Rapid Stage-Discharge Rating Curve Assessment Using Hydraulic Modeling in an Uncertainty Framework

Valentin Mansanarez1,2, Ida K. Westerberg3, Norris Lam1,2, and Steve W. Lyon1,2,4,5

1Department of Physical Geography, Stockholm University, Stockholm, Sweden, 2Bolin Centre for Climate Research, Stockholm University, Stockholm, Sweden, 3IVL Swedish Environmental Research Institute, Stockholm, Sweden, 4The Nature Conservancy, Delmont, NJ, USA, 5School of Environment and Natural Resources, Ohio State University–OARDC, Wooster, OH, USA

Abstract Establishing reliable streamflow time series is essential for hydrological studies and water-related decisions, but it can be both time-consuming and costly since streamflow is typically calculated from water level using rating curves based on numerous calibration measurements (gaugings). It can take many years of gauging data collection to estimate reliable rating curves, and even then extreme-flow estimates often still depend on rating curve extrapolation. Hydraulically modeled rating curves are a promising alternative to traditional methods as they can be rapidly derived with few concurrent stage-discharge gaugings. We introduce a novel framework for Rating curve Uncertainty estimation using Hydraulic Modelling (RUHM), based on Bayesian inference and physically based hydraulic modelling for estimating stage-discharge rating curves and their associated uncertainty. The framework incorporates information from the river shape, hydraulic configuration, and the control gaugings as well as uncertainties in the gaugings and model parameters. We explored the interaction of uncertainty sources within RUHM by (1) assessing its performance at two Swedish stations, (2) investigating the sensitivity of the results to the number and magnitude of the calibration gaugings, and (3) evaluating the importance of prior information on the model parameters. We found that rating curves with constrained uncertainty could be estimated using only three gaugings for either low or low and medium flows that have a high probability of occurrence, thereby enabling rapid rating curve estimation. Prior information about the water-surface slope-stage relation, obtainable from site surveys, was needed to adequately constrain uncertainty estimates.

Plain Language Summary Reliable streamflow time series are essential for water-related decisions. However, it can take several years and numerous measurements to establish a reliable streamflow time series, and these may still be associated with large uncertainty. To address these issues, we developed a novel framework that couples uncertainty assessment with hydraulic modeling of the relation between water level and streamflow at a hydrological monitoring station using information about the physical characteristics of the channel. This relation between water level and streamflow, known as the rating curve, is the basis for calculating streamflow time series from the water level time series measured at hydrological monitoring stations. We explored the interaction of different uncertainty sources on rating curve estimation at two Swedish stations and found that rating curves could be modeled with high confidence (i.e., low uncertainty) using only three observations for either low flows or low and medium flows. Since such flow conditions occur often and are easy to measure (at least relative to the rare and hard-to-measure high flows) our framework has an advantage over traditional approaches by potentially allowing for more rapid rating curve estimation.

1. Introduction

Reliable streamflow time series are essential but can be time-consuming and costly to establish and maintain. Water resource managers, policy makers, and researchers rely on accurate streamflow data—for example, for water quality monitoring (Harmel et al., 2006), flood management (Steinbakk et al., 2016), and detection of change in riverine systems (Juston et al., 2014). Errors in streamflow series may persist for years and lead to incorrect scientific conclusions (Westerberg et al., 2016; Wilby et al., 2017), as well as large societal costs due to long-lasting nonoptimal water management—as exemplified for the Norwegian hydropower sector by McMillan et al. (2017). Despite their importance, streamflow time series are rarely measured directly due to the difficulty or impossibility of measuring continuous discharge. Instead, streamflow time series are most often estimated by transforming a river’s stage into discharge using a model of the water surface-slope-stage relation. This process is known as stage-discharge rating curve estimation.
stage-discharge relation at the gauging site. Such a model is known as a rating curve and is empirically fitted to simultaneous stage-discharge measurements (gaugings) taken at different flow conditions (e.g., Herschy, 1993; Rantz, 1982; Schmidt, 2002).

Under steady flow conditions, rating curves are traditionally estimated using single- or multiple-section power law functions (World Meteorological Organization (WMO), 2010b; International Organization for Standardization, 2010). Such functions are based on hydraulic simplifications and rely on numerous gaugings for calibration and estimation of the rating curve uncertainty. The traditional methods are typically associated with several sources of uncertainty and have known limitations in extrapolation and uncertainty quantification (McMillan et al., 2012; Schmidt, 2002). More advanced methods for rating curve uncertainty estimation have therefore been developed to address these limitations (e.g., Coxon et al., 2015; Le Coz et al., 2014; McMillan & Westerberg, 2015; Morlot et al., 2014; Reitan & Petersen-Øverleir, 2009; Sikorska et al., 2013). A recent comparison study found that different uncertainty estimation methods give the largest differences at low and high flows, particularly in extrapolation and where the hydraulic control is unstable (Kiang et al., 2018).

It typically takes years of data collection efforts to assess a reliable rating curve and to define the associated uncertainty, particularly as there are often economic and practical constraints on the number of calibration gaugings that can be made. It is costly to make the frequent field measurement campaigns needed to cover all relevant flow conditions. Birgand et al. (2013) investigated discharge monitoring in small streams for short-term projects and found that 22 manual gaugings per year were needed to obtain acceptable rating curve and streamflow data. In addition, practical limitations typically affect high-flow gauging because (1) extreme-flow events happen infrequently and are often unpredictable; (2) many catchments have fast rainfall-runoff response and short peak-flow duration, giving little time for gauging (Westerberg et al., 2016); (3) high-flow events can be affected by unsteady flow (e.g., Mansanarez et al., 2016); and (4) peak flows can happen during the night or under extreme weather, making discharge measurements highly uncertain, unsafe, or even impossible to perform (WMO, 2010a). This means that high flows often need to be extrapolated and that the stage-discharge gaugings can have significant measurement uncertainty at high flows. Measurement uncertainty can also be large in relative terms at low flows, with typically lower uncertainty in the middle-flow range (McMillan et al., 2012). The stage-discharge relation can also vary over time (Guerrero et al., 2012; Tomkins, 2014), implying the need for reestablishing the rating curve (WMO, 2010b) and thus performing new discharge measurement campaigns across the full range of flows. As such, uncertainty in the discharge measurements themselves, in temporal variability of the stage-discharge relation, and in the establishment of the stage-discharge relation for ungauged parts contribute to often large uncertainty in streamflow estimates.

At the same time, the need for rapid assessment of rating curves for discharge data calculation is increasing as water resource systems are changing at an accelerating pace (Ceola et al., 2016; Montanari et al., 2013), increasing the need for discharge data in previously ungauged locations. In addition, for certain regions the frequency of flood events has increased due to climate change, leading to more frequent shifts in channel morphology (Hirsch & Archfield, 2015; Mallakpour & Villarini, 2015; Milly et al., 2005) and thus the need for more frequent rating curve changes.

Hydraulic modeling has proven to be a good alternative for more quickly deriving stage-discharge models for discharge estimation, especially for assessing more reliable high-flow rating relations in extrapolation beyond observation gaugings (e.g., Di Baldassarre & Claps, 2011; Lang et al., 2010). For example, Kean and Smith (2005) derived stage-discharge rating curves by applying their own hydraulic model at two inland rivers. They found accurate high-flow discharge estimations using only one calibration gauging at lower flows. Reistad et al. (2007) also estimated more reliable stage-discharge rating curves for high flows at four Norwegian sites using the Hydrologic Engineering Center River Analysis System (HEC-RAS) model with predefined Manning’s roughness values. In addition, hydraulic modeling efforts have been shown to be useful for investigating unsteady stage-discharge relations, such as rating curve hysteresis (Lee et al., 2017; Lee & Muste, 2017), and for developing rating curves for ephemeral streams (e.g., Bullard et al., 2007) or even at completely ungauged locations (Clayton & Kean, 2010).

