Considering rating curve uncertainty in water level predictions

A. E. Sikorska[1,2], **A. Scheidegger**[1], **K. Banasik**[2], **and J. Rieckermann**[1]

[1]Swiss Federal Institute of Aquatic Science and Technology (Eawag), Dept. of Urban Water Management,
Dübendorf, Switzerland
[2]Warsaw University of Life Sciences – SGGW, Dept. of Hydraulic Engineering, Warsaw, Poland

Correspondence to: A. E. Sikorska (as@annasikorska.eu)

Abstract. Streamflow cannot be measured directly and is typically derived with a rating curve model. Unfortunately, this causes uncertainties in the streamflow data and also influences the calibration of rainfall-runoff models if they are conditioned on such data. However, it is currently unknown to what extent these uncertainties propagate to rainfall-runoff predictions. This study therefore presents a quantitative approach to rigorously consider the impact of the rating curve on the prediction uncertainty of water levels. The uncertainty analysis is performed within a formal Bayesian framework and the contributions of rating curve versus rainfall-runoff model parameters to the total predictive uncertainty are addressed. A major benefit of the approach is its independence from the applied rainfall-runoff model and rating curve. In addition, it only requires already existing hydrometric data. The approach was successfully demonstrated on a small catchment in Poland, where a dedicated monitoring campaign was performed in 2011. The results of our case study indicate that the uncertainty in calibration data derived by the rating curve method may be of the same relevance as rainfall-runoff model parameters themselves. A conceptual limitation of the approach presented is that it is limited to water level predictions. Nevertheless, regarding flood level predictions, the Bayesian framework seems very promising because it (i) enables the modeler to incorporate informal knowledge from easily accessible information and (ii) better assesses the individual error contributions. Especially the latter is important to improve the predictive capability of hydrological models.

1 Introduction

Rational flood hazard management not only requires predictions of peak flows and associated inundation and water levels, but also information on their uncertainty (Montanari, 2007; Ramos et al., 2010). In hydrological flood forecasting, the problem of quantitative assessment and reduction of predictive uncertainties has been widely recognized (Renard et al., 2010; Wagener and Montanari, 2011). Recently, a few frameworks have been proposed to assess the total predictive uncertainty (e.g. Ajami et al., 2007; Beck, 1991; Del Giudice et al., 2013; Deltic et al., 2012; Kavetski et al., 2006; Montanari and Koutsoyiannis, 2012; Reichert and Mieleitner, 2009; Renard et al., 2011; Yang et al., 2007; Vrugt et al., 2008). Typically, the accuracy of model predictions and uncertainty estimates need to be assessed against calibration or validation data, which both may be uncertain due to measurement errors. However, it is not clear to what extent uncertainties in calibration data have an impact on the reliability of flood predictions (Domeneghetti et al., 2012).

Usually, calibration data refer to streamflows for rainfall-runoff (RR) models. Typically, the influence of the uncertainty in streamflow data for RR models on the predictive uncertainty is hardly assessed quantitatively by hydrologists in scientific literature. One reason for this is that streamflow is usually not measured directly but must be derived from other, directly measurable quantities such as a water level and velocity with the help of another model. Another is that modellers often only work with the derived quantities, such as streamflow, and not with the raw data. Implicitly, the uncertainty in calibration data themselves is assumed to be much smaller than that from imperfect rainfall information and to be therefore negligibly small (Di Baldassarre and Claps,

2011; Di Baldassarre and Montanari, 2009). For example, the World Meteorological Organisation (WMO, 2008) suggests that streamflow measurement errors of 5 % may be assumed. This, however, can only be valid when streamflow measurements are of a high quality, e.g. when derived by flow meters with the area-velocity method. This method links streamflow to a cross sectional area and a flow velocity, which are often obtained from manual measurements on a dense grid for each cross section of interest. It is probably the most widely used approach to derive streamflows when those are not directly measured (Di Baldassarre and Montanari, 2009; WMO, 2008). Unfortunately, the area-velocity method is impracticable in the field to obtain continuous or frequent measurements. The method becomes cost-inefficient and time consuming when numerous records are required because the grid measurements are labour-intensive. Therefore, streamflow is usually computed from water level only as that is simple to measure and has a small measurement error of 1–2 cm (WMO, 2008). Usually this water level-streamflow relation is modelled with a rating curve (RC) that is calibrated for a certain cross section on area-velocity measurements (Le Coz, 2012). The water level-streamflow relation could also be represented by a more sophisticated model such as a numerical hydraulic model (Di Baldassarre and Claps, 2011). Unfortunately, hydraulic models are less practical than rating curves because they require detailed data on the river channel properties, which are more often than not unavailable.

RC models, however, should not be considered error-free for several reasons. First, the RC is based on streamflow data that are uncertain as they are calculated with a model (e.g. area-velocity method). Second, the uncertainty of the RC is caused by parameter uncertainty and structural limitations of the RC. Third, temporary hydrological conditions such as a seasonal variation of vegetation within the cross section, stream bed dynamics, debris, ice jams in winter, unsteady flow conditions and the hysteresis effect add uncertainty to the calculated streamflows (e.g. Di Baldassarre et al., 2012). Fourth, in most situations the calibration data for rating curves is limited to normal conditions while flood forecasts usually focus on extreme events. This means that the RC must be extrapolated outside of the observed range (Pappenberger et al., 2006). Consequently, all these factors may introduce a large degree of uncertainty into the streamflow predictions. Unfortunately, although a number of recent publications have studied rating curves uncertainties, the contribution of the rating curve to prediction uncertainties has not yet been investigated systematically.

For instance, Di Baldassarre and Montanari (2009) investigated uncertainty present in river flow records when derived with the rating curve method and concluded that those may include errors of up to 25 % of estimates in the extrapolation range. Moreover, Domeneghetti et al. (2012) showed that those extrapolation errors dominate over all other sources of uncertainty in rating curves. The usage of rating curves to derive streamflows in the extrapolation range was further

investigated by Di Baldassarre and Claps (2011), who recommended using a numerical hydraulic approach to derive water level-streamflow curves for cross sections instead of a traditional extrapolation method. The main drawback of this approach is that it requires detailed data on the topology and input. In addition, it is also not free from errors due to (i) structural limitations of the hydraulic model, (ii) uncertainty about its parameters and (iii) measurement errors (Di Baldassarre et al., 2012). While exploring the uncertainty of calibration data is an ongoing issue, considerable progress has been made in the uncertainty assessment of hydrological predictive models. The main challenge lies in investigating the relevance of individual uncertainty sources. In this regard, several methods have been proposed to separate total prediction uncertainty into the individual contributions from (i) input uncertainty e.g. due to poor rainfall data (Kavetski et al., 2006; McMillan et al., 2011; Sikorska et al., 2012), (ii) model structure deficits (Reichert and Mieleitner, 2009; Renard et al., 2011), (iii) parameter uncertainty (Ajami et al., 2007; Vrugt et al., 2008) and (iv) measurement errors (Di Baldassarre and Montanari, 2009; McMillan et al., 2010). All of these studies, however, focus either on the analysis of the predictive uncertainty of hydrological models with rather crude assumptions on the uncertainty in calibration data or on the uncertainty of calibration data alone without considering the resulting uncertainty in hydrological predictions. However, a systematic approach to integrate both the rainfall-runoff and rating curve models is currently lacking. The only attempt towards integration was undertaken by McMillan et al. (2010), who investigated the impact of errors in streamflow measurements on the streamflow predictions informally. For the applied resampling approach to construct uncertainty intervals of the RC model no formal justification is provided. The subsequent calibration of the RR model is based on an informal likelihood with the help of Monte Carlo Markov chain algorithms. A consequence of the application of an informal likelihood function is that the resulting prediction uncertainty is mapped on the RR model parameters entirely. This prohibits an assessment of the importance of the different sources of uncertainty, such as rating curve and rainfall-runoff model parameters.

