Towards improved treatment of parameter uncertainty in hydrologic modeling

Vrugt, J.A.

Citation for published version (APA):

General rights
It is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), other than for strictly personal, individual use, unless the work is under an open content license (like Creative Commons).

Disclaimer/Complaints regulations
If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the Library know, stating your reasons. In case of a legitimate complaint, the Library will make the material inaccessible and/or remove it from the website. Please Ask the Library: http://uba.uva.nl/en/contact, or a letter to: Library of the University of Amsterdam, Secretariat, Singel 425, 1012 WP Amsterdam, The Netherlands. You will be contacted as soon as possible.

Download date: 17 Feb 2019

UvA-DARE is a service provided by the library of the University of Amsterdam (http://dare.uva.nl)
Towards improved treatment of parameter uncertainty in hydrologic modeling

Jasper Alexander Vrugt
Towards improved treatment of parameter uncertainty in hydrologic modeling
Towards improved treatment of parameter uncertainty in hydrologic modeling / J.A. Vrugt
Thesis Universiteit van Amsterdam – With ref. – With summaries in English and Dutch.

This study was carried out at the Centre for Geo-Ecological Research (ICG) and the Institute for Biodiversity and Ecosystem Dynamics (IBED) - Physical Geography, Faculty of Science, Universiteit van Amsterdam, The Netherlands.

The research was supported by the Earth and Life Sciences Council (ALW) with financial aid from the Netherlands Organization for Scientific Research (NWO).
Contents of this thesis

Acknowledgments ........................................................................................................... 5
1. Introduction .................................................................................................................. 9

Part I: Stating the Case – A Typical Parameter Estimation Problem

2. One-, Two-, and Three-Dimensional Root Water Uptake Functions for
   Transient Modeling .................................................................................................... 23

Part II: Single-objective Parameter Estimation

3. A Shuffled Complex Evolution Metropolis Algorithm for Optimization and
   Uncertainty Assessment of Hydrologic Model Parameters .................................... 51

4. Inverse Modeling of Large-Scale Spatially-Distributed Vadose Zone
   Properties Using Global Optimization .................................................................... 81

Part III: Recursive Parameter Estimation

5. Toward Improved Identifiability of Hydrologic Model Parameters:
   The Information Content of Experimental Data ..................................................... 119

Part IV: Multi-objective Parameter Estimation

6. Effective and Efficient Algorithm for Multi-Objective Optimization of
   Hydrologic Models .................................................................................................... 145

Part V: Simultaneous Parameter and State Estimation

7. Improved Treatment of Uncertainty in Hydrologic Modeling:
   Combining the Strengths of Global Optimization and Data Assimilation ............ 181

8. Epilogue ....................................................................................................................... 215

Summary (English) ......................................................................................................... 227

Summary (Dutch) .......................................................................................................... 233

List of publications ....................................................................................................... 239

References ..................................................................................................................... 241

Curriculum Vitae .......................................................................................................... 253
He calls me Charlie Mason
A stargazer am I
It seems that I was born
To chart the evening sky
They'd cut me out for baking bread
But I had other dreams instead
This baker's boy from the west country
Would join the Royal Society...

Now hold your head up, Mason
See America lies there
The morning tide has raised
The capes of Delaware
Come up and feel the sun
A new morning is begun
Another day will make it clear
Why your stars should guide us here...

Excerpt from: Mark Knopfler, Sailing to Philadelphia, Mercury Records 2000
ACKNOWLEDGMENTS

17 October 1998: I will never forget my excitement as I boarded an Icelandair Boeing at Schiphol Airport in The Netherlands, and headed off for the beginnings of an unimagined future. I certainly had no thoughts of where this journey would ultimately take me. Now, after about 200,000 kilometers of travel - and about 4.5 years since joining the University of Amsterdam as PhD-student, the first part of this unexpected journey has been completed.

A career outside Holland was not my original intention when I fled the wonderful scenery and canals of Amsterdam to pursue the uncertain dream of making a place for myself in the scientific arena. I had hoped to find a home and job close to family and friends, but after thinking about this lifestyle, I eventually opened myself for more challenging ambitions. After a few years of chasing these goals, I am delighted to present the first results of my journey in this thesis. This book provides a comprehensive chapter-by-chapter overview of the scientific research that has been conducted by me and colleagues during my PhD-appointment.

Successfully completing a PhD requires hard-work, dedication, creativity, but perhaps most importantly, teamwork. With great pleasure and enthusiasm, I have pleasantly worked with various individuals during the course of my PhD-research. I would, therefore, like to take the opportunity to thank all the members of the team, who made the first steps of my journey easier by exuding enthusiasm for my research and helping me along the way.

I would like to start by gratefully acknowledging my PhD-advisor Willem Bouten. Willem, your open mind, creativity, and critical attitude have always stimulated me to optimize my own functioning and performance. Due to your involvements from the early days on, I was eventually able to parlay an inexperienced researcher and writer into the beginnings of a career as scientist. You have also given me more than sufficient opportunity to explore myself and visit various experts in hydrologic sciences in the United States to further facilitate the progress of my research.

As I revisit the various stages of my PhD-research, I notice that the environment in Amsterdam was always created in such a way that I was able to optimally exploit my own talents. For this, I am very grateful to Koos Verstraten, my other advisor at the University of Amsterdam. Koos, I have always been impressed by your people managing skills and touched by your personal interest in me and my life. Moreover, you have afforded me many lessons in humility and how to manage people in a department with diverse interests, research disciplines and goals.
ACKNOWLEDGMENTS

In the second year of the PhD-journey (2001), I made a work visit of about six months, to the University of Arizona, Tucson to work closely with Hoshin Gupta and Soroosh Sorooshian, to facilitate meeting the objectives of the research. This all seems a long way from the start of my PhD-appointment, although the visit was foreseen in 1999 during the initial writing of the research proposal for the Netherlands Organization for Scientific Research. Hoshin, the free and unfettered exchange of ideas during our research conversations have built the framework for Chapters 3 and 5 in this thesis. Your active involvement and expertise and gift to explain complex mathematical systems engineering concepts in simple language have stimulated me in the development of the various papers in this thesis. Soroosh, due to your always-busy traveling schedule, we only had a few moments to meet in the last few years. Nonetheless, I would like to express my gratitude for your interests in my research and family and the financial support of SAHRA and other NSF projects. Your international reputation and skills as an entrepreneur continue to inspire me.

During the early stages of my first visit to Tucson, I had mysterious company. By pure-luck and a lot of persistence, I met Roya, my future wife (three days after my PhD-graduation we will get married). She and her extraordinary family and culture have made me morph from a Dutchman to a man of the world. Roya also made me finally come to accept and embrace a career outside Europe, far from childhood friends and family. I am very grateful to her and her family for their love, support, hospitality, generosity and most importantly, for taking me into their hearts along the way.

Throughout the summer of 1999 and 2003, I spent a total of nine months visiting Jan Hopmans at the University of California, Davis to work on unsaturated flow and transport modeling. These visits helped me lay the foundation for Chapters 2 and 4 in this thesis. Jan, much of the research in these two Chapters would not have been possible without your active involvement, suggestions, and financial support. I am very pleased with our mutual research interests and thankfully appreciate the opportunity you provided to visit Davis and work together.

On a personal note, I would like to express my appreciation to all of my colleagues in Amsterdam, Tucson, and Davis for their help, sympathy and humor. I am also gratefully indebted to the various co-authors of my papers for their hard-work and dedication, including all individuals who successfully applied the various developments in this thesis to other research disciplines. Sometimes I think it would have been nice to have all my published papers as chapters in this thesis, but then again, it may be a service to potential readers that they are
destined to remain in journals only. Thanks are also due to my friends for their labored efforts and help to build an enigmatic and versatile life outside the University.

Finally, I would like to thank my parents, two distinguished and unique members of the team. They gave me the opportunity and support to study and to be educated to the University level. Their unconditional love, trust, belief, and endless support, in all possible ways, led to this thesis.

Amsterdam, 8 May 2004.
Jasper Vrugt
CHAPTER 1

Introduction

1.1. General background

Hydrology involves the study of processes describing the movement of water on the earth's surface, from the land surface to the subsurface and from the subsurface to the surface to the atmosphere, as schematically outlined in Figure 1.1.

Figure 1.1. Schematic overview of the various components of the hydrologic cycle [graph taken from the website http://watercycle.gsfc.nasa.gov].

These processes occur on several time scales: from hours to days to weeks, for example, flash floods, water infiltration in soils, short-term droughts, and spring-time runoff events; to longer decadal to inter-decadal processes involving long-term droughts and changes in regional precipitation and water supply due to climate variability. Faced with the complexity and spatial
and temporal variability of hydrologic processes, and the difficulty of performing controlled experiments, a variety of hydrologic models have been developed to simulate the various components of the water cycle for a region of interest. These models are classified as conceptual, when water balance dynamics are represented by heuristic or empiric functions deemed to be qualitatively reasonable, or as physically-based when the functions are based on scientifically accepted principles [Kuczera, 1997]. Modeling is a framework for testing new theories and hypothesis in order to improve our understanding of the hydrological processes and how the different processes interact. Even the most elaborate physically-based model, however, cannot reflect the true complexity and heterogeneity of the processes occurring in the field, because to some degree, it must always conceptualize and aggregate complex interactions driven by a number of spatially distributed and highly interrelated water, energy, and vegetation processes by the use of only relatively simple mathematical equations. All hydrologic models are, therefore, to some degree lumped, and use effective parameters to characterize these spatial-temporal processes.

In order for a hydrologic model to reliably simulate a portion of the water cycle it is necessary to select appropriate values for the model parameters. Unfortunately, the estimation of the correct values of these effective parameters has not proved to be simple. As argued in the first paragraph, many (if not all) of the parameters in hydrologic models represent aggregated and lumped processes in space and time, which are not easily observable or measurable at the scale of interest, and therefore cannot readily be inferred from direct measurable quantities. Instead, the typical way to estimate the model parameters is to adjust them in such a way that the input-output behavior of the model approximates, as closely and consistently as possible, the underlying hydrologic system over some historical period of time. This process is known as model calibration. The parameters, which are estimated in this manner, represent effective conceptual representations of spatially and temporally heterogeneous watershed properties, insofar as they are related to inherent and invariant properties of the hydrologic system. After successful calibration the hydrologic model is ready to be used to predict the various components of the water cycle during future events under a set of naturally occurring circumstances. This modeling process is closely related to the development of understanding of hydrologic processes in the sense that a comparison of model predictions with observed system behavior (observations) is commonly used to test hypothesis about the operation of the system being modeled.
1.2. Classical paradigm of model calibration

A schematic overview of the model calibration process appears in Figure 1.2.

Consider a system \( \mathcal{S} \) for which a hydrologic model \( \eta \) is to be calibrated. Assume that the mathematical structure of the model is essentially predetermined and fixed and that realistic upper and lower bounds on each of the \( p \) model parameters can be specified prior to any modeling being undertaken. Let \( \mathbf{y} = \{ y_1, ..., y_n \} \) denote the vector of measurement data available at time steps 1, ..., \( n \) and let \( Y(\theta) = \{ y_1(\theta), ..., y_n(\theta) \} \) represent the corresponding vector of model output predictions using the model \( \eta \) with the parameter values \( \theta \). The difference between the model-simulated output and measured data can be represented by the residual vector, \( E \):

\[
E(\theta) = G[\hat{y}(\theta)] - G[y] = \{ e_1(\theta), ..., e_n(\theta) \}
\]  

(1.1)

where the function \( G(\cdot) \) allows for various user-selected linear or nonlinear transformations. The aim of model calibration now becomes finding those set of model parameters \( \theta \) such that the measure \( E \), commonly called the objective function, is in some sense forced to be as close to zero as possible. The development of a measure that mathematically measures the “size” of \( E(\theta) \)
is typically based on assumptions regarding the distributions of the measurement errors presented in the data.

The following development closely follows the work presented in Vrugt et al. [2003a] and Vrugt and Dane [2004]. The classical approach to estimating the parameters in Eq. (1.1) is to ignore input data uncertainty and to assume that the predictive model $\eta$ is a correct representation of the underlying physical data-generating system ($\mathbb{F}$). In line with classical statistical estimation theory, the residuals in Eq. (1.1) are then assumed to be mutually independent (uncorrelated), and Gaussian distributed with a constant variance. Under these circumstances, the traditional “best” parameter set in Eq. (1.1) can be found by minimizing the following additive simple least square (SLS) objective function with respect to $\theta$:

$$F_{\text{SLS}}(\theta) = \sum_{i=1}^{n} \varepsilon_i(\theta)^2$$  \hspace{1cm} (1.2)

This step has, in practice, proven to be quite difficult to carry out in a reliable and consistent manner, because there are typically a large number of parameters that need to be estimated, which usually either have similar or compensating effects on different parts of the simulated output. In combination with varying parameter sensitivity, this not only leads to considerable interaction among the model parameters, but perhaps more importantly, results in numerous local minima in the response surface (objective function mapped out in the parameter space).

### 1.3. Manual and automatic parameter estimation

In the process of parameter estimation or model calibration, the hydrologist adjusts the values of the model parameters such that the model is able to closely match the behavior of the real system it is intended to represent. In its most elementary form, this calibration is performed by manually adjusting the parameters while visually inspecting the agreement between observations and model predictions [Janssen and Heuberger, 1995]. In this approach, the "closeness" of the model with the measurements is typically evaluated in terms of several subjective visual measures, and a semi-intuitive trial-and-error process is used to perform the parameter adjustments [Boyle et al., 2000]. Because of the subjectivity and time-consuming nature of manual trial-and-error calibration, there has been a great deal of research into the development of automatic methods for calibration of hydrologic models [e.g., Gupta and Sorooshian, 1994].
Automatic methods for model calibration seek to take advantage of the speed and power of computers, while being objective and easier to implement than manual methods.

Many algorithms have been developed in the past to solve the nonlinear SLS optimization problem stated in Eq. (1.2). These algorithms may be classified as local search methodologies, when seeking for systematic improvement of the objective function using an iterative search starting from a single arbitrary initial point in the parameter space, or as global search methods in which multiple concurrent searches from different starting points are conducted within the parameter space. One of the simplest local search optimization methods, which is commonly used in the field of soil hydrology, is a Gauss-Newton (Levenberg-Marquardt) type of derivative based search [Marquardt, 1963]:

$$\theta^{(k+1)} = \theta^{(k)} - H(\theta^{(k)})^{-1} \nabla \eta(\theta^{(k)})'$$

(1.3)

where $\theta^{(k+1)}$ is the updated parameter set, and $\nabla \eta(\theta^{(k)})$ and $H(\theta^{(k)})$ denote the gradient and Hessian matrix, respectively, evaluated at $\theta = \theta^{(k)}$:

$$\nabla \eta(\theta) = \begin{pmatrix} \frac{\partial \eta}{\partial \theta_1}(\theta) \\ \vdots \\ \frac{\partial \eta}{\partial \theta_p}(\theta) \end{pmatrix}$$

$$H(\theta) = \begin{pmatrix} \frac{\partial^2 \eta}{\partial \theta_i \partial \theta_j}(\theta) \\ \vdots \end{pmatrix}_{i,j}$$

(1.4)

From an initial first guess of the parameters $\theta^{(0)}$, a sequence of parameter sets, $\{\theta^{(0)}, \ldots, \theta^{(k+1)}\}$, is generated that is intended to converge to the global minimum of $F(\theta)$ in the parameter space. If the model $\eta$ depends linearly on each parameter $\theta_j (j = 1, \ldots, p)$ the minimization problem stated in Eq. (1.2) reduces to a linear regression problem for which analytical solutions exist.

The derivative based search defined in Eqs. (1.3) and (1.4) will evolve towards the global minimum in the parameter space in situations where the response function (the objective function mapped out in the feasible parameter space) exhibits a topographical convex shape. However, practical experience with hydrologic models suggests that the objective function seldom satisfies these restrictive conditions. To illustrate the severity of the optimization problem, we closer examine a typical example in the field of vadose zone hydrology, which considers the inverse estimation of the soil water retention and unsaturated soil hydraulic conductivity characteristics using data from a transient one-step laboratory outflow experiment.
Figure 1.3 [taken from Vrugt and Bouten, 2002], presents the results of such an analysis, in terms of marginal probability distributions (histograms) for two soil hydraulic parameters in the vicinity of the global minimum of the parameter space.

![Figure 1.3](image)

**Figure 1.3.** Marginal probability distributions of two soil hydraulic parameters in the vicinity of the global minimum, (a) residual soil water content, and (b) inverse of air-entry value.

If the objective function would exhibit a convex shape in the entire parameter domain, the histograms for each of the soil hydraulic parameters would exhibit a clear Gaussian distribution with a single mode. However, the large number of different modes (local minima) for each of the parameters depicted in Fig. 1.3 is the most probable reason for the numerous reports in the literature of the inability to find a unique set of hydraulic parameter values using observed outflow dynamics from one-step outflow experiments [Kool et al., 1985; Parker et al., 1985; Toorman et al., 1992; van Dam et al., 1992; among others]. As the local gradient-based search algorithms are not designed to handle the peculiarities of the response surface illustrated in Fig. 1.3, they will terminate their search prematurely with their final solution essentially being dependent on the starting point in the parameter space. Another emerging problem, reported by Hopmans et al. [2002], is that some of the hydraulic parameters are typically highly correlated using observed outflow data, further lowering the chance of getting a single unique solution with local search methodologies.

The existence of non-uniqueness problems with local search methodologies has led soil hydrologists to argue that there is not sufficient information in the outflow measurements to enable a unique characterization of the soil hydraulic properties. So even with laboratory experiments, where a soil can be manipulated with great flexibility, non-uniqueness problems are often experienced [van Dam et al., 1992]. Seemingly, there was a widespread conviction that the best way to solve the non-uniqueness problem would be to add more and better measurements.
INTRODUCTION

[Elbing and Hopmans, 1993; van Dam et al., 1994]. However, research into data requirements has led to the understanding that the information content of the data is far more important than the amount of data used for the identification of the model parameters [Sorooshian et al., 1983; Gupta and Sorooshian, 1985; Yaço et al., 1996]. Vrugt et al. [2003a] argued, therefore, that it is not the information in the measurements that is lacking to obtain a unique set of parameters, but the fact that the classical local-search optimization procedures utilized in soil hydrology are typically not powerful enough to solve the problems illustrated in Fig. 1.3. Their arguments on uniqueness of parameters are not based on the convergence problems of the applied optimization methods, but on the shape of the posterior marginal distributions of the parameters. Hence, closer inspection of Fig. 1.3 demonstrates that for each of the parameters there is a single optimum with highest posterior probability in the global minimum. Indeed, although not shown here, the multivariate probability distribution of the parameters confirms that these regions with highest posterior probability for each parameter coincide.

Initial responses to similar failure problems of local search algorithms in the area of watershed model calibration, during the 1980s were to try and put the optimization problem onto a more rigorous statistical footing using maximum likelihood or Bayesian theory [Sorooshian and Dracup, 1980; Kuc^era, 1983]. However, these approaches did not directly address the causes of the inability to find the optimum for a selected objective function.

To be able to deal with the peculiarities of the response surface so evident in Fig. 1.3, we must, therefore, develop automatic calibration algorithms which can deal with the existence of multiple optima in the parameter space (with both small and large domains of attraction), discontinuous first derivatives and curving multi-dimensional ridges. These considerations inspired Duan et al. [1992] to develop a powerful, robust, and efficient, optimization procedure, entitled, the Shuffled Complex Evolution (SCE-UA) global optimization algorithm. By merging the strengths of the Downhill Simplex procedure [Nelder and Mead, 1965] with the concepts of controlled random search, systematic evolution of points in the direction of global improvement, competitive evolution [Holland, 1975], and complex shuffling, the SCE-UA algorithm represents a synthesis of the best features of several optimization strategies. The strength and reliability of the SCE-UA global optimization algorithm have since been independently tested and proven by numerous researchers and the algorithm is now extensively used world-wide [e.g., Duan et al., 1992; Sorooshian et al., 1993; Gan and Biftn, 1996; Kuc^era, 1997; Hogue et al., 2000; Boyle et al., 2000; among many others]. With the current available SCE-UA algorithm, one can now have confidence that the global minimum of a predefined objective function is found. However, the
method does not provide any information about the uncertainty associated with the final estimated parameters.

1.4. Limitations of current model calibration strategies

Despite the progress made, several contributions to the hydrologic literature in the past decade have brought into question the continued usefulness of the classical paradigm for model calibration [Beven and Binley, 1992; Gupta et al., 1998; Kavetski et al., 2003]. For the purpose of this discussion, we classify these into four categories:

(1) The classical model calibration approach typically ignores parameter uncertainty. The classical method is based on traditional statistical non-linear regression theory, which operates under the central assumption that the available model structure is correct, and therefore seeks to identify a unique “optimal” set of parameter estimates, which minimizes the SLS estimator in Eq. (1.2). Practical experience with the calibration of hydrologic models suggests, however, that it is typically difficult to find a unique “best” parameter set whose performance measure (objective function) obviates considerable of other feasible parameter sets;

(2) The approach assumes the model parameters to be time invariant and aggregates model performance errors over a large range of hydrologic behavior. It is not very difficult to see that temporal and/or spatial aggregation of model residuals into a single additive performance statistic, such as the SLS estimator in Eq. (1.2), results in considerable loss of important information that can be used to distinguish between competing parameter sets;

(3) The approach gives little guidance how to combine different types of data sets during model calibration. With the growing popularity of sophisticated physically-based hydrologic models, the complexity of the calibration problem has been multiplied many-fold. For instance, the latest hydrologic watershed or land-surface models simulate several output fluxes (e.g. water, energy, chemical constituents, etc.) for which measurement data are available, and all these data must be correctly utilized to ensure proper model calibration;
(4) The approach typically assumes that all the uncertainty in the input-output representation of the model can be attributed to uncertainty in the parameter estimates, thereby effectively neglecting input, output and model structural uncertainty. Hence, uncertainties in the modeling procedure not only stem from uncertainties in the parameter estimates, but also from measurement errors associated with the system input and output, and from model structural errors arising from the aggregation of spatially distributed real-world processes into a mathematical model.

Summarizing, current calibration techniques are not very well suited to the task of calibrating hydrologic models, and are particularly inadequate in the face of the emerging generation of multi-input-output distributed parameter hydrologic models.

1.5. Objectives of the PhD research

The preceding sections briefly described the historical development leading to current views on model calibration, and discussed some of the techniques that have been developed for estimating parameters, thereby enabling the model to mimic the behavior of the underlying hydrologic system. Finally, in the previous section the limitations of current model calibration strategies were established. In this section, the main objectives of the PhD research are summarized. Each of the listed objectives below have to be considered in the light of trying to improve upon the classical model calibration strategy, and form a response to each of the summarized points in the previous section.

(1) Development of methods, which not only estimate the traditional “best” parameter set, but simultaneously also generate a sample set of parameter values describing the probabilistic representation of the remaining parameter uncertainty. Instead of producing a single error-free model forecast, which is the case after conventional calibration, this posterior uncertainty can be used to produce probabilistic model outputs (most likely forecast and 95% confidence intervals at each time step). In this way, at least some of the uncertainty in the input-output representation of the model is represented in the model forecasts. Moreover, by analyzing the uncertainty of the parameters and their correlation structure induced in the posterior probability distribution, the hydrologist is able to better understand which complexity of the
model is warranted by the calibration data and which parameters to select for finding useful regionalization relationships (those which are well defined after calibration).

(2) Development of methods, which increase information retrieval from the data by recursively assimilating and processing the calibration data, thereby explicitly taking into account parameter uncertainty. Careful inspection of one-dimensional projections of the evolution of the high probability density region of the parameters, will not only reveal which measurements are most informative for the model parameters (useful information for optimal experimental design strategies), but perhaps more importantly, provide a way for checking for violations of the underlying assumption that parameters are constant;

(3) Development of methods, which explicitly recognize the multi-objective nature of the model calibration problem and use innovative ways to deal with different observation types, data sub-sets and global search algorithms to identify the non-inferior solution space or preferred solutions. By analyzing the tradeoffs in the fitting of different data sub-sets the hydrologist is able to better understand the limitations of the current model structure.

(4) Development of methods, which more completely treat input, output, parameter, and model structural uncertainty in hydrologic model calibration. Traditional calibration methods assign all the uncertainty in the input-output representation to uncertainty in the parameter estimates and miss the conceptual rigor to distinguish between the various sources of errors associated with the application of hydrologic models. We hypothesize that the implementation of such a strategy will not only result in meaningful prediction uncertainty bounds on the model simulations, but will also result in parameter estimates, which better represent system properties as they are less corrupted by modeling errors. This will further enhance the prospects of finding useful regionalization relationships.

