Parameter Estimation and Predictive Uncertainty Quantification in Hydrological Modelling

Dmitri Kavetski

Contents

D. Kavetski (✉)
School of Civil, Environmental and Mining Engineering, University of Adelaide, Adelaide, SA, Australia

School of Engineering, University of Newcastle, Callaghan, NSW, Australia

Department of Systems Analysis, Integrated Assessment and Modelling (SIAM), Eawag, Swiss Federal Institute of Aquatic Science and Technology, Dübendorf, Switzerland
e-mail: dmitri.kavetski@adelaide.edu.au

Q. Duan et al. (eds.), *Handbook of Hydrometeorological Ensemble Forecasting*,
https://doi.org/10.1007/978-3-642-40457-3_25-1

Abstract

The majority of hydrological and environmental models contain parameters that must be specified before the model can be used. Parameter estimation is hence a very common problem in environmental sciences and has received tremendous amount of research and industry attention. This chapter reviews some of the key principles of parameter estimation, with a focus on calibration approaches and uncertainty quantification. The distinct approaches of manual calibration, optimization, multi-objective optimization, and probabilistic approaches are described in terms of key theory and representative applications. Advantages and limitations of these strategies are listed and discussed, with a focus on their ability to represent parametric and predictive uncertainties. The role of posterior diagnostics to check calibration and model assumptions that impact on parameter estimation is emphasized. Auxiliary tricks and techniques are described to simplify the process of parameter estimation in practical applications. The chapter concludes with an outline of directions for ongoing and future research. It is hoped that this chapter will help hydrologists and environmental modellers get to the current state of research and practice in model calibration, parameter estimation, and uncertainty quantification.

Keywords

Hydrological model · Parameter estimation · Model calibration · Optimization · Bayesian inference · Uncertainty quantification

1 Introduction

Hydrological (rainfall-runoff) models are widely used in environmental sciences and engineering, including flood forecasting, water yield predictions, and so forth (e.g., Duan et al. 1992; Beven 1997; Lindstrom et al. 1997; Clark et al. 2008, 2015). In addition to being useful in their own right, predictions from hydrological models, particularly rainfall-runoff models, provide inputs to the planning and operation of water resource systems (Loucks et al. 1981). The scales of these applications vary from a single hillslope to entire continents (Archfield et al. 2015), and the prediction lead times vary from minutes in operational flood forecasting (Neal et al. 2012) to seasonal scales (Tuteja et al. 2011). Given the inherent uncertainty of environmental predictions, uncertainty quantification and risk assessment is another key aspect that is receiving increased attention in the literature (e.g., Vogel 2017; Reichert et al. 2015).

Hydrological models are often classified on a spectrum from black-box models to conceptual models to physical models. Typical black-box models are given by artificial neural networks (e.g., Govindaraju 2000; Kingston et al. 2008); physical "bottom-up" models can be defined as models based on contemporary understanding of physical laws (e.g., Freeze and Harlan 1969; Ivanov et al. 2004; Clark et al. 2015), with conceptual "top-down" models (e.g., Sivapalan et al. 2003a; Fenicia et al. 2011) somewhere in between these bookends. This classification provides useful guidance but is not always crisp, with most practical models not fitting neatly into a single category and instead exhibiting a mix of different modelling philosophies (e.g., Clark et al. 2011). Irrespective of their philosophy and mathematics, hydrological and environmental models almost always contain adjustable parameters, intended to describe the invariant properties of the system. Before a model can be used for simulation or prediction of a system of interest, its parameters must be specified.

This chapter deals with the problem of parameter estimation and uncertainty quantification. Broadly speaking, two types of estimation strategies can be distinguished: a priori and calibration (inverse modelling). A priori estimation seeks to assign parameter values based directly on observable physical quantities, e.g., soil properties, vegetation characteristics, and so forth (e.g., Koren et al. 2003). Calibration, in this work, refers to any procedure for estimating model parameters (and possibly their uncertainties) from available observations of quantities the model is supposed to predict (e.g., Tarantola 2005). A priori estimation tends to depend on model structure and physical basis (e.g., see the debates in Abbott et al. 2003; Pappenberger and Beven 2006). Calibration in this respect represents a more general mathematical operation. In principle any model can – or, as many have argued, should – be calibrated, yet in practice it is often a formidable challenge to calibrate a model suitable for extrapolation and reliably account for estimation uncertainties. An important subset of parameter estimation is estimation under data-scarce conditions, including in ungauged basins – these applications tend to use a combination of a priori estimation and calibration (e.g., see Hrachowitz et al. 2013).

In recognition of these challenges, parameter estimation in hydrology and environmental modelling is shifting from a reliance on manual expertise, especially in research applications – initially to the largely mathematical task of finding (optimizing) the "best" model parameters according to a given performance metric (e.g., Ibbitt and O'Donnell 1971; Gupta and Sorooshian 1985; Duan et al. 1992) – and ultimately to a more "holistic" treatment that seeks to reflect multiple competing objectives in model calibration (e.g., Gupta et al. 1998; Efstratiadis and Koutsoyiannis 2010), multiple sources of uncertainties (e.g., Beven and Binley 1992; Kavetski et al. 2002; Reichert and Mieleitner 2009; Renard et al. 2011), stringent diagnostics of model "realism" (e.g., Gupta et al. 2008; Clark et al. 2011), operational reliability (e.g., Krzysztofowicz 1999; Cloke and Pappenberger 2009; Wang et al. 2009; McInerney et al. 2017), and so forth.

The aims of this chapter are to review the main types of parameter estimation methods with a focus on calibration, to provide a rigorous but accessible summary of key ideas and methods, and to direct the interested reader to the rich scientific and operational literature. In the author's experience, there is often a disconnect between the intuitive objective function techniques used by practitioners and the statistically motivated likelihood functions used in the research literature. This chapter attempts to close this gap and provides a unified perspective that ties together seemingly distant techniques such as single- and multi-objective optimization, Bayesian inference, residual error diagnostics, and numerical model implementation aspects. Emphasis is placed on technical aspects and practical recommendations, including discussions of pros and cons of individual techniques. However, philosophical aspects are also relevant, and the practitioner should be aware of the types of assumptions being made and limitations arising thereof.

The chapter is structured as follows. Section 2 establishes the notation and background of parameter estimation, including a brief review of manual calibration and goodness-of-fit functions. Section 3 motivates automatic calibration – using digital computers rather than humans – and focuses on the optimization approach and its advantages, challenges, and limitations. Section 4 considers multiple competing objectives. Section 5 surveys the vast topic of probabilistic estimation and uncertainty quantification, with a focus on Bayesian techniques. Section 6 considers the critical topic of posterior diagnostics to check calibration assumptions. Section 7 picks up practicalities relevant to application, Sect. 8 describes ongoing research directions, and Sect. 9 wraps up with conclusions.

2 Basic Concepts of Parameter Estimation

2.1 Basic Setup of the Calibration Problem

A hydrological model $\mathcal{H}(\mathbf{x};\, \boldsymbol{\theta})$ simulates catchment streamflow $\mathbf{y}$ over a series of time steps t given forcing data $\mathbf{x}$ and parameters $\boldsymbol{\theta}$:

$$\mathbf{y} = \mathcal{H}(\mathbf{x};\boldsymbol{\theta}) \tag{1}$$

To keep notation simple for presentation purposes, it will be assumed that $\mathcal{H} = \{\mathcal{H}_t,\, t = 1, .., N_{\mathcal{H}}\}$ is a vector of length equal to the number of time steps in the forcing data $\mathbf{x} = \{\mathbf{x}_t,\, t = 1, .., N_{\mathbf{x}}\}$, i.e., $N_{\mathcal{H}} = N_{\mathbf{x}}$. A more general (but less transparent) notation can be deployed if the responses include streamflow at locations other than the catchment outlet, water depth levels, water quality, and so forth. Typical forcing required by hydrological models includes rainfall, potential evaporation, irrigation schedules, pumping schedules, and so forth. Most models contain internal states, typically storages across a collection of storage elements (conceptual models) or grid cells/finite elements (physical models discretized in space) (e.g., Singh and Woolhiser 2002; Fenicia et al. 2011; Clark et al. 2015, and many others).

In addition to inputs, outputs, and internal states, which are typically variables, models contain *parameters*, which are quantities intended to characterize the inherent properties of the modelled system (including the physical system *and* the observational system used to collect data). In Eq. (1), parameters are indicated as $\boldsymbol{\theta} = \{\theta_k,\, k = 1, .., N_{\boldsymbol{\theta}}\}$.

Parameters are typically defined subject to lower and upper bounds

$$\boldsymbol{\theta}^{(L)} \leq \boldsymbol{\theta} \leq \boldsymbol{\theta}^{(H)} \tag{2}$$

or more complex linear and nonlinear constraints (e.g., if parameters are mutually constrained).

The numbers of parameters in hydrological models vary widely. Lumped conceptual models, such as GR4J (Perrin et al. 2003), have just a handful of parameters intended to describe max storage values, routing characteristics, and groundwater exchange. Distributed physically based models such as SWAT (Arnold and Fohrer 2005) may have hundreds of parameters, describing soil hydraulic properties (such as conductivity), surface lags, crop growth rates, and so forth. The distinction between model parameters and states is not always clear-cut, and it frequently depends on the context in which the variables appear. For example, in some applications, parameters are defined as time- and/or state-dependent (e.g., Young 1998; Kuczera et al. 2006; Reichert and Mieleitner 2009; Young and Ratto 2009; Westra et al. 2012). Note that the term "model parameterization" is sometimes used to refer to the form of the model equations and their parametric dependencies and other times to the actual parameter values, which can cause confusion.

When parameter values are unknown, a model will generally be unable to reproduce even known data, let alone future unknown data. Hence, parameter estimation is among the first steps of deploying a model.

The following sections describe the two main parameter estimation strategies, namely, a priori estimation and calibration.

2.2 A Priori Estimation

A priori estimation refers to establishing parameter values from measured physical system properties. This strategy presupposes that model parameters have a sufficiently reliable physical interpretation (Abbott et al. 2003; Ivanov et al. 2004). For example, consider the specification of channel geometry in flood models – in the case of engineered channels, their width and length can be usually established from maps and other records. Parameter estimation in models of natural systems may require measurements and tests. For example, the hydraulic conductivity of soils, a parameter within physically based groundwater models such as MODFLOW (Harbaugh 2005), may be obtained from laboratory analysis of core samples, in situ tests, and/or geology maps (Fetter 1994). Other examples of a priori estimation might include the specification of channel roughness in hydraulic models using Manning's equation (e.g., Streeter and Wylie 1983).

A priori estimation can be effective, especially when modelling well-instrumented locations using equations that embody our best current understanding of environmental physics (e.g., Ivanov et al. 2004; Clark et al. 2015). In contrast, when working with conceptual models, it has proven difficult to reliably relate their parameters to available information (Koren et al. 2003; Duan et al. 2006), though in some cases, useful relationships appear possible (Samaniego et al. 2010). Parameter estimation from observable catchment characteristics is challenged by the tremendous spatial variability of soils and vegetation, both within and across basins (Miller and White 1999), as well as by the frequent problem of non-commensurability of modelled and observed quantities (Kuczera and Franks 2002). Another question relates to the estimation of parameter and predictive uncertainties; model structural errors are particularly difficult to estimate a priori without recourse to at least some observed data. Advances in physical process representation notwithstanding, it has been argued that models can only be described as "truly" physical if their parameters are specified independently from observed responses (Grayson et al. 1992). That said, even quantities currently seen to have a firm physical basis, such as Darcian hydraulic conductivity, were established empirically by fitting to experimental data (Brown 2002) – it can hence be argued that all practical environmental models begin their life as empirical quantities. This observation leads us to the general class of parameter estimation given by calibration (inverse modelling).

2.3 Calibration

The idea of model calibration is to find parameter values $\boldsymbol{\theta}^{(\mathrm{cal})}$ such that, given a set of observed ("known") inputs $\tilde{\mathbf{x}}$, the model $\mathcal{H}$ reproduces a set of known outputs $\tilde{\mathbf{y}} = \left\{\tilde{y}_t, t = 1,..,N_{\tilde{\mathbf{y}}}\right\}$, at least to a degree that is satisfactory for the application of interest. In algorithmic/mathematical notation:

$$\begin{aligned} \text{Calibration : Find } \boldsymbol{\theta}^{(\mathrm{cal})} \text{ such that } \mathcal{H}\left(\tilde{\mathbf{x}};\boldsymbol{\theta}^{(\mathrm{cal})}\right) \approx \tilde{\mathbf{y}} \\ \boldsymbol{\theta}^{(\mathrm{cal})} : \mathcal{H}\left(\tilde{\mathbf{x}};\boldsymbol{\theta}^{(\mathrm{cal})}\right) \approx \tilde{\mathbf{y}} \end{aligned} \tag{3}$$

Figure 1, in conjunction with Eq. (3), illustrates the basics of going from an uncalibrated to a calibrated model. While superficially simple, hydrological model calibration is a rather challenging task, especially once we recognize that a perfect model fit is unattainable and wish to characterize the attendant trade-offs and uncertainties. Different calibration approaches are then distinguished by aspects such as (i) how is the approximate equality in Eq. (3) expressed mathematically or even visually, (ii) how is the search for parameters conducted, (iii) how many sets of estimated parameter values are retained (e.g., to represent uncertainty), and so forth.

2.4 Manual Calibration

A hydrologist or engineer familiar with the model and catchment system of interest will often be able to find parameter values for which the model behaves in a reasonable way. For example, when working with a flood model, an engineer will generally try to match the flood peak magnitude, the total flow volume, and ideally

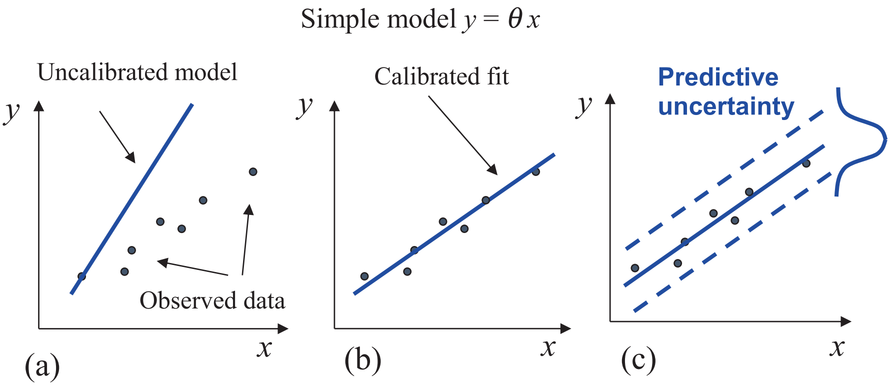

Fig. 1 Calibration concepts illustrated using a simple straight-line model. Panel *a* shows the uncalibrated model (poor value of model parameter θ), panel *b* shows the same model with a calibrated parameter θ, and panel *c* shows a more comprehensive application where model uncertainty is quantified

the timing (e.g., Maidment 1993). Characteristics such as the shape of recession would also be considered, though usually to a lesser extent as these aspects have less impact on flood damages. On the other hand, a water supply engineer working in an arid area may be far less interested in flood peaks but will try to get the model to reproduce low flows, which often contribute the most to cumulative flow volumes and hence to water availability. Finally, a hydrologist interested in understanding catchment dynamics may very well focus on the shape of recession, e.g., using master recession analysis (Tallaksen 1995).

Manual calibration allows the modellers to exploit their experience and hydrological understanding – which are formidable tools in the hands of an expert (Savenije 2009; Hrachowitz et al. 2014). Trade-offs in the ability of the model to reproduce different aspects of the data can be resolved based on the application objectives, once again exploiting expert judgment where available.

On the other hand, the subjectivity of manual calibration also creates inevitable weaknesses and limitations. Most notably, how do we establish if parameter set $\boldsymbol{\theta}^{(1)}$ is closer or further away from $\boldsymbol{\theta}^{(\mathrm{cal})}$ than parameter set $\boldsymbol{\theta}^{(2)}$? The eye of an experienced modeller can provide superb expert judgment, but the resulting non-transparency and irreproducibility pose problems, both in scientific and operational contexts (Hill et al. 2015). Nor is it obvious how should a hydrologist quantify and report the uncertainty in manually calibrated parameters, especially if these parameters have been selected on the basis of a fit to visual hydrograph characteristics.

