



HAL
open science

Impact of stage measurement errors on streamflow uncertainty

I. Horner, Benjamin Renard, Jérôme Le Coz, F. Branger, H.K. Mcmillan, G. Pierrefeu

► **To cite this version:**

I. Horner, Benjamin Renard, Jérôme Le Coz, F. Branger, H.K. Mcmillan, et al.. Impact of stage measurement errors on streamflow uncertainty. *Water Resources Research*, 2018, 54, pp.1952-1976. 10.1002/2017WR022039 . hal-02022803

HAL Id: hal-02022803

<https://hal.science/hal-02022803>

Submitted on 18 Feb 2019

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



RESEARCH ARTICLE

10.1002/2017WR022039

Impact of Stage Measurement Errors on Streamflow Uncertainty

I. Horner¹, B. Renard¹, J. Le Coz¹, F. Branger¹, H. K. McMillan^{2,3}, and G. Pierrefeu⁴

¹Irstea, UR RiverLy, 5 rue de la Doua CS 20244, 69625 Villeurbanne Cedex, France, ²National Institute of Water and Atmospheric Research, Hydrological Processes Group, PO Box 8602, Christchurch, New Zealand, ³San Diego State University, Department of Geography, San Diego, CA, USA, ⁴Compagnie Nationale du Rhône, Lyon, France

Key Points:

- Stage errors in gaugings and in stage time series affect streamflow computation
- Specific error models are introduced and used to compute streamflow time series uncertainties
- Effect of stage errors depends on site characteristics and averaging time intervals, systematic stage errors being the most impactful

Supporting Information:

- Supporting Information S1

Correspondence to:

I. Horner,
ivan.horner@irstea.fr;
J. Le Coz,
jerome.lecoz@irstea.fr

Citation:

Horner, I., Renard, B., Le Coz, J., Branger, F., McMillan, H. K., & Pierrefeu, G. (2018). Impact of stage measurement errors on streamflow uncertainty. *Water Resources Research*, 54, 1952–1976. <https://doi.org/10.1002/2017WR022039>

Received 11 OCT 2017

Accepted 7 FEB 2018

Accepted article online 26 FEB 2018

Published online 23 MAR 2018

Abstract Stage measurement errors are generally overlooked when streamflow time series are derived from uncertain rating curves. We introduce an original method for propagating stage uncertainties due to two types of stage measurement errors: (i) errors of the stage read during the gauging and (ii) systematic and nonsystematic (independent) errors of the recorded stage time series. The error models are generic and can be used for any probabilistic rating curve estimation method that provides an ensemble of rating curves. The new method is applied to a range of six contrasting hydrometric stations in France. Uncertainty budgets quantifying the contribution of various error sources to the total streamflow uncertainty are computed and compared for streamflow time series averaged at time intervals from hour to year. A sensitivity analysis is conducted on the stage time series error model to identify the most sensitive parameters. The results are site specific, which illustrates the key role played by the properties of both the hydrometric site and the gauged catchment. Across the range of sites, stage errors of the gaugings are found to have limited impact on rating curve uncertainty, at least for gaugings performed in fair conditions. Nonsystematic errors in the stage time series have a negligible effect, generally. However, systematic stage errors should not be neglected. Over the six hydrometric stations in this study, the 95% uncertainty component reflecting stage systematic errors (from ± 0.5 cm to ± 6.8 cm) alone ranged from 4% to 12% of daily average streamflow, and from 1% to 3% of yearly average streamflow as sensors were assumed to be recalibrated every 30 days. Perspectives for improving and validating the streamflow uncertainty estimation techniques are eventually discussed.

Plain Language Summary Water discharges in rivers are monitored by measuring the water level (known as stage) every 15 minutes or so and by converting it to discharge with a calibrated stage-discharge model (known as rating curve). We introduce an original method for computing the uncertainty of water discharge time series due to the measurement errors of the stage records used as inputs of the stage-discharge model. Systematic errors in stage measurements, for instance due to sensor drift or calibration errors, were found to produce non negligible discharge uncertainties. Over six hydrometric stations with varied conditions, the corresponding discharge uncertainty ranged from 4% to 12% of daily average discharge, and from 1% to 3% of yearly average discharge.

1. Introduction

Since streamflow time series serve as a basis for most hydrologic studies and water-related policy decisions, there has been a growing concern that they usually come without a full-fledged quantitative uncertainty analysis (Hamilton & Moore, 2012). Through real examples in Norway and New Zealand, McMillan et al. (2017) demonstrated that streamflow uncertainty analysis improves the economic efficiency and the public acceptance of water management policies. Streamflow values in a time series are seldom measured directly. Instead, a water level (known as “stage”) time series is measured and converted into a discharge time series through a stage-discharge model called a “rating curve.” The rating curve is estimated using paired stage and discharge measurements (known as “gaugings”) that are made occasionally by field hydrologists. Typically, on average 4–8 gaugings per year and per station are recommended by the French hydrological services (Puechberty et al., 2017). This number can be lower or higher depending on the stability of the river

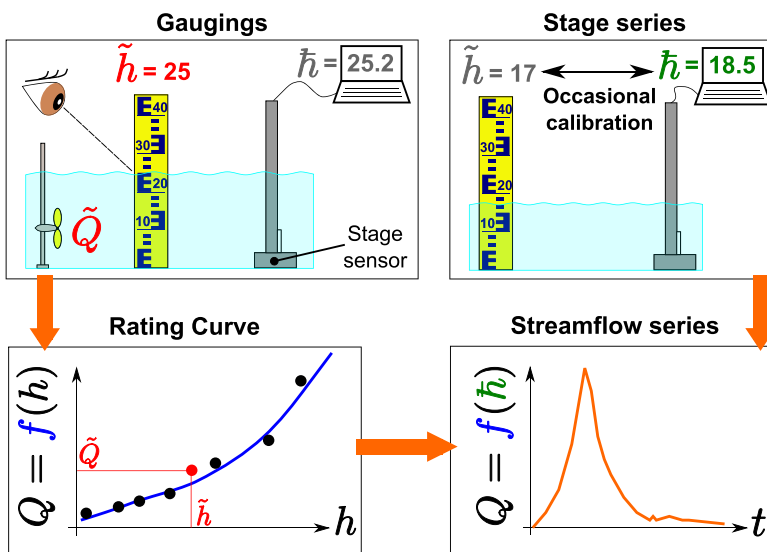


Figure 1. Schematic illustration of the process of computing a streamflow time series. (top left) Gaugings are used to establish (bottom left) a rating curve. (top right) The stage time series is then transformed into (bottom right) a streamflow time series using this rating curve. Importantly, note the distinct ways of measuring stage for gaugings and for the stage time series.

bed, the strategic importance of the station and other factors. The workflow typically followed to establish streamflow time series is illustrated in Figure 1. Two kinds of stage measurements must be distinguished: the discrete stage readings of the gaugings and the recorded stage time series measured by a stage sensor connected to a data logger.

The recommended practice is that the stage assigned to a gauging (\tilde{h} in Figure 1) is determined from a reading of the reference gauge, such as a staff plate, as required by, e.g., the Water Survey of Canada (F. Rainville, personal communication, 2017), the USGS (T. Kenney, personal communication, 2017), and the French hydrologic services (Puechberty et al., 2017). In some specific situations, the stage time series from the recorder may be used to assign a stage value to a gauging. This includes cases where direct reading of the reference gauge is impossible or difficult during floods, or when the stage fluctuates during the gauging. In any case, it is essential that the recorder is accurately aligned against the reference gauge. Most hydrometry manuals and publications are unfortunately silent on this issue. Nevertheless, assigning a stage to a gauging from a reading of the reference gauge is arguably the best practice and the procedure most commonly followed by field hydrologists around the world.

Stage time series (\hat{h} in Figure 1) are nowadays produced using a range of techniques, including traditional floats in a stilling well and pressure sensors (Petersen-Øverleir & Reitan, 2005), but also aerial ultrasonic and radar sensors, image analysis (Ran et al., 2016), and satellite telemetry (Birkinshaw et al., 2014; da Silva et al., 2010; Getirana & Peters-Lidard, 2013). During field visits, the stage differences between the recorder and the reference staff gauge are checked and the recorder is corrected if necessary: this is referred to as “occasional calibration” in Figure 1.

A number of stage error sources can be identified (see Hamilton & Moore, 2012; McMillan et al., 2012; Petersen-Øverleir & Reitan, 2005; Van der Made, 1982). Among the eight sources of stage measurement errors listed by Sauer and Turnipseed (2010), the first two affect the stage measurements using the reference gauge: *datum errors*, due to vertical movements and imperfect leveling of the staff gauge; and *gauge-reading errors*, due to poor lighting, very clear water, water surge, etc.

The other six sources of stage measurement errors affect the recorder only: *stage sensor errors*, related to the instrument performance and its measurement conditions; *water surface-to-sensor-to-recorder errors*, due to problems in the transmission of pressure, mechanical, or electric signals along the measuring system, especially through the intake of a stilling well if any; *hydraulically induced errors*, due to drawdown or buildup because of the interaction between the measuring structure and fast flows; *recorder errors*, due to mechanical issues or errors in the conversion of electric signals; *retrieval errors* when downloading the data to a field computer or a distant server; *verification errors* (or calibration errors), related to the correction of

the recorder against readings of the reference gauge. The magnitudes of all these error sources are difficult to quantify practically, as reflected by the comprehensive literature review published by McMillan et al. (2012). Uncertainties reported by users and manufacturers are often difficult to understand and compare due to the multiplicity of the studied effects, of the sources of information and of the underlying statistical assumptions, which are seldom made explicit.

While errors in the development and estimation of rating curves substantially contribute to streamflow uncertainties, most field hydrologists would confirm that errors in stage time series should not be neglected, and that their impact varies with the sensitivity of stage-discharge controls (Freestone, 1983; Petersen-Øverleir & Reitan, 2005). Accordingly, the uncertainties of stage time series were included in uncertainty propagation methods that are based on the first order Taylor series expansion of the stage-discharge model (Dymond & Christian, 1982; Herschy, 1999; ISO 1100-2:2010, 2010; Petersen-Øverleir & Reitan, 2005; Olivier et al., 2008; World Meteorological Organization, 2006). However, a major limitation of this approach is that it assumes independent stage errors and therefore neglects systematic errors, in particular those induced by the occasional calibration of the stage recorder.

Most importantly, stage errors have variable degrees of autocorrelation: such correlation is expected to affect the uncertainties of the streamflow averages at various time intervals, as well as statistical indicators of the hydrologic regime. Petersen-Øverleir and Reitan (2005) recognized the importance of correlated stage errors but did not include them in their computations. In fact, none of the uncertainty propagation methods we are aware of consider correlated stage errors. In the last decade, research has made substantial advances on rating curve uncertainty analysis (see Le Coz et al., 2014 for a review). Promising solutions have emerged, especially several approaches which can provide probabilistic results through the simulation of a large number of rating curves (referred to as an “ensemble”), within either Bayesian (Juston et al., 2014; Le Coz et al., 2014; Moyeed & Clarke, 2005; Reitan & Petersen-Overleir, 2009; Sikorska et al., 2013) or non-Bayesian (Coxon et al., 2015; McMillan & Westerberg, 2015; Morlot et al., 2014; Westerberg et al., 2011) frameworks. However, none of these published methods account for the propagation of the errors in input stage records, which are assumed to be negligible or are merely ignored.

The aims of this paper are therefore the following:

1. To introduce an original error model for continuously measured river stage.
2. To demonstrate how this new error model can be integrated into a full analysis of streamflow uncertainty to determine the relative contributions of stage uncertainty sources, at averaging time intervals from hour to year.
3. To apply the method to a range of contrasting sites with varied hydrologic regimes to quantify the importance of stage uncertainty to total streamflow uncertainty at these sites.
4. To investigate how sensitive the results are with respect to the main assumptions and parameters of the stage time series error model.

The method is based on generic stage error models that can be used with any probabilistic method that provides an ensemble of rating curves. Errors affecting the stages of the gaugings are taken into account, as well as the errors of the stage time series, decomposed into systematic and nonsystematic error components. The remainder of the paper is organized as follows. The new method is presented in section 2. The model of stage errors in gaugings is presented in section 2.1, and section 2.2 explains how they are accounted for in the estimation of the rating curve. The model of stage errors in time series and the uncertainty propagation to streamflow time series are presented in sections 2.3 and 2.4, respectively. In section 3, the new method is applied to a range of six contrasting hydrometric stations in France to assess the importance of stage errors in gaugings (section 3.2) and in time series (section 3.3). The sensitivity of the results to the main parameters of the stage time series error model is eventually investigated (section 3.4). Finally, the results and some perspectives for improving and validating the streamflow uncertainty estimation techniques are discussed in section 4.

2. Method for Propagating Stage Measurement Errors to Streamflow Time Series

2.1. Error Model of Stage and Discharge Measurements in Gaugings

A gauging j can be defined as a pair of two values $(\tilde{h}_j, \tilde{Q}_j)$ where \tilde{h}_j and \tilde{Q}_j are the measured stage and discharge, respectively. In this paper, we restrict our attention to stage measurements corresponding to direct

readings of the reference staff gauge, since this is the most common and recommended practice as explained in section 1. Both the stage and the discharge measurement errors are assumed to be independent and Gaussian with mean zero and standard deviations σ_j^h and σ_j^Q , respectively. For a given gauging j , the error model expresses \tilde{h}_j and \tilde{Q}_j as

$$\tilde{h}_j = h_j + \epsilon_j^h \text{ with } \epsilon_j^h \sim \mathcal{N}(0, \sigma_j^h) \quad (1)$$

$$\tilde{Q}_j = Q_j + \epsilon_j^Q \text{ with } \epsilon_j^Q \sim \mathcal{N}(0, \sigma_j^Q) \quad (2)$$

where h_j and Q_j are the unknown true stage and discharge, respectively.

Standard deviations σ_j^h and σ_j^Q have to be specified for each gauging. Typically, larger standard deviations are specified for larger stages or discharges. Discharge uncertainties in gaugings depend on the measurement procedure, the instrumentation and the site conditions (Le Coz et al., 2014). Typical uncertainty values can be found in the literature (McMillan et al., 2012). In most cases, stage uncertainties can be specified using expert knowledge. If available, experimental data such as repeated staff gauge readings can help estimate the standard deviation. As a last resort, information gathered in the literature can also be used to estimate the stage uncertainty of gaugings. Note that we chose this simple error model for stage errors in gaugings, as these measurements typically only contain datum and gauge-reading errors (see section 1). Continuously measured stage is subject to further error sources and is treated with a more complex error model (see section 2.3).