Note, that efficiency and effectiveness of these computerized algorithms/methods are important considerations.

It is important to stress, however, that the objectives of the PhD research should not be understood as that of simply developing methods for parameter estimation alone. The
development and application of these methods is a necessary step to be able to better distinguish between input, output, parameter, and model structural uncertainty in hydrologic modeling. This information will help direct resources towards model structural improvements and, as such, will help to increase our understanding and knowledge of hydrologic processes. The envisaged higher goal of the PhD-research is, therefore, to contribute to an improved understanding of the various hydrologic processes operating at or near the earth’s surface. The usefulness and applicability of these methods is, therefore, continuously demonstrated by application to a variety of different hydrologic models throughout the remainder of this thesis.

1.6. Outline of this thesis

This thesis consists of a series of eight closely related chapters, which combined address the listed objectives in the previous section. To state the case and illustrate a typical parameter estimation problem in hydrology, Chapter 2 involves the calibration of soil physical and root water uptake parameters in a physically-based three-dimensional unsaturated soil water flow model using measured spatial distributions of soil water content around a sprinkler-irrigated almond tree. This chapter serves as an illustrative example to demonstrate the capabilities of inverse modeling (parameter estimation), and highlights some of the limitations of conventional parameter estimation algorithms. They are computationally very demanding when solving for high-dimensional parameter problems. Moreover, the emphasize lies on the identification of a single best set of model parameters, thereby effectively neglecting the influence of possible sources of uncertainty on the final parameter estimates.

To overcome these limitations, Chapter 3 presents the Shuffled Complex Evolution Metropolis (SCEM-UA) global optimization algorithm, which is designed to provide an effective and efficient estimate of the traditional “best” parameter set and its underlying posterior distribution within a single optimization run. The SCEM-UA algorithm is an improvement over the original SCE-UA algorithm and uses a probabilistic offspring strategy instead of the deterministic Simplex method to search the parameter space. The power and applicability of the newly developed SCEM-UA algorithm is demonstrated by application to a hydrologic watershed model with typical conceptual components.

Chapter 4 presents another application of the SCEM-UA algorithm by inverse estimation of large-scale spatially distributed vadose zone properties using the solution of a physically-based three-dimensional distributed model combined with spatially distributed measured tile drainage data from the 9700 ha Broadview Water District in the San Joaquin Valley.
of California. To study the benefits of using a spatially distributed three-dimensional vadose zone model, the results of the three-dimensional model were compared with those obtained using a simple conceptual bucket model and a spatial-averaged one-dimensional unsaturated water flow model.

The primary focus of the previous Chapters has been on the "batch" calibration approach, which aggregates error residuals over a large range of hydrologic behavior. To reduce the information loss associated with traditional batch processing, Chapter 5 presents a recursive parameter estimation method, entitled the Parameter Identification Method based on the Localization of Information (PIMLI), which is especially designed to increase information retrieval from the calibration data and to infer the type and location of the measurements, which contain the most information for the specific model parameters.

Despite the progress made, practical experience with the calibration of hydrologic models suggests that single objective functions, no matter how carefully chosen, are often inadequate to properly measure all of the characteristics of the observed data deemed to be important. By employing a number of complementary criteria in the optimization procedure, and analyzing the tradeoffs in the fitting of these criteria, the hydrologist is able to better understand the limitations of current hydrologic model structures, and gain insights into possible model improvements. Chapter 6 presents the Multi-objective Shuffled Complex Evolution Metropolis (MOSCEM-UA) global optimization algorithm, which is capable of efficiently solving the multi-objective optimization problem for hydrologic models.

Chapter 7 surveys the conceptual limitations of the PIMLI, SCEM-UA, and MOSCEM-UA algorithms (amongst others) for distinguishing between input, output, parameter and model structural error, and presents SODA, a Simultaneous Optimization and Data Assimilation method, which more completely treats the various sources of errors in hydrologic model calibration. One of the goals of the SODA framework is to produce a time series of error residuals that have the desirable properties of constant variance and independence in time and space, so that unbiased estimates of the model parameters are obtained.

Finally, Chapter 8 presents an epilogue, in which the most important research questions that can now be forcefully addressed with the availability of the SCEM-UA, PIMLI, MOSCEM-UA, and SODA parameter and state estimation methods are highlighted and a view on future research in model calibration is explicated. A summary of all the preceding chapters is given in the conclusion section of this thesis.
Part I

Stating the Case –
A Typical Parameter Estimation Problem
CHAPTER 2

One-, Two-, and Three-Dimensional Root Water Uptake Functions for Transient Modeling

Abstract

Although solutions of multi-dimensional transient water flow can be obtained by numerical modeling, their application may be limited as root water uptake is generally considered to be one or two-dimensional only. This is especially the case for trees. The first objective of this Chapter is to test the suitability of a three-dimensional root water uptake model for the simultaneous simulation of transient soil water flow around an almond tree. The soil hydraulic and root water uptake parameters were optimized by minimizing the residuals between measured and simulated water content data. Water content was measured in a three-dimensional grid around a sprinkler-irrigated almond tree for a 16-day period, following irrigation. A second objective was to compare the performance and results of the three-dimensional flow model using with one- and two-dimensional root water uptake models. For this purpose, measured water contents were aggregated in the $x$- and $y$-direction in the one-dimensional case and in the radial direction for the two-dimensional uptake model. For the estimation of root water uptake model parameters, a Genetic Algorithm was used to estimate the approximate global minimum of the parameter space, whereas final parameters were determined using the Simplex optimization algorithm. With the optimized root water uptake parameters, simulated and measured water contents during the 16-day period were in excellent agreement for all root water uptake models. Most significantly, the spatial variation in flux density below the rooting zone decreased when reducing multi-dimensional root water uptake to fewer dimensions, thereby justifying the proposed multi-dimensional approach.

This Chapter has been published by: Jasper A. Vrugt, Mark T. van Wijk, Jan W. Hopmans, and Jirka Šimůnek, One-, two-, and three-dimensional root water uptake functions for transient modeling, Water Resources Research, 37(10), 2457-2470, 2001. © Reprinted with permission of the American Geophysical Union.
CHAPTER 2

2.1. Introduction and scope

From a hydrological perspective, water uptake by root systems and their spatial distribution may exert a large degree of control on the water fluxes to the atmosphere and the groundwater [Canadell et al., 1996]. For an improved understanding of the magnitude of these fluxes, accurate estimates of the temporal and spatial root water uptake patterns are needed. Clearly, quantification of root water extraction rates also contributes to an understanding of chemical fluxes in the vadose zone in both ecological and hydrological studies [Somma et al., 1998], as well as their control by vegetation. Interactions between roots and soil in the rhizosphere influence the quality and quantity of water transport in and exported from the vadose zone. An understanding of the interactions between the roots and surrounding soil, and solutes under a variety of changing environmental conditions, has large implications since it will lead to a decrease in contamination of subsurface and surface water resources by reducing the loss of fertilizers and other agrochemicals below the root zone [Clausnitzer and Hopmans, 1994; Clothier and Green, 1994]. Moreover, the rhizosphere might be responsible for accelerated breakdown of organic chemicals by biodegradation [Walton and Anderson, 1990] and phytoremediation [Nyer and Gatiloff, 1996], hence a thorough understanding of root function regarding uptake of water and associated solutes is warranted.

Actual root water uptake not only spatially depends on the root density distribution, but also on its temporal functioning as determined by soil water availability and soil salinity. In addition to water stress in periods of low water availability, root water uptake is also reduced when concentrations of soluble salts exceed plant-specific threshold values [Homaei, 1999]. In irrigated soils, particularly in arid and semi-arid regions, plants are generally subjected to both salinity and water stress. In these regions soil and water management practices are based on maintaining favorable soil water content and salinity status in the root zone, thereby minimizing periods of water stress while controlling leaching to minimize salinity stress.

The influence of plant root systems on water and chemical movement can be better understood using soil water simulation models, provided that accurate spatial and temporal root water uptake distributions are included [Musters and Bouteijn, 1999]. One of the earliest detailed quantitative studies of water extraction by a plant root was presented by Gardner [1960]. This microscopic model considered a single root to be equivalent to an infinitely long cylinder of uniform radius with water-absorbing properties. The steady-state soil water flow equation was solved analytically assuming radial flow and imposed root water uptake rates. Soil water matric head distributions around the idealized root were calculated. This concept was extended in later
papers [Gardner, 1964, 1965; Gardner and Ehlig, 1962] and proved to be very insightful; however, lacked practical applicability since the detailed geometry of the rooting system is difficult to measure and is time dependent. Consequently, most root water extraction terms have been developed using a macroscopic rather than a microscopic approach. Nonetheless, all macroscopic models preserve the essence of Gardner’s [1960] insight.

In the macroscopic approach of Richards’ equation, a sink term representing the water extraction of the entire root system, is included to describe transient multi-dimensional water flow [Whisler et al., 1968; Molz and Remson, 1970; Clausnitzer and Hopmans, 1994], according to:

\[
\frac{\partial \theta}{\partial t} = \nabla \cdot \left[ K \nabla (h - \xi) \right] - S(x, y, z, t) \tag{2.1}
\]

where \( \theta (L^3) \) is the volumetric soil water content, \( K (LT^{-1}) \) is the unsaturated hydraulic conductivity tensor, \( h (L) \) is the soil water matric head and \( \xi (L) \) is the depth which is included for vertical flow only, and \( S (L^3 T^{-1}) \) is the volumetric sink term, representing root water uptake as a function of both space and time. The benefits of such an approach are evident since it allows direct integration of root water uptake with transient soil water flow, and providing natural interactions between transpiration and root water extraction. That is, as long as we know the mechanisms with which to describe \( S \).

Though transient soil water flow in the vadose zone is often simulated in one, two and three spatial dimensions, root water uptake is general considered simply to be a function of the vertical dimension only. For uniform crops with a spatially uniform water uptake pattern, one-dimensional models may suffice. However, for row crops and tree lines, for example, a two-dimensional representation would be better. For isolated trees, such as apples and almonds in large monocultures, the process of water uptake is complex and a three-dimensional representation would therefore seem appropriate [Green and Clothier, 1999]. Flow models such as HYDRUS-3D, being an update of the SWMS_3D code of Šimůnek et al [1995] allow multi-dimensional root water uptake. However, spatial characterization of root water uptake data is generally lacking to support multi-dimensional root water uptake parameters. Moreover, available uptake models are largely limited to one dimension only [Feddes et al., 1976; Molz, 1981; Jarvis, 1989], and describe variations in water uptake with soil depth while allowing for reduction in uptake by soil water stress. Exceptions are the two-dimensional models proposed by Neuman et al. [1975], Warrick et al. [1980], Coelho and Or [1996] and most recently, Vrugt et al. [2001b].
In the past few years, computing capabilities have significantly improved the effectiveness of multi-dimensional, soil water flow models to study spatial and temporal patterns of root water uptake. Such a multi-dimensional approach in root water uptake is needed if uptake is varying in space thereby allowing a more accurate quantification of spatial variability of the soil water regime, including the water and solute flux densities below the rooting zone. The objective of this study is threefold. Firstly, we test the suitability of a three-dimensional model for the simultaneous, dynamic simulation of soil water flow and root water uptake. A three-dimensional finite-element grid over the considered soil domain serves to define the spatial distribution of soil physical properties and root characteristics and acts as a framework for the transient water flow model. Soil physical and root parameters are subsequently estimated using inverse modeling while using the measured spatial distribution of water contents around a sprinkler-irrigated almond tree during a 16-day period. A second objective was to compare the results of the three-dimensional analysis with numerical models describing soil water flow and root water uptake in one and two dimensions, with uptake parameters optimized using the same field data set. Finally, the third goal of the simulation study is to evaluate the improved prediction of the spatial variability of soil water flux taking into consideration the multi-dimensionality of root water uptake. Whereas verification of the presented multi-dimensional uptake model is limited to a single data set, the main goal of this study is to emphasize the importance of multi-dimensional root water uptake modeling in root zone domains with spatially distributed root water uptake.

2.2. Materials and methods

2.2.1. One-, two-, and three-dimensional root water uptake model

Recently, Vrugt et al. [2001b] proposed a one-dimensional root water uptake model, which was shown to be very flexible. It was based on the model by Raats [1974]:

$$\beta(\zeta) = \left(1 - \frac{\zeta}{Z_m}\right) \exp\left(-\frac{P_{\xi}}{Z_m}\left|\zeta^* - \zeta\right|\right)$$

(2.2)

where $\beta(\zeta)$ is a shape factor describing the spatial distribution of potential root water uptake with depth, $Z_m$ (L) is the maximum rooting depth, and $P_{\xi}$ and $\zeta^*$ (L) are empirical parameters. These parameters are included to provide for zero root water uptake at $\zeta = Z_m$ to account for asymmetrical root water uptake with depth, and also to allow for a maximum root water uptake
A TYPICAL PARAMETER ESTIMATION PROBLEM

rate at any depth, $Z_0$. The asymmetry in root water uptake with soil depth is determined by the ratio between $p_z$ for $z \leq z^*$ and the $p_z$-value for $z > z^*$. To reduce the number of parameters, $p_z$ is set to unity for values of $z > z^*$, whereas it is a fitted value for $z \leq z^*$. The value of $Z_0$, the depth of maximum uptake, can simply be calculated from the first derivative, or $\frac{d\beta(z)}{dz} \bigg|_{Z_0} = 0$.

Assuming axial symmetry in root water uptake, while using the same root water uptake model, leads to the following two-dimensional root water uptake model:

$$\beta(r, z) = \left(\frac{1 - \frac{r}{R}}{R}\right) \left(1 - \frac{z}{Z} \right) \exp \left[ - \left( \frac{p_z}{Z} |z^* - z| + \frac{p_z}{R} |r^* - r| \right) \right]$$  \hspace{1cm} (2.3)

where $R$ (L) is the maximum rooting length in the radial direction, $r$ (L) is the radial distance from the origin of the tree, and $p_z$ (-) and $r^*$ (L) are additional empirical parameters. Here $\beta(r, z)$ (-) denotes the two-dimensional spatial distribution of potential root water uptake. Vrugt et al. [2001b] showed that the root water uptake in Eq. (2.3) is extremely flexible and allows spatial variations of water uptake as influenced by non-uniform water application (e.g. drip irrigation) and root length density patterns.

Including an additional exponential term in Eq. (2.3) leads to a three-dimensional root water uptake model, which is expressed in Eq. (2.4) as:

$$\beta(x, y, z) = \left(1 - \frac{x}{X}\right) \left(1 - \frac{y}{Y}\right) \left(1 - \frac{z}{Z} \right) \exp \left[ - \left( \frac{p_x}{X} |x^* - x| + \frac{p_y}{Y} |y^* - y| + \frac{p_z}{Z} |z^* - z| \right) \right]$$  \hspace{1cm} (2.4)

where $X$ (L) and $Y$ (L) are the maximum rooting length in the $x$- and $y$-direction, $x$ (L) and $y$ (L) are the distances from the origin of the tree in the $x$- and $y$-direction, $p_x$ (-), $p_y$ (-), $x^*$ (L) and $y^*$ (L) are empirical parameters and $\beta(x, y, z)$ denotes the unitless three-dimensional spatial distribution of potential root water uptake. As in Eq. (2.1) we set $p_x$, $p_y$, and $p_z$ to unity for $z > z^*$, $x > x^*$ and $y > y^*$, respectively.

Denoting the normalized root water uptake, $S_n$ (L$^3$ T$^{-1}$), as the volume of water extracted per unit volume of soil with time, it follows that:
CHAPTER 2

\[
S_m(\zeta) = \frac{\beta(\zeta) T_{pot}}{\int_0^Z \beta(\zeta) d\zeta} \quad 1\text{-D description} \quad (2.5a)
\]

\[
S_m(r, \zeta) = \frac{\pi R^2 \beta(r, \zeta) T_{pot}}{2\pi \int_0^R \int_0^\zeta r \beta(r, \zeta) dr d\zeta} \quad 2\text{-D axial symmetry} \quad (2.5b)
\]

\[
S_m(x, y, z) = \frac{X^2 Y^2 \beta(x, y, z) T_{pot}}{\int_0^x \int_0^y \int_0^z \beta(x, y, z) dx dy dz} \quad 3\text{-D description} \quad (2.5c)
\]

So integration of any of the above expressions over the spatial domain leads to the result that cumulative potential root water uptake is equal to the potential transpiration, \( T_{pot} \).

To provide for root-water uptake under water-stressed conditions, a soil water stress response function was included [van Genuchten, 1987]:

\[
\gamma(h) = \frac{1}{\left( 1 + \left( \frac{h(x, y, z, t)}{b_{so}} \right)^p \right)} \quad (2.6)
\]

where \( h \) (L) is the soil water matric head at a particular spatial location, \( b_{so} \) (L) is the soil water pressure head at which root water uptake rate is reduced by 50\%, and \( p \) is a unitless fitting parameter. The parameter \( p \) is usually assumed to be 3 [van Genuchten and Gupta, 1993]. Finally, the actual root water uptake rate at any particular spatial location can be calculated from:

\[
S(h, x, y, z) = \gamma(h) S_m(x, y, z) \quad (2.7)
\]

where for an almond tree:

\[
T_{pot} = ET_{almond} - E_i \quad (2.8)
\]

where \( S(h, x, y, z) \) is the actual root water uptake and \( E_i \) (LT\(^{-1}\)) denotes soil evaporation. \( ET_{almond} \) defines the potential \( ET \) by the almond crop, and is computed from the product of \( K_i \).
and $ET_o$ where $K_c(\cdot)$ is the crop coefficient and $ET_o(\text{LT}^+)\text{ is the reference evapotranspiration.}$ Hence, the actual transpiration rate ($T_\alpha$) can be computed from:

$$T_\alpha = \int_0^Z S(b, \zeta) d\zeta$$  

one-dimensional (1-D) \hspace{1cm} (2.9a)

$$T_\alpha = \frac{2\pi}{\pi R^2} \int_0^Z \int_0^{R_z} S(b, r, \zeta) dr d\zeta$$  

axial-symmetrical (2-D) \hspace{1cm} (2.9b)

$$T_\alpha = \frac{1}{X_n Y_m} \int_0^X \int_0^Y \int_0^Z S(b, x, y, \zeta) dx dy d\zeta$$  

three-dimensional (3-D) \hspace{1cm} (2.9c)

### 2.2.2. Field description and measurements

The experimental plot of our almond orchard includes about one quarter of the wetted area of a micro-sprinkler irrigating a single almond tree [Koumanov et al., 1997; Vrugt et al., 2001b]. Hence, despite the spatial variability in irrigation amounts by the micro-sprinkler, we assumed that this instrumented area of 2.0 x 2.0 m (Figure 2.1) was representative for the almonds water. Twenty-five PVC neutron probe access tubes were installed in a square grid of 0.50 m spacing to a depth of 90 cm.

![Figure 2.1](image-url)  
**Figure 2.1.** Schematic overview of the experimental plot. Location of the neutron probe access tubes is indicated with the symbol "•". Circles approximate neutron probe measurement volumes, whereas curved lines indicate the averaging volumes for the two-dimensional simulations.
The neutron probe was calibrated from gravimetric measurements using soil samples collected at soil depths of 15, 30, 45, 60, 75 and 90 cm during and after access tube installation. Separate calibration curves were used for the 0 - 15 cm surface soil and the 30 – 90 cm soil depth interval. Standard errors of the estimate of volumetric water content curves were approximately 0.01 $m^3/m^3$ at the 15 cm depth interval and 0.02 $m^3/m^3$ at all other measurements depths. The field is slightly undulating, and the soil is a shallow gravelly loam [Andreu et al., 1997], overlaying a sloping high-density restricting clay layer at about the 90 - 120 cm soil depth. The studies by Andreu et al. [1997] and Kaufmanov et al. [1997] indicated that root water uptake during the growing season was mainly limited to the top 40-60 cm, and that drainage at the 90 cm depth occurred primarily by lateral flow along the sloping restricting clay layer.

The measurements were conducted during the September 13-29 period in the summer of 1995, after the micro-sprinkler system was used to wet up the whole soil profile. Neutron probe measurements were taken on September 13, immediately after the irrigation at 13:00, 15:00 and 18:00 hours, during the period September 14 through September 17, every 4 hours at 6:00, 10:00, 14:00, and 18:00 hours, and during the period September 18 through September 29, daily at approximately 10:00 hours. This resulted in 31 water content measurements at each spatial location. To test the one- and two-dimensional root water uptake model, the three dimensional local measurements of water content were reduced to one- and two dimensions, respectively. For one-dimension, all the water content measurements at a specific depth for all $x$- and $y$-locations were arithmetically averaged.

To simplify testing of the two-dimensional root water uptake model, the three dimensional grid measurements of water content needed to be reduced to two dimensions ($r$ and $z$). For this we assumed that (1) the root water uptake around the tree was axi-symmetrical and (2) that the measurement volume of the neutron probe water content measurements was a sphere with a constant radius of approximately 0.25 m.

For each depth interval the rectangular measurement grid of Fig. 2.1 was partitioned into five concentric 0.6 m wide circular strips with their origin determined by the neutron access pipe location closest to the tree trunk. Second, a radial average water content value was computed for each of the 5 soil areas (0.2-0.8 m, 0.8-1.4 m, 1.4-2.0 m, 2.0-2.6 m and 2.6-3.2 m concentric circles) using weighting factors for each neutron probe location with values equal to the fraction of the measurement volume fitting within the respective concentric soil area. We used 0.6 m wide strips for each of the five soil areas to ensure that enough water content measurements were contained within the respective strip. Moreover, the averaging using the 0.6 m wide strips gave the best agreement in total water depletion of the reduced two-dimensional domain as
compared to the original three-dimensional grid of water content measurements. Since the averaging procedure was applied to depth intervals of 0-0.15 m, 0.15-0.3 m, 0.3-0.45 m and 0.45-0.6 m during the September 13-29 period, the final two-dimensional map included 20 average water content values at each measurement time, namely 4 depth intervals and 5 radial distance increments.

In Figure 2.2 we present the measured three-dimensional volumetric water content distributions at three different times during the monitoring period, as illustrated by depth interval-averaged water content values.

![Figure 2.2](image)

**Figure 2.2.** Measured three-dimensional volumetric water content distributions at three different times during the monitoring period.

At the beginning of the period, volumetric water content values are about $0.24 \text{ m}^3 \text{ m}^{-3}$ at the soil surface and increase downward to about $0.28 \text{ m}^3 \text{ m}^{-3}$. Clearly, as time proceeds, the water content at the 60 cm depth is still relatively high, whereas a water uptake pattern is becoming apparent in the surface layers, with volumetric water content values ranging between $0.04$ and $0.08 \text{ m}^3 \text{ m}^{-3}$ in the top soil layers.

### 2.3. Soil water flow modeling

The adapted models that solved for the one-, two-, and three-dimensional solution of Eq. (2.1) were HYDRUS-1D [Simunek et al., 1998b], HYDRUS-2D [Simunek et al., 1999], and HYDRUS-3D (update of the SWMS_3D code of Simunek et al. [1995]). All three models use the Galerkin finite element method based on the mass conservative iterative scheme proposed by Celia et al. [1990].
CHAPTER 2

The unsaturated hydraulic properties for all three models are defined by [van Genuchten, 1980; Mualem, 1976]:

\[
\theta_s = \frac{\theta(a) - \theta_r}{\theta_s - \theta_r} = \left[ 1 + \left( \frac{\theta_r}{\theta_s} \right)^n \right]^{-m} ; \quad m = \frac{n-1}{n} \tag{2.10}
\]

and

\[
K(\theta) = K_s \left[ 1 - \left( 1 - \frac{\theta}{\theta_s} \right)^{-\gamma} \right]^2 \tag{2.11}
\]

where \( \theta_s \) is the saturated water content (L^3 L^-1), \( \theta_r \) is the residual water content (L^3 L^-1), \( \alpha \) (L^-1) and \( n \) (-) are curve shape parameters, and \( K_s \) \( (LT^-1) \) denotes the saturated hydraulic conductivity.