Although rating curves developed with hydraulic models have proven to be accurate (Lam et al., 2016; Lyon et al., 2015), uncertainties in input measurements (i.e., both discharge and other model inputs),
topography (Casas et al., 2006), and model parameters (roughness coefficient and boundary conditions) can affect the accuracy of the modeling results. Lang et al. (2010) show that stage-discharge gauging errors should be considered in hydraulic modeling as they impact on rating curve assessment in the extrapolation part. Di Baldassarre and Claps (2011) highlight the limitations of traditional rating curve methods in extrapolation and use a one-dimensional hydraulic model (HEC-RAS) as an alternative to constrain discharge. They show, using Generalized Likelihood Uncertainty Estimation (GLUE), that uncertainty in the Strickler coefficient in the hydraulic model should be taken into account when deriving hydraulically modeled rating curves. Furthermore, they point out the difficulty of assessing the uncertainty of hydraulically modeled rating curves and acknowledge that other sources of uncertainty should be considered, such as discharge measurement errors. While advantages are gained through hydraulic rating curve modeling by leveraging the physics of open channel flows to constrain uncertainty, additional uncertainty is introduced from the hydraulic model itself (e.g., through parameterization approximations) in combination with increased impact of uncertainties due to few calibration observations. There is therefore a need to investigate how uncertainties in hydraulic models and data impact on hydraulically modeled rating curves.

In this study, we explore the interaction of these sources of uncertainty through a systematic evaluation of hydraulic modeling of stage-discharge rating curves within a novel uncertainty framework. We take the physically based hydraulic modeling work developed by Kean and Smith (2005, 2010) to represent stage-discharge relations calibrated to observed flows. This modeling approach is integrated into a novel uncertainty framework for Rating curve Uncertainty estimation using Hydraulic Modelling (RUHM) capable of analyzing the impact of the various uncertainty sources through Bayesian inference. Our main aim is to develop and evaluate the hydraulic modeling uncertainty framework for rapidly estimating rating curves based on few calibration gaugings. Such rapidly estimated rating curves would provide better initial estimates of streamflow at previously ungauged sites or at gauged sites after stage-discharge alterations, thereby enabling correct scientific conclusions to be drawn faster as well as reduced costs to water management. Our three specific objectives were to (1) apply and test the framework at two Swedish sites, (2) assess the sensitivity of the results to the number and the magnitude of the calibration gaugings, and (3) evaluate the need for prior information on the model parameters.

2. Materials and Methods
2.1. Hydraulic Modeling of Rating Curves

Rating curves were empirically derived from the hydraulic model of Kean and Smith (2005, 2010). In this section, we present a brief overview of this hydraulic model with the main changes made to adapt the model for uncertainty assessment within RUHM. For a full model description (but without uncertainty assessment) we refer to several example applications in previous studies (e.g., Clayton & Kean, 2010; Kean & Smith, 2005; Kean & Smith, 2010; Lam et al., 2016; Lyon et al., 2015; Nathanson et al., 2012). The general required information for modeling hydraulically based rating curves is described here with site-specific information provided in section 2.3.

The physically based hydraulic model of Kean and Smith (2005, 2010) has been developed for relatively straight rivers that have width-to-depth ratios greater than about 7 (Kean & Smith, 2010) and gravel bed roughness elements smaller than the depth of flow. The model assumes streamwise steady uniform flow with logarithmic velocity profiles. As explained by Nathanson et al. (2012), the hydraulic model of Kean and Smith (2005) differs from standard one-dimensional flow models used for discharge estimations (e.g., HEC-RAS) in the channel roughness specification. Instead of modeling the roughness coefficient as a single parameter, and thus using an average value over the whole reach, the hydraulic model can incorporate different roughness characteristics along the stream reach. For example, it can take into account locations and values of high relative roughness, such as dense vegetation, along the banks or the presence of specific small roughness elements within the channel (e.g., boulders). The roughness of the channel is thus defined as a roughness height, \( z_0 \). If the streambed material is accessible at some location of the reach, the roughness height, \( z_0 \), can be approximated by \( z_0 = 0.1D_{s4} \), where \( D_{s4} \) represents the 84th percentile of the streamed particle size distribution (Whiting & Dietrich, 1990). Therefore, at locations within a stream reach where the roughness characteristics can be quantified...
through geometric measurements such as pebble count surveys, laser scans, or vegetation density surveys (Lam et al., 2015; 2016), fixed values of the roughness can be used in the hydraulic model. At locations where these roughness characteristics are not provided, the roughness height parameter, $z_0$, typically used to model the resistance to the flow, is considered an average value for the whole stream reach. Contrary to previous versions of the hydraulic model (e.g., Kean & Smith, 2010; Lam et al., 2016), the roughness height parameter, $z_0$, implemented in this study was no longer fixed manually or back-computed from discharge as in these aforementioned approaches. Instead, the roughness height parameter, $z_0$, was modeled as an independent parameter within the uncertainty framework. This is advantageous as it allows specifying the uncertainty of this parameter independently from the other parameters.

For computing flows, the physically based hydraulic model uses the river geometry (main channel plus floodplain) obtained through topographic and bathymetric surveys. Geometric measurements of the biological (i.e., woody vegetation) and physical (i.e., banks and streambed material) elements of the stream reach are used to quantify the roughness. A water-surface boundary condition (Kean & Smith, 2005) is used to allow discharge to be computed for a given stage value. Similarly to Lam et al. (2016), the water-surface slope, $S$, was modeled as a linear function of stage, $h$:

$$ S = \frac{S_2 - S_1}{h_2 - h_1} h + \frac{S_1 h_2 - S_2 h_1}{h_2 - h_1} $$

(1)

where $S_1$ and $S_2$ were two distinct average water-surface slope values for two distinct stages $h_1$ and $h_2$, respectively. The application of equation (1) requires at least two distinct slope values (i.e., $S_1$ and $S_2$) to compute the water-surface slope-stage relation.

After defining the topography (bathymetry) of the river, the roughness parameters, and the water-surface slope-stage relation, the hydraulic modeling procedure consists of computing vertical velocity profiles using the water-surface slope-stage relation (from equation (1)) for every submerged point of the river grid in the modeling domain. This is done by numerically solving the one-dimensional momentum flow equations for steady uniform flow for a given stage value. The governing mass conservation and momentum flow equations are:

$$ \begin{cases} 
\frac{\partial Q}{\partial x} = 0 \\
\frac{1}{2} \frac{\partial (u^2)_{av}}{\partial x} + \frac{\partial E}{\partial x} + \frac{1}{\rho} \frac{\partial \tau_b}{\partial R} = 0
\end{cases} $$

(2)

where $(u^2)_{av}$ (m$^2$/s$^2$) is the average component at the cross section of the square of the downstream velocity, $g$ (m$^2$/s) is the gravity acceleration, $x$ (–) is the downstream direction, $E$ (m) is the surface water elevation, $\rho$ (–) is the density of water, $(\tau_b)_{av}$ (m$^2$/s$^2$) is the perimeter-averaged shear stress, and $R$ is the hydraulic radius of the cross-sectional area.