In this manuscript, we propose a formal Bayesian approach to quantify the uncertainty in hydrological predictions of water levels by means of an integrated assessment of a rainfall-runoff (RR) model and the corresponding rating curve (RC). This enables the derivation of the predictive distribution of water levels and to simultaneously assess the uncertainty contribution of the RC to the total predictive uncertainty. For the first time, we compare the contribution of the RC to those of the parameters of the RR model formally. The ability to predict water level for given rainfall is important for design studies and risk assessments. Other methods that rely water level forecast on the known water level or streamflow (e.g. Coccia and Todini, 2011) are dedicated for operational forecasting and are not considered in this paper.

The proposed approach is readily applicable because no additional hydrometric data on a rating curve than those already existing are required. Due to the low data demand and the possibility to use informative prior distributions, it can also be applied in poorly gauged catchments.

The remainder of the manuscript is as follows. First, we demonstrate analytically what implicit assumptions are made by ignoring the RC uncertainty. Second, we detail our approach for the joint Bayesian analysis of the RC and the RR model uncertainty. Third, we demonstrate its feasibility in a case study on a small catchment in Poland (Warsaw, Sluzew Creek). Fourth, we discuss the strength and limitations of our approach and derive practical recommendations as well as directions for future research. Finally, we draw our main conclusions.

2 Methods

2.1 Water level-runoff model

Unfortunately, in many hydrological applications streamflow is not directly measurable, although reliable flood modelling and prediction ideally requires continuous or frequent measurements. Therefore, streamflow has to be derived from other, measurable variables by means of a model that converts them into streamflow. Usually, this is done with a water level-runoff model (LR) that relates the streamflow Q_t to the water level L_t:

$$Q_t = \mathrm{LR}(L_t, \boldsymbol{\theta}_{\mathrm{LR}}) + E_t^{\mathrm{LR}}, \qquad (1)$$

where $\boldsymbol{\theta}_{\mathrm{LR}}$ is the parameter vector of the LR and E_t^{LR} is an error term.

Typically, the LR model (Eq. 1) has empirical parameters that can only be calibrated if some simultaneous observations of L_t and Q_t are available; L_t is measured directly and Q_t usually indirectly by means of the area-velocity method (WMO, 2008). Thereby, uncertainties in Q_t are about 3–6 % on average but may increase to about 20 % under poor measurement conditions (Sauer and Meyer, 1992). The error term E_t^{LR} therefore represents uncertainties due to the computation of Q_t and due to structural limitations of the LR model. These are always present, if only due to the hysteresis effect, where the same L_t can be observed for different Q_t at the rising and the falling limb of a flood hydrograph.

2.2 Rainfall-runoff modelling

Rainfall-runoff models (RR) predict the streamflow Q_t based on input information $X_{1:t}$ that typically contains at least mean areal precipitation within the catchment. Every RR model can be written as

$$Q_t = \mathrm{RR}(X_{1:t}, \boldsymbol{\theta}_{\mathrm{RR}}) + E_t^{\mathrm{RR}}, \qquad (2)$$

where $\boldsymbol{\theta}_{\mathrm{RR}}$ is the parameter vector of the RR model and E_t^{RR} is an error term. In contrast to common practice, E_t^{RR} does not necessarily need to have an expected value of zero (Reichert and Mieleitner, 2009). Therefore, E_t^{RR} here represents structural deficits of the RR model and all other uncertainty not explicitly accounted for as input uncertainty. Usually, RR models are calibrated against "measured" streamflow (McMillan et al., 2010; Wagener and Montanari, 2011). The calibration, however, is complicated by the fact that the output of the RR model (Q_t) cannot be measured directly.

2.3 Standard rainfall-runoff model calibration procedure

RR models are typically calibrated in four steps:

1. The water level is measured and the corresponding streamflow is computed for few conditions, e.g. with the area-velocity approach described in Sect. 2.1.

2. Based on these data, the water level-runoff (LR) model is then calibrated. Mostly, a model according to the Eq. (1) is used and normally distributed errors with zero mean are assumed.

3. Streamflow data $\hat{Q}_t$ are calculated from the water level L_t using the previously calibrated LR at the best parameter estimates while neglecting the error of the LR.

4. The RR model is calibrated to match the computed streamflows $\hat{Q}_t$. This can be formalized as

$$\hat{Q}_t = \mathrm{LR}(L_t, \hat{\boldsymbol{\theta}}_{\mathrm{LR}}) = \mathrm{RR}(X_{1:t}, \boldsymbol{\theta}_{\mathrm{RR}}) + E_t^{\mathrm{RR}}, \qquad (3)$$

where $\hat{\boldsymbol{\theta}}_{\mathrm{LR}}$ is the parameter vector that led to the best fit of the LR at step 2.

This procedure might be suitable if only the "best fitting" parameters are of interest. However, for predictive uncertainty analysis it has two conceptual flaws. First, the error term of the LR model is "lost" in the third step, which can be seen by comparing Eqs. (1) and (2):

$$Q_t = \mathrm{LR}(L_t, \boldsymbol{\theta}_{\mathrm{LR}}) + E_t^{\mathrm{LR}} = \mathrm{RR}(X_{1:t}, \boldsymbol{\theta}_{\mathrm{RR}}) + E_t^{\mathrm{RR}}.$$

It is important to realize that the remaining error term in Eq. (3) cannot not include the LR uncertainty, as the RR model is calibrated against the *average* value of Q for a given water level. Therefore, E_t^{LR} is never "seen" by the RR model.

Second, the uncertainty of the estimated parameters $\hat{\boldsymbol{\theta}}_{\mathrm{LR}}$ is neglected. Unfortunately, both flaws might lead to overconfident predictions.

2.4 Modelling water level

The two problems of the RR model calibration procedure presented above can be circumvented by modelling the water level directly. To this end, instead of an RR model, a

rainfall-water level (RL) model is formulated that relates the inputs $X_{1:t}$ to the direct measurable water level $L_{1:t}$. Such a RL model can be constructed by combining the deterministic parts of the RR model (Sect. 2.2) with the inverse of the LR model (Sect. 2.1):

$$L_t = \text{RL}(X_{1:t}, \boldsymbol{\theta}_{\text{RL}}) + E_t^{\text{RL}} = \text{LR}^{-1}(\hat{Q}_t, \boldsymbol{\theta}_{\text{LR}}) + E_t^{\text{RL}} =$$
$$\text{LR}^{-1}(\text{RR}(X_{1:t}, \boldsymbol{\theta}_{\text{RR}}), \boldsymbol{\theta}_{\text{LR}}) + E_t^{\text{RL}}, \quad (4)$$

where the parameters of the RL model are denoted by $\boldsymbol{\theta}_{\text{RL}}$. E_t^{RL} represents an error term and now includes structural deficits of both, LR and RR submodels, as well as (a presumably small) measurement error of the water level. Additionally, E_t^{RL} will also compensate for all other uncertainty contributions that are not explicitly accounted for here. Indirectly, input uncertainty is also represented in E_t^{RL}.

2.5 Inference and predictive distribution

2.5.1 Likelihood function

Statistical parameter estimation uses a likelihood function, which requires assumptions of the error term E_t^{RL}. As it is well known that the residuals of hydrological models are heavily auto-correlated (e.g. Sikorska et al., 2012; Yang et al., 2007), we assume a continuous error process that is equivalent to a first order autoregressive process and combine it with a Box-Cox transformation as proposed by Yang et al. (2007). The advantage of such an error model, in contrast to the traditional Gaussian approach, which assumes independent and identically distributed errors with a zero mean, is that it helps to meet the underlying statistical assumptions with regard to temporal autocorrelation and heteroscedasticity (Sikorska, 2012; Sikorska et al., 2012; Yang et al., 2008). This lumped error model describes all error sources that are not explicitly acknowledged, here mainly input and structural uncertainty of the model. The error model parameters ($\boldsymbol{\theta}_{E^{\text{RL}}}$) are the asymptotic standard deviation (ERL1) and characteristic correlation time (ERL2) of the autoregressive process. For details we refer to Sikorska et al. (2012).

The proposed likelihood function has a frequentist interpretation and therefore a maximum likelihood estimation would be possible. However, RR models are usually overparameterized with correlated parameters (Beck, 1991; Wagener et al., 2004). For these reasons, the Bayesian inference is more suitable for hydrological models as it allows prior knowledge to be incorporated in the calibration process by means of a prior probability distribution of the model parameters. Thus, identifiability problems in the calibration process are avoided (Gelman et al., 2003). The same holds for the RL model.

