

CHAPTER 3

MODEL CALIBRATION IN WATERSHED HYDROLOGY

Koray K. Yilmaz^{1,2}, Jasper A. Vrugt³, Hoshin V. Gupta⁴,
and Soroosh Sorooshian⁵

¹*Earth System Science Interdisciplinary Center, University of Maryland, College Park, MD 20742, USA*

²*NASA Goddard Space Flight Center, Laboratory for Atmospheres, Greenbelt, MD 20771, USA*

³*Center for NonLinear Studies, Los Alamos National Laboratory, Los Alamos, NM 87545, USA*

⁴*Department of Hydrology and Water Resources, The University of Arizona, Tucson, AZ 85721, USA*

⁵*Center for Hydrometeorology and Remote Sensing, Department of Civil and Environmental Engineering, University of California, Irvine, CA 92697-2175, USA*

Hydrologic models use relatively simple mathematical equations to conceptualize and aggregate the complex, spatially distributed, and highly interrelated water, energy, and vegetation processes in a watershed. A consequence of process aggregation is that the model parameters often do not represent directly measurable entities and must, therefore, be estimated using measurements of the system inputs and outputs. During this process, known as model calibration, the parameters are adjusted so that the behavior of the model approximates, as closely and consistently as possible, the observed response of the hydrologic system over some historical period of time. This Chapter reviews the current state-of-the-art of model calibration in watershed hydrology with special emphasis on our own contributions in the last few decades. We discuss the historical background that has led to current perspectives, and review different approaches for manual and automatic single- and multi-objective parameter estimation. In particular, we highlight the recent developments in the calibration of distributed hydrologic models using parameter dimensionality reduction sampling, parameter regularization and parallel computing.

Finally, this chapter concludes with a short summary of methods for assessment of parameter uncertainty, including recent advances in Markov chain Monte Carlo sampling and sequential data assimilation methods based on the Ensemble Kalman Filter.

1. Introduction

Hydrologic models serve as important tools for improving our knowledge of watershed functioning (understanding), for providing critical information in support of sustainable management of water resources (decision making), and for prevention of water-related natural hazards such as flooding (forecasting/prediction). Hydrologic models consist of a general structure, which mathematically represents the coupling of dominant hydrologic processes perceived to control the general model structure is then used for simulation and/or prediction of the response of the watershed to various inputs. The mathematical expressions embodied within.

No matter how sophisticated and spatially explicit, all hydrologic models aggregate (at some level of detail) the complex, spatially distributed vegetation and subsurface properties into much simpler homogeneous storages with transfer functions that describe the flow of water within and between these different compartments. These conceptual storages correspond to physically identifiable control volumes in real space, even though the boundaries of these control volumes are generally not known. A consequence of this aggregation process is that most of the parameters in these models cannot be inferred through direct observation in the field, but can only be meaningfully derived by calibration against an input - output record of the watershed response. In this process, the parameters are adjusted in such a way that the model approximates, as closely and consistently as possible, the response of the watershed over some historical period of time. The parameters estimated in this manner are therefore effective conceptual representations of spatially and temporally heterogeneous properties of the watershed. Therefore, successful application of any hydrologic model depends critically on the chosen values of the parameters.

In this chapter, we review the current state-of-the-art of model calibration in watershed hydrology. We discuss manual and automatic

parameter estimation techniques for calibration of lumped and spatially distributed hydrologic models. Specific methods include the widely used SCE-UA (Shuffled Complex Evolution - University of Arizona) and MOCOM-UA (Multi Objective COMplex evolution - University of Arizona) approaches for single- and multi-objective model calibration, step-wise (MACS: Multi-step Automatic Calibration Scheme) and sequential parameter estimation methods (DYNIA: DYNAMIC Identifiability Analysis; and PIMLI: Parameter Identification Method based on the Localization of Information), and emerging simultaneous multi-method evolutionary search methods (AMALGAM: A Multi-Algorithm Genetically Adaptive Multiobjective). We highlight recent developments in the calibration of distributed hydrologic models containing spatially distributed parameter fields, using parameter dimensionality reduction sampling, parameter regularization and parallel computing. The chapter concludes with a short summary on methodologies for parameter uncertainty assessment, including Markov chain Monte Carlo sampling and sequential data assimilation using the Ensemble Kalman Filter (EnKF); here we discuss the RWM (Random Walk Metropolis), SCEM-UA (Shuffled Complex Evolution Metropolis-University of Arizona), DREAM (DiffeREntial Evolution Adaptive Metropolis) and SODA (Simultaneous Optimization and Data Assimilation) sampling algorithms. Note that, although our discussion is limited to watershed models, the ideas and methods presented herein are applicable to a wide range of modeling and parameter estimation problems.

2. Approaches to Parameter Estimation for Watershed Models

There are two major approaches to parameter estimation for watershed models. In *a priori* estimation, values of the model parameters are specified without recourse to the observed dynamic hydrologic response (e.g. streamflow) of the watershed under study. Calibration, on the other hand, involves the selection of a parameter set that generates model responses that reproduce, as closely as possible, the historically observed hydrologic response of a particular watershed. Therefore, calibration can only be performed when long-term historical measurements of input-state-output behavior of the watershed (including streamflow, precipitation and potential evaporation) are available.

Parameter estimation strategies are intimately tied to the degree of hydrologic process representation embedded within the model. Hydrologic models can be classified accordingly as *conceptual* or *physically-based*. Most hydrologic models in use today are of the *conceptual* type, that conceptualize and aggregate the complex, spatially distributed, and highly interrelated water, energy, and vegetation processes in a watershed into relatively simple mathematical equations without exact treatment of the detailed underlying physics and basin-scale heterogeneity. Typical examples of conceptual type models are the SAC-SMA (SACramento Soil Moisture Accounting) model,^{1,2} and the HBV (Hydrologiska Byråns Vattenbalansmodell) model³. Due to hydrologic process aggregation, the parameters in these models cannot generally be measured directly in the field at the desired scale of interest. Instead, when using conceptual type models, only the ranges of feasible parameter values can generally be specified *a priori* (perhaps with the combined knowledge of model structure and of dominant watershed processes). Calibration is then employed to select parameter estimates (from within the *a priori* defined ranges) that capture, as closely and consistently as possible, the historical record of the measured (target) hydrologic response of the watershed the model is intended to represent.

Spatially distributed *physically-based* hydrologic models contain a series of partial differential equations describing physical principles related to conservation of mass, momentum (and energy). Typical examples of physically-based models are MIKE-SHE⁴ (Système Hydrologique Européen) and KINEROS⁵ (KINematic Runoff and EROSion). Their spatially distributed physically-based structure provides two potential strengths: 1) the ability to account for the spatial variability of runoff producing mechanisms and 2) the ability to infer model parameter values directly from spatio-temporal data by establishing physical or conceptual relationships between observable watershed characteristics (e.g., geology, topography, soils, land cover, etc.) and the parameters for the hydrologic processes represented in the model.^{6,5,7,8,9} The latter will be defined as the *ORFDQ' priori* parameter estimation approach and is particularly valuable for implementing hydrologic models in poorly gauged and ungauged watersheds where local response data is sparse or non-existent. In general, parameters estimated via the local *a priori* approach will still require some degree of fine-tuning via a calibration approach to obtain effective values that account for the influencing factors, such as heterogeneity, emergent processes,

differences in scale between model (larger scale) and the embedded hydrologic theory (developed at point/small scale). This refinement process ensures proper consistency between the model input-state-output behavior and the available response data.^{10,11,12}

Another parameter estimation strategy, developed mainly for the implementation of conceptual type of models in ungauged basins, is the *a priori* approach.^{13,14,15,16,17} This approach involves the development of regional regression relationships between the model parameter values estimated for a large number of gauged basins (via calibration) and observable watershed characteristics (i.e. landcover and soil properties) at those locations. The idea is that these relationships can be used to infer parameter estimates for ungauged basins based on observable watershed characteristics. A major assumption of the regional *a priori* approach is that the calibrated model parameters are uniquely and clearly related to observable watershed properties. This assumption can be difficult to justify when many combinations of parameters are found to produce similar model responses due to parameter interaction,¹⁸ measurement uncertainty¹⁹ and model structural uncertainty²⁰, and can therefore result in ambiguous and biased relationships between the parameters and the watershed characteristics¹⁷. One way to improve the efficiency of the regionalized approach is to impose conditions (via watershed characteristics) on the calibrated parameters.²¹ In an alternative approach, Yadav et al.²² proposed regionalization of streamflow indices. In this approach, relationships between streamflow indices and physical watershed characteristics are established at the gauged locations. The regionalized flow indices, providing dynamic aspects of the ungauged watersheds, are then used to constrain hydrologic model predictions (and parameters). One advantage of this approach is that, the regionalized indices are independent of model structure and therefore can be used to constrain any watershed model.

The paragraphs above provide a broad overview of approaches commonly used by the hydrologic community to specify values of the parameters in hydrologic models, and in the following sections we therefore provide a more detailed overview of various calibration strategies that have been developed within the water resources context and have found widespread use in the hydrologic community. We discuss these methods within the

context of their historical development, including current and future perspectives.

2.1. Model Calibration

For a model to be useful in prediction, the values of the parameters need to accurately reflect the invariant properties of the components of the underlying system they represent. Unfortunately, in watershed hydrology, many of the parameters can generally not be measured directly, but can only be meaningfully derived through calibration against historical record of dynamic response (traditionally streamflow) data. Calibration is an iterative process in which the model parameters are adjusted so that the dynamic response of the model represents, as closely as possible, the observed target response (e.g. outlet streamflow) of the watershed. Figure 1 provides a schematic representation of the resulting model calibration problem. In this figure, \oplus represents observations of the forcing (rainfall) and streamflow response that are subject to measurement errors and uncertainty, and therefore may be different from the true values. Similarly, Φ represents the hydrologic model with functional response

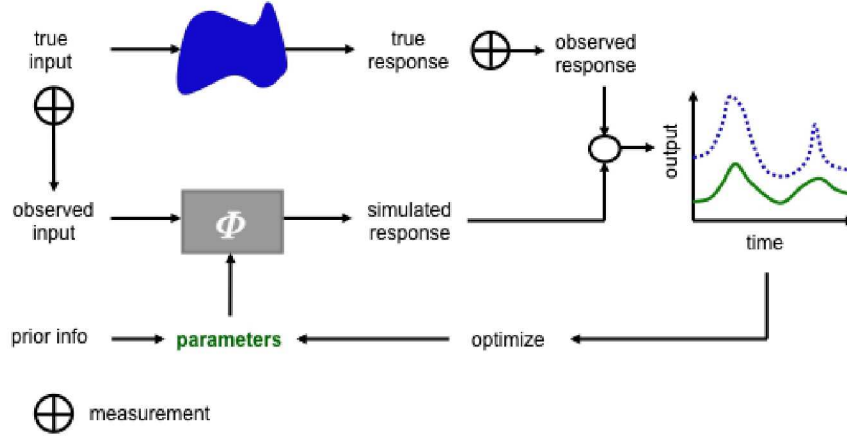


Fig. 1. A schematic representation of the general model calibration problem. The model parameters are iteratively adjusted so that the predictions of the model, Φ , (represented with the solid line) approximate as closely and consistently as possible the observed response (represented with dotted line).

Mathematically, the model calibration problem depicted in Figure 1 can be formulated as follows. Let $\tilde{S} = \Phi(\psi, X, \theta)$ denote the streamflow predictions $\tilde{S} = \{\tilde{s}_1, \dots, \tilde{s}_n\}$ of the model Φ with observed forcing X (rainfall and potential evapotranspiration), state vector ψ , and model parameters θ . Let $S = \{s_1, \dots, s_n\}$ represent a vector with n observed streamflow values. The difference between the model-predicted and measured streamflow can be represented by the residual vector E as:

$$E(\theta) = \{T(\tilde{S}) - T(S)\} = \{T(\tilde{s}_1) - T(s_1), \dots, T(\tilde{s}_n) - T(s_n)\} = \{e_1(\theta), \dots, e_n(\theta)\} \quad (1)$$

where $T(\cdot)$ allows for various monotonic (such as logarithmic) transformations of the model outputs.

Traditionally, we seek values for the model parameters that result in a minimal discrepancy between the model predictions and observations. This can be done by minimizing an aggregate summary statistic of the residuals:

$$F(\theta) = F\{e_1(\theta), e_2(\theta), \dots, e_n(\theta)\} \quad (2)$$

where the function $F(\theta)$ allows for various user-selected linear and non-linear transformations of the residuals and is interchangeably called a

3FULWHULRQ' □ □ 3PHDVXUH' □ RU □ 3REMHFWLYH □ IXQFWLRQ' □ LQ □ WKH □ ZDWHUVKH

literature. Note that in typical time series analysis the influence of the initial condition, ψ , on the model output diminishes with increasing distance from the start of the simulation. In those situations, it is common to use a spin-up period to reduce sensitivity to state-value initialization.