In the pre-processing phase, the soil domain was discretized into a rectangular grid of finite elements with similar size elements defining the spatial resolution, to avoid differences in truncation errors between the three simulation models. For all three simulation models, the vertical domain was 0.6 m deep. In HYDRUS-1D the vertical domain was discretized into 40 equidistant finite elements. For the two-dimensional model, the simulated flow domain was 3 m long in the radial direction, using a grid spacing of 0.05 m in the radial and 0.015 m in the vertical direction. For the three-dimensional simulations, the dimension of the soil domain was 0.15 by 2.65 m in the \( x \)-\( y \)-plane. The flow domain was discretized into a structured mesh of 13,500 blocks corresponding with 14,400 nodes. For the \( x \) and \( y \)-direction, three sequences of nodal distances of 4 x 0.10, 1 x 0.09 and 1 x 0.01 and one sequence of 5 x 0.10 was used, whereas for the \( x \)-direction three sequences of 2 x 0.05, 1 x 0.04 and 1 x 0.01 and one sequence of 3 x 0.05 was used. This structured mesh was used to avoid large initialization errors in water content and was a compromise between available computing time and nodal density.

2.3.1. Boundary and initial conditions

Figure 2.3 presents the daily estimated boundary conditions as function of time during the monitoring period. Numerical modeling requires estimates of potential transpiration (\( T_{pd} \)) and soil evaporation (\( E_s \)). As no direct measurement of the transpiration of the almond tree was available, we used the following approach. Daily reference evapotranspiration (\( ET_0 \)) were provided by a nearby weather station of the California Irrigation Management Information System (CIMIS).
Almond potential $ET_{almond}$ was calculated from $ET_0$ and the appropriate crop coefficients ($K$). Snyder and Pruitt [1988] recommended a value for $K_c$ of 0.91, corresponding to conditions of 60% canopy soil surface coverage for drip-irrigated trees in the Sacramento Valley. Ritchie's [1972] equation was used to estimate soil evaporation. The radiation interception was calculated using the empirical function for maize [Snyder et al., 1985], while we used an upper limit of stage 1 cumulative soil water evaporation of 6 mm, and a partitioning factor of 0.4 between stage 1 and stage 2 evaporation [Ritchie, 1972]. The potential transpiration of almond trees ($T_{pol}$) was obtained by subtracting the soil evaporation from $ET_{almond}$ (Eq. (2.8)). Although we agree that this approach for calculating $T_{pol}$ is quite simplistic, it is to be expected that errors in estimated daily $T_{pol}$ amounts are smaller than 10 percent. We should also note that the focus of this study was to demonstrate the development and functionality of spatially-distributed root water uptake models.

Due to the lack of accurate flux information, we assumed a unit hydraulic gradient at the lower boundary (gravity flow). This approach seemed most appropriate, since water balance calculations using the estimated ET and measured infiltration data indicated that a drainage term was required to match the measured soil water storage data change. The water content measurements immediately after the irrigation of September 13 were used as initial condition for all numerical simulations.
2.4. Parameter optimization by inverse modeling

In the inverse modeling stage of this study, a total of six (e.g. $Z_p$, $P_z$, $n$, $K_t$ and $h_0$), nine (e.g. $Z_m$, $P_z$, $n$, $K_t$ and $h_0$), and twelve (e.g. $X_m$, $P_z$, $n$, $K_t$ and $h_0$) root water uptake and soil hydraulic parameters were identified simultaneously using the one-, two-, and three-dimensional HYDRUS models, respectively. Despite measurement of the soil hydraulic properties of a nearby location in the same almond orchard [Andreu et al., 1997], the soil heterogeneity within the orchard led us to also optimize some of the soil hydraulic parameters, simultaneously with the respective root water uptake model parameters. While fixing the parameters $\Theta$ and $\alpha$ to reported values [Andreu et al., 1997] of 0.30 m$^3$ m$^{-3}$ and 9.4 m$^{-1}$, respectively, the soil hydraulic properties were assumed to be characterized by the fitting parameters $n$ and $K_t$ of Eqs. (2.10) and (2.11). Since some of the measured water content values were smaller than the residual water content values reported by Andreu et al. [1997], the residual water content ($\Theta$) was fixed to 0.0 m$^3$ m$^{-3}$.

Since optimization algorithms such as Levenberg-Marquardt or Simplex method are generally only applicable to identify a limited number of parameters, an alternative was needed to optimize the larger set of parameters of this study. Recently, it has been shown that Genetic Algorithms are a powerful tool for parameter identification, if the number of fitted parameters is large [Back, 1996; Wang, 1991; Holland, 1975]. The Genetic Algorithm (GA) is a search procedure based on the mechanics of natural selection and natural genetics that combines an artificial survival of the fittest with genetic operators [Holland, 1975]. The GA differs from other search methods as it searches among a population of parameter sets rather than the parameter values themselves using probabilistic transition rules. We applied the GA presented in Penny and Undfield [1995], with the small adaptation that the best performing parameter combination is not mutated in the next generation. We used a crossover percentage of 85 to ensure a relatively fast convergence to the global optimum, whereas a mutation factor of 0.15 was used to avoid optimized solutions in local minima. The population size, representing the number of first generation parameter combinations was set to 120, whereas the final optimized parameter combination was selected after 200 generations. Assuming that the residuals, representing errors between measured and optimized volumetric water content values, are Gaussian distributed, independent and homoscedastic (constant variance), the fitness of a chromosome was calculated by the following objective function, $OF$: 

34
A typical parameter estimation problem

\[ OF(b) = \sum_{j=1}^{N} \left[ \theta'(t_j) - \theta(t_j, b) \right]^2 \]  \hspace{1cm} (2.12)

where \( N \) is the number of observations, and \( \theta'(t) \) and \( \theta(t, b) \) denote the measured and predicted water content values, respectively, at time \( t \). The parameter vector, \( b \), characterizes the chromosome with the genes representing the fitting parameters. The allowable ranges of the parameters included in \( b \) for each numerical model are presented in Table 2.1.

### Table 2.1. Minimum (Min) and maximum (Max) values used in the parameter optimization with the genetic algorithm (GA) for the HYDRUS-1D, HYDRUS-2D, and HYDRUS-3D models.

<table>
<thead>
<tr>
<th>Par.</th>
<th>Unit</th>
<th>HYDRUS-1D Min.</th>
<th>HYDRUS-1D Max.</th>
<th>HYDRUS-2D Min.</th>
<th>HYDRUS-2D Max.</th>
<th>HYDRUS-3D Min.</th>
<th>HYDRUS-3D Max.</th>
</tr>
</thead>
<tbody>
<tr>
<td>( X_a )</td>
<td>[m]</td>
<td>0.00</td>
<td>1.00</td>
<td>1.00</td>
<td>5.00</td>
<td>1.00</td>
<td>5.00</td>
</tr>
<tr>
<td>( Y_a )</td>
<td>[m]</td>
<td>0.00</td>
<td>1.00</td>
<td>1.00</td>
<td>5.00</td>
<td>1.00</td>
<td>5.00</td>
</tr>
<tr>
<td>( Z_a )</td>
<td>[m]</td>
<td>0.10</td>
<td>15.00</td>
<td>0.10</td>
<td>15.00</td>
<td>0.10</td>
<td>15.00</td>
</tr>
<tr>
<td>( R_a )</td>
<td>[m]</td>
<td>1.00</td>
<td>5.00</td>
<td>1.00</td>
<td>5.00</td>
<td>1.00</td>
<td>5.00</td>
</tr>
<tr>
<td>( p_w )</td>
<td>[-]</td>
<td>0.00</td>
<td>1.00</td>
<td>0.00</td>
<td>1.00</td>
<td>0.00</td>
<td>1.00</td>
</tr>
<tr>
<td>( p_t )</td>
<td>[-]</td>
<td>0.00</td>
<td>1.00</td>
<td>0.00</td>
<td>1.00</td>
<td>0.00</td>
<td>1.00</td>
</tr>
<tr>
<td>( p_e )</td>
<td>[-]</td>
<td>0.00</td>
<td>1.00</td>
<td>0.00</td>
<td>1.00</td>
<td>0.00</td>
<td>1.00</td>
</tr>
<tr>
<td>( \alpha )</td>
<td>[m]</td>
<td>0.00</td>
<td>1.00</td>
<td>0.00</td>
<td>1.00</td>
<td>0.00</td>
<td>1.00</td>
</tr>
<tr>
<td>( \beta )</td>
<td>[m]</td>
<td>0.00</td>
<td>1.00</td>
<td>0.00</td>
<td>1.00</td>
<td>0.00</td>
<td>1.00</td>
</tr>
<tr>
<td>( \gamma )</td>
<td>[m]</td>
<td>0.00</td>
<td>1.00</td>
<td>0.00</td>
<td>1.00</td>
<td>0.00</td>
<td>1.00</td>
</tr>
<tr>
<td>( \delta )</td>
<td>[m]</td>
<td>0.00</td>
<td>1.00</td>
<td>0.00</td>
<td>1.00</td>
<td>0.00</td>
<td>1.00</td>
</tr>
<tr>
<td>( \zeta )</td>
<td>[m]</td>
<td>0.00</td>
<td>1.00</td>
<td>0.00</td>
<td>1.00</td>
<td>0.00</td>
<td>1.00</td>
</tr>
<tr>
<td>( \xi )</td>
<td>[m]</td>
<td>0.00</td>
<td>1.00</td>
<td>0.00</td>
<td>1.00</td>
<td>0.00</td>
<td>1.00</td>
</tr>
<tr>
<td>( \eta )</td>
<td>[-]</td>
<td>1.00</td>
<td>5.00</td>
<td>1.00</td>
<td>5.00</td>
<td>1.00</td>
<td>5.00</td>
</tr>
<tr>
<td>( K_s )</td>
<td>[cm d(^{-1})]</td>
<td>1.0 \times 10^{-2}</td>
<td>100.00</td>
<td>1.0 \times 10^{-2}</td>
<td>100.00</td>
<td>1.0 \times 10^{-2}</td>
<td>100.00</td>
</tr>
<tr>
<td>( b_{lu} )</td>
<td>[m]</td>
<td>-0.20</td>
<td>-10.00</td>
<td>-0.20</td>
<td>-10.00</td>
<td>-0.20</td>
<td>-10.00</td>
</tr>
</tbody>
</table>

Although GA's are an effective means of reaching the global minimum region, they are not necessarily efficient in finding the exact optimum location. Therefore, the results of the genetic algorithm were used as initial values for a subsequent SA to determine the local minimum of \( OF \) within the global minimum region as determined by the GA. Using a sensitivity analysis in which each parameter was varied with 10% of its final optimized value, while keeping the additional parameters fixed at their value determined by the GA, only those six parameters were fine-tuned that were most sensitive to model output. Both the GA and Simplex optimization were carried out using MATLAB, version 5.3 [The Mathworks, 1999].

The uncertainty of each optimized parameter, \( b_j, j = 1, \ldots, m \), was determined from the diagonal elements of the parameter covariance matrix \( C \) [Kool and Parker, 1988; Šimůnek and Hopmans, 2000], representing the estimate of the standard deviation, \( s_p \).
where \( s_j = \sqrt{c_j} \) \hspace{1cm} (2.13)

whereas model performance was evaluated by the Root Mean Square Error (RMSE), computed from,

\[
RMSE = \sqrt{\frac{\sum_{i=1}^{N} (\theta^*(t_i) - \theta(t_i, b))^2}{N - m}}
\] \hspace{1cm} (2.14)

where \( N \) and \( m \) denote the number of measurements and total number of parameters, respectively.

A single forward simulation of the HYDRUS-3D model requires between 5 and 60 minutes on a PIII-466 MHz computer, depending on the parameter combination provided. So the computational time for one computer to perform all 24,000 model runs for the GA optimization iterations was extremely long. Instead, we used 40 PIII-400 MHz slave computers, connected with one master computer to perform the optimizations.

### 2.5. Results and discussion

#### 2.5.1. Three-dimensional simulations

The parameter vector of the best performing “chromosome” of the final population after 200 generations of the three-dimensional root water uptake parameters using GA in combination with the HYDRUS-3D flow model is presented in Table 2.2. Also included are the final results after final tuning of the selected parameters using the SA with their confidence intervals, values for the derived parameters \( X_0, Y_0, \) and \( Z_0 \), as well as RMSE and \( R^2 \)-values. The optimum maximum rooting depth \( (Z_* \text{ m}) \) \((0.43 < Z_* < 0.45 \text{ m})\) is in excellent agreement with the results obtained by Koumanov et al. [1997] for the same experimental plot, confirming that active root-water uptake was limited to the top 40 cm only. The position of maximum root water uptake of the almond tree under non-stressed conditions \((X_0 = 1.53, Y_0 = 1.61, Z_0 = 0.27 \text{ m})\) agrees well with the surface area of maximum irrigation application by micro-sprinkling [Koumanov et al., 1997]. This was so, despite the location of the micro-sprinkler at the far corner along the tree row (see Fig. 2.1), as caused by non-uniform water applications during the growing season.
Table 2.2. Optimized parameter values and their 95% confidence intervals after optimization with the Genetic (GA) and Simplex algorithm (SA) for the HYDRUS-3D model.

<table>
<thead>
<tr>
<th>Par.</th>
<th>Unit</th>
<th>HYDRUS-3D</th>
<th>95% Confidence interval</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>GA</td>
<td>SA</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Root model</td>
<td></td>
</tr>
<tr>
<td>X*</td>
<td>[m]</td>
<td>3.31</td>
<td>†</td>
</tr>
<tr>
<td>Y*</td>
<td>[m]</td>
<td>2.65</td>
<td>2.60</td>
</tr>
<tr>
<td>Z*</td>
<td>[m]</td>
<td>0.43</td>
<td>0.45</td>
</tr>
<tr>
<td>f</td>
<td>[-]</td>
<td>1.86</td>
<td>†</td>
</tr>
<tr>
<td>p</td>
<td>[-]</td>
<td>2.62</td>
<td>†</td>
</tr>
<tr>
<td>s</td>
<td>[-]</td>
<td>2.57</td>
<td>†</td>
</tr>
<tr>
<td>X *=</td>
<td>[m]</td>
<td>1.92</td>
<td>†</td>
</tr>
<tr>
<td>Y *=</td>
<td>[m]</td>
<td>1.92</td>
<td>2.21</td>
</tr>
<tr>
<td>Z *=</td>
<td>[m]</td>
<td>0.35</td>
<td>0.40</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Soil hydraulic model</td>
<td></td>
</tr>
<tr>
<td>n</td>
<td>[-]</td>
<td>1.72</td>
<td>1.74</td>
</tr>
<tr>
<td>k</td>
<td>[cm d⁻¹]</td>
<td>1.82</td>
<td>1.60</td>
</tr>
<tr>
<td>b</td>
<td>[m]</td>
<td>-0.85</td>
<td>†</td>
</tr>
<tr>
<td>RMSE</td>
<td>[m³ m⁻³]</td>
<td>0.0183</td>
<td>0.0180</td>
</tr>
<tr>
<td>R²</td>
<td></td>
<td>0.91</td>
<td>0.92</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Derived parameters</td>
<td></td>
</tr>
<tr>
<td>X'</td>
<td>[m]</td>
<td>1.53</td>
<td>1.53</td>
</tr>
<tr>
<td>Y'</td>
<td>[m]</td>
<td>1.64</td>
<td>1.61</td>
</tr>
<tr>
<td>Z'</td>
<td>[m]</td>
<td>0.26</td>
<td>0.27</td>
</tr>
</tbody>
</table>

† Open space means parameter held constant to value found by genetic algorithm.

The optimized n-value of the soil hydraulic functions (n = 1.74) agrees well with the reported range of n-values (n = 1.44 – 1.99) for this soil obtained with the instantaneous profile method [Andreu et al., 1997]. The relatively high value of the bₚₚ-parameter (bₚₚ = -0.85 m) is an indication of the small water holding capacity of this coarse-textured soil. The optimized saturated conductivity (Kₛ = 1.60 cm d⁻¹) is lower than the reported range of 34.1 - 62.4 cm d⁻¹ measured under saturated conditions for the 0 - 60 cm depth interval [Andreu et al., 1997]. The lower optimized Kₛ-value is to be expected, since the measured experimental conditions were such that the soil-rooting zone was less than saturated, thereby eliminating the influence of the macropores on the estimated Kₛ. Using the same data set for their two-dimensional analysis, Vrugt et al. [2001b] showed that the HYDRUS-2D model was well able to predict water content dynamics for spatial locations, not included in the calibration period.

The measured water content values for all depths and measurement locations are correlated with simulated water content values using the final optimized parameter values in Figure 2.4. Measured values match the simulated values with an R²-value of 0.92. The overall RMSE-value of 0.018 m³ m⁻³ is low, considering that the standard errors of the neutron probe...
water content measurements is already 0.01 m³ m⁻³ (15 cm depth) and 0.02 m³ m⁻³ (all other soil depths).

![Figure 2.4](image.png)

**Figure 2.4.** Measured versus simulated soil water contents around the almond tree obtained using the calibrated HYDRUS-3D model.

Figure 2.5 presents three-dimensional maps of simulated water content and root water uptake intensity, $S_r(x, y, z)$, averaged for the indicated soil compartments (0-0.15 m, 0.15-0.30 m, 0.30-0.45 m, and 0.45-0.60 m) at the 3 different times of Sept. 18, 23, and 29. These maps were obtained from arithmetic averaging over all nodal values within each depth interval followed by interpolation using SURFER [Golden Software, 1996]. At the beginning of the period, after irrigation on Sept. 13, maximum actual water uptake rates approached $8 \times 10^{-4}$ m³ m⁻³ h⁻¹. As the soil becomes depleted in water, regions of maximum root water uptake shifted to other locations within the rooting zone where soil water was most readily available [Green and Clothier, 1999]. For example, close observation of Fig. 2.5 shows that the general root water uptake pattern changes with time from maximum uptake around $(X_0, Y_0, Z_0)$ towards the outside perimeter of the rooting volume, as caused by changes in soil water stress with time. Although there are differences in the spatial pattern of soil water content between simulated and measured water content values (compare Figs. 2.2 and 2.5), the magnitude of simulated water contents at the different depth intervals agreed well with the measured water contents.
A TYPICAL PARAMETER ESTIMATION PROBLEM

Figure 2.5. Simulated three-dimensional volumetric water content and potential root water uptake distributions at three times during the monitoring period.

In Figure 2.6a we present contour plots of the time averaged Root Mean Square Error (RMSE) of water content at the four depth intervals. These contour plots were computed using arithmetic averaging of all nodal values within the respective soil volume of the corresponding neutron probe measurement and thus resulted in twenty-five RMSE values for each depth interval.

Figure 2.6. a) Three-dimensional spatial distribution of Root Mean Square Errors of water content, b) Box-plot of RMSE for all measured locations and times.
CHAPTER 2

Although in general RMSE-values are small, relatively large error values are present at the bottom corner in the spatial domain near \( x = 2.4 \) m and \( y = 0.4 \) m. Differences between measured and simulated water content values are likely, because of model errors as caused by restrictive assumptions regarding the geometry of the rooting system, homogeneity of soil hydraulic properties within the spatial domain, and the prescribed root water uptake model. For example, the water-uptake model assumes a single region of maximum uptake, whereas in reality more regions within the rooting zone may show local maximum uptake as caused by water application non-uniformities and soil environmental factors affecting root growth. Figure 2.6b presents a Box-plot of the time averaged RMSE values for all 100 spatial locations (25 tubes * 4 depth intervals). The Box-plot shows the single outlier with RMSE = 0.04 at \( 1.9 < x < 2.4 \) m in the bottom right hand corner of the spatial domain as well as a clear clustering of RMSE values between 0.01 and 0.02 m\(^3\) m\(^{-3}\), with their magnitude about equal to the standard error of the water content measurements with the neutron probe. The various horizontal lines represent the minimum, maximum, 25% percentile, mean and 75% percentile of the RMSE values, respectively.

2.5.2. Dimensional effects on parameter optimization results

Final optimized parameter values after using the GA and fine-tuning with the SA for the HYDRUS-1D, HYDRUS-2D and HYDRUS-3D models and their 95% confidence intervals are presented in Table 2.3. Also included are the derived parameter values of \( X_0, Y_0 \) and \( Z_m \), and the fitting results as expressed by the RMSE and R\(^2\)-values. Since many errors may occur, including measurement, model and numerical errors, an uncertainty analysis of the optimized parameters makes up an important part of parameter estimation. Therefore, we included 95% confidence intervals for the optimized parameters, calculated using the Jacobian matrices and residuals for the final optimized solution.

As Table 2.3 shows, optimized parameter values for the maximum rooting depth, \( Z_m \) (0.41 to 0.49 m), \( n \) (1.74 to 1.91), \( b_3 \) (-0.53 to -0.85 m) and \( Z_o \) (0.27 to 0.28 m) are close between the three dimensional cases. It indicates that the information content of the water content measurements, whether aggregated or not, is the most robust for these parameters [Vrugt et al., 2001a]. To obtain convergence of the inverse solution it was essential to select initial values for the other parameters, close to their expected values [Inoue et al., 1998]. Problems with non-uniqueness of these parameters are caused by the presence of numerous local minima, as can occur when this many parameters are optimized simultaneously [Duan et al., 1992].
Table 2.3. Optimized parameter values and their 95\% confidence regions after Simplex algorithm (SA) for the HYDRUS-1D and -2D models. (For completeness we also report the HYDRUS-3D model results.)

<table>
<thead>
<tr>
<th>Par.</th>
<th>Unit</th>
<th>HYDRUS-1D</th>
<th>Optimized</th>
<th>CV</th>
<th>HYDRUS-2D</th>
<th>Optimized</th>
<th>CV</th>
<th>HYDRUS-3D</th>
<th>Optimized</th>
<th>CV</th>
</tr>
</thead>
<tbody>
<tr>
<td>X_0</td>
<td>[m]</td>
<td>3.31 ± 0.46</td>
<td>7.10</td>
<td></td>
<td>2.60 ± 1.18</td>
<td>3.46</td>
<td></td>
<td>0.49 ± 0.045</td>
<td>4.21</td>
<td></td>
</tr>
<tr>
<td>Y_0</td>
<td>[m]</td>
<td>0.41 ± 0.49</td>
<td>10.76</td>
<td></td>
<td>3.99 ± 0.49</td>
<td>6.22</td>
<td></td>
<td>0.45 ± 0.13</td>
<td>13.33</td>
<td></td>
</tr>
<tr>
<td>Z_m</td>
<td>[m]</td>
<td>1.86 ± 0.20</td>
<td>5.38</td>
<td></td>
<td>2.62 ± 0.15</td>
<td>2.86</td>
<td></td>
<td>2.98 ± 0.74</td>
<td>12.49</td>
<td></td>
</tr>
<tr>
<td>p_c</td>
<td>[-]</td>
<td>1.92 ± 0.30</td>
<td>7.81</td>
<td></td>
<td>2.21 ± 0.21</td>
<td>4.75</td>
<td></td>
<td>0.27 ± 0.04</td>
<td>7.09</td>
<td></td>
</tr>
<tr>
<td>p_s</td>
<td>[-]</td>
<td>1.96 ± 0.11</td>
<td>13.56</td>
<td></td>
<td>1.75 ± 0.11</td>
<td>3.26</td>
<td></td>
<td>1.74 ± 0.04</td>
<td>1.15</td>
<td></td>
</tr>
<tr>
<td>n</td>
<td>[-]</td>
<td>0.75 ± 0.16</td>
<td>10.93</td>
<td></td>
<td>0.75 ± 0.16</td>
<td>10.93</td>
<td></td>
<td>0.75 ± 0.16</td>
<td>10.93</td>
<td></td>
</tr>
<tr>
<td>K_h</td>
<td>[cm d^{-1}]</td>
<td>-0.53 ± 0.16</td>
<td>15.10</td>
<td></td>
<td>-0.53 ± 0.16</td>
<td>15.10</td>
<td></td>
<td>-0.53 ± 0.16</td>
<td>15.10</td>
<td></td>
</tr>
<tr>
<td>b_{so}</td>
<td>[m]</td>
<td>0.0068</td>
<td>0.0154</td>
<td>0.0180</td>
<td>0.98</td>
<td>0.91</td>
<td>0.92</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RMSE</td>
<td>[m^3 m^-3]</td>
<td>1.93</td>
<td>1.53</td>
<td>1.61</td>
<td>0.27</td>
<td>0.28</td>
<td>0.27</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Whereas all our optimization results for the numerical models indicate the depth of maximum root water uptake to be at around 0.25 m, the study by Andreu et al. [1997] concluded that maximum uptake occurred at the soil surface (0-15 cm) and decreased further down the soil profile. However, their study did not include soil evaporation as a possible mechanism of soil water depletion near the soil surface. The coefficient of variation (CV) indicates that for most parameters the confidence intervals at the minimum are typically small (CV<10\%).