2.6 Bayesian inference and predictions

Given a likelihood function $p_{\text{RL}}(L|\boldsymbol{\theta}_{\text{RR}}, \boldsymbol{\theta}_{\text{LR}}, X)$ of the RL model and the data $\{L, X\}$ (where L is the calibration data

and X is the input data), the prior parameter distribution $p(\boldsymbol{\theta}_{\text{RR}}, \boldsymbol{\theta}_{\text{LR}})$ is updated as

$$p(\boldsymbol{\theta}_{\text{RR}}, \boldsymbol{\theta}_{\text{LR}}|L, X) \propto p_{\text{RL}}(L|\boldsymbol{\theta}_{\text{RR}}, \boldsymbol{\theta}_{\text{LR}}, X)p(\boldsymbol{\theta}_{\text{RR}}, \boldsymbol{\theta}_{\text{LR}}). \quad (5)$$

The knowledge about the future realization of the water level L^p conditioned on past calibration and input data $\{L, X\}$ and future input X^p is described by

$$p(L^p|L, X, X^p) = \int\int p_{\text{RL}}(L^p|\boldsymbol{\theta}_{\text{RR}}, \boldsymbol{\theta}_{\text{LR}}, X^p)\cdot$$
$$p(\boldsymbol{\theta}_{\text{RR}}, \boldsymbol{\theta}_{\text{LR}}|L, X)d\boldsymbol{\theta}_{\text{LR}}d\boldsymbol{\theta}_{\text{RR}}. \quad (6)$$

For almost all models these distributions must be approximated by Markov Chain Monte Carlo methods (see Sect. 2.7).

2.7 Prior distribution of RR and RC model parameters

Several methods are available to define the prior distribution on RR submodel parameters without a need for calibration data. One possibility would be to use methods that derive model parameters from catchment properties, as described by Sikorska et al. (2012). In contrast, defining the prior distribution of the LR parameters, and in particular of the RC method, requires some field observations. An informative prior on the RC parameters can be easily obtained from already existing hydrometric measurements of cross-sectional average velocities and corresponding water levels. Here, we suggest calibrating the RC as in Eq. (1) with the standard maximum likelihood method. The distribution of the parameter estimator can then be derived using large sample size properties of the maximum likelihood estimator (e.g. Harrell, 2010) and can serve as a prior afterwards.

2.8 Cross-validation for predictions and prediction measures

To assess the predictive distribution of water levels, we performed a leave-one-out cross-validation (e.g. Wang and Robertson, 2011). Thereby a single event is randomly selected as a validation data and the remaining events as a model calibration set. The computation was repeated in order to use each event once to validate the model.

The efficiency of model predictions is measured by the Nash–Sutcliffe (NS) index (Nash and Sutcliffe, 1970) estimated for the best model prediction (mode of the posterior). The uncertainty bands are assessed by the data coverage DC (α) (also known as reliability, Del Giudice et al., 2013; Montanari and Koutsoyiannis, 2012; Sikorska, 2013) defined as the fraction of the data covered by the prediction interval given by the $\alpha/2$-quantile and the $(1-\alpha/2)$-quantile. Theoretically, the DC (α) should be larger or equal to 1-α.

2.9 Influence of the RC on total prediction uncertainty

To assess the influence of the RC on total prediction uncertainty, we compare it to the uncertainty of the RR model

by means of a sensitivity analysis (Table 1). To this end, we compare the total prediction uncertainty to two scenarios where either the RR parameters (A) or the RC parameters (B) are kept constant at the maximum of their posterior marginals. Remaining parameters are sampled from the posterior distribution conditional on the maximal posterior marginals of those parameters that are kept constant in each of two scenarios (RR in A and RC in B). Then, we compare the uncertainty of each scenario to that of the total predictive distribution. The reduction of uncertainty then indicates the relative importance of the RR and RC components. This comparison of prediction uncertainty is preferable to a local sensitivity analysis because it takes into account estimated mutual interactions between the parameters.

2.10 Implementation

The RL model and the inference procedure were implemented in R (R Development Core Team, 2011). The posterior probability distribution was sampled with the adaptive Monte Carlo Markov Chain (MCMC) algorithm proposed by Haario et al. (2001). Specifically, we used the implementation of Chivers (2012) to produce three chains with 100 000 samples each. The number of samples and chains resulted in a reasonable compromise between fully exploring the posterior distribution and fast computations.

The R script is available on request from the corresponding author.

3 Case study

3.1 Test catchment and data

For a case study, we chose the upper part of the Sluzew Creek catchment (Warsaw, Poland), which has an area of about $28.7 \, \text{km}^2$ (A_{red}: $18.3 \, \text{km}^2$) (see Fig. S1 in Supplement). Sluzew Creek is a third-degree watercourse and a tributary of the Wilanowka river, which flows into the Vistula River. The catchment is located in the lowland and is therefore rather flat, with an elevation from 95 m to 110 m above sea level and surface runoff dominated by land use characteristics. The Sluzew Creek catchment has undergone rapid urbanization in the last three decades and today urban areas cover 58.7 % of the catchment. As a consequence, it is strongly affected by floods dominated by torrential rainfalls which mostly occur during spring-summer seasons and in the lower (highly urbanized) part of the catchment (Sikorska and Banasik, 2010; Sikorska et al., 2012). Although the catchment is partly urbanized and a few anthropogenic hydraulic infrastructures are located along the stream, the streamflow is not disturbed during ordinary to middle-high flow conditions (Barszcz, 2009). At the analysed gauging profile an undisturbed streamflow is observed till the water level exceeds 180 cm (see Fig. 1). For more details on the case study, the reader is referred to Banasik et al. (2008) and Sikorska

Table 1. Uncertainty analysis scenarios. θ_{RR} – parameters of the RR submodel, θ_{RC} – parameters of the RC submodel, $\theta_{E^{\text{RL}}}$ – parameters of the RL lumped error model (E^{RL}).

Scenario	θ_{RR}	θ_{RC}	$\theta_{E^{\text{RL}}}$
A	x	o	x
B	o	x	x

Note: x – uncertainty explicitly accounted for, o – uncertainty neglected (parameters kept at the maximum of the posterior).

and Banasik (2010). As for most small catchments, no routine monitoring programme exists. We therefore performed our own monitoring program that consisted of regular measurement of precipitation data (3 locations) and stream water levels at the catchment outlet (Supplement). In addition, we measured a cross-sectional streamflow during a set of field experiments by means of the area-velocity method (WMO, 2008), see Fig. 1.

In total, data on 8 storm events were collected during 2011 which were all used in cross-validation (Sect. 2.6.). An empirical RC was constructed based on 11 water level-streamflow records using a power-law model and the recommended range for its extrapolation was set at 180 cm (see Fig. 1). This water level was not exceeded for any of the analysed storm events.

### 3.2 Water level-runoff submodel

As an LR in Eq. (1), we used a power law equation, which is widely used as a rating curve (Di Baldassarre and Claps, 2011; Domeneghetti et al., 2012). It fits our observations well (Fig. 1):

$$\text{LR}(L_t, \theta_{\text{RC}}) = \text{RC1} \, (L_t - \text{RC2})^{\text{RC3}} \Leftrightarrow \text{LR}^{-1}(Q_t, \theta_{\text{RC}}) =$$
$$\left(\frac{Q_t}{\text{RC1}} \right)^{1/\text{RC3}} + \text{RC2}, \qquad (7)$$

where the parameters of the LR submodel are here combined to $\theta_{\text{RC}} = \{\text{RC1}, \text{RC2}, \text{RC3}\}$. Alternatively, if a single power law equation does not fit the observed data sufficiently well, more sophisticated structure of a RC (Dottori et al., 2009) or a non-stationary RC (Westerberg et al., 2011) could be applied.