Boyle et al.²³ classified the process of model calibration into three levels of increasing sophistication. In level zero, approximate ranges for the parameter estimates are specified by physical reasoning that incorporates available data about static watershed characteristics (e.g. geology, soil, land cover, slope, etc.), via lookup tables or by borrowing values from similar watersheds. At this level, only crude *a priori* ranges of the parameters are estimated without conditioning on the input-output streamflow response data. In level one, the parameter ranges are refined by identifying and analyzing the characteristics of specific segments of the response data, that are thought to be controlled by a single or a group of parameter(s). In this level, the effects of parameter interactions are ignored. Finally, in level two, a detailed analysis of parameter interactions and performance trade-offs is performed using a carefully chosen representative period of historical data (i.e. calibration data) to further refine the parameter ranges or select a representative parameter set. This level involves a complex, iterative process of parameter adjustments to bring the simulated response as closely and consistently as possible to the observed watershed response.

Calibration can be further divided into two types, depending on whether this iterative process is being guided manually by an expert hydrologist or automatically by a computer following pre-defined algorithmic rules. These approaches are called *manual calibration* and *automated calibration*, respectively.

2.1.1. Overview of the Manual Calibration Approach

Manual calibration is a guided trial-and-error process performed by an expert hydrologist, involving complex knowledge-based analyses to match the perceived hydrologic processes in the watershed with their conceptual equivalents represented in the model structure. This interactive process can involve a variety of graphical interfaces and a multitude of performance measures to transform the historical data into information that will aid the hydrologist in decision-making. Although progressive steps during manual calibration of a model are generally established via pre-defined guidelines, the actual sequence of procedures

will vary based on the experience and training of the modeler, the properties of the data and characteristics of the watershed system being modeled.²³ Successful manual calibration, therefore, requires a good knowledge of the physical and response characteristics of the watershed, as well as a good understanding of the structure and functioning of the various model components and parameters. The major aim is to find values for the parameters that are consistent with the hydrologic processes they were designed to represent.²⁴ The consistency between model simulations of hydrologic behaviors for which observations are available is examined at various timescales and time-periods to try and isolate individual effects of each parameter. Hydrologic behaviors include, for example, annual averages to understand the long-term water balance dynamics, seasonal and monthly averages to identify the trends and low-high flow periods, extended recession periods to understand watershed baseflow characteristics and event based measures to analyze the shape and timing of floods.^{2,25} Manual calibration is expected to produce process-based (i.e. conceptually realistic) and reliable predictions. The National Weather Service (NWS) of the United States, for example, primarily uses a manual approach (assisted with automatic techniques) to estimate the parameters of their lumped hydrologic models used in operational streamflow forecasting. The manual calibration, however, is a time- and labor-intensive process involving many subjective decisions. Due to this subjectivity, different modelers will most likely produce different model parameter values for the same watershed. Another difficulty in manual calibration is the ever-increasing complexity of watershed models. As more parameters are added, it becomes more difficult to estimate them accurately; parameter interaction and compensating effects on the modeled variables make it difficult to estimate individual parameters. Note that Bayesian methods (see Sec. 6) provide a general framework for explicit use of expert knowledge/belief in the form of priors. An excellent and comprehensive discussion of manual calibration is given by three recent reviews.^{23,24,25}

2.1.2. Overview of Automated Calibration Approaches

Automated parameter identification (calibration) methods rely on an *a priori* model structure, optimization algorithm and one or more mathematical measures of model performance (often called objective function, criterion or measure) to estimate the model parameters using a

historical record of observed response data. The advantages of the automated calibration approach are not difficult to enumerate. Such methods use objective, rather than visual and subjective, measures for performance evaluation, and exploit the power and speed of computers to efficiently and systematically search the feasible model parameter space. Within this context, the goal has been to develop an objective strategy for parameter estimation that provides consistent and reproducible results independent of the user.²³ A potential disadvantage is that automatic calibration algorithms can, if not properly designed, return values of the model parameters that are deemed to be hydrologically unrealistic. It is, therefore, optimized parameters. A traditional approach, accepted as a minimum UHTXLUHPHQW IRU HYDOXDWLRQ RI SDUDPHWHU UREXVWQHVV LV modeled and observed response for an independent time period that is not used for calibration. This can be done using a split-sample approach and should cover a long enough time period so as to contain various ranges of hydrologic conditions (e.g. dry and wet periods), and, if possible, using multiple response data.^{26,27}

In the following sections, we will discuss single- and multi-objective model calibration strategies. Single objective methods utilize a single measure of closeness between the simulated and observed variables, whereas multi-objective approaches attempt to emulate the strengths of the manual calibration approach by simultaneously using multiple different measures of performance.

3. Single Criterion Automatic Calibration Methods

Traditional automated approaches to model calibration utilize a single mathematical criterion (sometimes defined as objective function or measure) WR VHDUFK IRU D XQLTXH 'EHVW' SDUDPHWHU VHW SQ function (Eq. 2) can be defined as an aggregate statistical summary TXDQWLILQJ WKH 'FORVRQ' EHWZHHQ WKH VLPXODWHG and observed hydrologic variable(s). The objective function largely influences the outcome of the automated calibration procedure and, hence, should be carefully selected based on the goal of the modeling. Traditionally, automated calibration problems seek to minimize the discrepancy between the model predictions and observations by minimizing the following additive simple least square (SLS) objective function (or its variations) with respect to θ :

$$F_{SLS}(\theta) = \sum_{i=1}^n e_i(\theta)^2 \quad (3)$$

Often the Root Mean Square Error (RMSE) criterion, is used because it has the same units as that of the variable being estimated (e.g. streamflow discharge). An RMSE objective function puts strong emphasis on simulating the high flow events; to increase the emphasis on the low flows, it can be used after performing a log-transformation of the output variables.

The popularity of the SLS function stems from its statistical properties; the SLS is an unbiased estimator under strong assumptions related to the distribution of the residuals -- i.e. the residuals are pairwise independent, have constant variance (homogenous) and normally distributed with a mean value of zero (see also Sorooshian and Gupta,²⁸ and Gupta et al.²⁹). However, the validity of these assumptions is questionable within a hydrologic context, as most hydrograph simulations published to date in the literature show significant non-stationarity of error residuals.^{30,31}

3.1. A Historical Perspective

A powerful optimization algorithm is a major requirement to ensure finding parameter sets that best fit the data. Optimization algorithms iteratively explore the response surface (a surface mapped out by the objective function values in the parameter space). Optimization algorithms developed in the past to solve the nonlinear SLS optimization problem stated in Eq. (3) may be classified as local search methodologies if they are designed to seek a systematic improvement of the objective function using an iterative search starting from a single arbitrary initial point in the parameter space, or as stochastic global search methods if multiple concurrent searches from different starting points are conducted within the parameter space. Local optimization approaches can be classified into two categories, based on the type of search procedure employed, namely derivative-based (gradient) and derivative-free (direct) methods.

Gradient-based methods make use of the estimates of the local downhill direction based on the first and/or second derivative of the response surface with respect to each individual model parameter.³² The simplest gradient-based algorithm is that of steepest descent, which searches along the first-derivative direction for iterative improvement of

the objective function. Newton-type methods, such as the Gauss-Newton family of algorithms, are examples of second derivative algorithms. Gupta and Sorooshian³³ and Hendrickson et al.³⁴ demonstrate how analytical or numerical derivatives can be computed for conceptual watershed models.

Derivative-based algorithms will evolve towards the global minimum in the parameter space in situations where the objective function exhibits a topographical convex shape in the entire parameter domain. Unfortunately, numerous contributions to the hydrologic literature have demonstrated that the response surface seldom satisfies these restrictive conditions, and exhibits multiple optima in the parameter space. Local gradient-based search algorithms are not designed to handle these peculiarities, and therefore often prematurely terminate their search at a final solution that is dependent on the starting point in the parameter space. Another related problem is that many of the hydraulic parameters typically demonstrate significant interactions, because of an inability of the observed experimental data to properly constrain all of the calibration parameters.

Direct search methods sample the value of the objective function in a systematic manner without computing derivatives of the response surface with respect to each parameter. Popular examples of direct search methods include the Simplex Method,³⁵ the Pattern Search Method³⁶ and the Rotating Directions Method of Rosenbrock³⁷. Many studies have focused on comparative performance analysis of local search methods for calibration of watershed models.^{38,39,40} Their general conclusion was that local search methods were not powerful enough to reliably find the best (global optimum) values of the watershed model parameters. The main limitation of local search methods is that, like gradient-based algorithms, their outcome is highly dependent on their initial starting point. For example, local search algorithms are prone to getting trapped in local basins of attraction (local minima) and, as argued above, may become confused in finding the preferred direction of improvement in the presence of threshold structures and other undesirable irregularities of the response surface.

Among others, Moore and Clarke⁴¹ and Sorooshian and Gupta⁴² pointed out that the causes of the above difficulties were mainly difficulties concerned with the underlying model structure and that local search methods were not powerful to do the job. Gupta and Sorooshian⁴³, for example, focused on model structural inadequacies in the SAC-SMA

(Sacramento Soil Moisture Accounting) model and showed that parameter identifiability can be improved by a careful re-parameterization of the percolation function. Other seminal contributions^{30,42} concluded that a properly chosen objective function, which can better recognize the stochastic nature of the errors in the calibration data (such as those derived using Maximum Likelihood Theory), can result in smoother response surfaces for which the global optimum is easier to identify. In later studies,^{30,42} the authors pointed out that streamflow measurements contain errors that are temporally auto-correlated and heteroscedastic (having non-constant, magnitude dependent variance) and also introduced a Heteroscedastic Maximum Likelihood Estimator (HMLE) that accounts for non-stationary error variance arising from various sources, including the rating curve used to convert the stage measurements (in units of height) into runoff rate (in units of discharge, e.g. cubic feet per day). In a parallel work, Kuczera⁴⁴ proposed a methodology based on Bayesian statistics to properly account for the measurement error properties of the data while predicting the confidence bounds for the parameter estimates. Recently, Kavetski et al.^{45,46} and Vrugt et al.^{47,48} have extended that approach to account for error in rainfall depths as well. Another commonly used methodology applies a parameterized power transformation to the streamflow data to stabilize the heteroscedastic measurement error:⁴⁹

$$\hat{y} = \frac{(y+1)^\lambda - 1}{\lambda} \quad (4)$$

where y and \hat{y} represent flows in the original and transformed spaces respectively and λ is the transformation parameter (a commonly used value is $\lambda \sim 0.0$ to 0.3).

Other researchers have investigated the requirements for calibration data and pointed out that the quantity and quality of calibration data play a critical role in controlling the success of the calibration procedure. These studies concluded that the informativeness of the data is far more important than the length and amount used for calibration.^{50,51,52,53,54}

The convergence problems encountered with local search algorithms have inspired researchers to develop and test global search algorithms for calibration of watershed models. While local optimization methods rely on a single initial point within the feasible parameter space to start the search, global optimization methods utilize multiple concurrent searches from different starting points to reduce the chance of getting stuck in a

function value so that the first point represents the smallest criterion value and the last point represents the largest criterion value (assuming that the goal is to minimize the criterion value).

- (3) Partitioning into complexes: partition the s points into p complexes, each containing m points. The complexes are partitioned such that the first complex contains every $p(k-1)+1$ ranked point, the second complex contains every $p(k-1)+2$ ranked point, and so on, where $k = 1, 2, \dots, m$.
- (4) Complex evolution: evolve each complex according to the competitive complex evolution algorithm, which is based on the Simplex downhill search scheme.³⁵ The evolution procedure JHQUHDWHV □ QHZ □ SRLQWV □ FDOOHG □ ³RIIVSULQJ' □ WKDW □ □ RQ □ DYHU within the improvement region.
- (5) Complex shuffling: combine the points in the evolved complexes into a single sample population; sort the sample population in order of increasing criterion value; shuffle (i.e. re-partition) the sample population into p complexes according to the procedure specified in Step (3).
- (6) Check for convergence: if any of the pre-specified termination criteria are satisfied, stop; otherwise, continue. Termination criteria can be specified as maximum number of iterations (or maximum number of shuffling) or parameter convergence.