The optimized soil water retention and unsaturated soil hydraulic functions of the HYDRUS-1D, HYDRUS-2D and HYDRUS-3D model are presented in Figure 2.7. Also included are the measured \((\theta, h)\) data using the multi-step outflow method from soil cores taken at the 30 cm soil depth for a nearby location.
Both the measured ($\theta$, $\theta$)-points and the optimized retention curves in Fig 2.7a clearly show the small water holding capacity of this shallow gravely soil. Whereas the optimized retention functions of the HYDRUS-2D and HYDRUS-3D model match extremely well, the HYDRUS-1D optimization shifts the water retention to an even more coarser-textured soil, increasing drainage as compared to the multi-dimensional flow models. Differences between independently measured ($\theta$, $\theta$)-points and the optimized curves mostly occur outside the range of experimental water content values ($\theta < 0.08$) and are most likely caused by the sensitivity of the optimized retention curve to the fixed residual water content value ($\theta = 0$). Clearly, parameters obtained with parameter estimation are to be used only within the measurement range for which they were determined [Irioue et al., 2000; Vrugt et al., 2001a]. Additionally, the optimized unsaturated hydraulic conductivity and measured ($\theta$, $\theta$) points (Fig. 2.7b) show the rapid decrease of the hydraulic conductivity with decreasing water content. The optimized saturated conductivity of 0.46 cm d$^{-1}$ of the SA optimization is lower than the reported range of 34.1 - 62.4 cm d$^{-1}$ measured under saturated conditions for the 0 - 60 cm depth interval by Andreu et al. [1997]. However, one should realize that the saturated hydraulic conductivity in this study is more a water balance parameter, controlling the magnitude of the lower boundary flux, than it is a soil physical parameter affecting soil water flow in the soil domain. Moreover, the saturated conductivity determined by Andreu et al. [1997] was measured under saturated conditions, when macro pores play a major role.

Although not presented, the parameter correlation matrix for the one, two- and three dimensional models using the off-diagonal terms of the covariance matrix showed that
correlations were typically low for the three-dimensional optimizations, but that parameter correlations increased when decreasing spatial dimensions (2-D and 1-D). For example, the correlation between the parameters $n$ and $b_{50}$ was high for the 2D-model. Low parameter correlations are important as they increase the likelihood of uniqueness of the final solution. The decrease in parameter correlation with increasing flow dimensions is likely caused by the corresponding increase in number of observations used in the optimization, relative to the increasing number of fitting parameters. As shown in Figure 2.8, the simulated water contents with the final parameter estimates obtained with the SA compare favorably with the corresponding measured water contents for both the HYDRUS-1D and the HYDRUS-2D model.

![Figure 2.8](image1.png)

**Figure 2.8.** Measured versus simulated soil water contents around the almond tree after parameter optimization using (a) HYDRUS-1D and (b) HYDRUS-2D flow models.

The increase in RMSE with increasing model dimension is caused by the significantly increased number of water content observations (3-D versus 2-D and 1-D) included in the objective function (Eq. (2.12)) with increasing spatial dimension of the optimization problem. This is especially the case as the number of fitting parameters increased only slightly with spatial dimension. Moreover, the averaging of the water contents used in the one- and two-dimensional simulations reduced the general water content variability, thereby decreasing the final RMSE-values.

Figure 2.9 presents a comparison between the optimized spatial distributions of potential root water uptake over the spatial domain as obtained using the one-, two- and three-dimensional root water uptake and flow models.
CHAPTER 2

Figure 2.9. Comparison of optimized spatial distributions of potential root water uptake in one-, two-, and three-dimensions.

Starting with the three-dimensional model results, spatial values of the optimized potential root water uptake function, $\beta(x,y,z)$, were arithmetically averaged and subsequently normalized in the radial direction (Eq. (2.5b)) to obtain an average $S_m(r,z)$. Subsequently, a similar averaging procedure was carried out in the radial direction to show the average $S_m(z)$. Using the two-dimensional flow and root uptake model, spatially averaging and normalization (Eq. (2.5a)) of $S_m(r,z)$ resulted in another average $S_m(z)$. The close match between the three one-dimensional potential root water uptake distributions in Fig. 2.9 indicates that the suggested aggregation of measured three-dimensional water content measurements to arrive at two- and one-dimensional root water uptake models is valid.

In Table 2.4 we present the different water balance components with their spatial variations (standard deviations) as simulated by the HYDRUS-1D, HYDRUS-2D and HYDRUS-3D models. Spatial variability in drainage flux was computed from simulated flux density values at the 55 cm soil depth. The agreement between the various listed water balance components for the different numerical models is satisfactory. As different numerical models are used with differences in aggregation of water content values between spatial dimensions, perfect agreement is unlikely. Cumulative soil evaporation and drainage components are typically small compared to total cumulative root water uptake.
A TYPICAL PARAMETER ESTIMATION PROBLEM

Table 2.4. Components of the soil water balance with standard deviation in parentheses: cumulative potential almond transpiration ($T_p^*$), soil evaporation ($E_v$), actual root water uptake ($T_r$) and drainage, using the HYDRUS-1D, -2D, and -3D models.

<table>
<thead>
<tr>
<th>Component</th>
<th>Unit</th>
<th>HYDRUS-1D</th>
<th>HYDRUS-2D</th>
<th>HYDRUS-3D</th>
</tr>
</thead>
<tbody>
<tr>
<td>$T_p^*$</td>
<td>[mm]</td>
<td>48.70</td>
<td>48.70</td>
<td>48.70</td>
</tr>
<tr>
<td>$E_v$</td>
<td>[mm]</td>
<td>6.99</td>
<td>6.99</td>
<td>6.49</td>
</tr>
<tr>
<td>$T_r$</td>
<td>[mm]</td>
<td>42.04</td>
<td>36.53 (35.18)</td>
<td>39.45 (39.60)</td>
</tr>
<tr>
<td>Drainage</td>
<td>[mm]</td>
<td>7.78</td>
<td>8.62 (5.33)</td>
<td>13.06 (8.11)</td>
</tr>
<tr>
<td>ΔStorage</td>
<td>[mm]</td>
<td>56.81</td>
<td>52.18</td>
<td>59.00</td>
</tr>
</tbody>
</table>

Differences between actual and potential transpiration are caused by water stress (differences between $T_p^*$ and $T_r$). As soil water storage is used in the objective function of Eq. (2.12), its value must remain approximately equal between simulations. Consequently, changing drainage amounts compensate for differences in root water uptake between 1, 2, and 3-D simulations. Most importantly, the standard deviation results in Table 2.4 shows that the spatial variation in drainage rate and root water uptake decreases when reducing multi-dimensional soil water flow and root water uptake to decreasing spatial dimensions. This may have large implications for chemical transport in root zones, as drainage rates and corresponding chemical transport rates will vary according to root water uptake distribution.

Figure 2.10 presents a detailed two-dimensional contour plot of the spatial variability of cumulative flux density (mm) during the September 18-29 monitoring period at the 55 cm soil depth, as computed from the HYDRUS-3D model.

![Cumulative Flux density contour plot](image-url)

**Figure 2.10.** Two-dimensional contour plot of spatial variability in cumulative drainage at the 0.55 m soil depth during the monitoring period.
Cumulative net soil water flow is downward (positive values), except for a small portion of the rooting zone domain at \((x,y) = (1.4,2.5)\). Although the results in Fig. 2.10 are influenced by the choice of the lower boundary condition, it clearly demonstrates that the spatial variability of the drainage rate below the rooting zone is large, with values increasing as corresponding root water uptake values decrease. The increasing accurate spatial description of root water uptake and soil water flow with increasing spatial dimension is essential to improve model predictions of water fluxes and contaminant transport through the vadose zone. Moreover, total chemical load to the groundwater will depend on local concentration and fluxes, and their spatial variability. Specifically, although the average chemical load can be small, as computed from average flux and concentration values in one-dimensional simulations, the actual chemical load can be much larger. For example, this is the case if local regions of high drainage rates, as controlled by low root water uptake, correspond with high concentration values.

2.5.3. One-dimensional root water uptake model in multi-dimensional flow modeling

The final analysis investigates the need for a multi-dimensional root water uptake approach in multi-dimensional flow modeling. For this purpose, we compare simulation results when including a one-dimensional root water uptake model in the multi-dimensional water flow modeling. For this purpose, the soil hydraulic parameters \((h_{so}, n, K)\) and root parameters \((Z_r, p_c, \zeta)\) were taken from the optimized HYDRUS-1D model. The final fitting results are expressed by the RMSE and \(R^2\)-values between measured and simulated water content values. These values, combined with the water balance components, including their spatial variability as determined by standard deviation values, are presented in Table 2.5.

Table 2.5. Components of the soil water balance (and standard deviation) obtained using one-dimensional root water uptake model in multi-dimensional flow model.

<table>
<thead>
<tr>
<th>Unit</th>
<th>HYDRUS-2D</th>
<th>HYDRUS-3D</th>
</tr>
</thead>
<tbody>
<tr>
<td>(T_{sw}) [mm]</td>
<td>48.70</td>
<td>48.70</td>
</tr>
<tr>
<td>(E_i) [mm]</td>
<td>6.99</td>
<td>6.25</td>
</tr>
<tr>
<td>(T_r) [mm]</td>
<td>42.46 (37.48)</td>
<td>41.97 (38.02)</td>
</tr>
<tr>
<td>Drainage [mm]</td>
<td>10.80 (4.58)</td>
<td>11.34 (6.44)</td>
</tr>
<tr>
<td>(\Delta\text{Storage}) [mm]</td>
<td>60.25</td>
<td>59.57</td>
</tr>
<tr>
<td>RMSE [m(^3) m(^{-3})]</td>
<td>0.0553</td>
<td>0.0627</td>
</tr>
<tr>
<td>(R^2)</td>
<td>0.61</td>
<td>0.57</td>
</tr>
</tbody>
</table>
As expected, correlations are low ($R^2$-values of about 0.6) and RMSE-values are large (0.05 – 0.06 m$^3$ m$^{-3}$), indicating that a one-dimensional root water uptake model is unable to capture soil moisture variations caused by spatial-variable root water uptake. Moreover, whereas total root water uptake and its variation are almost identical to the comparable HYDRUS1-D values in Table 2.4, total drainage was increased and spatial variability decreased. The inability to predict within root zone soil moisture variability using a one-dimensional root water uptake model in multi-dimensional rooting systems clearly favors the need for multi-dimensional root water uptake and flow models, if detail at this scale is required.

2.6. Conclusions

In this Chapter, we have presented a three-dimensional model for the simultaneous, dynamic simulation of soil water flow and root water uptake. After parameter optimization of the selected root water uptake model and soil hydraulic parameters, the agreement between simulated and measured water contents values during the 16-day period was good, with an overall time average RMSE value of 0.018 m$^3$ m$^{-3}$. These results are excellent bearing in mind that the standard error of the water content measurements was between 0.01 m$^3$ m$^{-3}$ and 0.02 m$^3$ m$^{-3}$. Subsequently, using the same field data set of multi-dimensional volumetric water content values, the results of the three-dimensional root water uptake model were compared with inverse modeling data, describing root water uptake and soil water flow in two and one-dimensions. Independently measured soil water retention data agreed favorably with the optimized retention curves using either one-, two- or three-dimensional root water uptake with corresponding multi-dimensional water flow models. The high value of the optimized water-stress parameter agreed with the low water holding capacity of the sandy field soil.

Optimized root water uptake distributions between one-, two-, and three-dimensional flow models with corresponding root water uptake models were almost identical. These results provide evidence that the presented spatial aggregation of soil moisture data is adequate for calibration purposes to arrive at effective root water uptake parameters. Also, when comparing water balance components between the 3 models, all models were in approximate agreement. However, major differences occurred for the spatial variation in root water uptake and drainage rates between one- and multi-dimensional models. This loss of information regarding variability of drainage rates and root water uptake clearly justifies the need for multi-dimensional root water uptake and flow models, especially when the fate and transport of chemicals below the rooting zone for single trees is of concern.
Part II

Single-objective Parameter Estimation
CHAPTER 3

A Shuffled Complex Evolution Metropolis Algorithm for Optimization and Uncertainty Assessment of Hydrologic Model Parameters

Abstract

Markov Chain Monte Carlo (MCMC) methods have become increasingly popular for estimating the posterior probability distribution of parameters in hydrologic models. However, MCMC methods require the a priori definition of a proposal or sampling distribution, which determines the explorative capabilities and efficiency of the sampler and therefore the statistical properties of the Markov Chain and its rate of convergence. In this Chapter, we present an MCMC sampler entitled the Shuffled Complex Evolution Metropolis algorithm (SCEM-UA), which is well suited to infer the posterior distribution of hydrologic model parameters. The SCEM-UA algorithm is a modified version of the original SCE-UA global optimization algorithm developed by Duan et al. [1992]. The SCEM-UA algorithm operates by merging the strengths of the Metropolis algorithm, controlled random search, competitive evolution, and complex shuffling in order to continuously update the proposal distribution and evolve the sampler to the posterior target distribution. Three case studies demonstrate that the adaptive capability of the SCEM-UA algorithm significantly reduces the number of model simulations needed to infer the posterior distribution of the parameters when compared with the traditional Metropolis-Hastings samplers.

CHAPTER 3

3.1. Introduction and scope

Hydrologic models often contain parameters that cannot be measured directly but which can be inferred by a trial-and-error (calibration) process that adjusts the parameter values to closely match the input-output behavior of the model to the real system it represents. Traditional calibration procedures, which involve “manual” adjustment of the parameter values, are labor-intensive, and their success is strongly dependent on the experience of the modeler. Automatic methods for model calibration, which seek to take advantage of the speed and power of computers while being objective and relatively easy to implement, have therefore become more popular [e.g., Boyle et al., 2000]. Since the early work reported by Dawdy and O’Donnell [1965], automatic calibration procedures have evolved significantly. However, many studies using such methods have reported difficulties in finding unique (global) parameter estimates [Johnston and Pilgrim, 1976; Duan et al., 1992; Sorooshian et al., 1993; Gan and Biftu, 1996].

Regardless of the methodology used, most hydrologic models suffer from similar difficulties, including the existence of multiple local optima in the parameter space with both small and large domains of attraction (a subregion of the parameter space surrounding a local minimum), discontinuous first derivatives, and curving multi-dimensional ridges. These considerations inspired Duan et al [1992] to develop a powerful robust and efficient global optimization procedure, entitled the Shuffled Complex Evolution (SCE-UA) global optimization algorithm. Numerous case studies have demonstrated that the SCE-UA algorithm is consistent, effective, and efficient in locating the optimal model parameters of a hydrologic model [e.g., Duan et al., 1992, 1993; Sorooshian et al., 1993; Luce and Cundy, 1994; Gan and Biftu, 1996; Tanakamaru, 1995; Kuczera, 1997; Hogue et al., 2000; Boyle et al., 2000].

While considerable attention has been given to the development of automatic calibration methods which aim to successfully find a single best set of parameter values, much less attention has been given to a realistic assessment of parameter uncertainty in hydrologic models. Estimates of hydrologic model parameters are generally error-prone, because the data used for calibration contain measurement errors and because the model never perfectly represents the system or exactly fits the data. Consequently, it is generally impossible to find a single point in the parameter space associated with good simulations; indeed, there may not even exist a well-defined region in the sense of a compact region interior to the prior parameter space. Although the SCE-UA global optimization algorithm can reliably find the global minimum in the parameter space, it still remains typically difficult, if not impossible, to find a unique “best” parameter set, whose performance measure differs significantly from other feasible parameter
sets within this region. Such poor parameter identifiability may result in considerable uncertainty in the model output and, perhaps more importantly, makes it virtually impossible to relate these parameter values to easily measurable soil or land-surface characteristics [Schaap et al., 1998; Duan et al., 2001; Vrugt and Bouten, 2002].

Only recently have methods for realistic assessment of parameter uncertainty in hydrologic models begun to appear in the literature. These include the use of a multi-normal approximation to parameter uncertainty [Kuczera and Mroczkowski, 1998], evaluation of likelihood ratios [Beven and Binley, 1992], parametric bootstrapping and Markov Chain Monte Carlo (MCMC) methods [e.g., Tarantola, 1987; Kuczera and Parent, 1998]. Because traditional statistical theory based on first-order approximations and multinormal distributions is typically unable to cope with the nonlinearity of complex models, MCMC algorithms have become increasingly popular as a class of general purpose approximation methods for problems involving complex inference, search, and optimization [Gilks et al., 1996]. An MCMC method is a stochastic simulation that successively visits solutions in the parameter space with stable frequencies stemming from a fixed probability distribution. A variety of MCMC samplers can be constructed for any given problem by varying the sampling or proposal distribution subject to conditions that ensure convergence to the posterior target distribution. These algorithms originally arose from the field of statistical physics where they were used as models of physical systems that seek a state of minimal free energy. More recently, MCMC algorithms have been used in statistical inference and artificial intelligence [Geman and Geman, 1984; Neal, 1993].

Recently, Kuczera and Parent [1998] used the Metropolis-Hastings (MH) algorithm [Metropolis et al., 1953; Hastings, 1970], the earliest and most general class of MCMC samplers, in a Bayesian inference framework to describe parameter uncertainty in conceptual catchment models. The MH algorithm is the basic building block of classical MCMC methods and requires the choice of a proposal distribution to generate transitions in the Markov Chain. The choice of the proposal distribution determines the explorative capabilities of the sampler and therefore the statistical properties of the Markov Chain and its rate of convergence. If the selected proposal distribution closely approximates the posterior target distribution, the Markov Chain that is sampled will rapidly explore the parameter space, and it will not take long to obtain samples that can be treated as independent realizations of the target distribution of interest. However, a poor choice of the proposal distribution will result in slow convergence of the Markov Chain and an inability to recognize when convergence to a limiting distribution has been achieved. For complex hydrologic models, there is usually very little a priori knowledge available about the location of the high-probability density region within the parameter space. The proposal
distribution should therefore express a great deal of initial uncertainty, thereby resulting in slow convergence to the final posterior target distribution (for example, Beven and Binley [1992] suggested imposing a uniform distribution over a large rectangle of parameter values). An important challenge, therefore, is to design MCMC samplers that exhibit fast convergence to the global optimum in the parameter space, while maintaining adequate occupation of the lower posterior probability regions of the parameter space.

To improve the search efficiency of MCMC samplers, it seems natural to tune the proposal distribution during the evolution to the posterior target distribution, using information inferred from the sampling history induced by the transitions of the Markov Chain. This Chapter describes an adaptive MCMC sampler, entitled the Shuffled Complex Evolution Metropolis algorithm (SCEM-UA), which is an effective and efficient evolutionary MCMC sampler. The algorithm, a modification of the original SCE-UA global optimization algorithm developed by Duan et al. [1992], operates by merging the strengths of the Metropolis algorithm [Metropolis et al., 1953], controlled random search [Price, 1987], competitive evolution [Holland, 1975], and complex shuffling [Duan et al., 1992] to continuously update the proposal distribution and evolve the sampler to the posterior target distribution. The stochastic nature of the Metropolis-annealing scheme avoids the tendency of the SCE-UA algorithm to collapse to a single region of attraction (i.e. the global minimum), while information exchange (shuffling) allows biasing the search in favor of better solutions.

This Chapter is organized as follows. In Section 3.2, we describe the Metropolis-Hastings algorithm and the new SCEM-UA algorithm for estimating the posterior probability distribution of hydrologic model parameters. Section 3.3 illustrates the power of both algorithms by means of three case studies with increasing complexity; here we are especially concerned with algorithm efficiency (particularly the number of simulations needed to converge to the stationary posterior distribution). Finally, Section 3.4 summarizes the methodology and discusses the results.

3.2. Search algorithms for assessment of parameter uncertainty

We are interested in hydrologic models that predict outputs from inputs. These models are indexed by parameters, which may (or may not) be physically interpretable. We assume that the mathematical structure of the model is essentially predetermined and fixed. Following Troutman [1985], the hydrologic model $\eta$ can be written as:

$$ y = \eta(\xi | \theta) + \varepsilon $$

(3.1)
where $\hat{y}$ is $N \times 1$ vector of model outputs, $\xi = (\xi_1, \xi_2, \ldots, \xi_N)$ is an $N \times r$ matrix of input values, $\theta = (\theta_1, \theta_2, \ldots, \theta_n)$ is a vector with $n$ unknown parameters, and $\varepsilon$ is a vector of statistically independent errors with zero expectation and constant variance $\sigma^2$. In the classical approach to model calibration, the goal is to find the best attainable values of the parameters $\theta$ such that the vector of error terms, $E(\theta) = \{\varepsilon_1(\theta), \ldots, \varepsilon_N(\theta)\}$, is in some sense forced to be as close to "zero" as possible [Gupta et al., 1998]. For this, the SCE-UA global optimization algorithm developed by Duan et al. [1992] has proven to be consistent, efficient, and effective.

The goal of searching for a single optimal representation of Eq. (3.1) is, however, questionable. For instance, to quote Kuczera and Parent [1998] "... as hydrologic models can be viewed as the result of a combination of conceptual and/or physically based transfer functions, no hydrologist should be naive enough to rely on a uniquely determined value for each of the model parameters $\theta$, whatever the skill and imagination of the modeler may be". Most likely, a search conducted on the feasible parameter space close to the global optimum will reveal many behavioral parameter sets with quite similar performance in reproducing the observed data. If we want to be able to regionalize or relate model parameters to easily measurable land or soil-surface characteristics, a prerequisite is that the parameters are unique, preferably having a small variance. From this perspective, it is necessary to infer the parameter uncertainty resulting from calibration studies.

While classical statistics consider the model parameters $\theta$ in Eq. (3.1) to be fixed but unknown, the Bayesian statistics treat them as probabilistic variables having a joint posterior probability density function (pdf), which captures the probabilistic beliefs about the parameters $\theta$ in the light of the observed data $y$. The posterior pdf $p(\theta | y)$ is proportional to the product of the likelihood function and the prior pdf. The prior pdf with probability density (or mass) function $p(\theta)$ summarizes information about $\theta$ before any data are collected. This prior information usually consists of realistic lower and upper bounds on each of the parameters, thereby defining the feasible parameter space, $\theta \in \Theta \subset \mathbb{R}^n$, and imposing a uniform (noninformative) prior distribution on this rectangle.

Assuming that the residuals are mutually independent, Gaussian distributed, with constant variance, the likelihood of parameter set $\theta$ for describing the observed data $y$ can be computed using [Box and Tiao, 1973]:

$$L(\theta | y) = \exp \left[ -\frac{1}{2} \sum_{i=1}^{N} \left( \frac{\varepsilon_i(\theta)}{\sigma} \right)^2 \right] \quad (3.2)$$
CHAPTER 3

Assuming a noninformative prior of the form \( p(\theta) \propto \sigma^{-1} \), Box and Tiao [1973] showed that the influence of \( \sigma \) can be integrated out, leading to the following form of the posterior density of \( \theta \):

\[
p(\theta | y) \propto \left[ M(\theta) \right]^{-1/2}
\]

(3.3)

where

\[
M(\theta) = \sum_{i=1}^{N} e(\theta)^2
\]

(3.4)

For more information about the Bayesian inference scheme, please refer to Box and Tiao [1973] and, more recently, to Thiemann et al. [2001].

3.2.1. Traditional first-order approximation

The classical approximation to obtain the posterior probability density function from Eq. (3.2) is to use a first-order Taylor series expansion of the non-linear model equations evaluated at the globally optimal parameter estimates \( \theta_{opt} \). The estimated multivariate posterior joint probability density function of \( \theta \) is then expressed as [Box and Tiao, 1973]:

\[
p(\theta | y) \propto \exp \left[ -\frac{1}{2\sigma^2} (\theta - \theta_{opt})^T X^T X (\theta - \theta_{opt}) \right]
\]

(3.5)

where \( X \) is the Jacobian or sensitivity matrix evaluated at \( \theta_{opt} \). The posterior marginal probability density function of \( \theta \) is therefore approximated by the normal distribution, \( N(\theta_{opt}, \sigma^2 \Sigma) \), where \( \Sigma_i \) is the \( i \)th diagonal element of the covariance matrix computed as \( \sqrt{(X^T X)^{-1}} \).