2.2. Uncertainty Analysis of the Rating Curve

The rating curve uncertainty is assessed using the existing BaRatin (Bayesian Rating curve) method (cf., Le Coz et al., 2014 for a detailed description). Note that it would be possible to use any other probabilistic method that provides an ensemble of rating curves. The BaRatin framework has been developed to build stage-discharge rating curves and assess their uncertainty. It is based on three main components:

1. *A model.* A mathematical representation of the stage-discharge relation is specified using preliminary hydraulic analysis of the hydrometric station. Prior distributions are defined for all the parameters based on a hydraulic analysis.
2. *Observations.* The gaugings (see section 2.1) are used as the uncertain observations.
3. *Inference.* Bayesian inference and MCMC (Markov Chain Monte Carlo) simulations are applied to compute and sample the joint posterior distribution of the rating curve parameters, which combines the prior knowledge of the rating curve parameters and the information contained in the gaugings.

In the following subsections describing the application of the BaRatin method, we consider two scenarios: (1) gaugings are considered free of stage errors and (2) gaugings are considered to include stage errors, which are propagated using a Monte Carlo approach.

2.2.1. Ignoring Stage Errors in Gaugings

We first assume that the stage measurements of the gaugings are free of errors: $\tilde{h}_j = h_j$. For a given gauging (h_j, \tilde{Q}_j) , the relation between stage h_j and discharge \tilde{Q}_j can be written as

$$\tilde{Q}_j = \underbrace{f(h_j|\theta)}_{\tilde{Q}_j} + \epsilon_j^f + \epsilon_j^Q \text{ with } \epsilon_j^f \sim \mathcal{N}(0, \gamma_1 + \gamma_2 \hat{Q}_j) \quad (3)$$

where f is the rating curve equation and θ its vector of parameters.

The structural error term, ϵ_j^f , accounts for imperfections of the rating curve equation as a model of the real stage-discharge relation (see Le Coz et al., 2014). It has two unknown parameters, γ_1 and γ_2 , which are estimated through the Bayesian inference, using wide uniform prior distributions $\mathcal{U}(0; 10, 000)$. The term ϵ_j^Q is the gauging discharge error term (see equation (2)). Errors ϵ^f and ϵ^Q are assumed to be mutually independent and also independent between gaugings.

The inferred posterior distribution of the parameters is explored using a MCMC sampling technique (see Renard et al., 2006). MCMC simulations give N_{sim} vectors of parameters $(\theta_i, \gamma_{1,i}, \gamma_{2,i})_{i=1:N_{sim}}$, or "samples." Each vector θ_i determines a possible rating curve i . The first half of them is discarded as a "burn-in" period, then every tenth parameter vector is selected to reduce autocorrelation between successive samples, and

therefore reduce computing time and storage issues. This results in n vectors of parameters subsampled from the N_{sim} vectors ($n \ll N_{sim}$): $(\theta_i, \gamma_{1,i}, \gamma_{2,i})_{i=1:n}$.

The “MaxPost” rating curve is computed with the vector of parameters θ_{MP} corresponding to the maximum of the posterior distribution. The MaxPost rating curve is thus defined as $Q_{MP} = f(h|\theta_{MP})$.

For a given stage value h , n values of discharge can be computed through the n possible rating curves (i.e., the n vectors of parameters); statistics on those values provide the parametric uncertainty, i.e., the uncertainty due to the estimation of the parameters of the stage-discharge model. The total uncertainty is obtained after adding a structural error sampled in a Gaussian distribution with mean zero and standard deviation $\gamma_{1,i} + \gamma_{2,i} \times f(h|\theta_i)$ to each rating curve i ($i=1, \dots, n$).

2.2.2. Accounting for Stage Errors in Gaugings

A Monte Carlo approach is used to propagate errors in stage measurements of gaugings. It is based on repeating the BaRatin estimation described in previous section 2.2.1 for many replicated sets of gaugings with additional stage errors sampled in a Gaussian distribution of mean zero and standard deviation σ_j^h specified from the stage uncertainty. More precisely, the following pseudo-algorithm describes this Monte Carlo approach:

For $k=1, \dots, m$:

- a. sample gaugings stage errors $(\epsilon_{j,k}^h)_{j=1, \dots, N_{gaugings}} \sim \mathcal{N}(0, \sigma_j^h)$,
- b. create the k th replicated set of gaugings $(h_{j,k}, \tilde{Q}_j)_{j=1, \dots, N_{gaugings}}$ with $h_{j,k} = \tilde{h}_j - \epsilon_{j,k}^h$, and
- c. apply the BaRatin estimation described in previous section 2.2.1, yielding MCMC-simulated vectors of parameters $(\theta_{i,k}, \gamma_{1,i,k}, \gamma_{2,i,k})_{i=1, \dots, n}$.

Merging MCMC-simulated parameter vectors obtained across the m Monte Carlo simulations yields an ensemble of $m \times n$ parameter vectors $(\theta_{i,k}, \gamma_{1,i,k}, \gamma_{2,i,k})_{i=1, \dots, n, k=1, \dots, m}$. The variability in this ensemble includes the effect of the uncertainty in the stage measurement of the gaugings. It can be used in the same way as the MCMC simulation described in previous section 2.2.1: thinning to avoid computing time and storage issues, computation of MaxPost parameters and rating curve, generation of ensembles of rating curves representing parametric or total uncertainty. In mathematical terms, the Monte Carlo samples $(\theta_{i,k}, \gamma_{1,i,k}, \gamma_{2,i,k})_{i=1, \dots, n, k=1, \dots, m}$ are realizations from the distribution with probability density function (pdf):

$$g(\theta, \gamma_1, \gamma_2) = \int p(\theta, \gamma_1, \gamma_2 | \mathbf{h}, \tilde{\mathbf{Q}}) p(\mathbf{h} | \tilde{\mathbf{h}}) d\mathbf{h} \tag{4}$$

The first term in the integral corresponds to the posterior pdf for a given realization of stages (step (c) in the algorithm above), while the second term is the Gaussian pdf corresponding to the stage error model (steps (a)–(b)). We refer to the discussions in Sikorska and Renard (2017) for a similar derivation and a discussion of alternative approaches.

2.3. Error Model of Stage Measurements in Stage Time Series

As stated in section 1, stage time series are affected by various error sources. We choose to consider two types of errors: nonsystematic errors (waves, instrumental noise) and systematic errors (unknown instrumental biases and calibration drifts over time). We introduce a simple error model with two error terms: a nonsystematic error δ_{NS}^h and a systematic error δ_S^h . Considering the measured stage time series $\tilde{h}(t)$ and the unknown true stage time series $h(t)$, the following stage time series error model can be introduced:

$$\tilde{h}(t) = h(t) + \delta_{NS}^h(t) + \delta_S^h(t) \tag{5}$$

At every time t , $\delta_{NS}^h(t)$ is sampled from a Gaussian distribution $\mathcal{N}(0, \sigma_{NS}^h)$. It is uncorrelated from one time step to another.

At specific times t_p , $\delta_S^h(t_p)$ is sampled from a Gaussian distribution $\mathcal{N}(0, \sigma_S^h)$. Between these specific times, the error is “systematic,” i.e., it varies in a deterministic way between the sampled errors. Two error behaviors are tested:

- 1. Constant error model. For $t_p \leq t < t_{p+1}$, $\delta_S^h(t) = \delta_S^h(t_p)$.
- 2. Linear error model. For $t_p \leq t < t_{p+1}$, $\delta_S^h(t) = \delta_S^h(t_p) + \frac{t-t_p}{t_{p+1}-t_p} \times (\delta_S^h(t_{p+1}) - \delta_S^h(t_p))$.

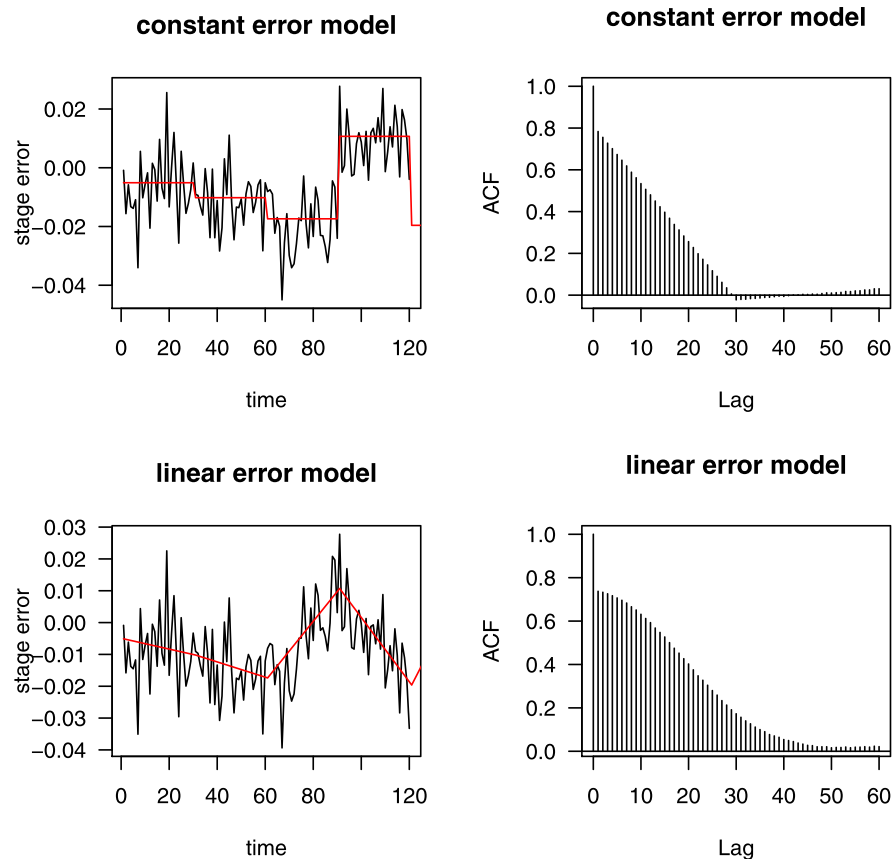


Figure 2. (left) Realizations of systematic (in red) and total (in black) errors associated with the (top) constant and (bottom) linear versions of the systematic error model; (right) autocorrelation functions (ACF) of total errors.

We stress that both error models generate autocorrelated errors. As an illustration, the left column of Figure 2 shows realizations of the systematic and the “total” (systematic + nonsystematic) errors, with the “constant” and “linear” versions of the systematic error model. By computing the autocorrelation of the total errors at various lags, we obtain the autocorrelation functions (ACFs) shown in the right column. These ACFs illustrate the highly autocorrelated nature of the total errors. Figure 3 further illustrates how these sampled errors corrupt the measured stage time series.

The approach of decomposing systematic and nonsystematic errors, as embedded into equation (5), leads to a model that can be interpreted in terms of operational practice, and whose properties can therefore easily be specified as discussed subsequently. By contrast, more standard models for autocorrelated errors (typically, autoregressive models) lead to a more challenging specification. For instance, using an order-1 autoregressive model would require specifying a value for the lag-one autocorrelation, which is difficult to relate to operational practices.

Other models for $\delta_S^h(t)$ may be set up but we lack the precise information on the recorder bias and drift that would be necessary for defending alternative models (for further discussion, see section 4.4). Ideally, the sampling times t_p correspond to recalibration of the stage sensor against the reference staff gauge. In the absence of such information, one can consider usual practices typical of a hydrologic service and of a hydrometric station to set a default resampling periodicity.

Equation (5) requires the specification of two standard deviations: σ_{NS}^h and σ_S^h . Depending on the site, different sources of information can be used to specify these two values:

1. Available data such as repeated measurements of the same stage (to specify σ_{NS}^h), observed differences between the sensor and the reference gauge (to specify σ_S^h), etc.

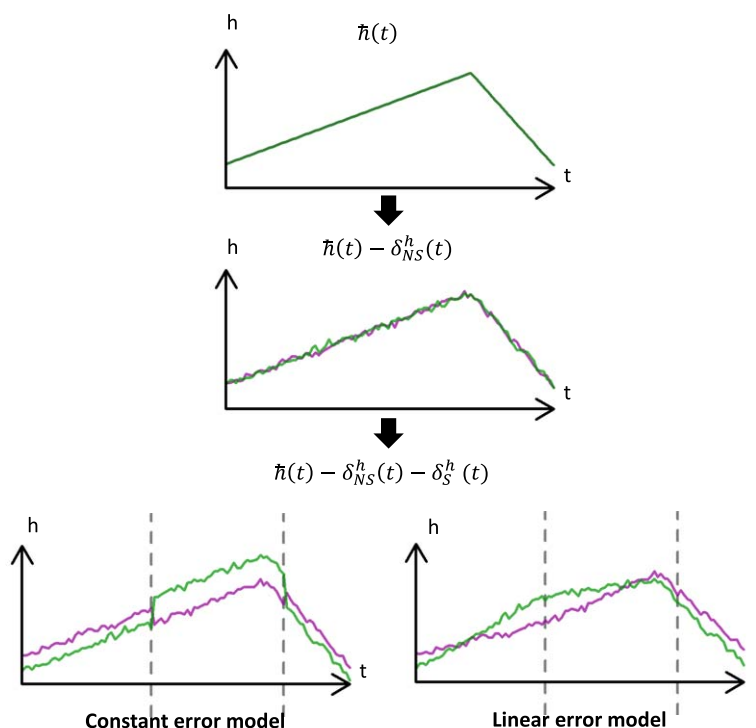


Figure 3. Example of two realizations (green and purple) for the nonsystematic, $\delta_{NS}^h(t)$, and systematic, $\delta_S^h(t)$, stage time series errors from a fictitious measured stage time series $\hat{h}(t)$. Dashed vertical lines show times t_p when the systematic error is resampled.

2. Expert knowledge on typical errors that may affect stage measurements.
3. Literature information.

2.4. Uncertainty Propagation From Stage to Streamflow

Using equation (5), n samples of the stage time series are computed from the measured stage time series. For simplicity and without loss of generality, we choose to have the same number n of samples for the stage time series as the n parameter vectors obtained using the BaRatin method (see section 2.2): $(\theta_i, \gamma_{1,i}, \gamma_{2,i})_{i=1:n}$. Then, n streamflow time series are computed using the following equation:

$$Q_i(t) = f(\underbrace{\hat{h}(t) - \delta_{NS,i}^h(t) - \delta_{S,i}^h(t)}_{\hat{Q}_i(t)} | \theta_i) + \epsilon_i^f(t) \quad \text{with } \epsilon_i^f \sim \mathcal{N}(0, \gamma_{1,i} + \gamma_{2,i} \hat{Q}_i(t)) \quad (6)$$

The computation process, illustrated in the left column of Figure 4 accounts for the various contributors to streamflow uncertainty:

1. sampling of $\delta_{NS,i}^h(t)$ reflects the stage uncertainty due to the nonsystematic errors affecting stage time series,
2. sampling of $\delta_{S,i}^h(t)$ reflects the stage uncertainty due to systematic errors affecting stage time series,
3. sampling of θ_i reflects the parametric uncertainty of the rating curve, and
4. sampling of $\epsilon_i^f(t)$ reflects the structural uncertainty of the rating curve.