Experience with the method has indicated that the effectiveness and efficiency of the SCE-UA algorithm is influenced by the choice of a few algorithmic parameters. Duan et al.^{66,73} performed sensitivity studies and suggested practical guidelines for selecting these algorithmic parameters according to the degree of difficulty of the optimization problem. The primary parameter to be selected is the number of complexes, p . The above studies showed that the dimension of the calibration problem (i.e. number of parameters to be optimized), n , is the primary factor determining the proper choice of p ; practically (if no other information is available) p is set to the greater value between 2 or n (see also Kuczera⁷²). The size of a complex, m , is generally chosen to be equal to $2n + 1$. Accordingly, the sample (population) size, s , becomes the product $p \times m$. The number of offspring, β , that can be generated by each independently evolving complex between two consecutive shuffles is the same as the complex size ($2n + 1$), the size of each sub-complex selected for generation of an offspring (via Simplex scheme) is $n + 1$ and defines

a first order approximation to the objective function space. The number of consecutive offspring generated by each sub-complex, α , is equal to 1. In selecting these algorithmic parameters, a balance between algorithm effectiveness and efficiency should be sought. For instance, selecting a large number of complexes increases the probability of converging to the global optimum, however at the expense of a larger number of simulations (and hence longer computational time). The SCE-UA code is available free of charge from the following web address: www.sahra.arizona.edu/software.

While significant progress has been made in the use of global optimization algorithms for parameter estimation, the current generation of optimization algorithms typically implements a single operator (i.e. Simplex search in the case of SCE-UA) for population evolution. Reliance on a single model of natural selection and adaptation presumes that a single method can efficiently evolve a population of potential solutions through the parameter space and work well for a large range of problems. However, existing theory and numerical benchmark experiments have demonstrated that it is impossible to develop a single universal algorithm for population evolution that is always efficient for a diverse set of optimization problems.⁷⁴ This is because, the nature of the response surface often varies considerably between different optimization problems, and often dynamically changes en route to the global optimal solution. It therefore seems productive to develop a search strategy that adaptively updates the way it generates offspring based on the local peculiarities of the response surface.

In light of these considerations, Vrugt and Robinson⁷⁵ and Vrugt et al.⁷⁶ have recently introduced a new concept of self-adaptive multi-method evolutionary search. This approach, termed as A Multi Algorithm Genetically Adaptive Method (AMALGAM), runs a diverse set of optimization algorithms simultaneously for population evolution and adaptively favors individual algorithms that exhibit the highest reproductive success during the search. By adaptively changing preference to individual algorithms during the course of the optimization, AMALGAM has the ability to quickly adapt to the specific peculiarities and difficulties of the optimization problem at hand. A brief algorithmic description of AMALGAM for solution of multi-objective optimization problems is given in Section 4. Synthetic benchmark studies covering a diverse set of problem features, including multimodality, ruggedness, ill-conditioning, non-separability, interdependence (rotation) and high-

dimensionality, have demonstrated that AMALGAM significantly improves the efficiency of evolutionary search.^{75,76} An additional advantage of self-adaptive search is that the need for algorithmic parameter tuning is reduced, thus increasing applicability to solving search and optimization problems in many different fields of study. An extensive algorithmic description of AMALGAM, including comparison against other state-of-the-art optimization methods, can be found in Vrugt and Robinson⁷⁵ and Vrugt et al.⁷⁶ The AMALGAM code is available from the second author upon request.

3.3. Limitations of Single Criterion Methods

Despite these algorithmic advances, automated model evaluation strategies that rely on a single regression-based aggregate measure of performance (e.g. RMSE) are, in general, weak and make it unnecessarily difficult to isolate the effects of different parameters on the model output.^{10,77,78} Hence, two different parameter combinations might give completely different streamflow responses, but result in very similar values of the objective function. This is undesirable. A major reason for this is the loss (or masking) of valuable information inherent in the process of projecting from the high dimension of the data set ($\mathcal{R}^{\text{Data}}$) down to the single dimension of the residual-based summary statistic (\mathcal{R}^1), leading to an ill-posed parameter estimation problem ($\mathcal{R}^{\text{Parameter}} < \mathcal{R}^{\text{Data}}$).⁷⁷ To avoid (or at least minimize) this problem, an optimization strategy must necessarily make use of multiple, carefully selected, measures of model performance, thereby more closely matching the number of unknowns (the parameters) with the number of pieces of information (the measures), resulting in a better-posed identification problem. There is, therefore, an urgent need to develop mathematical theory that more convincingly proves this line of thought and provides ways forward to improve parameter inference. This is especially pressing within the context of spatially distributed models that contain a manifold of parameters for which little compelling *a priori* information is available about appropriate values.

4. Multi-Criteria Automatic Calibration Methods

Multi-criteria analysis can be used to assimilate information from multiple non-commensurable (i.e. not measurable by the same standard)

sources.¹⁰ The goal is to increase the extraction of information content from the data (decrease the gap between measure and parameter dimensions), by properly expressing the different important aspects of model performance. For instance, a number of criteria can be formulated, each of which is sensitized to a specific watershed output flux (e.g. water, energy, chemical constituents) for which measurements are available.^{79,27} In principle, each criterion can also be designed to isolate a different characteristic behavior of some component of the physical system.⁸⁰ Note that the process of interactive manual-expert evaluation and calibration of a model, following a process-guided set of rules, actually follows a powerful (albeit somewhat subjective) multi-criteria approach, wherein a variety of graphical and numerical tools are used to highlight different aspects of model response.^{23,25,27} A major advantage of the automated multi-criteria approach is that various aspects of the manual calibration strategy can be absorbed into the calibration process, thus strengthening the physical basis of identified parameters. Many automated (or semi-automated) multi-criteria calibration strategies have been proposed for calibration of watershed models. These strategies can be broadly classified into simultaneous, step-wise and constraining approaches. These three approaches are discussed next.

4.1. Simultaneous Multi-criteria Calibration Approach

The simultaneous multi-criteria approach finds a set of solutions (so-
FDOOHG□³DUHWR□RSWLPDO'□UHJLRQ□□WKDW□VLPXOWDQHRXVO\□RSWLPL]H□
optimization run) and trade-off the performance of several user-selected
criteria that measure different aspects of model performance.^{10,81,82,11} In
general, the multi-criteria model constraining problem can be expressed
in the following form:¹⁰

$$\min_{w.r.t.\theta} F_{MC}(\theta) = \{F_1(\theta), F_2(\theta), \dots, F_n(\theta)\} \quad \theta \in \Theta \quad (5)$$

where $F_1(\theta), F_2(\theta), \dots, F_n(\theta)$ represent the different performance criteria summarizing information related to various components of the physical system. To solve this multi-criteria problem, model parameter sets (θ) are systematically sampled from their *a priori* region (Θ) in search of solutions that simultaneously minimize all of these criteria. As is well known, it is generally not possible to satisfy all of the criteria simultaneously. The solution to this minimization problem (Eq. 5) is

generally not unique but takes the form of a Pareto surface that characterizes the trade-offs in its ability to satisfy all of the competing criteria.^{10,81}

As a commonplace illustration, consider the migration of birds from Scandinavia to Africa and backwards that simultaneously considers flight time and energy use (Figure 2). For this situation, there is no single optimal solution. Rather, there is a family of tradeoff solutions along a Pareto surface (Figure 2b) in which improvement in one objective (say, reduction in flight time) comes only at the expense of increased energy use per day (second objective). In other words, moving from one solution to another along the Pareto surface will result in the improvement of at least one criterion while deteriorating at least one other. Analysis of the size and properties of the Pareto region can provide insights into possible model improvements as well as degrees of confidence in different aspects of the model predictions.^{10,84}

Many computational approaches to deriving efficient estimates of the Pareto solution set have been proposed.^{81,82,85,86} A pioneering algorithm that provides an approximation to the Pareto optimal region in a single

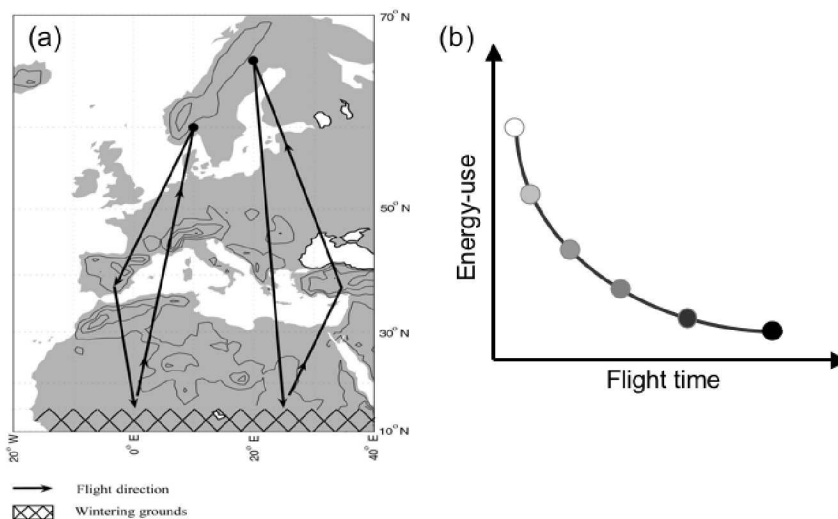


Fig. 2. An illustration of the Pareto optimality. a) Map showing migration paths of birds from Scandinavia to Africa and backwards. b) Birds trying to simultaneously optimize flight time and energy use. There exists a family of trade-off solutions along a curve FDOOHG□WKHRSULFDORJ(are Wright et al.⁸³)

optimization run is called the MOCOM-UA (Multi-Objective COMplex evolution algorithm) developed at the University of Arizona.⁸¹ MOCOM-UA is a general-purpose multi-objective global optimization strategy that is based on an extension of the SCE-UA population evolution method.⁶² It combines the strengths of the controlled random search method,⁵⁸ Pareto ranking⁸⁷ and a multi-objective downhill simplex search method.

In brief, the MOCOM-UA method starts with an initial sampling of a population of s points distributed randomly throughout the n -dimensional feasible parameter space, Θ . In the absence of prior information about the location of the Pareto region, a uniform sampling distribution is used. For each sampled point, the multi-criteria vector, $F_{MC}(\theta)$, is calculated and the population is ranked and sorted using the Pareto-ranking procedure presented by Goldberg⁸⁷. Simplexes of $n + 1$ points are then selected from the population according to a robust rank-based selection method suggested by Whitley⁸⁸. A multi-objective extension of the downhill simplex method is used to evolve each simplex in a multi-criteria improvement direction and generate new (better) points. Iterative application of the ranking and evolution procedures causes the entire population to converge towards the Pareto optimum. The procedure terminates automatically when the population converges to the Pareto Region with all points mutually non-dominated. The final population provides a fairly uniform approximation of the Pareto solution space $P(\Theta)$. The details of the MOCOM-UA algorithm can be found in Yapo et al.⁸¹ The MOCOM-UA code is available from the following web address: www.sahra.arizona.edu/software.

MOCOM-UA has been successfully applied to a number of multi-criteria calibration studies of hydrologic^{23,10} and land-surface models.^{89,90} However, some studies reported that the MOCOM-UA algorithm has a tendency to cluster the solutions in the center of Pareto region⁹¹ and may require a very large number of model runs for convergence.⁹⁰ The MOSCEM-UA (Multiobjective Shuffled Complex Evolution Metropolis - University of Arizona) method⁸² is especially designed to overcome some of these convergence problems, but its stochastic nature causes the Pareto solution set to contain irregularities with bumpy fronts / surfaces as a consequence.

The AMALGAM evolutionary search algorithm recently developed by Vrugt and Robinson⁷⁵, has shown to be the method of choice for solving multi-objective optimization problems. This method utilizes self-

adaptive multi-method evolutionary search and is more efficient and robust than existing search approaches including the commonly used Strength Pareto Evolutionary Algorithm (SPEA2)⁹² and Non-Dominated Sorting Genetic Algorithm (NSGA-II).⁹³ The AMALGAM algorithm is initiated using a random initial population P_0 of size N , generated using Latin hypercube sampling. Then, each parent is assigned a rank using the fast non-dominated sorting (FNS) algorithm⁹³. A population of offspring Q_0 , of size N , is subsequently created by using the multi-method search concept that lies at the heart of the AMALGAM method. Instead of implementing a single operator for reproduction, we simultaneously use k individual algorithms to generate the offspring, $Q_0 = \{Q_0^1, \dots, Q_0^k\}$. These algorithms each create a pre-specified number of offspring points, $N = \{N_t^1, \dots, N_t^k\}$, from P_0 using different adaptive procedures. After creation of the offspring, a combined population $R_0 = P_0 \cup Q_0$ of size $2N$ is created and R_0 ranked using FNS. By comparing the current offspring with the previous generation, elitism is ensured since all previous non-dominated members will always be included in R .^{93,94,95} Finally, members for the next population P_1 are chosen from subsequent non-dominated fronts of R_0 based on their rank and crowding distance.⁹³ The new population P_1 is then used to create offspring using the method described below, and the aforementioned algorithmic steps are repeated until convergence is achieved.