The laboriousness of manual calibration is another major practical downside – a human must select parameter values, run the model, inspect its output, suggest a new trial parameter set, rinse and repeat, and eventually decide when to stop. Clearly this is not only subjective but exhausting. Once again, these limitations become more pronounced in the case of large-scale national forecasting services, e.g., the US National Weather Service (NWS) (Demargne et al. 2014), the Australian Bureau of

Meteorology (Tuteja et al. 2017), and other agencies tasked with modelling and forecasting over hundreds and thousands of catchments spanning national- and continental-scale areas.

These limitations lead us to automatic calibration. But first we must solve the question of goodness-of-fit measures.

2.5 Goodness-Of-Fit Function as an Optimization Objective

The idea of a goodness-of-fit function Φ is to quantify how well the model reproduces the calibration data. Ideally we would like a perfect match of model to data, but we need to handle discrepancies in some reasonable way.

The most widely used goodness-of-fit function is the sum of squared errors (SSE), usually credited to Karl Gauss who developed it in the late eighteenth century (Merriman 1877):

$$\Phi_{\text{SSE}}(\boldsymbol{\theta}) = \Phi_{\text{SSE}}(\boldsymbol{\theta};\tilde{\mathbf{x}},\tilde{\mathbf{y}}) = \sum_{t=1}^{N_{\tilde{\mathbf{y}}}} \left(\tilde{y}_t - \mathcal{H}_t[\tilde{\mathbf{x}};\boldsymbol{\theta}]\right)^2 \tag{4}$$

where to avoid clutter, the dependence of Φ on $\tilde{\mathbf{x}}$ and $\tilde{\mathbf{y}}$ is omitted in the notation, and it is understood that, for time stepping models, $\mathcal{H}_t[\tilde{\mathbf{x}};\boldsymbol{\theta}]$ only depends on inputs $\tilde{\mathbf{x}}_{1:t}$ up to and including time step t. Intuitively, the SSE function penalizes discrepancies between model and observations in a "reasonable" way (a larger discrepancy at any time step lowers the goodness of fit) and has the historical advantage that it is easy to manipulate analytically.

Given a goodness-of-fit function, model calibration problem can be articulated as an optimization problem:

$$\boldsymbol{\theta}^{(\text{opt})} = \underset{\boldsymbol{\theta}}{\operatorname{argmin}}\, \Phi(\boldsymbol{\theta}) \tag{5}$$

subject to any external parameter constraints such as in Eq. (2).

In the optimization context given by Eq. (5), the goodness-of-fit function serves as the "objective" function – a naming convention that, perhaps unintentionally, hides that the choice of the error measure (e.g., the two-norm in Eq. (4)) is generally subjective. That said, Sect. 6 offers avenues to test these assumptions as part of the calibration process.

Figure 2 illustrates a typical least squares objective function of a hydrological model, shown as a cross section with respect to two parameters. Panel A provides an idealized schematic, with a well-defined optimum and smooth elliptic (quadratic) contours. Intuitively, the shape of the objective function indicates not just the optimal parameters but also parameter uncertainty: a peaky shape suggests well-defined parameters, whereas a flat shape indicates substantial uncertainty.

Figure 2 also illustrates parameter dependence – elongation of the objective function along certain parameter combinations. Parameter dependence is closely related to parameter identifiability: the objective function contours indicate

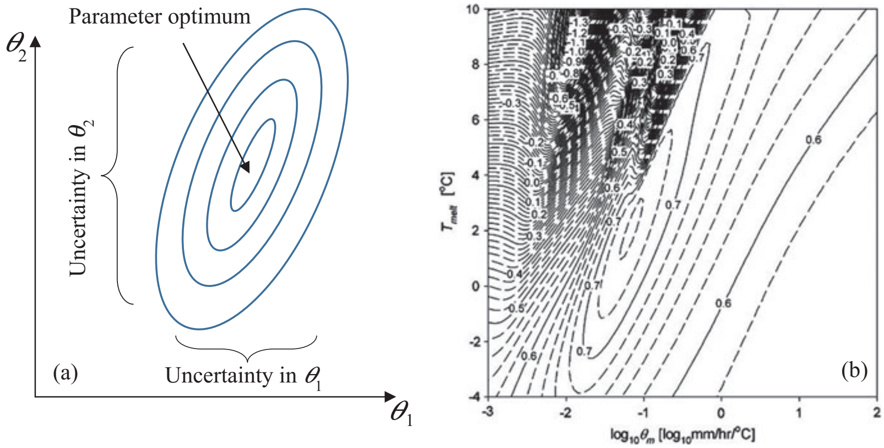

Fig. 2 Diagram of representative least squares objective functions. Panel *a* shows the idealized case of a two-parameter linear model, in which case the SSE objective function is exactly quadratic. Panel *b* (reproduced with permission from Elsevier) shows a real case example from the hydrological model case study of Kavetski et al. (2006c), where model nonlinearities lead to disturbances in the shape of the SSE

parameter sets that produce predictions "indistinguishable" from each other according to the goodness-of-fit function. For example, consider the straight-line model from Fig. 1: if we simultaneously increase its slope and reduce its intercept, the changes could compensate for each other and maintain the same goodness-of-fit value. Characterizing these parameter interactions can yield insight into model deficiencies and is an important goal of parameter calibration and uncertainty quantification.

The shape of SSE functions depends on the model $\mathcal{H}$. Eq. (4) indicates that, to the extent that a model is linear with respect to its parameters, i.e., $\partial^2\mathcal{H}/\partial\boldsymbol{\theta}^2 \approx \mathbf{0}$, its SSE objective function will be *quadratic* and have a single optimum irrespective of the data. In addition, a model that is smooth with respect to its parameters is guaranteed to have a smooth SSE objective function. In practice, most hydrological models exhibit nonlinearities, which induce irregularities in the objective function surface. Figure 2 panel b shows a well-behaved near-optimal region, as well as regions of irregular geometry and flat (insensitive) regions. In some instances, multiple optima can arise (e.g., Duan et al. 1992), which raises an even more immediate question of parameter identifiability than insensitive parameters.

The use of an objective function makes manual calibration more systematic, but its true power shines when used in conjunction with mathematical techniques such as optimization, which can find parameter optima analytically or numerically. For example, in the case of a straight-line model, $\mathcal{H}(x; a) = ax$, $\boldsymbol{\theta}^{(\text{opt})}$ is obtained analytically as

$$a^{(\mathrm{opt})} = \sum_{t=1}^{N_{\tilde{\mathbf{y}}}} \tilde{x}_t \tilde{y}_t \Big/ \sum_{t=1}^{N_{\tilde{\mathbf{y}}}} \tilde{x}_t^2 \tag{6}$$

Similar expressions exist for more general linear models. Nonlinear models, for which the SSE function cannot be optimized analytically, can be handled using numerical optimization (Sect. 3). Before considering these techniques, it is insightful to consider alternative objective functions.

2.6 Other Objective Functions: How Different Are They?

The SSE objective function makes intuitive sense but is not without some limitations. For example, its values (and units) are not easy to interpret or compare across time series of unequal length. Two SSE-derived functions are common in hydrology, the root mean squared error (RMSE) and the Nash-Sutcliffe efficiency (NSE).

The RMSE metric, widely used in engineering and physics, is defined as

$$\Phi_{\mathrm{RMSE}}(\boldsymbol{\theta}) = \sqrt{\frac{1}{N_{\tilde{\mathbf{y}}}} \sum_{t=1}^{N_{\tilde{\mathbf{y}}}} \left(\tilde{y}_t - \mathcal{H}_t[\tilde{\mathbf{x}};\boldsymbol{\theta}]\right)^2} = \sqrt{\frac{1}{N_{\tilde{\mathbf{y}}}} \Phi_{\mathrm{SSE}}(\boldsymbol{\theta})} \tag{7}$$

It offers the advantage of having the same units as the quantity of interest (e.g., m^3/s or mm/d in case of flowrates and catchment-average daily streamflow, respectively), as well as being scaled with respect to record length.

The NSE metric is a modification of the SSE function with a long tradition in hydrology (Nash and Sutcliffe 1970):

$$\Phi_{\mathrm{NSE}}(\boldsymbol{\theta}) = 1 - \frac{\sum_{t=1}^{N_{\tilde{\mathbf{y}}}} \left(\tilde{y}_t - \mathcal{H}_t[\tilde{\mathbf{x}};\boldsymbol{\theta}]\right)^2}{\sum_{t=1}^{N_{\tilde{\mathbf{y}}}} \left(\tilde{y}_t - \mathrm{ave}[\tilde{\mathbf{y}}]\right)^2} = 1 - b \times \Phi_{\mathrm{SSE}}(\boldsymbol{\theta}) \tag{8}$$

where $\mathrm{ave}[\tilde{\mathbf{y}}]$ is the sample mean of observed data. The asymptotic identity $\Phi_{\mathrm{NSE}}(\boldsymbol{\theta}) \underset{N\tilde{\mathbf{y}}\to\infty}{\to} 1 - \left(\frac{\Phi_{\mathrm{RMSE}}(\boldsymbol{\theta})}{\mathrm{sdev}[\tilde{\mathbf{y}}]}\right)^2$, where sdev[] denotes the standard deviation, elucidates that the NSE quantifies the fraction of streamflow variability captured by the hydrological model. Schaefli and Gupta (2007) suggest generalizing the NSE by replacing $\mathrm{ave}[\tilde{\mathbf{y}}]$ with a reference model, e.g., seasonal means, to provide a more informative and stringent benchmark.

The RMSE and NSE functions are related to the SSE kernel through monotonic transformations, and hence their optimal parameter sets (both local and global) are the same. Collectively, we shall refer to the optimization of these objective functions as least squares estimation. Connections to probabilistic estimation will be made in Sect. 5.2.

Genuinely different parameter estimates and objective function behavior are obtained with non-quadratic goodness-of-fit functions, e.g., the sum of absolute errors (SAE):

$$\Phi_{\mathrm{SAE}}(\boldsymbol{\theta}) = \sum_{t=1}^{N_{\tilde{\mathbf{y}}}} \left| \tilde{y}_t - \mathcal{H}_t[\tilde{\mathbf{x}};\boldsymbol{\theta}] \right| \tag{9}$$

An attractive feature of the SAE formulation is that it is more robust with respect to outliers. SSE squares individual errors, which tends to exaggerate the influence of outliers – the calibration can distort parameter values just to get the model closer to the outlier. The idea of robust regression is to reduce the leverage of individual points, and weighting functions exist that discount outliers altogether, such as Tukey's biweight and others (Press et al. 1992). That said, SAE functions are relatively uncommon in hydrology: their benefits are not always demonstrable, and the absolute value function is surely less smooth than SSE functions. Outlier detection and leverage analysis appear more attractive from the perspective of hydrological model setup and data analysis rather than just optimization (e.g., Wright et al. 2015; Hill et al. 2015).

The goodness-of-fit function framework allows the hydrologists to craft their own objective functions to reflect modelling objectives of interest – mimicking the hydrologist's intuition mentioned in Sect. 2.4. For example, SSE and SAE functions can be computed separately for high and/or low flows; response weights and transformations such as logarithmic can be used to emphasize the fitting of particular data points and so forth (Chapman 1970; Chiew et al. 1993; Pushpalatha et al. 2012). These aspects are revisited in Sect. 5.4 from a statistical perspective. Further examples of goodness-of-fit functions are provided by Legates and McCabe Jr. (1999).

Several questions arise at this stage. Do we need to restrict attention to a single goodness-of-fit function? Given that some objective functions, notably SSE, RMSE, and NSE, have the same optimum but a different shape (curvature), how can we unambiguously quantify parameter uncertainty? And more generally, how do we navigate the vast range of potential performance measures? The following sections will describe how to overcome some of the challenges and present objective functions and calibration approaches from a more systematic perspective.

3 Automatic Calibration Via Optimization

The idea of automatic calibration is to reduce the need for human intervention and tackle the calibration problem in Eq. (5) using mathematical algorithms. As even simple goodness-of-fit function will be impossible to optimize analytically for most hydrological models, numerical optimization is employed. A plethora of numerical optimization algorithms have been employed in hydrological calibration, ranging from local methods such as classical Newton and quasi-Newton methods that

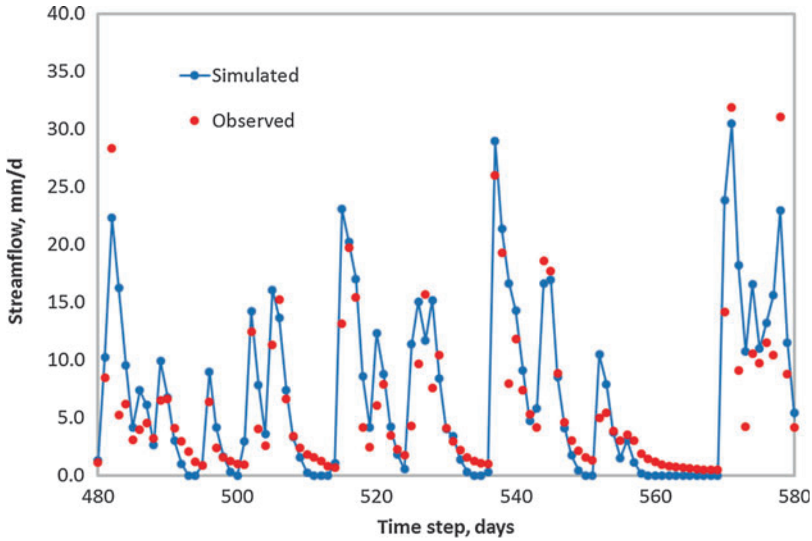

Fig. 3 Calibration of simple nonlinear reservoir bucket model, $dS/dt = P - kS^{\alpha} - E$, to the Maimai catchment, New Zealand

assume the objective function is smooth and near-quadratic (Gill et al. 1981) to global evolutionary methods such as the shuffled complex evolution (SCE) algorithm (Duan et al. 1992) and the dynamically dimensioned search (DDS) algorithm (Tolson and Shoemaker 2007) that make few if any such assumptions.

The selection of an optimization algorithm depends on the hydrological model and objective function. For example, Gauss-Newton-type algorithms are tailored to (possibly transformed) sum of squared errors (SSE) objective functions and are implemented in packages such as PEST (Doherty 2005), the Australian eWater Source platform (Welsh et al. 2013), and the BATEAU toolkit (Kavetski 2005) available to calibrate groundwater, water resources, and hydrological models.

In many cases, optimization works remarkably well. For example, a single-reservoir nonlinear model can be fitted to the Maimai catchment data using the Excel Solver tool, as shown in Fig. 3. For more complex modelling scenarios, calibration toolkits such as PEST (Doherty 2005) and BATEAU (Kavetski 2005) can be used, offering model coupling through ASCII input/output files and/or DLLs, visual interfaces, scripts, and other productivity features. Research is advancing into multi-start strategies and search randomization to increase the chance of finding the global optimum (e.g., Skahill and Doherty 2006; Kavetski et al. 2007; Tolson and Shoemaker 2007), as well as "multi-method" approaches that run multiple optimizers in parallel and pick the ones making the most progress (Vrugt and Robinson 2007).

That said, off-the-shelf optimization of hydrological models is not yet routinely attainable. Hydrological models with highly nonlinear dependence on their parameters have markedly non-quadratic objective functions, often exhibiting macroscale multi-optimality and microscale roughness (Duan et al. 1992). These problematic features are often exaggerated by fragile numerical implementations, such as the explicit Euler time stepping scheme, and by internal model thresholds (Kavetski and Kuczera 2007; Kavetski and Clark 2010). Under these conditions, gradient-based Newton-type algorithms typically converge only to the optimum nearest to the initial search point and generally behave erratically. Although current wisdom in hydrological modelling tends to favor evolutionary optimization methods, which tend to exhibit more robust global convergence and are less susceptible to microscale roughness, robust modifications of Newton-type methods offer the promise of comparable robustness at a much lower computational cost (Qin et al. 2018).