It is also possible to propagate one source of uncertainty only, as illustrated in the other columns of Figure 4. This will be useful to quantify the individual contribution of each uncertainty source, as will be described in section 2.5.3.

The *MaxPost* streamflow time series is computed using the *MaxPost* rating curve (see section 2.2) and the measured stage time series $\hat{h}(t)$ with no errors added: $Q_{MP}(t) = f(\hat{h}(t) | \theta_{MP})$. The *MaxPost* streamflow time series is the baseline that ignores any uncertainty, i.e., the current standard approach used in most hydrological modeling endeavors.

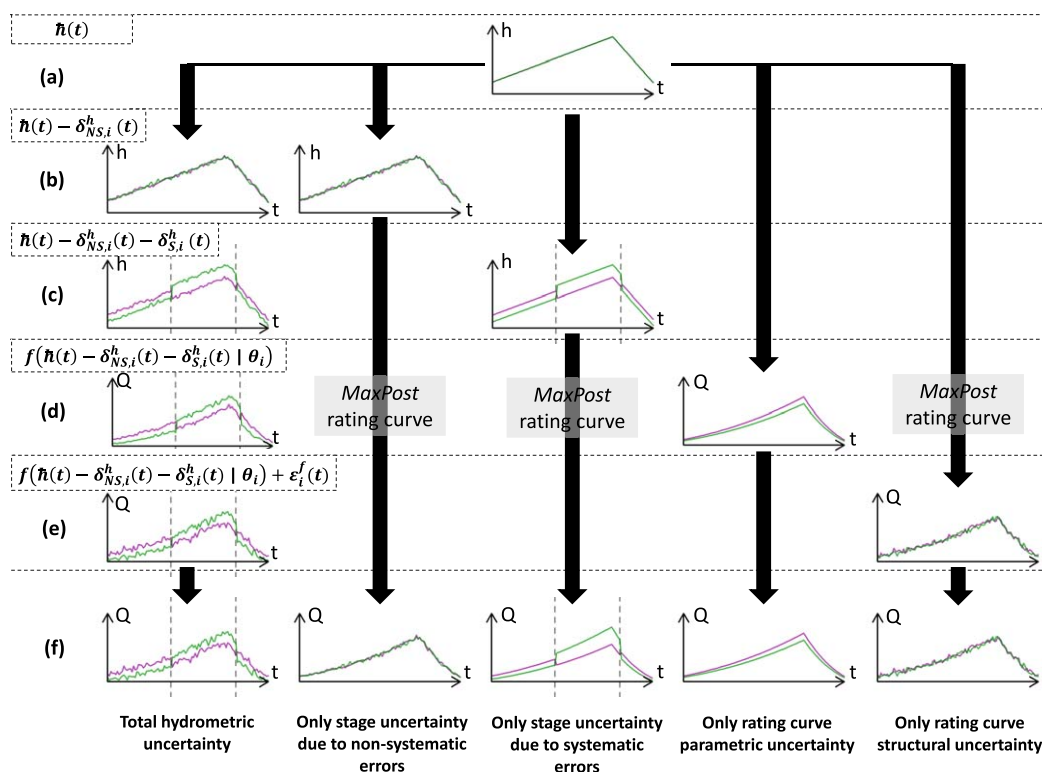


Figure 4. Schematic of the sampling process used to propagate all considered uncertainty sources. The two distinct colors (green and purple) are used to differentiate two realizations. The first column shows the procedure when all the uncertainty sources are considered: (a) measured stage time series, (b) adding the stage nonsystematic errors, (c) adding the stage systematic errors using the “constant error” model, (d) computation of the streamflow time series including the rating curve parametric uncertainty, and (e) adding the structural errors. The last row (f) shows the resulting streamflow time series samples. The other columns show the sampling procedure when only one uncertainty source is considered. Dashed verticals indicate the times when systematic stage errors are resampled. Measured stage $h(t)$ is intentionally shown as perfectly smooth, for clarity.

Note that gauging uncertainties do not appear in equation (6) and therefore play no direct role in this propagation step. Uncertainties in gaugings stages do not appear because the stage time series is used here, not the gaugings stages, and they are measured using distinct methods (see Figure 1). Uncertainties in gaugings discharges do not appear because the aim is to estimate the unknown true discharge, as opposed to the discharge that might have been measured with a gauging. However, gauging uncertainties play an indirect role as they were used to estimate the rating curve, and therefore affect the posterior distribution of $(\theta, \gamma_1, \gamma_2)$.

2.5. Postprocessing of Streamflow Time Series

2.5.1. Averaging of Streamflow Time Series to Various Time Intervals

The original time series may be recorded at constant or variable time steps. This instantaneous streamflow time series is often time-integrated using the trapezoidal rule to compute the water volumes of given time intervals. The water volumes are then divided by the time intervals to compute the mean streamflow values.

The uncertainty propagation described in the previous section results in n streamflow time series. At a given time step, uncertainty can be characterized by statistics computed on the n corresponding streamflow values (e.g., variance or quantiles to derive an uncertainty interval). Time averaging of an uncertain streamflow time series requires time averaging each of the n samples, which produces n samples of time-averaged streamflow time series. Exactly as the original time series, at a given time step, uncertainty is characterized by statistics computed on the n corresponding streamflow values.

2.5.2. The Total Uncertainty

Throughout the paper, uncertainty is generally expressed as the half-length of the 95% uncertainty interval divided by the *MaxPost* streamflow value. Using the 95% uncertainty interval is common practice in many measurement areas and has been recommended by the Hydrometric Uncertainty Guidance (ISO/TS25377:2007, 2009). The total uncertainty of streamflow data combines all the considered uncertainty sources: systematic and nonsystematic stage errors in time series and parametric and structural rating curve errors. For a given time step t , the total uncertainty U_T can be expressed as

$$U_T(t) = \frac{q_{0.975}[Q_{1,\dots,n}(t)] - q_{0.025}[Q_{1,\dots,n}(t)]}{2Q_{MP}(t)} \tag{7}$$

where $q_p[Q_{1,\dots,n}(t)]$ is the p -quantile of the n streamflow values and $Q_{MP}(t)$ is the *MaxPost* streamflow value at time t . Throughout the paper, U_T is expressed in percentage.

Note that the same way of expressing uncertainty can also be used with partial uncertainties, i.e., when not all of the uncertainty sources are included. In equation (7), the streamflow values $Q_{1,\dots,n}(t)$ can be replaced by the results of a simulation in which only one source of uncertainty is propagated (see Figure 4).

2.5.3. The Uncertainty Budget

While using probability intervals is the standard way of representing uncertainties, it is not convenient if one is interested in the relative contributions of various sources of uncertainty. This is because probability intervals are not additive, in the following sense: the width of an interval representing two sources of uncertainty is not equal to the sum of the widths of the two intervals representing a single source of uncertainty (in general, the former is smaller than the later). To see this, consider the simple case of two independent Gaussian random variables V_1 and V_2 (representing two sources of uncertainty), combined through a simple addition to yield the random variable $W = V_1 + V_2$. The total uncertainty can be represented by the half-length of the 95% probability interval for W , which is equal to $U_{Total} = 1.96\sigma_W = 1.96\sqrt{\sigma_{V_1}^2 + \sigma_{V_2}^2}$ (because of the Gaussian and independence assumptions). A partial uncertainty representing only the first source of uncertainty can be obtained by fixing the second random variable to a particular value v , yielding the half-length $U_1 = 1.96\sigma_{V_1+v} = 1.96\sigma_{V_1}$ (because v is a constant). Similarly, $U_2 = 1.96\sigma_{v+V_2} = 1.96\sigma_{V_2}$. Under this setup, we therefore have $U_{Total} < U_1 + U_2$, because $\sqrt{\sigma_{V_1}^2 + \sigma_{V_2}^2} < \sigma_{V_1} + \sigma_{V_2}$ (see Steinschneider et al., 2012 for a similar discussion). We note however that the variances are additive ($Var(W) = Var(V_1) + Var(V_2)$): it is therefore sensible to build an uncertainty budget representing the contribution of various sources based on variances, rather than probability intervals.

In this paper, the uncertainty budget is calculated using the full simulation where all uncertainty sources are considered (column 1 in Figure 4) and four simulations where only one source of uncertainty is propagated (see columns 2–5 in Figure 4). Consequently, we obtain five distinct ensembles of streamflow time series: (1) $Q_{1,\dots,n}(t)$, (2) $Q_{1,\dots,n}^{\sigma_{NS}}(t)$, (3) $Q_{1,\dots,n}^{\sigma_S}(t)$, (4) $Q_{1,\dots,n}^{\theta}(t)$, and (5) $Q_{1,\dots,n}^{(\gamma_1, \gamma_2)}(t)$. They stand for propagation of (1) all the uncertainty sources, (2) nonsystematic errors in the stage time series only, (3) systematic errors in the stage time series only, (4) rating curve model parameters only, and (5) rating curve structural errors only.

Let $Var(X_{1,\dots,n})$ denote the variance computed from n realizations from the random variable X . We define the uncertainty budget $UB(t)$ for a given time t as the following vector of size 4:

$$UB(t) = \left(\frac{Var(Q_{1,\dots,n}^{\sigma_{NS}}(t))}{Var(Q_{1,\dots,n}(t))}, \frac{Var(Q_{1,\dots,n}^{\sigma_S}(t))}{Var(Q_{1,\dots,n}(t))}, \frac{Var(Q_{1,\dots,n}^{\theta}(t))}{Var(Q_{1,\dots,n}(t))}, \frac{Var(Q_{1,\dots,n}^{(\gamma_1, \gamma_2)}(t))}{Var(Q_{1,\dots,n}(t))} \right) \tag{8}$$

In the paper, the uncertainty budget is shown as stacked color areas with specific colors for each source of uncertainty: blue for stage uncertainty due to nonsystematic errors, light blue for stage uncertainty due to systematic errors, light red for the parametric uncertainty and red for the structural uncertainty of the rating curve.

Note that the four components of the uncertainty budget do not add up to exactly one. This is because the total variance (denominator) is in general not equal to the sum of partial variances (numerators). Indeed, unlike in the simplistic example used at the beginning of this section, the various uncertainty sources are not combined through a simple addition but instead are propagated through a nonlinear model (the rating

curve). However, experience suggests that the sum of the 4 components is in general fairly close to 1 (see case study of section 3), and we therefore consider that the uncertainty budget is fit for purpose.

3. Case Study

3.1. Study Sites

The proposed method for propagating streamflow uncertainties is applied to a range of six contrasting hydrometric stations in France (Figure 5). The sites are selected to cover a range of conditions: sensitivity of the hydraulic controls, quality of the rating curve, number, range and uncertainty of the gaugings, hydrologic regimes, etc. All the hydrometric stations rely on a simple stage-discharge relation that was checked to be stable over at least 1 year. The numbers of gaugings reported hereafter and in Figure 5 are representative of that stage-discharge relation.

The selected sites for this study are succinctly described below:

1. *The Rhône River at Bognes*. This station monitors a large regulated river. A natural riffle controls the stage-discharge relation at low flows. For higher flows, a channel control (see Appendix A for definition) takes over. More than 130 gaugings are available, most of them done using an Acoustic Doppler Current Profiler (ADCP). Upstream river regulation influences most of the flow regime with constant streamflow periods that are the consequence of opening and closing of Génissiat Dam conduits. Floods are less affected by the regulation, though partial flood retention can be operated with the reservoir behind Génissiat Dam.
2. *The Ardèche River at Sauze*. Two controls are identified for this station: a natural riffle at low flow and a channel control for higher flows. The 40 available gaugings are mostly done using an ADCP. The hydrological regime is Mediterranean with fast floods and long periods with very low flows.
3. *The Ardèche River at Vallon*. This hydrometric station is located 30 km upstream of the Sauze hydrometric station. Two hydraulic controls are identified: a natural riffle at low flow and a channel control for higher flows. The 55 gaugings were mostly done using an ADCP except for the highest gaugings where noncontact techniques such as surface velocity radar (SVR) measurements were used. The flow regime is very similar to the one monitored by the Sauze station, given that there are no major tributaries between the two stations.
4. *The Blies River at Bliesbruck*. The station monitors a lowland catchment with slowly varying flows. The low flow control, a natural riffle, has a poor sensitivity, i.e., stage variations are small with respect to variations of discharge. The small number of gaugings makes the high-flow section of the rating curve quite uncertain.
5. *The Yzeron River at Taffignon*. This station monitors a small peri-urban catchment with long dry periods and sudden floods during the wet months. Despite being an artificial structure, the low-flow control has a poor sensitivity. Many gaugings (120) are available for this station.


Site	Bognes	Sauze	Vallon	Bliesbruck	Taffignon	Lancône
River	Rhône	Ardèche	Ardèche	Blies	Yzeron	Bévinco
Catchment [km^2]	10910	2258	1958	1815	129	53.6
Lat. N [°]	46.03771	44.31401	44.39790	49.11626	45.73422	42.60319
Long. E [°]	5.808570	4.551105	4.385578	7.177730	4.771914	9.379985
						
Q_{mean} [$m^3 \cdot s^{-1}$]	359	64.8	58.1	18.3	0.663	0.705
Q_{max} [$m^3 \cdot s^{-1}$]	1128	1834	1601	130	29	44
AM_{10} [$m^3 \cdot s^{-1}$]	157	6.41	3.77	6.1	0.006	0.056
Low flow control	natural	natural	natural	natural	artificial	artificial
Nb of gaugings	135	40	55	26	70	120
Gauged flows [$m^3 \cdot s^{-1}$]	2.29 - 1592	4.4 - 3180	2.74 - 1660	5.04 - 190	0.007 - 38.5	0.0037 - 24.8
Stage records since	1969	1980	1979	2007	1988	1969

Figure 5. Description of the six hydrometric stations studied. Q_{mean} , Q_{max} , and AM_{10} respectively stand for the interannual mean of annual means, annual maxima, and annual minima over 10 days. The streamflow values reported here are the official values published by the French national hydrometric services (date of retrieval: June 2016).

