

## Model parameter estimation with imprecise information

Wolfgang Rauch <sup>a,\*</sup>, Nikolaus Rauch <sup>b</sup> and Manfred Kleidorfer <sup>a</sup>

<sup>a</sup> University of Innsbruck, Unit of Environmental Engineering, Technikerstrasse 13, Innsbruck, A-6020, Austria

<sup>b</sup> University of Innsbruck, Interactive Graphics and Simulation Group, Technikerstrasse 13, Innsbruck, A-6020, Austria

\*Corresponding author. E-mail: Wolfgang.Rauch@uibk.ac.at

 WR, 0000-0002-6462-2832; NR, 0000-0003-1326-709X; MK, 0000-0002-4001-1711

### ABSTRACT

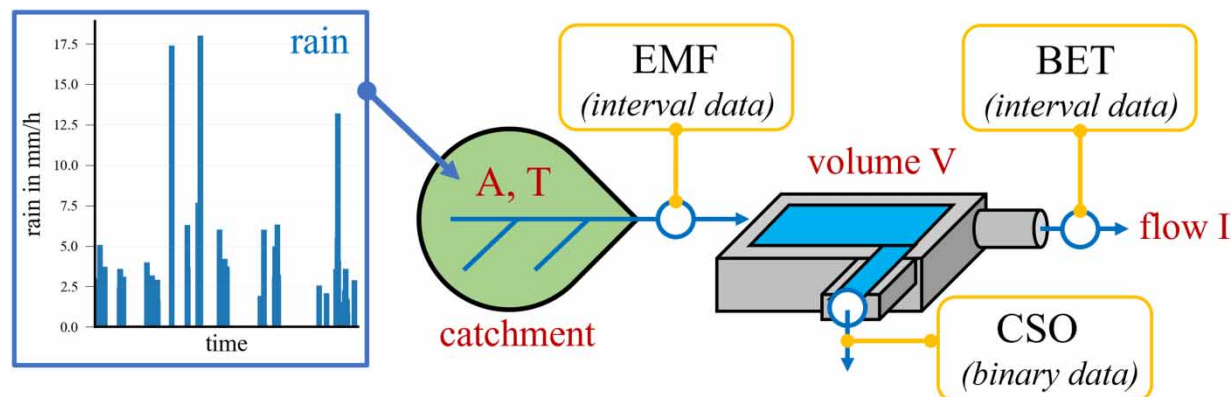
Model parameter estimation is a well-known inverse problem, as long as single-value point data are available as observations of system performance measurement. However, classical statistical methods, such as the minimization of an objective function or maximum likelihood, are no longer straightforward, when measurements are imprecise in nature. Typical examples of the latter include censored data and binary information. Here, we explore Approximate Bayesian Computation as a simple method to perform model parameter estimation with such imprecise information. We demonstrate the method for the example of a plain rainfall–runoff model and illustrate the advantages and shortcomings. Last, we outline the value of Shapley values to determine which type of observation contributes to the parameter estimation and which are of minor importance.

**Key words:** approximate Bayesian computation, binary data, calibration, imprecise information, inverse modeling, Shapley value

### HIGHLIGHTS

- Approximate Bayesian Computation is a likelihood-free approach for inverse parameter estimation.
- The method is both computationally simple and versatile to use imprecise information.
- Censored and binary data are useful observations for parameter estimation.
- Shapley values allow us to estimate the significance of a specific type of observation.
- The Monte-Carlo based method is computationally expensive.

### GRAPHICAL ABSTRACT



### INTRODUCTION

Model parameter estimation by inverse modeling is the adjustment of the unknown parameter values to get a match between model predictions and measurement data, i.e., observations (Refsgaard 1997; Šimůnek & De Vos 1999; Tschekner-Gratl *et al.* 2016). Classical inverse modeling relies on paired prediction – observation in the form of

This is an Open Access article distributed under the terms of the Creative Commons Attribution Licence (CC BY 4.0), which permits copying, adaptation and redistribution, provided the original work is properly cited (<http://creativecommons.org/licenses/by/4.0/>).

single-valued point data and the observations are assumed as being uncensored. Two approaches are popular to solve the calibration problem, i.e., first, the method of minimizing an objective function and second, maximum likelihood. The first method selects some objective function that is related to the deviations between observations and model predictions. The best fit of the parameter values is achieved by minimizing the objective function (Marquardt 1963; Madsen *et al.* 2002). The second method determines the best fit by maximizing the likelihood – which is the probability of the model predictions to match the observations (Beven & Binley 1992). Many textbooks are available on calibration by inverse modeling (Beck & Arnold 1977; Tarantola 2005; Beven 2010; Aster *et al.* 2018, etc.) and the reader is expected to be familiar with the approach.

For any practical application, it has to be seen that measurements are inaccurate in nature (see, for e.g., McMillan *et al.* 2018 for hydrological data) and – in principle – classical calibration methods are no longer straightforwardly applicable. Still, as long as single-valued point data (or a time series thereof) is available, the most common approach is to ignore errors in the parameter estimation and accept the resulting uncertainty in model prediction. More recently, authors such as Beven & Binley (1992), Abbaspour *et al.* (1997), Higdon *et al.* (2004), Gupta *et al.* (2005), Todini (2008), Dotto *et al.* (2011), Dotto *et al.* (2012) discuss methods to take the uncertainties in measurements as well as deficits in model structure into account (Deletic *et al.* 2012). The underlying idea is to assign uncertainties to the parameters so that the model predictions fit the observations as well as presumptions on inaccuracy and other prior information (Abbaspour *et al.* 2007). With increasing model complexity and linked or coupled models, the propagation of uncertainties through models becomes more important (Tscheikner-Gratl *et al.* 2019). At the same time, data collection and handling methods are improving. However, there are still uncertainties related to data availability (Hajihassanpour *et al.* 2023), and even when data availability is good, it may be difficult to process all available data sets (Tscheikner-Gratl *et al.* 2016). Finally, recent developments include the application of artificial intelligence algorithms to calibration and uncertainty assessment problems (Garzón *et al.* 2022), and attempts to assess long-term prediction uncertainties related to global change (Kourtis & Tsihrintzis 2021) and to improve robustness under deep uncertainty (Casal-Campos *et al.* 2018).