To ensure that the most offspring to the new population, we update $\{N_t^1, \dots, N_t^k\}$ according to:

$$N_t^i = N \frac{P_t^i}{N_{t-1}^i} \left[\sum_{i=1}^k \frac{P_t^i}{N_{t-1}^i} \right]^{-1} \quad i = 1, \dots, k \quad (6)$$

The term (P_t^i / N_{t-1}^i) is the ratio of the number of offspring points an algorithm contributes to the new population, P_t^i , and the corresponding number the algorithm created in the previous generation (N_{t-1}^i). The rest of the expression scales the reproductive success of an individual algorithm to the combined success of all the algorithms.

Benchmark results using a set of well-known multi-objective test problems show that AMALGAM is three to ten times more efficient than

existing multi-objective optimization algorithms. Initial applications to hydrologic parameter estimation problems have reported similar efficiency improvements.^{96,97}

Notwithstanding this progress made, approximation of the entire Pareto surface can be computationally too expensive. This is especially true within the context of distributed hydrologic modeling, for which approximate Pareto-optimal solutions can be obtained by lumping the various individual performance criteria within a single aggregate function through scalarization:

$$F_{agg}(\theta) = \sum_{i=1}^k \omega_i \tau_i F_i(\theta) \quad (7)$$

where k denotes the number of criteria, and ω_i and τ_i are scaling transformations, respectively, applied to each criterion. Here, the weights define the relative contributions of each criterion to the aggregate function, F_{agg} , such that $\omega_1 + \omega_2 + \dots + \omega_k = 1$, and can be subjectively assigned to place greater or lesser emphasis on some aspect of the model. Further, they can also be formalized to reflect our knowledge regarding the relative degree of uncertainty in the various measurement data sources.⁹⁸ It can be shown theoretically that if the Pareto region is convex, its shape can be well approximated by systematically varying the weights ω_i assigned to each of the criteria $F_i(\theta)$ over all possible values. Note that the above formulation also accounts for the fact that, in general, the various criteria $F_i(\theta)$ may not be directly commensurable and may, in fact, vary over very different orders of magnitude; the multiplier τ_i is used to compensate for this difference by transforming the criteria onto a commensurable scale.^{85,11,99}

4.2. Step-wise Multi-criteria Calibration Approach

The step-wise multi-criteria approach aims to reduce the dimension of the optimization problem by breaking it into several sub-problems that are handled in a step-by-step manner; each step considers different aspects of hydrologic response by transforming the flows,^{100,101} by focusing on different time periods,^{102,23,103,104} and/or different time scales.^{105,9,106} For example, the Multi-step Automatic Calibration Scheme^{100,101} (MACS) uses the SCE-UA global search algorithm and a step-by-step process to emulate some aspects of the progression of steps followed by an expert hydrologist during manual calibration. The step-

by-step process is as follows: (1) the lower zone parameters are calibrated to match a logarithmic transformation of the flows, thereby placing a strong emphasis on reproducing the properties of the low-flow portions of the hydrograph; (2) the lower zone parameters are subsequently fixed and the remaining parameters are optimized with the RMSE objective function to provide a stronger emphasis on simulating high-flow events; and (3) finally, a refinement of the lower zone parameters is performed using the log-transformed flows while keeping the upper zone parameters fixed at values derived during step two. The method has been tested for a wide variety of hydro-climatic regimes in the United States and has been shown to produce model simulations that are comparable to traditional manual calibration techniques.^{100,101}

Other studies that utilize step-wise procedures in the calibration of hydrologic model parameters also exist. Brazil¹⁰² estimated parameters of the SAC-SMA model using a combination of an interactive analysis of observed streamflow time series, an automated global search algorithm and a local search algorithm for fine-tuning. Similarly, Bingeman et al.²⁷ utilized a hybrid approach combining manual and automatic approaches in a step-wise strategy to calibrate and evaluate the spatially distributed WATFLOOD watershed model. Boyle et al.²³ utilized a simultaneous multi-criteria approach to generate a set of Pareto optimal solutions, showing performance trade-off in fitting different segments of the hydrograph, and then selected a parameter set from within the Pareto region that best satisfies two long-time-scale statistical measures of fit (mean and variance). In a more sophisticated approach, Pokhrel et al.¹⁰⁷ constructed the Pareto optimal solutions using traditional objective functions (i.e. RMSE and log-RMSE) and then employed 'signature measures' of model performance⁸⁰ to select a parameter set from within the solutions in close proximity to Pareto region. In an effort to improve the performance of the United States Geological Survey (USGS) Precipitation Runoff Modeling System, Leavesley et al.⁹ performed a step-wise approach which began by calibrating the parameters affecting water balance, followed by the parameters related to hydrograph timing and soils and vegetation, respectively. Harlin¹⁰⁸ and Zhang and Lindstrom¹⁰³ introduced process-oriented step-wise calibration strategies for the HBV hydrological model. Their procedures partitioned the flow time series into several periods, where specific hydrologic processes dominate, and linked these periods to specific parameter(s) during calibration. Turcotte et al.¹⁰⁵ developed a step-wise calibration strategy

for the HYDROTEL distributed model. In the procedure, they first calibrated the model parameters sensitive to objectives related to long timescales and then calibrated those parameters sensitive to short timescales. Their calibration steps included parameters related to large-scale water balances, evapotranspiration, infiltration capacity and routing, respectively. Shamir et al.¹⁰⁶ introduced a step-wise parameter estimation approach based on a set of streamflow descriptors that emphasize the dynamics of the streamflow record at different timescales. Fenicia et al.⁸⁴ showed how simultaneous and step-wise multi-criteria optimization strategies can help in understanding model deficiencies and hence guide model development. Wagener et al.¹⁰⁹ (DYNIA) and Choi and Beven¹⁰⁴ developed process-based approaches based on Generalized Sensitivity Analysis^{110,111} (GSA) to incorporate the time-varying nature of the hydrologic responses into model/parameter identification. Specifically, the DYNIA approach adopts the GSA procedure within a sequential Monte Carlo framework to locate periods of high identifiability for individual parameters and to diagnose possible failures of model structures. Vrugt et al.¹¹² presented a similar idea to assess the information content of individual observations with respect to the various model parameters. Their method, called PIMLI, merges the strength of Bayesian Inference with Random Walk Metropolis sampling to resample the parameter space each time the most informative measurement is added to the objective function. Results showed that only a very small number of streamflow measurements was actually needed for reliable calibration of a parsimonious five-parameter rainfall-runoff model. Specifically, about 95% of the discharge observations turned out to contain redundant information. Similar conclusions were drawn for case studies involving subsurface flow and transport models. The development of PIMLI and DYNIA has been inspired by Sequential Monte Carlo (SMC) schemes that provide a generalized treatment of time-varying parameters and states, and are increasingly being used for posterior tracking within the statistical and computational science literature.

4.3. Multi-criteria Constraining Approach

The basic concept of the constraining approach differs from automated calibration strategies, described above, in the sense that it considers 'consistency' of model structure/parameters as the ultimate goal, while the latter aims at optimality. In other words, the constraining approach seeks parameter estimates that are consistent with some minimal thresholds of performance on several criteria. Rooted in the concept of the Generalized Sensitivity Analysis (GSA),^{110,111} models/parameters are separated into behavioral/non-behavioral groups by comparing their performance with a subjectively selected threshold behavior. The models/parameters that present better performance than the selected acceptable representation of the system. The remaining models/parameters are rejected as non-behavioral. The Monte Carlo simulation-based GSA approach is at the core of many model identification and uncertainty estimation techniques (e.g. GLUE methodology of Beven and Binley¹¹³ and DYNIA methodology of Wagener et al.¹⁰⁹).

The main advantages of the constraining approach are its ease of implementation and use and its flexibility in incorporating additional performance criteria into the parameter estimation problem. Similar to the step-wise multi-criteria calibration approach (Sec. 4.2) several performance measures can be formulated and behavioral models/parameter sets can be selected considering a single or a group of performance measure(s).⁸⁰ The selected set (behavioral set) can be further analyzed using the concept of Pareto optimality to identify a

We illustrate the constraining approach with an overview of the step-wise semi-automated methodology, recently proposed by Yilmaz et al.⁸⁰ In Yilmaz et al.⁸⁰, parameters of the spatially distributed SAC-SMA model were constrained based on a set of hydrologically meaningful 'signature measures' target and extract hydrologically relevant (contextual) *information* from the outlet streamflow observations in ways that correspond to the following behavioral functions of a watershed system: (1) Maintain overall water balance, (2) Vertically redistribute excess rainfall between fast and slow runoff components, and (3) Redistribute the runoff in time (influencing hydrograph timing and shape). The selected signature measures were defined in terms of percent bias between the following hydrologically meaningful indices (calculated using simulated and observed outlet

streamflow): the percent bias in runoff ratio (%BiasRR), various properties of the flow duration curve (high-flow volume, %BiasFHV; slope of the mid-segment, %BiasFMS; low-flow volume, %BiasFLV; and median flow, %BiasFMM) and a simple index of watershed lag-time (%BiasTLag). One advantage of using signature-based measures in a constraining approach is that they can take on both positive and negative values, thereby indicating the direction of improvement. The procedure starts with establishing relationships between signature measures and parameters of the SAC-SMA model (via a procedure rooted in random sampling). These relationships are then used to constrain the ranges of parameters towards regions of signature measure improvement. In the second step, additional random samples are generated from the constrained parameter ranges and behavioral parameter sets are selected by establishing thresholds on the signature measures. There are various ways in which these thresholds can be defined. In Yilmaz et al.⁸⁰ the performance of the SAC-SMA model with *a priori* parameter values was used as benchmark. The procedure for selection of the behavioral parameter sets is depicted in Figure 3; the names of the signature measures are listed on the x-axis, and the y-axis shows their corresponding values. Each line along the x-axis represents a parameter set. The gray-dashed line represents the performance of the model using *a priori* parameters and the shaded region envelops the signature measure improvement region (i.e. behavioral region), defined as ± 1 times the *a priori* model performance. The solid line-triangle parameter combination falls entirely within the gray region, and therefore represents a behavioral parameter set.

5. Automated Calibration of Spatially Distributed Watershed Models

³LVWULEXWHG' □ ZDWHUVKHG □ PRCHG □ HGV □ The potential ability to simulate the spatial distribution of hydrologic processes in the watershed of interest, as well as to provide estimates of discharge volume along the entire length of the channel network so that, for instance, the dynamic evolution of flood inundation regions can be estimated.

Although their potential benefits are clear, the spatial complexity of these processes is perceived to be an obstacle to the proper identification of distributed model components and parameters, translating into

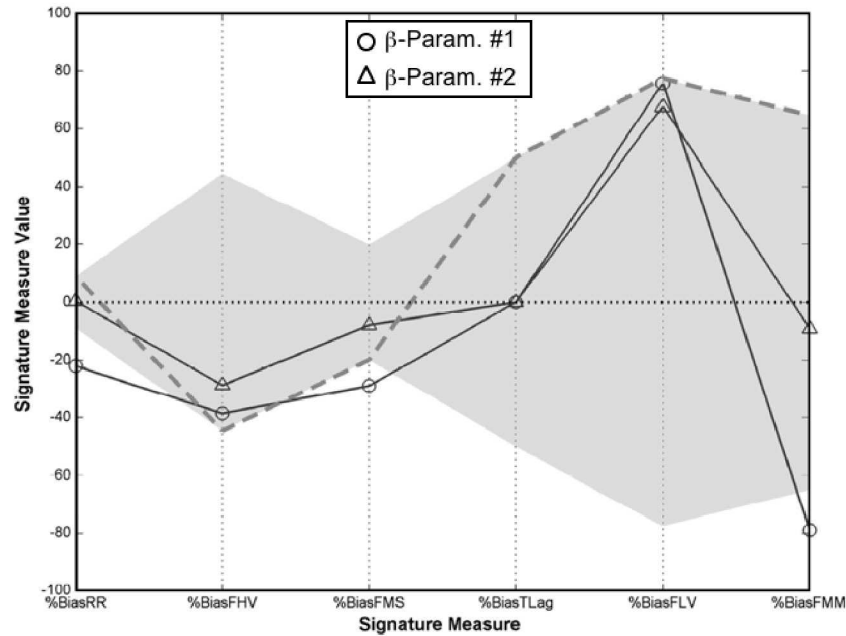


Fig. 3. Illustration of the parameter constraining approach proposed by Yilmaz et al.⁸⁰ Thresholds are selected based on the signature measure values for the model using *a priori* parameter sets (gray-dashed line). Behavioral region is defined as ± 1 times the *a priori* model performance (shaded region). Solid-line-triangle and solid-line-circle represent a behavioral and a non-behavioral parameter set, respectively. (After Yilmaz et al.⁸⁰)

significant predictive uncertainty in the model results.¹¹⁸ Therefore, controversy still persists regarding how such models should be implemented.^{119,120,121} For example, Phase One of the recent Distributed Model Intercomparison Project¹²² (DMIP-1) FRQFOXGHV □ WKDW □³UHOLDEOH □ DQG □ objective procedures for parameter estimation, data assimilation, and HUURU □ FRUHFVLRQ' □ VWLOO □ QHHG □ WR □ EH □ GHYHORSHG □

An important strength of distributed watershed models is that their parameters (and structure) can, in principle, be inferred from data about watershed physical characteristics via a local *a priori* parameter estimation approach (See Sec. 2 for details). However, in general, parameters estimated using a local *a priori* approach continue to require some degree of calibration (adjustment of the model parameter fields) to

ensure that the input-state-output behavior of such models remains consistent with the available data.^{10,11,12,123}

In calibration of watershed models having spatially distributed parameter fields, the number of parameter values to be estimated can be quite large, resulting in an ill-posed optimization problem. Note that the watershed outlet streamflow time series, the most widely used hydrologic variable in calibration of watershed models, represents the *integrated response* of the catchment, and therefore the effects of sub-catchment (local) scale variability in hydraulic properties has been averaged out to some extent. Therefore, it will typically only contain information about the watershed scale properties of the parameter fields (the mean values and the broad scale spatial patterns), rather than the smaller scale patterns of variation. As a consequence, the number of unknowns (parameters) will be larger than the one that can be identified based on the information content of the streamflow time series data, and hence the estimated parameters will most likely be unrealistic.