Table 1
Specification of the Properties of the Error Models for Gauging Stages and Stage Time Series

	Bognes	Sauze	Vallon	Bliesbruck	Taffignon	Lancone
<i>Gauging stages</i>						
σ_j^h (m)	0.01–0.1	0.01–0.1	0.01–0.1	0.005–0.05	0.005–0.05	0.005–0.05
<i>Stage times series</i>						
σ_{NS}^h (m)	0.01	0.01	0.01	0.01	0.0025	0.0025
σ_S^h (m)	0.034	0.015	0.015	0.015	0.0025	0.0025
Resampling periodicity (day)	30	30	30	30	30	30

6. *The Bévinco River at Lancone.* This station monitors a small mountainous Mediterranean catchment in Corsica. Streamflow is usually very small apart from a few flash floods. Hydraulic controls are a specific case: an artificial structure combines several concrete rectangular weirs. Many gaugings are available for this station with some significant scatter probably caused by errors affecting the gaugings.

As explained by Le Coz et al. (2014), the application of BaRatin requires performing a preliminary hydraulic analysis of each station in order to derive the rating curve equation and specify prior parameters. This is detailed in Appendix A.

The two following paragraphs describe the specification of (i) the gaugings discharges and stages uncertainties and (ii) the stage time series uncertainties. The properties of the stage error models are summarized in Table 1. Note that the stage error values gathered in Table 1 only apply for these hydrometric stations. They should not be seen or used as representative and/or typical stage error values.

Stage uncertainties in gaugings are considered for all stations. Due to the lack of information, the standard deviations are specified based on expert judgment for all the gaugings. For a given station, the specification is based on a minimum uncertainty value for the lowest gauging and a maximum uncertainty value for the highest gauging. This intends to reflect that errors in stage readings are generally larger in high-flow conditions. The uncertainty of each gauging is then calculated by linear interpolation with respect to the stage value. As for discharge uncertainty in gaugings, we used different relative uncertainties according to the

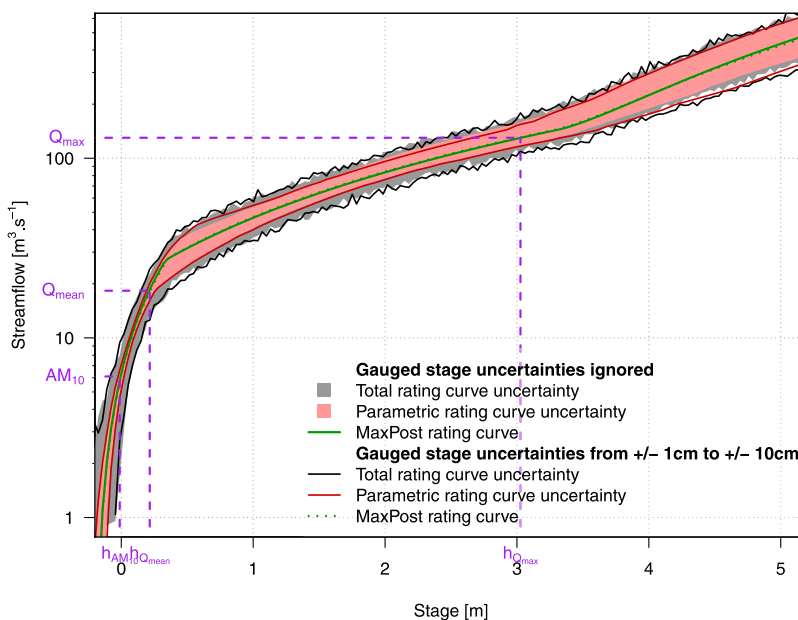


Figure 6. Comparison of the estimated rating curves and associated uncertainties at Bliesbruck with and without taking into account gaugings stage uncertainties. The dashed purple lines indicate low, medium and high streamflow values (AM_{10} , Q_{mean} , and Q_{max} , respectively, see Figure 5) and the corresponding stage values, $h_{AM_{10}}$, $h_{Q_{mean}}$, and $h_{Q_{max}}$.

technique used: $\pm 5\%$ for ADCP and volumetric method, $\pm 7\%$ for velocity area method, $\pm 15\%$ for surface velocity methods or when the method used is unknown.

Ideally, the nonsystematic and systematic errors in stage time series should also be varied with stage but this is practically impossible to specify due to the lack of observations on stage time series errors in high-flow conditions, as recalibration of the recorder is usually done in low flow conditions in France (cf., discussion in section 4.4). As a result, the uncertainties of stage time series are assumed to be constant and they are specified using expert knowledge at all sites. The only exception is Bognes for which a data set of observed differences between the reference gauge and the recorder is available: the uncertainty of systematic stage errors is therefore computed as the standard deviation of these differences. An average periodicity of 30 days for resampling the systematic errors is set for all the sites. If not explicitly stated, the constant error model is used. Finally, note that only 1 year of records is used for each station. The year was selected for each station in order to have no or very few missing values. The year 2013 was chosen for stations Bognes and Sauze and the year 2014 for stations Vallon, Bliesbruck, Taffignon, and Lancone.

3.2. Impact of Stage Errors in Gaugings

Whether the uncertainty of gaugings stages impacts the estimation of the rating curve is tested by comparing the results assuming error-free stages (section 2.2.1) with those of the Monte Carlo approach that propagates

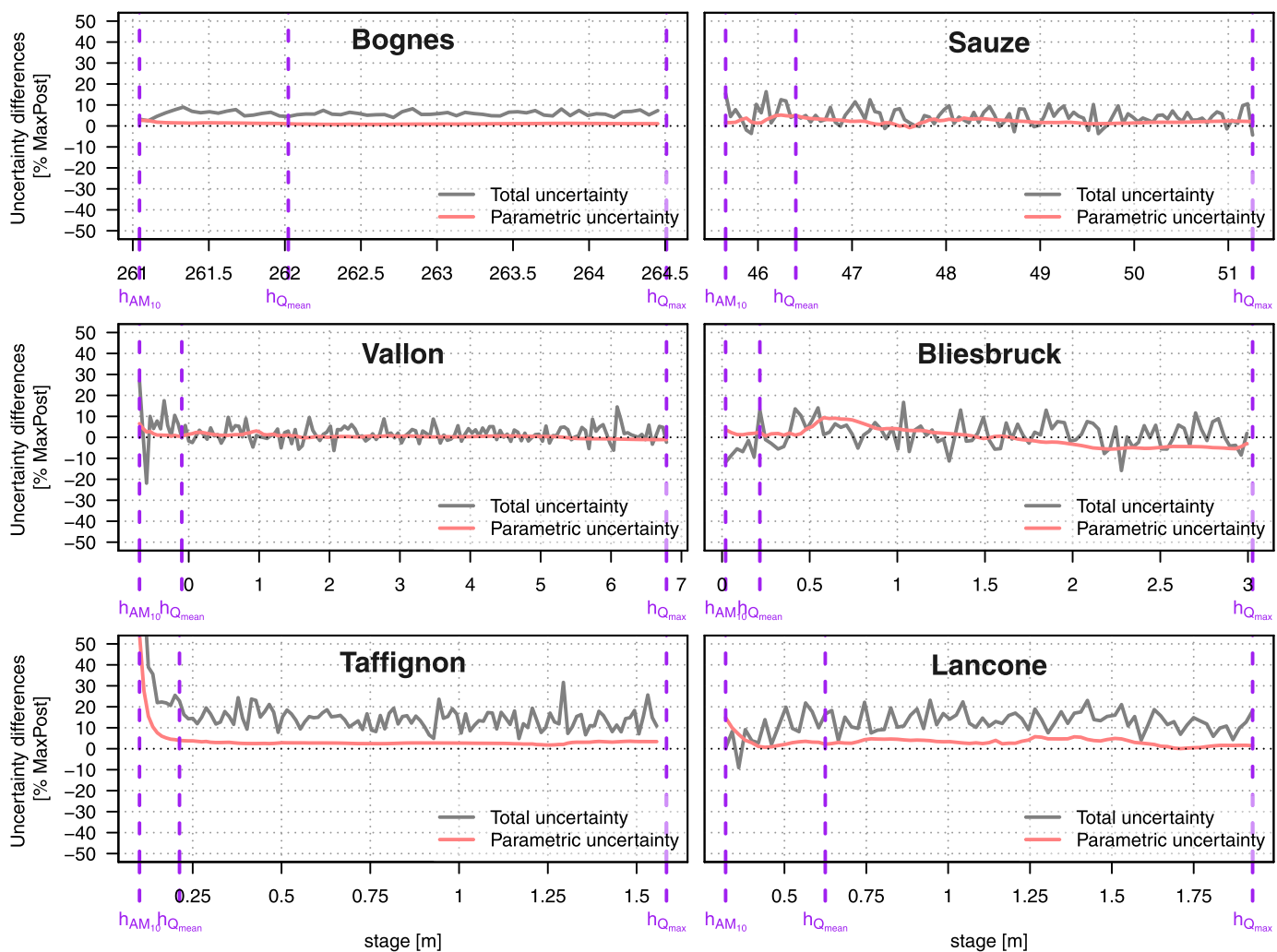


Figure 7. Relative differences between uncertainty U_T computed without gaugings stage errors and uncertainty U_T computed with gaugings stage errors (U_T is defined in equation (7)). The differences are normalized by the $MaxPost$ rating curve Q_{MP} computed without gaugings stage errors. The dashed purple lines illustrate stage values corresponding to low (AM_{10}), medium (Q_{mean}), and high (Q_{max}) streamflow values (see Figure 5).

stage errors of the gaugings (section 2.2.2). Results are first shown for one station (Figure 6) in detail and for all the six stations (Figure 7) in a summarized form. For the Bliesbruck station (Figure 6), differences with or without gaugings stage uncertainties are mostly negligible. One can note that parametric uncertainty for the highest stages is slightly higher when gaugings stage uncertainties are taken into account.

The six studied stations show varied increases of the total and parametric uncertainties of their rating curves when stage errors in gaugings are included (Figure 7). The increases in parametric uncertainty are usually limited to less than +5%, except at Bliesbruck near an ill-identified control transition and at Taffignon because low flows tend to nearly zero. Increases in total uncertainty may be more substantial, around +10 % typically, and even more for Taffignon and Lancone hydrometric stations. These two stations happen to have wide, shallow river channels, where water stage is much lower than the other sites, hence the relative importance of the gaugings stage errors.

The overall conclusion of this analysis is that gaugings stage uncertainties have a limited effect on the estimated rating curve and its parametric uncertainty, and a slightly larger effect on the total uncertainty (see Taffignon and Lancone in Figure 7). Of course this conclusion holds only for the uncertainties specified here, which we consider as representative of gaugings conducted in fair conditions. In specific situations where stage uncertainties in gaugings are much larger, the effect could be more pronounced.

3.3. Impact of Stage Errors in Time Series

First, we focus on one station (section 3.3.1) in order to properly introduce the results summed up for all the six stations in section 3.3.2. Note that throughout this section, the rating curves have been estimated with stage uncertainties in the gaugings being accounted for (see section 3.2).

3.3.1. The Bliesbruck Case

Figure 8 focuses on the Bliesbruck station for 1 year of data (2014). It displays the uncertainty results at daily intervals as a function of sorted daily streamflow averages. Very similar results are obtained with hourly streamflow averages. The dominant sources of uncertainty appear to be very different depending on the streamflow range: for streamflows lower than $30 \text{ m}^3 \text{ s}^{-1}$, the uncertainty is mainly due to stage systematic errors and to structural errors whereas for streamflows greater than $30 \text{ m}^3 \text{ s}^{-1}$, the parametric uncertainty is

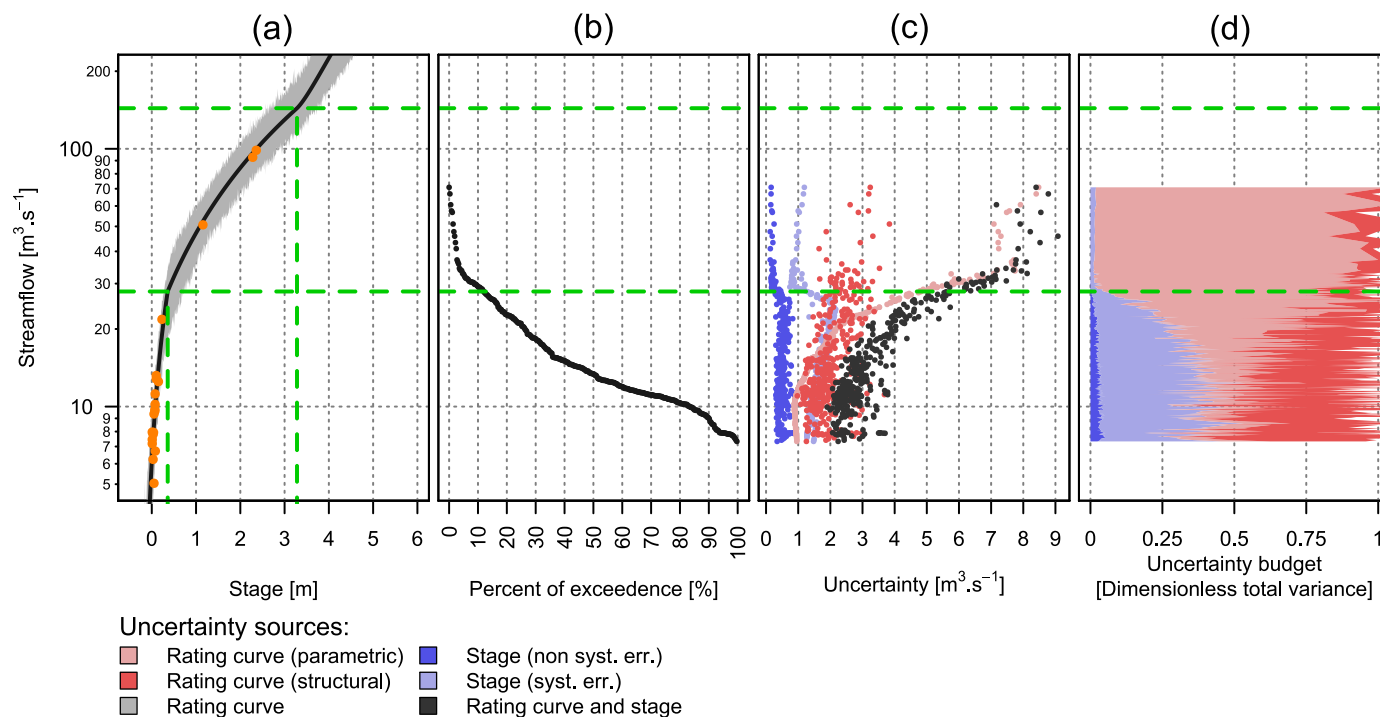


Figure 8. Uncertainty associated with the Bliesbruck station streamflow data: (a) *MaxPost* rating curve and associated total uncertainty with the gaugings in orange, (b) flow duration curve of daily streamflow data (derived from the *MaxPost* rating curve) for the year 2014, (c) corresponding sources of uncertainty (half-length of 95% uncertainty envelopes), and (d) corresponding uncertainty budget. The green dashed lines denote the transition between hydraulic controls.

dominant, due to the small number of gaugings available in this streamflow range. The importance of stage uncertainty at low flows is probably due to the poor sensitivity of the first control, i.e., stage being poorly sensitive to variations of discharge.