If the hydrologic model is linear (or very nearly linear) in its parameters, the posterior probability density region estimated by Eq. (3.5) can give a good approximation of the actual parameter uncertainty. However, for non-linear models (e.g., hydrologic models), this approximation can be quite poor [Kuczera and Parent, 1998; Vrugt and Bouten, 2002]. Besides exhibiting strong and non-linear parameter interdependence, the surface of \( p(\theta | y) \) can deviate significantly from the multi-normal distribution. It may also have multiple local optima and
discontinuous derivatives [Duan et al., 1992]. In view of these considerations, it is evident that an explicit expression of the joint and marginal probability density functions is often not possible. Fortunately, MCMC samplers are very well suited to dealing with the peculiarities encountered in the posterior pdf of hydrologic model parameters.

3.2.2. Monte Carlo sampling of posterior distribution: The SCEM-UA algorithm

Markov Chain schemes represent a general approach for sampling from the posterior probability distribution $p(\theta | y)$. A Markov Chain is generated by sampling $\theta^{(r+1)} \sim \pi(\theta | \theta^{(r)})$. This $\pi(\cdot)$ is called the transition kernel or proposal distribution of the Markov Chain. Consequently, $\theta^{(r+1)}$ depends only on $\theta^r$ and not on $\theta^0, \theta^1, .., \theta^{r-1}$. Ergodicity and convergence properties of MCMC algorithms to the posterior distribution have been intensively studied in recent literature, and conditions have been given for geometric convergence [Mengersen and Tweedie, 1996; Roberts and Tweedie, 1996]. In practice, this means that if one looks at the values generated by the Markov Chain, which are sufficiently far from the starting value, the successively generated parameter sets will be distributed with stable frequencies stemming from the posterior target distribution. Any statistical quantity of interest such as the probability density function and the various posterior moments can be evaluated from the generated pseudo sample. The most general and earliest MCMC algorithm, known as the Metropolis-Hastings (MH) algorithm [Metropolis et al., 1953; Hastings, 1970], is given as follows:

1. Randomly start at a location in the feasible parameter space, $\theta^0$, and compute the posterior density, $p(\theta^0 | y)$, relevant to this point according to Eq. (3.2) or (3.3).
2. Generate a new configuration $\theta^{(r+1)}$ from $\pi(\theta | \theta^{(r)})$, where $\theta^{(r+1)}$ is called a candidate point and $\pi$ is called the proposal distribution.
3. Evaluate $p(\theta^{(r+1)} | y)$ using Eq. (3.2) or (3.3) and compute $\Omega = p(\theta^{(r+1)} | y)/p(\theta^r | y)$.
4. Randomly sample a uniform label $Z$ over the interval 0 to 1.
5. If $Z \leq \Omega$, then accept the new configuration. However, if $Z > \Omega$, then reject the candidate point and remain at the current position, that is, $\theta^{(r+1)} = \theta^r$.
6. Increment $t$. If $t$ is less than a pre-specified number of draws, then return to step 2.
Note that the MH algorithm will always accept candidate points (jumps) into a region of higher posterior probability but will also explore regions with lower posterior probability with probability $Z$. Indeed, this algorithm is a MCMC sampler generating a sequence of parameter sets, {$\theta^{(0)}, \theta^{(1)}, ..., \theta^{(t)}$}, that converges to the posterior probability distribution, $p(\theta | y)$, for large $t$ [Gelman et al., 1995]. However, the shape and the size of the proposal distribution $g(\cdot)$ is known to be very crucial for the convergence properties of the Markov Chain; see for example Gilks et al., 1995; 1998. When the proposal distribution is too large, too many candidate points are rejected, and therefore the chain slowly covers the target distribution. On the other hand, when the proposal distribution is too small, too many candidate points are accepted, and the chain traverses slowly through the parameter space, thereby resulting in slow convergence.

Various approaches have been suggested to improve the efficacy and efficiency of MCMC samplers. For instance, the proposal distribution can be updated during the evolution to the posterior target distribution using information from the sampling history induced in the transitions of the Markov Chain [Gilks et al., 1996; Haario et al., 1999, 2001]. However, care must be taken to ensure that the adaptation process does not destroy the overall ergodicity of the Markov Chain or diminish the convergence rate. Some authors have suggested that performance might be improved by exchanging information among multiple samplers running in parallel [e.g., Geyer, 1991; Kass and Raftery, 1995]. One possibility, recently implemented in a hydrologic application by Kuczera and Parent [1998], is to periodically update the covariance structure of the proposal-jump distribution by selecting a sample of points generated by each multiple sequence. However, this approach is subject to the same difficulty encountered by any adaptive sampler—how to use information in a way that ensures convergence to the stationary target distribution and desirable asymptotic properties of kernel density estimates derived from the sampler. An important challenge is therefore to design samplers that rapidly converge to the global minimum in the parameters space, while maintaining sufficient occupation of the lower posterior probability regions of the parameter space.

In examining ways to increase information exchange between the sampled candidate points in the Markov Chain, it seems natural to consider the SCE-UA global optimization strategy developed by Duan et al. [1992]. Recently, Thyer et al. [1999] examined the use of simulated annealing, a probabilistic optimization technique, which is intimately related to the MH algorithm, in combination with a Simplex downhill search method (SA-SX) to calibrate parameters in a conceptual catchment model. Although the stochastic nature of their SA-SX algorithm avoided getting trapped in local minima in the parameter space, the SCE-UA algorithm, which operates with a population of points divided into subcomplexes spread out
over the feasible parameter space, was found to be more effective in searching the parameter space, especially when the dimension of the problem was increased. The SCE-UA approach has the desirable characteristic that it explicitly uses information about the nature of the response surface, extracted using the deterministic Simplex geometric shape, to direct the search into regions with higher posterior probability. Moreover, periodic shuffling of the population enhances survivability and performance by a sharing of the information gained independently by each community.

3.2.3. The Shuffled Complex Evolution Metropolis algorithm

The goal of the SCE-UA algorithm developed by Duan et al. [1992] is to find a single best parameter set in the feasible space. Therefore, it continuously evolves the population toward better solutions in the search space, relinquishing occupation of the regions of the parameter space with lower posterior density. This genetic drift, where the majority of the population converges towards a single mode, is typical of many evolutionary search algorithms. To prevent the collapse of the algorithm into the relatively small region of a single best parameter set, we have modified the SCE-UA algorithm. The new algorithm, entitled the Shuffled Complex Evolution Metropolis (SCEM-UA; developed in collaboration between the University of Amsterdam and the University of Arizona) algorithm is given below and is illustrated in the Figures 3.1 and 3.2.

(1) Generate sample – sample \( s \) points \( \{\theta_1, \theta_2, \ldots , \theta_J\} \) randomly from the prior distribution and compute the posterior density \( \{p(\theta^0 | y), p(\theta^2 | y), \ldots , p(\theta^0 | y)\} \) of each point using Eq. (3.2) or (3.3).

(2) Rank points – sort the \( s \) points in order of decreasing posterior density and store them in array \( D[1:s,1:n+1] \), where \( n \) is the number of parameters, so that the first row of \( D \) represents the point with the highest posterior density.

(3) Initialize parallel sequences – initialize the starting points of the parallel sequences, \( S^1, S^2, \ldots , S^q \), such that \( S^k = D[k,1:n+1] \), where \( k = 1,2, \ldots , q \).

(4) Partition into complexes – partition the \( s \) points of \( D \) into \( q \) complexes \( C^1, C^2, \ldots , C^q \), each containing \( m \) points, such that the first complex contains every \( q(j-1)+1 \) ranked point, the second complex contains every \( q(j-1)+2 \) ranked point of \( D \), and so on, where \( j = 1,2, \ldots , m \).
(5) Evolve each sequence – evolution of each of the parallel sequences according to the Sequence Evolution Metropolis algorithm outlined below.

(6) Shuffle Complexes – unpack all complexes $C$ back into $D$, rank the points in order of decreasing posterior density and reshuffle the $s$ points into $q$ complexes according to the procedure specified in Step 4.

(7) Check convergence – check the Gelman and Rubin (GR) convergence statistic. If convergence criteria are satisfied, stop; otherwise, return to step 5. The definition of the GR convergence statistic appears in section 3.2.6 of this manuscript.

---

**Figure 3.1.** Flow chart of the SCEM-UA algorithm.
The use of a large initial random sample provides an extensive exploration of the parameter space, thereby increasing the chance of finding the global optimum of the prescribed density function. The use of a number of parallel sequences with different starting points enables an independent exploration of the search space, thereby allowing that the optimization problem has more than one region of attraction, and enables the use of heuristic tests to judge whether convergence of the sequences to a limiting distribution has been achieved. The use of complexes enables the collection of information gained about the search space by each individual sequence during the evolution process. The shuffling of these complexes enhances the survivability of the sequences by a global sharing of the information gained independently by each parallel sequence. This series of operations results in a robust MCMC sampler that conducts a robust and efficient search of the parameter space.

One of the key components of the SCEM-UA algorithm is the Sequence Evolution Metropolis (SEM) algorithm, as mentioned in Step 5. This algorithm produces new candidate points in each of the parallel sequences $S^k$ by generating draws from an adaptive proposal distribution by using the information induced in the $m$ samples of $C^k$. An outline of the SEM algorithm is given below (see also Figure 3.2):

(I) Compute the mean $\mu^k$ and covariance structure $\Sigma^k$ of the parameters of $C^k$. Sort the $m$ point in complex $C^k$ in order of decreasing posterior density and compute $\Gamma^k$, the ratio of the posterior density of the first ("best") to the posterior density of the last ("worst") member of $C^k$.

(II) Compute $\alpha^k$, the ratio of the mean posterior density of the $m$ points in $C^k$ to the mean posterior density of the last $m$ generated points in $S^k$.

(III) If $\alpha^k$ is smaller than a predefined likelihood ratio, $T$, generate a candidate point, $\theta^{(r+1)}$, by using a multi-normal distribution centered on the last draw, $\theta^r$, of the sequence $S^k$, and covariance structure $\epsilon^2\Sigma^k$, where $\epsilon$ is a predefined jump rate. Go to Step V, otherwise continue with Step IV.

(IV) Generate offspring, $\theta^{(r+1)}$, by using a multi-normal distribution with mean $\mu^k$ and covariance structure $\epsilon^2\Sigma^k$, and go to Step V.

(V) Compute the posterior density, $p(\theta^{(r+1)} | y)$, of $\theta^{(r+1)}$ using Eq. (3.2) or (3.3). If the generated candidate point is outside the feasible parameter space set $p(\theta^{(r+1)} | y)$ to zero.

(VI) Compute the ratio $\Omega = p(\theta^{(r+1)} | y) / p(\theta^r | y)$ and randomly sample a uniform label $Z$ over the interval 0 to 1.
(VII) If \( Z \) is smaller than or identical to \( \Omega \), then accept the new candidate point. However, if \( Z \) is larger than \( \Omega \), reject the candidate point and remain at the current position in the sequence, that is, \( \theta^{(r+1)} = \theta^{(r)} \).

(VIII) Add the point \( \theta^{(r+1)} \) to the sequence \( S^k \).

(IX) If the candidate point is accepted, replace the best member of \( C^k \) with \( \theta^{(r+1)} \), and go to Step X, otherwise, replace the worst member \( \theta^{(m)} \) of \( C^k \) with \( \theta^{(r+1)} \), provided that \( \Gamma^k \) is larger than the predefined likelihood ratio, \( T \), and \( p(\theta^{(r+1)} | y) \) is higher than the posterior density of the worst member of \( C^k \).

(X) Repeat the steps 1-VIII \( L \) times, where \( L \) is the number of evolution steps taken by each sequence before complexes are shuffled.

---

**Figure 3.2.** Flow chart of the SEM strategy employed in the SCEM-UA algorithm.
In the SEM algorithm, candidate points are generated using an adaptive multi-normal proposal distribution with mean identical to the current draw in the sequence and covariance matrix corresponding to the structure induced in the \( m \) points of complex \( k \). However, in situations where the mean posterior density of the last \( m \) generated points in sequence \( k \) is significantly smaller than the mean posterior density of the \( m \) points in the corresponding complex \( k \), the center of the proposal distribution is temporarily switched to the mean of the points in the complex. This particular feature in the SEM algorithm (Step IV) significantly reduces the chance that individual sequences get stuck in a local non-productive region of attraction, thereby further improving the mixing of the sequences. After generating a new candidate point, the posterior density relevant to this point is computed and the Metropolis-annealing [Metropolis et al., 1953] criterion is used to judge whether the candidate point should be added to the current sequence or not. Finally, the last Step (IX) in the SEM algorithm considers which member of the current complex \( k \) should be replaced with the point \( \theta^{(n^*)} \). When the candidate point is accepted, \( \theta^{(n^*)} \), automatically replaces the best member of the complex. However, when the candidate point is rejected, \( \theta^{(n^*)} \) replaces the worst point in complex \( k \) provided that \( \Gamma^k \) is larger than the predefined likelihood ratio, \( T \), and the posterior density relevant to \( \theta^{(n^*)} \) is higher than the posterior density corresponding to the worst point of complex \( k \). Hence, when \( \Gamma^k \) is larger than some prior defined large number (i.e. \( T > 10^5 \)), there is sufficient reason to believe that the covariance of the proposal distribution is specified too big as members with a too low probability are still present in \( C^* \). Replacement of the worst member of \( C^* \) in this particular situation will facilitate convergence to a limiting distribution.

In contrast with traditional MCMC samplers, the SCEM-UA algorithm is an adaptive sampler, where the covariance of the proposal or sampling distribution is periodically updated in each complex during the evolution to the posterior target distribution using information from the sampling history induced in the transitions of the generated sequences. Of course, it is not clear from the algorithm presented above whether the proposed procedure for updating the proposal distribution in view of the past history of the chains will result in an ergodic chain with desirable asymptotic properties of the kernel density estimates derived from the sampler [Haario et al., 1999, 2001]. However, an empirical (experimental) investigation of the ergodicity of the SCEM-UA strategy has revealed that the algorithm performs very well, as demonstrated by different case studies presented in this Chapter.

The SCEM-UA algorithm is different from the original SCE-UA algorithm presented by Duan et al. [1992] in two important ways. Both modifications are necessary to prevent the search from becoming mired in a small basin of attraction and thus to arrive at the correct posterior
target distribution. First, the downhill Simplex method in the Competitive Complex Evolution algorithm outlined by Duan et al. [1992] is replaced by a Metropolis-annealing covariance based offspring approach, thereby avoiding a deterministic drift towards a single mode. Second, the SCEM-UA algorithm does not further subdivide the complex into subcomplexes during the generation of the offspring (candidate points) and uses a different replacement procedure, to counter any tendency of the search to terminate occupations in the lower posterior density region of the parameter space.

3.2.4. Selection of algorithmic parameters in the SCEM-UA algorithm

The SCEM-UA algorithm contains two algorithmic parameters that need to be specified by the user. These are, the number of complexes/sequences, \( q \), and the population size, \( s \), which in turn also determine the number of points within each complex \( (m = s/q) \). For simple problems with an uncorrelated or correlated Gaussian target distribution, relatively small population sizes \( (s \leq 100) \) and a small number of generated sequences/complexes \( (q \leq 5) \) will usually suffice. However, in the case of complex-shaped posterior probability density distributions, like highly non-linear banana-shaped distributions, we recommend the use of larger population sizes \( (s \geq 250) \) and a larger number of parallel sequences/complexes \( (q \geq 10) \) to be able to precisely capture the complex shape of the covariance structure. Specific information about the number of sequences/complexes and the population size can be found in the case study section of this Chapter. Additionally, the SEM algorithm contains three algorithmic parameters which values must be chosen carefully. The SCEM-UA algorithm employed for the different case studies reported in this Chapter used the values \( L = (m/10) \) and \( T = 10^6 \). As a basic choice, the value of the jump rate, \( c_s \), was set to \( 2.4/\sqrt{n} \) [Gelman et al., 1995]. Preliminary sensitivity analyses of the SCEM-UA algorithm indicated that these values for the algorithmic parameters work well for a broad range of applications.

3.2.5. Comparison of SCEM-UA algorithm against traditional MH samplers

To enable a direct comparison in performance between the MH and SCEM-UA algorithm, the \( q \) parallel sequences in the MH sampler were initialized using the \( q \) points that exhibited the highest posterior density in the original \( s \) points of the population. Moreover, the proposal distribution for the traditional MH sampler was set identical to the covariance structure of the random initialized population of points in the feasible parameter space. We argue that this is fair,
SINGLE-OBJECTIVE PARAMETER ESTIMATION

considering the fact that a uniform prior over the predefined feasible parameter space is usually the only information we have about the location of the high-posterior probability density region. To benchmark against more modern MCMC techniques, case study 2 also contains the results for other stable state of the art MH samplers, which are known to maintain ergodicity.

3.2.6. Convergence of MCMC samplers

An important issue in MCMC sampling is convergence of the sampler to the stationary posterior distribution (Step 7 of the SCEM-UA algorithm outlined in section 3.2.3). Theoretically, a homogeneous sampler converges in the limit as \( t \to \infty \), but in any applied problem one must determine how many draws to make with the sampler. Gelman and Rubin [1992] developed a quantitative convergence diagnostic, \( \sqrt{\text{SR}} \), which they call the Scale Reduction score, based on the within and between chain (sequence) variances:

\[
\sqrt{\text{SR}} = \sqrt{\frac{g-1}{g} + \frac{q+1}{q} \frac{B}{W}}
\]  

(3.6)

where \( g \) is the number of iterations within each sequence, \( B \) is the variance between the \( q \) sequence means and \( W \) is the average of the \( q \) within-sequences variances for the parameter under consideration respectively. Note that the product of \( q \) and \( g \) is identical to the total number of draws, \( t \), with the MH- or SCEM-UA sampler. A score close to 1 for \( \sqrt{\text{SR}} \) for each of the parameters indicates convergence. However, because a score of unity is difficult to achieve, Gelman and Rubin [1992] recommend using values less than 1.2 to declare convergence to a stationary distribution.

3.3. Case studies

We compare the power and applicability of the MH and Shuffled Complex Evolution Metropolis algorithms for three case studies with increasing complexity. The first is a synthetic study using a simple one-parameter bimodal posterior probability distribution. This illustrates the ability of both search algorithms to infer the known posterior target distribution. The second case study explores the effectiveness and efficiency of the MH and SCEM-UA samplers for approximating a strongly non-linear banana-shaped posterior probability distribution and investigates the ergodicity of the SCEM-UA sampler. Finally, the third case study involves assessment of
CHAPTER 3

parameter uncertainty using a five-parameter conceptual rainfall-runoff model. In case studies 2 and 3, we are especially concerned with algorithm efficiency, particularly the number of simulations needed to converge to the stationary posterior distribution.

3.3.1. Case study I: A simple bimodal probability distribution

We investigate the applicability of the MH algorithm and SCEM-UA algorithm for assessment of parameter uncertainty in the presence of bimodality. Consider the following bimodal probability density function:

\[ p(\theta) = \frac{1}{\sqrt{2\pi}} \exp\left[-\frac{1}{2} \theta^2\right] + \frac{2}{\sqrt{2\pi}} \exp\left[-\frac{1}{2} (2\theta - 8)^2\right] \]  

(3.7)

which is the sum of two Gaussian probability distributions, each having a different well-defined optimum as demonstrated in Figure 3.3. The plot shows the bimodal character of the posterior distribution, with a high probability region in the area corresponding to the most likely parameter value \((\theta = 4)\) and another high-density region at \((\theta = 0)\) well separated from the first cluster.

![Figure 3.3](image)

Figure 3.3. Bimodal probability distribution and histogram of 4,000 samples generated using (a) the Metropolis and (b) the Shuffled Complex Evolution Metropolis algorithms.

Fig. 3.3 presents a histogram of 4000 samples generated using the MH and SCEM-UA algorithm outlined in section 3.2. Both algorithms generated 5 parallel sequences \((q = 5)\), each with 1,000 samples using a population size of 50 points. The first 200 samples of each sequence were discarded, because it is unlikely that the initial draws come from the stationary distribution.
needed to construct the posterior estimates. An acceptable scale reduction score of less than 1.2 indicated approximate convergence. For both algorithms, it is evident that they are able to successfully infer the posterior target distribution defined in Eq. (3.7).

### 3.3.2. Case study II: A two-dimensional banana-shaped posterior target distribution

This case study explores the effectiveness and efficiency of the MH and SCEM-UA algorithms for inferring a two-dimensional highly non-linear banana-shaped posterior target distribution. In this study, we are especially concerned with the ergodic properties of the SCEM-UA sampler. If the proposed SCEM-UA algorithm is able to generate a useful approximation of the highly complex non-linear banana-shaped test distribution in this study, it seems reasonable to conjecture that the sampler is suited to construct accurate posterior estimates for the parameters in hydrologic models.

The non-linear banana-shaped distribution is constructed from the standard multivariate Gaussian distribution as follows. Let \( f \) be the density of the multivariate normal distribution, \( N(0, \Sigma) \) with the covariance matrix given by \( \Sigma = \text{diag}(100, 1, \ldots, 1) \). The twisted Gaussian density function with non-linearity parameter \( b > 0 \) is given by:

\[
    f_b = f \circ \phi_b
\]

where the function \( \phi_b \) is:

\[
    \phi_b(\theta) = (\theta_1, \theta_2 + b\theta_1^2 - 100b, \theta_3, \ldots, \theta_n)
\]

The non-linearity of function \( \phi_b \) increases with \( b \). In our test, we applied the value \( b = 0.1 \) to yield a strongly twisted banana-shaped target distribution. Due to the complexity of the posterior density surface of this test distribution, the population size in the SCEM-UA algorithm was set to 1000 points, and the number of parallel sequences was set to 10. The feasible parameter space was taken to be a uniform distribution between -100 and 100 for each parameter. The test cases reported in this Chapter have been performed in the dimensions 2 and 8.
In the two-dimensional case, the evolution of the Gelman–Rubin convergence diagnostic for the parameters $\theta_1$ and $\theta_2$ using the MH and SCEM-UA algorithms is illustrated in Figures 3.4a-b, respectively.

Due to random initializations of the starting points of the parallel sequences in the feasible parameters space, the scale reduction factor for the first 10,000 simulations is quite large using the MH sampler (Fig. 3.4a). Thereafter, the convergence diagnostic for both parameters narrows down quickly and continues to widen and narrow intermittently. Finally, after approximately 45,000 simulations, the plot suggests that the parallel sequences have converged to a stationary posterior distribution for both parameters ($\sqrt{\text{SR}} < 1.2$). In the case of the SCEM-UA algorithm, the periodic updating of the covariance structure of the sampling-proposal distribution significantly improves the explorative capabilities of the sampler and the mixing of the sequences. Consequently, far less iterations are needed with the SCEM-UA algorithm than with
the traditional MH sampler to achieve convergence to a stationary posterior distribution. Indeed, the evolution of the convergence diagnostic depicted in Fig. 3.4b demonstrates convergence to a posterior distribution after approximately 10,000 simulations.

Although the results in Fig. 3.4 demonstrate a faster convergence rate of the SCEM-UA sampler over the traditional MH sampler, it is important to check whether the former sampler has converged to the true prior defined two-dimensional banana-shaped posterior test distribution and hence has the right ergodic properties. Table 3.1 summarizes the evolution of the average Euclidean distance of the "true" means and standard deviations of the prior defined probability distribution from those respective values estimated with the SCEM-UA algorithm as function of the number of function evaluations. Each number in the table denotes an average over 100 independent runs.

Table 3.1. Statistical properties of the two-dimensional banana-shaped posterior distribution as function of the number of function evaluations with the MH and SCEM-UA sampler. Mean(|E|) denotes the average Euclidean distance of the SCEM-UA derived mean values from their true values at origo, thus Mean(|E|) = \left( \sum_{i=1}^{n} (E_i)^2 \right)^{\frac{1}{2}}. In a similar way, Std(|E|) denotes the average Euclidean distance of the SCEM-UA derived standard deviations from their true values at origo. To reduce the influence of sampling variability, the presented statistics denote averages over 100 independent runs.

<table>
<thead>
<tr>
<th>No. Evaluations</th>
<th>MH</th>
<th>SCEM-UA</th>
</tr>
</thead>
<tbody>
<tr>
<td>10,000</td>
<td>7.85</td>
<td>1.59</td>
</tr>
<tr>
<td>20,000</td>
<td>2.54</td>
<td>1.00</td>
</tr>
<tr>
<td>30,000</td>
<td>2.94</td>
<td>1.05</td>
</tr>
<tr>
<td>40,000</td>
<td>2.51</td>
<td>0.84</td>
</tr>
<tr>
<td>50,000</td>
<td>2.28</td>
<td>0.96</td>
</tr>
<tr>
<td>60,000</td>
<td>1.84</td>
<td>0.95</td>
</tr>
<tr>
<td>70,000</td>
<td>1.58</td>
<td>0.87</td>
</tr>
<tr>
<td>80,000</td>
<td>1.68</td>
<td>0.81</td>
</tr>
</tbody>
</table>

The results in Table 3.1 demonstrate that the average Euclidean distance of the SCEM-UA estimated mean values and standard deviations from the true values at origo slowly approaches zero, suggesting that the sampler provides a correct estimation of the posterior target distribution and hence does not collapse to a small region of highest posterior density. More significantly, the posterior moments derived with the adaptive SCEM-UA sampler compare favorably well with identical counterparts derived using the MH sampler. This serves as
numerical evidence that the SCEM-UA algorithm has the right ergodic properties and hence provides correct simulation of the target distribution.