3.3 Conceptual rainfall-runoff submodel

We applied a simple, conceptual, event-based RR model that combines the SCS-CN method (Mishra and Singh, 2010) to separate the effective rainfall from the total precipitation with an instantaneous form of unit hydrograph model (IUH) (Khaleghi et al., 2011; Nash, 1957). The parameters of the applied RR model, $\theta_{\text{RR}} = \{\text{RR1}, \text{RR2}, \text{RR3}, \text{RR4}\}$, are: catchment area (RR1), maximal potential retention of a catchment (RR2), retention time of a linear reservoir (RR3)
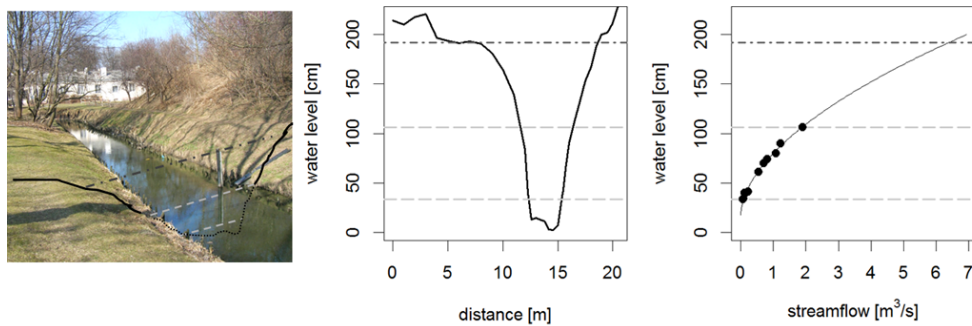
Fig. 1. Monitoring cross section (left and middle) and prior information on the rating curve (right); the reference point for both is a cease-to-streamflow reference level; dashed grey lines depict observation range, upper grey line cuts off a justifiable extrapolation range from the valid bathymetric profile (till flood plains); right figure: black dots illustrate measured water level-streamflow relations, black solid line presents prior mean rating curve.

and the number of identical linear reservoirs (RR4). The RR model based on the SCS-CN method, due to the limited number of parameters, is a common choice to model the direct surface runoff in rainfall-runoff processes in (small) catchments with no long time series available (e.g. Seibert, 1999; Sikorska et al., 2012). Lumped models are justified when the modelling focus lies in the catchment outlet only (Coutu et al., 2012). Such models, due to their simplicity and transparency of modelled patterns, are frequently used as an application example for uncertainty analysis approaches (e.g. Blöschl and Montanari, 2010; Dotto et al., 2012; Sadegh and Vrugt, 2013; Seibert and McDonnell, 2013; Uhlenbrook et al., 1999).

3.4 Formulating prior knowledge on the RR and LR model parameters

The prior distribution for the parameters of the RR and LR submodel has been derived from catchment characteristics as described in Sikorska et al. (2012). For the RC submodel parameters θ_{RC}, the prior was inferred from reference measurements of water level and velocity (Sect. 2.5.3) shown in Fig. 1. To allow for a fair comparison of RC and RR error contributions, we only used data from a relatively short period. This should avoid bias due to seasonal or long-term changes, such as changes in the catchment land use and surface properties or cross-section geometry. The prior for the parameters of the error model ($\theta_{E^{RL}}$) is more difficult to specify because they do not have a direct physical interpretation. To express this lack of knowledge, we selected rather wide distributions (see also Table 2). Correlation between parameters was only considered in the prior for θ_{RC} as their interaction is known from the maximum likelihood estimation (see Sect. 2.5.2). The other parameters were assumed to be independent, which is common practice in uncertainty studies (e.g. Reichert and Schuwirth, 2012; Sikorska et al., 2012).
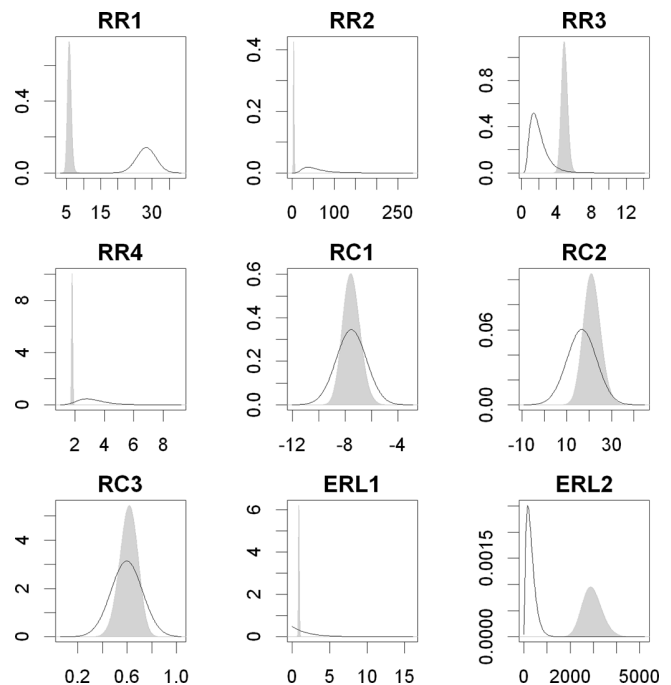
Fig. 2. Prior (solid line) and posterior parameter distribution (grey area); RR parameters (θ_{RR}), RC parameters (θ_{RC}) and E^{RL} lumped error model parameters ($\theta_{E^{RL}}$).

3.5 Results

3.5.1 Results of the statistical inference

The RL model described in Sect. 2.4 was calibrated and validated using the leave-one-out cross-validation method (Sect. 2.6.) with all eight recorded rainfall-water level events. All parameters (θ_{RR}, θ_{RC}, $\theta_{E^{RL}}$) were inferred simultaneously and the posterior parameter distribution was obtained from a calibration where all eight events were used simultaneously. For the parameters of the Box-Cox transformation,

Table 2. Prior distribution derived for the Sluzew Creek. θ_{RR} – parameters of the RR submodel, θ_{RC} – parameters of the RC submodel, $\theta_{E^{RL}}$ – parameters of the RL model error term (E^{RL}).

Model/ parameter/name and meaning		Distribution	Expected value (E) and standard deviation (SD)
Rainfall-water level (RL) model (θ_{RL})			
Rainfall-runoff (RR) submodel (θ_{RR})			
RR1	area of a catchment [km^2]	Normal	$E = 28.3$; SD $= 2.8$
RR2	max potential retention of a catchment [mm]	Log-normal	$E = 55$; SD $= 33$
RR3	retention time of a linear reservoir [h]	Log-normal	$E = 2.0$; SD $= 1.0$
RR4	number of linear reservoirs [–]	Log-normal	$E = 3.2$; SD $= 1.0$
Rating curve (RC) submodel (θ_{RC})			
RC1	coefficient, or streamflow scale [–]		$E = -7.5$; SD $= 1.1$
RC2	location parameter, or cease to streamflow-water level, in units of the water level, e.g. [cm]	Multivariate normal	$E = 16.8$; SD $= 6.6$
RC3	exponent, linked to the type and shape of the hydraulic control [–]		$E = 0.6$; SD $= 0.1$
Lumped error model (E^{RL}) ($\theta_{E^{RL}}$)			
ERL1	asymptotic standard deviation of the errors [cm]	Gamma	$E = 2$; SD $= 2$
ERL2	characteristic correlation time of the autoregressive process [min]	Gamma	$E = 300$; SD $= 200$
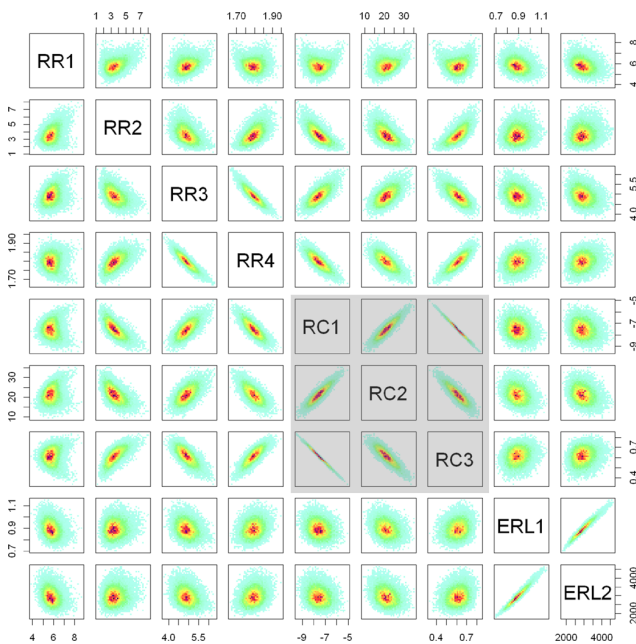
Fig. 3. Posterior parameters correlations; RR1:RR4 – RR parameters (θ_{RR}), RC1:RC3 – RC parameters (θ_{RC}) and ERL1:ERL2 – E^{RL} lumped error model parameters ($\theta_{E^{RL}}$).

we used $\lambda_1 = 0.5$ and $\lambda_2 = 0$, which proved to be a good assumption for this catchment (Sikorska et al., 2012).