However, consider the following questions:

1. Identifiability problems. For example, Jakeman and Hornberger (1993) reported that typical rainfall-runoff data can support the identification of at most a "handful" of parameters in a lumped conceptual model. The optimization of distributed models solely to input-output data at the endpoints of their domain is clearly questionable – how would such data support the attribution of water flows through multiple internal pathways?
2. The very idea of looking solely for the global optimum at the expense of everything else can be questioned – nominally "slightly worse" optima may also be relevant and in some cases may provide more "realistic" model performance. In cases of pronounced multi-optimality, which could arise in case of grossly over-parameterized models, the optimum that becomes global may ultimately depend on data errors, subtle interplay of internal pathways, objective function idiosyncrasies, etc.
3. How do we estimate parameter uncertainty? For example, consider the endpoints of multiple optimization sequences – can these be treated as indicative of parameter uncertainty? Ultimately this approach gauges the effectiveness of the optimization algorithm and (potentially) the presence of multiple optima – somewhat counterintuitively, it would fail if the optimization algorithm is sufficiently robust to find the global optimum from most initial points. To estimate parameter uncertainty due to data and model errors, the shape of the objective function in the vicinity of the optimum should be investigated, e.g., using χ^2 methods (Press et al. 1992), which are related to the statistical ideas of Sect. 5.2.
4. A single objective is mathematically convenient but does not reflect the reality that multiple attributes might be of interest, e.g., low and high flows, timing of peaks, flow volumes, etc. (Sect. 2.4). In principle, a single-objective function can be constructed as a composite of multiple terms, e.g., separate SSE for low and high flows, water quality, etc., and melded together using weights. This approach goes some way toward recognizing the diverse nature of modelling objectives but is not quite "truly" multi-objective.

For these reasons, single-objective optimization on its own cannot be considered a complete solution to the calibration problem, even if it happens to be successful in terms of finding the global optimum.

4 Multi-Objective Optimization

Multi-objective optimization seeks to find the optimum of multiple objective functions simultaneously:

$$\boldsymbol{\theta}^{(\mathrm{opt})} = \underset{\boldsymbol{\theta}}{\operatorname{argmin}} \left[\Phi_1(\boldsymbol{\theta}), \Phi_2(\boldsymbol{\theta}), \ldots, \Phi_{N_\Phi}(\boldsymbol{\theta})\right] \tag{10}$$

It is well known that the optima of multiple general functions will not coincide except in very special cases (notably if the model is perfect or at least flexible enough to meet every objective – unlikely!). Instead, trade-offs arise between the degree to which individual objectives are optimized. Multi-objective optimization revolves around the concept of a “non-dominated” solution, which is a solution such that none of its corresponding objective function values can be improved without worsening at least one other objective. The Pareto front is defined as the set of non-dominated solutions.

Multi-objective optimization is a huge field of research in engineering, sciences, and mathematics; see Gupta et al. (1998) and the thorough review by Efstratiadis and Koutsoyiannis (2010) in the context of hydrological model calibration. Examples of multi-objective algorithms used in hydrology include MOSCEM (Vrugt et al. 2003), AMALGAM (Vrugt and Robinson 2007), and generalizations of the DDS algorithm (Asadzadeh and Tolson 2013). More generally, multi-objective optimization can be used to incorporate performance metrics other than those that quantify the model fit to observed data. For example, water resource model optimization may include economic objectives, pollution factors, and so forth. These setups may not be directly relevant to hydrological model calibration per se but are frequently used in the setup of management models where cost-benefit analysis is a major consideration (e.g., Marchi et al. 2014).

The ensemble of parameter sets comprising the Pareto front can be seen as representing parameter nonuniqueness associated with the (nonunique) choice of objective function. However, the interpretation of the Pareto front spread as parameter uncertainty is questionable. For example, the Pareto front does not, by itself, represent sources of uncertainty such as observation errors in the data, etc. Some advances along the direction of combining probabilistic and multi-objective techniques have been reported by Reichert and Schuwirth (2012) and warrant further investigation.

5 Probabilistic/Statistical Uncertainty Quantification

It is well known that hydrological modelling is affected by multiple sources of uncertainty, entering at every stage of the modelling process. For example, rainfall observations are subject to substantial sampling errors (e.g., McMillan et al. 2011), and streamflow observations are affected by rating curve errors (e.g., Westerbeg et al. 2010), not to mention the approximation of natural systems by mathematical models

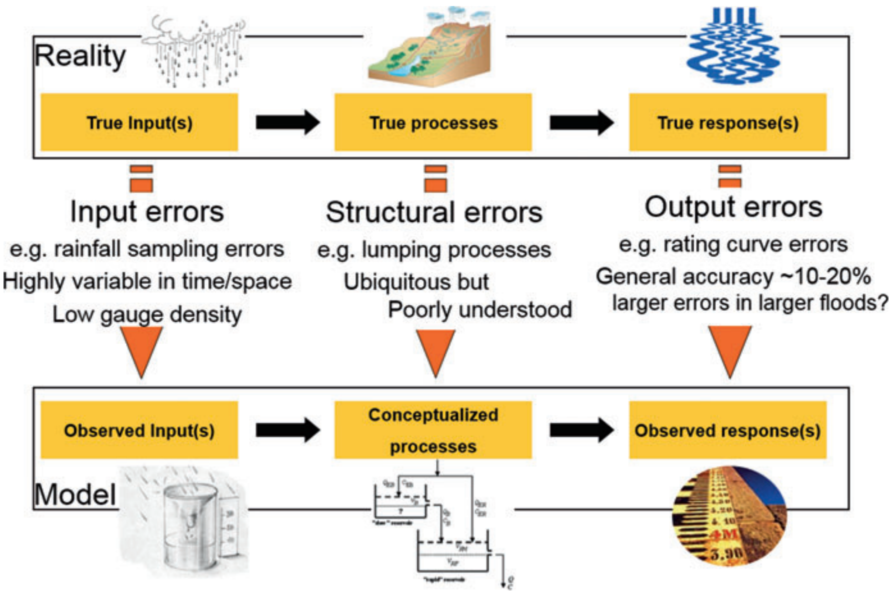

Fig. 4 Sources of uncertainty affecting parameter estimation in hydrological models

(e.g., Beven 2005; Renard et al. 2010). These sources of uncertainty are depicted schematically in Fig. 4.

Uncertainty is often classified as epistemic (i.e., due to model imperfections arising from incomplete knowledge of reality) versus aleatory (i.e., due to inherent randomness of the underlying phenomenon) – a distinction that is insightful yet not always clear-cut. As noted by Ang and Tang (2007), both types of uncertainty are tractable using probabilistic analysis, where uncertainty is described using probability theory. In this chapter, our primary focus is on the Bayesian paradigm, which provides a particularly appealing avenue for combining different sources of information.

5.1 Bayesian Inference: General Principles

Bayesian inference is a general class of probabilistic techniques, based on the premise that uncertainty in any quantity – including in model parameters – can be represented using random variables (probability distributions). Bayesian inference revolves around the posterior distribution of quantities of interest, $p(\boldsymbol{\theta}|\ \mathbf{D})$, which is given by Bayes equation:

$$p(\boldsymbol{\theta}|\mathbf{D}) = \frac{p(\mathbf{D}|\boldsymbol{\theta})p(\boldsymbol{\theta})}{p(\mathbf{D})} \tag{11}$$

Bayesian inference requires two key conceptual ingredients: a prior $p(\boldsymbol{\theta})$ and a likelihood function $p(\mathbf{D}|\ \boldsymbol{\theta})$. The term $p(\mathbf{D})$ is independent from $\boldsymbol{\theta}$ and represents a normalizing constant, defined by the total probability integral $p(\mathbf{D}) = \int p(\mathbf{D}|\ \boldsymbol{\theta})p(\boldsymbol{\theta})\, d\boldsymbol{\theta}$; while it is often a bear to compute, luckily this is not necessary in most modern Bayesian implementations (Sect. 5.3).

The prior distribution $p(\boldsymbol{\theta})$ is intuitively defined as what is known about parameters $\boldsymbol{\theta}$ *before* data $\mathbf{D}$ has been observed. Prior information can come from multiple sources, including previous quantitative studies, qualitative expert judgment, and combinations of multiple such sources. For example, in flood frequency analysis, it is common to use regional information derived from previous studies in neighboring locations (e.g., Hailegeorgis and Alfredsen 2017). When developing rating curve models, priors may come from the analysis of hydraulic controls (Le Coz et al. 2014). In hydrologic models, it is common to use the admissible range of parameter values to specify flat (non-informative) priors. In some cases, when working with widely used models such as GR4J, it may be reasonable to specify the prior based on parameter values inferred in previous calibrations – either worldwide or in nearby or similar locations (Perrin et al. 2001).

The likelihood function $p(\mathbf{D}|\ \boldsymbol{\theta})$ represents, loosely speaking, the probability of observing the data $\mathbf{D}$ given a particular set of model parameters $\boldsymbol{\theta}$. To obtain this, we need to specify a statistical model that may have "reasonably" generated the observed data. We are as free to specify this "reasonable" model just as hydrologists are free to specify their bucket model – guided by knowledge and intuition – and making practical judgments to simplify where appropriate. Section 5.2 will walk the interested reader through such a derivation. Section 6 will consider how to test calibration and model assumptions.

More formally, the likelihood function is the probability density function of the data-generating model, evaluated at the observed data, and viewed as a function of the model parameters. To emphasize its primary argument, the likelihood function is often written as $\mathcal{L}(\boldsymbol{\theta};\mathbf{D})$. This notation is convenient for defining multiple likelihood functions depending on the specific assumptions made, e.g., as in Sects. 5.2 and 5.4.

The idea of Bayesian inference is that we start with a vague initial knowledge (the prior) and use the information contained in the data to refine this knowledge and obtain a (hopefully) sharper posterior. In this respect, posterior knowledge represents prior knowledge updated using the data. This principle is illustrated schematically in Fig. 5.

Note that in Bayesian methods, the outcome of the inference is the entire posterior distribution – in contrast to single-objective optimization, where the outcome of the inference is a single parameter set. That said, in practice, Bayesian posteriors are often summarized using properties such as the (posterior) mean, mode, or median, and their uncertainty (spread) is often summarized using the posterior covariance and so forth (Sect. 5.3).

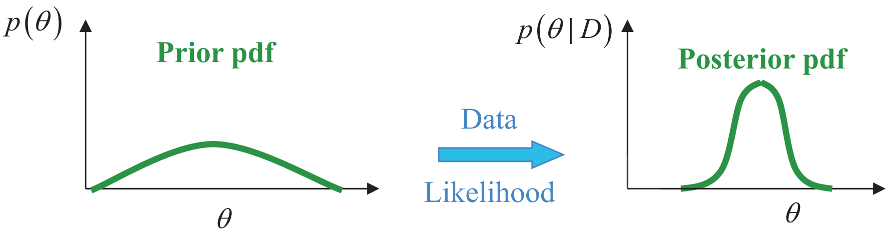

Fig. 5 Conceptual schematic of Bayesian inference. Parameter uncertainty is expressed using probability distributions. The combination of the prior and the likelihood yields the posterior, which represents the results of the inference

5.2 Least Squares Techniques as Gaussian Error Models

Bayesian inference often appears mysterious to modellers with a deterministic modelling background, especially when focusing solely on Eq. (11). Let us demystify Bayesian inference with a basic example.

Suppose we hypothesize that the original (deterministic) hydrological model provides a description of the observed data that is accurate on average but subject to random errors. In this case, the probabilistic model can be articulated as

$$\mathbf{Y} = \mathcal{H}(\tilde{\mathbf{x}};\boldsymbol{\theta}) + \boldsymbol{\mathcal{E}} \tag{12}$$

where the term $\boldsymbol{\mathcal{E}}$ Eis the residual error, intended to represent the effect of all source of uncertainty contributing to differences between observed and modelled streamflow values.

Suppose these residual errors are independent and identically distributed (iid) Gaussian with zero mean and variance σ_ε^2:

$$\varepsilon_t \sim \mathcal{N}\left(0, \sigma_\varepsilon^2\right) \tag{13}$$

Figure 6 illustrates this conceptualization. Equations (12 and 13) are referred to as the *error model* equations, as they provide a description of the errors affecting the model simulations and predictions.

The likelihood function $\mathcal{L}_{\mathrm{SLS}}(\boldsymbol{\theta}, \sigma_\varepsilon;\tilde{\mathbf{y}},\tilde{\mathbf{x}})$ corresponding to this error model can be derived as

$$\mathcal{L}_{\mathrm{SLS}}(\boldsymbol{\theta},\sigma_\varepsilon;\tilde{\mathbf{y}},\tilde{\mathbf{x}}) = p(\tilde{\mathbf{y}}|\boldsymbol{\theta},\tilde{\mathbf{x}},\sigma_\varepsilon) \underset{\text{indep}}{=} \prod_{t=1}^{N_{\tilde{\mathbf{y}}}} p(\tilde{y}_t|\boldsymbol{\theta},\sigma_\varepsilon,\tilde{\mathbf{x}}) \underset{\text{identic\&Gauss}}{=} \prod_{t=1}^{N_{\tilde{\mathbf{y}}}} f_{\mathcal{N}}\left(\tilde{y}_t;\mathcal{H}_t(\tilde{\mathbf{x}};\boldsymbol{\theta}),\sigma_\varepsilon^2\right) \tag{14}$$

where $f_{\mathcal{N}}(y;\mu,\sigma^2)$ denotes the Gaussian probability density function (pdf). The annotation “indep” refers to simplifications arising from the independence assumption and “identic & Gauss” to the identical Gaussian assumption.

Equation (14) can be also articulated in terms of the residuals:

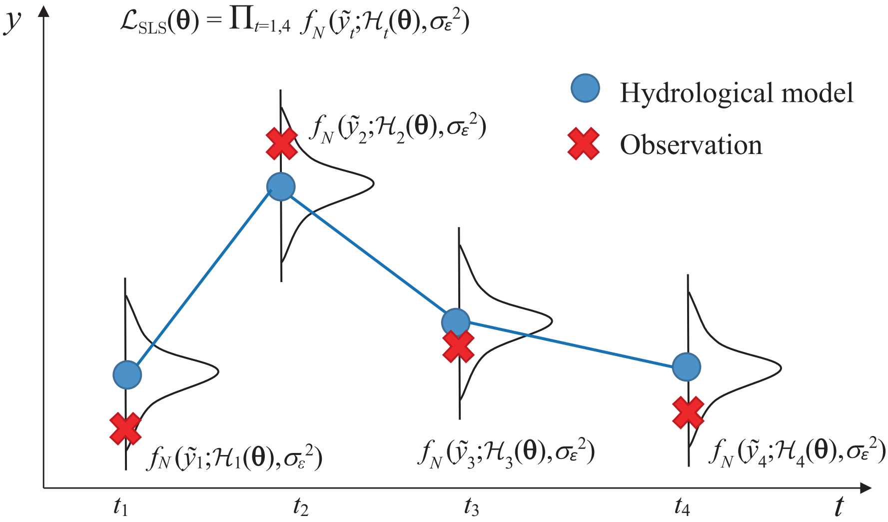

Fig. 6 Probabilistic model corresponding to Bayesian least squares inference. A Gaussian error model is assumed around each value simulated by the hydrological model, with the error variance being an error model parameter. The likelihood function is then defined as the (Gaussian) probability density of observation values of streamflow within this (joint) distribution, which in this case is the product of the (Gaussian) densities of individual observations

$$\mathcal{L}_{\text{SLS}}(\boldsymbol{\theta},\sigma_\varepsilon;\tilde{\mathbf{y}},\tilde{\mathbf{x}}) = \prod_{t=1}^{N_{\tilde{\mathbf{y}}}} f_N(\tilde{y}_t - \mathcal{H}_t(\tilde{\mathbf{x}};\boldsymbol{\theta});0,\sigma_\varepsilon) = \prod_{t=1}^{N_{\tilde{\mathbf{y}}}} f_N\left(\varepsilon_t[\tilde{\mathbf{y}};\tilde{\mathbf{x}},\boldsymbol{\theta}];0,\sigma_\varepsilon^2\right) \tag{15}$$

$$\varepsilon_t = \tilde{y}_t - \mathcal{H}_t(\tilde{\mathbf{x}};\boldsymbol{\theta}) \tag{16}$$

Note that the change of variables from $\mathbf{Y}$ to $\boldsymbol{\mathcal{E}}$ in the pdfs given by Eqs. (14) and (15) requires a Jacobian term $\partial\boldsymbol{\varepsilon}/\partial\mathbf{y}|_{\mathbf{y}=\tilde{\mathbf{y}}}$; however, this term is unity in view of Eq. (16) and is hence omitted.