However, there are instances where point measurements are no longer available, i.e., the information on the system performance cannot be measured exactly but the information is imprecise. In statistical terms we can consider imprecise observations as being censored (meaning that only part of the information is available – see for e.g., Gijbels 2010). This is a common problem in various urban drainage applications. For example, in deterioration models of ageing infrastructure, censored data are often discussed in relation to the problem that data are only available for parts of the system. If this introduces a bias into the data, e.g., if only data-within an area is collected, or only older pipes with certain materials are inspected, this can lead to poorly positioned data driven models (Scheidegger *et al.* 2011). Other options are simplified data such as proxy data or binary data (Wani *et al.* 2017). However, despite being imprecise such information is still considered to be hard data – as compared to soft data. Seibert & McDonnell (2002) define soft data as ‘qualitative knowledge from the experimentalist that cannot be used directly as exact numbers but that can be made useful through fuzzy measures of model simulation and parameter value acceptability’. Thus - in this context – one can turn soft data into imprecise information by subjective judgement of associated probabilities.

Different approaches are pursued for parameter estimation using the – above mentioned – imprecise information: e.g., Otto & Bandemer (1986) use fuzzy theory, Wang (2019) evidence theory, while Sankararaman & Mahadevan (2012) provide a general framework based on both least square optimization and likelihood assessment. Wani *et al.* (2017) develop a methodology for censored and binary data by defining a likelihood function and using Bayesian interference for the estimation.

In this paper we demonstrate the use of a likelihood-free approach for this problem, that is Approximate Bayesian Computation (ABC) (Nott *et al.* 2012; Vrugt & Sadegh 2013; Kavetski *et al.* 2018). The virtue of the approach is its intuitive angle toward using imprecise information and the simple computational implementation. In the following, we first present the theoretical background and then apply the methodology to two case studies. The first one is a simple but real-world parameter estimation problem that arose in wastewater-based epidemiology, while the second shows the application to a semi-virtual urban drainage catchment – using measured rain series as input data. In order to present a unified framework, we also demonstrate how to use Shapley values to identify which observations are supporting the parameter estimation and which are of little importance.

## METHODS

### Typology of imprecise information

A universal classification of incomplete, censored or collapsed data for parameter estimation is likely too complex to be helpful and thus the following typology aims at situations that arise in typical urban catchment hydrology (Figure 1):

- Interval data

The data are bracketed, i.e., the measurement value lies within an interval defined with lower and upper bounds. Typically, the boundary values are estimated as confidence intervals from observations or prior knowledge, e.g., CI 2.5–97.5%. Also, soft data can be transferred into (hard) interval data. An example would be the expectation that data from a highly unreliable sensor can still be bracketed within a prediction uncertainty.

- Probability distribution

This information is given for the classical description of measurement uncertainty and assumes that observations can be described with a probability distribution. For example, the uncertainty in a sensor measurement could be assumed to be normally distributed. Note that for a uniform distribution, the information is equal to interval data.

- Binary information

The measurement quantity is either Yes/No or value A/value B. The simplest form is the information on the state of a system state, e.g., for a certain rain event, we either experience a combined sewer overflow or not or flooding on the surface or not.

- Boundary information

The data are censored in such a way that measurements can only be lower or upper a defined boundary of the true value. For example, the runoff from an urban catchment cannot be lower than the dry weather baseflow. Boundary information is a special case of interval data.

### Maximizing similarity

The traditional approach in inverse modeling (denoted minimizing objective function) defines first an objective function such as root mean squared error (RMSE) or – for hydrological systems – Nash-Sutcliffe efficiency (NSE)

$$NSE = 1 - \frac{\sum (Y_o^t - Y_m^t)^2}{\sum (Y_o^t - \bar{Y}_o)^2}$$

where  $t$ =time from 1..T;  $Y_o$  is observation,  $Y_m$  model prediction and  $\bar{Y}_o$  is the mean of the observations. The objective function is a measure of the success of the calibration. Next the optimal parameter set  $\theta$  is determined by maximizing  $NSE$  toward best fit  $NSE = 1$ , which is the same as minimizing  $RMSE$ , using repeated model simulations with varying  $\theta$ .

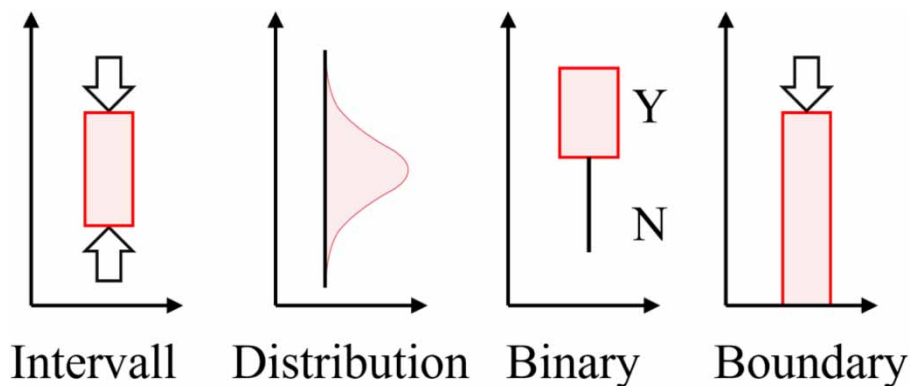


Figure 1 | Schematic representation of imprecise information.

It is possible to adapt this traditional approach in inverse modeling for imprecise information, by using the following algorithm:

1. Transform the imprecise observations  $Y_O$  into a matching distribution  $q(Y_O)$  (e.g., uniform between lower/upper bounds for interval data) and define an initial set of parameter values  $\Theta$  as well as a (prior) distribution for the parameters  $\pi(\Theta)$
2. Compute the resulting probability distribution of the model predictions  $p(Y_m)$  e.g., by sampling from  $\pi(\Theta)$  using Monte-Carlo Simulation
3. Compute a similarity metric of the two distributions  $q(Y_O)$  and  $p(Y_m)$
4. Repeat Steps 2 and 3 with increasing improvement of  $\Theta$  until the similarity is maximized

Note that we have to replace the objective function with the similarity metric of two distributions. We are using here the distribution-free overlapping index  $\eta[0, 1]$  (Pastore & Calcagni 2019) as being both simple to implement and robust in performance:

$$\eta(q, p) = \int \min[f_q(x), f_p(x)] dx$$

### Approximate Bayesian Computation

As demonstrated by Wani *et al.* (2017) the maximum likelihood method can be extended to use censored and binary information for parameter estimation. Here we take a slightly different approach (Nott *et al.* 2012), which is denoted as likelihood-free Bayesian inference or ABC. The method was originally introduced in 1990 as a rejection technique thus bypassing the requirement to compute a likelihood function (Tavaré *et al.* 1997; Pritchard *et al.* 1999). The prototype rejection algorithm has been suggested as follows (adapted from Ke & Tian 2019):

Given the observations  $Y_O$  and a prior distribution for the parameters  $\pi(\Theta)$