In general, the marginal posterior distributions of the model parameters (θ_{RR}, θ_{RC}) show a similar shape as the prior but, as expected, have smaller variances (Fig. 2).

The two RR parameters (θ_{RR}) indicate that the average rainfall-runoff process in Sluzew Creek is described by about 1.8 reservoirs (RR4) with a relatively short retention time (RR3) of 4.9 h. Two other RR parameters (RR1 and RR2) suggest that, first, during heavy rainfalls only a fraction of the catchment area, which is probably impervious and closely located to the stream, contributes to the surface runoff (RR1). Second, during intensive precipitation the catchment retention is less important for surface runoff (RR2). This is reasonable for a small and urbanized catchment, where the response of the catchment to heavy rainfalls is expected to be rapid. Note, however, that the estimation of the RR model parameters is not the scope of this paper. For more detailed discussions of the case study results, the reader is referred to Sikorska et al. (2012).

The posterior of the RC parameters (θ_{RC}) is very similar to the prior. This was more or less expected because measuring rainfall input and water level output is not the ideal experimental context to learn about the RC parameters. This also emphasizes the importance of obtaining an informative prior distribution as described in Sect. 2.5.3. Additionally, a strong correlation between all RC parameters was observed (see Fig. 3). Moreover, we observed a significant correlation between RR and RC parameters (Fig. 3). Intuitively, this can be explained by a mutual compensation of both submodels.

Finally, for both of the lumped error model parameters ($\theta_{E^{RL}}$) information was gained from the data. However, the interpretation of these parameters is not straightforward since they do not have an obvious physical meaning. This is further discussed below.
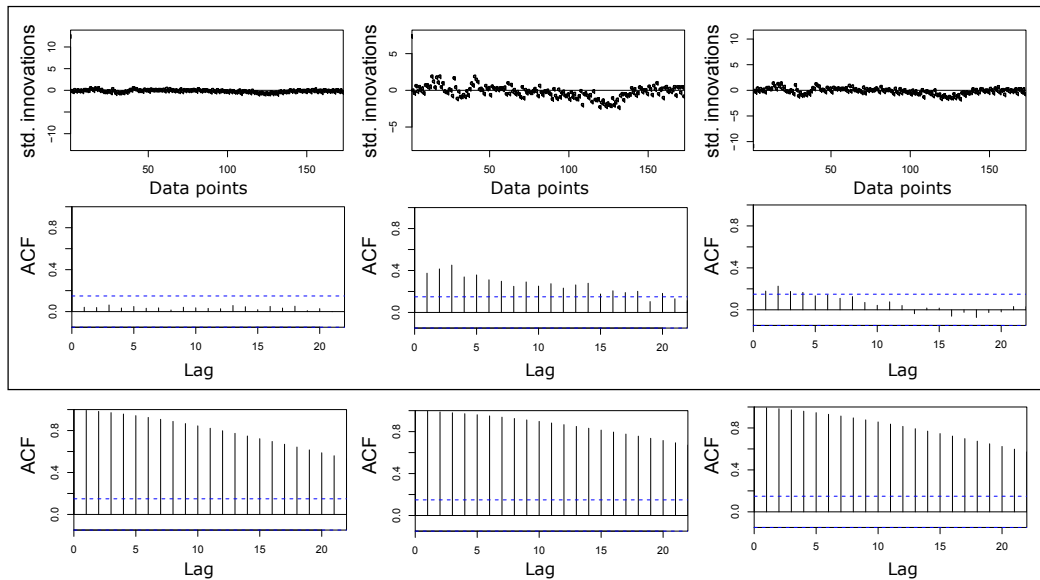
Fig. 4. Diagnostic plot of innovations and residuals for chosen events (No. 5, 6, 7); lumped error model: sequence of innovations (top row) and autocorrelation function (ACF) of standardized innovations (middle row); traditional error model: ACF of standardized residuals (bottom row).

Figure 4 presents a diagnostic analysis of model innovations in the continuous autoregressive error model for selected events. This error model does not assume that the residuals are independent and identically distributed (i.i.d.), which is necessary for the standard Gaussian error model. Instead, the innovations of the stochastic process have to be i.i.d. The autocorrelation function of the innovations is shown in the top row in Fig. 4. While for some events a slight autocorrelation still remains (see also Supplement), the statistical assumptions are much better fulfilled than for the assumption of a standard Gaussian error model (Fig. 4, bottom row). The residuals were computed as a difference between observed and simulated values corresponding to the best model prediction. In addition, the innovations are less heteroscedastic compared to the residuals (Fig. 4, middle row). Details regarding the statistical assumptions are also further discussed below.

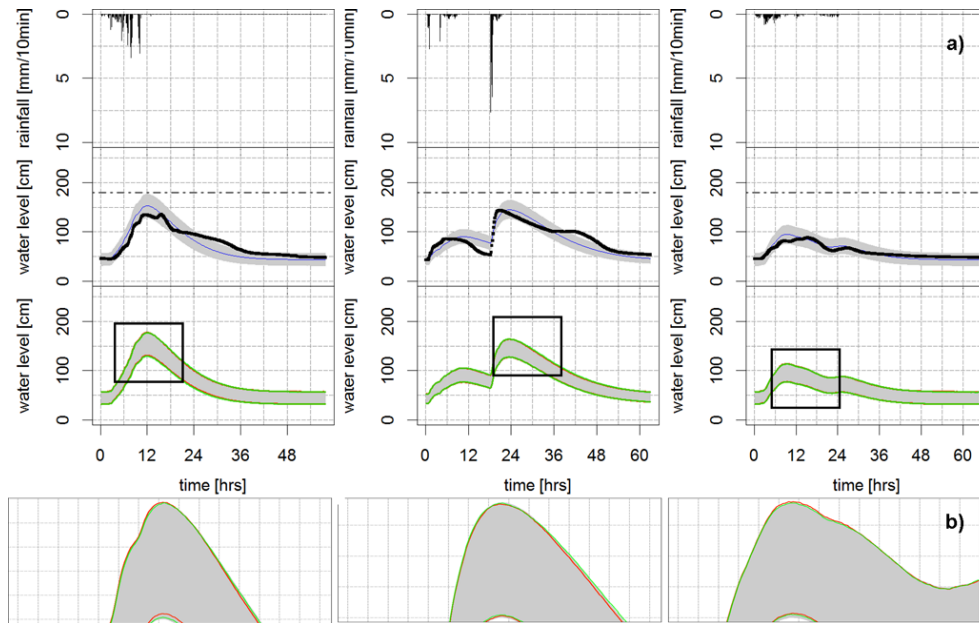### 3.5.2 Total predictive uncertainty and model performance

To approximate the total predictive uncertainty, a Monte Carlo simulation with 100 000 runs was performed, drawing repeatedly from the full posterior parameter distribution obtained from the leave-one-out cross-validation (Sect. 2.6). Therefore, the predictive uncertainty bands for each event are the result from a calibration without this event; eight independent MCMC chains were generated for every calibration set of seven events and validated on the remaining one (see Sect. 2.8). The 2.5 % and 97.5 % quantiles were computed

and the corresponding 95 % predictive uncertainty bands for three events are presented in Fig. 5a (middle row – grey polygons); solid blue lines correspond to the predictions using the mode of the posterior density.