Part of the solution to the stabilization of ill-posed problems of this kind is to recognize that the spatially distributed elements of the model parameter fields are not, in fact, independent entities that can take on arbitrary values in the parameter space. Instead, their values are somehow related to the spatial distributions of various hydrologically relevant watershed characteristics, including, for example, geology, soil type, vegetation, topography, etc. (e.g. Grayson and Blöschl¹²⁴). Recognition of these dependencies can facilitate implementation of regularization relationships that help to constrain the dimensionality of the parameter estimation problem (e.g. Pokhrel et al.¹²⁵; Doherty and Skahill¹²⁶).

Regularization is a mathematical technique that allows stabilization of the otherwise ill-posed (over-parameterized) parameter estimation problems through introduction of additional information about the parameters.^{127,128,129,130} Two kinds of regularization techniques are in common use and these help to: 1) improve conditioning of the optimization problem through use of additional information and/or 2) reduce the dimensionality of the parameter search space.^{126,131,125} An example of the former technique is called Tikhonov Regularization (TR),¹²⁷ which employs a penalty function approach¹³² wherein the original parameter dimension is retained, while the shape of the global objective function to be optimized, $F_G(\theta, \mu)$, is modified by a regularization objective function, $F_{par}(\theta)$:

$$F_G(\theta, \mu) = F(\theta) + \mu \cdot F_{par}(\theta) \quad (8)$$

where $F_{par}(\theta)$ represents a penalty applied on the solutions that deviate from satisfying the regularization constraints. The regularization (or penalty) parameter, μ , is the weight assigned to the regularization constraints; greater values assigned to μ will result in improved parameter reasonableness and stability, although possibly at the cost of a poorer fit between model outputs and measurements (e.g. Yilmaz¹³³). Sometimes, an appropriate value for μ can be assigned through knowledge of the underlying uncertainty associated with the various components of $F_{par}(\theta)$. Such information, however, is often unavailable and thus alternative methods have been developed (e.g. Doherty¹³⁰; Mertens et al.⁹⁸; Yilmaz¹³³). An example of a regularization technique with the potential to significantly reduce the dimensionality of the estimable parameter space is the Truncated Singular Value Decomposition (TSVD). This technique only focuses on dominant directions and excludes search directions that are associated with negligible eigenvalues with little or no function sensitivity.^{134,135,136} Either separately or together, Tikhonov regularization and TVSD have been used to address high-dimensional inverse modeling problems in hydrology.^{137,138,139,126,140} Vrugt et al.¹⁴¹ have recently completed the development of a global optimization algorithm with self-adaptive subspace learning that uses TVSD to continuously update the dominating eigenvectors estimated from an evolving population of solutions. Initial results demonstrate significant efficiency improvements in solving high-dimensional parameter estimation problems (more than 200 parameters). A beta version of this algorithm has been developed in MATLAB and can be obtained from the second author upon request.

A simple and commonly used regularization technique to reduce the dimensionality of the spatially distributed parameters of watershed models seeks to characterize and preserve the pattern of relative spatial variation provided by the (local) *a priori* parameter fields (or maps). In its simplest form, one scalar multiplier per *a priori* parameter field is used to vary only the mean level of the parameter field, while constraining the values to remain within their pre-defined physical ranges (see Bandaragoda et al.¹⁴² and White et al.¹⁴³, among others). More sophisticated approaches utilize both additive and multiplicative terms,^{144,9} non-linear transformations via a single parameter,⁸⁰ or more complex approaches¹²⁵ that are combinations of the above. These

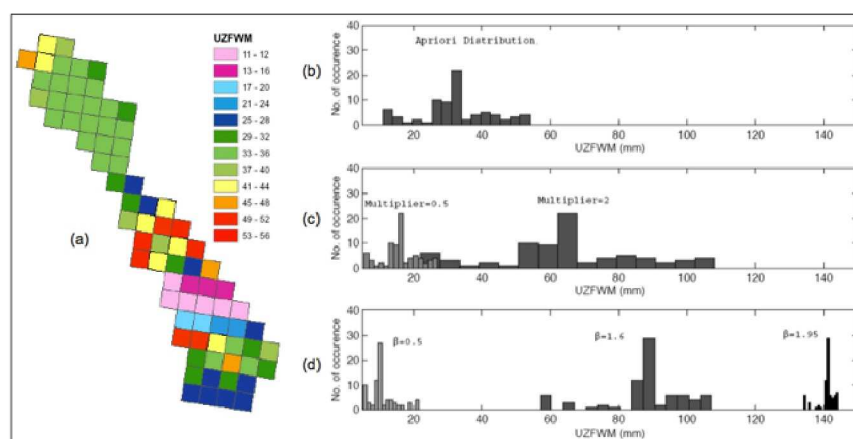


Fig. 4. The effect of using transformation functions on the distribution of UZFWM parameter values. (a) Map showing the *a priori* field of the UZFWM parameter (size of the upper layer free water tank) within the distributed version of the SAC-SMA model setup for the Blue River Basin, near Blue, Oklahoma. Histograms showing the distribution of parameter values within UZFWM parameter field: (b) *a priori* specified values, (c) parameter values transformed using scalar multipliers, (d) parameter values transformed using the transformation function proposed by Yilmaz et al.⁸⁰

space globally, but when the search progresses the method settles down with emphasis on local parameter refinements. The transition from global to local search is achieved by dynamically and probabilistically reducing the number of search dimensions (i.e. the number of model parameter values being perturbed) in the neighborhood of the current best estimate. The DDS algorithm has many elements in common with evolutionary strategies, and has only one algorithmic parameter. This is the neighborhood perturbation size, which is calculated from a multi-normal distribution with standard deviation equal to some fraction of the initial range of each decision variable. The user specified inputs are the initial solution and the maximum number of model evaluations. The DDS algorithm is very simple to implement, and has been used in calibration of a rainfall runoff model¹⁴⁵, and a land surface scheme.¹⁴⁸ The DDS algorithm can also be used to assess parameter uncertainty by using multiple trails with different starting points and analyze their joint sampling paths.¹⁴⁹ Such an approach violates first-order Markovian properties and can therefore only provide a rough estimate of parameter uncertainty (see next Section). Recent work has compared the

performance of DDS with SCE-UA for a range of studies.^{145,150,151} Behrangi et al.¹⁵⁰ has introduced an alternative SCE-UA algorithm that is at least as efficient as DDS when a limited budget of function evaluations is considered.

Constrained Optimization using Response Surfaces^{146,147} (CORS) is a response surface approximation method. A response surface model is a multivariate approximation of the black box objective function and used as surrogate for optimization in situations where function evaluations are computationally expensive. In the CORS method the next point (parameter combination) is chosen to be the one that minimizes the current response surface model subject to various constraints including that the new point must be of some minimum distance from previously evaluated points. This distance is sequentially reduced from a high value (global search) at the start of the search to a low value (local search) at termination. This method has shown to converge nicely to the global optimum for a set of continuous functions.

6. Treatment of Parameter Uncertainty

One major weakness of the automated calibration methods discussed earlier in this chapter is their underlying treatment of the uncertainty as being primarily attributable to the model parameters, without explicit treatment of input, output and model structural uncertainties. It is well known, however, that the uncertainties in the watershed modeling procedure stem not only from uncertainties in the parameter estimates, but also from measurement errors associated with the system inputs and outputs, and from model structural errors arising from the aggregation of spatially distributed real-world processes into a relatively simple mathematical model. Not properly accounting for these errors results in residuals that exhibit considerable variation in bias (non-stationarity), variance (heteroscedasticity) and correlation structures under different hydrologic conditions and, hence, undermines our efforts to derive meaningful parameter estimates that properly mimic the target hydrologic processes.

One response to the treatment of uncertainty in the parameter combination and adopt a Bayesian viewpoint, which allows the identification of a distribution specifying the uncertainty regarding the values of the model parameters. Bayesian statistics have recently found

increasing use in the field of hydrology for statistical inference of parameters, state variables and model output prediction.^{152,153,154,155,156,157}

For a recent review, see Liu and Gupta¹⁵⁷. The Bayesian paradigm provides a simple way to combine multiple probability distributions using Bayes theorem. In a hydrologic context, this method is suited to systematically address and quantify the various error sources within a single cohesive, integrated and hierarchical method.

To successfully implement the Bayesian paradigm, sampling methods that can efficiently summarize the posterior probability density function (pdf) are needed. This distribution combines the data likelihood with a prior distribution using Bayes theorem, and contains all the desired information to make statistically sound inferences about the uncertainty of the individual components in the model. Unfortunately, for most practical hydrologic problems, this posterior distribution cannot be obtained by analytical means or by analytical approximation. For this reason, researchers commonly resort to iterative approximation methods, such as Markov Chain Monte Carlo (MCMC) sampling, to generate a sample from the posterior pdf.

6.1. Random Walk Metropolis (RWM) algorithm

The basis of the MCMC method is a Markov chain that generates a random walk through the search space with stable frequencies stemming from a fixed probability distribution. To visit configurations with a stable frequency, an MCMC algorithm generates trial moves from the current ('old') position of the Markov chain

$$\alpha(\theta_{t-1}, \mathcal{G}) = \begin{cases} \min\left(\frac{\pi(\mathcal{G})}{\pi(\theta_{t-1})}, 1\right) & \text{if } \pi(\theta_{t-1}) > 0 \\ 1 & \text{if } \pi(\theta_{t-1}) = 0 \end{cases} \quad (9)$$

where

hydrology is the Shuffled Complex Evolution Metropolis (SCEM-UA) global optimization algorithm.¹⁵⁵ This method is a modified version of the original SCE-UA global optimization algorithm,⁶² and runs multiple chains in parallel to provide a robust exploration of the search space. These chains communicate with each other through an external population of points, which are used to continuously update the size and shape of the proposal distribution in each chain. The MCMC evolution is repeated until the \hat{R} statistic of Gelman and Rubin¹⁵⁹ indicates convergence to a stationary posterior distribution. This statistic compares the between- and within-variance of the different parallel chains.

Numerous studies have demonstrated the usefulness of the SCEM-UA algorithm for estimating (nonlinear) parameter uncertainty. However, the method does not maintain detailed balance at every single step in the chain, casting doubt on whether the algorithm will appropriately sample the underlying pdf. Although various benchmark studies have reported very good sampling efficiencies and convergence properties of the SCEM-UA algorithm, violating detailed balance is a reason for at least some researchers and practitioners not to use this method for posterior inference. An adaptive MCMC algorithm that is efficient in hydrologic applications, and maintains detailed balance and ergodicity therefore remains desirable.

6.2. Differential Evolution Adaptive Metropolis (DREAM)

Vrugt et al.¹⁶⁰ recently introduced the Differential Evolution Adaptive Metropolis (DREAM) algorithm. This algorithm uses differential evolution as genetic algorithm for population evolution, with a Metropolis selection rule to decide whether candidate points should replace their respective parents or not. DREAM is a follow up to the DEMC (Differential Evolution Markov Chain) method of ter Braak¹⁶¹, but contains several extensions to increase search efficiency and acceptance rate for complex and multimodal response surfaces with numerous local optimal solutions. Such surfaces are frequently encountered in hydrologic modeling. The method is presented below.

1. Draw an initial population

8. Compute the Gelman-Rubin \hat{R}_j convergence diagnostic¹⁵⁹ for each dimension $j = 1, \dots, d$ using the last 50% of the samples in each chain.

