evaluated at vector $\boldsymbol{u}$ (size N).

The posterior distribution is then computed up to a constant of proportionality using Bayes' theorem:

$$p\left(\boldsymbol{\theta}_{\mathrm{RR}}, \boldsymbol{\theta}_{\mathrm{RC}}, \boldsymbol{\phi}, \boldsymbol{\gamma}|\boldsymbol{h}, \boldsymbol{x}\right) \propto p\left(\boldsymbol{h}|\boldsymbol{\theta}_{\mathrm{RR}}, \boldsymbol{\theta}_{\mathrm{RC}}, \boldsymbol{\phi}, \boldsymbol{\gamma}, \boldsymbol{x}\right) \cdot p\left(\boldsymbol{\theta}_{\mathrm{RR}}, \boldsymbol{\theta}_{\mathrm{RC}}, \boldsymbol{\phi}, \boldsymbol{\gamma}\right) \tag{25}$$

The prior distribution for RC-related parameters $\boldsymbol{\theta}_{\mathrm{RC}}$ and $\boldsymbol{\gamma}$ is set to the posterior distribution obtained after calibration of the RC using gaugings at stage 1 (eq. 16). For the parameters of the RR model ($\boldsymbol{\theta}_{\mathrm{RR}}$), priors are case-specific and related to the RR model and available information. For the parameters $\boldsymbol{\phi}$ governing the properties of RR structural errors, wide non-informative priors are used.

Note that the RS model is calibrated against time series with observed stages. However, during the evaluation both the output of the RS model, stage, and the output of the RR model, discharge, will be examined. This is possible thanks to the explicit treatment of RC and RR errors.

3.3. Calibration strategies

The posterior distribution in eq. 25 corresponds to a full calibration strategy, schematized in Figure 1: the parameters related to both the RR model and the RC are estimated together, thus enabling interactions between them and hence assessing how RC uncertainties impact the estimation of RR parameters. In particular, both parametric ($\boldsymbol{\theta}_{\text{RC}}$) and structural ($\boldsymbol{E}$) uncertainties of the RC are accounted for. In order to understand in more depth the impact of these two types of uncertainty, we also implement incomplete calibration strategies, where some uncertainty sources are ignored. As shown in Table 1, these strategies are the following:

1. Strategy NoS ignores RC structural uncertainty. This corresponds to assuming that $\boldsymbol{E} = 0$, which is achieved by using $\boldsymbol{\Sigma}^{(\text{RC})} = 0$ in eq. 20. A similar representation of RC uncertainty has been used by Steinbakk [55] in the context of flood frequency analysis, and by Sikorska et al. [26] in the context of model calibration.
2. Strategy NoP ignores RC parametric uncertainty. This is achieved by removing $\boldsymbol{\theta}_{\text{RC}}$ from the list of inferred parameters. The RC is therefore used with a fixed parameter vector $\hat{\boldsymbol{\theta}}_{\text{RC}}$, taken as the maxpost estimate (i.e. the vector maximizing the stage-1 posterior of eq. 16). This strategy is similar to the representation of RC uncertainty used by e.g. Thyer at al. [37] or Renard et al. [6].
3. Strategy NoPNoS ignores both RC parametric and structural uncertainty, hence using both a fixed parameter vector $\hat{\boldsymbol{\theta}}_{\text{RC}}$ and setting $\boldsymbol{\Sigma}^{(\text{RC})} = 0$. In this strategy, there is no explicit representation of RC uncertainty, which corresponds to the most widely-used approach in hydrological modeling (*standard uncertainty estimation approach*).
4. Strategy FULL* is similar to the full strategy, except that the prior for RC parameters $\boldsymbol{\theta}_{\text{RC}}$ is truncated. More precisely, we set the prior pdf to zero outside of 95% probability intervals for each component of $\boldsymbol{\theta}_{\text{RC}}$. This strategy stongly limits the possible interactions between $\boldsymbol{\theta}_{\text{RC}}$ and other inferred parameters. It guarantees that after calibration of the RS model, the RC parameters will still be within the 95% credibility intervals derived by calibrating the RC to gaugings. Note that bluntly truncating the prior as done here makes the resulting distribution unnormalized; however this is not problematic in the Bayesian-MCMC context of this paper since the posterior only needs to be known up to a normalizing constant.

5. Similarly, strategy NoS* is a variation of the NoS strategy, with a truncated prior for $\boldsymbol{\theta}_{\mathrm{RC}}$.

4. Case study: the Ardèche river at Meyras

4.1. Ardèche catchment

The river Ardèche is a right tributary of the River Rhône and has it sources in the Massif Central in France (Figure 2). The gauging station Meyras, located at 318 m a.s.l., controls an area of 98.43 km^2. The mean elevation of this catchment is 899 m a.s.l., with the highest point located at 1467 m a.s.l. The catchment is quite steep with an average slope of 23.4 % and it is in 68% covered by forests [56]. The average annual precipitation, estimated based on fifty years of observations at the station Péreyres (840 m a.s.l.), is 1774 mm/yr in this region, whereof approximately 40% is lost to evaporation. With the yearly mean daily temperature equal to 9.25°C and the snowfall ratio of less than 3% of the annual precipitation, the snow processes can be neglected to model this catchment.

4.2. Rating curve

As a RC model (eq. 1), we use a piecewise combination of power functions of the form $q = a(h-b)^c$. This combination is defined by the succession of hydraulic controls governing the stage-discharge relationship, as explained in more details by Le Coz [23]. At the Meyras gauging station, three controls can be identified (Figure 3). Low flows are first governed by a natural gravel riffle (control 1). When the stage gets above a certain level, this riffle is drowned and a channel control takes over (control 2). Finally, for very high stage values, the main channel may be full and some flow may also occur in the floodplain (control 3). This configuration leads to the following rating curve equation:

$$f_{\mathrm{RC}}\left(h_t, \boldsymbol{\theta}_{\mathrm{RC}}\right) = \begin{cases} a_1\left(h_t - b_1\right)^{c_1} & \text{if } \kappa_1 < h_t \leq \kappa_2 \text{ (control 1)} \\ a_2\left(h_t - b_2\right)^{c_2} & \text{if } \kappa_2 < h_t \leq \kappa_3 \text{ (control 2)} \\ a_2\left(h_t - b_2\right)^{c_2} + a_3\left(h_t - b_3\right)^{c_3} & \text{if } \kappa_3 < h_t \text{ (control 2 + 3)} \end{cases} \tag{26}$$

where $(\kappa_1, \kappa_2, \kappa_3)$ are the (unknown) activation stages for each control. Note that parameters b_1, b_2 and b_3 do not need to be inferred because they can be deduced by continuity of the RC as shown in eq. 27 below. Consequently, the parameters of the rating curve are $\boldsymbol{\theta}_{\mathrm{RC}} = (\kappa_1, a_1, c_1, \kappa_2, a_2, c_2, \kappa_3, a_3, c_3)$ where the relationships between κ and b are as follows:

$$b_1 = \kappa_1; b_2 = \kappa_2 - \left(\frac{a_1}{a_2} \cdot (\kappa_2 - b_1)^{c_1}\right)^{\frac{1}{c_2}}; b_3 = \kappa_3 \tag{27}$$

The parameters $\boldsymbol{\theta}_{\mathrm{RC}}$ are related to physical characteristics of the gauging section, which opens the possibility to specify informative priors. For instance, the first control by a natural riffle can be approximated using a rectangular weir formula, as shown in Table 2. This formula indicates that the exponent c_1 should be close to 1.5. Moreover, the parameter a_1 is linked to the weir width B_w and to a discharge coefficient C_r. The width can be approximated at [1] $B_w = (8 \pm 2)$ m, while literature suggests values of the coefficient $C_r = 0.4 \pm 0.1$ (see [48, 23]). These two uncertainties can be combined by using the uncertainty propagation formula recommended by the Guide to the Expression of Uncertainty in Measurement [57]. This yields the Gaussian prior distribution for a_1 shown in Table 2. Lastly, the elevation of the weir crest, which defines the activation stage κ_1, is estimated at $\kappa_1 = (-0.05 \pm 0.05)$ m.
A similar approach can be used to specify priors for parameters of controls 2 and 3, using the Manning-Strickler formula for wide rectangular channels (see Table 2). For the main channel, the Strickler coefficient is set to $K_S = (25 \pm 2.5)\ \mathrm{m}^{1/3} \cdot \mathrm{s}^{-1}$, the channel width to $B_w = (15 \pm 2.5)$ m and the slope to $S = (3 \pm 1)\ \mathrm{m} \cdot \mathrm{km}^{-1}$. For the floodplain, we use $K_S = (15 \pm 2.5)\ \mathrm{m}^{1/3} \cdot \mathrm{s}^{-1}$, $B_w = (30 \pm 5)$ m and $S = (3 \pm 1.25)\ \mathrm{m} \cdot \mathrm{km}^{-1}$. This completes the prior specification shown in Table 2.

4.3. Rainfall-runoff model (HBV)

The rainfall-runoff process within the Ardèche catchment is modelled with a HBV model [58, 59, 60]. The HBV consists of four main routines responsible for modelling snow dynamics, soil moisture, runoff response, and flow routing in the channel. Because snow processes can be neglected in this catchment, we use a simplified version of the HBV model, i.e., with an inactive snow component. To further simplify the model, we model the catchment as a single subcatchment without any elevation-dependent correction factors for inputs. This further reduces the number of inferred parameters to 6 (Table 3).

[1] in the notation $x \pm s$, s is the standard deviation.

Such a simplified HBV model requires mean areal precipitation and long term evaporation estimates as input, while temperature data responsible for modeling the snow component are not strictly required. In this study, the HBV model is run at hourly time steps. Since the HBV model was not applied before in this catchment, no previous knowledge was available for its parameters. Thus, we formulate prior for each HBV parameter as a uniform distribution restricted to possible ranges that were defined for each parameter independently (Table 3).