Figure 9 shows the streamflow uncertainty results through time for the year 2014 for several time averaging intervals. Time averaging has a clear impact on the total uncertainty and the uncertainty budget: the total uncertainty decreases with time averaging because nonsystematic errors are averaged out. This can be clearly seen in the uncertainty budget: the structural uncertainty which is modeled as nonsystematic, predominant at hourly intervals is almost negligible at weekly intervals. Besides, the uncertainty due to systematic stage errors is very important at monthly intervals but it is almost negligible in the annual mean. This is a consequence of resampling the systematic stage errors every 30 days, which induces their averaging out in longer time intervals.

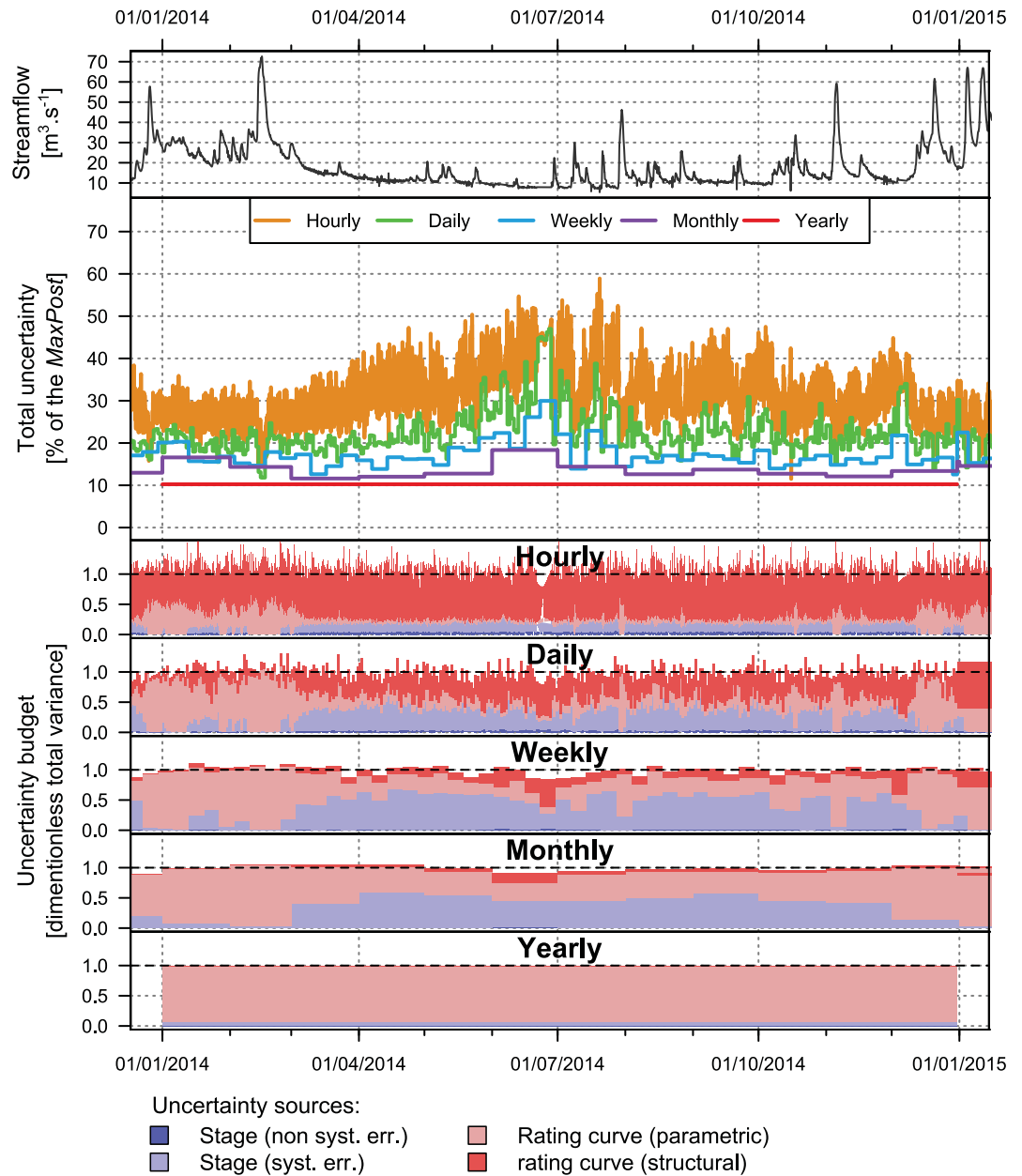


Figure 9. Streamflow uncertainty estimates through year 2014 at Bliesbruck. From top to bottom, the hourly *MaxPost* streamflow time series, the total uncertainty for various time averaging intervals, and the uncertainty budgets for the same time averaging intervals.

3.3.2. The Uncertainty Budget of a Range of Stations

Figure 10 summarizes for all 6 stations what was investigated in detail for the Bliesbruck station. For each station and each time step of a given year, the uncertainty budget $UB(t)$ is computed. The median uncertainty budget is then computed by computing the median of each component of $UB(t)$ across all time steps of the year. Similarly, the median total uncertainty U_T over the same year is computed. In Figure 10, the results are presented for different time averaging intervals. The ranges of the median uncertainty components over the studied range of stations are presented in Table 2. Systematic stage errors affect all the six studied stations. They are even predominant in some cases especially at daily, weekly and monthly intervals. On the other hand, nonsystematic stage errors are always small at instantaneous intervals and negligible for time-averaged streamflow time series.

Again, the averaging out of nonsystematic errors (both stage nonsystematic errors and rating curve structural errors) is clearly visible for all stations. In particular, the structural uncertainty, considered as nonsystematic (see section 2.2.1), is predominant at instantaneous intervals for all stations. It becomes almost negligible when considering data at monthly or yearly intervals. The total median uncertainty is also greatly reduced: from $\pm 32\%$ (instantaneous) to $\pm 4.6\%$ (yearly) at Taffignon or from $\pm 36\%$ to $\pm 10\%$ at Bliesbruck. Figure 10 also shows the importance of resampling the systematic stage errors which has an impact on

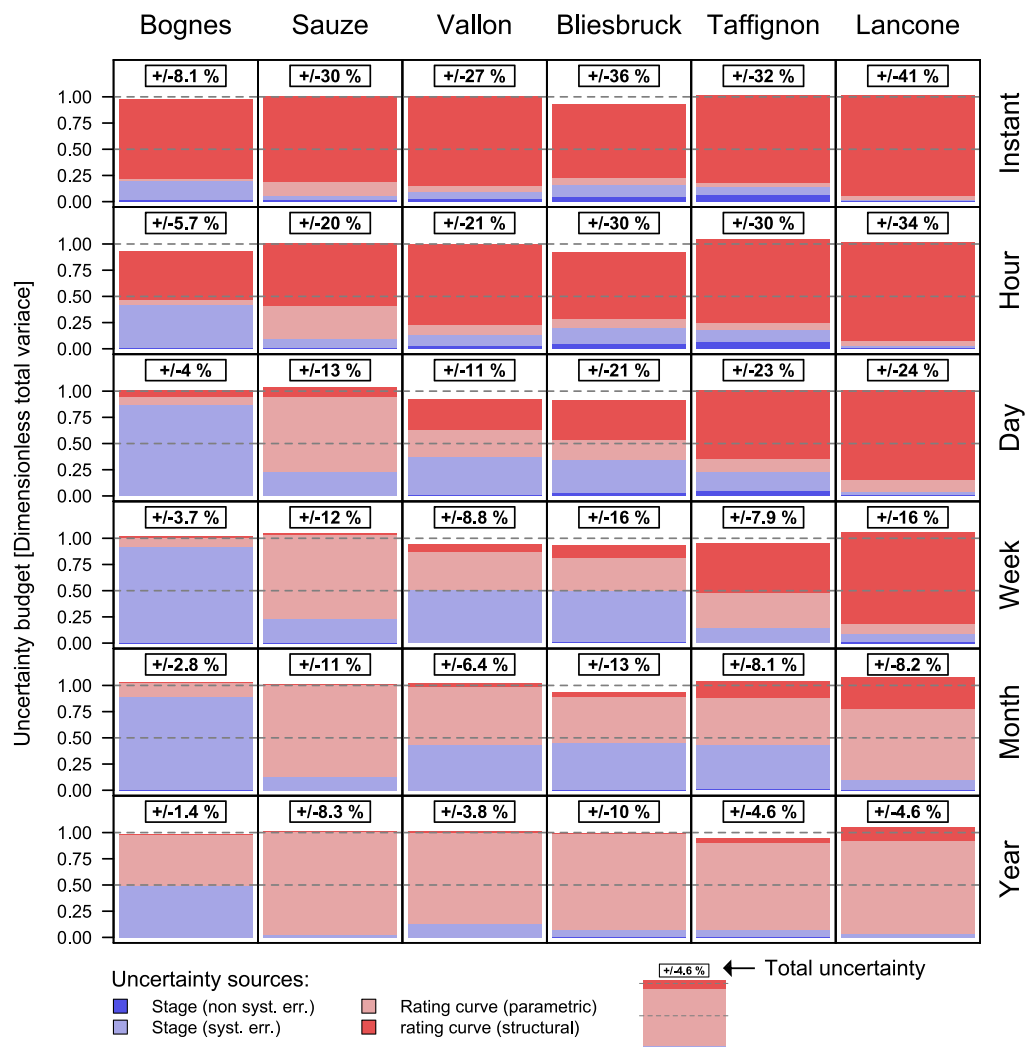


Figure 10. One year median of the uncertainty budget (dimensionless fraction of total variance) for six stations (columns) at different time averaging intervals (rows). The median total uncertainty in % of the $MaxPost$ is also shown.

Table 2
Ranges of Median Uncertainties (*U*, in Percentage of MaxPost Streamflow) Over the Studied Range of Stations

Time averaging intervals	<i>U</i> total	<i>U</i> stage (nonsyst. err.)	<i>U</i> stage (syst. err.)	<i>U</i> rating curve (parametric)	<i>U</i> rating curve (structural)
Instant	8.1–41	1.1–8.8	3.6–12	1.1–12	7.1–40
Hour	5.7–34	0.62–8.8	3.6–12	1.1–12	4.2–33
Day	4.0–24	0.15–5.6	3.7–12	1.0–12	1.0–23
Week	3.7–16	0.061–1.5	3.0–12	1.0–11	0.41–15
Month	2.8–13	0.032–1.2	2.3–9.0	1.0–10	0.22–4.3
Year	1.4–10	0.013–0.19	0.82–2.9	1.0–9.6	0.085–1.7

Note. Total uncertainty and the partial uncertainties are presented for various time intervals.

yearly streamflow averages. Indeed, an important decrease of stage systematic errors contributions is observed between monthly and yearly averaged streamflows for all the stations.

In addition to the influence of low flow hydraulic control sensitivity, the between-sites variations shown in Figure 10 are also probably linked to other sources of variability that impact the uncertainty estimation: for example, the hydrologic regime, the gauging strategies (gauging techniques, number and flow coverage) and the types and number of controls. The importance of hydraulic control sensitivity as illustrated in the Bliesbruck case can also be investigated in Figure 10. As an illustration, the same stage uncertainties were defined for Bliesbruck, Vallon, and Sauze (see Table 1). Of these three, Bliesbruck has the least sensitive low flow control and Sauze the most sensitive one. As can be seen in Figure 10, this correlates with the relative impacts of stage systematic errors on streamflow uncertainties. Note that Sauze and Vallon are similar stations that mainly differ in the sensitivity of their low flow hydraulic controls: they have the same hydrologic regime, two controls of same nature and similar number of gaugings.

Uncertainties affecting the stage time series are evaluated using expert knowledge of the different sites and we believe that the estimated standard deviations that we use are realistic. The results in Figure 10 are therefore likely to be representative of this source of uncertainty which can be predominant (e.g., Bognes), as large as the rating curve uncertainty (e.g., Taffignon, Bliesbruck, and Vallon) or minor (e.g., Sauze, Lancone). The generality of these results is further discussed in section 4.5.

3.4. Sensitivity Analysis of the Stage Time Series Error Model

The stage time series error model in equation (5) requires specifying several parameters: (1) the nonsystematic error standard deviation, (2) the systematic error standard deviation, (3) the average resampling periodicity of systematic errors, and (4) the choice between constant and linear versions of the model. As such specifications are mostly based on expert knowledge and may influence the resulting streamflow uncertainties due to stage

errors, a sensitivity analysis on these parameters is conducted. The sensitivity analysis is done on the Bliesbruck case for which the specifications used so far (see Table 1) are recalled in the first row of Table 3. The reference experiment is computed using these original specifications. The specification is changed for one component at a time to conduct individual experiments (see Table 3). The results of all these experiments are then compared with the reference experiment. The results are summed up for the year 2014 as the median total uncertainty and the median uncertainty budget in Figure 11.

The standard deviation controlling systematic errors has the strongest effect on streamflow uncertainties (experiment 2, Figure 11b). For all time averaging intervals except yearly, the total uncertainty increases with this standard deviation. The maximum increase is observed at weekly intervals as the uncertainty goes from ±17% to ±27% (from experiment 0 to experiment 2.2). The total uncertainties at yearly intervals are similar although the uncertainty budgets are not. This is probably because the contribution of the stage systematic errors remains small with respect to the total uncertainty. Uncertainties

Table 3
Setup of the Sensitivity Analysis

	σ_{NS}^h (m)	σ_S^h (m)	Periodicity (days)	Syst. error model
<i>Reference</i>				
Experiment 0	0.01	0.015	30	Constant
<i>Effect of nonsystematic errors standard deviation</i>				
Experiment 1.1	0.005	0.015	30	Constant
Experiment 1.2	0.02	0.015	30	Constant
<i>Effect of systematic error standard deviations</i>				
Experiment 2.1	0.01	0.0075	30	Constant
Experiment 2.2	0.01	0.03	30	Constant
<i>Effect of resampling periodicity</i>				
Experiment 3.1	0.01	0.015	7	Constant
Experiment 3.2	0.01	0.015	150	Constant
<i>Effect of systematic error behavior</i>				
Experiment 4	0.01	0.015	30	Linear

Note. Each row shows the specified parameters for each experiment.