1. Generate a sample from the prior distribution  $\Theta' \sim \pi(\Theta)$
2. Simulate using  $\Theta'$  to get  $Y_m \sim p(Y_m|\Theta')$
3. Accept the sample if  $Y_m = Y_O$  for binary data  $Y_m \in [Y_m^{lower\ bound}, Y_m^{upper\ bound}]$  for interval data otherwise reject the sample
4. Repeat 1. to 3. until the required number of posterior samples are obtained

The result is a posterior distribution of the model parameters  $\Theta$  that match the imprecise information. Note that the ABC methodology usually applies a summary statistic for acceptance (Ke & Tian 2019) much in the same way as the objective function mentioned above. However, it is actually the virtue of the simple ABC rejection scheme that the information on the acceptance of a sample has not been transformed by complex functions but can be used directly. Compared to standard Bayesian inference the benefit is, that no assumption of the distribution of the model residuals is necessary. It is thus evident that the simple and basic reject scheme is well suited for the assessment of imprecise information.

The basic ABC rejection method may result in a huge computational burden if the acceptance rate is low, as a suitable number of posterior samples is necessary. Just for the sake of an estimate: given an acceptance rate of 1% and aiming to achieve  $10^2$  posterior samples would already require  $10^2/0.01 = 10^4$  simulations. Thus, many improvements to the simple rejection scheme have been proposed, most importantly MCMC-ABC (Marjoram *et al.* 2003) and sequential Monte-Carlo (SMC) methods (Sisson *et al.* 2007). In order not to overload the paper we refrain here from discussing (and implementing) possible improvements but use the plain rejection scheme.

Another point that needs to be addressed is the statistics on the resulting posterior distribution of the parameters. For assessment of the parameter estimation and comparison of methods, one might not be able to use the full posterior distribution of the parameters but needs a resulting set of values  $\Theta_R$ . If a summary statistic is available, one can resort to using this information to determine the best-fitting parameter set. However, if we have more than one parameter and lack such summary statistics, this is no longer straightforward. Intuitively – and ignoring covariance – we can either use the median (50%ile values) or the mode (maximum density value) of the posterior probability density function. For the problems at hand, the median values gave more convincing results and have thus been used throughout.

### Shapley value

Especially for imprecise measurements, it happens in practice that different sources of information (types of observations, i.e., input time series) are available and used for parameter estimation. The general attitude is, that the more the better.

However, this simple approach falls short as inverse model parameter estimation is usually nonlinear. Depending on both the information available and the structure of the model (including modeling aim) it is important to identify which observations support the parameter estimation and which are of little importance. Lundberg & Lee (2017) and Sundararajan & Najmi (2020) demonstrate that Shapley values (Shapley 1953) are a measure of the importance of a specific feature (here observation) to model predictions.

The concept has originated from cooperative game theory and aims to distribute the payoff among several players in a coalition game (i.e., players can form coalitions in order to increase the payoff from the game). A marginal contribution is the increase of the payoff when a player joins. Shapley values quantify the contribution of each player to the total payoff with a weighted sum of its marginal contribution to all combinations of players (i.e., all possible coalitions). The Shapley value  $\phi$  of a player  $i$  is thus computed according to the following relation (Shapley 1953):

$$\phi_i(v) = \sum_{S \subset N} \frac{(S-1)!(n-S)!}{n!} (v(S) - v(S-i))$$

where  $N = \{1, \dots, n\}$  set of  $n$  players,  $S$  subset of  $N$  (indicating all possible coalitions with  $S$  denoting the number of players in each coalition) and  $v$  = characteristic (objective) function. In our framework a player is representing a certain type of observation (input time series) and the payoff is a measure to quantify the result of the parameter estimation. For the latter, we, for example, can use the metric of the objective function, i.e., *RMSE* or *NSE*. Note that the interpretation of Shapley values can require a complex framework – especially for a larger number of  $n$  (Lundberg & Lee 2017) but for the problem at hand, we can resort to a direct comparison of the Shapley values.

### Computational implementation

The computational burden in Monte-Carlo based procedures limits applications to fast models and/or requires the use of efficient algorithms and parallel computing. The numerical scheme was coded in ANSI C and tested on a PC with an Apple M1 Max 10-Core @3.22 GHz processor and 32 GB LPDDR5-6400 RAM. The runtime of processing the catchment hydrology case study is 70.3 sc for 200,000 Monte-Carlo samples – for the calibration rain series with a length 3 months and 10 min timestep resolution. Since the implementation is numerically efficient, the performance is fast enough for analysis. Parallelization of the code would further improve the computational speed as the computational complexity is  $O(N)$  with  $N$  being the number of Monte-Carlo samples.

## RESULT

### Case study 1 – Wastewater-based epidemiology

The first case study addresses a problem that occurred during the COVID-19 pandemic: According to the general concept of wastewater-based epidemiology, the measured virus concentration at a certain point in the sewer system (usually at the wastewater treatment plant) relates to the total number of infected (thus virus shedding) persons in the upstream catchment (see for e.g., Rauch *et al.* 2022). Assuming that each infected person is shedding a certain load of genetic material per time ( $L_{shed}$ ) into the sewer system, and neglecting all uncertainties and losses we get a simplified measure of the infected persons in the watershed (*IP*):

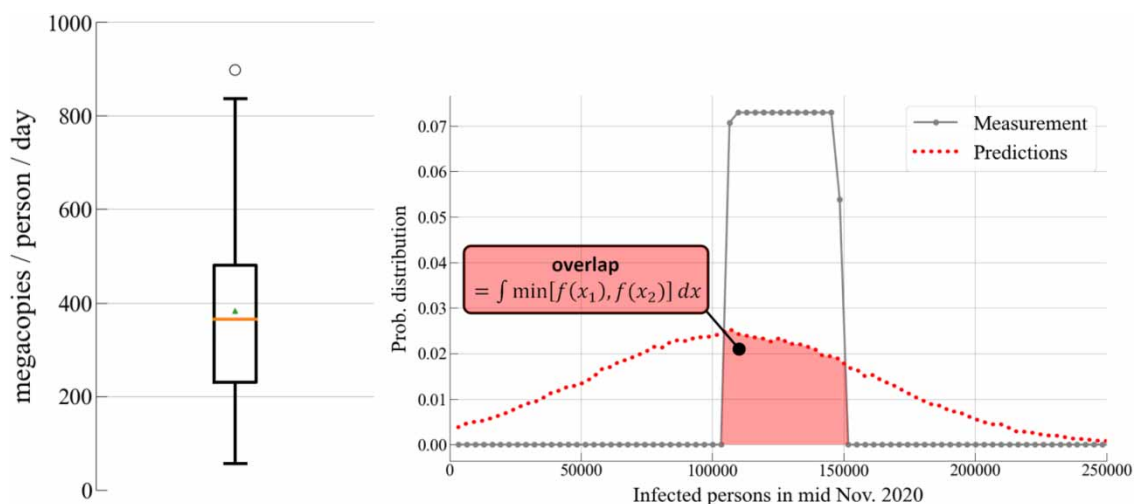
$$IP = \frac{c_{virus} \times Q}{L_{shed}} = \frac{L_{virus}}{L_{shed}}$$

where  $Q$  = flow volume in L/P/d;  $c_{virus}$  = measured virus concentration in gene copies/L,  $L_{virus}$  = measured virus load in gene copies/P/d,  $IP$  = number of infected persons in the watershed and  $L_{shed}$  is the individual virus shedding load in gene copies/ $P_{infected}/d$ .