4.4. Calibration data

4.4.1. stage 1: rating curve calibration

To infer the RC parameters $\boldsymbol{\theta}_{\mathrm{RC}}$ and $\boldsymbol{\gamma}$, we use 41 gaugings made between 2001 and 2008, for a period with no noticeable shift of the RC. For each gauged discharge, we assume a constant relative uncertainty of $\pm 3.5\%$, i.e. for a gauged discharge equal to q_t, the standard deviation δ_t in eq. 5 is set to $\delta_t = 0.035 \cdot q_t$. The gaugings and their uncertainty can be seen in Figure 4b.

4.4.2. stage 2: rainfall-stage model calibration

The RS model described with the Eq. 13 requires mean areal precipitation at the hourly time step as input. Yet, the stage observations at the gauging station are recorded by the limnigraph with unequal time steps adequate to the current dynamics of flow processes (i.e., between 1 hour and 10/15 days). Hence, we chose to use directly these data instead of converting them into the hourly estimates, which would yield additional errors due to the stage approximation. Note that this involves interpolating the HBV-discharge simulations on the temporal (irregular) grid used for stage values. Using irregularly spaced data is possible with the correlated error term on the hydrological model introduced (Eqs. 8 and 9).

4.5. Results: rating curve calibration (stage 1)

Figure 4a shows the prior RC resulting from the hydraulic analysis of the gauging station (Table 2). Figure 4b shows the posterior RC and illustrates the uncertainty reduction resulting from the information brought by the gaugings. The posterior RC is overall quite precise, especially for stages smaller than 1 m. For such relatively small stages, parametric uncertainty is only a small part of the total uncertainty, which is hence dominated by structural uncertainty. For stage values beyond 1 m, total uncertainty increases, mostly due to an increase of parametric uncertainty which becomes

dominant for such high stages. In particular the parameter κ_3 representing the activation stage of the third control is not precisely estimated (between 1 m and 1.5 m, see green band in Figure 4b).
The posterior distribution of RC parameters $\boldsymbol{\theta}_{\mathrm{RC}}$ and $\boldsymbol{\gamma}$ obtained after this first stage is now being used as a prior distribution for the second stage. Note that the posterior on RC is in fact represented with Monte Carlo samples. Hence, to specify the prior distribution for the second stage of calibration, we fit a multivariate Gaussian distribution to the Monte Carlo samples from the first stage. The resulting corresponding marginal distributions can be seen as gray boxplots in Figure 5.

4.6. Results: parameter estimates (stage 2)

Posteriors for the RS model for all six calibration strategies are plotted as boxplots against prior information (obtained from stage 1) in Figure 5. For parameters of the RR and RC sub-models and of the structural error of the RR model, we observe that parameters tend to form three groups in terms of their posterior behaviours. These groups are shaped as follows: (1) calibration strategies FULL and NoS, (2) NoP and NoPNoS, and (3) FULL* and NoS*, as seen in the figure. It appears that this grouping is driven by the way of accounting for RC parametric uncertainty, i.e.: fully accounting (group 1), non-accounting (group 2), and accounting but within the constrained truncated prior (group 3). The grouping effect is obviously not visible for parameters responsible for the RC structural uncertainty ($\boldsymbol{\gamma}$) as these parameters are excluded from the inference in the strategies NoS, NoPNoS and NoS*.

4.6.1. Hydrological model

With respect to the HBV parameters ($\boldsymbol{\theta}_{\mathrm{RR}}$, i.e., $PERC$:$MAXBAS$, top two rows in Figure 5), which mainly control the response and routing function, we specifically observe that posterior parameters vary among three pattern groups and particularly between FULL and NoPNoS strategy (*standard uncertainty estimation approach*). We observe that using a simplified description of errors as in NoPNoS leads to different values of inferred model parameters than when explicitly representing all major contributing sources. Such modified parameters of the RR model should mostly transfer to altered discharge simulations (being an intermediate step within the RS model) and might lead to biased estimates. Confronting posterior ranges of different calibration strategies indicates that generally parametric uncertainty of the RR model is very similar for most parameters in all strategies. The interpretation of individual parameter uncertainty is however difficult due to their interactions. Therefore, not the uncertainty of individual parameters but rather the resulting parametric uncertainty in predicted discharge is our major interest (see Sect. 4.8.2).

Diagnosis of the structural error model of the RR model ($\boldsymbol{\phi}$) show that the error standard deviation (ϕ_1) is the smallest for the FULL and the largest for NoPNoS strategy (Fig. 5). This seems logical as in the FULL strategy the total residual variance is decomposed into two contributing sources originating from the RR model ($\boldsymbol{\phi}$) and from the RC model ($\boldsymbol{\gamma}$, see further below), while in the strategy NoPNoS all variance is explained with $\boldsymbol{\phi}$ only. Hence, only this error can be increased to capture the mismatch between the observed and the simulated stage. Posterior error standard deviations of all other calibration strategies lie between these two strategies. This seems reasonable as they represent transitional steps between FULL and NoPNoS strategies in terms of the level of the variance decomposition from the simplest strategy (NoPNoS) towards the most complex strategy (FULL). As it also seems logical, excluding RC parameters ($\boldsymbol{\theta}_{\mathrm{RC}}$) from the inference results in an increased error ϕ_1 (NoP and NoPNoS) in comparison to strategies which include $\boldsymbol{\theta}_{\mathrm{RC}}$ into the inference (NoS) even if restricted prior is used (FULL* or NoS*). The error autocorrelation length (ϕ_2) generally follows the behavior of the error standard deviation and is the longest for strategy NoP and NoPNoS, and the shortest for NoS* and FULL*.

4.6.2. Rating curve model

Posterior RC parameters ($\boldsymbol{\theta}_{\mathrm{RC}}$) are presented in three bottom rows in Figure 5 (parameters k_1 till c_3). Note that RC parameters in strategies NoP and NoPNoS are not altered during the inference and are kept at the values of maxpost from the calibration stage 1 from section 4.5. For other four strategies, again two groups of parameter behaviors can be observed.

Specifically, using informative but unbounded prior for RC parameters $\boldsymbol{\theta}_{\mathrm{RC}}$ during the inference (FULL and NoS) results in a significant shift of posteriors often outside of the 95% prior bounds. This effect appears to be a result of a possible compensation for other uncertainty sources, specifically for the one originating from the RR model parameters. As it appears, nine parameters of the RC in addition to six parameters of the RR model gives a higher level of freedom for modifying model simulations to match stage observations. It is worth recalling that although RC parameters are related to physical characteristics of the gauging station and informative prior is used, prior information on the third control is very imprecise as it is constrained with only very few gaugings (see section 4.5). Indeed, posteriors on parameters of the third control are strongly modified during the inference. As all RS parameters are inferred at the same time, the possible compensation between parameters of the RR and the RC sub-model cannot be avoided given unbounded priors on all parameters being inferred. This shifting of the RC parameters outside of hydraulically reasonable boundaries is expected to have a consequence on the shape of an updated RC (see section 4.7).

Using a truncated prior on RC parameters indeed prevents from a strong modification of RC parameters (FULL* and NoS*). As it is visible in the Figure 5, posteriors attempt to move towards the values from unbounded strategies but remain within the 95% limits set. This also results in a smaller RC parametric uncertainty in strategies FULL* and NoS* than in unbounded strategies FULL and NoS.

Finally, the structural error of the rating curve ($\boldsymbol{\gamma}$) varies in different calibration strategies. As it is represented with two parameters, the combined structural error of the RC cannot be easily quantified from estimated posteriors. We observe, however, an inverse relationship between its behavior and the behavior of the RR structural error. This relationships seems also logical as both structural and parametric uncertainties of the RC are decomposed from the total uncertainty in FULL and FULL* strategies.

4.7. Results: updated rating curve (stage 2)

Updated RCs for four strategies accounting for RC parametric uncertainty (i.e., FULL, NoS, FULL* and NoS*) are plotted in Figure 6 with uncertainty bands (blue polygons) against the prior (red polygons). RCs for strategies NoP and NoPNoS are not plotted as their parameters are not altered during the calibration. As expected, a strong shift in the RC posterior parameters observed for strategies which use non-bounded prior (FULL and NoS) leads to a strong modification of the RC shape. This effect is especially visible in the range of the third control for which the updated RC distinctly transcends the prior ranges (red polygons) by pushing the RC towards assigning smaller discharge values for the same stages. As previously mentioned, the prior for parameter inference on the third RC control is established with very few measures and thus is very uncertain (see Figure 4), which allows for freely modifying these parameters. It is clear that setting bounded priors on RC parameters in strategies FULL* and NoS* prevents from destroying the RC shape which remains within the 95% prior limits.

This issue of using the RC parameters to compensate for limitations of the RR model has clearly serious implications for using such updated RCs and will be further discussed in section 5.1.

4.8. Results: predictive uncertainty (stage 2)

4.8.1. Total uncertainty bands

Total uncertainty bands (TUB) for the FULL strategy are plotted in Figure 7 for stages and in Figure 8 for discharges (top panels). For both variables TUB appear to be reasonable as they cover most of the data points and are smaller for low flow and higher for high flow conditions (assessed visually). The smaller uncertainty during low flows is more apparent for discharges than for stages.