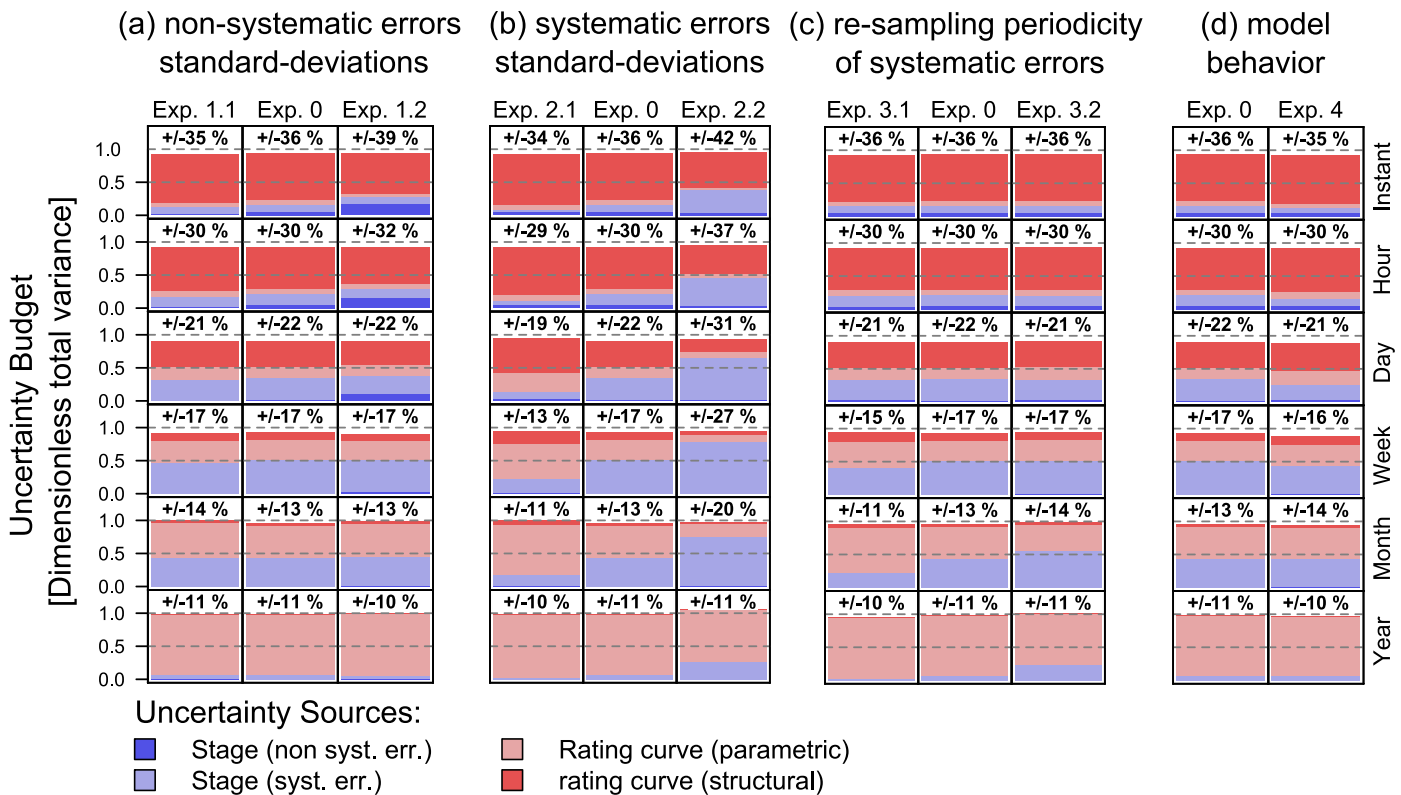


Figure 11. One year (2014) median of the uncertainty budget (dimensionless fraction of total variance) for the Bliesbruck station with various specifications for the stage error model (columns) and different time averaging intervals (rows). The median total uncertainty in % of the *MaxPost* is also shown. The column names refer to the experiments listed in Table 3.

arising from systematic stage errors are almost negligible for the reference experiment 0 whereas it brings a contribution of about 25% in experiment 2.2. The specification of the resampling period of systematic errors has virtually no effect on streamflow uncertainties for short time averaging intervals, up to daily intervals (see experiment 3, Figure 11c). However, the effect is more pronounced for longer intervals. In particular, the contributions of stage systematic errors decrease as soon as the time averaging interval exceeds the specified periodicity. This demonstrates again the importance of the systematic nature of stage errors for time-averaged data. The specification of nonsystematic errors standard deviations also impacts the resulting uncertainties of streamflow time series but to a much lesser extent (experiment 1, Figure 11a). The differences are mostly seen for short time intervals (up to daily). But in any case the contribution of this source of uncertainty remains small. The two tested systematic error models, constant and linear, show very little differences in the resulting streamflow uncertainties (experiment 4, Figure 11d). This might be explained by the similarity of the autocorrelation functions resulting from both error models (see Figure 2). Using a linear approach seems to slightly decrease the contribution of errors from the stage time series.

Overall, the sensitivity analysis highlights the importance of a reliable specification of the standard deviation of systematic errors. Moreover, depending on the time averaging scale of interest, additional efforts should be put either on the specification of the standard deviation of nonsystematic errors for short averaging intervals or on the specification of the systematic errors resampling periodicity for longer time averaging intervals. Finally, the results show little effect of the chosen systematic error model on the streamflow time series uncertainties.

4. Discussion

4.1. Does Uncertainty in Gaugings Stages Impact Rating Curve Estimation?

The impact of discharge errors in gaugings on rating curve uncertainties has frequently been investigated in the literature (cf., review by Le Coz et al., 2014) whereas the impact of stage errors in gaugings has largely

been neglected. Coxon et al. (2015) and McMillan and Westerberg (2015) did assume a gauging stage error of ± 5 mm with a uniform distribution, i.e., much smaller uncertainties than in this study. Our results suggest that their impact on the rating curve uncertainty is limited, at least on the parametric uncertainty and for stage uncertainties corresponding to fair gauging conditions. Our results also suggest that this impact may differ across hydrometric stations, those characterized by low stage values due to channel configuration (wide, flat, shallow rivers) being more affected by stage errors in gaugings. Further investigations could be undertaken to identify more precisely the main factors governing the impact of stage errors in gaugings on rating curves, including a broader set of hydrometric stations or even using synthetic data. The direct impacts of gaugings stage errors on the rating curve are likely only determined by local characteristics of the hydrometric stations. However, their indirect impacts on averaged streamflow time series are also determined by the hydrological regime (e.g., flow duration curve).

In this paper, we used a simple Monte Carlo approach to propagate the uncertainty of gaugings stages into the estimated rating curve. We note that an alternative approach is possible, based on “error-in-variables” models (in a hydrological context, see e.g., the BATEA approach of Kavetski et al. (2006)). Such models would consider the unknown true stages as latent random variables, and would attempt to estimate them as part of the inference, constraining these estimated true values to remain consistent with the observed values and their uncertainty. Since this approach might be prone to various issues that lie well beyond the scope of this paper (nonidentifiability, interaction with other sources of errors, see e.g., Renard et al., 2010; Thyer et al., 2009), we favored the simple Monte Carlo propagation here. However, further investigations are required to compare the two approaches and evaluate whether or not they lead to similar conclusions.

4.2. Does Uncertainty in the Stage Time Series Play an Important Role in Streamflow Uncertainty?

The contribution of errors in the stage time series to the uncertainty in the streamflow time series is considerable in many cases. Our results show that systematic stage errors may be one of the main contributors to streamflow uncertainty, especially for temporally averaged streamflow (cf., Table 2). Over the six hydrometric stations in this study, the uncertainty component reflecting stage systematic errors ranged from 4% to 12% of daily average streamflow, and from 1% to 3% of yearly average streamflow as sensors were assumed to be recalibrated every 30 days. On the other hand, nonsystematic stage errors have a negligible contribution in most cases. The relative contribution of stage uncertainty strongly depends on the importance of rating curve uncertainty: high uncertainties in the rating curve will result in a smaller relative contribution of stage errors. This contribution also depends on the management of the hydrometric station. In particular, results from the sensitivity analysis indicate that the periodicity with which the sensor is recalibrated and the amplitude of these recalibrations are very important factors. The uncertainty analysis can therefore help improving the management of the hydrometric station. The hydraulic sensitivity, i.e., how much stage varies with a given change in streamflow, also plays a key role as it magnifies or mitigates stage errors. Further investigations, not shown here, conducted using synthetic data indicated that, depending on the sensitivity of hydraulic controls, the effect of stage uncertainty may represent between 8% and 20% of the total uncertainty in monthly averaged streamflow data (Horner et al., 2017).

4.3. On the Assumption That Measurement Uncertainties Are Known

In this paper, all measurement uncertainties are assumed to be known before the estimation of the rating curve and its use to compute discharge time series. This corresponds to the use of known values for the standard deviations (σ_j^h, σ_j^Q) for gaugings and ($\sigma_{NS}^h, \sigma_S^h$) for the stage time series. The rationale behind this assumption is that measurement uncertainties exist independently of the rating curve. It is therefore possible to estimate them before estimating and using the rating curve. We stress that many operational methods, even standards, exist for this purpose. For instance, the method described in the ISO 748 and ISO 1088 standards, and its variants proposed by Le Coz et al. (2012), Muste et al. (2012), Cohn et al. (2013), and Despax et al. (2016) allow computing the uncertainties of gaugings performed with the velocity-area method. Standards are also available for other gauging methods (e.g., ISO 9555 on tracer dilution, ISO 24578 on mobile-boat ADCP, ISO 1070 on slope-area discharge estimates) and they all include a clause on their uncertainty analysis. Several methods have been published recently for computing the uncertainty of mobile-boat ADCP gaugings (González-Castro et al., 2016; Moore et al., 2017; Mueller, 2016). Le Coz et al. (2016) also showed that the uncertainty of stream gauging techniques can be quantified using in-situ repeated measures experiments.

Unlike measurement uncertainties, structural uncertainty is unknown and is estimated along with the rating curve, through the unknown parameters γ_1 and γ_2 governing the standard deviation of structural errors $\sigma_j^f = \gamma_1 + \gamma_2 \hat{Q}_j$. Moreover, we generally assume no prior knowledge on structural uncertainty, through the use of wide uniform priors for γ_1 and γ_2 . This opposite treatment of measurement and structural uncertainties is a direct consequence of the fundamentally different natures of measurement and structural errors: while the former exists independently of the rating curve, the latter is defined as the difference between the rating curve discharge and the (unknown) true discharge. It is therefore difficult, if not impossible, to estimate its properties before having estimated the rating curve.

The fact that measurement uncertainties are known and structural uncertainty is inferred implies that the latter may compensate for misspecifications of the former. Essentially, structural uncertainty is estimated so as to “close the uncertainty balance”: when measurement uncertainty is underestimated, structural uncertainty tends to be overestimated (and vice-versa), thus leading to a misrepresentation of the origin of uncertainty. In turn, this may lead to severe underestimations or overestimations in the uncertainty affecting discharge time series, since different sources of uncertainty have completely different behaviors with respect to temporal averaging, as largely illustrated in this paper. Consequently, it is important that measurement uncertainties be specified as realistically as possible, using whenever possible formal methods such as those discussed above rather than rule-of-thumb values.

In cases for which the specification of realistic measurement uncertainties is deemed impossible, it is tempting to add the standard deviations (σ_j^h, σ_j^o) to the list of unknown parameters that need to be inferred, possibly constrained by some informative prior distributions. This corresponds to an “uncertainty about the uncertainty” situation, coined *uncertainty*² by Juston et al. (2014). In our opinion, this strategy is fraught with difficulties for the following reasons:

1. In the absence of very informative priors on the standard deviations (σ_j^h, σ_j^o) , the problem is likely to be ill-posed: indeed, the various unknown standard deviations interact with each other and are hence non-identifiable, as demonstrated by Renard et al. (2010); Renard et al. (2011) in the context of hydrologic model calibration;
2. Using very informative priors may lead to little difference with simply assuming that the standard deviations are known, but leads to a more challenging inference problem due to the introduction of additional unknowns.

We claim that the first priority should be to encourage field hydrologists to quantify and communicate measurement uncertainties, based on proven methods and standards. In this respect, deriving methods that can explicitly make use of these measurement uncertainties in subsequent analyses is a strong incentive, as it demonstrates that measurement uncertainties are useful—if not necessary.

4.4. Improving the Error Model for Stage Time Series

The proposed error model for stage time series might be considered as too simplistic given the complex nature of errors affecting stage measurements. We acknowledge this limitation, especially for the systematic part of the model related to the stage sensor drift. In our opinion, the prerequisite to improve this part of the model is to improve our understanding of the physical processes governing the sensor drift. The simple models used here—constant or linear systematic error between successive recalibrations—simply correspond to our limited understanding of how the drift behaves between recalibrations. While the sensitivity analysis suggests that these two models yield very similar results, alternative models may have more impact. Among the possible sources of systematic errors listed by Sauer and Turnipseed (2010), the effect of water temperature variation on pressure transducers or the representativeness of the measured stage with respect to the hydraulically relevant spatially averaged flow depth could be considered. Other sources such as small animals resting on a float or electrical signal conversion errors seem less amenable to statistical modeling. Finally, the model for nonsystematic stage errors could also benefit from improvements, e.g., by introducing heteroscedasticity to account for larger waves during high flows.

A practical strategy to improve our understanding of sensor drift would be to perform frequent (e.g., daily) readings of the reference gauge for a few selected stations in order to capture the drift’s dynamics. It would then be possible to assess whether the drift is a smooth process (possibly depending on covariates such as temperature) or a more jumpy process, where sudden drifts appear either randomly or following specific

events such as floods. Specific probabilistic models could then be constructed to describe this behavior, while retaining a physical interpretation. We stress that in our opinion, increasing the sophistication of the probabilistic model is only valuable if it corresponds to an improved understanding of the drift's dynamics. In the absence of any physical interpretation, specifying realistic values for the parameters of this model is problematic, if not impossible.