If we substitute the Gaussian pdf expression into Eq. (15) and take logs, we get

$$\log\mathcal{L}_{\text{SLS}}(\boldsymbol{\theta},\sigma_\varepsilon;\tilde{\mathbf{y}},\tilde{\mathbf{x}}) = -\frac{N_{\tilde{\mathbf{y}}}}{2}\log 2\pi - N_{\tilde{\mathbf{y}}}\log\sigma_\varepsilon - \frac{1}{2\sigma_\varepsilon^2}\Phi_{\text{SSE}}(\boldsymbol{\theta};\tilde{\mathbf{x}},\tilde{\mathbf{y}}) \tag{17}$$

which establishes the close correspondence of probabilistic inference under Gaussian assumptions with the SSE objective function in Eq. (4). For example, it can be readily shown that the parameter set $\boldsymbol{\theta}^{(\text{SLS})}$ that maximizes the (log-) likelihood function in Eq. (17) also minimizes the SSE objective function. This equivalence holds whether the error variance σ_ε^2 is inferred or assumed known. For this reason, we can refer to Eqs. (14, 15, 16, and 17) as Bayesian least squares inference. The derivations needed to arrive at these equations put the choice of error norm in Eq. (4) on a more defined theoretical basis and in doing so highlight its implicit assumptions (here, iid Gaussian errors).

If the primary interest is in the model parameters $\boldsymbol{\theta}$, the error parameter σ_ε can be integrated out (Kavetski et al. 2006a):

$$\log \mathcal{L}_{\mathrm{SLS2}}(\boldsymbol{\theta};\tilde{\mathbf{y}},\tilde{\mathbf{x}}) = \log \int \mathcal{L}_{\mathrm{SLS}}(\boldsymbol{\theta},\sigma_\varepsilon;\tilde{\mathbf{y}},\tilde{\mathbf{x}})\, \mathrm{d}\sigma_\varepsilon \propto \frac{N_{\tilde{\mathbf{y}}}+2}{2} \log \Phi_{\mathrm{SSE}}(\boldsymbol{\theta};\tilde{\mathbf{x}},\tilde{\mathbf{y}}) \tag{18}$$

which still retains the SSE kernel within the (now lower-dimensional) likelihood function. A similar expression holds for certain non-rectangular priors on σ_ε.

5.3 Tools for Analyzing Bayesian Posteriors

The formulation of the posterior is only the first step of the inference. The equation defining the posterior distribution in effect plays the role of the "objective function" in Bayesian estimation. The same intuitive ideas established in Sect. 2.5 for interpreting the shape of the objective function apply to the posterior distribution – near-optimal regions suggest the most likely values of the parameters, the spread of the distribution is indicative of posterior parameter uncertainty, and elongated contours sloping with respect to the parameter axes indicate parameter correlations/dependencies. Once we have derived the functional form of the posterior, how do we use it to get a parameter value and its uncertainty?

In some simple cases, where the likelihood and prior are given by "conjugate" distributions – which for the most part are common textbook distributions such as Gaussian, Gamma, and so forth – the posterior will itself come out as a known distribution, with parameters derived from the parameters of the prior and likelihood (Box and Tiao 1992). However, this simplification is seldom possible with most hydrological models.

In practice, Bayesian posteriors are either summarized by their estimated mode (optimum) and covariance or explored wholesale using Markov Chain Monte Carlo (MCMC) algorithms.

The posterior mode can be found using optimization as $\widehat{\boldsymbol{\theta}} = \mathrm{mode}[\boldsymbol{\theta}] = \underset{\boldsymbol{\theta}}{\mathrm{argmax}}\, p(\boldsymbol{\theta}|\,\tilde{\mathbf{y}})$, i.e., directly treating the Bayesian posterior as an objective function (cf Sect. 5.2). The posterior covariance can then be approximated as $\mathrm{cov}[\boldsymbol{\theta}] \approx -\mathbf{H}_{\boldsymbol{\theta}}^{-1}[\log p(\boldsymbol{\theta}|\,\tilde{\mathbf{y}})]_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}}$ where $\mathbf{H}_{\boldsymbol{\theta}} = \partial^2/\partial\boldsymbol{\theta}^2$ is the Hessian (second derivative) matrix operator (Gelman et al. 1998), which can itself be approximated using finite differences. The Hessian matrix reflects the curvature of the posterior distribution (objective function) – peaky posteriors have large negative curvature and hence represent little posterior uncertainty, whereas flat posteriors have near-zero curvature and hence represent large uncertainty.

MCMC sampling provides an alternative – and very powerful – numerical technique for uncertainty quantification. A challenge of Bayesian posteriors is that they seldom take the form of common distributions. MCMC is a numerical sampling technique that (asymptotically) produces samples from the (possibly un-normalized) probability density function it is applied to (Gelman et al. 1998). It is most relevant to

(i) nonstandard distributions for which off-the-shelf samplers are not available, (ii) high-dimensional distributions, and (iii) distributions with pdf known only up to a constant of proportionality. These are precisely the features of most Bayesian posteriors, making MCMC a Bayesian's best friend! MCMC algorithms used in hydrology include multistage implementations of the classic Metropolis algorithm (Thyer et al. 2009), methods based on differential evolution (Vrugt et al. 2009a), and many others.

It is worth noting the important distinction between the choice of likelihood function and prior versus the choice of tools used to compute and analyze the posterior distribution. Just as the use of an optimization algorithm such as SCE to find the posterior mode does not make SCE into a "Bayesian algorithm" neither does the routine use of MCMC to sample from posterior distributions make MCMC into a "Bayesian algorithm." MCMC itself is *not* a Bayesian technique – it is a general numerical method for sampling from any probability distribution. Provided the MCMC algorithm is at all convergent, its results are determined by the function it is applied to, not the MCMC algorithm itself. For this reason, statements such as "we calibrated our model using MCMC" are about as informative (while still technically correct) as "we calibrated our model using MATLAB" – what should be reported first and foremost are the equations and assumptions defining the posterior distribution. A poorly chosen optimizer or MCMC sampler will surely degrade even the best posed inference, but algorithmic sophistication can hardly rescue a calibration from a poorly chosen objective function.

The next sections describe two distinct strategies for Bayesian inference, namely, aggregational and decompositional approaches, which are distinguished by the way they attempt to represent uncertainty.

5.4 Aggregational Methods

Aggregational approaches attempt to describe all sources of error using a single term. The simple least squares technique from Sect. 5.2 represents the prototypical implementation of this idea. However, its iid assumptions are questionable (e.g., Sorooshian and Dracup 1980). For example, errors of hydrological models typically exhibit heteroscedasticity, meaning larger errors in larger flows, which invalidates the assumption of identical distribution. In addition, errors typically exhibit persistence, meaning multiple consecutive errors of similar sign and magnitude, which invalidates the independence assumption. These statistical features can and should be reflected in the likelihood function.

Heteroscedasticity can be dealt with using weighted least squares and transformed least squares, in which case the assumption of identical distribution is applied to "normalized" residuals $\boldsymbol{\eta}$, rather than to raw residuals $\boldsymbol{\varepsilon}$:

$$\eta_t \sim \mathcal{N}\left(0, \sigma_\eta^2\right) \tag{19}$$

The use of weights (weighted least squares) corresponds to defining the normalized residuals as

$$\eta_t(\tilde{\mathbf{y}};\boldsymbol{\theta},\tilde{\mathbf{x}}) = \frac{\tilde{y}_t - \mathcal{H}_t[\tilde{\mathbf{x}};\boldsymbol{\theta}]}{\sigma_{\varepsilon(t)}} \tag{20}$$

which effectively gives some data points more influence than others. Under the heteroscedastic assumption that large flows have larger errors, data points corresponding to peak flows should receive *reduced* weight, so it is more accurate to say the influence is being "balanced."

If the residuals are assumed to represent all sources of error, their statistical properties are generally unknown, and additional assumptions are required, e.g.,

$$\sigma_{\varepsilon(t)} = a + b\,\mathcal{H}_t \tag{21}$$

where a and b are unknown parameters inferred along with the hydrological model parameters (Evin et al. 2014). Alternatively, the weights could be specified a priori, e.g., as in the PEST package (Doherty 2005). In other words, probabilistic inference often introduces parameters *in addition* to those of the hydrological model itself.

The use of response transformations (possibly with their own parameters $\boldsymbol{\theta}_Z$) represents an alternative strategy, where

$$\eta_t(\tilde{\mathbf{y}},\boldsymbol{\theta},\tilde{\mathbf{x}};\boldsymbol{\theta}_Z) = Z(\tilde{y}_t;\boldsymbol{\theta}_Z) - Z(\mathcal{H}_t[\tilde{\mathbf{x}};\boldsymbol{\theta}];\boldsymbol{\theta}_Z) \tag{22}$$

A common choice of transformation Z is the Box-Cox transformation:

$$Z(y;\lambda) = \begin{cases} \dfrac{y^{\lambda}-1}{\lambda} & \text{if } \lambda \neq 0 \\ \log y & \text{if } \lambda = 0 \end{cases} \tag{23}$$

which includes as special cases the logarithmic transformation ($\lambda = 0$), the inverse transformation ($\lambda = -1$), the square-root transformation ($\lambda = 0.5$), and, trivially, the null transformation ($\lambda = 1$).

Despite superficial differences, weighting and transformational strategies are closely related. Consider the likelihood function formulated in terms of normalized residuals $\boldsymbol{\eta}$:

$$\log\mathcal{L}_{\mathrm{H}}\left(\boldsymbol{\theta},\sigma_\eta;\tilde{\mathbf{y}},\tilde{\mathbf{x}},\boldsymbol{\theta}_Z\right) = p\left(\tilde{\mathbf{y}}|\boldsymbol{\theta},\tilde{\mathbf{x}},\sigma_\eta,\boldsymbol{\theta}_Z\right) = \left.\frac{\partial\boldsymbol{\eta}}{\partial\mathbf{y}}\right|_{\mathbf{y}=\tilde{\mathbf{y}}} \times f_{\mathcal{N}}\left(\boldsymbol{\eta}(\tilde{\mathbf{y}};\boldsymbol{\theta},\boldsymbol{\theta}_Z,\tilde{\mathbf{x}});\mathbf{0},\sigma_\eta^2\mathbf{I}\right) \tag{24}$$

where the Jacobian term $\partial\boldsymbol{\eta}/\partial\mathbf{y}|_{\mathbf{y}=\tilde{\mathbf{y}}}$ accounts for the change of variables from $\mathbf{y}$ to $\boldsymbol{\eta}$ and $\mathbf{I}$ is the identity matrix. Taylor series can be used to establish a first-order equivalence of linear weights in Eq. (21) and the log transformation given by Eq. (23) with $\lambda = 0$ (McInerney et al. 2017). This equivalence holds in terms of variances, but there are important differences in terms of skew and kurtosis that can impact practical performance (e.g., Schoups and Vrugt 2010; McInerney et al. 2017).

Persistence can be dealt with by incorporating autoregressive terms, e.g., the simplest AR(1) assumption yields

$$\eta_t = \phi \eta_{t-1} + W_t \tag{25}$$

$$W_t \sim \mathcal{N}\left(0, \sigma_w^2\right) \tag{26}$$

If the AR(1) assumption is applied to residuals after the Box-Cox transformation, the likelihood function is

$$\log \mathcal{L}_{\mathrm{H}}(\boldsymbol{\theta}, \phi, \sigma_w; \tilde{\mathbf{y}}, \tilde{\mathbf{x}}, \lambda) = p(\tilde{\mathbf{y}} | \boldsymbol{\theta}, \tilde{\mathbf{x}}, \phi, \sigma_w, \lambda) = \left.\frac{\partial \mathbf{w}}{\partial \mathbf{y}}\right|_{\mathbf{y}=\tilde{\mathbf{y}}} \times f_{\mathcal{N}}\left(\mathbf{w}(\tilde{\mathbf{y}}; \boldsymbol{\theta}, \tilde{\mathbf{x}}, \phi, \lambda); \mathbf{0}, \sigma_w^2 \mathbf{I}\right) \tag{27}$$

where we allow for the autocorrelation parameter ϕ to be inferred along with the error variance σ_w^2 while keeping the BC transformation parameter λ fixed.

The practicalities of representing heteroscedasticity and persistence within the likelihood function are often subtle and have a major impact on parameter estimation. First, the order of treatment is important – it is best to start by stabilizing the residual variance using a transformation (or weighting) and then treating persistence (Evin et al. 2013). Second, the parameters of the error models, notably the error variance and autocorrelation, can be inferred either jointly with the hydrological parameters or in a separate post-processing step. Evin et al. (2014) considered a post-processing approach that first estimated the hydrological parameters under the assumption of no persistence and then separately estimated the error variance and autocorrelation. Although the joint approach is a more pure application of the Bayesian paradigm, it can suffer from multi-way interactions between the mass-balance parameters, the autocorrelation coefficient, and the error variance, which lead to poor quality predictions; the post-processing approach appears more robust, as seen in Fig. 7 adopted from Evin et al. (2014). Third, in terms of transformation parameter, values of the Box-Cox λ in the range 0–0.5 appear to provide the best empirical performance (McInerney et al. 2017), with trade-offs arising between the reliability, precision, and bias of the resulting predictions.

Many aspects of residual error modelling are of interest beyond Eqs. (19, 20, 21, 22, 23, 24, 25, 26, and 27). For example, Gaussian assumptions can be replaced with more general skewed power exponential (SEP) distribution, which allows for skewness and kurtosis parameters (Schoups and Vrugt 2010), and AR(1) assumptions can be replaced with more general autoregressive models (e.g., Morawietz et al. 2011). The treatment of zero and near-zero flows has been investigated using approaches such as censoring (Wang and Robertson 2011) and "zero-flow inflation" (Smith et al. 2010). Residual error models based on mixtures (Schaefli et al. 2007) and conditioned on covariates other than streamflow (e.g., Pianosi and Raso 2012) have also been investigated.

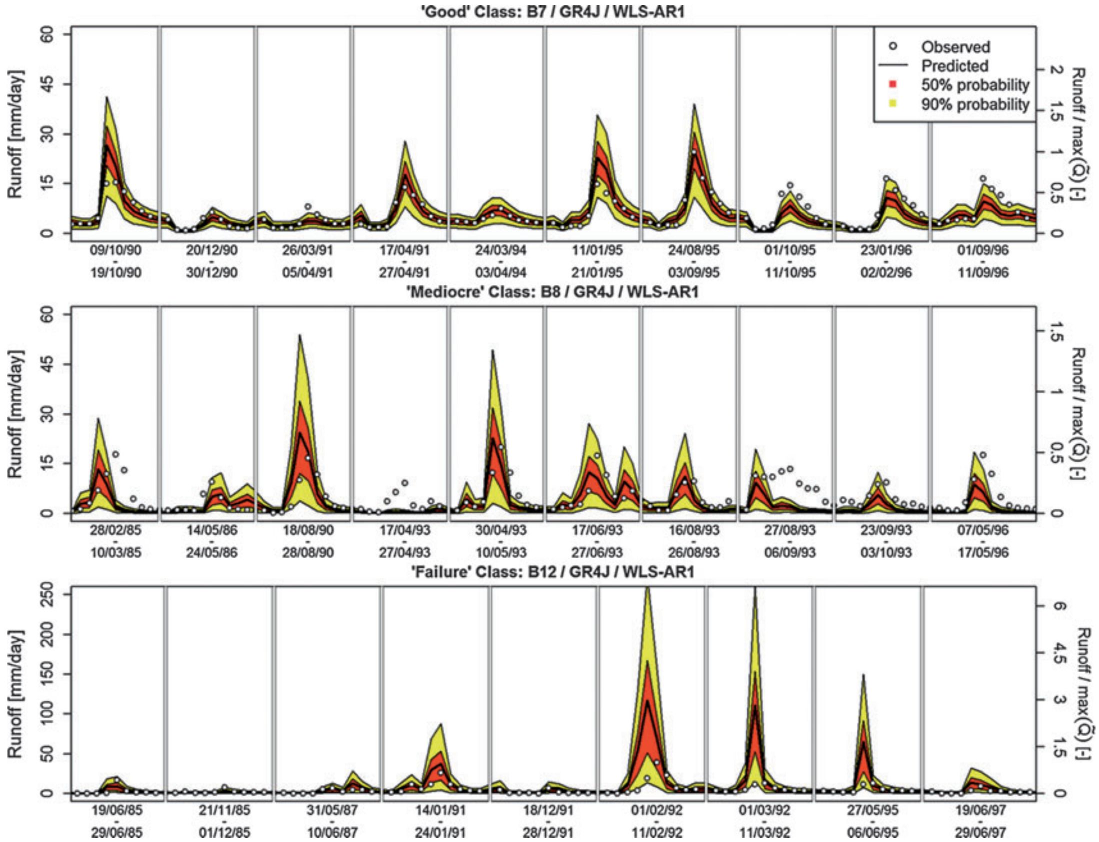

Fig. 7 Representative hydrological predictions obtained in the case studies of Evin et al. (2014), where an aggregational approach was employed with all sources of error represented using a single residual error term. The likelihood function was formulated accounting for heteroscedasticity and persistence. Three classes of results are presented, ranging from “good” to “mediocre” and “failure” categories, based on predictive performance of the parameter estimation approach. The notation B7, B8 and B12 refers to the French Broad, English River and San Marcos catchments in the USA. Reproduced with permission from John Wiley and Sons

Aggregated strategies are well suited to operational applications, as the resulting inference can produce reliable and precise predictions at a low-moderate cost in terms of algorithm complexity and computational effort. In other words, aggregated strategies allow the modeller to focus on the pragmatic goal of overall predictive uncertainty quantification (e.g., Krzysztofowicz 1999; Lerat et al. 2015). The next section details decompositional approaches – which are more ambitious in their objectives yet are also harder and more expensive to implement.