9. If $\hat{R}_j < 1.2$ for all j , stop, otherwise go to CHAIN EVOLUTION.

At every step, the points in

6.3. Sequential Data Assimilation

In the past few years, ensemble-forecasting techniques based on Sequential Data Assimilation (SDA) methods have become increasingly popular due to their potential ability to explicitly handle the various sources of uncertainty in environmental modeling. Techniques based on the Ensemble Kalman Filter¹⁶³ (EnKF) have been suggested as having the power and flexibility required for data assimilation using nonlinear models. In particular, Vrugt et al.¹⁶⁴ recently presented the Simultaneous Optimization and Data Assimilation (SODA) method, which uses the EnKF to recursively update model states while estimating time-invariant values for the model parameters using the SCEM-UA¹⁵⁵ optimization algorithm. A novel feature of SODA is its explicit treatment of errors due to parameter uncertainty, uncertainty in the initialization and propagation of state variables, model structural error and output measurement errors. The development below closely follows that of Vrugt et al.¹⁶⁴

To help facilitate the description of the classical Kalman Filter (KF), we start by writing the model dynamics as a stochastic equation. In keeping with Figure 1, consider a model Φ in which the discrete time evolution of the state vector ψ is described with:

$$\psi_{i+1} = \Phi(\psi_i, X_i, \theta) + q_i \quad (14)$$

where X represents the observed forcing (e.g. boundary conditions), θ is the vector of parameter values, i denotes time and q_i is a dynamical noise term representing errors in the conceptual model formulation. This stochastic forcing term flattens the probability density function of the states during the integration. We assume that the model predictions are related to its internal state according to:

$$\tilde{s}_i = H(\psi_i) + \varepsilon_i \quad \varepsilon_i \sim N(0, \sigma_i^0) \quad (15)$$

where $H(\cdot)$ is the measurement operator, which maps the state space into measurement or model output space, and also has a random additive error ε_i , called the measurement error. Note that σ_i^0 denotes the error deviation of the measurements, and ψ^* represents the true model states. At each measurement time, when an output observation becomes available, the output forecast error z_i is computed:

$$z_i = s_i - H(\psi_i^f) \quad (16)$$

and the forecasted states, ψ_i^f , are updated using the standard KF analysis equation:

$$\psi_i^u = \psi_i^f + K_i(s_i - H(\psi_i^f)) \quad (17)$$

where ψ_i^u is the updated or analyzed state, and K_i denotes the Kalman gain. The size of the gain directly depends on the size of the measurement and model error.

The analyzed state then recursively feeds the next state propagation step into the model:

$$\psi_{i+1}^f = \Phi(\psi_i^u, X_i, \theta) \quad (18)$$

The virtue of the KF method is that it offers a very general framework for segregating and quantifying the effects of input, output, and model structural error in watershed modeling. Specifically, uncertainty in the model formulation and observational data are specified through the stochastic forcing terms q and ε , whereas errors in the input data are quantified by stochastically perturbing the elements of X .

The SODA method is an extension of traditional techniques in that it uses the Ensemble Kalman Filter (EnKF) to solve for equations (14) \pm (18). The EnKF uses a Monte Carlo (MC) method to generate an ensemble of model trajectories from which the time evolution of the probability density of the model states and related error covariances are estimated.¹⁶⁵ The EnKF avoids many of the problems associated with the traditional extended Kalman Filter (EKF) method. For example, there is no closure problem as is introduced in the EKF by neglecting contributions from higher-order statistical moments in the error covariance evolution. Moreover, the conceptual simplicity, relative ease of implementation and computational efficiency of the EnKF make the method an attractive option for data assimilation in the meteorologic, oceanographic and hydrologic sciences.

In summary, the EnKF propagates an ensemble of state vector trajectories in parallel, such that each trajectory represents one realization of generated model replicates. When an output measurement is available, each forecasted ensemble state vector ψ_i^f is updated by means of a linear updating rule in a manner analogous to the Kalman Filter. A detailed description of the EnKF method, including the algorithmic details, is found in Evensen¹⁶³ and so will not be repeated here.

7. Parallel Computing

The traditional implementation and application of the local and global optimization methods, discussed herein, involves sequential execution of the algorithm using the computational power of a single Central Processing Unit (CPU). Such an implementation works acceptably well for relatively simple optimization problems, and those optimization problems with models that do not require much computational time to execute. However, for high-dimensional optimization problems involving complex spatially distributed models, such as the ones frequently used in the field of environmental science, this sequential implementation needs to be revisited. Most computational time required for calibrating parameters in watershed models is spent running the model code and generating the desired output. Thus, there should be large computational efficiency gains from parallelizing the algorithm so that independent model simulations are run on different nodes in a distributed computer system. General-purpose parallel versions of SCE-UA, MOCOM, MOSCEM-UA, AMALGAM and DREAM have been developed in Octave and can be obtained from the second author upon request.

8. Summary and Conclusions

This chapter reviewed the current state-of-the-art of model calibration in watershed hydrology. The current status in watershed modeling is that, regardless of their type (i.e. conceptual/physically-based) and structural complexity (e.g. spatially lumped/distributed), watershed models still need to be calibrated to improve the reliability of their simulation/prediction performance. In calibration, model parameters are systematically adjusted in such a way that the behavior of the model approximates, as closely and consistently as possible, the historical observed response of the watershed system under study. In parallel with advances in computing resources and measurement techniques, hydrologic models are becoming increasingly complex, not only representing the spatial heterogeneity of a watershed but also simulating various internal states and fluxes. This increase in complexity is typically accompanied by a decrease in identifiability (or justification) of the components and parameters in these models due to problems with direct observation and noisy data. This leads to highly uncertain model predictions. There is, therefore, a pressing need for better approaches to model calibration.

We discussed manual and automatic parameter estimation techniques for calibration of lumped and spatially distributed watershed models. Historically, manual calibration has been the common practice for model calibration. Its main strength is the use of expert knowledge to match the perceived hydrologic processes in the watershed with their equivalents in the model, and thus they help to prescribe consistent values for the parameters. However, manual calibration is time and labor intensive, involving many subjective decisions. The era of digital computing and advances in hardware power and speed has led to the development of automated procedures that are more objective to implement and use, and more efficient in systematically searching the parameter space of the model parameters. We have discussed the historical development of single criterion methods and provided details of the widely used SCE-UA global optimization algorithm. The major limitation of automated algorithms is that they rely on a single regression-based aggregate measure of model performance, which often leads to an ill-posed parameter estimation problem. This is due to a loss of information when projecting from the high-dimensional data space down to the single dimension of the residual-based summary statistic. For highly complex models with many interacting parameters, this approach is often poor at discriminating between the effects of individual parameters on the simulated model output. This, therefore, unnecessarily introduces non-uniqueness and equifinality in parameter estimation.

More sophisticated automated (or semi-automated) methods for parameter adjustment have been developed within the framework of multi-criteria theory, thereby more closely matching the number of unknowns (the parameters) with the number of pieces of information (the measures). A major advantage of the multi-criteria approach is that various aspects of the manual calibration strategy can be absorbed into the calibration process, thus strengthening the physical basis of the identified parameters. Although multi-criteria methods are widely used in the calibration of watershed models and powerful algorithms have been developed (among these we discussed MOCOM-UA, MACS, DYNIA, PIMLI, AMALGAM and MOSCEM) in the past decade or so, how to properly incorporate expert knowledge into automatic parameter estimation still remains difficult, and an active area of research. To better realize the strengths of multi-criteria optimization requires a shift from statistical regression-based performance measures towards more powerful and hydrologically relevant (contextual) measures of

information.^{77,80} One possible way to develop these signature-based spatio-temporal data and by relating these to their causal mechanisms. Signature measures that quantitatively summarize these signature patterns can be used to estimate plausible model parameters and further point towards possible causes of model failure and guide model improvement strategies. More research is warranted on how to 1) properly construct different and complementary signature measures of model performance that are diagnostic,^{77,80} while being representative of the watershed response behaviors that are deemed important to reproduce with the model; and 2) integrate these signature measures into automated calibration techniques while keeping the resulting multi-criteria optimization problem computationally tractable.

We have also discussed the calibration of spatially distributed watershed models. Calibration of such models is complicated by their large number of parameters. Part of the solution to the stabilization of ill-posed problems of this kind is to recognize that the spatially distributed elements of the model parameter fields are not, in fact, independent entities that can take on arbitrary values in the parameter space. Instead, their values are somehow related to the spatial distributions of various hydrologically relevant watershed characteristics including, for example, geology, soil type, vegetation, topography, etc. Recognition of these dependencies can facilitate implementation of regularization relationships that help to constrain the dimensionality of the parameter estimation problem. We would like to emphasize that although application of regularization techniques in calibration of complex watershed models is in its infancy (e.g. Pokhrel et al.¹²⁵), considerable progress has already been reported in adjacent fields, such as groundwater and subsurface hydrology. Methods and approaches developed therein might be of great use and inspiration to solve parameter estimation problems in watershed hydrology. Particularly appealing are approaches that attempt to 1) improve conditioning of the optimization problem through use of additional information (penalty functions; e.g. Tikonov and Arsenin¹²⁷); 2) reduce the dimensionality of the optimization problem using concepts of sensitivity and principle component analysis (e.g. Tonkin and Doherty¹³⁹), as discussed briefly herein; and 3) representer-based calibration methods.¹⁶⁶ Further, the availability of parallel computing and powerful optimization algorithms that can efficiently handle a high number of parameters in complex

distributed hydrologic models is of particular importance for developing rigorous model calibration strategies for such models.⁷⁵

Finally, this chapter has discussed emerging methods for parameter uncertainty estimation. Uncertainty quantification is currently receiving a surge in attention in hydrology, as researchers try to better understand what is well and what is not well understood about the watersheds that are being studied and as decision makers push to better quantify accuracy and precision of model predictions. Various methodologies have been developed in the past decade to better treat uncertainty. We have highlighted the standard RWM for posterior exploration, and have discussed recent advances in MCMC sampling by combining multi-chain population evolution with the genetic algorithm Differential Evolution and subspace sampling. We have also summarized emerging state-space filtering methods that rely on the Ensemble Kalman Filter to specifically treat forcing (input), output, parameter and state error, within a recursive estimation framework.

Among emerging approaches for uncertainty estimation not discussed in this chapter include methods such as Bayesian Estimation of Structure (BEST),¹⁶⁷ Bayesian Model Averaging (BMA),^{168,169,170,171} and its maximum likelihood (ML) variant, MLBMA^{172,173,174}. Such methods seem particularly appealing because they help to account for model structural uncertainty. BMA and MLBMA, in particular, are relatively easy to implement, accounting jointly for conceptual model and parameter uncertainty by considering a set of alternative models that may be very different from each other while being based on a common set of data.

Acknowledgments

The first author is supported by the Earth and Space Science (ESS) program (Stephen Ambrose) and PMM (Ramesh Kakar) of NASA Headquarters. The second author is supported by a J. Robert Oppenheimer Fellowship from the LANL postdoctoral program. Partial support for the first and third authors was provided by SAHRA, the Science and Technology Center for Sustainability of semi-Arid Hydrology and Riparian Areas under NSF-STC grant EAR-9876800 and by the US National Weather Service Office of Hydrology under grant NA04NWS462001.

References

1. R. J. C. Burnash, R. L. Ferral and R. A. McGuire, in *A Generalized Streamflow Simulation System-Conceptual Modeling for Digital Computers*. Technical Report of the Joint Federal-State River Forecast Center, (Department of Water Resources, State of California and National Weather Service, 1973), pp. 204.
2. R. J. C. Burnash, in *Computer Models of Watershed Hydrology*, Ed. V. P. Singh (Water Resources Publications, Colorado, 1995), p. 311.
3. S. Bergstrom, in *Computer Models of Watershed Hydrology*, Ed. V. P. Singh (Water Resources Publications, Colorado, 1995), p. 443.
4. J. C. Refsgaard and B. Storm, in *Computer Models of Watershed Hydrology*, Ed. V. P. Singh (Water Resources Publications, Colorado, 1995), p. 809.
5. D. A. Woolhiser, R. E. Smith and D. C. Goodrich, USDA, Agricultural Research Service Report No: ARS-77, pp. 130 (1990).
6. J. C. Refsgaard, *J. Hydrol.* **198**, 69 (1997)
7. V. Koren, M. Smith, D. Wang and Z. Zhang, in *15th Conference on Hydrology*, AMS Conference Proceeding (AMS, Long Beach, California, 2000), p. 103.
8. V. Koren, M. Smith and Q. Duan, in *Calibration of Watershed Models*, Eds. Q. Duan, S. Sorooshian, H. Gupta, H. Rosseau, R. Turcotte (Water Science and Applications, American Geophysical Union, 2003), p. 239.
9. G. H. Leavesley, L. E. Hay, R. J. Viger and S. L. Markstrom, in *Calibration of Watershed Models*, Eds. Q. Duan, S. Sorooshian, H. Gupta, H. Rosseau, R. Turcotte (Water Science and Applications, American Geophysical Union, 2003), p. 255.
10. H. V. Gupta, S. Sorooshian and P. O. Yapo, *Water Resour. Res.*, **34(4)**, 751 (1998).
11. H. Madsen, *Adv. Water Resour.* **26**, 205 (2003).
12. S. Reed, V. Koren, M. Smith, Z. Zhang, F. Moreda, D.-J. Seo, and DMIP participants, *J. Hydrol.* **298**, 27 (2004).
13. F. A. Abdulla and D. P. Lettenmaier, *J. Hydrol.* **197**, 230 (1997).
14. C. E. M. Sefton and S. M. Howarth, *J. Hydrol.* **211**, 1 (1998).
15. W. Fernandez, R. M. Vogel and S. Sankarasubramanian, *Hydrol. Sci. J.* **45(5)**, 689 (2000).
16. T. Wagener, H. S. Wheater and H. V. Gupta, (Imperial College Press, London, UK, 2004), p. 300.
17. T. Wagener and H. S. Wheater, *J. Hydrol.* **320**, 132 (2006).