4.5. Generality of the Results

The results presented in this paper have been obtained with the BaRatin method (Le Coz et al., 2014) for estimating the rating curve and its uncertainty. However, the stage error models and the Monte Carlo approach we proposed are derived independently of the BaRatin method, so that many other rating curve uncertainty estimation methods could be used in principle. In particular, comparing the results of various methods would be beneficial to assess the generality of the results presented herein. Indeed, at least part of the observed impact (or lack thereof) of stage errors on streamflow uncertainty might be specific to the method used to estimate the rating curve.

One questionable assumption made by the BaRatin method is that rating curve structural errors are independent from one time step to the next. This ignores their partly systematic nature. We refer the reader to the papers by Le Coz et al. (2014) and Sikorska and Renard (2017) for thorough discussions on the reasons behind this assumption and the difficulty to propose a more general structural error model. Due to this assumption, the contribution of structural uncertainty tends to decrease with the temporal averaging of streamflow (e.g., daily to monthly values, see Figure 10). A rating curve estimation method based on different premises, such as the voting-point method of McMillan and Westerberg (2015), might lead to quite different uncertainty budgets, and therefore to a different estimation of the relative contribution of stage uncertainty.

4.6. Verification of Uncertainty Estimates

As emphasized by, e.g., Hall et al. (2007), uncertainty estimates should be verified (or "validated"), because they rely on the many assumptions embedded in the error models. This is typically achieved through some form of split-sampling: some data are left out of the calibration sample, and the consistency of these data with the estimated uncertainty is evaluated through specialized diagnostics (e.g., PIT diagrams, see e.g., Laio & Tamea, 2007; Thyer et al., 2009).

In the particular context of rating curve estimation, gaugings are the only direct observations of streamflow. Split-sampling verification can be applied using the gaugings, and in our experience results suggest an acceptable reliability. However, this is rather weak verification, because gaugings correspond to sporadic measurements of instantaneous discharge. It is therefore only informative of the reliability of the total uncertainty affecting instantaneous discharge estimations. A more stringent verification would also involve verifying the reliability of the individual uncertainty components (stage systematic and nonsystematic errors, rating curve parametric and structural errors), along with the reliability of the uncertainty affecting temporally averaged streamflow. Unfortunately, in natural rivers, daily or monthly streamflows are never measured directly, so that there is in general no reference against which the temporally averaged streamflow derived from the rating curve can be compared. In some rare cases, it could be possible to compare two independent estimates of the temporally averaged streamflow series: for instance, streamflow is continuously measured using ultrasonic transit time systems in the penstocks of some dams, which could be compared with the streamflow estimated at a classical hydrometric station located just downstream. However, comparing two uncertain series, both being affected by a mix of systematic and nonsystematic errors, is a methodological challenge that would require further developing available verification tools such as the PIT diagram.

5. Conclusion

The objective of this paper was to propose a method for quantifying the contribution of various sources of stage errors to the total streamflow uncertainty. We introduced an original error model for propagating stage time series errors through an uncertain stage-discharge rating curve following an ensemble approach. The error model of stage time series considers two uncertainty components: systematic and nonsystematic

errors. The rating curve uncertainty estimation also takes into account the stage uncertainty of gaugings. The method was applied to a range of six contrasting sites.

Results from the six studied stations showed that considering stage uncertainty in gaugings has in general a limited effect on rating curve estimates and on the parametric uncertainty. Increases in the total rating curve uncertainty were most of the time less than +10% but could be more substantial for hydrometric stations characterized by low flow depths (small catchments, catchments in arid/semiarid environments). Systematic errors in stage time series proved to be the most important type of stage errors to consider. Using realistic uncertainty specifications, we found that the systematic stage errors were predominant for weekly and monthly streamflow averages at most stations. By contrast, the nonsystematic stage errors contributions were almost always negligible.

Our results emphasize the importance of considering the systematic nature of errors for averaged streamflow data, i.e., the stage uncertainty due to systematic errors and the parametric uncertainty of the rating curve. Nonsystematic errors—nonsystematic stage errors and the rating curve structural errors (assumed here to be nonsystematic)—are quickly averaged out with time averaging. For annual averages, the parametric uncertainty was predominant in most cases because systematic stage errors are resampled several times per year (corresponding to recalibration of the stage sensor), hence are also averaged out.

The need for a better understanding of stage error sources and dependencies, and better methods for validation of stage error estimates, was highlighted in this paper. This should be the subject of future studies. Also, the importance of hydraulic sensitivity should be investigated as it may play a key role in the contribution of stage time series errors. Future studies could also focus on the propagation of stage time series errors in the calculation of various hydrological signatures (e.g., Westerberg & McMillan, 2015) and hydrological statistics (e.g., Steinbakk et al., 2016), and more generally, in any use of streamflow data such as the estimation and validation of hydrological models.

Considering errors in stage time series is clearly important in most cases but their relative contribution to the uncertainty budget is site specific: it depends on the channel characteristics, the information available to build the rating curve, the station management, the time interval at which streamflow data are averaged and the streamflow range. Because the impacts of stage errors are site specific, a quantitative uncertainty analysis of each site individually is necessary. Such uncertainty analysis help the data producer making decisions for computing the rating curve and the streamflow time series. Moreover, all the measurement processes may be improved once the main sources of errors have been identified (e.g., adapt the stage sensors recalibration frequency). Finally, the data user can make his/her own decisions in a transparent way.

Appendix A: Hydraulic Analysis of the Hydrometric Stations

A1. General Information

A detailed description of BaRatin is available in the BaRatinAGE software documentation (available for download at https://forge.irstea.fr/projects/baratinage_v2/). The hydraulic analysis of a hydrometric station leads to the identification of controls that are active for different stage ranges. Any type of control is approximated by a power function as

$$Q(h) = a(h - b)^c \tag{A1}$$

where Q and h stand for streamflow and stage, respectively. Parameters a , b , and c are referred to as the coefficient, offset and exponent of the control, respectively. Implicitly, when $h < b$, $Q(h) = 0$.

The two types of controls that are at play for the sites studied in this paper are wide rectangular channel controls and rectangular section controls. A channel control is approximated using a Manning-Strickler equation for uniform, wide rectangular channel:

$$Q(h) = KB\sqrt{S}(h - b)^{5/3} \tag{A2}$$

where B , K , S , and b stand for the channel width, Strickler coefficient, slope, and mean bed elevation, respectively.

A rectangular section control is modeled as a rectangular weir using the following equation:

$$Q(h) = C\sqrt{2g}B(h - b)^{3/2} \tag{A3}$$

where g is the gravity acceleration and C , B , and b stand for the discharge coefficient, weir crest width, and elevation, respectively.

For the sites studied, the geometries of the controls were estimated mostly from aerial and ground photographs. Topographic measurements were also used when available. The application of BaRatin to the sites studied in this paper is described in Horner et al. (2015a, 2015b). Note that no priors are specified for offsets b in the equations above because they are calculated by continuity of the stage-discharge relation. When an uncertainty is specified as $\pm x$ it means that the corresponding standard deviation is $x/2$ (if the distribution is Gaussian).

A2. Bognes

At the lowest stages, the stage-discharge relation is controlled by a critical cross-section downstream of the stage recorder. The very irregular control section can be approximated by a rectangular weir (width: 45 m \pm 25; discharge coefficient: 0.45 \pm 0.05; crest level: 259 m \pm 0.5).

Above an uncertain stage threshold, a channel control takes over (activation stage: 260.75 m \pm 1.25; width: 65 m \pm 15; slope: 0.003 \pm 0.002; Strickler coefficient: 30 m^{1/3} s⁻¹ \pm 5).

Therefore, the stage-discharge rating curve is defined as a two-segment equation:

$$Q(h) = \begin{cases} a_1(h-b_1)^{c_1} & \text{if } h \leq k_2 \\ a_2(h-b_2)^{c_1} & \text{if } h \geq k_2 \end{cases} \tag{A4}$$

The corresponding prior distributions are Gaussian $\mathcal{N}(\mu, \sigma)$ with mean μ and standard deviation σ :

	k_i	a_i	c_i
Control $i = 1$	$\mathcal{N}(259, 0.25)$	$\mathcal{N}(90, 25)$	$\mathcal{N}(1.5, 0.025)$
Control $i = 2$	$\mathcal{N}(260.75, 0.625)$	$\mathcal{N}(107, 23.5)$	$\mathcal{N}(1.67, 0.025)$

A3. Sauze

Two controls are at play at this hydrometric station. A low flow critical cross section is approximated as a narrow rectangular weir (width: 40 m \pm 20; discharge coefficient: 0.4 \pm 0.05; crest level: 44.5 m \pm 1). At higher flow (activation stage: 48 m \pm 2), the stage-discharge relation is controlled by the river channel (width: 92 m \pm 30; slope: 0.0013 \pm 0.0008; Strickler coefficient: 30 m^{1/3} s⁻¹ \pm 5).

Therefore, the stage-discharge rating curve is defined as a two-segment equation (see equation (A4)). The corresponding prior distributions are as follows:

	k_i	a_i	c_i
Control $i = 1$	$\mathcal{N}(44.5, 0.5)$	$\mathcal{N}(71, 18.5)$	$\mathcal{N}(1.5, 0.025)$
Control $i = 2$	$\mathcal{N}(48, 1)$	$\mathcal{N}(100, 24)$	$\mathcal{N}(1.67, 0.025)$

A4. Vallon

The hydraulic analysis suggests two controls: a section control (width: 50 m \pm 20; discharge coefficient: 0.45 \pm 0.1; crest level: -1 m \pm 0.2) for low flows replaced (activation stage: 0.25 m \pm 0.75) by a channel control (width: 60 m \pm 15; slope: 0.0003 \pm 0.0002; Strickler coefficient: 30 m^{1/3} s⁻¹ \pm 10) for higher flows.

Therefore, the stage-discharge rating curve is defined as a two-segment equation (see equation (A4)). The corresponding prior distributions are as follows:

	k_i	a_i	c_i
Control $i = 1$	$\mathcal{N}(-1, 0.1)$	$\mathcal{N}(100, 23)$	$\mathcal{N}(1.5, 0.025)$
Control $i = 2$	$\mathcal{N}(0.25, 0.375)$	$\mathcal{N}(31.2, 8.35)$	$\mathcal{N}(1.67, 0.025)$

A5. Bliesbruck

At low flow, the stage-discharge relation is controlled by the river bed directly downstream of the stage sensor (width: 40 m ± 3; discharge coefficient: 0.4 ± 0.1; crest level: -0.5 m ± 0.4). When stage exceeds an uncertain threshold (activation stage: 0.5 m ± 0.5), a river channel which is much more narrow downstream of the station takes over. A hydraulic model was used to estimate a_2 parameter (model result when considering a range of realistic Strickler coefficient values). For the highest discharge (activation stage: 3.25 m ± 0.5), a control by the floodplain is added. It is modeled as a river channel (a_3 was estimated using a hydraulic model).

Therefore, the stage-discharge rating curve is defined as a three-segment equation:

$$Q(h) = \begin{cases} a_1(h-b_1)^{c_1} & \text{if } h \leq k_2 \\ a_2(h-b_2)^{c_2} & \text{if } k_2 \leq h \leq k_3 \\ a_2(h-b_2)^{c_2} + a_3(h-b_3)^{c_3} & \text{if } h \geq k_3 \end{cases} \quad (A5)$$

The corresponding prior distributions are as follows:

	k_i	a_i	c_i
Control $i = 1$	$\mathcal{N}(-0.5, 0.1)$	$\mathcal{N}(71, 9.5)$	$\mathcal{N}(1.5, 0.025)$
Control $i = 2$	$\mathcal{N}(0.5, 0.25)$	$\mathcal{N}(13, 2.5)$	$\mathcal{N}(1.67, 0.025)$
Control $i = 3$	$\mathcal{N}(3.25, 0.25)$	$\mathcal{N}(70, 15)$	$\mathcal{N}(1.67, 0.025)$

A6. Taffignon

A wide concrete rectangular weir (width: 8.25 m ± 0.5; discharge coefficient: 0.4 ± 0.05; crest level: 0.075 m ± 0.025) controls the stage-discharge relation for almost all the stage values. Above the activation stage (1.3 m ± 0.2), the water overflows the left bank. This second control is modeled as a channel control (width: 6 m ± 2; slope: 0.005 ± 0.005; Strickler coefficient: 25 m^{1/3} s⁻¹ ± 7) and is added to the control by the concrete weir that is still active in the main channel.

Therefore, the stage-discharge rating curve is defined as a two-segment equation:

$$Q(h) = \begin{cases} a_1(h-b_1)^{c_1} & \text{if } h \leq k_2 \\ a_1(h-b_1)^{c_1} + a_2(h-b_2)^{c_2} & \text{if } h \geq k_2 \end{cases} \quad (A6)$$

The corresponding prior distributions are as follows:

	k_i	a_i	c_i
Control $i = 1$	$\mathcal{N}(0.075, 0.0125)$	$\mathcal{N}(14.6, 1)$	$\mathcal{N}(1.5, 0.025)$
Control $i = 2$	$\mathcal{N}(1.3, 0.1)$	$\mathcal{N}(10.6, 3.5)$	$\mathcal{N}(1.67, 0.025)$

A7. Lancone

The stage sensor is located directly upstream of a complex weir made of four different weirs of various widths and elevations. Each weir is activated when stage exceeds the crest level and adds to the already activated weirs:

1. width: 1.9 m ± 0.5; crest level: 0.2 m ± 0.1,
2. width: 3 m ± 0.3; crest level: 0.6 m ± 0.1,
3. width: 11 m ± 3; crest level: 1.1 m ± 0.1, and
4. width: 11 m ± 3; crest level: 1.3 m ± 0.1

For all these controls, a discharge coefficient of 0.4 ± 0.05 is chosen.

A last control corresponding to the overflow over rocks located on the left side is also added. It is modeled as a weir (width: 4 m ± 0.5; crest level: 1.6 m ± 0.25; discharge coefficient: 0.4 ± 0.1).