5.5 Decompositional Methods

Decompositional approaches attempt to explicitly disentangle the contributions of individual sources of error, such as those seen in Fig. 4 (e.g., Krzysztofowicz 1999; Kavetski et al. 2002; Seo et al. 2006; Huard and Mailhot 2008; Vrugt et al. 2008; Reichert and Mieleitner 2009). An example of the decompositional approach in hydrological modelling is given by the Bayesian total error analysis (BATEA) (Kavetski et al. 2006a; Kuczera et al. 2006). Decompositional approaches require more data and are more complex than aggregated approaches. Notably, Renard et al. (2010) established that the decomposition is inherently ill-posed in the absence of (approximate) prior knowledge; as noted by Beven (2005), the problem can be conceptualized as inferring individual error terms $\boldsymbol{\varepsilon}_x$, $\boldsymbol{\varepsilon}_y$, and $\boldsymbol{\varepsilon}_{\mathcal{H}}$ associated with input data errors, output data errors, and model structural errors, respectively,

$$\boldsymbol{\varepsilon} = \boldsymbol{\varepsilon}_x + \boldsymbol{\varepsilon}_y + \boldsymbol{\varepsilon}_{\mathcal{H}} \tag{28}$$

which is clearly ill-posed without at least some information about at least two of the three error terms.

Renard et al. (2011) tackled this challenge in a case study of the Yzeron catchment (France), where a dense rain gauge network, R13H, was available over a 2-year period. Data from R13H was exploited using conditional simulation (Tompson et al. 1989) to develop a prior error model for a sparse rain gauge network, R3D, active over a much longer time period. The priors for parameters describing streamflow observation errors were obtained using rating curve error analysis (Thyer et al. 2009). The representation of model structural error is another major challenge in a decompositional approach. Unlike data errors, which in principle can be estimated by comparison to a more accurate data set, there is no analogous concept for model structural errors – model comparison experiments do not provide much evidence of a particular model being consistently more accurate than others (e.g., Duan et al. 2006). Renard et al. (2011) treated model structural errors as what's left behind after the other errors have been characterized. At the cost of setting them up, decompositional approaches can offer fascinating insights into the relative contributions of different sources of error to total predictive uncertainty. For example, when using GR4J and the R3D data set to predict streamflow in the Yzeron catchment, structural errors dominated rainfall-induced errors, as shown in Fig. 8 adopted from Renard et al. (2011). These insights can guide efforts to improve the predictions, e.g., in this instance by prioritizing improvements to model structure over looking for

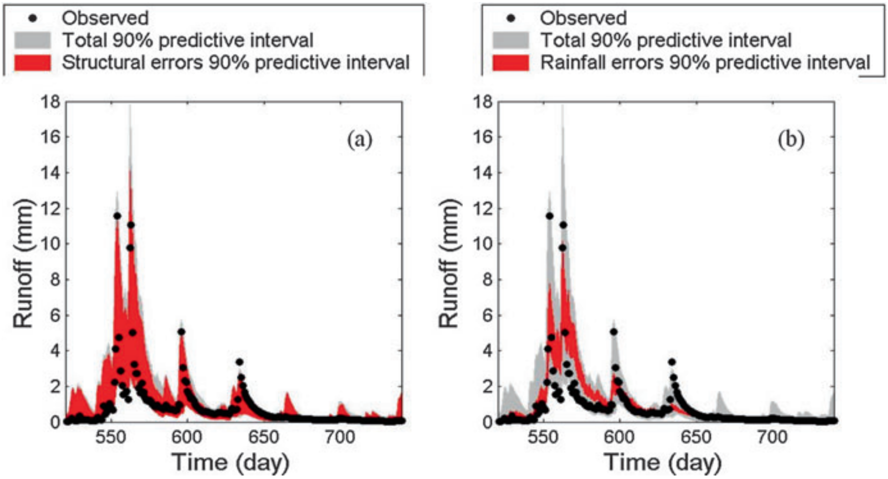

Fig. 8 Insights from the application of the BATEA decompositional parameter estimation approach in the Yzeron catchment, France (Renard et al. 2011). In this case study, structural errors of the GR4J model appear to dominate streamflow uncertainty due to the effects of rainfall errors. Figure reproduced from Renard et al. (2009)

more accurate data sets (a statement specific to this case study and not intended to detract from the general importance of data to hydrology!).

5.6 Methods Other than Bayesian and Other than Probabilistic

The presentation of probabilistic inference thus far has focused primarily on Bayesian methods, which tend to be commonly used for hydrological models. This section provides a brief summary of other related techniques.

Frequentist techniques are common in statistical estimation, particularly in flood frequency analysis. The method of moments (MoM) can be used whenever the parameters of a distribution can be related to the moments of the calibration data (e.g., Salas 1993). Maximum likelihood (ML) methods (e.g., Martins and Stedinger 2000) are broadly similar to Bayesian methods (barring some philosophical differences in the interpretation of probability) but (generally) do not allow for a prior distribution. Procedurally, MoM and ML proceed by first estimating the optimal parameters and then estimating parametric uncertainty – as opposed to Bayesian inference where the modeller must first derive the posterior distribution and then summarize/use it. In many rainfall-runoff model applications, there is little difference between frequentist and Bayesian approaches, as the influence of the prior is close to negligible for typical lengths of data. However, it is less clear how to develop a decompositional approach, e.g., analogous to BATEA but without using priors to handle the disaggregation exemplified by Eq. (28). Priors can also be valuable in data-sparse contexts.

The generalized likelihood uncertainty estimation (GLUE) (Beven and Binley 1992) is an estimation technique common in conceptual hydrological modelling. GLUE is often described as an "informal" technique, in the sense that it does not seek to construct probabilistic descriptions of uncertainty such as the error models in Sects. 5.2 and 5.4. For example, many GLUE publications have used the Nash-Sutcliffe efficiency *as if* it were a likelihood function and have relied solely on parametric uncertainty (without a residual error model) when generating prediction limits (e.g., Freer et al. 1996). The motivation for GLUE has evolved since its original development and has more recently focused on the challenges of describing epistemic uncertainty using probability theory. The solutions suggested by GLUE have elicited much debate, from the role of parameters and parametric uncertainty in modelling to whether prediction limits should satisfy probabilistic criteria such as enveloping a prescribed proportion of observations (e.g., Mantovan and Todini 2006; Stedinger et al. 2008; Beven 2006; Beven et al. 2012; Clark et al. 2012). Despite many divergent perspectives, there is also important commonality. The concept of equifinality, central to the original motivation for GLUE (Beven 2006), corresponds broadly to non-identifiability and ill-posedness (Sects. 2.1 and 5.3). More recently, the GLUE community has drawn attention to "disinformative data" (Beven and Westerberg 2011), which in the context of probabilistic techniques represents data that violates the assumed error models. Some work has attempted to bridge the gap between GLUE and Bayesian methods (e.g., Vrugt et al. 2009b; Nott et al. 2012; Kavetski et al. 2018).

Ultimately, conceptual and algorithmic similarities will necessarily arise between all calibration approaches that work with (optimize and/or sample) functions of the form $\mathcal{L}(\boldsymbol{\theta};\tilde{\mathbf{y}})p(\boldsymbol{\theta})$. From this perspective, genuine differences can only arise from the way $\mathcal{L}(\boldsymbol{\theta};\tilde{\mathbf{y}})$ and $p(\boldsymbol{\theta})$ are constructed: the use of probability theory leads to Bayesian (and maximum likelihood) methods and probabilistic prediction, whereas the use of other principles, such as fuzzy set theory (Freer et al. 2004) and others, leads to correspondingly different interpretations (Smith et al. 2008). For this reason, a modeller that wishes to obtain prediction limits that have a probabilistic interpretation is best advised to use the tools of probability theory. But the extent to which a probabilistic interpretation is a desirable attribute of an estimate or prediction is a much deeper question. This chapter takes the perspective that probability theory is a suitable – indeed advantageous – platform for describing the data and model uncertainties of relevance to scientific and engineering applications (Ang and Tang 2007); the interested reader is directed to discussions in the broader scientific community (e.g., de Finetti 1964; Oberkampf et al. 2004; O'Hagan and Oakley O'Hagan and Oakley 2004; Reichert et al. 2015, among others).

6 Model Diagnostics as Part of Parameter Estimation

The specific task of parameter estimation cannot be seen in isolation from the broader modelling process of hypothesis testing, refinement, and prediction. As vividly seen in Sects. 2, 3, 4, and 5, parameter estimation is necessarily based on assumptions, such as the applicability of data for a priori estimation, the selection of an objective function for optimization or the selection of an error model for probabilistic inference. In order to gain confidence that parameter estimation has been successful, these assumptions should be scrutinized and, if necessary, replaced. In the absence of such checks, there is little guarantee that, as eloquently noted by Kirchner (2006), the calibrated models are not merely dancing like mathematical marionettes to the tune of the calibration data. The topic of posterior diagnostics is extensive in its own right; this chapter highlights some of the key principles and practical implementations but does not attempt to be truly comprehensive.

We first consider what kind of model fit can be expected after calibration and how should uncertainty limits behave. To this end, Fig. 9 shows the results of a GR4J model calibration. For didactic reasons, synthetic data with Gaussian errors was used, so that the inference assumptions are met by construction. Should parametric uncertainty encompass the observations? In Fig. 9b, they clearly do not – does this mean something is wrong? To answer this question, consider the error model underlying the inference – Eqs. (12 and 13) clearly assume that differences between the model simulations and the observed values are explained by an additive noise term. Therefore, if we want to construct probability limits that encompass the data, we need to account for the residual errors. Adding this term now produces the expected results, as seen in Fig. 9c. Another questionable aspect is the "white noisiness" of predictive limits seen in Fig. 9c – this is a consequence of the independence assumption in the residual error model. Figure 9d illustrates, for a different synthetic dataset, the much smoother and realistic (for a streamflow prediction) behavior of the autocorrelated error model in Eq. (25). A different perspective is taken in an approach such as GLUE, where no probabilistic error model is used (Sect. 5.6). In this case, parametric uncertainty is used to describe all sources of error, which in turn requires a much flatter ("lenient") pseudo-likelihood function. Hence, depending on the assumptions made by the estimation framework, different behaviors may be expected, and diagnostics much be crafted accordingly.

The individual assumptions underlying error models must also be tested. For example, if the Box-Cox Gaussian AR(1) error model is used, the modeller should test that (i) normalized residuals are approximately homoscedastic (e.g., no dependence on magnitude of simulated response), (ii) innovations are approximately independent (e.g., using PACF plots), and (iii) innovations are approximately Gaussian (e.g., using histograms and Gaussian QQ plots) (e.g., Thyer et al. 2009; Schoups and Vrugt 2010; Morawietz et al. 2011; McInerney et al. 2017). Model diagnostics represent a form of hypothesis testing and are most effective when applied in a structured and systematic way (Clark et al. 2011). For example, diagnostics are most inquisitive and informative when applied to stratified data, e.g., by flow magnitude, season, etc., as this can yield more insights into individual model deficiencies without being masked by averaging effects, etc.

What period should these diagnostics be applied to? Testing over the calibration period is generally weak. Arguably, one cannot claim a model to be "predicting" a quantity (e.g., data period) already explicitly used in its calibration – at best, the model is "simulating" or "reproducing" data already known to it. This perspective

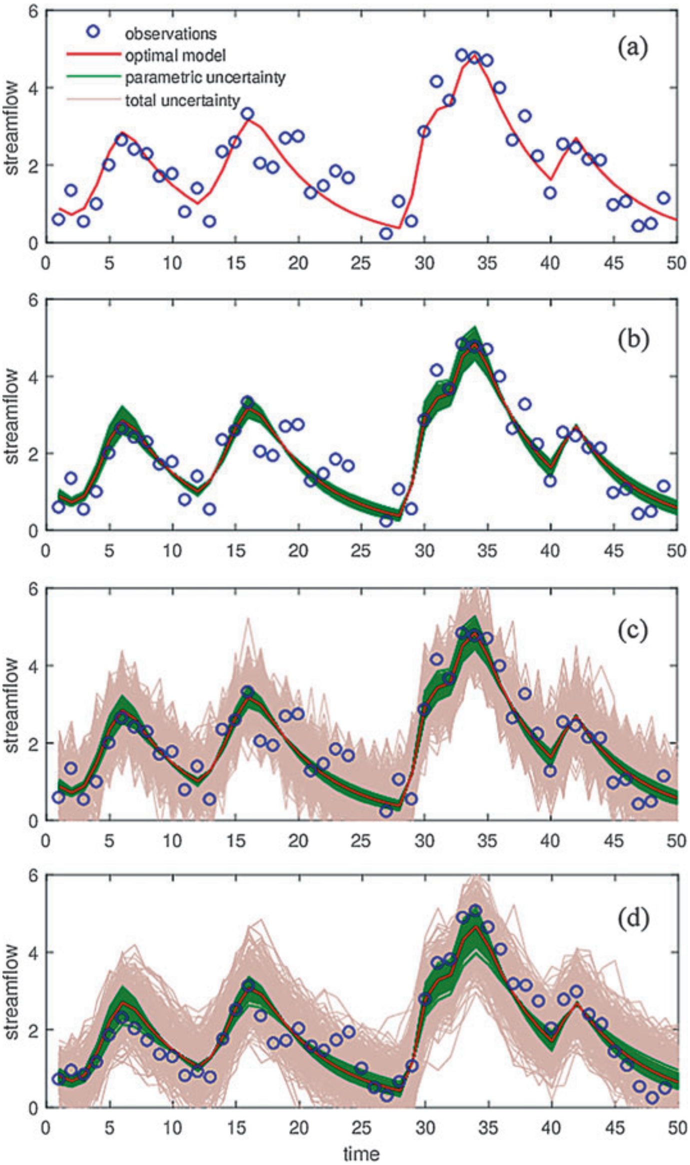

Fig. 9 Optimization, parametric uncertainty, and total predictive uncertainty in a simple Bayesian least squares framework. Panel *a* shows the optimal predictions; panel *b* shows the parametric uncertainty. Under Gaussian error model assumption, discrepancies between observed and simulated streamflows are described by the (Gaussian) residual error. Hence, residual error uncertainty, shown in panel *c*, must be added in order for the predictive limits to envelop the observed data. With reference to Fig. 6, panel *c* represents the use of the probability model to generate the predictive distribution and reconcile it against the actual observations. Panel d illustrates (using a different data set) the much "smoother" behavior of autocorrelated error models, which reduce the unrealistic jitter in hydrograph replicates generated using white noise error models

leads to the concept of a "validation" period, e.g., split-sample calibration where the available data is split into calibration and validation periods (Kuczera and Franks 2002), and more complex cross-validation setups (Tuteja et al. 2017). Validation mimics the way the model will be used in practice and arguably offers the best chance to detect deficiencies in the model and calibration. However, testing on a new period can create genuine complications in the case of non-stationarities (e.g., land-use change and/or climate variability) (Westra et al. 2012) – which highlights the formidable challenge of environmental prediction. Even the very semantics of the term "validation" have been questioned (e.g., Konikow and Bredehoeft 1992), on the grounds that it can provide a misleading impression of the model's abilities to make predictions.

Testing on validation periods can detect instances of over-parameterization, where a model has been over-fitted to spurious features in the calibration data and performs poorly on new data. A simple example is the fitting of a high-degree polynomial to a few data points. While any model can in principle be over-fitted, complex highly parameterized models are more susceptible, notably models based on neural networks (Kingston et al. 2008), but also physically based distributed models calibrated solely to catchment-average rainfall and runoff (e.g., Grayson et al. 1992; Jakeman and Hornberger 1993). Parameter estimation – whether via calibration or a priori estimation – is hence inevitably an exercise in balancing model complexity with available data (e.g., Fenicia et al. 2008).