Note, however, that both samplers generate slightly biased estimates of the posterior moments as compared to their "true" values of the predefined banana-shaped test distribution. This is also demonstrated in Figures 3.5a-b that present scatterplots of the \( \theta_1, \theta_2 \) sampled MH and SCEM-UA points, respectively, that were generated after convergence of the sequences to a stationary posterior distribution has been achieved. The dark contour line refers to the theoretical 68.3 and 95% confidence regions of the predefined banana-shaped posterior target distribution. Both the MH and SCEM-UA sampled points and posterior moments are consistent with the test target distribution.

![Figure 3.5](image.png)

**Figure 3.5.** A scatterplot of the \( \theta_1, \theta_2 \) generated samples after convergence has been achieved to a stationary posterior distribution using, (a) the Metropolis, and (b) the Shuffled Complex Evolution Metropolis algorithms. The dark dashed lines indicate the one-dimensional 68.3 and 95% confidence regions of the parameters.

The scatterplots presented in Figs. 3.5a-b demonstrate that with an identical number of iterations, the population sampled using the SCEM-UA algorithm is much more diverse than the population sampled using the traditional MH sampler. However, especially in the case of the MH sampler, the sampling density at the extreme tails of the banana-shaped distribution is rather sparse, suggesting that the algorithms experience problems in exploring the lower posterior density region of the parameter space. This explains the slightly biased posterior moments presented in Table 3.1. Clearly, periodically updating of the covariance structure of the proposal distribution in view of the history induced in the transitions of the Markov Chains does not only improve the convergence rate of the sampler but also significantly increases the diversity of the
SINGLE-OBJECTIVE PARAMETER ESTIMATION

The final sampled population. Hence, a more diverse population yields better estimates of the final statistical moments of the posterior distribution and as such is an additional advantage of the SCEM-UA algorithm over traditional MH-samplers.

The transitions of parameter $\theta_i$ in 3 of the 5 parallel sequences (Markov Chains) during the evolution of the MH and SCEM-UA samplers to the stationary posterior distribution is illustrated in Figures 3.6a-b, respectively. For clarity, the three different parallel sequences are coded with different symbols.

![Graph](image)

Figure 3.6. Transitions of the parameter $\theta_i$ in 3 of the 5 parallel generated Markov Chains during the evolution to the banana-shaped posterior target distribution using the (a) MH, and (b) SCEM-UA algorithm. For more explanation, please refer to the text.

The 1-D scatterplots of the sampled parameter space demonstrate that, at early stages during the evolution, the individual sequences tend to occupy different regions of the posterior surface. This low mixing of the paths, especially in the case of the MH sampler (Fig. 3.6a), is associated with a relatively high value for the scale reduction factor (see Fig. 3.4), indicating poor convergence. At a later stage during the evolution, all of the individual sequences have been able to fully explore the banana-shaped posterior target distribution, thereby resulting in a scale reduction factor smaller than 1.2, indicating convergence to a stationary distribution. Note, however, that the mixing of the MH-generated Markov Chains is quite poor, suggesting that the proposal distribution used to sample with the MH sampler was too large. Hence, the transitions
in the Markov Chain reveal a low diversity; too many candidate points are rejected and therefore
the chain slowly covers the posterior target distribution. Graphical examination of the transitions
and mixing of the different symbolic coded paths yields a similar picture about the convergence
status of the sampler as the value of the convergence diagnostic presented in Fig. 3.4. Practical
experience with other case studies also suggests that the GR convergence diagnostic is useful to
test whether convergence to a limiting posterior distribution has been achieved even when the
parallel chains/sequences are not fully independent, as in the case of the SCEM-UA sampler.
Although the SCEM-UA algorithm is an adaptive sampler, which continuously updates the
proposal distribution based on the information induced in the history of the sampled points, the
sampler does not collapse to a single region of highest attraction. Graphical examination of the
sampled parameter space demonstrates that the SCEM-UA sampler maintains occupation at the
extreme tails of the hyperbolic banana-shaped distribution during the evolution. This ensures an
asymptotic convergence to the desirable kernel density estimates and again serves as empirical
evidence that the generated Markov Chains-sequences are ergodic.

To test whether the SCEM-UA algorithm also provides a correction simulation of the
target distribution in higher dimensions, the algorithm was run in various dimensions up to \( n = 500 \). The results of extensive tests in dimension \( n = 8 \) are summarized in Table 3.2. To
benchmark against other modern adaptive MH samplers, the results of the Adaptive Proposal (AP) and Adaptive Metropolis (AM) algorithms, as developed by Haario et al. [1999, 2001], are
also included. These algorithms are known to maintain ergodicity.

Table 3.2. Statistical properties of the 8-dimensional banana-shaped posterior distribution as estimated with the
SCEM-UA algorithm. Mean(\(|E|\)) and Std(\(|E|\)) denote the average distance of the SCEM-UA derived mean
and standard deviations from their true values at origo calculated over 100 repetitions. For more explanation, please
refer to the text.

<table>
<thead>
<tr>
<th>Method</th>
<th>AP</th>
<th>AM</th>
<th>SCEM-UA</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean((</td>
<td>E</td>
<td>))</td>
<td>4.85</td>
</tr>
<tr>
<td>Std((</td>
<td>E</td>
<td>))</td>
<td>4.20</td>
</tr>
</tbody>
</table>

The results in Table 3.2 demonstrate that the SCEM-UA algorithm also generates a correct
simulation of the banana shaped posterior target distribution in the 8-dimensional test case. The
SCEM-UA algorithm has a consistent better performance than the AP algorithm and a quite
similar performance as the AM algorithm. However, the exploration of the parameter space with
a number of parallel sequences, rather than a single sequence search strategy as employed in the
AP and AM algorithms, enables the SCEM-UA algorithm to deal with optimization problems that contain more than one region of attraction. Additionally, the shuffling procedure implemented in the SCEM-UA algorithm enhances the survivability of the sequences by a global sharing of the information gained independently by each parallel sequence. Consequently, as opposed to the AP and AM algorithms developed by Haario et al. [1999,2001] the SCEM-UA algorithm does not require an optimization technique to first locate the high probability density region in the parameter space. Although not explicitly demonstrated we have successfully applied the SCEM-UA algorithm up to $n = 500$ dimensions.

The empirical results presented here illustrate three important findings. First, in the case of a strong banana-shaped curvature of the posterior distribution of the parameters, the SCEM-UA algorithm is successfully able to infer the posterior target distribution. Second, graphical examination of the sampled parameter space and numerical analyses of the statistical properties of the Markov Chain demonstrate that the generated sequences are ergodic, thereby ensuring asymptotic convergence to the desirable kernel density estimates of the posterior target distribution. These results give confidence that the SCEM-UA algorithm will likely yield accurate estimates of the posterior moments for complex problems such as are usually found in hydrologic modeling. Third, periodic tuning of the sampling–proposal distribution by use of a local search direction (sequences), constructed using global information exchange (shuffling), can help make improvements in mixing and therefore in the efficiency (speed of convergence) of MCMC samplers.

3.3.3. Case study III: The HYMOD model

This case study illustrates the usefulness of the MH and SCEM-UA algorithms to hydrologists who are especially concerned with a realistic assessment of prediction uncertainty on hydrologic responses. For this purpose, we used HYMOD, a 5-parameter conceptual rainfall-runoff model (see Figure 3.7), introduced by Boyle [2000] and recently used by Wagener et al. [2001]. The HYMOD model consists of a relatively simple rainfall excess model, described in detail by Moore [1985], connected with two series of linear reservoirs (three identical quick and a single reservoir for the slow response) and requires the optimization of five parameters to observed streamflow data: the maximum storage capacity in the catchment, $C_{mx} \,(L)$, the degree of spatial variability of the soil moisture capacity within the catchment, $b_{txp} \,(\%)$, the factor distributing the flow between the two series of reservoirs, $\alpha \,(\%)$, and the residence time of the linear quick and slow reservoirs, $R_q \,(T)$ and $R_s \,(T)$, respectively.
In keeping with previous studies [e.g., Thiemann et al., 2001], approximately 11 years (28 July 1952 to 30 September 1962) of hydrologic data from the Leaf River watershed were used for model calibration. This humid 1944 km² watershed, located north of Collins, Mississippi, has been investigated intensively [e.g., Brazil, 1988; Sorooshian et al., 1993; Boyle et al., 2000; Thiemann et al., 2001]. The data, obtained from the National Weather Service Hydrology Laboratory (HL), consist of mean areal precipitation (mm/day), potential evapotranspiration (mm/day), and streamflow (m³/s). This data set was used to test the efficiency and effectiveness of the MH and SCEM-UA algorithms for estimating the posterior distribution of the model parameters and to assess prediction uncertainty on the hydrologic responses. The prior uncertainty ranges of the parameters are defined in Table 3.3.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Unit</th>
<th>Min.</th>
<th>Max.</th>
</tr>
</thead>
<tbody>
<tr>
<td>$C_{\text{max}}$</td>
<td>[mm]</td>
<td>1.000</td>
<td>500.000</td>
</tr>
<tr>
<td>$b_{\text{op}}$</td>
<td>[-]</td>
<td>0.100</td>
<td>2.000</td>
</tr>
<tr>
<td>Alpha</td>
<td>[-]</td>
<td>0.100</td>
<td>0.990</td>
</tr>
<tr>
<td>$R_s$</td>
<td>[d]</td>
<td>0.000</td>
<td>0.100</td>
</tr>
<tr>
<td>$R_q$</td>
<td>[d]</td>
<td>0.100</td>
<td>0.990</td>
</tr>
</tbody>
</table>

To reduce sensitivity to state value initialization, a 65-day warm-up period was used, during which no updating of the posterior density defined in Eq. (3.3) was performed. Moreover, we used a population size of 250 points and we applied a Box-Cox power transformation with $\lambda = 0.3$ to the measured and HYMOD predicted streamflow to avoid problems with heteroscedastic and non-Gaussian error distributions.

The evolution of the Gelman and Rubin scale-reduction convergence diagnostic for each of the model parameters using the MH and SCEM-UA algorithms is illustrated in Figures 3.8a-b.
Both sampling algorithms generated 5 parallel sequences ($q=5$), each with 6,000 samples. Note the different scales on the $y$-axes of the two plots. Clearly, the SCEM-UA algorithm is more efficient in traversing the parameter space, with convergence to a stationary posterior distribution ($\sqrt{\text{SR}} < 1.2$) for each of the HYMOD model parameters achieved after approximately 4,000 simulations. In contrast, the MH sampler is far from convergence to a stationary distribution, with the scale-reduction factor for each of the parameters larger than 2, even after performing 30,000 simulations. To understand why, consider Figures 3.9a-b, which present the evolution of samples generated in three parallel launched sequences, using either the MH or SCEM-UA algorithm.

**Figure 3.8.** Evolution of the Gelman and Rubin scale-reduction factor for the parameters in the HYMOD model using 11 years of runoff data (1952-1962) for the Leaf River watershed for (a) the Metropolis-Hastings, and b) the Shuffled Complex Evolution Metropolis algorithm.

**Figure 3.9.** Markov Chain Monte Carlo $C_{\text{max}}$-samples generated in three parallel sequences using either (a) the Metropolis-Hastings, or (b) the Shuffled Complex Evolution Metropolis algorithm.
CHAPTER 3

Because no information about the location of the highest probability density region in the parameter space is exchanged between parallel sequences launched using the MH- sampler, it remains difficult for the sequences (chains) to mix and terminate their occupation in regions of the parameter space with a low posterior density. Due to this slow mixing rate, the posterior moments derived from the samples generated in each of the parallel chains differ appreciably between the different chains, and the scale-reduction factor remains far above the threshold value of 1.2, indicating lack of convergence. Given the ability of the SCEM-UA algorithm to exchange information about the search space gained by the different parallel launched sequences, this increases the explorative capabilities of the sampler and therefore the traversing speed of the chains through the feasible parameter space. This behavior is evident in Fig. 3.9b. Consequently, the population sampled using the SCEM-UA algorithm is more diverse. Periodic shuffling of the complexes in the SCEM-UA algorithm ensures sharing of information gained independently by each community about the nature of the posterior distribution and therefore increases the traversal through the parameter space. This allows us to make more sound inferences about the nature of the posterior probability density function.

Although not presented in this Chapter, we also performed a variety of experiments with the MH sampler to speed up convergence to the posterior distribution. One of those experiments was to periodically update the covariance structure of the jump-proposal distribution using a sample within each sequence, and to use this covariance structure to generate new candidate points in each parallel sequence. In this way, the information gained by each individual local sampler is more thoroughly exploited. However, using this approach, a significantly larger number of simulations were needed, as compared to the SCEM-UA algorithm, to achieve convergence (typically 10,000 simulations).

Figure 3.10 presents the posterior marginal probability density distributions for each of the HYMOD model parameters inferred for the Leaf River watershed using the samples generated with the SCEM-UA algorithm. While the histograms of Alpha and the slow-and quick-tank recession parameters, Rq, and Rs, exhibit an approximately normal Gaussian distribution, the posterior density distributions for the other model parameters reveal the existence of several modes. This multi-modality suggests the presence of multiple regions of attraction in the posterior surface and illustrates the severity of the optimization problem, even in the case of this simple and parsimonious 5-parameter conceptual rainfall runoff model using more than 10 years of daily streamflow data.
Note, however, that the parameter sets sampled using the SCEM-UA algorithm occupy a relatively small range, interior to their uniform prior distributions (e.g., Table 3.3), which indicates that the HYMOD model parameters are reasonably well identifiable from measured time series of runoff.

In Table 3.4, we present the final posterior moments derived using those samples that were generated with the SCEM-UA algorithm after convergence to the stationary target distribution has been achieved. For this, the first 600 simulations of each parallel sequence were discarded (i.e., $\sqrt{SR} > 1.2$). Also included are the most likely parameter values, describing the approximately 11 years of runoff data of the Leaf River watershed, identified using the SCE-UA global optimization algorithm [Duan et al., 1992]. As stated earlier, the posterior standard deviations and correlation coefficients between the sampled parameters depict that the parameters of the HYMOD model are well identifiable using measured runoff data. A direct comparison between the optimal parameter values derived using the original SCE-UA global optimization algorithm and the SCEM-UA algorithm demonstrates that the latter algorithm is
CHAPTER 3

not only able to conveniently derive the posterior distribution of the model parameters, but also successfully identifies the most likely parameter values within this high-density region.

Table 3.4. Shuffled Complex Evolution Metropolis posterior mean (Mean), standard deviation (Std), coefficient of variation (CV), and correlation coefficients between the generated samples for the HYMOD model parameters using 11 years of runoff data (1952-1962) for the Leaf River watershed. Also included is the parameter set found with highest posterior probability using the SCEM-UA (SCEM) algorithm and the original SCE-UA (SCE) global optimization algorithm. The corresponding units are presented in Table 3.3.

<table>
<thead>
<tr>
<th>Par.</th>
<th>Mean</th>
<th>Std.</th>
<th>CV</th>
<th>Cmax</th>
<th>kmax</th>
<th>Alpha</th>
<th>R</th>
<th>Ry</th>
<th>SCEM</th>
<th>SCE</th>
</tr>
</thead>
<tbody>
<tr>
<td>b</td>
<td>0.38</td>
<td>0.026</td>
<td>7.24</td>
<td>---</td>
<td>1.00</td>
<td>0.85</td>
<td>-0.44</td>
<td>-0.40</td>
<td>253.41</td>
<td>253.63</td>
</tr>
<tr>
<td>Alpha</td>
<td>0.84</td>
<td>0.0091</td>
<td>1.18</td>
<td>---</td>
<td>1.00</td>
<td>0.85</td>
<td>-0.37</td>
<td>-0.38</td>
<td>0.38</td>
<td>0.38</td>
</tr>
<tr>
<td>R</td>
<td>0.0027</td>
<td>0.0051</td>
<td>19.44</td>
<td>---</td>
<td>——</td>
<td>——</td>
<td>1.00</td>
<td>0.27</td>
<td>0.0029</td>
<td>0.0030</td>
</tr>
<tr>
<td>Ry</td>
<td>0.46</td>
<td>0.0028</td>
<td>0.61</td>
<td>---</td>
<td>——</td>
<td>——</td>
<td>——</td>
<td>1.00</td>
<td>0.46</td>
<td>0.46</td>
</tr>
</tbody>
</table>

Clearly, this feature is an additional benefit of the SCEM-UA algorithm because it makes superfluous the two-step procedure in which the global optimum in the parameter space is first identified, followed by launching parallel MH samplers from this starting point to identify parameter uncertainty.

Finally, Figure 3.11 illustrates how the results of the SCEM-UA algorithm can be translated into estimates of hydrograph prediction uncertainty, using data from WY 1953.

Figure 3.11. (a) Hydrograph prediction uncertainty associated with the most likely set derived using the SCEM-UA algorithm. The light-gray region denotes model uncertainty, whereas parameter uncertainty is indicated with the dark-gray region. The dots correspond to the observed streamflow data. (b) Hydrograph prediction uncertainty associated with the uncertainty in the model (light-gray) and parameter estimates (dark-gray region) for the WY 1953. The dots correspond to the observed streamflow data.
Figure 3.11a shows the observed streamflows (dots), the 95% hydrograph prediction uncertainty associated only with the posterior distribution of the parameter estimates (dark-gray region), and the 95% hydrograph prediction uncertainty associated with the total error (light-gray region) in terms of the model residuals (computed by subtracting the hydrograph prediction having the highest posterior probability). Figure 3.11b shows the same information in the streamflow space. Note that the 95% total prediction uncertainty ranges bracket the observed flows during the period, but are quite large, indicating considerable uncertainty in the predictions given the current model structure and the data used to condition the model. Further, the hydrograph prediction uncertainty associated only with the posterior distribution of the parameter estimates (dark gray) does not include the observations and displays bias on the long recessions, suggesting that the model structure may be in need of further improvement.

3.4. Summary

This Chapter has presented a MCMC sampler, which is well suited for the practical assessment of parameter uncertainty in hydrologic models. The sampler, entitled the Shuffled Complex Evolution Metropolis algorithm, merges the strengths of the Metropolis-Hastings algorithm, controlled random search, competitive evolution, and complex shuffling to evolve a population of sampled points to an approximation of the stationary posterior distribution of the parameters. There are two differences between the SCEM-UA algorithm and the original SCE-UA algorithm presented by Duan et al. [1992]. These modifications prevent the search from becoming mired in a small basin of attraction and facilitate convergence to a stationary posterior target distribution. The first modification involves replacement of the downhill Simplex method by a Metropolis-annealing covariance-based offspring approach, thereby avoiding a deterministic drift towards a single mode. Second, the SCEM-UA algorithm does not further divide the complex into subcomplexes during the generation of the offspring and uses a different replacement procedure, to counter any tendency of the search to terminate occupations in the lower posterior density region of the parameter space.

The efficiency and effectiveness of the newly developed SCEM-UA algorithm for estimating the posterior distribution of the parameters was compared with the traditional MH sampler for three case studies of increasing complexity. The first case study, a simple bimodal probability distribution, showed that the SCEM-UA algorithm does indeed successfully infer a known posterior target distribution. The second and third case study explored the effectiveness and efficiency of the algorithm for assessing parameter uncertainty in a highly non-linear banana-
shaped posterior test distribution and a conceptual watershed model. Both studies clearly demonstrated that the SCEM-UA is efficient, needing smaller numbers of simulations than the MH algorithm for realistic assessment of parameter uncertainty. The ability of the SCEM-UA algorithm to exchange information between parallel launched sequences increases the traversing speed of the chains through the feasible parameter space. Besides inferring the posterior distribution of the model parameters, an additional advantage of the SCEM-UA algorithm is that it simultaneously identifies the most likely parameter values within this high-density region. This makes superfluous the two-step procedure in which the global optimum in the parameter space is first identified, followed by launching parallel MH samplers from this starting point to estimate parameter uncertainty.
CHAPTER 4

Inverse Modeling of Large-Scale Spatially-Distributed Vadose Zone Properties Using Global Optimization

Abstract

Computational capabilities have evolved to a point, where it is possible to use multi-dimensional physically-based hydrologic models to study spatial and temporal patterns of water flow in the vadose zone. However, models based on multi-dimensional governing equations have only received limited attention, in particular because of their computational, distributed input, and parameter estimation requirements. The aim of the present Chapter is to explore the usefulness and applicability of the inverse method to estimate vadose zone properties using the solution of a physically-based, distributed three-dimensional model combined with spatially distributed measured tile drainage data from the 3880 ha Broadview Water District (BWD) in the San Joaquin Valley of California. The inverse problem is posed within a single criterion Bayesian framework and solved by means of the computerized Shuffled Complex Evolution Metropolis (SCEM-UA) global optimization algorithm. To study the benefits of using a spatially distributed three-dimensional vadose zone model, the results of the 3-D model were compared with those obtained using a simple storage-based bucket model and a spatially-averaged one-dimensional unsaturated water flow model for a 2-year period. District-wide results demonstrate that measured spatially distributed patterns of drainage data contain only limited information for the identification of vadose zone model parameters, and are particularly inadequate to identify the soil hydraulic properties. In contrast, the drain conductance, and a soil matrix bypass coefficient were well determined, indicating that the dominant hydrology of the BWD was determined by drain system properties and preferential flow. Despite the significant CPU time needed for model calibration, results suggest that there are advantages in using physically-based hydrologic models to study spatial and temporal patterns of water flow at the scale of a watershed. These models not only generate consistent forecasts of spatially-distributed drainage data during the calibration and validation period, but also possess unbiased predictive capabilities with respect to measured groundwater table depths not included in the calibration.

1. Introduction and scope

The predictive capability of unsaturated flow and transport models relies heavily on accurate estimates of the soil water retention and unsaturated soil hydraulic characteristics at the application scale of the model. To enable such accurate soil physical characterization, methodologies need to be developed that allow a rapid, reliable, and cost-effective estimation of the hydraulic properties of the considered soil domain, including its spatial variability. Most of the early work reporting on the estimation of hydraulic properties of unsaturated soils has focused on relatively small soil samples using static or steady-state flow experiments. These static or steady-state flow experiments have the advantage of being relatively simple to implement. However, these methods are typically time consuming and require restrictive initial and boundary conditions to satisfy the assumptions of the corresponding analytical solutions.

Significant advances in computational capabilities have resulted in Inverse Modeling (IM) applications for the estimation of soil hydraulic properties from small soil cores [Durner et al., 1997; Hopmans et al., 2002a; Vrugt and Dane, 2004]. However, when using an IM approach, the soil hydraulic properties can no longer be estimated by direct inversion, but are determined using an iterative solution, thereby placing a heavy demand on computational resources. In this iterative process, the soil water retention and unsaturated soil hydraulic conductivity characteristics are indirectly determined from repeated numerical simulations of the governing Richards' equation:

\[ C(h_m) \frac{\partial h_m}{\partial t} = \nabla \cdot [K(h_m) \nabla (h_m + \xi)] + A(x, y, z, t) \]  \hspace{1cm} (4.1)

thereby minimizing the difference between the observed and model predicted flow variables such as water content and fluxes. Using one-, two-, or three-dimensional forms, this transient equation solves for soil water matric potential, water content and water flux density as a function of time and space. In Eq. (4.1), \( C \text{ (L}^1 \text{)} \) denotes the soil water capacity, \( K \text{ (LT}^{-1} \text{)} \) is the unsaturated hydraulic conductivity tensor, \( h_m \text{ (L)} \) is the soil water matric head, \( \xi \text{ (L)} \) denotes the gravitational head to be included for the vertical flow component only, and \( A \text{ (L}^2 \text{L}^3 \text{T}^{-1}) \) is the volumetric sink term, representing sources and/or sinks of water. For isotropic soils, \( K \) simplifies to a scalar that is a function of both \( h_m \) and the spatial coordinates. Boundary conditions must be included to allow for specified soil water potentials or fluxes at all boundaries of the simulated unsaturated soil domain. Moreover, user-specified initial conditions and time-varying source/sink terms need
be specified. Both, the soil water retention and unsaturated hydraulic conductivity functions (referred to as soil hydraulic functions) are highly nonlinear, with $h_m$ and $K$ varying many orders of magnitude over the water content range that significantly contributes to water flow.