Widths of TUB for all other strategies are plotted in Figure 9 for both stages (top) and discharges (bottom). The TUB width in the FULL strategy is used as a reference. Widths of TUB for all other strategies are represented with respect to the FULL TUB width and thus are plotted as curves (a value larger than one representing a TUB width larger than that of the FULL strategy). The top panel of Figure 9 shows that for stage, the TUB widths of all strategies are larger than that of the FULL strategy for almost the entire calibration and validation periods. Specifically, during low flow periods (e.g. around the vertical red line), TUB widths are larger than that of the FULL strategy by a factor of up to 2, while during high flows this factor decreases to about 1.25. Similar patterns are observed with respect to discharges apart for the strategy NoS*, for which TUB width is similar to the FULL strategy on average.

Although the effect of obtaining the smallest TUB width for the FULL strategy is visible for both stage and discharge, it has greater implications for modeling discharge. Much smaller TUB for FULL strategy clearly demonstrates a benefit compared to the strategy NoPNoS (*standard uncertainty estimation approach*). This finding indicates that accounting for both (structural and parametric) RC uncertainties allows for removing these uncertainty parts from the total discharge uncertainty and this results in narrower TUB in comparison to strategies which do not present such ability (NoPNoS, NoP, NoS). Using bounded prior on RC parameters results in wider TUB in comparison to their respective unbounded strategies, which confirms that the structural error of the RR model is used for compensation of other unrepresented uncertainty components.

4.8.2. Uncertainty contributors

An explicit representation of different uncertainty components within the TUB (i.e. of rating curve and of hydrological model) allows for their relative assessment. Depending on the strategy, these are parametric and structural uncertainty of the RR model and/or the RC model. Clearly, the most interesting is the FULL strategy which makes it possible to assess all four uncertainty components in predictions of stages and two components in prediction of discharges. Note that by an explicit representation of the parametric and structural uncertainties of RC in the FULL strategy, these uncertainty components can be decomposed from the TUB and thus are not propagated on the discharge (since the aim is to predict the "true" discharge and not the RC-estimated one). On the contrary, not accounting for structural or parametric uncertainty of the RC does not allow for removing these uncertainty parts from the total uncertainty and thus they will be implicitly propagated on the discharge simulations. Uncertainty contributions for the FULL strategy are presented visually in Figure 7 for stages and in Figure 8 for discharges (bottom panels).

With respect to stages, it can be seen that the structural error of the RR model ($\boldsymbol{\phi}$) represents the majority of the total uncertainty while the next major contributor is the structural error of the RC model. Parametric uncertainty of the RR model and of the RC model are both less relevant. These contributions vary slightly over time and the contribution of the RC structural error is slightly higher during recession periods, whereas the contribution of the RC parametric error increases during high flows. The contribution of the parametric uncertainty of the RR model is higher during high flows and successive recession periods, while it is smaller during low flows. In a similar fashion, the structural error of the RR model accounts for the majority of the total uncertainty of discharge prediction.

The visual assessment of uncertainty contributions is accompanied by the time-averaged relative contributions of each uncertainty source for all six strategies and these are presented in Table 4. Generally, we observe quite stable uncertainty contributions of different error sources in all six calibration schemes. For all calibration strategies, the structural error of the RR model explains the majority of the total uncertainty which is ranging from 81% in the FULL strategy to 94% in NoPNoS for stages, and from 92% in FULL to

94% for NoP, NoPNoS, FULL* and NoS* for discharges. Both the parametric uncertainties of the RR and of the RC model vary only insignificantly and are much less relevant than other two uncertainty components. Hence, the change in contribution shares is thus mainly due to structural uncertainties of both the RC but mostly the RR model. The latter component is thus used to compensate for all unrepresented uncertainty source(s).

5. Discussion

5.1. Feasibility of accounting for rating curve uncertainty through a rainfall-stage model

The approach proposed in this paper to explicitly account for RC uncertainty in the calibration of a RR model is to include both RC and RR parameters within a rainfall-stage (RS) model. However, the results of the case study show that the initial RC (established using gaugings) is strongly modified after calibration of the RS model (strategies FULL and NoS), unless a restrictive truncated prior is used for RC parameters (strategies FULL* and NoS*). We consider that the extent to which the RC is modified is hardly defensible; we therefore do not consider this modification as a meaningful improvement of the RC, but rather as a sign that the results produced by strategies FULL and NoS should be taken with caution.

It is of interest to further discuss this issue in terms of the information content used in the two successive calibration stages. The initial RC is established using the information brought by 41 independent gaugings, along with the prior information derived from the hydraulic analysis (the latter being informative but still quite imprecise). The posterior distribution of calibration stage 1 reflects this quantity of information. During calibration stage 2, this posterior is used as a prior, but the RC can be further modified by the information brought by more than 1000 stage values used for calibration of the RS model. At first sight, the information imbalance between 41 gaugings vs. 1000 stage values may explain why the RC is strongly modified by the calibration of the RS model (well beyond the prior constraint induced by the gaugings). However, one should keep in mind the following points:

1. Since the RR structural error uses an autocorrelation component, the information content of these 1000 stage values does not correspond to that of 1000 independent data;

2. The information contained in the 41 gaugings is only used to estimate the RC, while the information contained in the 1000 stage values is used to infer both the RC and the RR model.

These clarifications notwithstanding, the strong modification of the RC is a sign that the error models we used do not convincingly weight the information brought by the gaugings and the stage time series. We see at least two avenues to improve this:

1. Improve the error models, as discussed in the next section 5.2;
2. Do not re-estimate the RC during calibration stage 2. This can be achieved by means of a propagation approach, as discussed in section 5.3

5.2. Limitation of the error models

The RC structural error model in Eq. 3 assumes independent errors, which is questionable at least for time steps close to each other. In principle, it is feasible to avoid this independence assumption e.g. by using an autocorrelation component. However, identifying an autocorrelation structure based on gaugings is difficult in practice, if not impossible, because gaugings are made too sporadically. Typically two successive gaugings are separated by weeks or months, which makes shorter autocorrelation structures non-identifiable. While implementing dedicated high-frequency gauging strategies might be feasible, we do not see any obvious solution with existing operational gauging datasets.

Unlike RC structural errors, RR structural errors are not assumed independent and instead an explicit autocorrelation component is used (Eq. 8). This autocorrelation structure is identifiable because the stage time series is sampled at a high frequency. However, due to the particular dynamics of the RR model, even this autocorrelation structure is too simplistic. In particular, autocorrelation properties are likely very different during dry periods and rainy periods, when quick-flow components are activated. As input error is implicitly encompassed into the model bias, these different properties cannot be distinguished with the RR error model used here and the inferred bias is "averaged" over dry and wet conditions. More flexibility should hence be added to the autocorrelation component to allow distinguishing these distinct properties, for instance by making a bias dry/wet period dependent or input-related (for further discussion on input error see Sect. 5.5).

Moreover, a common limitation of both RC and RR structural error models is the lack of a systematic component. A structural error is indeed defined as the difference between the model prediction (forced with perfect inputs) and the unknown truth. For a given set of inputs, this error is likely to have a non-zero mean, because it is (at least partly) due to model structural deficits that will systematically manifest themselves when the model is forced with similar inputs. Such a non-zero mean can also be interpreted as a "conditional bias" (conditional to the inputs and initial conditions). The fact that the structural error models we used ignore this conditional bias (as do the error models we are aware of in the literature) probably explains the undesired modification of the RC discussed in section 5.1: the calibration can only use parameters $\boldsymbol{\theta}_{\mathrm{RR}}$ and $\boldsymbol{\theta}_{\mathrm{RC}}$ (whose modification induces a systematic difference in model prediction) to minimize this conditional bias. Deriving an error model that explicitly describes the conditional bias is an important perspective in our opinion, but also a challenging one: its formulation and identifiability from the data in the absence of prior knowledge are open questions. Finally, we note that this discussion has some links with the problem of describing epistemic errors with statistical models, which motivated the development of "informal" likelihoods for hydrological [61] and rating curve [39] models.

5.3. An alternative: propagating RC uncertainty

An alternative to the approach used in this paper is to propagate RC uncertainty by performing many calibrations of the RS model, with each calibration using a distinct RC. As an illustration, consider the box-plots shown in Figure 10. They have been obtained by performing 5 calibrations using the strategy NoP, where parameters $\boldsymbol{\theta}_{\mathrm{RC}}$ are fixed to 5 distinct values randomly chosen in the MCMC simulations of calibration stage 1. For the sake of simplicity, we demonstrate this approach on the basis of only one RR and RC parameter (θ_{RR} and θ_{RC} respectively). One given box-plot represents the uncertainty in RR parameter θ_{RR}, conditional on one particular RC. Merging all 5 box-plots together allows "'unconditioning", i.e. representing the total uncertainty in RR parameter θ_{RR}, given all plausible RCs. In a similar fashion, "unconditional" estimates are derived for all RR parameters $\boldsymbol{\theta}_{\mathrm{RR}}$. This propagation approach has been used for instance by Steinbakk [55] or Petersen-Øverleir [34] in a flood frequency analysis context.