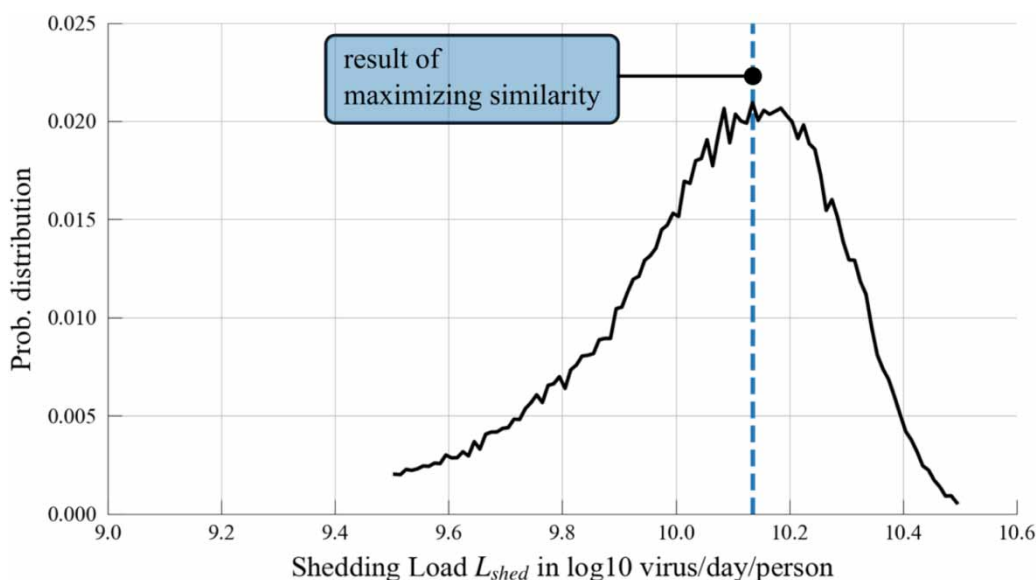
The individual virus shedding load per *IP* is an interesting parameter in pandemic management as a proxy for the severity of virus variants and the immunity of the population. Given we have information on the number of infected persons and measured virus load ( $L_{virus}$ ) we can compute  $L_{shed}$  directly. However, in reality, the information is imprecise at best. For Austria there is interval data (95% CI 2.6–3.5%) available for the percentage of infected persons in the total population in the period 12th–14th November 2020 (Rauch *et al.* 2024). For viral load measurement, we do have data from 22 wastewater

treatment plants in the same period – covering 4.1 Mill. PE, i.e., approximately 50% of the Austrian population (see Figure 2 Left). It is estimated that the measurements ( $L_{virus}$ ) can be expressed as normally distributed with  $\mu = 366$  and  $\sigma = 188$  - both in Mcopies/P/d).

For the task at hand, we thus have uncertain input data ( $L_{virus} \sim \mathcal{N}(366, 35\,344)$ ) and interval data as observations ( $Y_O \sim \mathcal{U}(106\,600, 147\,600)$  infected Persons). For the sought parameter we assume a uniform prior distribution ( $L_{shed} \sim \mathcal{U}(10^{9.5}, 10^{10.5})$  in copies/P/d). Figure 2 plots the resulting distribution for the optimal parameter value for the Maximum-Similarity approach ( $L_{shed} = 10^{10.15}$ ), while Figure 3 gives the distribution of the accepted parameter space for ABC with  $L_{shed} = 10^{10.135}$  providing the highest density. It is obvious that both approaches give quite coinciding results with a relative error of 0.1% for the exponent.



**Figure 2** | Wastewater-based epidemiology – Left: Boxplot of virus load per population served by each treatment works of 22 treatment plants in Austria in November 2020. Right: maximize similarity approach: Plot of computed probabilities of infected persons for the optimal parameter  $L_{shed} = 10.15 \log_{10}$  Virus/day/Person. The shaded area is the overlap (similarity index) with the uniform distribution of the measurement (CI of the prevalence survey in November 2020).



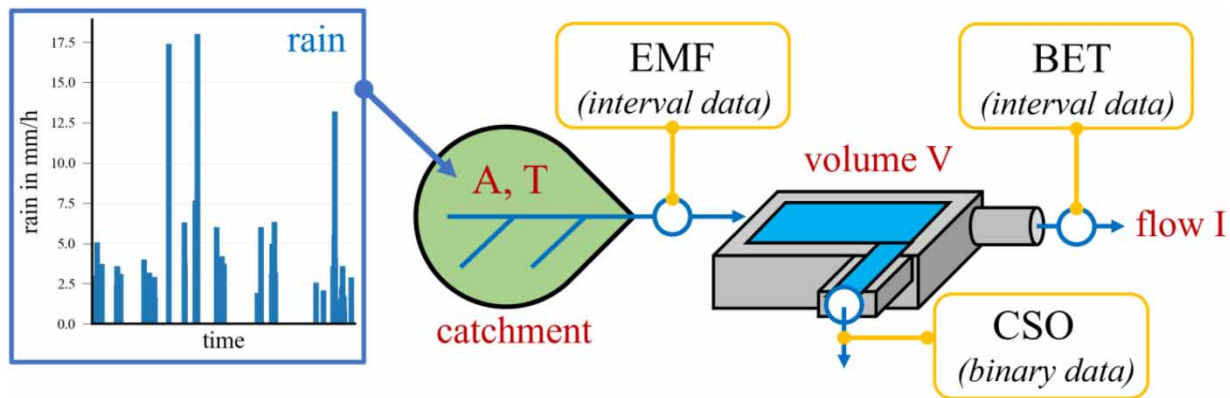
**Figure 3** | Wastewater-based epidemiology – ABC approach: Plot of computed probabilities of accepted samples of parameter  $L_{shed}$  in  $\log_{10}$  Virus/day/Person. The highest probability ( $L_{shed} \sim 10.135$ ) matches the result of the Maximizing Similarity method ( $L_{shed} = 10.15$ ) quite well.