The 95 % total predictive uncertainty bands obtained, when accounting for both RR and RC parameters uncertainty, are on average 15 % higher than peak water levels during rainfall-runoff events. Maximum deviations are up to 50 % higher than the observations.

For all events, the data coverage DC $(1 - 0.95)$ is 0.7 (see Sect. 2.7), whereas 26 % of data points lie above and 4 % below the upper and lower limits, respectively. The uncertainty bands properly cover most of the events, except events 2 and 8, for which larger deviations were obtained (see Supplement). The deviations for event No. 2 can be explained by changing external factors such as additional water discharges from sewage systems or overland flows. In contrast, the deviations for event No. 8 nicely illustrate the consequences of extrapolating the rating curve beyond its justifiable range by the upper limit of the uncertainty interval (Supplement): the approximated uncertainty bands are clearly overestimated. Such a high water level as predicted by the model would most likely not occur in reality because of overland flow outside the flood plains (compare to the Fig. 1). As this process cannot be modelled accurately with the applied RL, data coverage is poor and the prediction uncertainty bands are not reliable (see Discussion). Excluding these two events, DC $(1 - 0.95)$ is 0.86.

Fig. 5. (**a** and **b**) Predicted water levels in the Sluzew Creek using the posterior parameter distribution (events No. 5, 6, 7). (**a**) Total predictive uncertainty. Middle row: observations (dotted black lines) and predicted water levels corresponding to the mode posterior values (solid blue lines). Grey areas present 95 % total prediction uncertainty bands, dashed grey horizontal line cuts an extrapolation range for the RC from Fig. 1. Bottom: influence of the RC parameters. Grey areas describe 95 % total prediction uncertainty bands, red lines illustrate 95 % limits for the predictive uncertainty bands whilst ignoring uncertainty in RR parameters (scenario B), green lines illustrate 95 % limits for the predictive uncertainty bands whilst ignoring uncertainty in RC parameters (scenario A). (**b**) Zoom in of the uncertainty bands in the range of the runoff peaks. Visibly similar contribution of the RR and RC uncertainty.

The obtained average Nash–Sutcliffe index for the best model prediction (mode) during the validation mode was 0.61 over all events and 0.69 when excluding events No. 2 and 8. This can be considered as satisfying for the simple hydrological model applied here.

3.5.3 Influence of the RC parameter

The contribution of the RC model parameter to the total predictive uncertainty is shown in Fig. 5, and is assessed under two scenarios as described in Sect. 2.7 (Table 1).

The corresponding predictive uncertainties were found to be almost of the same relevance for both scenarios A and B (RR vs. RC). A difference in the contributions is less than 1 % (mean) in the validation mode, with a slight dominance of the RC uncertainty (scenario B). This can be visually seen on the bottom of Fig. 5a, where both RR and RC uncertainties intervals lie close to each other and to the total predictive uncertainty limits. This would suggest, first, that the uncertainty in RR and RC parameters leads almost to the same predictive uncertainties of water level in the Sluzew Creek catchment, at least at this monitored cross section. Second, for this particular case study the uncertainty of RC and RR parameters around their mode could also be neglected since both contribute much less to the total predictive uncertainty than the uncertainty of the runoff-water level model structure

alone (bottom row of Fig. 5a and b). This, however, is not transferrable to other case studies, and a previous estimation of the importance of the parameter uncertainties is difficult. Therefore, the analysis which we suggest should be repeated for the case study of interest. In addition, the model structure error contribution is not easy to interpret since the error model lumps all structural errors into one process: in RC, RR and RL itself, and likewise for other uncertainties which are not explicitly considered here, such as the input uncertainty. Further explanation is provided in the Discussion (point ii).

4 Discussion

In the study presented, we proposed an approach to assess the uncertainty of water level predictions with consideration of the uncertainty of the rating curve. To better interpret the results, we would like to discuss (i) the specific water level predictions for Sluzew Creek, (ii) methodological aspects, joint uncertainty assessment and its limitations, and (iii) implications for practical applications and future perspectives.

Regarding the case study results, generally, the interpretation of estimated parameters is always tenuous as parameters lose (some degree of) their physical meaning during a calibration process if the model structure is not perfect (Wagener and Gupta, 2005). The posterior distribution of the

RR parameters suggests that the Sluzew Creek catchment responds rapidly to heavy rainfalls and that only a part of the catchment contributes to the streamflow observed in the stream. This may be explained by the fact that, as in every event-based model, only the direct runoff which occurs during the first phase is modelled, while a slower runoff due to the catchment retention is omitted. Also other discharges that are drained by the canalization network are not modelled.

These findings correspond to the results of a previous study from the same catchment that used a different data set and a different model (Sikorska et al., 2012). As that study focused on *streamflow* predictions, whereas here we investigate *water levels*, the parameter estimates cannot be directly compared. Specifically, the parameters compensate differently for the structural limitations of the hydrological models.

In this case we found that the contribution to the total predictive uncertainty of the RC and RR parameters is small, with a slight dominance of the RC submodel. However, those findings are strongly case related and cannot be directly transferred to other catchments. Therefore, for future studies we suggest performing first the uncertainty analysis proposed by us to pinpoint individual contributions. Clearly, the largest contribution still remains the structural uncertainty of the RL model itself. The uncertainty of the LR model stems, however, not only from the uncertainty about its parameters. Additional uncertainty contributions are errors due to structural limitations of the LR and measurement errors of the water level. These are included in the autocorrelated error term, which also lumps all uncertainties of the RR model that are not explicitly considered, such as the input uncertainty in rainfall data. Therefore, these uncertainties cannot be separated here and we only compare the uncertainty contribution from the *parameters* of the two submodels (RC versus RR). Although we did not attempt to separately assess all uncertainty contributions, it is conceptually straightforward to extend this framework with an explicit model for input uncertainty (e.g. rainfall multipliers) (Kavetski et al., 2006). Practically, the computational effort could be limiting.

Additionally, the comparably large contribution of the model structure uncertainty of the simplified rainfall-runoff model is interesting. Such simplified models only have a few parameters. They are therefore convenient for flood predictions when only limited data are available so that the application of structurally more complex models is not possible. As our case shows, simple models can predict flood events satisfactorily as long as the rainfall-runoff in the catchment follows conventional rainfall-runoff processes i.e. in natural catchments or when the streamflow is not disturbed due to external factors (e.g. hydraulic infrastructure). However, all models are limited in predicting extreme flood events where unforeseen interactions occur, e.g. external processes that are not included in the model structure. This also explains why the statistical assumptions with respect to the innovations in the applied lumped error model are sometimes violated.

Where this is critical, different error models or transformations could have also been investigated (Del Giudice et al., 2013).

It must be also stressed that the applied RR model, as an event-based model, is limited to model only the rainfall-runoff process within the catchment while omitting other water balance components such as groundwater, base flow and evapotranspiration. Such simplified lumped models are especially useful when only the output of the catchment, but not the processes within the catchment, is of interest.

In our view, the simplified model structure also explains the dominating uncertainty contribution of the lumped error model. On the one hand, the limited-model structure causes large systematic errors in the predictions. On the other hand, a few model parameters (here: seven) have a relatively well-defined prior. Together with the larger number of observations (here: ca. 2000 data points), this results in a very narrow posterior distribution. While parameter uncertainty gets smaller as more data are available, the existing model structure deficits, as well as input errors, remain the same; hence the variance of the lumped error term remains the same, i.e. large.

The approach presented is only useful if the water level is the quantity of interest. While this is the case in many situations, namely for predicting flood hazard or inundation risk, other applications require streamflow predictions, e.g. sizing a culvert or operating a reservoir. If the water level is modelled, streamflow is an internal state of the RL model for which no probabilistic statements can be made. Conceptually, an extension that enables the calculation of the predictive distribution of streamflow is possible so that the streamflow is inferred for all points in time. A similar problem is solved with rainfall multipliers when input uncertainty is considered explicitly and the "true" rainfall must be inferred (see Sikorska et al., 2012). Rainfall multipliers represent a single correction factor per rain event, which is estimated simultaneously with the model parameters. This is very useful to reduce the number of inferred parameters. However, in the case of inferring streamflow, a similar useful simplification is not obvious because streamflow cannot meaningfully be divided into events. This requires further research.