18. N. McIntyre, H. Lee, H. S. Wheater, A. Young and T. Wagener, *Water Resour. Res.* **41**, W12434 doi:10.1029/2005WR004289 (2005).
19. D. Kavetski, S. W. Franks, G. Kuczera, in *Calibration of Watershed Models*, Eds. Q. Duan, S. Sorooshian, H. Gupta, H. Rosseau, R. Turcotte (Water Science and Applications, American Geophysical Union, 2003), p. 49.
20. K. J. Beven, *Water Sci. Tech.* **52(6)**, 167 (2005).
21. J. Göttinger and A. Bárdossy, *J. Hydrol.* **(333)**, 374 (2007).
22. M. Yadav, T. Wagener and H. Gupta, *Adv. Water Resour.* **30**, 1756 (2007).
23. D. P. Boyle, H. V. Gupta and S. Sorooshian, *Water Resour. Res.* **36(12)**, 3663 (2000).
24. Hydrologic Research Center, (Video Series produced for the National Weather Service, Hydrologic Research Center, 1999).
25. M. B. Smith, D. P. Laurine, V. Koren, S. Reed and Z. Zhang, in *Calibration of Watershed Models*, Eds. Q. Duan, S. Sorooshian, H. Gupta, H. Rosseau, R. Turcotte (Water Science and Applications, American Geophysical Union, 2003), p. 133.
26. M. Mroczkowski, G. P. Raper and G. Kuczera, *Water Resour. Res.* **33(10)**, 2325 (1997).
27. A. K. Bingeman, N. Kouwen and E. D. Soulis, *J. Hydrol. Eng.* **11(5)**, 451 (2006).
28. S. Sorooshian and H. V. Gupta, in *Computer Models of Watershed Hydrology*, Ed. V. P. Singh (Water Resources Publications, Colorado, 1995), p. 23.
29. H. V. Gupta, K. J. Beven and T. Wagener, in *Encyclopedia of Hydrological Sciences*, Ed. M. G. Anderson (John Wiley & Sons Ltd., Chichester, UK, 2006).
30. S. Sorooshian and J. A. Dracup, *Water Resour. Res.* **16(2)**, 430 (1980).
31. K. J. Beven, *Rainfall-Runoff Modeling ± The Primer*, (Wiley, Chichester, 2001), pp. 360.
32. Y. Bard, *Nonlinear Parameter Estimation*, (Academic Press, 1974), pp. 341.
33. V. K. Gupta and S. Sorooshian, *Water Resour. Res.* **21(4)**, 473 (1985).
34. J. D. Hendrickson, S. Sorooshian and L. Brazil, *Water Resour. Res.* **24(5)**, 691 (1988).
35. J. A. Nelder and R. Mead, *Comp. J.* **7**, 308 (1965).
36. R. Hooke and T. A. Jeeves, *J. Assoc. Comput. Mach.* **8(2)**, 212 (1961).
37. H. H. Rosenbrock, *Comp. J.* **3**, 175 (1960).

38. R. P. Ibbitt and T. O'Donnell, *J. Hydraul. Eng-ASCE*, **97(HY9)**, 1331 (1971).
39. P. R. Johnston and D. Pilgrim, *Water Resour. Res.* **12(3)**, 477 (1976).
40. G. Pickup, *Hydrol. Sci. B.* **12(2)**, 257 (1977).
41. R. J. Moore and R. T. Clarke, *Water Resour. Res.* **17(5)**, 1367 (1981).
42. S. Sorooshian and V. K. Gupta, *Water Resour. Res.* **19(1)**, 251 (1983).
43. V. K. Gupta and S. Sorooshian, *Water Resour. Res.* **19**, 269 (1983).
44. G. Kuczera, *J. Hydrol.* **103**, 229 (1988).
45. D. Kavetski, G. Kuczera and S. W. Franks, *Water Resour. Res.* **42**, W03407, doi:10.1029/2005WR004368 (2006).
46. D. Kavetski, G. Kuczera and S.W. Franks, *Water Resour. Res.* **42**, W03408, doi:10.1029/2005WR004376 (2006).
47. J. A. Vrugt, C. J. F. ter Braak, H. V. Gupta and B. A. Robinson, *Stoch. Environ. Res. Risk Assess.*, doi:10.1007/s00477-008-0274-y (2008).
48. J. A. Vrugt, C. J. F. ter Braak, M. P. Clark, J. M. Hyman and B. A. Robinson, *Water Resour. Res.* (In Press, 2008).
49. G. E. P. Box and D. R. Cox, *J. Roy. Stat. Soc.* **B26**, 211 (1964).
50. G. Kuczera, *Water Resour. Res.* **18(1)**, 146 (1982).
51. S. Sorooshian, V. K. Gupta and J. L. Fulton, *Water Resour. Res.* **19(1)**, 251 (1983).
52. V. K. Gupta and S. Sorooshian, *J. Hydrol.* **81**, 57 (1985).
53. P. O. Yapo, H. V. Gupta and S. Sorooshian, *J. Hydrol.* **181**, 23 (1996).
54. J. A. Vrugt, H. V. Gupta, S. Sorooshian, T. Wagener and W. Bouten, *J. Hydrol.* **325(1-4)**, 288, doi:10.1016/j.hydrol.2005.10.041 (2006).
55. D. C. Karnopp, *Automatica* **1**, 111 (1963).
56. S. F. Masri, G. A. Bekey and F. B. Safford, *Appl. Math. Comput.* **7**, 353 (1980).
57. L. E. Brazil and W. F. Krajewski, presented in *Conference on Engineering Hydrology* (Hydraulics Division American Society of Civil Engineers, Williamsburg, VA, August 3-7, 1987).
58. W. L. Price, *J. Optimiz. Theory App.* **40(3)**, 333 (1983).
59. O. Klepper, H. Scholten and J. P. G. van de Kamer, *J. Forecasting* **10**, 191 (1991).
60. S. Kirkpatrick, C. D. Gelatt Jr. and M. P. Vecchi, *Science* **220**, 671 (1983).
61. M. Thyer, G. Kuczera and B. C. Bates, *Water Resour. Res.* **35(3)**, 767 (1999).

62. Q. Duan, V. K. Gupta and S. Sorooshian, *Water Resour. Res.* **28**,1015 (1992).
63. Q. J. Wang, *Water Resour. Res.* **27(9)**, 2467 (1991).
64. J. A. Vrugt, M. T. van Wijk, J. W. Hopmans and J. Simunek, *Water Resour. Res.* **37(10)**, 2457 (2001).
65. Q. Duan, in *Calibration of Watershed Models*, Eds. Q. Duan, S. Sorooshian, H. Gupta, H. Rosseau, R. Turcotte (Water Science and Applications, American Geophysical Union, 2003), p. 89.
66. Q. Duan, V. K. Gupta and S. Sorooshian, *J. Optimiz. Theory App.* **76(3)**, 501 (1993).
67. S. Sorooshian, Q. Duan and V. K. Gupta, *Water Resour. Res.* **29(4)**, 1185 (1993).
68. J. H. Holland, *Adaptation in Natural and Artificial Systems* (University of Michigan Press, Ann Arbor, 1975).
69. T. Y. Gan and G. F. Biftu, *Water Resour. Res.* **32(7)**, 2161 (1996).
70. C. H. Luce and T. W. Cundy, *Water Resour. Res.* **30(4)**, 1057 (1994).
71. H. Tanakamaru, *Trans. JSIDRE* **178**, 103 (1995).
72. G. Kuczera, *Water Resour. Res.* **33(1)**, 177 (1997).
73. Q. Duan, S. Sorooshian and V. K. Gupta, *J. Hydrol.* **158**, 256 (1994).
74. D. Wolpert and W. G. MacReady, *IEEE Trans. Evol. Comp.* **1**, 67 (1999).
75. J. A. Vrugt and B. A. Robinson, *Proc. Natl. Acad. Sci. USA* **104**, 708, doi:10.1073/pnas.0610471104, 2007.
76. J. A. Vrugt, B. A. Robinson and J. M. Hyman, *IEEE Trans. Evolut. Comput.* **1-17**, 10.1109/TEVC.2008.924428 (In Press, 2008)
77. H. V. Gupta, T. Wagener and Y. Liu, *Hydrol. Process.*, doi:10.1002/hyp.6989 (2008).
78. T. Wagener and H. V. Gupta, *Stoch. Environ. Res. Risk Assess.* **19(6)**, 378, doi:10.1007/s00477-005-0006-5 (2005)
79. T. Meixner, H. V. Gupta, L. A. Bastidas and R. C. Bales, in *Calibration of Watershed Models*, Eds. Q. Duan, S. Sorooshian, H. Gupta, H. Rosseau, R. Turcotte (Water Science and Applications, American Geophysical Union, 2003), p. 213.
80. K. K. Yilmaz, H. V. Gupta and T. Wagener, *Water Resour. Res.* **44**, W09417, doi:10.1029/2007WR006716 (2008)
81. P. O. Yapo, H. V. Gupta and S. Sorooshian, *J. Hydrol.* **204**, 83 (1998).

82. J. A. Vrugt, H. V. Gupta, L. A. Bastidas, W. Bouten and S. Sorooshian, *Water Resour. Res.* **39(8)**, 1214, doi:10.1029/2002WR001746, (2003).
83. J. A. Vrugt, J. van Belle and W. Bouten, *J. Avian Biology* **38**, 432, doi:10.1111/j.2007.0908-8857.03909 (2007).
84. F. Fenicia, H. H. G. Savenije, P. Matgen and L. Pfister, *Water Resour. Res.* **43**, W03434, doi:10.1029/2006WR 005098 (2007).
85. H. Madsen, *J. Hydrol.* **235**, 276 (2000).
86. P. Reed, B. S. Minsker and D. E. Goldberg, *Water Resour. Res.* **39(7)**, 1096 (2002).
87. D. E. Goldberg, *Genetic Algorithms in Search, Optimization and Machine Learning*, (Addison-Wesley Publishing Company, Reading, MA, 1989), pp. 412.
88. D. Whitley, in *Proceedings of the Third International Conference on Genetic Algorithms*, Arlington, VA, (Morgan Kaufman Publishers, San Mateo, CA, 1989), p. 116.
89. Y. Liu, *Ph.D. Dissertation*, (Dept. of Hydrol. and Water Resour., Univ. of Arizona, Tucson, Arizona, USA, 2003), pp. 236.
90. H. V. Gupta, L. A. Bastidas, S. Sorooshian, W. J. Shuttleworth and Z. L. Yang, *J. Geophys. Res.* **104 (D16)**, 19491 (1999).
91. H. V. Gupta, L. Bastidas, J. A. Vrugt and S. Sorooshian, in *Calibration of Watershed Models*, Eds. Q. Duan, S. Sorooshian, H. Gupta, H. Rosseau, R. Turcotte (Water Science and Applications, American Geophysical Union, 2003), p. 125.
92. E. Zitzler, M. Laumanns, and L. Thiele, Department of Electrical Engineering, Swiss Federal Institute of Technology (ETH) Zurich, TIK-Report 103, (2001).
93. K. Deb, A. Pratap, S. Agarwal, and T. Meyarivan, *IEEE Trans. Evol. Comp.* **6**, 182 (2002).
94. E. Zitzler, K. Deb, and L. Thiele, *Evol. Comp.* **8**, 173 (2000).
95. E. Zitzler, and L. Thiele, *IEEE Trans. Evol. Comp.* **3**, 257 (1999).
96. J. A. Vrugt, P. H. Stauffer, T. Wöhling, B. A. Robinson and V. V. Vesselinov, *Vadose Zone J.* **7(2)**, 843, doi:10.2136/vzj2007.0078 (2008).
97. T. Wöhling and J. A. Vrugt, *Water Resour. Res.* (In Press, 2008).
98. J. Mertens, H. Madsen, L. Feyen, D. Jacques, J. Feyen, *J. Hydrol.* **294**, 251 (2004).
99. M. B. Butts, J. T. Payne, M. Kristensen and H. Madsen, *J. Hydrol.* **298**, 242 (2004).