Formally, the propagation approach leads to the following pdf representing uncertainty in RR parameters $\boldsymbol{\theta}_{\mathrm{RR}}$:

$$p_{propa}\left(\boldsymbol{\theta}_{\mathrm{RR}}|\boldsymbol{h}\right)=\int p\left(\boldsymbol{\theta}_{\mathrm{RR}}|\boldsymbol{h},\boldsymbol{\theta}_{\mathrm{RC}}\right)p\left(\boldsymbol{\theta}_{\mathrm{RC}}\right)\,\mathrm{d}\boldsymbol{\theta}_{\mathrm{RC}} \tag{28}$$

By contrast, the approach used in this paper represents uncertainty in RR parameters $\boldsymbol{\theta}_{\mathrm{RR}}$ using its marginal posterior distribution, defined as follows:

$$\begin{aligned}p\left(\boldsymbol{\theta}_{\mathrm{RR}}|\boldsymbol{h}\right)&=\int p\left(\boldsymbol{\theta}_{\mathrm{RR}},\boldsymbol{\theta}_{\mathrm{RC}}|\boldsymbol{h}\right)\,\mathrm{d}\boldsymbol{\theta}_{\mathrm{RC}}\\&=\int p\left(\boldsymbol{\theta}_{\mathrm{RR}}|\boldsymbol{h},\boldsymbol{\theta}_{\mathrm{RC}}\right)p\left(\boldsymbol{\theta}_{\mathrm{RC}}|\boldsymbol{h}\right)\,\mathrm{d}\boldsymbol{\theta}_{\mathrm{RC}}\end{aligned} \tag{29}$$

The difference between the two approaches appears clearly in these equations: the latter uses the information contained in the stage calibration data to update the inference of RC parameters (term $p\left(\boldsymbol{\theta}_{\mathrm{RC}}|\boldsymbol{h}\right)$ in Eq. 29), while the former ignores this information and only uses the prior RC estimates (term $p\left(\boldsymbol{\theta}_{\mathrm{RC}}\right)$ in Eq. 28), i.e. the RC inferred with gaugings only.

Future work should investigate the pros and cons of each approach. The propagation approach is akin to repeating the NoP approach many times, except that the RC parameters are not fixed at their maxpost estimate, but are rather sampled from the prior distribution derived from the analysis of rating curve data. The advantage compared to NoP is that RC parametric uncertainty is not ignored. However, an obvious drawback of the propagation approach is its computational cost, since a potentially costly calibration has to be repeated many times.

5.4. Limitation of the proposed approach in terms of time steps

The approach proposed in this paper uses the inverse of the RC to derive a RS model. This only makes sense if the RC is invertible, or in other words, if the stage-discharge relationship can be represented by a bijective function. This may not be the case under some particular circumstances (e.g. hydraulic hysteresis or variable backwater effects). But even more generally, the stage-discharge relationship can only be represented by a bijective function at a nearly-instantaneous time step. Consider for instance a given daily-averaged discharge value: for this particular day, an infinity of stage time series could lead to the same daily discharge. Consequently, it is not possible to relate this daily discharge to a single daily stage indicator (e.g. daily mean/median/etc.).

Consequently, the approach proposed here is restricted to time steps for which the within-step variability of stage can be neglected. Whether and how the approach can be extended for larger time steps remains unclear yet. A possible strategy would be to define a RC between e.g. daily-averaged stage and discharge, equipped with a stochastic component in order to account for the non-uniqueness of the daily discharge associated with a given daily stage. The variability of this stochastic component would directly depend of the within-step variability of stage.

5.5. Towards a complete decomposition of input/output/structural errors

Deriving a complete uncertainty framework that allows explicitly representing all uncertainty sources remains a major challenge of hydrologic modeling. Several methodological frameworks have been proposed for this purpose, e.g., SODA [62], BATEA [63], Kalman and particle filters (e.g. [64, 65]) and many more. These methodological frameworks need to be equipped with specific error models to describe the various sources of uncertainty (input, output and structural errors), and realistic error models are a prerequisite for a meaningful uncertainty analysis. Moreover, specifying precise and accurate prior distributions to characterize input and output errors is another prerequisite to limit the interactions between the various error sources (e.g., [6, 15]). Consequently, studies focusing on a specific uncertainty component are valuable to derive realistic error models and investigate their properties. For instance, previous research was devoted to investigate properties of input errors (e.g., [13, 15]) and their impact on model calibration. Following the same line of thought, we focus in this paper on rating curve output errors, and their impact on model calibration. The specific error models we propose could later be included into a more general framework such as SODA, BATEA or a Kalman/particle filter. Finally, we stress that there is no unique solution to uncertainty estimation in hydrologic modelling. Instead, varied and flexible error models are necessary to adapt to the objective of the study, the available information, etc. As an illustration, we note that the output error model we propose in this paper requires a significant amount of information (hydraulic analysis of the gauging station, gaugings and their uncertainty). While this allows making valuable use of local information, it is primarily adapted to the detailed analysis of a small number of catchments. This information may not be available for larger-scale analyses that may involve hundreds or thousands of catchments. In this case, an alternative error model would need to be considered (see, e.g., the nonparametric discharge uncertainty estimate of Vrugt et al. [62]).

6. Conclusions

In this work, we develop a Bayesian approach to probabilistically represent parametric and structural uncertainties of the rating curve in the estimation of the hydrological model. To achieve this, we couple the hydrological model with the inverse rating curve yielding the *rainfall-stage model* that is calibrated in the stage and not in the discharge space. Such a model description enables us for explicitly representing and quantifying uncertainties associated with both the hydrological and the rating curve model in the total uncertainty of stage and discharge predictions. For a case study in France, we consider six different calibration strategies with a different representation level of rating curve uncertainties (parametric and/or structural). Our results show that a) ignoring rating curve uncertainty leads to visible changes in hydrological model parameters, and b) structural uncertainty of the hydrological model dominates other uncertainty sources. The major limitation of the current method arises from a strong modification of the rating curve shape if rating curve parameters are re-estimated during the calibration of the rainfall-stage model and unbounded prior is used. We see this problem to be related to the shortcomings of the error models used to describe correlated errors of the hydrological model and structural errors of the rating curve. Thus, the next step should be to test the method with a more advanced description of errors and/or to explore the proposed alternative of propagating rating curve parametric uncertainty in more detail.

Acknowledgments

The support of the Ambassade de France en Suisse in the enforcement of this research is thankfully acknowledged. The authors thank the editor Harrie-Jan Hendricks-Franssen and three anonymous reviewers for their useful comments, which helped improving the manuscript.

References

[1] A. Montanari, What do we mean by uncertainty? the need for a consistent wording about uncertainty assessment in hydrology, Hydrol. Process. 21 (6) (2007) 841–845. doi:10.1002/hyp.6623.

[2] M.-H. Ramos, T. Mathevet, J. Thielen, F. Pappenberger, Communicating uncertainty in hydro-meteorological forecasts: mission impossible?, Meteor. Appl. 17 (2) (2010) 223–235. doi:10.1002/met.202.

[3] A. E. Sikorska, A. Scheidegger, K. Banasik, J. Rieckermann, Bayesian uncertainty assessment of flood predictions in ungauged urban basins for conceptual rainfall-runoff models, Hydrol. Earth Syst. Sci. 16 (4) (2012) 1221–1236. doi:10.5194/hess-16-1221-2012.

[4] N. K. Ajami, Q. Duan, S. Sorooshian, An integrated hydrologic bayesian multimodel combination framework: Confronting input, parameter, and model structural uncertainty in hydrologic prediction, Water Resour. Res. 43 (1) (2007) W01403. doi:10.1029/2005WR004745.

[5] G. Kuczera, B. Renard, M. Thyer, D. Kavetski, There are no hydrological monsters, just models and observations with large uncertainties!, Hydrol. Sci. J. 55 (6) (2010) 980–991. doi:10.1080/02626667.2010.504677.

[6] B. Renard, D. Kavetski, G. Kuczera, M. Thyer, S. W. Franks, Understanding predictive uncertainty in hydrologic modeling: The challenge of identifying input and structural errors, Water Resour. Res. 46 (5) (2010) W05521. doi:10.1029/2009WR008328.

[7] H. McMillan, T. Krueger, J. Freer, Benchmarking observational uncertainties for hydrology: rainfall, river discharge and water quality, Hydrol. Process. 26 (2012) 4078–4111. doi:10.1002/hyp.9384.

[8] A. E. Sikorska, J. Seibert, Value of different precipitation data for flood prediction in an alpine catchment: A bayesian approach, J. Hydrol. - (-) (2016) in press. doi:10.1016/j.jhydrol.2016.06.031.

[9] A. Montanari, D. Koutsoyiannis, A blueprint for process-based modeling of uncertain hydrological systems, Water Resour. Res. 48 (9) (2012) W09555. doi:10.1029/2011WR011412.

[10] D. Kavetski, G. Kuczera, S. W. Franks, Bayesian analysis of input uncertainty in hydrological modeling: 2. application, Water Resour. Res. 42 (3) (2006) W03408. doi:10.1029/2005WR004376.

[11] J. Vrugt, C. ter Braak, M. Clark, J. Hyman, B. Robinson, Treatment of input uncertainty in hydrologic modeling: Doing hydrology backward with Markov chain Monte Carlo simulation, Water Resour. Res. 44 (2008) W00B09. doi:http://dx.doi.org/10.1029/2007WR006720.

[12] B. Renard, D. Kavetski, E. Leblois, M. Thyer, G. Kuczera, S. W. Franks, Toward a reliable decomposition of predictive uncertainty in hydrological modeling: Characterizing rainfall errors using conditional simulation, Water Resour. Res. 47 (11) (2011) n/a–n/a, w11516. doi:10.1029/2011WR010643.

[13] H. McMillan, B. Jackson, M. Clark, D. Kavetski, R. Woods, Rainfall uncertainty in hydrological modelling: An evaluation of multiplicative error models, J. Hydrol. 400 (1-2) (2011) 83–94. doi:10.1016/j.jhydrol.2011.01.026.

[14] P. Reichert, N. Schuwirth, Linking statistical bias description to multi-objective model calibration, Water Resour. Res. 48 (9) (2012) W09543. doi:10.1029/2011WR011391.

[15] D. Del Giudice, C. Albert, J. Rieckermann, P. Reichert, Describing the catchment-averaged precipitation as a stochastic process improves parameter and input estimation, Water Resour. Res. 52 (4) (2016) 3162 – 3186. doi:10.1002/2015WR017871.

[16] G. Di Baldassarre, A. Montanari, Uncertainty in river discharge observations: A quantitative analysis, Hydrol. Earth Syst. Sci. 13 (6) (2009) 913–921.