Finally, we note that diagnostics are generally based on comparing simulated and observed responses and relating any deficiencies to the parameter estimates. In this respect, response diagnostics work with "tangible" quantities (to an extent) and seek to draw conclusions about parameters, which ultimately are "intangible" quantities. Mantovan et al. (2007) go as far as to refer to parameters as "abstract devices," which is of course far from ideal from the perspective of physical interpretability of models. Parameters, ultimately, do hold insights into systems, e.g., residence time, etc. (Fenicia et al. 2010). Nevertheless, it should be clear that parameter estimation is not as dependable as the prediction of observable quantities – while it may be reasonable to rely on validated prediction of quantities such as streamflow, relying on the corresponding parameter values requires a longer leap of faith!

7 Practicalities

This section lists some empirical "tricks" to supplement the theory of parameter estimation in hydrology.

7.1 Parameter Transformations

Parameter transformations often improve numerical algorithm performance when working with highly non-quadratic objective functions. Figure 10 shows an example from hydrological modelling, where a "banana-shaped" objective function becomes much better behaved when expressed in log-transformed parameter space. In some estimation problems, transformations are simply essential (Thyer et al. 2002). Parameters of hydrological models often benefit from log transformations, especially when appearing in exponents and multiplicative factors. Note the fundamental distinction between transforming parameters and responses: the former is purely a numerical device to improve the shape of the objective function "as seen" by an analysis method, whereas the latter yields a genuinely different objective function.

7.2 Impact of Model Non-smoothness/Discontinuities

Section 3 alluded to the difficulties posed by non-smooth models. The effect can be dramatic, especially on optimization algorithms that rely on gradients to establish the search direction. Figure 11 shows an example of a discontinuous objective function, due to internal model thresholds. When working with non-smooth models, hydrologists have two options: either use a more robust but slower algorithm or modify the model to remove the thresholds (Kavetski and Kuczera 2007). As seen in Fig. 11, model smoothing can successfully remove discontinuities from the objective function and substantially simplify the estimation process. Interestingly, previous work on smoothing often reported an improvement in model performance, which suggests that, at least on the large scale, environmental dynamics are not as threshold-driven as the models themselves (Kavetski and Clark 2010). For this reason, Hill et al. (2015) recommend a more concerted effort by environmental modellers to use robust numerics and smooth constitutive functions; this philosophy has been adopted in

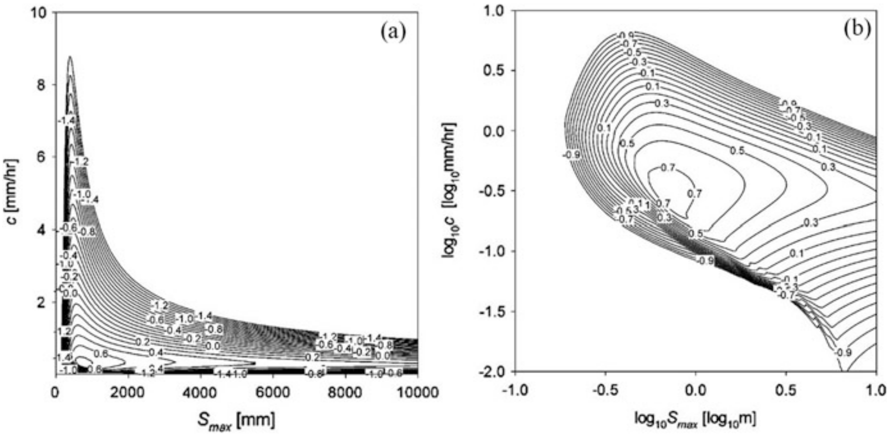

Fig. 10 Illustration of parameter transformation to improve the conditioning of the objective function (Kavetski et al. 2006c). Panel *a* shows a least squares objective function exhibiting a banana-type shape. Panel *b* shows the same objective function plotted in log-transformed parameter space, exhibiting a much better-behaved near-quadratic shape. Reproduced with permission from Elsevier

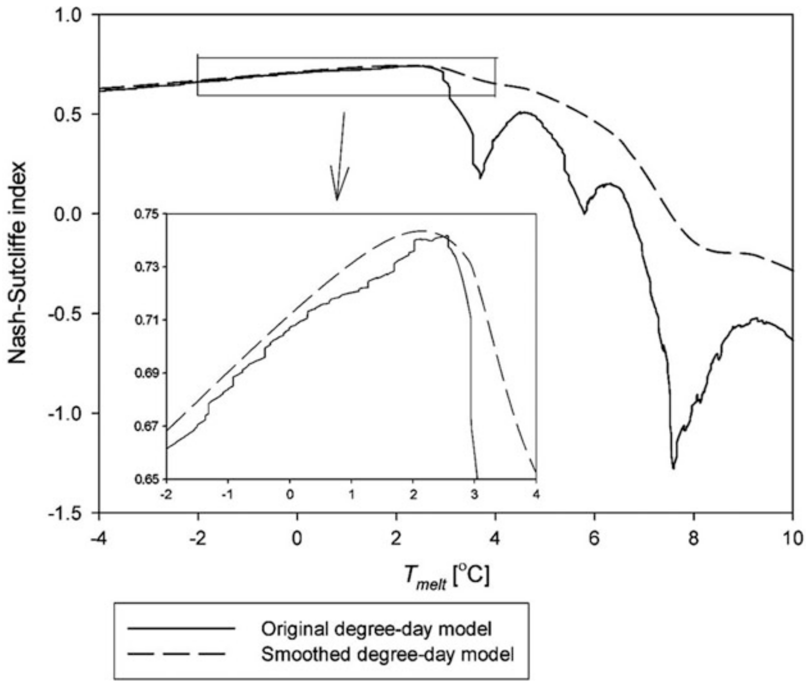

Fig. 11 Effect of non-smooth models on the objective function, illustrated using results from the case study of Kavetski et al. (2006b). Both macro- and microscale irregularities are visible. Model smoothing removes these artefacts and yields a remarkably well-behaved surface. Reproduced with permission from Elsevier

modelling frameworks such as FUSE (Clark et al. 2008), SUPERFLEX (Fenicia et al. 2011), SUMMA (Clark et al. 2015), and RAVEN (Craig et al. 2017).

7.3 Initial Conditions: Estimate or Warm-Up

Most hydrological models are dynamic in time, and initial state values must be specified before the model can be deployed. In practice, initial model states are unknown. Three options can be considered:

(i) Use a warm-up period after setting the initial conditions to some arbitrary values (e.g., 50% full). Warm-up periods are easy to implement but can chew up a lot of data in slow-responding catchments. The warm-up period is sufficiently long if the objective function (and hence calibrated parameters) exhibits little sensitivity to the (arbitrary) initial states.

(ii) Estimate initial conditions along with the model parameters. This approach does not waste data but can distort parameter values by favoring the fitting of the initial data period, where the fitted initial conditions in effect provide another degree of freedom.

(iii) Estimate initial conditions using other considerations, e.g., by solving the model equations for the steady-state storage values. This approach avoids the limitations of approaches (i) and (ii) but has the drawback that the theoretical steady state may not be representative of actual catchment conditions at the beginning of the calibration period. This approach is best used to inform the selection of initial values to shorten (but not avoid) the warm-up period.

7.4 Estimation of Expensive Models

Distributed models, such as MODFLOW (Harbaugh 2005), SWAT (Arnold and Fohrer 2005), and MHM (Samaniego et al. 2010), are increasingly used in environmental work to generate distributed predictions. These benefits accrue at substantial computational costs, with model runs taking as long as minutes or even hours per simulation. Computational costs inevitably impose restrictions on parameter estimation. For example, multi-start optimization and MCMC sampling may be severely limited or perhaps precluded altogether. When working with expensive models, it may be necessary to set a computational budget for analyses such as optimization, e.g., as implemented in the DDS optimizer (Tolson and Shoemaker 2007); parallel computing offers a pragmatic way to reduce wall-clock runtimes. Highly parameterized models can be handled using parameter regularization (Doherty 2003), model emulation (e.g., Albert 2012; Laloy et al. 2013), multi-scale parameter estimation techniques (Samaniego et al. 2010), and/or multistage estimation approaches where parameters are calibrated step-by-step rather than all at once (Fenicia et al. 2016).

8 Research Directions

Parameter estimation is a huge field with many unresolved challenges. This section lists some (not all) research directions of interest from practical and scientific perspectives.

8.1 Operational Improvements

Environmental and water agencies are increasingly interested in tackling challenging streamflow forecasting problems, including temporally consistent ("seamless") predictions over seasonal (3 months) lead times (e.g., Tuteja et al. 2011). Achieving these outcomes requires robust treatment of error persistence (e.g., Evin et al. 2014), balancing predictive reliability and precision (McInerney et al. 2017), and finding hydrological model parameters that perform well at multiple spatial and temporal scales (e.g., Samaniego et al. 2010). Similar requirements hold when seeking spatially coherent predictions over large catchment systems and river networks. In addition, since forecasting models are often calibrated using observed input-output data (e.g., rainfall-streamflow) but produce response forecasts (e.g., streamflow 3 months ahead) using forecasted forcings (e.g., rainfall from a numerical weather prediction model), better integration of multiple models representing individual sources of uncertainty is also of interest – this is one of the operational motivations for decompositional approaches (Sect. 5.5).

8.2 Sparse-Data Problems

Many locations around the globe are poorly gauged or ungauged. For example, in Australia, as much as 90–95% of the subcatchments of the Murray-Darling basin are ungauged (e.g., Chiew and Siriwardena 2005). Modelling these locations requires estimating model parameters from a combination of local properties (if available) and extrapolation from "similar" catchments. This was the theme of the Prediction in Ungauged Basins (PUB) decade (Sivapalan et al. 2003b). Indirect calibration approaches include non-concomitant calibration, where input and output data from different time periods are utilized. Spectral methods (e.g., Montanari and Toth 2007; De Vleeschouwer and Pauwels 2013; Schaefli and Kavetski 2017) and signature calibration (e.g., Yilmaz et al. 2008; Shafii and Tolson 2015; Westerberg and McMillan 2015; Fenicia et al. 2018) are of interest; e.g., signatures computed from simulated responses in Period A can be compared to the corresponding signatures of observed data in Period B. Computationally, signature calibration and uncertainty quantification can be approached using the fascinating class of methods known as approximate Bayesian computation (e.g., Nott et al. 2012; Vrugt and Sadegh 2013; Kavetski et al. 2018) – these methods avoid the need to derive the likelihood function in closed form and instead require sampling from the underlying probability model (Toni et al. 2009). Estimation in ungauged basins, where no data is available for model calibration, is being investigated using various parameter regionalization approaches (e.g., Bulygina and Gupta 2009; Hrachowitz et al. 2013).

8.3 Recursive Estimation and Data Assimilation

In many cases, calibration data is not available all at once and/or arrives in real time. For example, in applications such as flood forecasting, exploiting real-time information such as local observations and/or satellite imagery is of interest (e.g., Neal et al. 2007; Reichle 2008; Hostache et al. 2010; Giustarini et al. 2016; Revilla-Romero et al. 2016). In the context of parameter estimation, one can then distinguish between batch estimation (where the entire data set is used at once) and recursive estimation (where data is ingested sequentially, e.g., one data point at a time).

One of the simplest recursive estimation algorithms is the classic Kalman filter (KF) (Kalman 1960), which treats the problem of a linear model under conditions of Gaussian errors. The KF equations comprise a propagation (forecast) step and a correction (assimilation) step; the latter can be derived as a Bayesian inference not similar to the least squares problem but with a prior given by the posterior from the previous (forecast) step. The beauty of the KF is that its equations have an elegant and computationally fast solution (Gelb 1974). Since the assumptions of model linearity and Gaussian errors are restrictive, various generalizations of the KF equations have been proposed, including extended and ensemble Kalman filters and particle filters (e.g., Arulampalam et al. 2002; Vrugt et al. 2005, 2013; Weerts and El Serafy 2006); development, application, and improvement of these real-time techniques is of practical interest.

9 Summary and Conclusion

Parameter estimation in hydrological modelling is a common scientific and operational task and has received a tremendous amount of research and industry attention. This chapter has reviewed the broad classes of parameter estimation techniques, namely, a priori estimation and calibration (inverse modelling), with an emphasis on parameter estimation through calibration. Strategies reviewed include manual calibration, optimization, multi-objective optimization, and probabilistic estimation, with Bayesian inference receiving most attention. Manual calibration offers the ability to exploit expert understanding of the model and fit features in an intuitive way. However, reliance on expert knowledge makes it nontransparent and difficult to reproduce independently. The formulation of a goodness-of-fit function allows a modeller to quantify model performance for a given parameter set and lends itself to automatic implementation using optimization algorithms. Multi-objective optimization avoids one of the main limitations of single-objective work and allows exploring trade-offs between different aspects of the model fit (e.g., low vs high flows, timing of peaks, etc.). Probabilistic estimation, on the other hand, allows reflecting the uncertainty in the modelling process, which leads to parameter uncertainty. Bayesian inference supports probabilistic estimation from observed data while allowing for the use of additional information through prior distributions. Bayesian estimation can be implemented within aggregational approaches – where all sources of uncertainty are lumped into a single residual error term – or decompositional approaches, where there is an attempt to disentangle the effects of individual sources of errors (such as observational errors in rainfall forcings and streamflow responses and model structural errors). Irrespective of the calibration strategy, its assumptions must be scrutinized using posterior diagnostics, including tests for assumptions such as error heteroscedasticity, persistence, Gaussianity, and so forth. Practical implementations may also benefit from tricks such as parameter transformations and model smoothing. Directions of ongoing and future research include improvements in error model specification and the development of approaches for parameter estimation under sparse-data and ungauged conditions.

References

M.B. Abbott, V.M. Babovic, J.A. Cunge, Reply to comment by Beven et al on "Towards the hydraulics of the hydroinformatics era" by Abbott et al. J. Hydraul. Res. **41**(3), 333–336 (2003)

C. Albert, A mechanistic dynamic emulator. Nonlinear Anal. Real World Appl. **13**(6), 2747–2754 (2012)

A.H.-S. Ang, W.H. Tang, *Probability Concepts in Engineering: Emphasis on Applications to Civil and Environmental Engineering* (Wiley, Hoboken, 2007)

S.A. Archfield, M. Clark, B. Arheimer, L.E. Hay, H. McMillan, J.E. Kiang, J. Seibert, K. Hakala, A. Bock, T. Wagener, W.H. Farmer, V. Andréassian, S. Attinger, A. Viglione, R. Knight, S. Markstrom, T. Over, Accelerating advances in continental domain hydrologic modeling. Water Resour. Res. **51**(12), 10078–10091 (2015)

J.G. Arnold, N. Fohrer, SWAT2000: Current capabilities and research opportunities in applied watershed modelling. Hydrol. Process. **19**(3), 563–572 (2005)

M.S. Arulampalam, S. Maskell, N. Gordon, T. Clapp, A tutorial on particle filters for online nonlinear/non-Gaussian Bayesian tracking. IEEE Trans. Signal Process. **50**(2), 174–188 (2002). Special Issue on Monte Carlo Methods for Statistical Signal Processing

M. Asadzadeh, B.A. Tolson, Pareto archived dynamically dimensioned search with hypervolume-based selection for multiobjective optimization. Eng. Optim. **45**(12), 1489–1509 (2013)

K. Beven, TOPMODEL: A critique. Hydrol. Process. **11**(9), 1069–1085 (1997)

K. Beven, On the concept of model structural error. Water Sci. Technol. **52**, 167–175 (2005)

K.J. Beven, A manifesto for the equifinality thesis. J. Hydrol. **320**, 18–36 (2006)

K.J. Beven, A.M. Binley, The future of distributed models: Model calibration and uncertainty prediction. Hydrol. Process. **6**, 279–298 (1992)

K. Beven, I. Westerberg, On red herrings and real herrings: Disinformation and information in hydrological inference. Hydrol. Process. **25**, 1676–1680 (2011)

K. Beven, P. Smith, I. Westerberg, J. Freer, Comment on "Pursuing the method of multiple working hypotheses for hydrological modeling" by P. Clark et al. Water Resour. Res. **48**, W11801 (2012)

G.E.P. Box, G.C. Tiao, *Bayesian Inference in Statistical Analysis* (Wiley, New York, 1992)

G.O. Brown, Henry Darcy and the making of a law. Water Resour. Res. **38**(7), 1–12 (2002)