Using equation (2) and the definition of the perimeter-averaged shear velocity $u^*$(m/s), the vertical velocity $u$ (m/s) is calculated as follows:

$$ u = \beta_r u^* = \beta_r \left( \frac{\tau_b}{\rho} \right)^{1/2} $$

(3)

where $\beta_r$ (–) is a nondimensional roughness coefficient. Under the assumption of streamwise steady flow conditions, the shear stress is equal to:

$$ \tau_b = \rho g y S_f $$

(4)

where $y$ (m) is the local flow depth and $S_f$ (–) is the friction slope equal to the water-surface slope $S$ for uniform flow.
According to Kean and Smith (2005), under steady flow conditions, \( \beta_r \) can be written as follows:

\[
\beta_r = \frac{\ln \left( \frac{z_0}{\kappa} \right) - 0.74}{\kappa}
\]

(5)

where \( \kappa \) is the von Karman constant equal to 0.408 (Long et al., 1993). Combining equations (4) and (5) with equation (3), \( u \) at any point can be computed under uniform flow as follows:

\[
u = \ln \left( \frac{z_0}{\kappa} \right) - 0.74 \kappa \frac{\sqrt{gS}}{uQ_i/C_{16}/C_{17}}
\]

(6)

The flow accelerations are only resolved in one dimension (Kean & Smith, 2005), while the velocity is numerically computed in two dimensions (vertically and streamwise). Thus, for a given stage value, equation (6) is applied for every submerged point of the topographic grid (Kean & Smith, 2010) and is numerically solved considering a boundary-fitted scheme (Wobus et al., 2008) within a semi-implicit method for pressure-linked equations similar to that outlined by Patankar (1980). Discharge is then computed by multiplying the modeled cross-sectional velocities with the cross-sectional area for the corresponding stage at a defined gauge location. Finally, the rating curve is derived by repeating this modeling procedure for different values of stage.

### 2.2. Uncertainty Framework Using Bayesian Inference

We used Bayesian inference to assess uncertainty in the hydraulically modeled rating curves. The following section describes the uncertainty estimation part of the RUHM framework; we refer to Gelman et al. (2004) for a full general description of Bayesian inference.

#### 2.2.1. Parametrization

Let \( \theta = (\beta_1, \beta_2, \beta_3) \) denote the inferred model parameters. They correspond, in this order, to the parameters of the hydraulic model (section 2.1): the roughness height \( z_0 \) and the two water-surface slope values, \( S_1 \) and \( S_2 \), used in the water-surface slope model (equation (1)). These three parameters are assumed independent.

#### 2.2.2. Likelihood Computation

The gaugings \( \left( \hat{h}_i, \bar{Q}_i \right)_{i=1,N} \) are seen as \( N \) estimates of the real values \( (h_i, Q_i)_{1,N} \) of stage and associated discharge. We assumed that stage errors are negligible, but not discharge errors:

\[
\begin{align*}
\hat{h}_i &= h_i \\
\bar{Q}_i &= Q_i + \epsilon_{Q,i} \text{ with } \epsilon_{Q,i} \sim N(0, u_{Q,i})
\end{align*}
\]

(7)

where the discharge errors \( \epsilon_Q = (\epsilon_{Q,1}, ..., \epsilon_{Q,N}) \) are assumed independent and the standard deviations \( u_{Q,i} \) (i.e., the uncertainties of the discharge measurements) are assumed to be known. Where no site-specific information about measurement uncertainties is available with the gauging data, uncertainty values can be assigned based on typical literature values from, for example, measurement error analyses (e.g., Despax et al., 2016), or in situ intercomparisons (e.g., Le Coz et al., 2016). Assuming the standard deviations to be known is a simplifying assumption that is commonly used in rating curve uncertainty analysis (we refer to Le Coz et al., 2014, for an extensive discussion of this assumption).

The true discharge is then written as the discharge predicted by the hydraulic model \( f \):

\[
Q_i = f(h_i|\theta)
\]

(8)

Combining equations (7) and (8) yields the following stage-discharge relation between the observed gaugings:

\[
\bar{Q}_i = f(\hat{h}_i|\theta) + \epsilon_{Q,i} \text{ with } \epsilon_{Q,i} \sim N(0, u_{Q,i})
\]

(9)

Note that in equations (8) and (9), contrary to other Bayesian methods using traditional power law rating curves (e.g., Reitan & Petersen-Øverleir, 2009; Le Coz et al., 2014), we only considered the parametric and
data uncertainty in this study. This was mainly because we use too few gaugings to be able to estimate the parameters of a statistical error model.

The likelihood function, \( L \), expresses the probability of the observed data, given the hydraulic model and its parameters, thereby allowing information about the inferred model parameters to be estimated from the observed data. From equation (9) this likelihood, \( L \), of observed discharge values \( \bar{Q} \) is written as follows:

\[
L(\bar{Q} | \theta, h) = \prod_{j=1}^{N} p_{\text{norm}}(\bar{Q}_j | h_j, u_{Q_j})
\]  

(10)

where \( \bar{Q} = (\bar{Q}_1, ..., \bar{Q}_N) \) is the gauged discharge and \( p_{\text{norm}}(z | \mu, \sigma) \) denotes the probability density function of a Gaussian distribution with mean \( \mu \) and standard deviation \( \sigma \) that is evaluated at some value \( z \).

2.2.3. Prior Specification

Bayesian inference allows using prior distributions to include more information about the inferred parameters \( \theta \) that is independent from the observed data. Independent prior distributions were used for the three parameters in the hydraulic model, which leads to the following joint prior distribution:

\[
p(\theta) = \prod_{i=1}^{3} p(\theta_i)
\]  

(11)

where \( p(\theta_i) \) is the marginal prior distribution of the \( i \)th parameter. The specification of the prior distributions is site specific and is described below in section 2.3.

2.2.4. Posterior Inference Using Markov Chain Monte Carlo Sampling

According to Bayes’ theorem, information from the observed data (i.e., the likelihood function given in equation (10)) can be combined with the prior information (equation (11)), yielding the posterior distribution, given a constant of proportionality:

\[
p(\theta | \bar{Q}, h) \propto p(\theta) L(\bar{Q} | \theta, h)
\]  

(12)

The posterior distribution was explored using Markov chain Monte Carlo (MCMC) sampling to estimate possible values of each parameter \( \theta \) in the hydraulic model and thereby possible rating curves. We used an adaptive Gibbs/Metropolis sampler from Renard et al. (2006). In the MCMC sampling, two initial stages were realized to adapt the sampler properties to the inference problem. Four parallel chains of 20,000 simulations were performed. The first quarter of these simulations were discarded as a burn-in period. The convergence was checked by verifying the Gelman-Rubin criterion (Gelman & Rubin, 1992) on the remaining 15,000 simulations in each chain and was found to be \( \leq 1.2 \) for all the inferred parameters. The remaining 15,000 simulations were thinned by a factor of 10 to restrict computing time and attenuate the strong autocorrelation in the raw MCMC samples. The implementation of the MCMC sampler (number of simulations, burn-in, and thin factors) was investigated (analysis not shown) by simulating longer MCMC chains, and no noticeable loss of information was found.

2.3. Case Study: Two Swedish Sites and Their Prior Specifications

We applied the framework to two Swedish sites, the Röån and Nybro stations. The Nybro station presented more challenging nonideal conditions as backwater effects at high flows had been observed during several field visits.

2.3.1. The Röån River at Röån, Sweden

The Röån River is a tributary of the Ångermanälven River. At the Röån station, the Röån River has a mean annual discharge of 5.8 m\(^3\)/s and a catchment area of 584 km\(^2\). The hydrometric station (location: 63°38′29.7″N 16°45′11.6″E) is managed by the Swedish Meteorological and Hydrological Institute (SMHI) and is located 30 m downstream of a pedestrian bridge (Figure 1).