[17] G. Di Baldassarre, P. Claps, A hydraulic study on the applicability of flood rating curves, Hydrology Research 42 (1) (2011) 10–19. doi:10.2166/nh.2010.098.

[18] WMO, Guide to hydrological practice, volume i, hydrology – from measurement to hydrological information, 6th edn, World Meteorological Organisation, Geneva, Switzerland, 2008, p. pp. 296.

[19] J. Le Coz, A literature review of methods for estimating the uncertainty associated with stage-discharge relations (2012).

[20] R. Clarke, Uncertainty in the estimation of mean annual flood due to rating-curve indefinition, J. Hydrol. 222 (1-4) (1999) 185–190. doi:10.1016/S0022-1694(99)00097-9.

[21] T. Morlot, C. Perret, A.-C. Favre, J. Jalbert, Dynamic rating curve assessment for hydrometric stations and computation of the associated uncertainties: Quality and station management indicators, J. Hydrol. 517 (2014) 173–186. doi:http://dx.doi.org/10.1016/j.jhydrol.2014.05.007.

[22] A. Petersen-Øverleir, A. Soot, T. Reitan, Bayesian rating curve inference as a streamflow data quality assessment tool, Water Resour. Manag. 23 (9) (2009) 1835–1842. doi:10.1007/s11269-008-9354-5.

[23] J. Le Coz, B. Renard, L. Bonnifait, B. F., R. Le Boursicaud, Combining hydraulic knowledge and uncertain gaugings in the estimation of hydrometric rating curves: A bayesian approach, J. Hydrol. 509 (2014) 573 – 587. doi:http://dx.doi.org/10.1016/j.jhydrol.2013.11.016.

[24] H. McMillan, J. Freer, F. Pappenberger, T. Krueger, M. Clark, Impacts of uncertain river flow data on rainfall-runoff model calibration and discharge predictions, Hydrol. Process. 24 (10) (2010) 1270–1284. doi:10.1002/hyp.7587.

[25] Q. Shao, J. Lerat, G. Podger, D. Dutta, Uncertainty estimation with bias-correction for flow series based on rating curve, J. Hydrol. 510 (2014) 137 – 152. doi:http://dx.doi.org/10.1016/j.jhydrol.2013.12.025.

[26] A. E. Sikorska, A. Scheidegger, K. Banasik, J. Rieckermann, Considering rating curve uncertainty in water level predictions, Hydrol. Earth Syst. Sci. 17 (11) (2013) 4415–4427. doi:10.5194/hess-17-4415-2013.

[27] V. B. Sauer, R. W. Meyer, Determination of error in individual discharge measurements (1992).

[28] A. Domeneghetti, A. Castellarin, A. Brath, Assessing rating-curve uncertainty and its effects on hydraulic model calibration, Hydrol. Earth Syst. Sci. 16 (4) (2012) 1191–1202. doi:10.5194/hess-16-1191-2012.

[29] G. Coxon, J. Freer, I. K. Westerberg, T. Wagener, R. Woods, P. J. Smith, A novel framework for discharge uncertainty quantification applied to 500 UK gauging stations, Water Resour. Res. (2015) 5531 – 5546doi:10.1002/2014WR016532.

[30] G. Di Baldassarre, F. Laio, A. Montanari, Effect of observation errors on the uncertainty of design floods, Phys. Chem. Earth, Parts A/B/C 42–44 (2012) 85 – 90.

[31] J. Guerrero, I. K. Westerberg, S. Halldin, C.-Y. Xu, L.-C. Lundin, Temporal variability in stage–discharge relationships, J. Hydrol. 446–447 (2012) 90 – 102. doi:http://dx.doi.org/10.1016/j.jhydrol.2012.04.031.

[32] I. Horner, J. Le Coz, B. Renard, F. Branger, G. Pierrefeu, et al., Accounting for stage measurement errors in the uncertainty analysis of streamow records, J. Hydrol (2016) in prep.

[33] G. Kuczera, Correlated rating curve error in flood frequency inference, Water Resour. Res. 32 (7) (1996) 2119–2127. doi:10.1029/96WR00804.

[34] A. Petersen-Øverleir, T. Reitan, Accounting for rating curve imprecision in flood frequency analysis using likelihood-based methods, J. Hydrol. 366 (1-4) (2009) 89–100, cited By 20. doi:10.1016/j.jhydrol.2008.12.014.

[35] M. Lang, K. Pobanz, B. Renard, E. Renouf, E. Sauquet, Extrapolation of rating curves by hydraulic modelling, with application to flood frequency analysis, Hydrol. Sci. J. 55 (6) (2010) 883–898. doi:10.1080/02626667.2010.504186.

[36] I. K. Westerberg, T. Wagener, G. Coxon, H. K. McMillan, A. Castellarin, A. Montanari, J. Freer, Uncertainty in hydrological signatures for gauged and ungauged catchments, Water Resour. Res. 52 (3).

[37] M. Thyer, B. Renard, D. Kavetski, G. Kuczera, S. W. Franks, S. Srikanthan, Critical evaluation of parameter consistency and predictive uncertainty in hydrological modeling: A case study using bayesian total error analysis, Water Resour. Res. 45 (12) (2009) W00B14. doi:10.1029/2008WR006825.

[38] I. Westerberg, J.-L. Guerrero, J. Seibert, K. Beven, S. Halldin, Stage-discharge uncertainty derived with a non-stationary rating curve in the choluteca river, honduras, Hydrol. Process. 25 (4) (2011) 603–613. doi:10.1002/hyp.7848.

[39] H. K. McMillan, I. K. Westerberg, Rating curve estimation under epistemic uncertainty, Hydrol. Process. 29 (7) (2015) 1873–1882. doi:10.1002/hyp.10419.

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

[41] K. Beven, A. Binley, The future of distributed models: Model calibration and uncertainty prediction, Hydrol. Process. 6 (1992) 279–298. doi:10.1002/hyp.3360060305.

[42] Y. Liu, J. Freer, K. Beven, P. Matgen, Towards a limits of acceptability approach to the calibration of hydrological models: Extending observation error, J. Hydrol. 367 (1–2) (2009) 93 – 103. doi:http://dx.doi.org/10.1016/j.jhydrol.2009.01.016.

[43] D. Kavetski, G. Kuczera, S. W. Franks, Bayesian analysis of input uncertainty in hydrological modeling: 1. Theory, Water Resour. Res. 42 (3) (2006) W03407. doi:10.1029/2005WR004368.

[44] H. Sellami, I. La Jeunesse, S. Benabdallah, M. Vanclooster, Parameter and rating curve uncertainty propagation analysis of the swat model for two small mediterranean catchments, Hydrol. Sci. J. 58 (8) (2013) 1635–1657. doi:10.1080/02626667.2013.837222.

[45] M. Thyer, B. Renard, D. Kavetski, G. Kuczera, Can hydrological model predictions be improved by developing streamflow measurement error models using rating curve data?, STAHY workshop - Advances in Statistical Hydrology, 2010.

[46] G. E. Uhlenbeck, L. S. Ornstein, On the theory of the Brownian Motion, Phys. Rev. 36 (1930) 823.

[47] A. E. Sikorska, D. Del Giudice, K. Banasik, J. Rieckermann, The value of streamflow data in improving TSS predictions – Bayesian multi-objective calibration, J. Hydrol. 530 (-) (2015) 241–254. doi:10.1016/j.jhydrol.2015.09.051.

[48] R. W. Herschy, Hydrometry: principles and practice, John Wiley & Sons Ltd, 1998.

[49] ISO, Hydrometry – measurement of liquid flow in open channels using current-meters or floats (2007).

[50] J. Le Coz, B. Camenen, X. Peyrard, G. Dramais, Uncertainty in open-channel discharges measured with the velocity–area method, Flow Meas. Instrum. 26 (0) (2012) 18–29.

[51] P. M. Pelletier, Uncertainties in the single determination of river discharge: a literature review, Can. J. Civil Eng. 15 (5) (1988) 834–850.

[52] T. A. Cohn, J. E. Kiang, R. R. M. Jr., Estimating discharge measurement uncertainty using the interpolated variance estimator, J. Hydrol. Eng. 139 (5) (2013) 502–510.

[53] R. Linsley, M. Kohler, Hydrology for Engineers, McGraw Hill, London, 1988.

[54] H. Haario, E. Saksman, J. Tamminen, An adaptive metropolis algorithm, Bernoulli 7 (2001) 223–242.

[55] G. H. Steinbakk, T. L. Thorarinsdottir, T. Reitan, L. Schlichting, S. Hølleland, K. Engeland, Propagation of rating curve uncertainty in design flood estimation., Water Resour. Res. (2016) in press-doi:10.1002/2015WR018516.

[56] M. Adamovic, I. Braud, F. Branger, J. W. Kirchner, Assessing the simple dynamical systems approach in a mediterranean context: application to the ardéche catchment (france), Hydrol. Earth Syst. Sci. 19 (5) (2015) 2427–2449. doi:10.5194/hess-19-2427-2015.

[57] J. C. for Guides in Metrology, Evaluation of measurement data – guide to the expression of uncertainty in measurement (2010).

[58] S. Bergström, The HBV model - its structure and applications, no. RHO 4, SMHI Hydrology, Norröping, 1992, p. 35 pp.

[59] J. Seibert, Estimation of parameter uncertainty in the HBV model, Nord. Hydrol. 28 (4/5) (1997) 247–262.

[60] J. Seibert, M. J. P. Vis, Teaching hydrological modeling with a user-friendly catchment-runoff-model software package, Hydrol. Earth Syst. Sci. 16 (9) (2012) 3315–3325. doi:10.5194/hess-16-3315-2012.