Regarding flood level predictions, the Bayesian framework seems very promising because it enables the modeler to incorporate informal knowledge from easily accessible information. In addition, the uncertainty of the LR model and a rating curve in particular may be significant for poorly gauged stream gauges. For practical applications, it is important to update it frequently to reduce the uncertainty of the LR model. This is especially important for dynamic catchments, where cross sections change seasonally or with changing land use. To avoid such problems in a practical setting, remote sensing data from satellites could be incorporated to reduce the uncertainties of already existing rating curves (Di Baldassarre and Uhlenbrook, 2012). Also popular nowadays are social networks (e.g. Facebook or Twitter),

which may be used; e.g. flood-observer groups[1] might provide valuable information to calibrate models a posteriori.

Another important point concerns predictions beyond the valid extrapolation range of a rating curve. This is always challenging and should be done based on a cross-section analysis. Predicted uncertainties outside of a reasonable range cannot be treated as reliable. This could be improved by performing more streamflow-water level measurements during flood flows, especially during flood peaks. Improving the observations is an obvious way to improve flood predictions. In this regard, flood predictions and the uncertainty due to model parameters could generally benefit from gathering more calibration data. However, under changing conditions of the catchment (e.g. urbanization), its characteristics cannot be considered stationary. This is a general problem of every model applied for long-term predictions. A model calibrated under certain (stationary) conditions cannot forecast the catchment response under (unknown) changed conditions but only makes predictions for the given situation (Blöschl and Montanari, 2010). If one is interested in modelling the catchment behaviour for different conditions, the parameters have to be modified accordingly.

Another strategy to improve the predictive capability of the model can be the reduction of hydrological model structure deficits. Although the model accuracy usually increases, this must not necessarily lead to reducing the predictive uncertainty because of increasing parameter uncertainty (Sikorska, 2013). Yet, to effectively reduce structural deficits of the model the associated uncertainties of model predictions need to be explicitly decomposed. This requires going beyond the lumped error model and assessing the influence of the individual uncertainty contributions. Reichert and Schuwirth (2012) and Reichert and Mieleitner (2009) have suggested possible procedures that seem promising.

The value of our work is that we provide a method to systematically incorporate the uncertainty of the calibration data in rainfall-runoff modelling and showed the deficits of the usually applied calibration procedure of RR models. This is especially important since uncertainty analysis or, more generally, flood predictions cannot be assumed reliable if they rely on unreliable data. The proposed procedure can be combined with approaches to further decompose the predictive uncertainty.

5 Conclusions

In this study, we proposed an innovative approach to quantify the complete uncertainty in water levels predictions by means of an integrated assessment of the rainfall-runoff model and the corresponding water level-runoff model, which typically is a rating curve. Specifically, we use a formal Bayesian framework to assess the uncertainty contributions of the parameters of the rainfall-runoff and water level-runoff models to the total predictive uncertainty. By modelling water levels directly, we avoid the unjustifiable assumption that the calibration data are free of errors. Based on our main results we conclude that:

– Our approach is formulated generally and not limited to the rainfall-runoff and rating curve submodels presented.

– In addition, it is not data-demanding since it requires only already existing hydrometric data on a rating curve. Using informative prior distributions makes it also applicable to poorly gauged catchments.

– For structurally simple models, the fulfilment of statistical assumptions is, arguably, not always perfect. However, the autocorrelated lumped error model fulfils the underlying statistical assumptions much better than the traditional i.i.d. error model.

– As expected, our results demonstrate that predicted water levels are unrealistic and usually overestimated when the rating curve is extrapolated outside the permissible range. This range is not necessarily equal to the measurement range, especially for irregular or complicated bathymetric profiles. Therefore, it is crucial to continuously update the applied rating curve.

– In the case study presented, the uncertainty contribution from the rating curve parameters was as relevant as that from the rainfall-runoff model parameters. However, such uncertainty contributions are strongly case-related and greatly depend on the available monitoring data, chosen submodel structures, and catchment and cross-section properties. Therefore they cannot be generalized. In our view, to assess the uncertainty contribution from a rating curve requires repeating the uncertainty analysis proposed in this study.

– The main limitation of the approach presented is that it is limited to water level predictions. Future research should, on the one hand, concentrate on extending the approach to streamflow, which is often not measured directly. On the other hand, it is important to further improve the assessment of the individual uncertainty contributions to obtain better flood predictions.

Supplementary material related to this article is available online at http://www.hydrol-earth-syst-sci.net/ 17/4415/2013/hess-17-4415-2013-supplement.pdf.

---

[1]E.g. user group "Flood Group UK", https://www.facebook. com/floodgroupuk?ref=ts&fref=ts.

References

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

Banasik, K., Hejduk, L., and Barszcz, M.: Flood flow consequences of land use changes in a small urban catchment of Warsaw, in: 11th International Conference on Urban Drainage, 31, 10 pp., 2008.

Barszcz, M.: Forecast of probably flows caused by heavy rainfall on urbanized drainage basin of Służew Stream, Scientific Review Engineering and Environmental Sciences, 4, 3–21, 2009.

Beck, M. B.: Principles of Modelling, Water Sci. Technol., 24, 1–8, IWA Publishing, 1991.

Blöschl, G. and Montanari, A.: Climate change impacts – throwing the dice?, Hydrol. Process., 24, 374–381, doi:10.1002/hyp.7574, 2010.

Chivers, C.: MHadaptive: General Markov Chain Monte Carlo for Bayesian Inference using adaptive Metropolis-Hastings sampling. R package, available at: http://www.R-project.org (last access: 4 March 2013), 2012.

Coccia, G. and Todini, E.: Recent developments in predictive uncertainty assessment based on the model conditional processor approach, Hydrol. Earth Syst. Sci., 15, 3253–3274, doi:10.5194/hess-15-3253-2011, 2011.

Coutu, S., Del Giudice, D., Rossi, L., and Barry, D.: Parsimonious hydrological modeling of urban sewer and river catchments, J. Hydrol., 464–465, 477-484, doi:10.1016/j.jhydrol.2012.07.039, 2012.

Del Giudice, D., Honti, M., Scheidegger, A., Albert, C., Reichert, P., and Rieckermann, J.: Improving uncertainty estimation in urban hydrological modeling by statistically describing bias, Hydrol. Earth Syst. Sci., 17, 4209–4225, doi:10.5194/hess-17-4209-2013, 2013.

Deletic, A., Dotto, C. B. S., McCarthy, D. T., Kleidorfer, M., Freni, G., Mannina, G., Uhl, M., Henrichs, M., Fletcher, T. D., Rauch, W., Bertrand-Krajewski, J. L., and Tait, S.: Assessing uncertainties in urban drainage models, Phys. Chem. Earth 42–44, 3–10, doi:10.1016/j.pce.2011.04.007, 2012.

Di Baldassarre, G. and Claps, P.: A hydraulic study on the applicability of flood rating curves, Hydrol. Res., 42, 10–19, doi:10.2166/nh.2010.098, 2011.

Di Baldassarre, G. and Montanari, A.: Uncertainty in river discharge observations: a quantitative analysis, Hydrol. Earth Syst. Sci., 13, 913–921, doi:10.5194/hess-13-913-2009, 2009.

Di Baldassarre, G. and Uhlenbrook, S.: Is the current flood of data enough? A treatise on research needs for the im-

provement of flood modelling, Hydrol. Process., 26, 153–158, doi:10.1002/hyp.8226, 2012.

Di Baldassarre, G., Laio, F., and Montanari, A.: Effect of observation errors on the uncertainty of design floods, Phys. Chem. Earth, 42–44, 85–90, doi:10.1016/j.pce.2011.05.001, 2012.

Domeneghetti, A., Castellarin, A., and Brath, A.: Assessing ratingcurve uncertainty and its effects on hydraulic model calibration, Hydrol. Earth Syst. Sci., 16, 1191–1202, doi:10.5194/hess-16-1191-2012, 2012.