The discharge station has been operating since 1943, and the current stable period is valid since March 1995. During winter the station is sometimes affected by backwater effects due to ice jam downstream. The effect of ice jams was not considered in this study or in the official SMHI rating curve. The official curve was calibrated using 13 gaugings well-distributed across the flow range from 1.81 to 19.99 m\(^3\)/s. In addition, we
conducted one independent gauging in October 2013 with a measured discharge value of 9.23 m$^3$/s. This gauging and two other SMHI gaugings from February 2014 and April 2015 were used as calibration data in the example application presented in section 3.1. These measurements were performed with an acoustic Doppler current profiler (ADCP) or current meter, but we had no complete information about the quality of each gauging. To remove any possible effects in our subsequent sensitivity analyses due to different discharge measurement errors, the same uncertainty of ±5% was used for all gaugings in this study. This value was based on literature values for ADCPs (Lee et al., 2014) with which the majority of all gaugings at the two stations were measured.

The topography of the river was surveyed by ADCP in October 2013 and the river banks were modeled using the airborne laser scanning data (LiDAR) collected by Lantmäteriet (Swedish Land Survey) in June 2012 (see Lam et al., 2016, for full study site and data collection description). During the topographic survey, tree stem density surveys were performed to determine the flow resistance from woody vegetation (e.g., birch, pine, and spruce) on the banks at the Röån site. Two classes were identified representing small and large diameters of the woody vegetation within the reach area (Lam, 2017). The location and roughness values derived from these surveys were fixed in the hydraulic model. The streambank material within the study reach was composed of sand and gravel. Therefore, a uniform distribution between 0 and 0.3 m was used as uncertain prior information on the roughness height parameter, $z_0$ (Table 1).
Two water-surface slope surveys were performed, where elevation measurements of the water surface were taken where it intersected the bank along the reach: the first was at the same time as the topographic survey (October 2013) and the second in May 2016. Linear regressions on these data were made to determine prior distributions for the water-surface slope parameters as follows. A uniform measurement error of ±0.023 m was accounted for within every elevation value collected in the two surveys (Lam, 2017). We then propagated this uniform measurement error for the elevation values in a Monte Carlo analysis, fitting a linear regression to each sampled set of elevation data. The resulting distributions of regression parameter values were used to estimate normal prior distributions on the two water-surface slope parameters (Table 1).

### 2.3.2. The Voxnan River at Nybro, Sweden

The Voxnan River catchment, located in central Sweden, is gauged at the Nybro station (location: 61°21′46.3″ N, 15°31′30.9″ E) and is also managed by the SMHI. At Nybro, the Voxnan River has a catchment area of 2,251 km² and a mean annual discharge of 25.6 m³/s. The hydrometric station is located just downstream of a motor vehicle bridge (Figure 1). Upstream this bridge, the channel is fairly straight with moderately steep and stable banks without evidence of excessive bank erosion.

As with the Röån station, during winter, the station can be affected by variable backwater due to ice jams downstream, and this effect was not considered in the rating curve calibration. Backwater influence due to the immediate vicinity of the bridge (see Figure 1) has also been observed in the modeled reach. There is no explicit modeling of these backwater effects (e.g., Mansanarez et al., 2016) in the official rating curve, but it is considered indirectly through the rating curve calibration. The official rating curve has been calibrated by the SMHI using 13 gaugings and is considered valid and stable since 1991. Two more recent discharge measurements at medium flow were performed by the SMHI in November 2014 and April 2016 and agree with this calibrated rating curve.

In addition to these 15 SMHI gaugings, we performed two independent gaugings in October 2013 and May 2016. These two gaugings together with the SMHI gauging in April 2016 were used as calibration data in an example application of the hydraulic model (section 3.1) to illustrate model results in detail. These discharge measurements were performed in good conditions using ADCPs.

In September 2013 two stem density surveys were performed to determine the flow resistance from woody vegetation (e.g., birch, pine, and spruce) on the banks at the Nybro site. During the topographic survey, vegetation stem densities on the banks were also measured (Lam, 2017). The resulting vegetation values (roughness and location) were considered as fixed values in the hydraulic model (Figure 1). The streambed material in the channel is made of sand and small gravel. A uniform distribution between 0 and 0.3 m was set as prior for the roughness height parameter \( z_0 \) (Table 1).

Two water-surface slope surveys were performed in April 2014 and May 2016. The same Monte Carlo procedure using linear regressions as conducted for the Röån site was made on these survey data to obtain normal prior distributions of the two water-surface slope parameters (using the same uniform error of ±0.023 m on the elevation measurements).

The topographic information for the modeled stream reach was obtained by combining surveys of the stream bank elevation, conducted in September 2013, and stream bathymetry, conducted with ADCP measurements in October 2013 (Figure 1). The combined topographic survey consisted of 16 channel cross-

| Inferred parameter | Physical parameter | Unit | Nybro | Röån |
|--------------------|--------------------|------|-------|------|
| \( \theta_1 \)     | \( z_0 \)          | m    | U(0; 0.3) | U(0; 0.3) |
| \( \theta_2 \)     | \( S_1 \)          | –    | N(−3.41 × 10⁻⁵; 1.27 × 10⁻⁵) | N(−4.10 × 10⁻⁵; 1.59 × 10⁻⁵) |
| \( \theta_3 \)     | \( S_2 \)          | –    | N(−1.25 × 10⁻⁴; 6.24 × 10⁻⁵) | N(−5.00 × 10⁻⁴; 1.78 × 10⁻⁴) |

Note. The symbol \( \mathcal{N}(\mu; \sigma) \) denotes a normal distribution with mean \( \mu \) and standard deviation \( \sigma \). The symbol \( \mathcal{U}(a; b) \) denotes a uniform distribution on the interval \([a; b]\).
sections and was located just upstream of the bridge along a straight 290-m-long and 30-m-wide reach (length-to-width ratio of 9.7). Within the stream bank survey, elevation points were collected with a robotic total station and an adjustable prism rod from the top of the stream banks to about 1 m into the stream channel from the channel bank (mean length of 7 m). About 18 points per cross section were collected with a mean linear point spacing of 0.39 m. For the bathymetric information, the ADCP was deployed to coincide with the stream bank survey, and each cross section was surveyed twice as the instrument was taken across the river and back. This resulted in 32 individual bathymetric transects (Figure 1) with an average of 915 depth measurements per transect, a mean transect length of 30 m, and a mean transect width of 3.10 m. A maximum recorded depth of 5.25 m was observed with a mean depth of 2.28 m. The combined topographic survey was then interpolated into a uniformly spaced curvilinear grid surface. Ninety-two cross sections were extracted from this surface and used to model the river geometry (one cross section every 3.2 m), where each cross section contained 251 grid points (spacing of 12 cm).

The topography survey did not cover the SMHI staff gauge (located just downstream the motor vehicle bridge). Therefore, a stage offset of 0.8 m was used to relate the stage data recorded by the SMHI with our hydraulically modeled rating curve results. This offset was based on staff gauge observations between the SMHI’s gauge and a temporary gauge that we installed upstream during the topography survey (at the reference cross section for which the rating curve results were calculated). No uncertainty in this offset was considered.

2.4. Evaluation of the Modeled Results and Sensitivity Analyses

We first analyzed modeling results using three example gaugings for calibration at each site to illustrate application and model results in detail. We then analyzed the performance of the framework for a wide range of calibration gauging data sets to be able to draw general conclusions. We investigated the sensitivity of the framework to the amount of information provided by different configurations of calibration gaugings and parameter priors to analyze capabilities under different monitoring and prior information conditions. Specifically, we used the sensitivity analyses to investigate two questions of importance for the practical usefulness of the framework for rapid rating curve assessment: (1) How many calibration gaugings are needed and at what flow conditions do they need to be taken? And (2) which parameters require informative prior information?