[61] K. Beven, J. Freer, Equifinality, data assimilation, and uncertainty estimation in mechanistic modelling of complex environmental systems using the GLUE methodology, J. Hydrol. 249 (2001) 11–29. doi:10.1016/S0022-1694(01)00421-8.

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

[63] D. Kavetski, S. W. Franks, G. Kuczera, Confronting input uncertainty in environmental modelling, Calibration of watershed models (2003) 49 – 68doi:10.1029/WS006p0049.

[64] A. Weerts, G. El Serafy, Particle filtering and ensemble Kalman filtering for state updating with hydrological conceptual rainfall-runoff models, Water Resour. Res. 42 (2006) W09403. doi:10.1029/2005WR004093.

[65] P. Salamon, L. Feyen, Disentangling uncertainties in distributed hydrological modeling using multiplicative error models and sequential data assimilation, Water Resour. Res. 46 (2010) W12501. doi:10.1029/2009WR009022.

[66] G. E. P. Box, D. R. Cox, An analysis of transformations revisited, rebutted, R. J. Am. Stat. Ass. 77 (1982) 209–210.

Table 1: Calibration strategies.

Strategy	$\boldsymbol{\theta}_{\mathrm{RC}}$		E_t		$\pi(\boldsymbol{\theta}_{\mathrm{RC}})$	
	infer	fix	active	zero	full	truncated
FULL	✓		✓		✓	
NoS	✓			✓	✓	
NoP		✓	✓			N/A
NoPNoS		✓		✓		N/A
FULL*	✓		✓			✓
NoS*	✓			✓		✓

Table 2: Prior distributions for rating curve parameters. The Manning–Strickler equation is a simplified version valid for wide rectangular channels.

Control	Idealized formula	$\pi(\kappa)$	$\pi(a)$	$\pi(c)$
Control 1 natural riffle	Rectangular weir $q = \underbrace{C_r B_w \sqrt{2g}}_{a}(h - \underbrace{h_0}_{b})^{\overbrace{1.5}^{c}}$	$N(-0.05, 0.05^2)$	$N(14, 5^2)$	$N(1.5, 0.025^2)$
Control 2 main channel	Manning–Strickler equation $q = \underbrace{K_S B_w \sqrt{S}}_{a}(h - \underbrace{h_0}_{b})^{\overbrace{5/3}^{c}}$	$N(0.1, 0.05^2)$	$N(20, 5^2)$	$N(1.67, 0.025^2)$
Control 3 floodplain	Manning–Strickler equation	$N(1.2, 2^2)$	$N(25, 7.5^2)$	$N(1.67, 0.025^2)$

Table 3: HBV parameters being inferred during calibration and their prior.

Parameter	Significance [unit]	Prior min	Prior max
PERC	Percolation threshold parameter [$mm\ h^{-1}$]	0	2
UZL	Groundwater runoff threshold parameter [mm]	0	100
K0	Recession coefficient of the 1st storage [h^{-1}]	0	0.4
K1	Recession coefficient of the 2nd storage [h^{-1}]	0	0.2
K2	Recession coefficient of the 3rd storage [h^{-1}]	0	0.1
MAXBAS	Length of the triangular weighing function [h]	1	10

Table 4: Time-averaged relative contribution (in %) of each source of uncertainty.

Prediction of	stage				discharge	
Calibration strategy	$\boldsymbol{\theta}_{\mathrm{RR}}$	B_t	$\boldsymbol{\theta}_{\mathrm{RC}}$	E_t	$\boldsymbol{\theta}_{\mathrm{RR}}$	B_t
FULL	6	81	5	8	8	92
NoS	7	89	4	0	7	93
NoP	6	88	0	6	6	94
NoPNoS	6	94	0	0	6	94
FULL*	5	87	2	6	6	94
NoS*	6	92	2	0	6	94

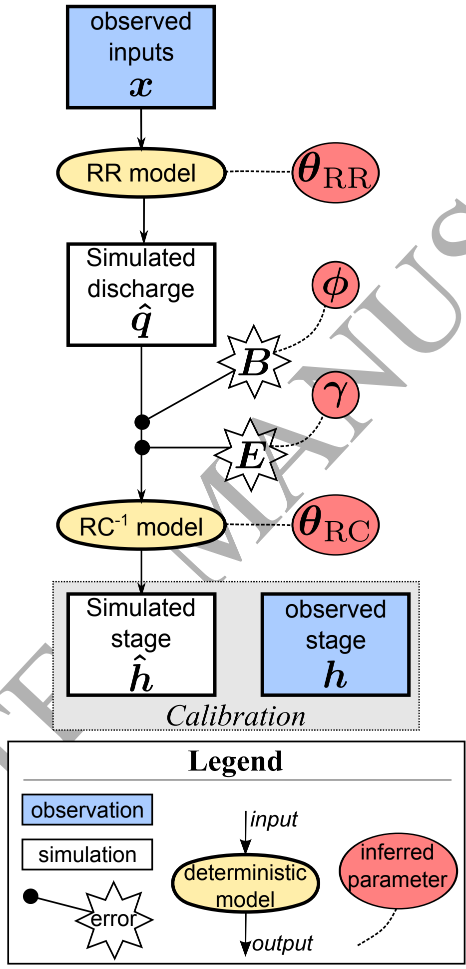


Figure 1: Schematic of the full calibration strategy.

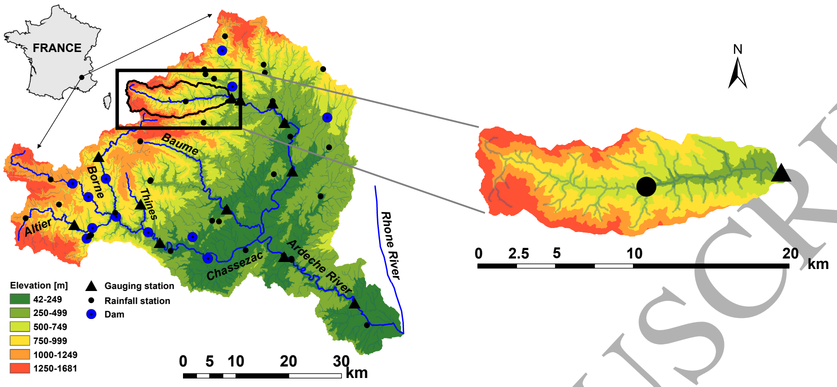


Figure 2: Overview of the catchment Ardèche. Left panel: entire catchment until its confluence with the Rhone River; right panel: Ardèche catchment at Meyras gauging station. Modified from Adamivic et al. [56].

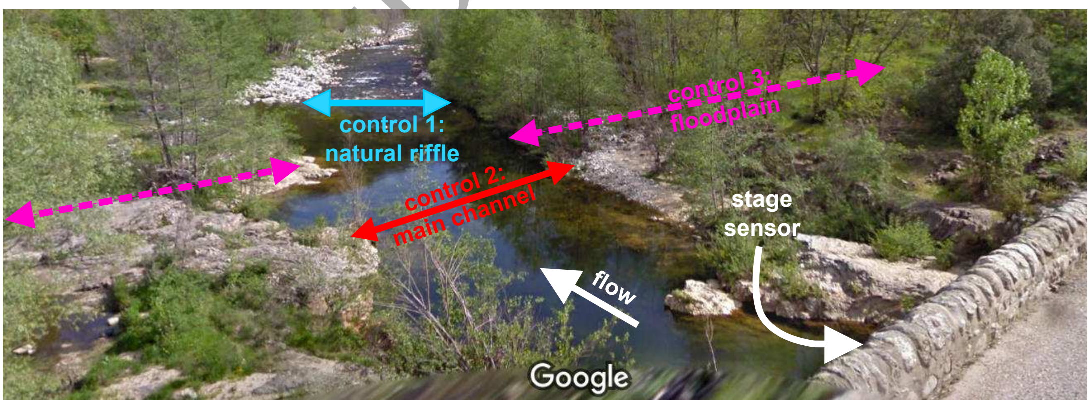


Figure 3: View of the gauging station for the Ardèche at Meyras. Picture from Google Maps, taken in April 2010.

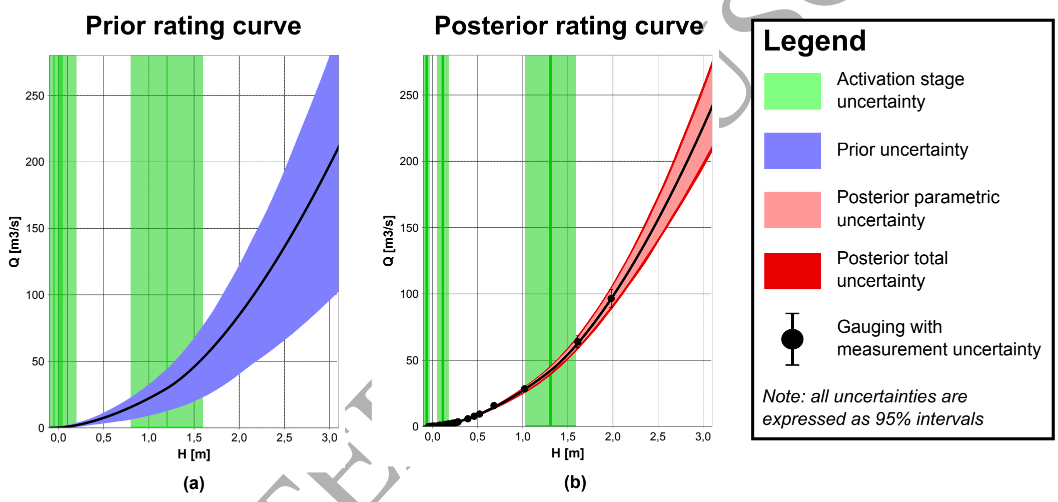


Figure 4: Prior (a) and posterior (b) rating curves in the 1st stage of calibration.