Research on the applicability and suitability of the inverse approach towards identification of the soil hydraulic properties has focused primarily on five issues, (i) the type of transient experiment and kind of prescribed initial and boundary conditions suited to yield a reliable characterization of the soil hydraulic properties [Hopmans et al., 2002a; van Dam et al., 1992, 1994; Ciollaro and Romano, 1995; Santini et al., 1995; Šimůnek and van Genuchten, 1996, 1997; Šimůnek et al., 1998a; Romano and Santini, 1999; Durner et al., 1997; Wildenschild et al., 2001], (ii) the determination of the appropriate quantity and most informative kind of observational data [Zachman et al. 1981; Kool et al., 1985; Parker et al., 1985; Kool and Parker, 1988; Valiantzas and Kerkides, 1990; Toorman et al., 1992; Eching and Hopmans, 1993; Eching et al., 1994; among others], (iii) the selection of an appropriate model of the soil hydraulic properties [Zachman et al, 1982; Russo, 1991; Zurmühl and Durner, 1998], (iv) the construction and weighting of multiple sources of information in an objective function [van Dam et al., 1994; Hollenbeck and Jensen, 1998; Vrugt and Bouten, 2002], and (v) the adoption and development of Bayesian and multiple-criteria parameter estimation strategies that can be used to quantify the uncertainty (probabilistic and multi-objective) associated with the inversely estimated soil hydraulic properties [Kool and Parker, 1988; Hollenbeck and Jensen, 1998; Vrugt and Bouten, 2002; Vrugt et al., 2003a; Vrugt and Dane, 2004]. With these developments, the capabilities and limitations of the inverse approach for the identification of soil hydraulic properties from laboratory soil cores may be considered reasonably well understood. Despite this progress made, still little is known about the suitability of the inverse approach for the identification of vadose zone properties at larger spatial scales.

It has been a major challenge to integrate these small-scale measurements of soil hydraulic properties in hydrologic models that apply across a range of spatial and temporal scales [Gelhar, 1986; Grayson and Blöschl, 2001]. In most applications, prediction of soil-water dynamics at larger spatial scales uses soil hydraulic properties determined from laboratory core or small field plot measurements, and are included in hydrologic models with a grid or element size much larger than the core or field plot scale. Because of the high nonlinearity of the soil hydraulic functions, their application across spatial scales is inherently problematic. Specifically, the averaging of processes determined from discrete small-scale samples may not be representative of the key hydrologic processes of the larger spatial domain. In addition, the dominant hydrologic flow processes may vary between spatial scales, so that potentially different models need to be used to describe water flow at the soil pedon, field, or watershed scale.
Typically, in hydrologic studies of large spatial dimensions, one may apply a deterministic approach, using a distributed physically-based model with upscaled effective soil properties [Blöschl et al., 1995] or use stochastic modeling. A stochastic model preserves the small-scale characteristics of the measurement, but provides estimates of effective properties at the larger spatial scale after accounting for spatial heterogeneity of hydraulic properties. Stochastic approaches to upscale soil hydrologic processes from the local to the field scale include analytical models, based primarily on perturbation approximations of Richards’ equation [Zhang, 2002; amongst others]. Alternatively, numerical stochastic models have used Monte-Carlo (MC) simulations to derive effective field scale hydraulic properties and to predict field scale hydraulic behavior based on local scale measurements [Hopmans and Stricker, 1989; Harter and Yeh, 1998; Harter and Zhang, 1999]. Any of these approaches can become computationally intensive, requiring a large number of model simulations.

Instead of a formal upscaling technique that incorporates nonlinear effects on upscaled soil properties applicable across a range of spatial domains, we here propose using a deterministic inverse modeling approach [Hopmans et al., 2002b]. To estimate effective vadose zone parameters, a parameter optimization technique will be applied that is consistent regarding the spatial and temporal scale of the measurement and model parameter support. However, unlike in small-scale experiments, boundary and initial conditions at the larger spatial scales are not as clearly defined because direct measurement techniques are mostly not available. Furthermore, the data available to characterize large-scale vadose processes are sparse, both in space and time. Both these factors lead to significant uncertainty when estimating effective large-scale parameters. Therefore, the application of deterministic inverse modeling at the watershed scale must account for the uncertainty of the estimated large-scale parameters and the associated prediction uncertainty. This requires the use of statistically based parameter estimation algorithms. An additional key element of the proposed distributed approach is that the applied vadose zone model must be able to simulate the key hydrologic processes that dominate the larger spatial domain, by using appropriate effective hydrologic parameters. This requires the use of a process-based model that incorporates functional distributed vadose zone parameters that can account for the relevant observed hydrologic processes.

Current computational capabilities have evolved to a point, where it is now possible to use multi-dimensional physically-based watershed models to study spatial and temporal patterns of water flow in the vadose zone [Beven, 2001; Hydrogeologic, 2001; Madsen, 2003]. With the availability of powerful personal computers, efficient computational methods, and sophisticated GIS, remote sensing and advanced visualizations tools, the hydrologic community is beginning
to take advantage of the potential and utility of these physically-based numerical models. With few exceptions, these models are based on complex multi-dimensional governing equations. They have received limited attention, primarily due to their computational, distributed input, and parameter estimation requirements.

Considerable progress has been made in the application of automated optimization algorithms to estimate hydrologic model parameters across a range of spatial scales. However, emphasis is mostly placed on the estimation of a single optimal set of model parameters, thereby effectively neglecting the influence of parameter uncertainty. Such uncertainties arise mainly from the inability of the calibration process to uniquely identify a single optimal parameter set, from measurement errors associated with system input and output and from model structure errors. The hydrologic community is increasingly aware that hydrologic model identification and evaluation procedures should explicitly include uncertainty estimates [Kuczera and Parent, 1998; Bates and Campbell, 2001; Thyer et al., 2002; Vrugt et al., 2003a; this thesis; 2004]. To acknowledge the presence of parameter uncertainty and to develop a tool which can be used to estimate this uncertainty, Vrugt et al. [Chapter 3 of this thesis], recently developed the Shuffled Complex Evolution Metropolis (SCEM-UA) global optimization algorithm. The SCEM-UA algorithm is a general purpose global optimization algorithm that provides an efficient estimate of the most likely parameter set and its underlying posterior probability distribution within a single optimization run. The algorithm is an extension of the SCE-UA population evolution method developed by Duan et al. [1992].

The aim of the present Chapter is to explore the usefulness and applicability of this inverse method to estimate vadose zone parameters at the small catchment and watershed scale by using spatially-distributed tile drainage data as calibration targets. We hypothesize that the proposed inverse modeling approach will significantly improve our understanding of unsaturated water flow at larger spatial and temporal scales. To test the proposed model calibration approach, we selected the 3880-ha Broadview Water District (BWD), located on the west side of the San Joaquin Valley of California. The BWD has been the subject of various investigations [Vaughan and Corwin, 1994, Vaughan et al., 1995 and 1999; Bourgault et al., 1997; Corwin et al., 1999]. These research efforts have resulted in a comprehensive measurement data set of spatially distributed, weekly tile drainage flows and groundwater table depths throughout the district. This dataset provides a unique opportunity to study spatial and temporal patterns of soil water flow by inverse modeling. As in other watersheds, however, this data set includes considerable uncertainties, arising from unknown soil properties, limited information about the spatial variations in rainfall, crop transpiration and soil evaporation across BWD, and the unknown
CHAPTER 4

spatial distribution of groundwater flow. In this Chapter, we compare three different mathematical models (representing different levels of model complexity) and consider three spatial resolution scales (field, drainage unit, and water district scale) for their ability to minimize uncertainty in the calibration parameters while also minimizing model prediction errors.

The remainder of this Chapter is organized by sections. Section 4.2 discusses the BWD including an overview of the measurements that are available for model calibration, presents a condensed description of the MODHMS and BUCKET hydrologic models, and describes the SCEM-UA algorithm, which is used to solve for the single criterion optimization problem. In Section 4.3, we explore the usefulness and applicability of the combined MODHMS - inverse methodology for the identification of vadose zone properties across a range of spatial scales using varying model complexity and spatial resolution of the boundary conditions. Fully integrated three-dimensional solutions of the Richards’ equation (4.1) with spatially distributed boundary conditions are compared with results from a simplified conceptual bucket model and with one-dimensional solutions of the unsaturated flow equation with upscaled, spatially averaged boundary conditions. Finally, a summary with conclusions is presented in Section 4.4.

4.2. Materials and Methods

4.2.1. Field site

The 3880 ha (9700-acres) Broadview Water District (BWD) is part of the San Joaquin Valley watershed and is located approximately 100 km west of Fresno. The district meets several criteria that are intrinsically important to the objectives of this study:

1. The district’s size is typical for mid-sized watersheds, much larger than the typical field scale;
2. An extensive tile-drain system underlies much of the district, which creates tight coupling between vadose zone, shallow groundwater, and drainage fluxes;
3. A comprehensive data set of measured input and output data is available, which relates the measured surface boundary conditions to spatial drainage patterns, measured in sumps of 25 tile-drained drain units, and groundwater table depths.

The western San Joaquin Valley is topographically flat with southwest-northeast sloping deposits with slopes less than 1%. The geomorphic landscape consists of a series of alluvial fans that have been deposited by intermittent streams originating in the Coast Ranges to the west. The BWD is located on the Panoche Creek alluvial fan.
These alluvial deposits are derived from marine sedimentary calcareous and gypsiferous shales and sandstones of the Coast Ranges [Harradine, 1950]. They are underlain by the Corcoran clay, a lacustrine clay deposit that extends throughout the region, and lies at depths of 244 m (valley margin) to 30 m (near river) below the surface [Belitz, 1988]. The alluvial soils derived from the
Coast Range alluvium are generally fine-textured soils, with an average clay content of 50%. The clay fraction is dominated by montmorillonite, with significant swelling and shrinking properties, accounting for most of the cation exchange behavior. Organic matter contents are less than 1%. Most soils are calcareous and gypsiferous.

### 4.2.1.1. Field description

Figure 4.2 presents a schematic overview of the BWD. Most of the water district is divided into quarter sections, each with an approximate area of 64 ha (160-acres). Some smaller fields exist near the district boundary that have irregular shapes. The total district consists of about 60 tile-drained 64-ha agricultural fields. Tile drains are installed at depths of 1.6 to 2 m. Horizontal drain spacings range from 100 m to 200 m. Sumps are located in the northeast corner of each of the drainage units for collection and disposal of drainage water.

**Broadview Water District**

**LEGEND**
- **agricultural fields**
- **drainage unit**
- **not drained**
- **model grid**
- **wells**

Figure 4.2. Schematic overview of the BWD. The thick lines denote the field borders and outline individual agricultural fields. The color coding and numbers refer to individual drainage units, which may consist of a single field, two adjacent fields, or all four agricultural fields in a 1.6 km by 1.6 km (one square mile) land section. The grid of dashed black lines refers to the structured mesh of the MODHMS model. Groundwater observation wells are indicated with circles.
The 8 fields shown in white (numbered "26"), are without a tile-drain system. We assumed that their irrigation would not affect the water balance of the surrounding fields or drainage units, and were not included in our analysis. Table 1 lists the cropping areas for 1995 and 1996. The main crops are cotton, tomatoes, alfalfa, melons and wheat.

Table 4.1. Crop areas for the BWD for 1995 and 1996.

<table>
<thead>
<tr>
<th>Crop</th>
<th>1995</th>
<th>1996</th>
</tr>
</thead>
<tbody>
<tr>
<td>Alfalfa Seed</td>
<td>348.8</td>
<td>356.1</td>
</tr>
<tr>
<td>Cotton</td>
<td>1972.5</td>
<td>2370.2</td>
</tr>
<tr>
<td>Fallow</td>
<td>184.1</td>
<td>27.5</td>
</tr>
<tr>
<td>Garbanzo Beans</td>
<td>54.6</td>
<td>364.2</td>
</tr>
<tr>
<td>Melons</td>
<td>284.5</td>
<td>461.3</td>
</tr>
<tr>
<td>Oats and Melons</td>
<td>22.3</td>
<td>60.7</td>
</tr>
<tr>
<td>Tomatoes</td>
<td>806.1</td>
<td></td>
</tr>
<tr>
<td>Vetch Oats</td>
<td>20.2</td>
<td></td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td>3648.2</td>
<td>3648.2</td>
</tr>
</tbody>
</table>

4.2.1.2. Measurements

The district maintains irrigation turnouts in the southwest corner of each field. Fields are either furrow or sprinkler irrigated. Water supplied from these irrigation turnouts \( (I) \) was measured every two days (bi-daily) with an approximate precision of about 120 m\(^3\) (0.1 acre-feet). The cumulative drainage water from each drainage unit, \( Q_{\text{drain}} \) was measured weekly at 25 sumps. Water draining into the sumps is pumped into drainage ditches and metered to the same precision as the irrigation water. Additionally, groundwater table measurements were measured in various wells (see Fig. 4.2) at irregular time intervals. A detailed description of the measurements in the BWD is presented in Vaughan et al. [1999].

Weekly rainfall amounts and reference evapotranspiration \( ET_0 \) were provided by the Firebaugh weather station of the California Irrigation Management Information System (CIMIS). Evapotranspiration was estimated by two separate procedures for the growing and for the fallow seasons. For the growing-season, field-specific actual crop evapotranspiration values \( (ET) \) were estimated using the standard crop coefficient approach [Allen et al., 1998]:

\[
ET = K_c ET_0
\]
where $K_c$ is the crop coefficient, and $ET_o$ is reference evapotranspiration for grass. Local information on time-varying crop coefficients and crop-specific planting dates were taken from Snyder et al. [1989]. Actual evapotranspiration rates were distributed over the root zone assuming a trapezoidal root distribution function. A piecewise linear crop water stress function [Feddes et al., 1978] was used that reduces crop transpiration if the soil water pressure falls below -3 m. Typical parameter values for various crops were taken from van Dam et al. [1997].

Soil evaporation during the fallow season, $E_s$, was estimated external to the models described below using a one-dimensional unsaturated flow model (HYDRUS-1D, Šimůnek et al., 1998b) that solves (1) subject to measured daily rainfall events and daily potential evapotranspiration rates as the upper boundary condition and a constant water table at 2 m as lower boundary condition. Soil hydraulic properties were set identical to those for the MODHMS simulations (see below).

These upper boundary fluxes were computed at the field scale. In some of the case studies, we used upscaled boundary fluxes at the individual drainage unit scale and at the BWD scale. These were computed as area-weighted averages of the individual field scale fluxes: irrigation, $I$, precipitation, $P$, actual crop transpiration, $ET_a$, and soil evaporation, $E_s$ (all expressed in units of (L/T)). Average depth to groundwater within each drainage unit was estimated by averaging of the available local groundwater table measurements (Fig. 4.2). Figure 4.3 shows average weekly boundary fluxes and drainage flow (mm wk⁻¹) at the BWD scale for the 1995 and 1996 water years (1 October, 1994 through 30 September, 1996).

![Figure 4.3. Weekly calculated soil surface boundary conditions for 1995 and 1996; (Time 0 corresponds to October 1994). Also included is the spatially averaged weekly measured drainage data of the BWD.](image-url)
Because of poor data quality, Drainage Unit 5 was not further used throughout the remainder of this study, thereby reducing the number of drainage units to 24.

While the measurement dataset, by all practical standards, is exceptionally rich in its spatio-temporal resolution, it does not account for actual irrigation and crop water uptake non-uniformity or soil variability at the sub-field scale, uncertainty in irrigation application timing between various sub-sections of individual fields, and non-uniformity of precipitation and ET at the sub-district scale. A key objective of this Chapter is to determine whether such a dataset contains sufficient information to accurately identify the vadose zone model parameters, despite potentially measurement and boundary condition errors.

### 4.2.2. Hydrologic models

In the past three decades, there has been an evolution in hydrologic watershed models for simulation of both single-event and continuous hydrologic processes. Earlier models quantified various hydrologic components using simplified procedures, including the unit hydrograph method, empirical formulas, and analytical equations. These so-called lumped parameter models do not consider spatially-distributed processes.

Moving from lumped parameter to distributed parameter approaches, numerical watershed models are now available that are based on multi-dimensional governing equations that allow for heterogeneous parameter distribution and heterogeneous boundary conditions. It is not yet clear whether these physically-based distributed models are better predictors than lumped parameter models [see Beven, 1989, and Grayson et al., 1992 for an extensive discussion]. The application of physically-based models to larger spatial scales is inherently problematic, because of the potential limitations of available model equations as applied to heterogeneous hydrologic systems, the lack of a theory of subgrid scale integration, and because of problems of dimensionality in parameter calibration [Beven, 1989]. These models usually employ, to various degrees, some form of lumping or spatio-temporal averaging over the local, plot, and field scale, thereby putting into question whether their parameters can be physically interpreted [Feddes et al, 1993].

Specifically, one may question the validity of Richards’ equation to describe soil water flow at spatial scales representing field or watershed scales. It is our view, however, that like any other hydrologic model, flow models based on Richards’ equation are applied as an approximation to the underlying hydrologic system to characterize the dynamics of spatially distributed drainage flow as a response to variable surface boundary conditions. Some may
question whether Richards’ equation is valid at all, even at the much smaller spatial scales of laboratory columns or small field plots often used to simulate flow and transport. Addressing this issue is beyond the scope of this Chapter.

The presented approach takes the point-of-view that the modeling of large-scale vadose zone systems is hampered by a lack of accurate spatially-distributed information on boundary conditions and hydrological measurements and parameters that would be required for a full physical application and interpretation of the governing physically-based Eq. (4.1). Instead, it is proposed to use statistically based inverse modeling to estimate the probability distribution of “effective” hydrological parameters that pertain to the inherent large spatial and temporal scales of the hydrologic domain. The effective parameters are calibration parameters and their posterior (calibrated) distributions reflect the uncertainties associated with model assumptions and boundary conditions.

In this Chapter, we compare parameter uncertainty ranges and predictive capability of a physically-based and a simplified bucket model. The models were applied to explicitly determine whether a physically-based, distributed hydrologic modeling approach is preferable over the bucket model given the spatial and temporal scales of the data collected at BWD. Furthermore, two case studies are implemented with different spatial resolution of the model parameters. In case study I, spatially uniform parameters are estimated at the BWD scale. In case study II, inverse modeling is used to obtain spatially distributed parameter sets at the drainage unit scale, resulting in one parameter set for each of 24 drainage units.

4.2.2.1. Physically-based soil water flow modeling

Eq. (4.1) is solved numerically using a mass-lumped fully implicit finite-difference method with adaptive time stepping. The dependence of $K$ and $h_m$ on $\theta$ is represented by van Genuchten-Mualem type models (see Section 4.2.2.3). The nonlinearities arising from these functions are handled with Newton-Raphson linearization. The adapted numerical model used in this study is MODHMS [HydroGeoLogic, 2001], an extension of the finite difference groundwater flow model MODFLOW [McDonald and Harbaugh, 1988]. The model constitutes a distributed fully coupled surface/vadose zone/groundwater flow model, based on state-of-the-art nonlinear computational algorithms and integrated with GIS-based graphical user interfaces.

For the three-dimensional model, the spatial domain of the BWD was divided into 1536 200x200 m square cells, using approximately 16 grid cells per agricultural field (see Fig. 4.2). In both the three-dimensional and the one-dimensional, vertical model, the soil profile was
discretized into 14 layers using 10 top layers with a thickness of 0.30 m (1 feet) and four remaining layer thicknesses of 0.60 m (2 feet), 0.60 m, 0.9 m (3 feet), and 0.9 m, resulting in a total of 21504 cells.

The initial head distribution throughout the profile was estimated by assigning a drainage unit specific water content value, $\theta_{ini}$, to the top nodal point of the finite difference grid, assuming hydrostatic equilibrium and a uniform soil. Subsurface drainage to the tile drains was simulated using a head-dependent function, with drain discharge proportional to the head above the drain and a drain-conductance parameter:

$$
Q = D_r \left[ H - (Z_0 - D_d) \right] \quad H \geq Z_0 - D_d \\
Q = 0 \quad H < Z_0 - D_d
$$

where $Q$ (L T$^{-1}$) is the subsurface drainage, $D_r$ (T$^{-1}$) is drain conductance, $H$ (L) is total hydraulic head in the layer that contains the drain, $Z_0$ (L) is land surface elevation, and $D_d$ (L) is the drain depth. Since the exact drain depth, $D_d$, and drain conductance ($D_r$) for each drainage unit were unknown, they were considered to be calibration parameters.

4.2.2 Boundary conditions

The fully integrated 3-D approach uses spatially distributed upper boundary conditions at field-scale resolution. The three-dimensional approach allows for lateral water flow through the unsaturated zone, for example, when fields are irrigated in sections, and also allows for continuously varying water table at the resolution of the finite difference grid (sub-field scale). In case-study I, the simulation domain of the three-dimensional model is the BWD, whereas in case-study II, the simulation domain of the three-dimensional model is the drainage unit. The one-dimensional flow model was used to represent the water district scale (case-study I) or the drainage-unit scale (case-study II) using appropriate upscaled boundary fluxes.

We assumed that no flow occurred across the lateral and lower boundaries of each of the models. A previous study in Broadview Water District [Vaugban et al., 1999] suggested that ignoring regional groundwater flows may lead to an underestimation of drainage flow. However, we found that most drainage units did not have drainage flows outside the growing season, indicating little or no regional groundwater contribution that may be considered as "baseflow". In addition, detailed water balance computations of BWD and each drainage unit, as well as various preliminary model calibrations runs with different lower boundary conditions, including
a head dependent flux lower boundary condition, demonstrated that the flux across the bottom of the 6.0-m deep soil profile was negligible compared to the surface and drainage fluxes. For these calculations, total hydraulic head gradients were estimated using piezometric head observations from deep wells surrounding the BWD. Furthermore, a preliminary annual water balance of the entire district indicated that the magnitude of regional groundwater flows between the BWD and surroundings districts was small relative to the measurement accuracy of the other water balance components.

4.2.2.3. Soil hydraulic properties

In order to solve Eq. (4.1), it is necessary to specify the water retention and unsaturated soil hydraulic conductivity function. These hydraulic relationships were defined by the van Genuchten-Mualem (VGM) parameters [van Genuchten, 1980; Mualem, 1976]:

\[ S_e - \theta_r - \theta_f = \left( \frac{1 + (\alpha |b_n|)^\gamma}{\gamma} \right)^\nu \quad b_n < 0 \]
\[ \frac{K(S_e)}{K_o} = \left( \frac{1 - S_e^{1/n}}{1 - \theta_r^{1/n}} \right)^2 \quad b_n \geq 0 \]

where \( S_e \) is the effective water saturation, \( \theta_r \) and \( \theta_f \) (L^3 L^{-3}) denote the saturated and residual water content, \( \alpha \) (L^1) and \( n \) (dimensionless) are curve shape parameters for the soil water retention curve, \( K_o \) denotes the saturated hydraulic conductivity, and \( \gamma \) is a unit-less fitting parameter for the unsaturated soil hydraulic conductivity function. In MODHMS, the parameter \( \gamma \) is fixed at a value of 0.5. As outlined hereafter, rather than considering all VGM parameters as calibration parameters, spatially-distributed soil hydraulic functions will be estimated using calibrated scaling factors.