2.4.1. How Many Calibration Gaugings Are Needed and at What Flow Conditions Do They Need to Be Taken?

Rating curve calibration typically requires low-, middle-, and high-flow gaugings to obtain reliable results, and typically more extreme gaugings (i.e., those with a lower probability of occurrence) provide more information as they constrain the upper and lower parts of the curve where extrapolation may often lead to large errors (see, e.g., Kiang et al., 2018). To be able to investigate different gauging scenarios that represented these issues, we first grouped the gaugings by flow magnitude using three classes based on flow duration curves (showing the probability of exceedance of discharge). For each site, the flow duration curve was computed using the discharge time series corresponding to the official rating curve. Based on the shape of the flow duration curves, we defined high, middle, and low discharge classes as corresponding to probabilities of exceedance under 0.1, between 0.1 and 0.7 and over 0.7, respectively. The gaugings for each station were then grouped according to these classes and used to define the different gauging scenarios together with the probability of occurrence of each gauging that was calculated from the streamflow time series.

Gauging scenarios representative of both ideal and practical stream monitoring practices were then developed for assessing model sensitivity to the calibration data configuration for each site (Table 2). Scenarios investigating the value of more/less extreme gaugings in each class (GS1 and GS2) were developed. We also investigated scenarios representing the best spread in discharge across the flow range (GS3) and the added value of one extra low-probability high/medium/low-flow gauging in addition to three gaugings of high probability (GS4, GS5, and GS6). These scenarios that were based on at least one gauging in each discharge class were compared to scenarios using the three gaugings with the highest probability overall (GS10), the three highest-probability medium-flow gaugings (GS12), and the three overall highest/lowest gaugings (GS11 and GS13). These contrasting scenarios allowed us to investigate the most appropriate gauging strategy to be used with the hydraulic model and possible impacts of having nonideal gauging data for calibration.
In addition to these scenarios using only three or four gaugings, we investigated the impact of the number of gaugings by comparing the scenarios based on one gauging in each class (GS1 and GS2, each having three gaugings in total) with two scenarios using two gaugings in each class (GS7 and GS8, each having six gaugings in total). We also investigated the scenario where all available gaugings were used in calibration (GS9). We used this scenario approach instead of a full-scale randomly sampled MC analysis (1) because it is more representative of realistic gauging strategies and (2) because of the long time required to run multiple scenarios in the framework (calibration with three gaugings takes about 2.5 days and six gaugings takes about 1 week). The posterior rating curve uncertainty results calibrated for each scenario were then evaluated using the remaining gaugings for evaluation.

2.4.2. Which Parameters Require Informative Prior Information?

Apart from the gaugings, information for calibrating the model in the uncertainty framework can be provided from the prior parameter distributions. We investigated the value of using informative versus noninformative prior parameter information for reducing uncertainty in the modeled rating curve results. We performed four simulations with noninformative prior distributions (i.e., a uniform prior distributions on \([0; 10]\) for the roughness height \(z_0\) and two uniform distributions on \([-1; 0]\) for the two slope parameters \(S_1\) and \(S_2\)) and compared the posterior parameter distributions to the results from calibration using informative prior distributions (Table 1). We used the same three calibration gaugings as in the example application (see section 3.1) to enable direct comparison with those results. For the three first simulations, the prior distribution of only one parameter at a time was changed into a wide uniform noninformative distribution. For the fourth simulation we investigated using noninformative priors on the two water-surface slope parameters and an informative prior only on the roughness height.

3. Results

3.1. Example Application of the Uncertainty Framework to the Two Swedish Sites

In this section we present example applications of the uncertainty framework using three gaugings covering low and medium flows for calibration to first illustrate the results in detail for each site. The general

---

**Table 2**

*Gauging Scenarios (GS) Used in the Sensitivity Analyses*

| Gauging scenario number | Gauging scenario name | Description | Total number of gaugings |
|-------------------------|-----------------------|-------------|--------------------------|
| GS1                     | Highest class probability | The gauging with the highest probability of occurrence per discharge class | 3 |
| GS2                     | Lowest class probability | The gauging with the lowest probability of occurrence per discharge class | 3 |
| GS3                     | Best spread            | Best spread in discharge values across the discharge range using one gauging per discharge class | 3 |
| GS4                     | GS1 + high flow        | Scenario GS1 with one extra gauging in the high-flow class with the lowest probability of occurrence | 4 |
| GS5                     | GS1 + low flow          | Scenario GS1 with one extra gauging in the low-flow class with the lowest probability of occurrence | 4 |
| GS6                     | GS1 + medium flow       | Scenario GS1 with one extra gauging in the medium-flow class with the lowest probability of occurrence | 4 |
| GS7                     | 2 high-probability gaugings | The two gaugings with the highest probability of occurrence per discharge class | 6 |
| GS8                     | 2 low-probability gaugings | The two gaugings with the lowest probability of occurrence per discharge class | 6 |
| GS9                     | All gaugings           | All the available gaugings | 14 at Röån and 17 at Nybro |
| GS10                    | Highest probability     | The three gaugings with the highest probability of occurrence irrespective of discharge class | 3 |
| GS11                    | High flow               | The three gaugings with the highest discharge values irrespective of discharge class | 3 |
| GS12                    | Medium flow             | The three medium-flow gaugings with the highest probability of occurrence | 3 |
| GS13                    | Low flow                | The three/four gaugings with the lowest discharge values irrespective of discharge class | 3 at Nybro and 4 at Röån* |

*We used four gaugings instead of three for Röån in this scenario as the Markov chain Monte Carlo did not converge when using only three gaugings.
The performance of the framework is then evaluated for a wide range of possible calibration gauging data sets in the following sensitivity analysis section.

Calibration to only three gaugings in these example applications clearly reduced the rating curve uncertainty intervals of the uncalibrated model (i.e., the prior uncertainty bands in Figure 2). For both sites, the posterior rating curve uncertainty intervals were smaller than the prior intervals by a factor of at least 10 on average (Figure 2). For medium and high flows, including in the extrapolation regions at Röån, the errors to the evaluation gaugings were less than ±15% (Figure 3) and the widths of the 95% rating curve uncertainty intervals relative to the maximum posterior (MAP) rating curve were also constrained (<±10% at Röån and <±15% at Nybro). At the extrapolated high flows the posterior results at Nybro were, however, worse than those at Röån because the relative errors to the observations increased with discharge. This was most likely because of the observed backwater influence at Nybro at high flows as the estimated results yielded a higher discharge for the same stage value than both the official rating curve and the two highest gaugings. Backwater conditions raise the water level downstream of the modeled reach, which is consistent with the higher stage for the same discharge value in the gaugings compared to the modeled results. The hydraulic model is based on the assumption of uniform flow and does not account for this effect. The model followed the third highest ADCP gauging, but according to the SMHI (B. Göransson, personal communication, 2018–09–09), this discharge value should likely be around 10% lower due to problems in the surface velocity extrapolation.

**Figure 2.** Stage-discharge representations in normal and logarithmic scales of the prior and posterior rating curve results from the hydraulic modeling uncertainty framework applied to the Röån and Nybro sites in Sweden. MAP = maximum posterior.
extrapolation below the lowest calibration gaugings, the uncertainty intervals were larger relative to the MAP and the gaugings (Figure 3). Larger relative uncertainties are expected at low flows: McMillan et al. (2012) review previous studies and report typical magnitudes of ±50–100%, which are similar to those found here. For Nybro the extrapolated low-flow 95% uncertainty intervals were underestimated but still close to the gaugings, whereas the uncertainty distribution for Röån was underestimated by at least 15% (Figure 3).