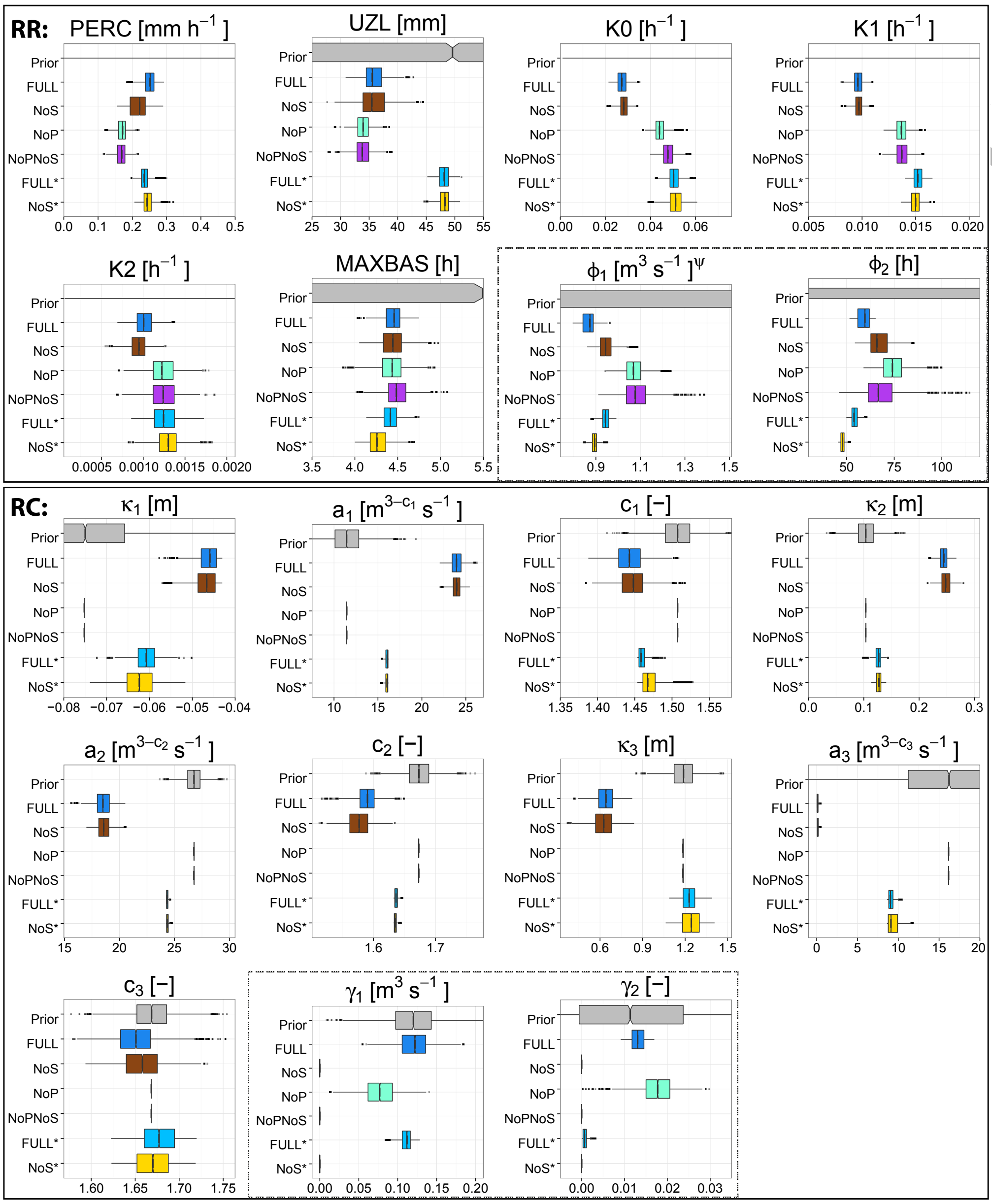


Figure 5: Calibration of the rainfall-stage model (stage 2): boxplots of prior (gray) and posterior (colored) distributions obtained with the six calibration strategies for the RR model (top panel) and RC model (bottom panel). Both error models (of RR and RC) are marked in the dashed boxes.

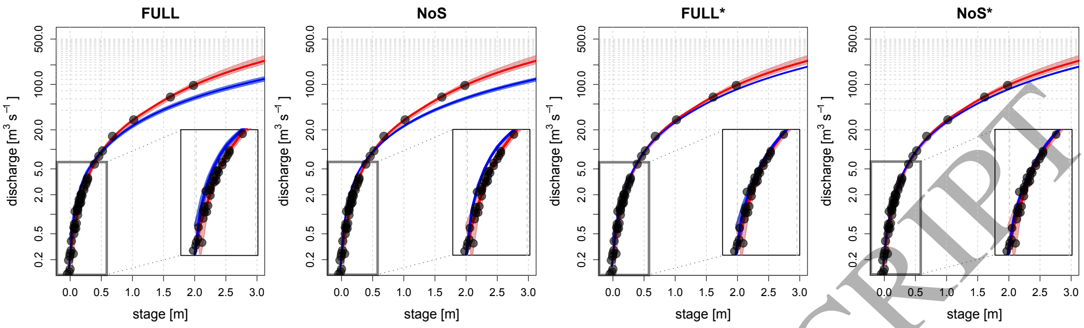


Figure 6: Comparison of rating curves before (red) and after (blue) calibration of the rainfall-stage model (stage 2) for the four calibration strategies accounting for RC parametric uncertainty.

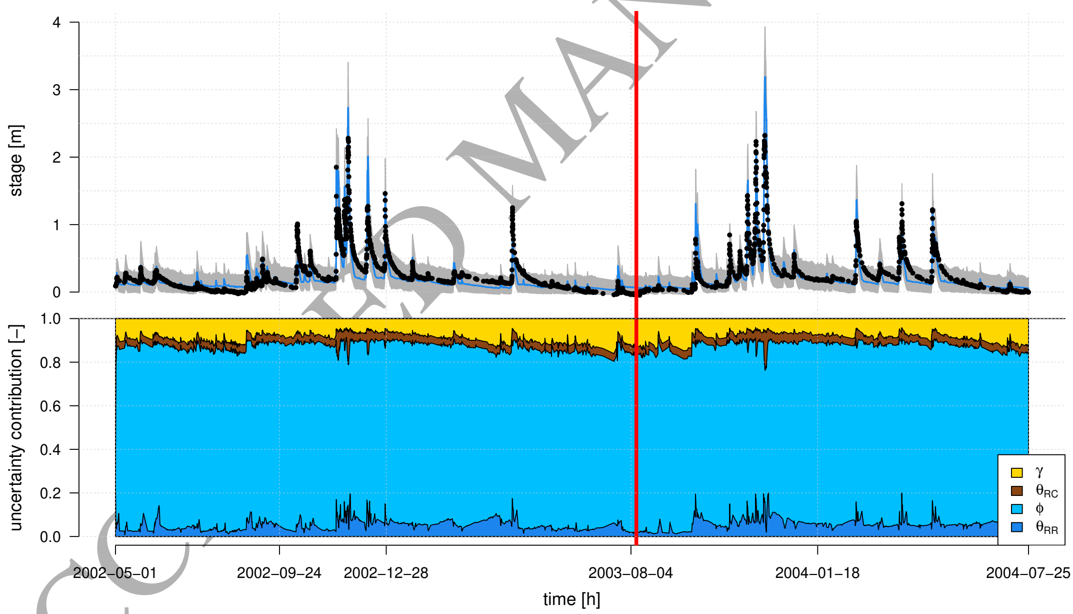


Figure 7: Stage prediction using the FULL calibration strategy. Top panel shows predicted vs. observed stage along with 95% intervals representing the total uncertainty. Bottom panel shows the relative contribution of each source of uncertainty. The calibration period is before the vertical line. Note the irregular time step of observed stages as demonstrated on the x-axis.

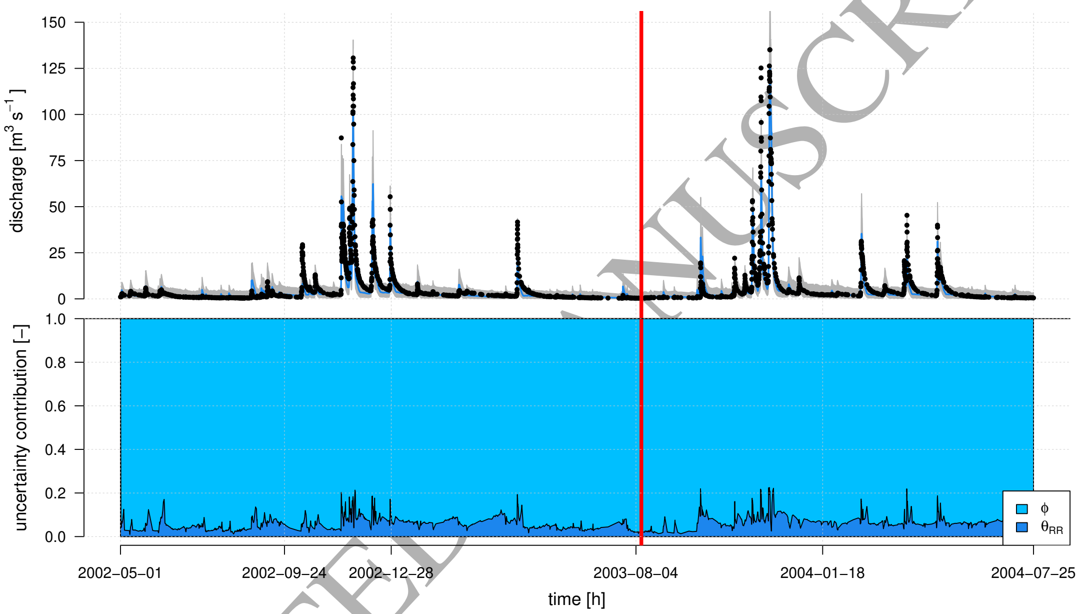


Figure 8: Discharge prediction using the FULL calibration strategy. Top panel shows predicted vs. observed discharge along with 95% intervals representing the total uncertainty. Bottom panel shows the relative contribution of each source of uncertainty. The calibration period is before the vertical line. Note that the irregular time step of discharges results from the irregular time step of observed stages.