Dotto, C. B. S, Mannina, G., Kleidorfer, M., Vezzaro, L., Henrichs, M., McCarthy, D. T., Freni, G., Rauch, W., and Deletic, A.: Comparison of different uncertainty techniques in urban stormwater quantity and quality modelling, Water Res., 46, 2545–2558, 2012.

Dottori, F., Martina, M. L. V., and Todini, E.: A dynamic rating curve approach to indirect discharge measurement, Hydrol. Earth Syst. Sci., 13, 847–863, doi:10.5194/hess-13-847-2009, 2009.

Gelman, A., Carlin, J. B., Stern, H. S., and Rubin, D. B.: Bayesian Data Analysis, 2nd Edn., Chapman & Hall, Boca Raton, Florida, 2003.

Haario, H., Saksman, E., and Tamminen, J.: An adaptive Metropolis algorithm, Bernoulli, 7, 223–242, 2001.

Harrell, F. E.: Regression Modeling Strategies: With Applications to Linear Models, Logistic Regression, and Survival Analysis (Springer Series in Statistics), Springer, New York, 2010.

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

Khaleghi, M. R., Gholami, V., Ghodusi, J., and Hosseini, H.: Efficiency of the geomorphologic instantaneous unit hydrograph method in flood hydrograph simulation, Catena, 87, 163–171, doi:10.1016/j.catena.2011.04.005, 2011.

Le Coz, J.: A literature review of methods for estimating the uncertainty associated with stage-discharge relations, WMO, 2012.

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

McMillan, H., Jackson, B., Clark, M., Kavetski, D., and Woods, R.: Rainfall Uncertainty in Hydrological Modelling: An Evaluation of Multiplicative Error Models/Mendeley, J. Hydrol., 400, 83–94, doi:10.1016/j.jhydrol.2011.01.026, 2011.

Mishra, S. K. and Singh, V. P.: Soil Conservation Service Curve Number (SCS-CN) Methodology, 1st Edn., Springer, Netherlands, 2010.

Montanari, A.: What do we mean by "uncertainty"? The need for a consistent wording about uncertainty assessment in hydrology, Hydrol. Process., 21, 841–845, doi:10.1002/hyp.6623, 2007.

Montanari, A. and Koutsoyiannis D.: A blueprint for processbased modeling of uncertain hydrological systems, Water Resour. Res., 48, W09555, doi:10.1029/2011WR011412, 2012.

Nash, J. E.: The form of instantaneous unit hydrograph, Int. Assoc. Sci. Hydrol., 45, 114–121, 1957.

Nash, J. and Sutcliffe, J.: River flow forecasting through conceptual models part I : A discussion of principles, J. Hydrol., 10, 282–290, doi:10.1016/0022-1694(70)90255-6, 1970.

Pappenberger, F., Matgen, P., Beven, K. J., Henry, J. B., Pfister, L., and De Fraipont, P.: Influence of uncertain boundary conditions

and model structure on flood inundation predictions, Adv. Water Resour., 29, 1430–1449, 2006.

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

R Development Core Team: R: A language and environment for statistical computing, available at: http://www.R-project.org (last access: 4 March 2013), 2011.

Reichert, P. and Mieleitner, J.: Analyzing input and structural uncertainty of nonlinear dynamic models with stochastic, time-dependent parameters, Water Resour. Res., 45, W10402, doi:10.1029/2009WR007814, 2009.

Reichert, P. and Schuwirth, N.: Linking statistical bias description to multiobjective model calibration, Water Resour. Res., 48, W09543, doi:10.1029/2011WR011391, 2012.

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

Renard, B., Kavetski, D., Leblois, M., Thyer, M., Kuczera, G., and Franks, S. W.: Toward a reliable decomposition of predictive uncertainty in hydrological modeling: Characterizing rainfall errors using conditional simulation, Water Resour. Res., 47, W11516, doi:10.1029/2011WR010643, 2011.

Sadegh, M. and Vrugt, J. A.: Approximate Bayesian Computation in hydrologic modeling: equifinality of formal and informal approaches, Hydrol. Earth Syst. Sci. Discuss., 10, 4739–4797, doi:10.5194/hessd-10-4739-2013, 2013.

Sauer, V. B. and Meyer, R. W.: Determination of error in individual discharge measurements, USGS Open File Report, USGS, Washington, DC, 1992.

Seibert, J.: Conceptual runoff models – fiction or representation of reality? Acta Univ. Ups., Comprehensive Summaries of Uppsala Dissertations from the Faculty of Science and Technology 436, 52 pp., Uppsala, ISBN 91-554-4402-4, 1999.

Seibert, J. and McDonnell, J.: Gauging the Ungauged Basin: The Relative Value of Soft and Hard Data, J. Hydrol. Eng., doi:10.1061/(ASCE)HE.1943-5584.0000861, 2013.

Sikorska, A. E.: Interactive comment on "Bayesian uncertainty assessment of flood predictions in ungauged urban basins for conceptual rainfall-runoff models" by A. E. Sikorska et al., 2012, Hydrol. Earth Syst. Sci. Discuss., 8, C6284–C6310, 2012.

Sikorska, A. E.: Uncertainty analysis of rainfall-runoff predictions for a small urbanized basin, doctoral thesis, Faculty of Civil and Environmental Engineering, Warsaw University of Life Sciences – SGGW, 2013.

Sikorska, A. and Banasik, K.: Parameter identification of a conceptual rainfall-runoff model for a small urban catchment, Annals of Warsaw University of Life Sciences-SGGW, Land Reclamation, 42, 279–293, 2010.

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

Uhlenbrook, S., Seibert, J., Leibundgut, C., and Allan Rodhe, A.: Prediction uncertainty of conceptual rainfall-runoff models caused by problems in identifying model parameters and structure, Hydrolog. Sci. J., 44, 779–797, doi:10.1080/02626669909492273, 1999.

Vrugt, J. A., Braak, C. J. F., Gupta, H. V., and Robinson, B. A.: Equifinality of formal (DREAM) and informal (GLUE) Bayesian approaches in hydrologic modeling?, Stoch. Env. Res. Risk A, 23, 1011–1026, doi:10.1007/s00477-008-0274-y, 2008.

Wagener, T. and Gupta, H. V.: Model identification for hydrological forecasting under uncertainty, Stoch. Env. Res. Risk A., 19, 378–387, doi:10.1007/s00477-005-0006-5, 2005.

Wagener, T. and Montanari, A.: Convergence of approaches toward reducing uncertainty in predictions in ungauged basins, Water Resour. Res., 47, W06301, doi:10.1029/2010WR009469, 2011.

Wagener, T., Gupta, H. V., and Sorooshian, S.: Stochastic formulation of a conceptual hydrological model, in: Hydrology: science and practice for the 21st century, Proceedings of the British Hydrological Society International Conference, Imperial College, London, 12–16 July 2004, 398–405, 2004.

Wang, Q. J. and Robertson, D. E.: Multisite probabilistic forecasting of seasonal flows for streams with zero values occurences, Water Resour. Res., 47, W02546, doi:10.1029/2010WR009333, 2011.

Westerberg, I., Guerrero, J.-L., Seibert, J., Beven, K. J., and Halldin, S.: Stage-discharge uncertainty derived with a non-stationary rating curve in the Choluteca River, Honduras, Hydrol. Process., 25, 603–613, doi:0.1002/hyp.7848, 2011.

WMO: Guide to Hydrological Practice, Volume I, Hydrology – From Measurement to Hydrological Information, 6th Edn., World Meteorological Organisation, Geneva, Switzerland, 2008.

Yang, J., Reichert, P., Abbaspour, K. C., and Yang, H.: Hydrological modelling of the Chaohe Basin in China: Statistical model formulation and Bayesian inference, J. Hydrol., 340, 167–182, 2007.

Yang, J., Reichert, P., Abbaspour, K., Xia, J., and Yang, H.: Comparing uncertainty analysis techniques for a SWAT application to the Chaohe Basin in China, J. Hydrol., 358, 1–23, doi:10.1016/j.jhydrol.2008.05.012, 2008.