The roughness height parameter \( z_0 \) was well identified with peaked posterior density distributions at both sites (Table 1 and Figure 4), which agreed well with the field-estimated values (Lam, 2017; Lam et al., 2016). The two posterior water-surface slope parameters \( S_1 \) and \( S_2 \) were more uncertain at Röån; the 95% interval for the \( S_1 \) posterior density distributions was constrained to around 57% of the prior distribution and the \( S_2 \) to 68%, whereas at Nybro the \( S_2 \) parameter had a well-constrained posterior density (95% interval of 29% of the prior, Figure 4). Multimodality effects in the posterior parameter densities were encountered at Röån when using few calibration gaugings (see, e.g., the Röån example application in Figure 4). The multimodality mainly stems from the combination of only using a few calibration gaugings and relatively wide priors, which means that little information is supplied to the model in the calibration. These effects were reduced or disappeared when using more gaugings.

### 3.2. Sensitivity Analysis: How Many Calibration Gaugings Are Needed and at What Flow Conditions?

The sensitivity analyses for Röån are presented below, and the corresponding results for the Nybro station are reported in the supporting information (Figures S1–S3). Overall, for the gauging scenarios that have at
least one gauging per discharge class (GS1–GS9), the posterior results showed good agreement with the evaluation gaugings at Röån; the MAP errors were within 12% for high flows, 20% for medium flows, and within 30% for low flows (Figure 5) and thereby match typical values from other studies (McMillan et al., 2012). The posterior rating curve uncertainty distributions were also well constrained with 95% intervals relative to the MAP smaller than ±10% for medium and high flows and within ±30% for low flows (Figure 6). At Nybro, for medium and low flows the results were similar to those at Röån, but with higher errors in extrapolation around the lowest evaluation gauging (up to 90% error in GS1 and GS6; Figure S2). For high flows the results were worse at Nybro (similar to the example application described in section 3.1), with errors to observations up to 25% in the gauged range and positive errors up to 50% in extrapolation. This is consistent with the observed backwater influences at high flows, which would result in a higher stage for the same discharge value in the observations than in the model that is based on the assumption of uniform flow and therefore no backwater influences.

When investigating the required number of calibration gaugings and their distribution across the flow range for the gauging scenarios having at least one gauging per discharge class (i.e., GS1–GS9), already the GS1 scenario using the gaugings with the highest probability of occurrence per discharge class showed good results compared to the evaluation gaugings and in terms of constrained uncertainty bounds at Röån (Figures 5 and 6). Subsequently, adding one calibration gauging at medium or high flow (scenarios GS4 and GS6) led to small improvements for high flows compared to GS1, whereas adding a lower gauging (GS5) had a large improvement on the low-flow rating curve results. At Nybro, the GS1 results (Figure S2) were similar to the example application (Figure 3) and adding the medium-flow gauging showed little effect, whereas both the low- and high-flow gaugings improved the estimation. Using the lowest probability of

![Figure 4. Marginal prior and posterior densities of the parameters in the hydraulic rating curve model. Note that these plots show the posterior density and not the likelihood (to be able to compare to the prior distributions). Note also that the maximum posterior (MAP) is not the same as the mode of the marginal posterior probability density function in the figure because of the smoothing from the kernel density estimator (this is because the mode occurs in a very narrow region).](image-url)
occurrence gaugings (GS2) and those with the best spread across the flow range (GS3) instead of the highest-probability gaugings (GS1) led to improvement mainly for low flows for both stations. For these two scenarios (GS2 and GS3), there was little change at Röån above low flows, but better high-flow and poorer medium-flow results at Nybro (Figures 5 and S2). The strategy of using two low/high-probability gaugings in each class (GS7 and GS8, i.e., in total six gaugings) instead of one (GS1 and GS2, i.e., in total three gaugings) led to some improvement at Röån and also at Nybro, whereas there was little further improvement between these results compared to when using all the gaugings (GS9).

When using the gauging strategies that did not have calibration gaugings covering all three discharge classes (i.e., GS10–GS13), there were larger differences in the results (Figures 7 and S1). Using the three gaugings...
with the highest probability of occurrence overall (GS10) gave good results, similar to GS1. The poorest strategies of all were the two that used only high-flow or medium-flow gaugings, which resulted in poor performance and/or wide uncertainty intervals at both stations. However, using only the lowest three gaugings (four at Röån because the MCMC chains did not converge for three gaugings in this case) resulted in relatively good rating curve results (uncertainties up to ±20%) at both stations. At Nybro these were among the best results for any scenario (but still biased for part of the flow range), while at Röån, discharges at medium and high flows had a small consistent underestimation (MAP errors between −5% and −15%). This shows that low flows are valuable for hydraulic model calibration as they can have a strong ability to constrain the rating curve uncertainty across higher flow ranges. This is likely because the topography and the fixed roughness values from the vegetation surveys in the model are more representative of high-flow than of low-flow conditions (e.g., Kean & Smith, 2005), so that they provide enough information to constrain high flows, but not low flows.

Taken all together, these results show that the hydraulically modeled rating curves in this uncertainty framework could be acceptably calibrated using only three gaugings for either only low flows or low and medium

Figure 6. Sensitivity analysis of the hydraulically modeled rating curve uncertainty framework to the number and configuration of calibration data applied to the Röån River at Röån: Relative errors of estimated discharges, $Q_{est}$, compared to maximum posterior (MAP) discharges, $Q_{MAP}$. $Q_{MAP}$ values are represented using a logarithmic scale. Green lines represent the discharge values of the calibration data, black lines represent the boundaries between the low-, medium-, and high-flow discharge classes. The gauging scenarios (GS) are explained in Table 2.
Figure 7. Sensitivity analysis of the hydraulically modeled rating curve uncertainty framework to the distribution of the calibration gauging data across the flow range applied to the Röån River at Röån. (left) Relative errors of estimated discharges, $Q_{est}$, compared to discharge observations, $Q_{Obs}$. (middle) Relative errors of estimated discharges, $Q_{est}$, compared to maximum posterior (MAP) discharges, $Q_{MAP}$. The green lines represent the calibration gaugings. (right) Rating curve uncertainty results with official rating curve. All discharges are represented in logarithmic scale. The black lines in the left and middle panels represent the boundaries between the low-, medium-, and high-discharge classes. MAP errors could not be estimated for the lowest flow in GS11 because at this stage the MAP sample gave positive water-surface slopes. The gauging scenarios (GS) are explained in Table 2.
flows. This means that a rating curve can be established rapidly as high-flow discharges (that occur more rarely) are not necessarily needed for constraining the rating curve uncertainty.

3.3. Sensitivity Analysis: Which Parameters Require Well-Defined Prior Information?

It was possible to infer the roughness height parameter, \( z_0 \), from the calibration data set when an informative prior was used for either one of the water-surface slope parameters, \( S_1 \) or \( S_2 \) (Figure 8a). When noninformative priors were used for both \( S_1 \) and \( S_2 \), this resulted in a wide and shifted posterior density distribution for the roughness parameter. However, in this latter case (noninformative \( S_1 \) and \( S_2 \)) the prior information on the roughness parameter can still help to reduce correlation effects with the other parameters and also constrain identification to a physically meaningful range.

For the water-surface slope parameters \( S_1 \) and \( S_2 \), an informative prior on at least one of these parameters was clearly required, with only slightly wider parameter distributions, in this case in the lower ranges for \( S_1 \) and \( S_2 \) and in the upper range for \( z_0 \), compared to when informative priors were used for both parameters (Figures 8b and 8c). When informative priors on the water-surface slope parameters were not provided, the posterior densities were both shifted and less precise compared to when using at least one informative prior. These results show that some information on the water-surface slope (e.g., direct field measurements or remotely sensed observations) is required for constraining rating curve uncertainty in the RUHM framework.