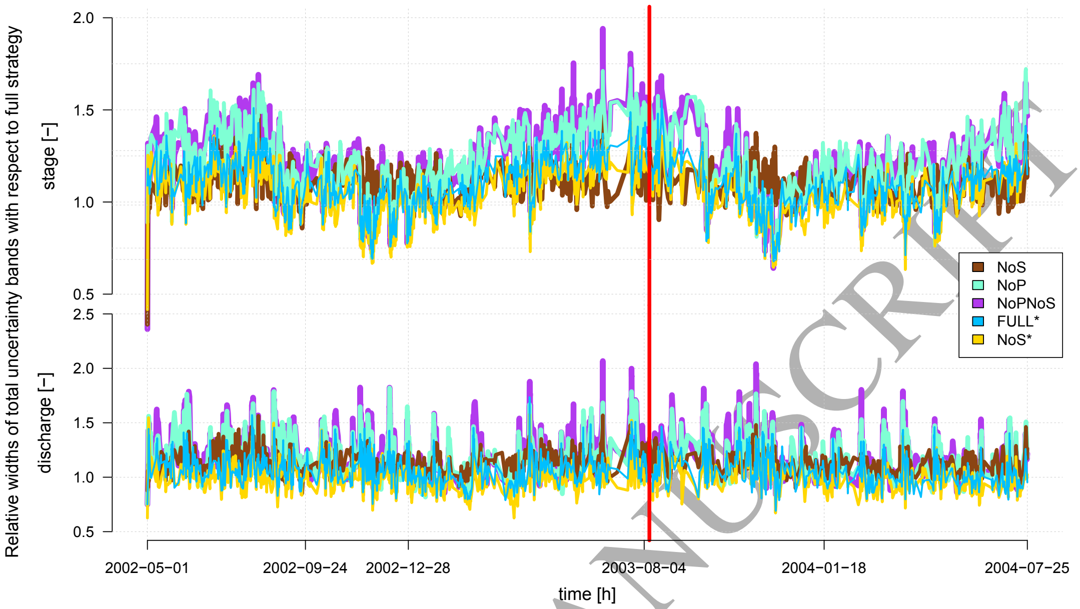


Figure 9: Comparison of total uncertainty for six calibration strategies. Each curve shows the ratio of 95% interval widths between the considered strategy and the reference strategy (FULL) in stage (top) and discharge (bottom) space. The calibration period is before the vertical line. Note the irregular time steps of both stages and discharges.

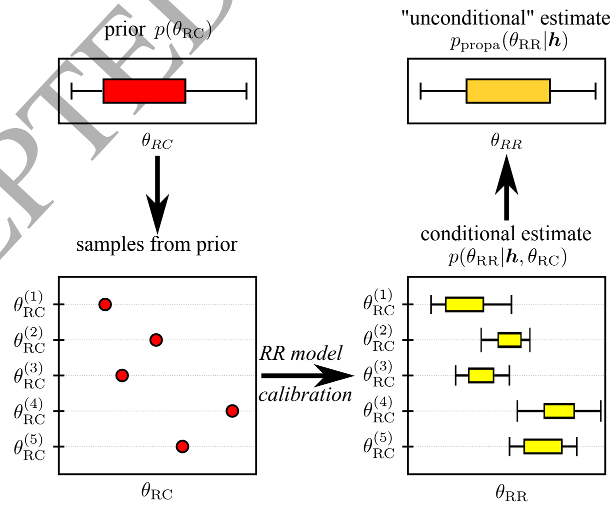


Figure 10: Schematic overview of the RC parametric uncertainty propagation approach.

Appendix A. Box-Cox transformation $\psi(\cdot)$

The Box-Cox transformation [66] with parameters λ_1 and λ_2 can be written as follows:

$$\psi\left(y\right)=\begin{cases}\frac{(y+\lambda_2)^{\lambda_1}-1}{\lambda_1} & \text{if } \lambda_1\neq 0\\ \ln\left(y+\lambda_2\right) & \text{if } \lambda_1=0\end{cases} \tag{A.1}$$

Parameter $\lambda_2 \geq 0$ aims at ensuring that the term $y+\lambda_2$ remains positive. Note that for $(\lambda_1=1,\lambda_2=1)$, the Box-Cox transformation is the identity, while for $(\lambda_1=0,\lambda_2=0)$ it simplifies to a logarithmic transformation. Typically parameter λ_1 is taken between 0 and 1.
The inverse of the Box-Cox transform and its derivative can be written as follows:

$$\psi^{-1}\left(\dot{y}\right)=\begin{cases}\left(\lambda_1\cdot\dot{y}+1\right)^{1/\lambda_1}-\lambda_2 & \text{if } \lambda_1\neq 0\\ \exp\left(\dot{y}\right)-\lambda_2 & \text{if } \lambda_1=0\end{cases} \tag{A.2}$$

$$\psi'\left(y\right)=\left(y+\lambda_2\right)^{\lambda_1-1} \tag{A.3}$$

Appendix B. Likelihood computation for the RS model

The task is to derive the joint pdf of $(h_{t_1},...,h_{t_N})$, where h_t is given by eq. 13 (recalled below in a simplified form):

$$h_t=f_{\mathrm{RC}}^{-1}\left(\psi^{-1}\left[\psi\left(\hat{q}_t\right)-B_t\right]+E_t\right) \tag{B.1}$$

The first step is to use a first-order approximation of the backward transform ψ^{-1} based on a first-order Taylor expansion, whose general form can be written as:

$$f(x+e)\approx f(x)+f'(x)\times e \tag{B.2}$$

Applied to the function ψ^{-1} in eq. B.1, this yields:

$$\psi^{-1}\left[\psi\left(\hat{q}_t\right)-B_t\right]\approx\psi^{-1}\left[\psi\left(\hat{q}_t\right)\right]-(\psi^{-1})'\left[\psi\left(\hat{q}_t\right)\right]\times B_t \tag{B.3}$$

We then use here the inverse-derivative rule:

$$(\psi^{-1})'(z)=\frac{1}{\psi'(\psi^{-1}(z))} \tag{B.4}$$

Plugging this back into eq. B.3 yields:

$$\psi^{-1}\left[\psi\left(\hat{q}_t\right) - B_t\right] \approx \hat{q}_t - \frac{B_t}{\psi'(\psi^{-1}\left[\psi(\hat{q}_t)\right])} = \hat{q}_t - \frac{B_t}{\psi'(\hat{q}_t)} \tag{B.5}$$

Finally, eq. B.1 becomes:

$$h_t \approx f_{\mathrm{RC}}^{-1}\left(\underbrace{\hat{q}_t - \frac{B_t}{\psi'(\hat{q}_t)} + E_t}_{Z_t}\right) \tag{B.6}$$

The second step is to deduce the joint pdf of $(h_{t_1}, ..., h_{t_N})$ from that of $(Z_{t_1}, ..., Z_{t_N})$. We use the change-of-variables formula for this purpose, which can be written in general terms as follows. Let $\boldsymbol{y} = (y_1, ..., y_N) = \boldsymbol{r}(x_1, ..., x_N)$, where $\boldsymbol{r}$ is a one-to-one transformation. The pdf of $\boldsymbol{y}$ can be deduced from the pdf of $\boldsymbol{x}$ using the following formula:

$$p_{\boldsymbol{y}}(y_1, ..., y_N) = p_{\boldsymbol{x}}\left(\boldsymbol{r}^{-1}(\boldsymbol{y})\right)\left|\det\left(J_{\boldsymbol{r}^{-1}}(\boldsymbol{y})\right)\right| \tag{B.7}$$

where $J_{\boldsymbol{r}^{-1}}(\boldsymbol{y})$ is the $N \times N$ Jacobian matrix (partial derivatives) of the inverse transform $\boldsymbol{r}^{-1}$.

Applying the change-of-variables formula above to the transformation $(h_{t_1}, ..., h_{t_N}) = \left(f_{\mathrm{RC}}^{-1}(Z_{t_1}), ..., f_{\mathrm{RC}}^{-1}(Z_{t_N})\right)$ yields the following formula:

$$\begin{aligned} p_{\boldsymbol{h}}(h_{t_1}, ..., h_{t_N}) &= p_{\boldsymbol{Z}}\left(f_{\mathrm{RC}}(h_{t_1}), ..., f_{\mathrm{RC}}(h_{t_N})\right)\left|\det\begin{pmatrix} f'_{RC}(h_{t_1}) & & 0 \\ & ... & \\ 0 & & f'_{\mathrm{RC}}(h_{t_N}) \end{pmatrix}\right| \\ &= p_{\boldsymbol{Z}}\left(f_{\mathrm{RC}}(h_{t_1}), ..., f_{\mathrm{RC}}(h_{t_N})\right)\prod_{k=1}^{N}\left|f'_{\mathrm{RC}}(h_{t_k})\right| \end{aligned} \tag{B.8}$$

which corresponds to the likelihood function from section 3.2 (eq. 24).