N. Bulygina, H. Gupta, Estimating the uncertain mathematical structure of a water balance model via Bayesian data assimilation. Water Resour. Res. **45**, W00B13 (2009)

T.G. Chapman, Optimization of a rainfall-runoff model for an arid zone catchment, in *I.A.S.H.-UNESCO Symposium on the Results of Research on Representative and Experimental Basins*, (IASH-AISH Publ, Wellington, 1970), pp. 126–144

F.H. Chiew, L. Siriwardena, Estimation of SIMHYD parameter values for application in ungauged catchments, in *MODSIM 2005 International Congress on Modelling and Simulation*, ed. by A. Zerger, R.M. Argent (Modelling and Simulation Society of Australia and New Zealand, Melbourne, Australia, 2005), pp. 2883–2889

F.H.S. Chiew, M.J. Stewardson, T.A. McMahon, Comparison of six rainfall-runoff modelling approaches. J. Hydrol. **147**, 1–36 (1993)

M.P. Clark, A.G. Slater, D.E. Rupp, R.A. Woods, J.A. Vrugt, H.V. Gupta, T. Wagener, L.E. Hay, Framework for understanding structural errors (FUSE): A modular framework to diagnose differences between hydrological models. Water Resour. Res. **44**, W00B02 (2008). https://doi.org/10.1029/2007WR006735

M.P. Clark, D. Kavetski, F. Fenicia, Pursuing the method of multiple working hypotheses for hydrological modeling. Water Resour. Res. **47**, W09301 (2011)

M.P. Clark, D. Kavetski, F. Fenicia, Reply to comment by K. Beven et al. on "Pursuing the method of multiple working hypotheses for hydrological modeling". Water Resour. Res. **48**, W11802 (2012)

M.P. Clark, B. Nijssen, J.D. Lundquist, D. Kavetski, D.E. Rupp, R.A. Woods, J.E. Freer, E.D. Gutmann, A.W. Wood, L.D. Brekke, J.R. Arnold, D.J. Gochis, R.M. Rasmussen, A unified

approach for process-based hydrologic modeling: 1. Modeling concept. Water Resour. Res. **51**(4), 2498–2514 (2015)
H.L. Cloke, F. Pappenberger, Ensemble flood forecasting: A review. J. Hydrol. **375**, 613–626 (2009)
J. Craig, et al., Raven User's and Developer's manual v2.7, http://www.civil.uwaterloo.ca/jrcraig/Raven/. (University of Waterloo, 2017)
B. de Finetti, Foresight: Its logical laws, its subjective sources, in *Studies in Subjective Probability*, ed. by H.E. Kyburg (Wiley, New York, 1964), pp. 93–158
N. De Vleeschouwer, V.R.N. Pauwels, Assessment of the indirect calibration of a rainfall-runoff model for ungauged catchments in Flanders. Hydrol. Earth Syst. Sci. **17**, 2001–2016 (2013)
J. Demargne, L. Wu, S.K. Regonda, J.D. Brown, H. Lee, M. He, D.J. Seo, R. Hartman, H.D. Herr, M. Fresch, J. Schaake, Y. Zhu, The science of NOAA's operational hydrologic ensemble forecast service. Bull. Am. Meteorol. Soc. **95**(1), 79–98 (2014)
J. Doherty, Ground water model calibration using pilot points and regularization. Ground Water **41**, 170–177 (2003)
J. Doherty, *PEST: Model Independent Parameter Estimation*, 5th edn. (Watermark Numerical Computing, Brisbane, 2005)
Q. Duan, S. Sorooshian, V. Gupta, Effective and efficient global optimization for conceptual rainfall-runoff models. Water Resour. Res. **28**(4), 1015–1031 (1992)
Q. Duan, J. Schaake, V. Andreassian, S.W. Franks, G. Goteti, H.V. Gupta, Y.M. Gusev, F. Habets, A. Hall, L. Hay, T. Hogue, M. Huang, G. Leavesley, X. Liang, O.N. Nasonova, J. Noilhan, L. Oudin, S. Sorooshian, T. Wagener, E.F. Wood, Model parameter estimation experiment (MOPEX): An overview of science strategy and major results from the second and third workshops. J. Hydrol. **320**(1–2), 3–17 (2006)
A. Efstratiadis, D. Koutsoyiannis, One decade of multi-objective calibration approaches in hydrological modelling: A review. Hydrol. Sci. J. **55**(1), 58–78 (2010)
G. Evin, D. Kavetski, M. Thyer, G. Kuczera, Pitfalls and improvements in the joint inference of heteroscedasticity and autocorrelation in hydrological model calibration. Water Resour. Res. **49**, 4518–4524 (2013)
G. Evin, M. Thyer, D. Kavetski, D. McInerney, G. Kuczera, Comparison of joint versus postprocessor approaches for hydrological uncertainty estimation accounting for error autocorrelation and heteroscedasticity. Water Resour. Res. **50**, 2350–2375 (2014)
F. Fenicia, H.H.G. Savenije, P. Matgen, L. Pfister, Understanding catchment behavior through stepwise model concept improvement. Water Resour. Res. **44**, W01402 (2008)
F. Fenicia, S. Wrede, D. Kavetski, L. Pfister, L. Hoffmann, H. Savenije, J.J. McDonnell, Impact of mixing assumptions on mean residence time estimation. Hydrol. Process. **24**(12), 1730–1741 (2010). (Special Issue on Residence Times and Preferential Flows)
F. Fenicia, D. Kavetski, H.H.G. Savenije, Elements of a flexible approach for conceptual hydrological modeling: Part 1. Motivation and theoretical development. Water Resour. Res. **47**, W11510 (2011)
F. Fenicia, D. Kavetski, H.H.G. Savenije, P. L, From spatially variable streamflow to distributed hydrological models: Analysis of key modeling decisions. Water Resour. Res. **52**, 954–989 (2016)
F. Fenicia, D. Kavetski, P. Reichert, C. Albert, Signature-domain calibration of hydrological models using approximate Bayesian computation: Empirical analysis of fundamental properties. Water Resour. Res. in press, https://doi.org/10.1002/2017WR021616 (2018)
C.W. Fetter, *Applied Hydrogeology*, 3rd edn. (Prentice-Hall, Upper Saddle River, 1994)
J. Freer, K. Beven, B. Ambroise, Bayesian estimation of uncertainty in runoff prediction and the value of data: An application of the GLUE approach. Water Resour. Res. **32**(7), 2161–2173 (1996)
J.E. Freer, H. McMillan, J.J. McDonnell, K.J. Beven, Constraining dynamic TOPMODEL responses for imprecise water table information using fuzzy rule based performance measures, J. Hydrol. **291**(3–4), 254–277 (2004)

R.A. Freeze, R.L. Harlan, Blueprint for a physically-based, digitally-simulated hydrologic response model. J. Hydrol. **9**, 237–258 (1969)
A. Gelb (ed.), *Applied Optimal Estimation* (MIT Press, Cambridge, MA, 1974)
A. Gelman, J.B. Carlin, H.S. Stern, D.B. Rubin, *Bayesian Data Analysis* (Chapman and Hall, London, 1998)
P.E. Gill, W. Murray, M.H. Wright, *Practical Optimization* (Academic, London, 1981)
L. Giustarini, R. Hostache, D. Kavetski, M. Chini, G. Corato, S. Schlaffer, P. Matgen, Probabilistic flood mapping using synthetic aperture radar data. IEEE Trans. Geosci. Remote Sens. **54**(12), 6958–6969 (2016)
R.S. Govindaraju, Artificial neural networks in hydrology. I: Preliminary concepts. J. Hydrol. Eng. **5**(2), 115–123 (2000)
R.B. Grayson, I.D. Moore, T.A. McMahon, Physically based hydrologic modeling: 2. Is the concept realistic? Water Resour. Res. **28**(10), 2659–2666 (1992)
V.K. Gupta, S. Sorooshian, The automatic calibration of conceptual catchment models using derivative-based optimization algorithms. Water Resour. Res. **21**(4), 473–485 (1985)
H.V. Gupta, S. Sorooshian, P.O. Yapo, Toward improved calibration of hydrologic models: Multiple and noncommensurable measures of information. Water Resour. Res. **34**(4), 751–763 (1998)
H.V. Gupta, T. Wagener, Y. Liu, Reconciling theory with observations: Elements of a diagnostic approach to model evaluation. Hydrol. Process. **22**, 3802–3813 (2008)
T.T. Hailegeorgis, K. Alfredsen, Regional flood frequency analysis and prediction in ungauged basins including estimation of major uncertainties for mid-Norway. J. Hydrol. **9**, 104–126 (2017)
A.W. Harbaugh, MODFLOW-2005, the U.S. Geological Survey modular ground-water model – the Ground-Water Flow Process, U.S. Geological Survey Techniques and Methods 6-A16 (2005)
M.C. Hill, D. Kavetski, M.P. Clark, M. Ye, M. Arabi, D. Lu, L. Foglia, S. Mehl, Practical use of computationally frugal model analysis methods. Groundwater **54**(2), 159 (2015)
R. Hostache, X. Lai, J. Monnier, C. Puech, Assimilation of spatially distributed water levels into a shallow-water model. Part II: Use of a remote sensing image of Mosel River. J. Hydrol. **390**(3–4), 257–268 (2010)
M. Hrachowitz, H.H.G. Savenije, G. Blöschl, J.J. McDonnell, M. Sivapalan, J.W. Pomeroy, B. Arheimer, T. Blume, M.P. Clark, U. Ehret, F. Fenicia, J.E. Freer, A. Gelfan, H.V. Gupta, D.A. Hughes, R.W. Hut, A. Montanari, S. Pande, D. Tetzlaff, P.A. Troch, S. Uhlenbrook, T. Wagener, H.C. Winsemius, R.A. Woods, E. Zehe, C. Cudennec, A decade of predictions in ungauged basins (PUB) – A review. Hydrol. Sci. J. **58**(6), 198–1255 (2013)
M. Hrachowitz, O. Fovet, L. Ruiz, T. Euser, S. Gharari, R. Nijzink, J. Freer, H.H.G. Savenije, C. Gascuel-Odoux, Process consistency in models: The importance of system signatures, expert knowledge, and process complexity. Water Resour. Res. **50**(9), 7445–7469 (2014)
D. Huard, A. Mailhot, Calibration of hydrological model GR2M using Bayesian uncertainty analysis. Water Resour. Res. **44**, W02424 (2008)
R.P. Ibbitt, T. O'Donnell, Designing conceptual catchment models for automatic fitting methods, in *Mathematical Models in Hydrology Symposium*, IAHS-AISH Publication No. 101(2) (1971), pp. 461–475
V.Y. Ivanov, E.R. Vivoni, R.L. Bras, D. Entekhabi, Catchment hydrologic response with a fully distributed triangulated irregular network model. Water Resour. Res. **40**(11), W11102 (2004). https://doi.org/10.1029/2004WR003218
A.J. Jakeman, G.M. Hornberger, How much complexity is warranted in a rainfall-runoff model? Water Resour. Res. **29**(8), 2637–2649 (1993)
R.E. Kalman, A new approach to linear filtering and prediction problems. J. Basic Eng. **82**(1), 35–45 (1960)
D. Kavetski, Analysis of input data uncertainty and numerical robustness in conceptual rainfall-runoff modelling, PhD Thesis, Faculty of Engineering and Built Environment, University of Newcastle (2005)

D. Kavetski, M.P. Clark, Ancient numerical daemons of conceptual hydrological modeling. Part 2: Impact of time stepping schemes on model analysis and prediction. Water Resour. Res. **46**, W10511 (2010). https://doi.org/10.1029/2009WR008896

D. Kavetski, G. Kuczera, Model smoothing strategies to remove microscale discontinuities and spurious secondary optima in objective functions in hydrological calibration. Water Resour. Res. **43**, W03411 (2007). https://doi.org/10.1029/2006WR005195

D. Kavetski, S. Franks, G. Kuczera, Confronting input uncertainty in environmental modelling, in *Calibration of Watershed Models*. Water Science and Application Series 6, ed. by Q.Y. Duan, H.V. Gupta, S. Sorooshian, A. Rousseau, R. Tourcotte. (American Geophysical Union, Washington, DC, 2002), pp. 49–68

D. Kavetski, G. Kuczera, S.W. Franks, Bayesian analysis of input uncertainty in hydrological modeling: 1. Theory. Water Resour. Res. **42**(3), W03407 (2006a)

D. Kavetski, G. Kuczera, S.W. Franks, Calibration of conceptual hydrological models revisited: 1. Overcoming numerical artefacts. J. Hydrol. **320**(1–2), 173–186 (2006b)

D. Kavetski, G. Kuczera, S.W. Franks, Calibration of conceptual hydrological models revisited: 2. Improving optimisation and analysis. J. Hydrol. **320**(1–2), 187–201 (2006c)

D. Kavetski, G. Kuczera, M. Thyer, B. Renard, Multistart Newton-type optimisation methods for the calibration of conceptual hydrological models, In Proceedings of Oxley, L. and Kulasiri, D. (eds) MODSIM 2007 International Congress on Modelling and Simulation, Christchurch, New Zealand. (Modelling and Simulation Society of Australia and New Zealand, 2007)

D. Kavetski, F. Fenicia, P. Reichert, C. Albert, Signature-domain calibration of hydrological models using approximate Bayes computation: Theory and comparison to existing applications. Water Resour. Res. in press, https://doi.org/10.1002/2017WR020528 (2018)

G.B. Kingston, H.R. Maier, M.F. Lambert, Bayesian model selection applied to artificial neural networks used for water resources modeling. Water Resour. Res. **44**, W04419 (2008)

J.W. Kirchner, Getting the right answers for the right reasons: Linking measurements, analyses, and models to advance the science of hydrology. Water Resour. Res. **42**(3), W03S04 (2006). https://doi.org/10.1029/2005WR004362

L.F. Konikow, J.D. Bredehoeft, Ground-water models cannot be validated. Adv. Water Resour. **15**, 75–83 (1992)

V. Koren, M. Smith, Q. Duan, Use of a priori parameter estimates in the derivation of spatially consistent parameter sets of rainfall-runoff models, in *Calibration of Watershed Models*, ed. by Q. Duan, H.V. Gupta, S. Sorooshian, A.N. Rousseau, R. Turcotte (AGU Press, Washington, DC, 2003)

R. Krzysztofowicz, Bayesian theory of probabilistic forecasting via a deterministic hydrologic model. Water Resour. Res. **35**(9), 2739–2750 (1999)

G. Kuczera, S. Franks, Testing hydrologic models: Fortification or falsification? in *Mathematical Modelling of Large Watershed Hydrology*, ed. by V.P. Singh, D.K. Frevert (Water Resources Publications, Littleton, 2002)

G. Kuczera, D. Kavetski, S.W. Franks, M. Thyer, Towards a Bayesian total error analysis of conceptual rainfall-runoff models: Characterising model error using storm-dependent parameters. J. Hydrol. **331**(1–2), 161–177 (2006)

E. Laloy, B. Rogiers, J.A. Vrugt, D. Mallants, D. Jacques, Efficient posterior exploration of a high-dimensional groundwater model from two-stage Markov chain Monte Carlo simulation and polynomial chaos expansion. Water Resour. Res. **49**(5), 2664–2682 (2013)

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

D.R. Legates, G.J. McCabe Jr., Evaluating the use of "goodness-of-fit" measures in hydrologic and hydroclimatic model validation. Water Resour. Res. **35**(1), 233–241 (1999)

J. Lerat, C. Pickett-Heaps, D. Shin, S. Zhou, P. Feikema, U. Khan, R. Laugesen, N. Tuteja, G. Kuczera, M. Thyer, D. Kavetski, Dynamic streamflow forecasts within an uncertainty framework for 100 catchments in Australia, in *Hydrology and Water Resources Symposium:*