**Case study 2 – Catchment hydrology**

In the second example, we test the ABC approach for a synthetic four-parameter rainfall–runoff model, containing as parameters the impervious catchment area  $A$  in  $m^2$ , reservoir coefficient  $T$  in sec, retention basin volume  $V$  in  $m^3$  and maximum flow in the interceptor to the downstream treatment plant  $I$  in  $m^3/s$  (see Figure 4 Left). The model computes only rainfall–runoff – adding dry weather flow is conceptually simple but does not add to this study.

The model computes the hydrograph from the catchment by means of a linear reservoir (parameters  $A$  and  $T$ ) and the combined sewer overflow (CSO) structure is represented by means of a mass balance model (parameters  $V$  and  $I$ ). For implementation and application examples see Kleidorfer *et al.* (2009a, 2009b) and Dotto *et al.* (2011). As rain input, we use a 3-month period for calibration (see Figure 4 Right) and a 1-year rain series (different year, i.e., no overlap in the information) for verification.

For a reference situation the parameter values (see Table 1) are chosen to reflect a standard combined sewer catchment according to Austrian guidelines (Kleidorfer & Rauch 2011). Using this reference model, three times a series of imprecise observations are computed, which are then used as input in the calibration process. The idea is to use the ABC method to inverse estimate the reference parameter values from these observations. As prior information, the parameters are assumed to be uniformly distributed  $\Theta' \sim \mathcal{U}(\Theta_{LB}, \Theta_{UB})$  see Table 1 for lower/upper boundaries.



**Figure 4** | Left: Schematic of synthetic rainfall–runoff model – Calibration of four parameters (area  $A$ , reservoir coefficient  $T$ , basin volume  $V$ , and interceptor flow to the wastewater treatment ( $I$ )). Observations as imprecise information: EMF, CSO and BET. Right: Input data for calibration – precipitation over 3 months in mm/h.

**Table 1** | (a) Reference model parameter values, prior estimates (Lower-upper boundary of uniform distribution) and resulting values (median values of distribution) from calibration for different types of observations. (b) Metric of model performance for the verification situation

	Reference	Prior LB	Prior UB	CSO	EMF	BET	All
(a) Parameter							
$A$ in $m^2$	10000.0	5000.0	12000.0	9845.0	10337.0	9689.0	10499.0
$T$ in s	1800.0	1000.0	3000.0	2114.0	1757.0	1701.0	1735.0
$V$ in $m^3$	15.0	10.0	25.0	17.1	17.5	15.9	18.5
$I$ in $m^3/s$	0.00200	0.00100	0.00400	0.00168	0.00250	0.00166	0.00189
(b) Result							
NSE				0.745	0.708	0.787	0.841
RMSE				0.430	0.460	0.394	0.339
relDev Par.				0.122	0.119	0.079	0.093
reDev CSO				0.000	0.125	0.063	0.063

For imprecise measurement, we assume the following three data series of observation. Note that all three observations are not continuous time series data, but summarize information to a single data point per rain event. A rain event is defined by an antecedent dry weather period of 240 min:

- CSO – combined sewer overflow: This is the binary information if a CSO event has occurred for a rain event or not (Wani *et al.* 2017) which is fairly simple information to record (Kleidorfer *et al.* 2009a, 2009b).
- EMF – empirical maximum flow: If the runoff from the catchment exceeds a value of  $5 \cdot I$  (i.e., for extreme events) we assume that there is uncertain information available (relative error  $< 50\%$ ) on the maximum flow in the catchment during that rain event. Such information could be from a crowd source or simple surrogate sensors. For matching with reality, we further assume that the information is only recorded randomly in 80% of the cases.
- BET – basin emptying time: If the retention basin has filled up, one can easily get information on the time required for emptying, i.e., directly by using a water level sensor or by using indirect information such as pumping time (if used for basin emptying) or flow measurement at the inlet of a treatment plant (an inflow above the base level indicates basin emptying). The emptying time of the basin can be seen as a form of software sensor. We assume that the observation is recorded as soon as the emptying time exceeds 60 min, i.e., if the basin is approximately half full. This threshold is chosen arbitrarily but with the intention to limit the signal to significant events. We further assume that the observation is inaccurate with a deviation of 60 min, i.e., a sample is accepted if  $|BET_m - BET_O| < 60 \text{ min}$

Table 1 shows the result of the model parameter estimation effort for the types of observations available (CSO, EMF and BET) where ALL denotes the situation when all three observations are used. Calibration is done for the calibration rain series. Model performance is then computed by using the resulting parameter set (mean values of posterior samples) and simulating for the verification rain series. For performance metrics we use RMSE and NSE of the CSO time series, mean relative deviation of the parameters and relative deviation from the true number of CSO events. It is interesting to note that adequate calibration results can be achieved already for each of the observations in itself. Consequently, imprecise information, e.g., binary information on CSO events, is a valuable source of information for parameter estimation.