Figure 8. Prior sensitivity analysis applied to the Röån River at Röån: Posterior and prior marginal densities of hydraulically modeled rating curve parameters for fully or partially informative priors: (a) The roughness parameter \( z_0 \) and the (b) first and (c) second water-surface slope parameters \( S_1 \) and \( S_2 \), respectively. Note that these plots show the posterior density and not the likelihood (to be able to compare to the prior distributions). Note also that the maximum posterior (MAP) is not the same as the mode of the marginal posterior probability density function in the figure because of the smoothing from the kernel density estimator (the mode occurs in a very narrow region).
3.4. Propagation of Rating Curve Uncertainty to Streamflow Uncertainty

In Figure 9 we illustrate how the rating curve uncertainty propagates to streamflow uncertainty for the GS10 scenario (i.e., calibration to the three gaugings that have the highest likelihood to be gauged) for the two stations. There is a significant reduction of streamflow uncertainty compared to the prior parameter distribution for the calibrated model results based on these three low- and medium-flow gaugings. The calibrated streamflow time series has a narrow uncertainty and closely matches the official Swedish Meteorological and Hydrological Institute rating curve at Röån. At Nybro there is also a narrow uncertainty but a larger deviation at high flows compared to the official rating curve. If these sites had been truly ungauged, the calibration to these three high-probability gaugings, which could straightforwardly be gauged within a year, would therefore have provided an excellent estimate of the rating curve at Röån. At Nybro there are good results for lower flows but biases at high flows because of the backwater influence. Obtaining good-quality streamflow time series data within a year’s monitoring for a previously ungauged site would enable more robust results and potentially lower costs in many types of research and water management projects reliant on streamflow data (McMillan et al., 2018).

4. Discussion

4.1. The Use of Hydraulically Modeled Rating Curves for Rapid Rating Curve Assessment

This work provides a first study of how uncertainties in both measurement data and model parameters affect rating curve and streamflow estimation with hydraulic models. We found that when using an appropriate...
4.2. Modeling Framework, Method Limitations, and Further Developments

4.2.1. Applicability to Different River Conditions

We applied the RUHM framework to two relatively straight uniform channels, and it should be applicable for sites with similar river characteristics (uniform bed material, fairly straight channel without flow obstructions, and relatively low-density vegetation). The Nybro site presented more challenging conditions as nonuniform flow was observed at high flows due to backwater effects and there was scatter in the high-flow gaugings. Future work could investigate the applicability and development of the framework across a wider range of site conditions and stage-discharge behavior such as overbank flow in floodways with denser vegetation for which the flow may no longer stay uniform (e.g., Kean & Smith, 2005). Nonuniform flow would be an important source of uncertainty to consider in the framework when dealing with, for example, turbulent flows or hysteresis due to overbank flow in compound channels with different roughness (Sellin, 1964; Smart, 1992). Constraining rating curve uncertainty with this framework at low and medium flows could, for example, be more challenging on steeper and/or smaller rivers with large rocks in the channel so that turbulent flow occurs and the model assumptions start to be violated. More precise topography surveys, a more complex water-surface slope function or hydraulic model, and/or more calibration gaugings could be necessary to ensure reliable rating curve estimations in such nonideal flow situations. The water-surface slope function considered in this current study was developed as a linear relation with stage, meaning that only two parameters are needed for the modeling. While this simplistic representation facilitated rapid implementation of the modeling for our analysis and limits the number of inferred parameters, it could potentially be improved to take into account more complex water-surface slope-stage relations (e.g., non-linear relations) or nonuniform water-surface profiles (such as backwater effects). In mobile-bed rivers where erosion and sedimentation lead to frequent rating curve changes, new methods of monitoring bed level changes (DeWeese et al., 2017) offer possibilities to include such effects directly in rating curve estimation via hydraulic modeling.

4.2.2. Framework for Uncertainty Estimation

Many sources of error affect rating curve uncertainty estimation regardless of the modeling approach, and for these we used typical assumptions of uncertainty estimation techniques for traditional rating curves. We refer to Kiang et al. (2018) for a full discussion of these issues and focus the discussion on the error sources specific to hydraulic modeling and our framework. We explicitly accounted for uncertainties in the hydraulic model parameters and in the discharge gauging data by using independent distributions. This enabled us to include prior information about the uncertainty in each of these sources from
knowledge about site configuration, survey, and measurement accuracies. The prior parameter distributions were assumed independent, but strong collinearity was observed between the posterior parameter distributions. This likely comes from the linear slope-stage model and the physical relation between roughness and water-surface slope. It may improve parameter identification if the framework was further developed to account for the dependency of these two parameters, either in the parameterization of the inference or in the water-surface slope model. However, it would mean that other prior information sources (e.g., additional water-surface slope surveys) and/or more calibration data are needed, which makes the framework less straightforward to apply with limited prior information/data.

Multimodality in the posterior parameter densities were encountered at Röån for a few of the simulations having a low number of gaugings (e.g., Figure 4). The multimodality mainly stems from the low amount of information supplied to the model in the calibration, that is, using only a few calibration gaugings in combination with relatively wide priors. We tested and saw that these effects were reduced or disappeared when providing more information (i.e., more gaugings or narrower priors). The effect of using more gaugings is shown in the paper; however, it was not motivated to use narrower priors in our analyses because the priors should reflect the available field knowledge (therefore, these results are not shown). In applications where multimodal posterior parameter densities are observed, this shows that more gaugings or better prior parameter information would be useful to reduce uncertainty. In such cases, the parameter estimation could also be improved in the future by using specifically designed MCMC samplers (such as parallel tempering techniques; e.g., Ballnus et al., 2017).

An uncertainty source that was not considered in the framework is the topography data. Casas et al. (2006) and Legleiter et al. (2011) show that accurate topography measurements of the river shape are required for constraining hydraulically modeled rating curves. In contrast, Lyon et al. (2015) found an acceptable level of error when using coarse-resolution topography data instead of high-resolution data. We investigated the impact of the topography surface grid on the rating curve results at Nybro where higher-resolution topography data were available. We found that using only one cross section every 3.2 m was enough to ensure accuracy in the rating curve results (original spacing 0.2 m). More spaced out cross sections along the channel resulted in similar results at high flows but yielded biased results at low and medium flows. A combined precise topography located around the reference cross section with more spaced out cross sections upstream and downstream of the reference cross section also gave promising results. This combination could thus potentially be a good compromise setup for assuring accuracy while lowering computing time (which increased from 3 days to 1.5 months when using the full cross-section spacing of 0.2 m as opposed to 3.2 m, that is, 1,450 instead of 92 cross sections). Uncertainty in the river topography could be explicitly included in the RUHM framework in future versions.

Besides the accuracy of the topography, we found that the number of gaugings and the number of MCMC samples linearly increased the computing time. For most applications, the few days of computing time it takes to estimate a rating curve is irrelevant compared to the many years it can take to obtain a reliable rating curve with traditional approaches.

Several of the evaluation observations were outside the uncertainty intervals, including for the full-gauging scenario (GS9). However, this is expected as the Bayesian error model only accounts for data and parameter uncertainty and not model structural uncertainty (as there are too few data for error model parameter estimation). The uncertainty bounds from our framework should therefore not be expected to give a full representation of rating curve uncertainty but rather a useful first estimate that can later be improved using more gaugings. Lastly, it should be noted that when using only a few gaugings for calibration, the results are more sensitive to the quality of the individual gaugings; using poor-quality gaugings would significantly increase the posterior uncertainty and could bias the results.