The Art and Science of Water, (Engineers Australia, Barton, ACT, Australia, 2015), pp. 1396–1403
G. Lindstrom, B. Johansson, M. Persson, M. Gardelin, S. Bergstrom, Development and test of the distributed HBV-96 hydrological model. J. Hydrol. **201**, 272–288 (1997)
D.P. Loucks, J.R. Stedinger, D.A. Haith, *Water Resource Systems Planning and Analysis* (Prentice-Hall, Englewood Cliffs, 1981)
D.R. Maidment, *Handbook of Hydrology* (McGraw-Hill, New York, 1993)
P. Mantovan, E. Todini, Hydrological forecasting uncertainty assessment: Incoherence of the GLUE methodology. J. Hydrol. **330**(1–2), 368–381 (2006)
P. Mantovan, E. Todini, M.L.V. Martina, Reply to comment by Keith Beven, Paul Smith and Jim Freer on "Hydrological forecasting uncertainty assessment: Inconherence of the GLUE methodology". J. Hydrol. **338**, 319–324 (2007)
A. Marchi, E. Salomons, A. Simpson, A. Zecchin, H. Maier, Z. Wu, C. Stokes, W. Wu, G.C. Dandy, The battle of the water networks II (BWN-II). J. Water Resour. Plann. Manage. **140**, 04014009:04014001–04014009:04014014 (2014)
E.S. Martins, J.R. Stedinger, Generalized maximum-likelihood generalized extreme-value quantile estimators for hydrologic data. Water Resour. Res. **36**(3), 737–744 (2000)
D. McInerney, M. Thyer, D. Kavetski, J. Lerat, G. Kuczera, Improving probabilistic prediction of daily streamflow by identifying Pareto optimal approaches for modeling heteroscedastic residual errors. Water Resour. Res. **53**, 2199–2239 (2017)
H. McMillan, B. Jackson, M. Clark, D. Kavetski, R. Woods, Rainfall uncertainty in hydrologic modelling: An evaluation of multiplicative error models. J. Hydrol. **400**, 83–94 (2011)
M. Merriman, On the history of the method of least squares. Analyst **4**(2), 33–36 (1877)
D.A. Miller, R.A. White, A conterminous United States multi-layer soil characteristics data set for regional climate and hydrology modeling. Earth Interact. **2**, 2 (1999)
A. Montanari, E. Toth, Calibration of hydrological models in the spectral domain: An opportunity for scarcely gauged basins? Water Resour. Res. **43**, W05434 (2007)
M. Morawietz, C.-Y. Xu, L. Gottschalk, L.M. Tallaksen, Systematic evaluation of autoregressive error models as post-processors for a probabilistic streamflow forecast system. J. Hydrol. **407**(1–4), 58–72 (2011)
J.E. Nash, J.V. Sutcliffe, River flow forecasting through conceptual models. Part 1 – A discussion of principles. J. Hydrol. **10**, 282–290 (1970)
J.C. Neal, P.M. Atkinson, H.C. W, Flood inundation model updating using an ensemble Kalman filter and spatially distributed measurements. J. Hydrol. **336**, 401–415 (2007)
J. Neal, G. Schumann, P. Bates, A subgrid channel model for simulating river hydraulics and floodplain inundation over large and data sparse areas. Water Resour. Res. **48**, W11506 (2012)
D.J. Nott, L. Marshall, J. Brown, Generalized likelihood uncertainty estimation (GLUE) and approximate Bayesian computation: What's the connection? Water Resour. Res. **48**, W12602 (2012)
W.L. Oberkampf, J.C. Helton, C.A. Joslyn, S.F. Wojtkiewicz, S. Ferson, Challenge problems: Uncertainty in system response given uncertain parameters. Reliab. Eng. Syst. Saf. **85**(1–3), 11–19 (2004)
A. O'Hagan, J. Oakley, Probability is perfect, but we can't elicit it perfectly. Reliab. Eng. Syst. Saf. **85**(1–3), 239–248 (2004)
F. Pappenberger, K.J. Beven, Ignorance is bliss: Or seven reasons not to use uncertainty analysis. Water Resour. Res. **42**, W05302 (2006). https://doi.org/10.1029/2005WR004820
C. Perrin, C. Michel, V. Andreassian, Does a large number of parameters enhance model performance? Comparative assessment of common catchment model structures on 429 catchments. J. Hydrol. **242**(3–4), 275–301 (2001)
C. Perrin, C. Michel, V. Andreassian, Improvement of a parsimonious model for streamflow simulation. J. Hydrol. **279**(1–4), 275–289 (2003)
F. Pianosi, L. Raso, Dynamic modeling of predictive uncertainty by regression on absolute errors. Water Resour. Res. **48**, W03516 (2012)

W.H. Press, B.P. Flannery, S.A. Teukolsky, W.T. Vetterling, *Numerical Recipes in Fortran-77: The Art of Scientific Computing* (Cambridge University Press, Cambridge, 1992)
R. Pushpalatha, C. Perrin, N.L. Moine, V. Andréassian, A review of efficiency criteria suitable for evaluating low-flow simulations. J. Hydrol. **420**, 171–182 (2012)
Y. Qin, D. Kavetski, G. Kuczera, A robust Gauss-Newton algorithm for the optimization of hydrological models: 2. Benchmarking against industry-standard algorithms. Water Resour. Res. in review, https://doi.org/10.1029/2017WR022489 (2018)
P. Reichert, J. Mieleitner, Analyzing input and structural uncertainty of nonlinear dynamic models with stochastic, time-dependent parameters. Water Resour. Res. **45**, W10402 (2009)
P. Reichert, N. Schuwirth, Linking statistical bias description to multiobjective model calibration. Water Resour. Res. **48**, W09543 (2012)
P. Reichert, S.D. Langhans, J. Lienert, N. Schuwirth, The conceptual foundation of environmental decision support. J. Environ. Manag. **154**, 316–332 (2015)
R.H. Reichle, Data assimilation methods in the Earth sciences. Adv. Water Resour. **31**(11), 1411–1418 (2008)
B. Renard, E. Leblois, G. Kuczera, D. Kavetski, M. Thyer, S. Franks, Characterizing errors in areal rainfall estimates: Application to uncertainty quantification and decomposition in hydrologic modelling. H2009: 32nd Hydrology and Water Resources Symposium, Newcastle (Engineers Australia, Barton ACT, 2009), pp. 505–516
B. Renard, D. Kavetski, M. Thyer, G. Kuczera, S.W. Franks, Understanding predictive uncertainty in hydrologic modeling: Le challenge of identifying input and structural errors. Water Resour. Res. **46**, W05521 (2010). https://doi.org/10.1029/2009WR008328
B. Renard, D. Kavetski, E.T. Leblois, M. Thyer, G. Kuczera, S.W. Franks, Toward a reliable decomposition of predictive uncertainty in hydrological modeling: Characterizing rainfall errors using conditional simulation. Water Resour. Res. **47**(11), W11516 (2011)
B. Revilla-Romero, N. Wanders, P. Burek, P. Salamon, A. de Roo, Integrating remotely sensed surface water extent into continental scale hydrology. J. Hydrol. **543**(Pt B), 659–670 (2016)
J.D. Salas, Analysis and modeling of hydrologic time series, in *Handbook of Hydrology*, ed. by D.R. Maidment (McGraw-Hill, New York, 1993), pp. 19.11–19.72
L. Samaniego, R. Kumar, S. Attinger, Multiscale parameter regionalization of a grid-based hydrologic model at the mesoscale. Water Resour. Res. **46**(5), W05523 (2010)
H.H.G. Savenije, The art of hydrology. Hydrol. Earth Syst. Sci. **13**, 157–161 (2009)
B. Schaefli, H.V. Gupta, Do Nash values have value? Hydrol. Process. **21**(15), 2075–2080 (2007)
B. Schaefli, D. Kavetski, Bayesian spectral likelihood for hydrological parameter inference. Water Resour. Res. **53**, 6857–6884 (2017)
B. Schaefli, D.B. Talamba, A. Musy, Quantifying hydrological modeling errors through a mixture of normal distributions. J. Hydrol. **332**, 303–315 (2007)
G. Schoups, J.A. Vrugt, A formal likelihood function for parameter and predictive inference of hydrologic models with correlated, heteroscedastic and non-Gaussian errors. Water Resour. Res. **46**, W10531 (2010)
D.-J. Seo, H.D. Herr, J.C. Schaake, A statistical post-processor for accounting of hydrologic uncertainty in short-range ensemble streamflow prediction. Hydrol. Earth Syst. Sci. **3**, 1987–2035 (2006)
M. Shafii, B.A. Tolson, Optimizing hydrological consistency by incorporating hydrological signatures into model calibration objectives. Water Resour. Res. **51**(5), 3796–3814 (2015)
V.P. Singh, D.A. Woolhiser, Mathematical modeling of watershed hydrology. J. Hydrol. Eng. **7**(4), 270–292 (2002)
M. Sivapalan, G. Bloschl, L. Zhang, R. Vertessy, Downward approach to hydrological prediction. Hydrol. Process. **17**(11), 2101–2111 (2003a)
M. Sivapalan, K. Takeuchi, S.W. Franks, V.K. Gupta, H. Karambiri, V. Lakshmi, X. Liang, J.J. McDonnell, E.M. Mendiondo, P.E. O'Connell, T. Oki, J.W. Pomeroy, D. Schertzer, S. Uhlenbrook, E. Zehe, IAHS decade on predictions in ungauged basins (PUB). Hydrol. Sci. J. **48**(6), 857–880 (2003b)

B.E. Skahill, J. Doherty, Efficient accommodation of local minima in watershed model calibration. J. Hydrol. **329**, 122 (2006). in press

P. Smith, K.J. Beven, J.A. Tawn, Informal likelihood measures in model assessment: Theoretic development and investigation. Adv. Water Resour. **31**(8), 1087–1100 (2008)

T. Smith, A. Sharma, L. Marshall, R. Mehrotra, S. Sisson, Development of a formal likelihood function for improved Bayesian inference of ephemeral catchments. Water Resour. Res. **46**(12), W12551 (2010). https://doi.org/10.1029/2010WR009514

S. Sorooshian, J.A. Dracup, Stochastic parameter estimation procedures for hydrologic rainfall-runoff models: Correlated and heteroscedastic error cases. Water Resour. Res. **16**(2), 430–442 (1980)

J.R. Stedinger, R.M. Vogel, S.U. Lee, R. Batchelder, Appraisal of the generalized likelihood uncertainty estimation (GLUE) method. Water Resour. Res. **44**, W00B06 (2008)

V.L. Streeter, E.B. Wylie, *Fluid Mechanics*, First SI Metric Edition. (McGraw-Hill, Singapore, 1983)

L.M. Tallaksen, A review of baseflow recession analysis. J. Hydrol. **165**, 349–370 (1995)

A. Tarantola, *Inverse Problem Theory and Methods for Model Parameter Estimation* (Society for Industrial and Applied Mathematics, Philadelphia, 2005)

M. Thyer, G. Kuczera, Q.J. Wang, Quantifying parameter uncertainty in stochastic models using the Box-Cox transformation. J. Hydrol. **265**(1–4), 246–257 (2002)

M. Thyer, B. Renard, D. Kavetski, G. Kuczera, S. Franks, S. Srikanthan, Critical evaluation of parameter consistency and predictive uncertainty in hydrological modelling: A case study using Bayesian total error analysis. Water Resour. Res. **45**, W00B14 (2009)

B.A. Tolson, C.A. Shoemaker, Dynamically dimensioned search algorithm for computationally efficient watershed model calibration. Water Resour. Res. **43**, W01413 (2007)

A.F.B. Tompson, R. Ababou, L.W. Gelhar, Implementation of the 3-dimensional turning bands random field generator. Water Resour. Res. **25**(10), 2227–2243 (1989)

T. Toni, D. Welch, N. Strelkowa, A. Ipsen, M.P.H. Stumpf, Approximate Bayesian computation scheme for parameter inference and model selection in dynamical systems. J. R. Soc. Interface **6**(31), 187–202 (2009)

N.K. Tuteja, D. Shin, R. Laugesen, U. Khan, Q. Shao, E. Wang, M. Li, H. Zheng, G. Kuczera, D. Kavetski, G. Evin, M. Thyer, A. MacDonald, T. Chia, B. Le, *Experimental Evaluation of the Dynamic Seasonal Streamflow Forecasting Approach* (Australian Bureau of Meteorology, Melbourne, 2011)

N.K. Tuteja, S. Zhou, J. Lerat, Q.J. Wang, D. Shin, D.E. Robertson, Overview of communication strategies for uncertainty in hydrological forecasting in Australia, in *Handbook of Hydrometeorological Ensemble Forecasting*, ed. by Q. Duan, F. Pappenberger, J. Thielen, A. Wood, H.L. Cloke, J.C. Schaake (Springer, Berlin/Heidelberg, 2017), pp. 1–19

R.M. Vogel, Stochastic watershed models for hydrologic risk management. Water Secur. **1**, 28–35 (2017)

J.A. Vrugt, B.A. Robinson, Improved evolutionary optimization from genetically adaptive multimethod search. Proc. Natl. Acad. Sci. U. S. A. **104**(3), 708–711 (2007)

J.A. Vrugt, M. Sadegh, Toward diagnostic model calibration and evaluation: Approximate Bayesian computation. Water Resour. Res. **49**(7), 4335–4345 (2013)

J.A. Vrugt, H.V. Gupta, L.A. Bastidas, W. Bouten, S. Sorooshian, Effective and efficient algorithm for multiobjective optimization of hydrologic models. Water Resour. Res. **39**(8), 1214 (2003)

J.A. Vrugt, C.G.H. Diks, H.V. Gupta, W. Bouten, J.M. Verstraten, Improved treatment of uncertainty in hydrologic modeling: Combining the strengths of global optimization and data assimilation. Water Resour. Res. **41**(1), W01017 (2005)

J.A. Vrugt, C.J.F. ter Braak, M.P. Clark, J.M. Hyman, B.A. Robinson, Treatment of input uncertainty in hydrologic modeling: Doing hydrology backward with Markov chain Monte Carlo simulation. Water Resour. Res. **44**, W00B09 (2008)

J.A. Vrugt, C.J.F. ter Braak, C.G.H. Diks, B.A. Robinson, J.M. Hyman, D. Higdon, Accelerating Markov chain Monte Carlo simulation by differential evolution with self-adaptive randomized subspace sampling. Int. J. Nonlinear Sci. Numer. Simul. **10**(3), 273–290 (2009a)

J.A. Vrugt, C.J.F. ter Braak, H.V. Gupta, B.A. Robinson, Equifinality of formal (DREAM) and informal (GLUE) Bayesian approaches in hydrologic modeling? Stoch. Env. Res. Risk A. **23**(7), 1011–1026 (2009b)

J.A. Vrugt, C.J.F. ter Braak, C.G.H. Diks, G. Schoups, Hydrologic data assimilation using particle Markov chain Monte Carlo simulation: Theory, concepts and applications. Adv. Water Resour. **51**, 457–478 (2013)

Q.J. Wang, D.E. Robertson, Multisite probabilistic forecasting of seasonal flows for streams with zero value occurrences. Water Resour. Res. **47**, W02546 (2011)

Q.J. Wang, D.E. Robertson, F.H.S. Chiew, A Bayesian joint probability modeling approach for seasonal forecasting of streamflows at multiple sites. Water Resour. Res. **45**(5), W05407 (2009)

A.H. Weerts, G.Y.H. El Serafy, Particle filtering and ensemble Kalman filtering for state updating with hydrological conceptual rainfall-runoff models. Water Resour. Res. **42**, W09403 (2006)

W.D. Welsh, J. Vaze, D. Dutta, D. Rassam, J.M. Rahman, I.D. Jolly, P. Wallbrink, G.M. Podger, M. Bethune, M.J. Hardy, J. Teng, J. Lerat, An integrated modelling framework for regulated river systems. Environ. Model Softw. **39**, 81–102 (2013)

I. Westerberg, J.-L. Guerrero, J. Seibert, K.J. Beven, S. Halldin, Stage-discharge uncertainty derived with a non-stationary rating curve in the Choluteca River, Honduras. Hydrol. Process. **25**(4), 603–613 (2010)

I.K. Westerberg, H.K. McMillan, Uncertainty in hydrological signatures. Hydrol. Earth Syst. Sci. **19**(9), 3951–3968 (2015)

S. Westra, M. Thyer, M. Leonard, D. Kavetski, M. Lambert, A strategy for diagnosing and interpreting hydrological model nonstationarity, Water Resources Research, **50**(6), 5090–5113 (2014)

D.P. Wright, M. Thyer, S. Westra, Influential point detection diagnostics in the context of hydrological model calibration. J. Hydrol. **527**, 1161–1172 (2015)

K.K. Yilmaz, H.V. Gupta, T. Wagener, A process-based diagnostic approach to model evaluation: Application to the NWS distributed hydrologic model. Water Resour. Res. **44**, W09417 (2008)

P. Young, Data-based mechanistic modelling of environmental, ecological, economic and engineering systems. Environ. Model Softw. **13**(2), 105–122 (1998)

P.C. Young, M. Ratto, A unified approach to environmental systems modeling. Stoch. Env. Res. Risk A. **23**(7), 1037–1057 (2009)