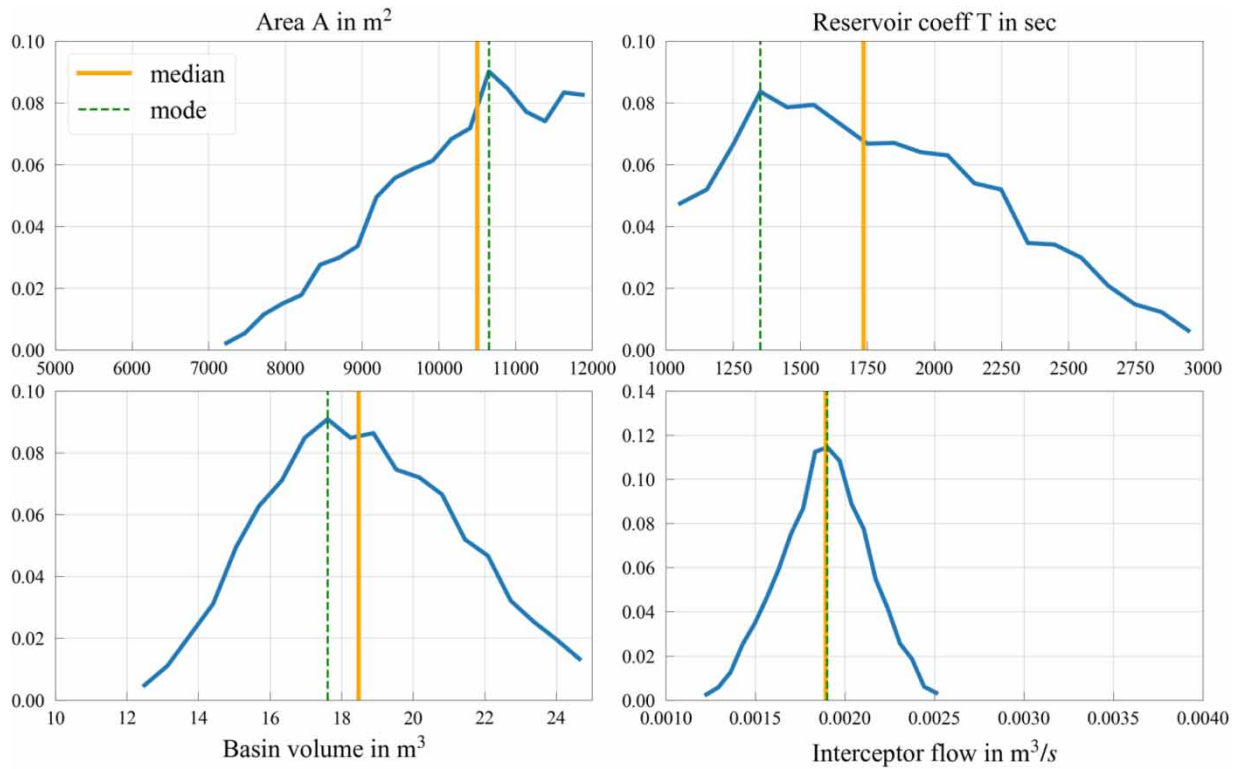
Figure 5 exemplifies the distributions of the accepted parameters in the ABC routine for the ALL situation as well as median and mode values. The boundaries of the x-axis reflect the lower/upper boundaries of the uniform distribution of the prior information. (Note that the boundaries of the uniform distribution  $\mathcal{U}(\theta_{LB}, \theta_{UB})$  are chosen arbitrarily in such a way that the true values are not their mean.) The probability density functions allow interpretation of the identifiability of each parameter – the more uniform the distribution, the less identifiable. Figure 5 also shows the problem of selecting the resulting set of parameter values  $\theta_R$ : Imprecise information inherently does not allow to formulate a summary statistic in a practical situation (the metric of model performance in Table 1 can be only computed as we know the reference system). In this case, we cannot use one unifying metric to identify the most fitting parameter set but need to select a suitable value in the posterior probability density function of each parameter, e.g., median or mode. For this case study it was found that the median value gives slightly better results for the parameter estimation and has thus been used throughout. An alternative option would be to use the entire parameter distribution(s) in the forward modeling approach, get a distribution of model outputs and calculate statistics on model output (median, selected percentiles).

### Significance of observations by means of Shapley values

Unlike standard calibration situations, there is usually more than one type of observation available when using imprecise information for model calibration. However, for nonlinear systems, it is not immediately obvious if a certain type of observation is actually contributing to the calibration or not. It is thus important to identify which observations support the parameter estimation and which are of little importance. The key problem for this estimation of significance is again related to the absence of a summary statistic: When using imprecise information, we cannot readily identify one unifying metric to express the system performance. E.g., when using the observation type CSO, all accepted samples must reproduce the exact number of CSO events – otherwise, they would have been rejected – and thus we have no means to determine the best fit. For EMF and BET the situation is alike but there is an additional random element included for the rejection.

One possible option to overcome this inherent obstacle of imprecise information is to use all observation types available. Using again the CSO example, an error is then defined as the number of CSO events that were incorrectly classified. As an appropriate metric, we use here *NRMSE* – normalized root mean square error – where *NE* denotes the number of events that are identified in each of the observation types. Subscript *obs* stands for observation and *m* for model. In this case study we





**Figure 5** | Probability density function of the accepted samples in the ABC scheme for the situation ALL (using CSO, EMF and BET).

investigate three observation types, thus  $n = [1, \dots, n_{Obs}]$  and  $n_{Obs} = 3$ .

$$NRMSE = \sqrt{\frac{1}{n_{Obs}} \sum_{n=1}^{n_{Obs}} (NE_{Obs}(n) - NE_m(n))^2} \cdot \frac{n_{Obs}}{\sum_{n=1}^{n_{Obs}} NE_{Obs}(n)}$$

In the following, we compute the Shapley values for each of the three observation types (CSO, EMF and BET) for the metric  $1 - NRMSE$  (thus turning the metric as positive in the interval  $[0 - 1]$ ). Table 2(b) lists all possible combinations of the three types, with the marginal contribution indicating the increase in the metric if a player (here observation type) is added. e.g., for the combination EMF-CSO-BET (3rd line in Table 2(b)), the marginal for the observation type CSO is computed as metric for the combination EMF-CSO (0.872 – fourth line in Table 2(a)) minus the combination EMF (0.876 – 2nd line in Table 2(a)), thus  $-0.004$ . The Shapley value is then the mean of all marginal values. Table 2(b) indicates that EMF and BET are giving slightly better results than CSO and should be preferred as observation input in this case study. The negative values in the marginal distributions of the CSO observation further indicate that the joint use of CSO with both other observation types is subpar to the use of only EMF and BET. Note that this information is gained by only analyzing the accepted samples in the ABC scheme.

In order to test the result of the scheme a scenario analysis is conducted for the different combinations of observation types using the information from the calibration effort. Here we use the metric  $NSE$  computed for the time series of the CSO – reference versus calibrated system and test for the same combinations as for the Shapley values. Table 2(a) corroborates that the best results are achieved when applying only EMF and BET observations. From this comparison, we conclude that Shapley values can contribute to identifying the significance of a certain type of observation in a parameter calibration scheme – with only the imprecise information that is available. The advantage of the Shapley value approach as compared to scenario analysis is that all possible combinations are assessed in a formal setting.