Therefore, the stage-discharge rating curve is defined as a five-segment equation:

$$Q(h) = \begin{cases} a_1(h-b_1)^{c_1} & \text{if } h \leq k_2 \\ a_2(h-b_2)^{c_2} + a_1(h-b_1)^{c_1} & \text{if } k_2 \leq h \leq k_3 \\ a_3(h-b_3)^{c_3} + a_2(h-b_2)^{c_2} + a_1(h-b_1)^{c_1} & \text{if } k_3 \leq h \leq k_4 \\ a_4(h-b_4)^{c_4} + a_3(h-b_3)^{c_3} + a_2(h-b_2)^{c_2} + a_1(h-b_1)^{c_1} & \text{if } k_4 \leq h \leq k_5 \\ a_5(h-b_5)^{c_5} + a_4(h-b_4)^{c_4} + a_3(h-b_3)^{c_3} + a_2(h-b_2)^{c_2} + a_1(h-b_1)^{c_1} & \text{if } h \geq k_5 \end{cases} \quad (A7)$$

The corresponding prior distributions are as follows:

	k_i	a_i	c_i
Control $i = 1$	$\mathcal{N}(0.2, 0.05)$	$\mathcal{N}(3.4, 0.5)$	$\mathcal{N}(1.5, 0.025)$
Control $i = 2$	$\mathcal{N}(0.6, 0.05)$	$\mathcal{N}(5.3, 0.45)$	$\mathcal{N}(1.5, 0.025)$
Control $i = 3$	$\mathcal{N}(1.1, 0.05)$	$\mathcal{N}(19.5, 2.9)$	$\mathcal{N}(1.5, 0.025)$
Control $i = 4$	$\mathcal{N}(1.3, 0.05)$	$\mathcal{N}(19.5, 2.9)$	$\mathcal{N}(1.5, 0.025)$
Control $i = 5$	$\mathcal{N}(1.6, 0.125)$	$\mathcal{N}(7.1, 1.25)$	$\mathcal{N}(1.5, 0.025)$

Acknowledgments

We are grateful and we thank Mark Thyer, three anonymous reviewers, and the Associate Editor for the quality, relevance and helpfulness of their remarks. The development of the method was funded by Irstea, the Compagnie nationale du Rhône (CNR), and the French National Hydrologic Services (SCHAPI). BaRatin uses the DMSL Fortran library developed by Dmitri Kavetski (University of Adelaide, Australia). The FloodScale project was funded by the French National Research Agency (ANR) under contract ANR 2011 BS56 027 and contributed to the HyMeX program. Collaboration between NIWA and Irstea was enhanced by the PHC Dumont D’Urville project 34185SH (2015–2016) and the Irstea grant and NIWA support for the scientific visits of J. Le Coz and F. Branger at NIWA. All the hydrologic services who have provided data and information are gratefully thanked: CNR, SPC Grand Delta, SPC Rhin-Sarre, DREAL Corse, DREAL Auvergne-Rhône-Alpes. Gaugings and stage data for all the stations studied in this paper (Bognes, Sauze, Vallon, Bliesbruck, Taffignon, and Lancone) are provided as supporting information.

References

Birkinshaw, S. J., Moore, P., Kilsby, C. G., O’Donnell, G. M., Hardy, A. J., & Berry, P. A. M. (2014). Daily discharge estimation at ungauged river sites using remote sensing. *Hydrological Processes*, 28(3), 1043–1054.

Cohn, T., Kiang, J., & Mason, R. (2013). Estimating discharge measurement uncertainty using the interpolated variance estimator. *Journal of Hydraulic Engineering*, 139(5), 502–510.

Coxon, G., Freer, J., Westerberg, I. K., Wagener, T., Woods, R., & Smith, P. J. (2015). A novel framework for discharge uncertainty quantification applied to 500 UK gauging stations. *Water Resources Research*, 51, 5531–5546. <https://doi.org/10.1002/2014WR016532>

da Silva, J. S., Calmant, S., Seyler, F., Rotunno Filho, O. C., Cochonneau, G., & Mansur, W. J. (2010). Water levels in the Amazon basin derived from the ERS 2 and ENVISAT radar altimetry missions. *Remote Sensing of Environment*, 114, 2160–2181.

Despax, A., Perret, C., Garçon, R., Hauet, A., Belleville, A., Le Coz, J., et al. (2016). Considering sampling strategy and cross-section complexity for estimating the uncertainty of discharge measurements using the velocity-area method. *Journal of Hydrology*, 533, 128–140.

Dymond, J. R., & Christian, R. (1982). Accuracy of discharge determined from a rating curve. *Hydrological Sciences Journal/Journal des Sciences Hydrologiques*, 27(4), 493–504.

Freestone, H. J. (1983). Sensitivity of flow measurement to stage errors for New Zealand catchments. *Journal of Hydrology New Zealand*, 22(2), 175–181.

Getirana, A. C. V., & Peters-Lidard, C. (2013). Estimating water discharge from large radar altimetry datasets. *Hydrology and Earth System Sciences*, 17(3), 923–933.

González-Castro, J., Buzard, J., & Mohamed, A. (2016). *RiverFlowUA—A package to estimate total uncertainty in ADCP discharge measurements by FOTSE—with an application in hydrometry*. Paper presented at the IAHR River Flow 2016 Conference (pp. 715–723).

Hall, J., O’Connell, E., & Ewen, J. (2007). On not undermining the science: Coherence, validation and expertise. Discussion of Invited Commentary by Keith Beven *Hydrological Processes*, 20, 3141–3146 (2006). *Hydrological Processes*, 21(7), 985–988.

Hamilton, A. S., & Moore, R. D. (2012). Quantifying uncertainty in streamflow records. *Canadian Water Resources Journal*, 37(1), 3–21. <https://doi.org/10.4296/cwrj3701865>

Hersch, R. W. (1999). *Hydrometry: Principles and practices*. Chichester, UK: John Wiley.

Horner, I., Le Coz, J., & Renard, B. (2015a). *Estimation des incertitudes sur les hydrogrammes avec BaRatin* [Estimation of streamflow time series uncertainties using BaRatin, in French] (Technical report). Irstea.

Horner, I., Le Coz, J., Renard, B., & Branger, F. (2017). *Impact de la sensibilité des contrôles hydrauliques sur les incertitudes hydrométriques* [The impact of control sensitivity on hydrometric uncertainties, in French]. Paper presented at the SHF Hydrométrie 2017 Conference, Lyon, France, March 14–15.

Horner, I., Le Coz, J., Renard, B., Mansanarez, V., Pierrefeu, G., Le Boursicaud, R., et al. (2015b). *BaRatin: Propagation des incertitudes aux chroniques de débit* [BaRatin: Propagation of uncertainties to streamflow time series, in French] (Technical report). Irstea.

ISO 1100-2:2010. (2010). *Hydrometry—Measurement of liquid flow in open channels—Part 2: Determination of the stage-discharge relation* (28 pp.). Geneva, Switzerland: International Organization for Standardization.

ISO/TS25377:2007 (2009). *Hydrometric uncertainty guidance (HUG)*.

Juston, J., Jansson, P.-E., & Gustafsson, D. (2014). Rating curve uncertainty and change detection in discharge time series: Case study with 44-year historic data from the Nyangores River, Kenya. *Hydrological Processes*, 28, 2509–2523. <https://doi.org/10.1002/hyp.9786>

Kavetski, D., Kuczera, G., & Franks, S. W. (2006). Bayesian analysis of input uncertainty in hydrological modeling: 1. Theory. *Water Resources Research*, 42, W03407. <https://doi.org/10.1029/2005WR004368>

Laio, F., & Tamea, S. (2007). Verification tools for probabilistic forecasts of continuous hydrological variables. *Hydrology and Earth System Sciences*, 11(4), 1267–1277.

Le Coz, J., Blanquart, B., Pobanz, K., Dramais, G., Pierrefeu, G., Hauet, A., et al. (2016). Estimating the uncertainty of streamgauging techniques using in situ collaborative interlaboratory experiments. *Journal of Hydraulic Engineering*, 142(7), 04016011. [https://doi.org/10.1061/\(ASCE\)HY.1943-7900.0001109](https://doi.org/10.1061/(ASCE)HY.1943-7900.0001109)

Le Coz, J., Camenen, B., Peyrard, X., & Dramais, G. (2012). Uncertainty in open-channel discharges measured with the velocity-area method. *Flow Measurement and Instrumentation*, 26, 18–29.

Le Coz, J., Renard, B., Bonnifait, L., Branger, F., & Le Boursicaud, R. (2014). Combining hydraulic knowledge and uncertainty gaugings in the estimation of hydrometric rating curves: A Bayesian approach. *Journal of Hydrology*, 509, 573–587.

- McMillan, H., Krueger, T., & Freer, J. (2012). Benchmarking observational uncertainties for hydrology: Rainfall, river discharge and water quality. *Hydrological Processes*, 26(26), 4078–4111.
- McMillan, H., Seibert, J., Petersen-Øverleir, A., Lang, M., White, P., Snelder, T., et al. (2017). How uncertainty analysis of streamflow data can reduce costs and promote robust decisions in water management applications. *Water Resources Research*, 53, 5220–5228. <https://doi.org/10.1002/2016WR020328>
- McMillan, H. K., & Westerberg, I. K. (2015). Rating curve estimation under epistemic uncertainty. *Hydrological Processes*, 29(26), 1873–1188. <https://doi.org/10.1002/hyp.10419>
- Moore, S. A., Jamieson, E. C., Rainville, F., Rennie, C. D., & Mueller, D. S. (2017). Monte Carlo approach for uncertainty analysis of acoustic Doppler current profiler discharge measurement by moving boat. *Journal of Hydraulic Engineering*, 143(3), 04016088.
- Morlot, T., Perret, C., Favre, A.-C., & Jalbert, J. (2014). Dynamic rating curve assessment for hydrometric stations and computation of the associated uncertainties: Quality and station management indicators. *Journal of Hydrology*, 517(26), 173–186.
- Moyeed, R. A., & Clarke, R. T. (2005). The use of Bayesian methods for fitting rating curves, with case studies. *Advances in Water Resources*, 28, 807–818.
- Mueller, D. S. (2016). *QRev—Software for computation and quality assurance of acoustic Doppler current profiler moving-boat streamflow measurements—User’s manual*. Reston, VA: U.S. Geological Survey.
- Muste, M., Lee, K., & Bertrand-Krajewski, J.-L. (2012). Standardized uncertainty analysis for hydrometry: A review of relevant approaches and implementation examples. *Hydrological Sciences Journal*, 57(4), 643–667.
- Olivier, A., Pierrefeu, G., Scotti, M., & Blanquart, B. (2008). *Incertitudes sur les débits issus des courbes de tarage [uncertainty of discharge measured with relating curve between level and discharge, in French]*. Paper presented at the SHF 2008 conference, Hydrological measurements and uncertainties. Paris, France, April 1–2.
- Petersen-Øverleir, A., & Reitan, T. (2005). Uncertainty in flood discharges from urban and small rural catchments due to inaccurate head measurements. *Nordic Hydrology*, 36(3), 245–257.
- Puechberty, R., Perret, C., & Poligot-Pitsch, S. (coord.) (2017). *Charte qualité de l’hydrométrie. Guide de bonnes pratiques [Quality plan for hydrometry. Good Practice Guide, in French]* (83 pp.). France: Ministère de l’environnement, de l’énergie et de la mer.
- Ran, Q., Li, W., Liao, Q., Tang, H., & Wang, M. (2016). Application of an automated LSPIV system in a mountainous stream for continuous flood flow measurements. *Hydrological Processes*, 30(17), 3014–3029. <https://doi.org/10.1002/hyp.10836>
- Reitan, T., & Petersen-Øverleir, A. (2009). Bayesian methods for estimating multi-segment discharge rating curves. *Stochastic Environmental Research and Risk Assessment*, 23(5), 627–642. <https://doi.org/10.1007/s00477-008-0248-0>
- Renard, B., Garreta, V., & Lang, M. (2006). An application of Bayesian analysis and Markov chain Monte Carlo methods to the estimation of a regional trend in annual maxima. *Water Resources Research*, 42, W12422. <https://doi.org/10.1029/2005WR004591>
- Renard, B., Kavetski, D., Leblois, E., Thyer, M., Kuczera, G., & Franks, S. W. (2011). Toward a reliable decomposition of predictive uncertainty in hydrological modeling: Characterizing rainfall errors using conditional simulation. *Water Resources Research*, 47, W11516. <https://doi.org/10.1029/2011WR010643>
- Renard, B., Kavetski, D., Thyer, M., Kuczera, G., & Franks, S. W. (2010). Understanding predictive uncertainty in hydrologic modeling: The challenge of identifying input and structural errors. *Water Resources Research*, 46, W05521. <https://doi.org/10.1029/2009WR008328>
- Sauer, V. B., Turnipseed, D. P. (2010). *Stage, measurement at gaging stations* (Tech. Methods, Book 3, chap. A7, 45 pp.). Reston, VA: U.S. Geological Survey. Retrieved from <http://pubs.usgs.gov/tm/tm3-a7/>
- Sikorska, A. E., & Renard, B. (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.
- Sikorska, A. E., Scheidegger, A., Banasik, K., & Rieckermann, J. (2013). Considering rating curve uncertainty in water level predictions. *Hydrology and Earth System Sciences Discussions*, 10(3), 2955–2986.
- Steinbakk, G. H., Thorarinsdottir, T. L., Reitan, T., Schlichting, L., Holleland, S., & Engeland, K. (2016). Propagation of rating curve uncertainty in design flood estimation. *Water Resources Research*, 52, 6897–6915. <https://doi.org/10.1002/2015WR018516>
- Steinschneider, S., Polebitski, A., Brown, C., & Letcher, B. H. (2012). Toward a statistical framework to quantify the uncertainties of hydrologic response under climate change. *Water Resources Research*, 48, W11525. <https://doi.org/10.1029/2011WR011318>
- Thyer, M., Renard, B., Kavetski, D., Kuczera, G., Franks, S. W., & Srikanthan, S. (2009). Critical evaluation of parameter consistency and predictive uncertainty in hydrological modeling: A case study using Bayesian total error analysis. *Water Resources Research*, 45, W00B14. <https://doi.org/10.1029/2008WR006825>
- van der Made, J. E. (1982). Determination of the accuracy of water level observations. In *Proceedings of the Exeter Symposium* (Vol. 134, pp. 172–184). IAHS Publications.
- Westerberg, I., Guerrero, J.-L., Seibert, J., Beven, K. J., & Halldin, S. (2011). Stage-discharge uncertainty derived with a non-stationary rating curve in the Choluteca River, Honduras. *Hydrological Process*, 25, 603–613. <https://doi.org/10.1002/hyp.7848>
- Westerberg, I., & McMillan, H. (2015). Uncertainty in hydrological signatures. *Hydrology and Earth System Sciences*, 12, 4233–4270. <https://doi.org/10.5194/hessd-12-4233-2015>
- World Meteorological Organization. (2006). *Technical regulations, Vol. III, Hydrology* (WMO-No. 49, 112 pp.).