100. T. S. Hogue, S. Sorooshian, H. V. Gupta, A. Holz and D. Braatz, *J. Hydrometeor.* **1**, 524 (2000).
101. T. S. Hogue, H. V. Gupta and S. Sorooshian, *J. Hydrol.* **320**, 202 (2006).
102. L. E. Brazil, *Multilevel Calibration Strategy for Complex Hydrologic Simulation Models* (Ph.D. dissertation, Colo. State Univ., Fort Collins, 1988), pp. 217.
103. X. Zhang and G. Lindström, *Hydrol. Process.* **11**, 1671 (1997).
104. T. H. Choi and K. Beven, *J. Hydrol.* **332**, 316 (2007).
105. R. Turcotte, A. N. Rousseau, J.-P. Fortin, J.-P. Villeneuve, in *Calibration of Watershed Models*, Eds. Q. Duan, S. Sorooshian, H. Gupta, H. Rosseau, R. Turcotte (Water Science and Applications, American Geophysical Union, 2003), p. 153.
106. E. Shamir, B. Imam, H. Gupta and S. Sorooshian, *Water Resour. Res.* **41(6)**, doi:10.1029/2004WR003409 (2005).
107. P. Pokhrel, K. K. Yilmaz and H. V. Gupta, *J. Hydrol.* (In Press, 2009).
108. J. Harlin, *Nordic Hydrol.* **22**, 15 (1991).
109. T. Wagener, N. McIntyre, M. J. Less, H. S. Wheater and H. V. Gupta, *Hydrol. Process.* **17(2)**, 455 (2003).
110. G. M. Hornberger and R. C. Spear, *J. Environ. Manage.* **12**, 7 (1981).
111. R. C. Spear and G. M. Hornberger, *Water Res.* **14**, 43 (1980).
112. J. A. Vrugt, W. Bouten, H. V. Gupta and S. Sorooshian, *Water Resour. Res.* **38(12)**, 1312, doi:10.1029/2001WR001118 (2002).
113. K. J. Beven and A. M. Binley, *Hydrol. Process.* **6**, 279 (1992).
114. V. Y. Ivanov, E. R. Vivoni, R. L. Bras and D. Entekhabi, *J. Hydrol.* **298(1±4)**, 80 (2004).
115. T. M. Carpenter and K. P. Georgakakos, *J. Hydrol.* **298**, 61 (2004).
116. R. E. Smith, D. C. Goodrich, D. A. Woolhiser and C. L. Unkrich, in *Computer Models of Watershed Hydrology*, Ed. V. P. Singh (Water Resources Publications, Colorado, 1995), p. 697.
117. V. Koren, S. Reed, M. Smith, Z. Zhang and D. J. Seo, *J. Hydrol.* **291**, 297 (2004).
118. K. J. Beven and J. E. Freer, *J. Hydrol.* **249**, 11 (2001).
119. K. J. Beven, *J. Hydrol.* **105**, 157 (1989).
120. K. J. Beven, *Hydrol. Process.* **16**, 189 (2002).

121. R. B. Grayson, I. D. Moore and T. A. McMahon, *Water Resour. Res.* **28(10)**, 2659 (1992).
122. M. B. Smith, D. J. Seo, V. I. Koren, S. Reed, Z. Zhang, Q. Duan, S. Cong, F. Moreda and R. Anderson, *J. Hydrol.* **298(1±4)**, 4 (2004).
123. B. E. Vieux and F. G. Moreda, in *Calibration of Watershed Models*, Eds. Q. Duan, S. Sorooshian, H. Gupta, H. Rosseau, R. Turcotte (Water Science and Applications, American Geophysical Union, 2003), p. 267.
124. R. Grayson and G. Bloschl, Eds., *Spatial Patterns in Catchment Hydrology: Observations and Modelling* (Cambridge University Press, Cambridge, UK, 2000) ISBN 0-521-63316-8.
125. P. Pokhrel, H. V. Gupta and T. Wagener, *Water Resour. Res.* **44**, W12419, doi:10.1029/2007WR006615 (2008).
126. J. Doherty and B. E. Skahill, *J. Hydrol.* **327**, 564 (2006).
127. A. Tikhonov and V. Arsenin, *Solutions of Ill-posed Problems* (V. H. Winston, 1977).
128. V. Cristina, F. Barbosa, J. B. Silva, *Geophysics* **59(1)**, 57 (1994).
129. H. W. Engl, M. Hanke and A. Neubauer, *Regularization of Inverse Problems*, (Kluwer, Dordrecht, 1996)
130. J. Doherty, *Ground Water* **41(2)**, 170 (2003).
131. J. Linden, B. Vinsonneau and K. J. Burnham, in *Proceedings of the 18th International Conference on Systems Engineering*, (IEEE Computer Society, Las Vegas Nevada, 2005), p. 112.
132. R. Fletcher, *IMA J. Appl. Math.* **15**, 319 (1974).
133. K. K. Yilmaz, *Ph.D. Dissertation* (Department of Hydrology and Water Resources, University of Arizona, Tucson, USA, 2007), pp. 263.
134. G. Demoment, *IEEE Trans. Acoustics Speech Signal Proc.* **37(12)**, 2024 (1989).
135. C. L. Lawson and J. R. Hanson, *Classics in Applied Mathematics* **15**, (Society for Industrial and Applied Mathematics, Philadelphia, PA, 1995), pp 337.
136. R. Weiss and L. Smith, *Water Resour. Res.* **34(4)**, 647 (1998).
137. T. H. Skaggs and Z. J. Kabala, *Water Resour. Res.* **30(1)**, 71 (1994).
138. E. E. van Loon and P. A. Troch, *Hydrol. Process.* **16**, 531 (2002).

139. M. J. Tonkin and J. Doherty, *Water Resour. Res.* **41**, W10412, doi:10.1029/2005WR003995 (2005).
140. E. Keating, J. Doherty, J. A. Vrugt and Q. Kang, *Water Resour. Res.* (submitted, 2008).
141. J. A. Vrugt, J. Doherty, C. J. F. ter Braak and M. P. Clark, *Proc. Natl. Acad. Sci. USA* (to be submitted, 2009).
142. C. Bandaragoda, D. Tarboton and R. Woods, *J. Hydrol.* **298**(1±4), 178 (2004).
143. L. W. White, B. E. Vieux and D. Armand, *Adv. Water Resour.* **26**, 337 (2003).
144. B. E. Vieux, *Distributed Hydrologic Modeling using GIS* (Water Science and Technology Library, Kluwer Academic Publishers, The Netherlands, 2001), pp. 293.
145. B. A. Tolson, and C. A. Shoemaker, *Water Resour. Res.* **43**, W01413, doi:10.1029/2005WR004723 (2007).
146. R. G. Regis and C. A. Shoemaker, *J. Global Optim.* **37**(1), 113 (2007).
147. R. G. Regis and C. A. Shoemaker, *J. Global Optim.* **31**(1), 153 (2005).
148. P. F. Dornes, J. W. Pomeroy, A. Pietroniro, and D. L. Verseghy, *J. Hydrometeor.* **9**, 789 (2008).
149. B. A. Tolson, and C. A. Shoemaker, *Water Resour. Res.* **44**, W04411, doi:10.1029/2007WR005869 (2008).
150. A. Behrang, B. Khakbaz, J. A. Vrugt, Q. Duan, and S. Sorooshian, *Water Resour. Res.* **44**, W12603, doi:10.1029/2007WR006429 (2008).
151. B. A. Tolson, and C. A. Shoemaker, *Water Resour. Res.* **44**, W12604, doi:10.1029/2008WR006862 (2008).
152. G. Kuczera and E. Parent, *J. Hydrol.* **211**, 69 (1998).
153. B. C. Bates and E. P. Campbell, *Water Resour. Res.* **37**(4), 937 (2001).
154. K. Engeland and L. Gottschalk, *Hydrol. Earth Syst. Sci.* **6**(5), 883 (2002).
155. J. A. Vrugt, H. V. Gupta, W. Bouten and S. Sorooshian, *Water Resour. Res.* **39**(8), 1201, doi:10.1029/2002WR001642 (2003).
156. L. Marshall, D. Nott and A. Sharma, *Water Resour. Res.* **40**, W02501, doi:10.1029/2003WR002378 (2004).
157. Y. Liu and H. V. Gupta, *Water Resour. Res.* **43**, W07401, doi:10.1029/2006WR005756 (2007).
158. W. K. Hastings, *Biometrika* **57**, 97 (1970).

159. A. Gelman and D. B. Rubin, *Stat. Sci.* **7**, 457 (1992).
160. J. A. Vrugt, C. J. F. ter Braak, M. P. Clark, J. M. Hyman and B.A. Robinson, *Water Resour. Res.* **44**, W00B09, doi:10.1029/2007WR006720. (2008).
161. C. J. F. ter Braak, *Stat. Comput.* **16**, 239 (2006).
162. G. O. Roberts and J. S. Rosenthal, *Stat. Sci.* **16**, 351 (2001).
163. G. Evensen, *J. Geophys. Res.* **99**, 10143 (1994).
164. J. A. Vrugt, C. G. H. Diks, H. V. Gupta, W. Bouten and J. M. Verstraten, *Water Resour. Res.* **41**, W01017, doi:10.1029/204WR003059 (2005).
165. G. Evensen, *J. Geophys. Res.* **97**, 17905 (1992).
166. J. R. Valstar, D. B. McLaughlin, C. B. M. T. Stroet and F. C. van Geer, *Water Resour. Res.* **40(5)**, W05116 (2004).
167. N. Bulygina and H. V. Gupta, *Water Resour. Res.* **45**, W00B13, doi:10.1029/2007WR006749 (2009).
168. J. A. Hoeting, D. Madigan, A. E. Raftery and C. T. Volinsky, *Stat. Sci.* **14(4)**, 382 (1999).
169. A. E. Raftery, T. Gneiting, F. Balabdaoui and M. Polakowski, *Monthly Weather Rev.* **133**, 1155 (2005).
170. J. A. Vrugt, M. P. Clark, C. G. H. Diks, Q. Duan, B. A. Robinson, *Geophys. Res. Lett.* **33**, L19817, doi:10.1029/2006GL027126 (2006).
171. J. A. Vrugt and B. A. Robinson, *Water Resour. Res.* **43**, W01411, doi:10.1029/2005WR004838 (2007).
172. S. P. Neuman, *Stoch. Environ. Res. Risk Assess.* **17(5)**, 291 (2003).
173. M. Ye, S. P. Neuman and P. D. Meyer, *Water Resour. Res.* **40(5)**, W05113, doi:10.1029/2003WR002557 (2004).
174. M. Ye, S. P. Neuman, P. D. Meyer and K. Pohlmann, *Water Resour. Res.* **41**, W12429, doi:10.1029/2005WR004260 (2005).

Yilmaz, K. K., Vrugt, J.A., Gupta, H.V., Sorooshian, S. 2009: Model Calibration in Watershed Hydrology, In: Advances in Data-based Approaches for Hydrologic Modeling and Forecasting, Eds: B. Sivakumar and R. Berndtsson, World Scientific

SUMMARY

Hydrologic models use relatively simple mathematical equations to conceptualize and aggregate the complex, spatially distributed, and highly interrelated water, energy, and vegetation processes in a watershed. A consequence of process aggregation is that the model parameters often do not represent directly measurable entities and must, therefore, be estimated using measurements of the system inputs and outputs. During this process, known as model calibration, the parameters are adjusted so that the behavior of the model approximates, as closely and consistently as possible, the observed response of the hydrologic system over some historical period of time. This Chapter reviews the current state-of-the-art of model calibration in watershed hydrology with special emphasis on our own contributions in the last few decades. We discuss the historical background that has led to current perspectives, and review different approaches for manual and automatic single- and multi-objective parameter estimation. In particular, we highlight the recent developments in the calibration of distributed hydrologic models using parameter dimensionality reduction sampling, parameter regularization and parallel computing. Finally, this chapter concludes with a short summary of methods for assessment of parameter uncertainty, including recent advances in Markov chain Monte Carlo sampling and sequential data assimilation methods based on the Ensemble Kalman Filter.