**Table 2** | (a) Metrics of the parameter estimation 1-NRMSE – computed from posterior samples and NSE computed for CSO time series – reference versus calibrated system. (b) Shapley values computed for the metric 1-NRMSE

(a) Combination	1-NRMSE	NSE	
CSO	0.797	0.745	
EMF	0.876	0.698	
BET	0.833	0.789	
CSO-EMF	0.872	0.781	
CSO-BET	0.861	0.828	
EMF-BET	0.910	0.866	
CSO-EMF-BET	0.904	0.845	
(b) Combination	Marginal CSO	Marginal EMF	Marginal BET
CSO-EMF-BET	0.797	0.075	0.032
CSO-BET-EMF	0.797	0.043	0.064
EMF-CSO-BET	- 0.004	0.876	0.032
EMF-BET-CSO	- 0.006	0.876	0.034
BET-CSO-EMF	- 0.015	0.043	0.833
BET-EMF-CSO	- 0.006	0.077	0.833
Mean	0.261	0.332	0.305

## RECOMMENDATION

For parameter estimation with imprecise information, the following sequence of steps is recommended:

- (1) Irrespective of the following procedure, prior information (lower/upper bounds of values, distribution) on the parameters is to be determined. Likewise, information on the type of available imprecise information is to be gathered as well as information on the measured data and its quality (time series of information).
- (2) Determine if the problem can be addressed by using the classical concept of maximizing similarity, i.e., if the imprecise information can be easily expressed in statistical terms. Only if this is not the case, use ABC.
- (3) For maximizing similarity, the underlying minimization procedure results in a clear estimate of the parameter values. On the other hand, for ABC the result is a sample of acceptable solutions. The actual values of the parameters have then to be determined from this sample, usually by addressing key indicators of the probability density function of the values.
- (4) If several types of imprecise information are available, it is recommended to check which combination of the information gives the best results. Intuitively this can be done via scenario analysis, i.e., testing for each combination of information sources, but the use of Shapley values is advocated as providing a formal and established framework.

## CONCLUSION

Model parameter estimation frequently suffers from the lack of precise information. Traditional calibration methods need to be enhanced to cope with such uncertainties. While there are different options available, the likelihood-free ABC is a simple, yet efficient method to deal with imprecise information. In this paper, we demonstrated:

- the result of the ABC method matches the reference method of maximizing similarity for the real-world case study ‘waste-water-based epidemiology’
- good calibration results can be achieved with ABC for a four-parameter rainfall–runoff model if binary and/or censored information are available. The quality of the parameter estimation is computed for a verification situation.
- 3 different examples of imprecise information are exemplified, i.e., binary data on CSO events, censored data on maximum flow in the catchment and emptying time of the basin.
- even if the quality of the calibration cannot be estimated accurately for imprecise information, Shapley values of the posterior distribution of accepted samples in the ABC scheme indicate the significance of a specific type of observation

- a shortcoming is the computational burden in the Monte-Carlo-based procedure. This limits applications to fast models and/or requires the use of efficient algorithms and parallel computing.

## DATA AVAILABILITY STATEMENT

Data cannot be made publicly available; readers should contact the corresponding author for details.

## CONFLICT OF INTEREST

The authors declare there is no conflict.

## REFERENCES

- Abbaspour, K. C., van Genuchten, M. T., Schulin, R. & Schlaeppi, E. 1997 A sequential uncertainty domain inverse procedure for estimating subsurface flow and transport parameters. *Water Resources Research* **33**, 1879–1892.
- Abbaspour, K. C., Yang, J., Maximov, I., Siber, R., Bogner, K., Mieleitner, J., Zobrist, J. & Srinivasan, R. 2007 Modelling hydrology and water quality in the pre-alpine/alpine Thur watershed using SWAT. *Journal of hydrology* **333** (2–4), 413–430.
- Aster, R. C., Borchers, B. & Thurber, C. H. 2018 *Parameter Estimation and Inverse Problems*. Elsevier, Amsterdam, The Netherlands.
- Beck, J. V. & Arnold, K. J. 1977 *Parameter Estimation in Engineering and Science*. John Wiley and Sons, New York.
- Beven, K. 2010 *Environmental Modelling: An Uncertain Future?* CRC Press, Boca Raton, Florida.
- Beven, K. & Binley, A. 1992 The future of distributed models: Model calibration and uncertainty prediction. *Hydrological Processes* **6**, 279–298.
- Casal-Campos, A., Sadr, S. M. K., Fu, G. & Butler, D. 2018 Reliable, resilient and sustainable urban drainage systems: An analysis of robustness under deep uncertainty. *Environmental Science & Technology* **52** (16), 9008–9021.
- 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. & Tait, S. 2012 Assessing uncertainties in urban drainage models. *Physics and Chemistry of the Earth, Parts A/B/C* **42–44**, 3–10.
- Dotto, C. B. S., Kleidorfer, M., Deletic, A., Rauch, W., McCarthy, D. T. & Fletcher, T. D. 2011 Performance and sensitivity analysis of stormwater models using a Bayesian approach and long-term high resolution data. *Environmental Modelling & Software* **26** (10), 1225–1239.
- Dotto, C. B. S., Mannina, G., Kleidorfer, M., Vezzaro, L., Henrichs, M., McCarthy, D. T., Freni, G., Rauch, W. & Deletic, A. 2012 Comparison of different uncertainty techniques in urban stormwater quantity and quality modelling. *Water Research* **46** (8), 2545–2558.
- Garzón, A., Kapelan, Z., Langeveld, J. & Taormina, R. 2022 Machine learning-based surrogate modeling for urban water networks: Review and future research directions. *Water Resources Research* **58** (5), e2021WR031808.
- Gijbels, I. 2010 Censored data. *Wiley Interdisciplinary Reviews: Computational Statistics* **2** (2), 178–188.
- Gupta, H. V., Beven, K. J. & Wagener, T., 2005 Model calibration and uncertainty estimation. In: *Encyclopedia of Hydrological Sciences* (Anderson, M. G., ed.). Wiley, New York, pp. 2015–2031.
- Hajihassanpour, M., Kesserwani, G., Pettersson, P. & Bellos, V. 2023 Sampling-based methods for uncertainty propagation in flood modeling under multiple uncertain inputs: Finding Out the most efficient choice. *Water Resources Research* **59** (7), e2022WR034011.
- Higdon, D., Kennedy, M., Cavendish, J. C., Cafeo, J. A. & Ryne, R. D. 2004 Combining field data and computer simulations for calibration and prediction. *SIAM Journal on Scientific Computing* **26** (2), 448–466.
- Kavetski, D., Fenicia, F., Reichert, P. & Albert, C. 2018 Signature-domain calibration of hydrological models using approximate Bayesian computation: Theory and comparison to existing applications. *Water Resources Research* **54** (6), 4059–4083.
- Ke, Y. & Tian, T. 2019 Approximate Bayesian Computational Methods for the Inference of Unknown Parameters. 2017 *MATRIX annals*, 515–529.
- Kleidorfer, M. & Rauch, W. 2011 An application of Austrian legal requirements for CSO emissions. *Water Science and Technology* **64** (5), 1081–1088.
- Kleidorfer, M., Deletic, A., Fletcher, T. D. & Rauch, W. 2009a Impact of input data uncertainties on urban stormwater model parameters. *Water Science and Technology* **60** (6), 1545–1554.
- Kleidorfer, M., Möderl, M., Fach, S. & Rauch, W. 2009b Optimization of measurement campaigns for calibration of a conceptual sewer model. *Water Science and Technology* **59** (8), 1523–1530.
- Kourtis, I. M. & Tsihrintzis, V. A. 2021 Adaptation of urban drainage networks to climate change: A review. *Science of The Total Environment* **771**, 145431.
- Lundberg, S. M. & Lee, S. I. 2017 A unified approach to interpreting model predictions. *Advances in Neural Information Processing Systems* **30**, 4765–4774.
- Madsen, H., Wilson, G. & Ammentorp, H. C. 2002 Comparison of different automated strategies for calibration of rainfall-runoff models. *Journal of Hydrology* **261** (1–4), 48–59.
- Marjoram, P., Molitor, J., Plagnol, V. & Tavaré, S. 2003 Markov chain Monte Carlo without likelihoods. *Proceedings of the National Academy of Sciences of the United States of America* **100** (26), 15324–15328.