### 4.3. Implications for Water Management and Hydrological Studies

The core idea behind the uncertainty framework developed in this paper is to be able to rapidly establish rating curves at new locations or existing stations with changes in the hydraulic control, with an acceptable level of accuracy and an estimate of uncertainty. Quickly obtaining reliable streamflow data for new sites has great value to both long- and short-term research projects (Birgand et al., 2013): It enables robust scientific conclusions to be drawn about process understanding and model representations of hydrological systems.
Acknowledgments
This work was financially supported by the Swedish Research Council Formas (Svenska Forskningsrådet Formas; Project Number 942-2015-321). We acknowledge Dmitri Kavetski (University of Adelaide, Australia) for the use of his numerical Fortran library (DMSL). We also acknowledge Keith Beven and Jason W. Kean for earlier discussions on this work. We thank the SMHI (Swedish Meteorological and Hydrological Institute) for sharing the gauging data sets and stage records for the Nybro and Röån stations. Of the Swedish Research Council Formas (Svenska Forskningsrådet Formas; Project Number 942-2015-321). We acknowledge Dmitri Kavetski (University of Adelaide, Australia) for the use of his numerical Fortran library (DMSL). We also acknowledge Keith Beven and Jason W. Kean for earlier discussions on this work. We thank the SMHI (Swedish Meteorological and Hydrological Institute) for sharing the gauging data sets and stage records for the Nybro and Röån stations. Official daily discharge can be directly downloaded from the SMHI's website (https://vattenwebb.smhi.se/station/). Stage records and gauging data set are available upon request (smhi@smhi.se). Data and rating curve results from this study can be downloaded from the Bolin Centre Database (https://bolin. su.se/data/mansanarez-2019). The RUHM framework executable is available for research purposes upon request (ida.westerberg@ivl.se).

References
Ballin, B., Hug, S., Hatz, K., Görlitz, L., Hasenauer, J., & Theis, F. J. (2017). Comprehensive benchmarking of Markov chain Monte Carlo methods for dynamical systems. BMC Systems Biology, 11(1), 63. https://doi.org/10.1186/s12918-017-0433-1
Birgand, F., Lellouche, G., & Appelboom, T. W. (2013). Measuring flow in non-ideal conditions for short-term projects: Uncertainties associated with the use of stage discharge rating curves. Journal of Hydrology, 503, 186–195. https://doi.org/10.1016/j.jhydrol.2013.09.007
Bullard, J. E., McIntosh, G. H., & Martin, P. (2007). Establishing stage–discharge relationships in multiple–channelled, ephemeral rivers: A case study of the Diamantina River, Australia. Geographical Research, 45(3), 233–245. https://doi.org/10.1111/j.1745-5871.2007.00457.x
Casas, A., Benito, G., Thordendruff, V. R., & Rico, M. (2006). The topographic data source of digital terrain models as a key element in the accuracy of hydraulic flood modelling. Earth Surface Processes and Landforms, 31(4), 444–456. https://doi.org/10.1002/esp.1178
Ceola, S., Montanari, A., Krueger, T., Dyer, F., Kreibich, H., Westerberg, I. K., et al. (2016). Adaptation of water resources systems to changing society and environment—A statement by the International Association of Hydrological Sciences. Hydrological Sciences Journal, 61(16), 2803–2817. https://doi.org/10.1080/02626667.2016.1230674
Clayton, J. A., & Weis, J. W. (2010). Establishing a multi-scale stream gaging network in the Whitewater River Basin, Kansas, USA. Water Resources Management, 24(13), 3641–3664. https://doi.org/10.1007/s11269-010-9624-x
Conn, G., Free, 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

Water Resources Research 10.1029/2018WR024176
Montanari, A., Young, G., Savenije, H. H. G., Hughes, D., Wagener, T., Ren, L. L., et al. (2013). “Panta Rhei—Everything flows”: Change in hydrology and society—The IAHS scientific decade 2013–2022. Hydrological Sciences Journal, 58(6), 1256-1275. https://doi.org/10.1080/02626667.2013.809088

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, 173–186. https://doi.org/10.1016/j.jhydrol.2014.05.007

Nathanson, M., Kean, J. W., Grabs, T. J., Seibert, J., Laudon, H., & Lyon, S. W. (2012). Modelling rating curves using remotely sensed LiDAR data. Hydrological Processes, 26(9), 1427–1434. https://doi.org/10.1002/hyp.8225

Patankar, S. V. (1980). Numerical heat transfer and fluid flow. 197 pp. Washington, D.C.: Hemisphere.

Rantz, S. E. (1982). Measurement and computation of stream discharge discontinuity in composite channel flows. Journal of Hydrologic Research, 18(1), 295–302. https://doi.org/10.1080/02626667.1982.11927112

Reitan, T., & Petersen-Overlie, 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

Schmidt, A. R. (2002). Analysis of stage-discharge relations for open-channel flows and their associated uncertainties (Doctoral dissertation). Retrieved from IDEALS (http://hdl.handle.net/2142/83191), University of Illinois, Urbana–Champaign, USA.

Seibert, J., & Beven, K. J. (2009). Gauging the ungauged basin: how many discharge measurements are needed? Hydrology and Earth System Sciences, 13(6), 883–892. https://doi.org/10.5194/hess-13-883-2009

Sellin, R. H. J. (1964). A laboratory investigation into the interaction between the flow in the channel of a river and that over its flood plain. La Houille Blanche, 7, 793–802. https://doi.org/10.1051/lhb/1964044

Sikorska, A. E., Scheidegger, A., Banasik, K., & Rieckermann, J. (2013). Considering rating curve uncertainty in water level predictions. Hydrology and Earth System Sciences, 17(11), 2955–2986. https://doi.org/10.5194/hess-17-4415-2013

Smart, G. M. (1992). Stage-discharge discontinuity in composite flood channels. Journal of Hydraulic Research, 30(6), 817–833. https://doi.org/10.1080/00221689209489912

Steinbak, G. H., Thorarinsson, T. L., Reitan, T., Schlichting, L., & Helleland, & Engeland, K. (2016). Propagation of rating curve uncertainty in design flood estimation. Water Resources Research, 52, 6897–6915. https://doi.org/10.1002/2015WR018516

Tomkins, K. M. (2014). Uncertainty in streamflow rating curves: Methods, controls and consequences. Hydrological Processes, 28(3), 464–481. https://doi.org/10.1002/hyp.9567

Westerberg, I. K., Wagener, T., Coxon, G., McMillan, H. K., Castellarin, A., Montanari, A., & Freer, J. (2016). Uncertainty in hydrological signatures for gauged and ungauged catchments. Water Resources Research, 52, 1847–1865. https://doi.org/10.1002/2015WR017635

Whiting, P. J., & Dietrich, W. E. (1990). Boundary shear stress and roughness over mobile alluvial beds. Journal of Hydraulic Engineering, 116(12), 1495-1511. https://doi.org/10.1061/(ASCE)0733-9429(1990)116:12(1495)

Wilby, R. L., Clifford, N. J., De Luca, P., Harrigan, S., Hillier, J. K., Hodgkins, R., et al. (2017). The ‘dirty dozen’ of freshwater science: Detecting then reconciling hydrological data biases and errors. WIREs Water, 4(3), e1209. https://doi.org/10.1002/wat2.1209

Wobus, C. W., Kean, J. W., Tucker, G. E., & Anderson, R. S. (2008). Modeling the evolution of channel shape: Balancing computational efficiency with hydraulic fidelity. Journal of Geological Research, 113, F02004. https://doi.org/10.1029/2007JF000914

World Meteorological Organization (2010a). Manual on stream gauging. Fieldwork, WMO-No. 1044, 1, 260.

World Meteorological Organization (2010b). Manual on stream gauging. Computation of discharge, WMO-No. 1044, vol. II, 195 pp.

Erratum

In the originally published version of this article, in section 2.2.4, the sentence, “We used Metropolis/Metropolis sampler from Renard et al. (2006)” was corrected to “We used an adaptive Gibbs/Metropolis sampler from Renard et al. (2006).” The sentence has been updated in text and this version may be considered the authoritative version of record.