- Marquardt, D. W. 1963 An algorithm for least-squares estimation of nonlinear parameters. *Journal of the Society for Industrial and Applied Mathematics* **11** (2), 431–441.
- McMillan, H. K., Westerberg, I. K. & Krueger, T. 2018 Hydrological data uncertainty and its implications. *Wiley Interdisciplinary Reviews: Water* **5** (6), e1319.
- Nott, D. J., Marshall, L. & Brown, J. 2012 Generalized likelihood uncertainty estimation (GLUE) and approximate Bayesian computation: What's the connection? *Water Resources Research* **48** (12). <http://dx.doi.org/10.1029/2011WR011128>.
- Otto, M. & Bandemer, H. 1986 Calibration with imprecise signals and concentrations based on fuzzy theory. *Chemometrics and Intelligent Laboratory Systems* **1** (1), 71–78.
- Pastore, M. & Calcagni, A. 2019 Measuring distribution similarities between samples: A distribution-free overlapping index. *Frontiers in Psychology* **10**, 1089.
- Pritchard, J. K., Seielstad, M. T., Perez-Lezaun, A. & Feldman, M. W. 1999 Population growth of human Y chromosomes: A study of Y chromosome microsatellites. *Molecular Biology and Evolution* **16** (12), 1791–1798.
- Rauch, W., Schenk, H., Insam, H., Markt, R. & Kreuzinger, N. 2022 Data modelling recipes for SARS-CoV-2 wastewater-based epidemiology. *Environmental Research* **214**, 113809.
- Rauch, W., Schenk, H., Rauch, N., Harders, M., Oberacher, H., Insam, H., Markt, R. & Kreuzinger, N. 2024 Estimating actual SARS-CoV-2 infections from secondary data. *Scientific Reports* **14** (1), 6732.
- Refsgaard, J. C. 1997 Parameterisation, calibration and validation of distributed hydrological models. *Journal of Hydrology* **198** (1–4), 69–97.
- Sankararaman, S. & Mahadevan, S. 2012 Model parameter estimation with imprecise and unpaired data. *Inverse Problems in Science and Engineering* **20** (7), 1017–1041.
- Scheidegger, A., Hug, T., Rieckermann, J. & Maurer, M. 2011 Network condition simulator for benchmarking sewer deterioration models. *Water Research* **45** (16), 4983–4994.
- Shapley, L. S. 1953 A value for n-person games. In: *Contributions to the Theory of Games II* (Kuhn, H. W. & Tucker, A. W., eds). Princeton University Press, Princeton, pp. 307–317.
- Šimůnek, J. & De Vos, J. A. 1999 Inverse optimization, calibration and validation of simulation models at the field scale. In: *Modelling Transport Processes in Soils at Various Scales in Time and Space* (Feyen J. & Wiyo K., eds.). Wageningen Pers, Wageningen, The Netherlands, pp. 431–445.
- Sisson, S. A., Fan, Y. & Tanaka, M. M. 2007 Sequential Monte Carlo without likelihoods. *Proceedings of the National Academy of Sciences of the United States of America* **104** (6), 1760–1765.
- Seibert, J. & McDonnell, J. J. 2002 On the dialog between experimentalist and modeler in catchment hydrology: Use of soft data for multicriteria model calibration. *Water resources research* **38** (11), 23–1.
- Sundararajan, M. & Najmi, A. 2020 The many Shapley values for model explanation. In *International Conference on Machine Learning*. PMLR, pp. 9269–9278.
- Tarantola, A. 2005 *Inverse Problem Theory and Methods for Model Parameter Estimation*. SIAM. ISBN 978-0-89871-792-1.
- Tavaré, S., Balding, D., Griffith, R. & Donnelly, P. 1997 Inferring coalescence times from DNA sequence data. *Genetics* **145**, 505–518.
- Todini, E. 2008 A model conditional processor to assess predictive uncertainty in flood forecasting. *International Journal of River Basin Management* **6** (2), 123–137.
- Tscheikner-Gratl, F., Zeisl, P., Kinzel, C., Leimgruber, J., Ertl, T., Rauch, W. & Kleidorfer, M. 2016 Lost in calibration: Why people still do not calibrate their models, and why they still should—a case study from urban drainage modelling. *Water Science and Technology* **74** (10), 2337–2348.
- Tscheikner-Gratl, F., Bellos, V., Schellart, A., Moreno-Rodenas, A., Muthusamy, M., Langeveld, J., Clemens, F., Benedetti, L., Rico-Ramirez, M. A., de Carvalho, R. F., Breuer, L., Shucksmith, J., Heuvelink, G. B. M. & Tait, S. 2019 Recent insights on uncertainties present in integrated catchment water quality modelling. *Water Research* **150**, 368–379.
- Vrugt, J. A. & Sadegh, M. 2013 Toward diagnostic model calibration and evaluation: Approximate Bayesian computation. *Water Resources Research* **49** (7), 4335–4345.
- Wang, C. 2019 Evidence-theory-based uncertain parameter identification method for mechanical systems with imprecise information. *Computer Methods in Applied Mechanics and Engineering* **351**, 281–296.
- Wani, O., Scheidegger, A., Carbajal, J. P., Rieckermann, J. & Blumensaat, F. 2017 Parameter estimation of hydrologic models using a likelihood function for censored and binary observations. *Water Research* **121**, 290–301.

First received 31 January 2024; accepted in revised form 27 May 2024. Available online 11 June 2024