4.2.2.4. Scaling of soil hydraulic properties

The scaling approach of Miller and Miller [1956] has been extensively used to characterize soil hydraulic spatial variability and to develop a standard methodology to assess the variability of soil hydraulic functions and their parameters [Peck et al., 1977; Hopmans and Stricker, 1989; Clausnitzer,
et al., 1991; Tuli et al., 2001]. The single objective of scaling is to coalesce a set of hydraulic relationships into a single reference curve using scaling factors that describe the set as a whole. Using the scaling factors, the soil water retention and hydraulic conductivity curve of any drainage unit $i$ can be related to the reference hydraulic functions $h_{m,ref}(\theta)$ and $K_{r,ref}(\theta)$ using:

$$b_{m,i} = \frac{b_{m,ref}}{\chi_i}$$ \hspace{1cm} (4.6)

and

$$K_i = \chi_i^2 K_{r,ref}$$ \hspace{1cm} (4.7)

The hydraulic VGM parameters of the reference soil hydraulic functions for the BWD were estimated with the neural network model ROSETTA of Schaap et al. [1998] for a clay soil, which is the dominant soil type in the district [NRC3, 2003]. The ROSETTA computed values are listed in Table 4.2, and represent effective soil hydraulic parameters throughout the 6.0-m profile.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Unit</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\theta_0$</td>
<td>[cm$^3$ cm$^{-3}$]</td>
<td>0.45</td>
</tr>
<tr>
<td>$\theta_s$</td>
<td>[cm$^3$ cm$^{-3}$]</td>
<td>0.10</td>
</tr>
<tr>
<td>$\alpha$</td>
<td>[cm$^{-1}$]</td>
<td>0.02</td>
</tr>
<tr>
<td>$n$</td>
<td>[-]</td>
<td>1.25</td>
</tr>
<tr>
<td>$K_r$</td>
<td>[cm d$^{-1}$]</td>
<td>14.82</td>
</tr>
</tbody>
</table>

Although the physical basis for application of the scaling concept to scales of hundreds of meters and larger is not obvious, the scaling factor approach provides an effective means to describe spatial variability of soil hydraulic functions between drainage units, using a single parameter. In the absence of available data on the spatial distribution of soils within and between drainage units, and to limit the number of calibration parameters, we treated the soil in each drainage unit as homogeneous.
4.2.2.5. Preferential transport to tile drains

Water flow transport in the vadose zone as expressed by Eq. (4.1) does not describe the accelerated transport of water through preferential flow paths. Recent investigations by Kohler et al. [2001, 2003] demonstrated that preferential flow can be an important hydrologic component in agricultural field soils. A significant fraction of the infiltrated water may move through soil cracks and along roots, thereby short-circuiting the soil matrix pore space and reaching groundwater much faster than predicted from Richards' equation.

Preferential flow in structured media is usually described using dual-porosity or dual-permeability models [Šimůnek et al., 2003]. These type of models require a relatively large number of input parameters, and is not warranted because of the spatial and temporal resolution of the calibration data. For example, the dual-permeability model of Gerke and van Genuchten [1993] may require 16 parameters. To avoid problems of overparametrization, we initially considered a simpler, bimodal variant of the multimodal pore-size distribution model developed by Mohanty et al. [1997], to account for potentially accelerated movement of water to a tile drain. This model simulates a rapid increase in the hydraulic conductivity near saturation by dividing the soil matrix into a capillary-dominated and a non-capillary dominated flow domain. However, preliminary optimization runs for each of the drainage units in the BWD demonstrated that the parameters in the pore size distribution model of Mohanty et al. [1997] were not well determined by calibration to spatially distributed drainage data. At the spatial and time scales considered here, the simulation results were nearly identical to those obtained using a traditional solution of the Richards' equation with the unimodal VGM model (Eqs. (4.1) with (4.4) and (4.5)). Hence, we implemented the following lumped preferential flow mechanism in all considered models:

\[
Q_{\text{quick}}(t) = f_y [P(t) + I(t)]
\]  

(4.8)

where \( Q_{\text{quick}} \) is the drainage by preferential flow, \( P \) denotes the weekly rainfall, \( I \) is the infiltrating irrigation water, and \( f_y \) is the fraction of applied water that bypasses the soil matrix, to be determined during model calibration. The corresponding water volume does not contribute to soil storage changes but moves directly in the tile drains. Similar approaches to simulate preferential water flow were presented by Jarvis [1994] and Šimůnek et al. [2003]. In this study, potential tail water generated by field run-off of applied irrigation water into the drain sumps is accounted for by the quick-flow component as well. Thus, the total subsurface discharge to the tile drains of each drainage unit was computed by adding the quick flow

\[
Q_{\text{total}} = Q_{\text{quick}} + Q_{\text{slow}}
\]
component to the drainage resulting from matrix flow into the groundwater. In a subsequent paper, Schoups et al. [2004] will further elaborate on the validity of this model concept by comparing observed and model predicted drainage salinity concentrations.

**4.2.2.6. Model Calibration and Evaluation**

We initially selected the following five calibration parameters: \( \chi, D_x, D_y, f_u \) and \( \theta_{sw} \). To increase flexibility in the retention function close to saturation we included \( \theta_{sw} \) as an additional calibration parameter. Moreover, since the model simulations were highly sensitive to the estimated \( ET_a \), we included a crop \( ET \) correction parameter, \( f_a \) as another calibration parameter, thereby allowing adjustment of computed actual transpiration rates to acknowledge the potential uncertainty associated with reported crop coefficients. For example, differences in values for crop coefficients as reported in Snyder et al. [1989] and Allen et al. [1988] can be about 25%. Moreover, the calibrated effective crop coefficient values may effectively correct for water balance errors, caused for example by neglecting regional groundwater flow across the BWD and drainage unit boundaries. The final simulated actual \( ET \), was computed from the product of \( f \) and \( ET \). These seven effective parameters make up the parameter set, \( \Theta \). The values of \( \Theta \) were estimated by calibration using inverse modeling, against the spatially distributed, weekly drainage data that serve as calibration targets. The prior uncertainty ranges for each of these parameters are defined in Table 4.3, and apply to both the BWD and drainage unit scale.

**Table 4.3.** Calibration parameters for the MODHMS model, including their prior uncertainty ranges.

<table>
<thead>
<tr>
<th>Par.</th>
<th>Unit</th>
<th>Description</th>
<th>Min</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \chi )</td>
<td>[-]</td>
<td>Scaling factor</td>
<td>0.01</td>
<td>10.0</td>
</tr>
<tr>
<td>( D_x )</td>
<td>[d (^{-1})]</td>
<td>Drain conductance</td>
<td>0.00006</td>
<td>0.19</td>
</tr>
<tr>
<td>( D_y )</td>
<td>[cm]</td>
<td>Drain depth</td>
<td>60.0</td>
<td>450.0</td>
</tr>
<tr>
<td>( \theta_{iw} )</td>
<td>[cm(^3)cm(^{-3})]</td>
<td>Initial water content</td>
<td>0.15</td>
<td>0.40</td>
</tr>
<tr>
<td>( \theta_s )</td>
<td>[cm(^3)cm(^{-3})]</td>
<td>Saturated water content</td>
<td>0.25</td>
<td>0.50</td>
</tr>
<tr>
<td>( f_a )</td>
<td>[-]</td>
<td>Crop-coefficient adjustment factor</td>
<td>0.70</td>
<td>1.30</td>
</tr>
<tr>
<td>( f_b )</td>
<td>[-]</td>
<td>Bypass fraction</td>
<td>0.00</td>
<td>0.30</td>
</tr>
</tbody>
</table>

We note that the calibrated vadose zone model parameters, including the soil hydraulic functions, represent effective soil properties, whose values can generally not be obtained by direct measurements [Feddes et al., 1993].
As is common in inverse modeling [Hill, 1998], we use the first part of the available dataset (1995) for model calibration and the second half of the dataset (1996) for model evaluation (validation). Such a split sample test provides the only means to test the validity of the calibrated parameters. To further examine the validity of the spatially distributed physically-based hydrologic models we also evaluated the predictive capabilities of these models by comparing measured and simulated groundwater table depths at various locations within the BWD. As these spatially distributed water table depths were not used during the calibration, this is a much stronger test of the internal consistency of the model than the model validation based on measured drainage dynamics alone.

### 4.2.3. Storage-based bucket model

We compared the performance of the one- and three-dimensional MODHMS models with a simple bucket-type lumped storage-based model, BUCKET. For the bucket model, the vadose zone is conceptualized by two independent compartments in parallel, each characterized by the instantaneous water balance:

$$\frac{dS}{dt} = In - Out$$  \hspace{1cm} (4.9)  

where $S$ (L) denotes soil water storage, $In$ (LT$^{-1}$) is the sum of all incoming fluxes, and $Out$ (LT$^{-1}$) is the sum of all outgoing fluxes using weekly time intervals. The first compartment represents the matrix flow domain, whereas the second compartment is the preferential flow domain. No interaction is allowed between the two compartments. The explicit forward-in-time finite-difference approximation of the water balance equation for the matrix flow domain results in:

$$S(t + \Delta t) = S(t) + \Delta t \left[ (1 - f) \left[ P(t) + I(t) \right] - f \cdot ET_r(t) - E_s(t) - Q_{slm}(t) \right]$$  \hspace{1cm} (4.10)  

where $\Delta t$ (T) denotes the time step, and $Q_{slm}$ (LT$^{-1}$) is the slow (matrix) component of drainage. Drainage out of the matrix domain, $Q_{slm}$ was calculated as follows:

$$Q_{slm}(t) = \begin{cases} a \cdot \exp(b \left[ S(t) - S_{min} \right]) & \text{if } S(t) > S_{min} \\ 0 & \text{if } S(t) \leq S_{min} \end{cases}$$  \hspace{1cm} (4.11)  

98
where $S_{\text{min}}$ (L) is the minimum storage needed to initiate drain flow from the matrix domain, $a$ (LT$^{-1}$) denotes the minimum attainable drainage rate, and $b$ (L$^{-1}$) is an additional fitting parameter describing the rate of drainage as a function of soil water storage. Eq. (4.10) is solved for each time step starting from initial water storage, $S_{\text{in}}$, using weekly time steps. Total drainage to the tile-drain is computed according to:

$$Q_{\text{drain}} = Q_{\text{slow}} + Q_{\text{quick}}$$

in which the quick-flow component ($Q_{\text{quick}}$) is modeled the same way as in the physically-based model. In conclusion, the bucket model contains six parameters, $S_{\text{in}}, f_{c}, S_{\text{min}}, f_{b}$, $a$, and $b$. For each of the 25 drainage units, the prior ranges for each of these parameters are listed in Table 4.4.

### Table 4.4. Calibration parameters of the conceptual BUCKET model, including their prior uncertainty ranges

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Unit</th>
<th>Description</th>
<th>Min</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td>$S_{\text{in}}$</td>
<td>[cm]</td>
<td>Initial storage</td>
<td>0.00</td>
<td>500.00</td>
</tr>
<tr>
<td>$S_{\text{min}}$</td>
<td>[cm]</td>
<td>Minimum soil water storage</td>
<td>0.00</td>
<td>500.00</td>
</tr>
<tr>
<td>$a$</td>
<td>[cm d$^{-1}$]</td>
<td>Minimum attainable drainage</td>
<td>0.00</td>
<td>10.0</td>
</tr>
<tr>
<td>$b$</td>
<td>[cm$^{-1}$]</td>
<td>Drainage parameter</td>
<td>0.00</td>
<td>1.00</td>
</tr>
<tr>
<td>$f_{c}$</td>
<td>[-]</td>
<td>Crop-coefficient adjustment factor</td>
<td>0.70</td>
<td>1.30</td>
</tr>
<tr>
<td>$f_{b}$</td>
<td>[-]</td>
<td>Bypass fraction</td>
<td>0.00</td>
<td>0.30</td>
</tr>
</tbody>
</table>

### 4.2.4. Parameter optimization by inverse modeling

While much research has been devoted to developing automated procedures for calibration of lumped parameter models, automated parameter estimation for distributed physically-based watershed models is a new area of research [Madsen, 2003]. Physically-based distributed hydrologic models potentially contain a large number of model parameters which need to be estimated by calibration against a historical record of data. Specifically, for each of the drainage units within the BWD six or seven parameters need to be estimated, depending on which model is used. To improve parameter optimization methods, they need to be applicable for cases with large number of calibration parameters and must provide parameter uncertainties with corresponding model uncertainties. The SCEM-UA algorithm is a general purpose global optimization algorithm that provides an efficient estimate of the most likely parameter set and its underlying posterior probability distribution within a single optimization run.
When using the SCEM-UA algorithm to retrieve the posterior distribution of the model parameters, the posterior density of each parameter set needs to be estimated. In this study, we used the following standard density function [Box and Tiao, 1973]:

$$p(\Theta | y) \propto \prod_{j=1}^{N} \left[ \frac{\varepsilon_j(\Theta)^2}{\sigma^2} \right]^{-\frac{1}{2}}$$

(4.13)

in which $p(\Theta | y)$ denotes the posterior density (assumed uniform a-priori) given the parameter set $\Theta$ and observed data $y$, and $\varepsilon$ is a $N \times 1$ vector of residuals, which are calculated by computing the difference between the observed data $y$ and the corresponding model predictions, $\hat{y}$, of size $N$. In the remaining part of this Chapter, the stationary posterior distribution of the parameters is also referred to as the High Probability Density (HPD) region of the parameter space. The marginal parameter uncertainty ranges reported in the remainder of this Chapter correspond to the 0.5% and 99.5% percentiles of the stationarity posterior probability distribution.

The posterior probability distribution for each of the optimization runs reported in this Chapter was estimated using 3,000 trials with the SCEM-UA algorithm, resulting in an equal number of generated parameter sets and model runs. Convergence to a stationary posterior distribution was monitored by comparison of first and second order statistical moments for each of the parameters in the different SCEM-UA generated sequences. To reduce the computational time needed to perform each optimization run for the three-dimensional solution of the Richards' equation, the SCEM-UA algorithm was partly rewritten to facilitate parallel optimization using a pre-specified number of computers. Specifically, in this study we used 10 Pentium IV 2.8 GHz computers to perform each of the 3,000 MODHMS-3D model runs in parallel.

4.3. Results and discussion

In this section, we illustrate and discuss the results obtained when performing the various optimizations with the SCEM-UA algorithm. In particular, we explore the usefulness and applicability of the inverse method for the identification of soil and drainage system properties at the drainage unit and BWD district scale. Throughout this section we are especially concerned with the benefits of using a fully integrated three-dimensional solution of the Richards' equation.
APPLICATION OF THE SCEM-UA ALGORITHM

(MODHMS-3D) by comparing its results with those obtained using the simplified conceptual bucket model (BUCKET) and the one-dimensional solution of the unsaturated flow equation (MODHMS-1D).

To verify whether improved predictions of spatially distributed drainage data can be made by implementing finer spatial grid discretization and finer temporal discretization, we compared MODHMS-1D simulation results using bi-daily boundary conditions in combination with 100 nodes in the vertical direction (Calibration I) with those using the weekly boundary conditions and 14 nodal values with depth (Calibration II), as used in all subsequent modeling exercises. As an example, we present a comparison of Drainage Unit 4 in Figure 4.4.

![Figure 4.4](image)

**Figure 4.4.** Influence of spatial and temporal discretization on performance of the MODHMS-1D model for Drainage Unit 4, (a) Bi-daily boundary conditions in combination with 100 nodes in the vertical direction, and (b) weekly boundary conditions and 14 nodes in vertical direction. The dashed lines define the drainage rates for the most likely parameter set, whereas the gray shaded area in both plots denotes the MODHMS-1D prediction uncertainty ranges associated with the HPD region of the parameter space. The solid circles correspond to the drain flow observations.

The 2 cases produce nearly identical RMSE values although the higher temporal discretizations produced improved temporal dynamics in the drainage discharge. In part, this is caused by outliers in the observations that cannot be simulated, irrespective of the discretization scheme.
4.3.1. Case study I: Spatially-uniform (BUCKET and MODHMS-1D) and spatially-distributed boundary conditions (MODHMS-3D) with spatially-uniform calibration parameters

The first case study investigates the ability of BUCKET and MODHMS-1D and -3D to predict spatially distributed drainage data based on district-wide uniform parameters. To accomplish the calibration for the one-dimensional BUCKET and MODHMS-1D model, the district-average boundary conditions were used (Fig. 4.3), whereas the boundary conditions were spatially distributed across agricultural fields for the MODHMS-3D simulations. For each of the three models, the likelihood function of Eq. (4.13) included the discharge data of each of the 24 drainage units during the 1995 calibration year, resulting in a total of 1248 observations. The results of this calibration are summarized in Figures 4.5, 4.6 and 4.7.

Figure 4.5 presents parameter plots for each of the members in the HPD region of the parameter space for the (a) BUCKET, (b) MODHMS-1D, and (c) MODHMS-3D model. The parameters were scaled according to their prior uncertainty ranges defined in Table 4.3 and 4.4 to yield normalized ranges. Additionally, Figure 4.5b presents scatterplots of simulated versus observed district-wide drainage data. These scatter plots are created by plotting the model-predicted weekly average drainage values corresponding to the most likely parameter set against the corresponding observed values for each of the 24 different drainage units. The results of Fig. 4.5 highlight several important observations. First, the parameters $a$, $b$, $f_r$, and $f_h$ of the BUCKET model and the parameters $D_n$ and $f_b$ of the MODHMS-models are very well determined by calibration to spatially distributed drainage data, as shown by the narrow range of the HPD region in the prior physically plausible space for each of these parameters. Secondly, there is considerable uncertainty associated with the other parameters, regardless of the modeling approach. Particularly, the soil hydraulic parameters cannot be adequately characterized. Thus, the scaling factor and saturated water content in the MODHMS-1D and -3D models are poorly defined, meaning that “acceptable” model simulations are found over a wide-range of parameter values. It appears that much of the hydrologic regime of the BWD can be explained by a combination of drain system characteristics and preferential flow mechanisms.
Figure 4.5. Case I: Normalized uncertainty plots for each of the parameters of the (a) BUCKET, (b) MODHMS-1D, and (c) MODHMS-3D models, constructed using the results for the whole BWD calibration. The calibration parameters for each model are listed along the x-axis, while the y-axis corresponds to the parameter values, scaled according to their prior uncertainty ranges (defined in Table 4.3 and 4.4) to yield normalized ranges between zero and one. Each gray line across the graph represents one member of the HPD region. The solid black line going from left to right across the plot corresponds to the mode of the posterior distribution. The scatterplots at the right hand side depict simulated versus observed district-wide drainage data. The dashed lines represent the 1:1 line.

The scatterplots in Fig. 4.5b show that the conceptual BUCKET and MODHMS-1D model do not capture the observed drainage discharge dynamics as well as the MODHMS-3D model. The large variability in measured drainage flows within the BWD was primarily caused by the temporal variability in the start of irrigation events between different drainage units. Therefore, when distributing boundary conditions across fields in the MODHMS-3D simulations, the comparison improves slightly. Overall, all three models greatly underestimate drainflows at high discharge events, thereby causing some bias in the computed water balance. Note, however, that this bias is not due to neglected dynamical processes in the BUCKET and...
MODHMS models, but a natural consequence of the Bayesian density function implemented in this study. The classical calibration approach, which minimizes the squared residuals between model predictions and measurements, may give small RMSE values, but at the expense of considerable model bias [see Figure 5 in Boyle et al., 2000]. Parameter sets with minimal variance (lowest RMSE) tend to have a strong bias, whereas sets having close to zero bias have somewhat larger RMSE values.

The inability of the inverse procedure to resolve a single, relatively unique set of hydraulic parameters from measured spatially distributed drainage data is further illustrated in Figure 4.6, which presents the soil water retention (Figs. 4.6a and 4.6b) and unsaturated soil hydraulic conductivity functions (Figs. 4.6c and 4.6d) for each of the members of the HPD region (Fig. 4.5).

**Figure 4.6.** Case I: Soil water retention and unsaturated soil hydraulic conductivity functions corresponding to each of the members of the HPD region of Fig. 4.5 for (a,b) MODHMS-1D and (c,d) MODHMS-3D. The solid and dashed black lines correspond to the hydraulic functions of the most likely parameter set and reference curve, respectively.
The significant uncertainty associated with the fitted soil water retention and unsaturated soil hydraulic conductivity functions, further supports the view that the soil hydraulic properties - at the spatial and temporal scales considered - are not well determined by calibration to measured drainage data. Consequently, when the hydrologic response is dominated by effective properties, such as drain depth, drain conductance and bypass flow, soil hydraulic properties may play only a minor role. We also note that the most optimal hydraulic properties, indicated with the black line, are well removed from the center of the prediction uncertainty bounds, suggesting a lognormal distribution of the scaling factor in the HPD region. Another significant and interesting observation is that the identifiability of the hydraulic parameters does not improve when increasing the spatial model dimension from 1 to 3. A similar conclusion was reported in previous work when comparing inversely estimated root water uptake parameters obtained with a one-, two-, and three-dimensional soil water flow model and spatially distributed soil water content data [Vrugt et al., in Chapter 2 of this thesis].

To assess the relative contributions of the quick and slow-flow components to total simulated drainage, Figure 4.7a elucidates the BUCKET computed terms of the total drainage flow.

![Figure 4.7a](image)

**Figure 4.7.** Case I: BUCKET model - simulated drainage and soil water storage, \( S \), over the 1995 calibration year corresponding to the most likely identified parameter set, (a) slow-flow, quick-flow, and total drainage, and (b) soil water storage.

For completeness, Fig. 4.7b presents the corresponding BUCKET computed temporal changes in soil water storage, \( S(t) \), from Eq. (4.10). Notice that the quick-flow component is generally
small compared to the slow-flow component. Indeed, for each of the numerical models, the quick-flow component \( Q^{\text{const}} \) constitutes between 1 and 13% of the total simulated drainage flow \( Q^{\text{sim}} \) to the tile-drains.

### 4.3.2. Case study II: Spatially distributed boundary conditions with spatially distributed calibration parameters

In this case study we decreased the horizontal scale represented by the parameters, thereby increasing the number of estimated parameters. Individual parameters were estimated for each of the 24 drainage units, thereby distributing the parameters over the BWD. Preliminary optimization runs showed that water fluxes between adjacent drainage units were negligible (smaller than 5% of the total drainage unit drainage). Drainage units were, therefore, modeled individually. This approach is computationally not only more attractive, but also increases the prospects of finding the preferred parameter solutions for each of the drainage units, as only a limited number of parameters need to be estimated in each optimization run. Hence, depending on whether the BUCKET or MODHMS models were used, 6 or 7 parameters were identified for each drainage unit by calibration to the weekly discharge data. The results are summarized in Figures 4.8, 4.9, and 4.10.

Figure 4.8 presents scatterplots of simulated drainage flows, derived with the most likely parameter set, versus their observed values for the (a) BUCKET, (b) MODHMS-1D, and (d) MODHMS-3D models for the total set of drainage units.

![Figure 4.8. Case II-calibration: Scatterplot of simulated versus observed drainage flow values for all drainage units, when assigning different parameter sets to each unit using the (a) BUCKET, (b) MODHMS-1D, and (c) MODHMS-3D models.](image-url)
APPLICATION OF THE SCEM-UA ALGORITHM

Compared to Case Study I, predictions of spatially distributed drainage data not only capture spatial patterns but perform much better in predicting temporal variations of individual drainage flows. In general, the MODHMS-3D model performs best and has the smallest RMSE. The larger RMSE of MODHMS-1D is the result of integrating boundary conditions (i.e. \( I \) and \( ET \)) to the entire drainage unit rather than modeling individual fields. At that reduced level of spatial resolution, the conceptual representation of the vadose zone by the BUCKET model exhibits fairly equal predictive capabilities. The benefits associated with the use of a three-dimensional physically-based hydrologic model seem, therefore, relatively marginal, even when comparing the temporal dynamic behavior between the three models. However, the BUCKET model is computationally by far the most efficient, requiring less than 1 minute to calibrate one drainage unit for one year of discharge data.

To gain more insights into the performance of each model, Figure 4.9 presents the discharge simulation for a set of four representative drainage units (ranging from good to bad fit), using the most likely parameter set for the BUCKET (solid line), MODHMS-1D (dashed line), and MODHMS-3D (dotted line) model.

![Figure 4.9](image1.png)

**Figure 4.9.** Case II-calibration: Drainage rates (mm week\(^{-1}\)) for a representative set of 4 drainage units (calibration case II) using the most likely parameter set for the BUCKET (solid line), MODHMS-1D (dashed line), and MODHMS-3D models (dotted line).
The numbering used in the top-right corner of each of the small plots in Fig. 4.9 refers to the adopted numbering and color coding in Fig. 4.2. Additionally, the top left corner of each plot presents summary statistics in terms of the RMSE of the residuals of the calibration period for each of the three models. Calibration summary statistics for all 24 drainage units, including those of Fig. 4.9, are listed in Table 4.5.

Table 4.5. Root Mean Square Error of the Drainflow Predictions of Each of the Numerical Models for the Different Drainage Units of the BWD. Cal: Calibration; Val: Validation.

<table>
<thead>
<tr>
<th>Drainage Unit</th>
<th>BUCKET</th>
<th>MODHMS-1D</th>
<th>MODHMS-3D</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Cal</td>
<td>Val</td>
<td>Cal</td>
</tr>
<tr>
<td>1</td>
<td>1.83</td>
<td>3.24</td>
<td>1.43</td>
</tr>
<tr>
<td>2</td>
<td>2.86</td>
<td>3.16</td>
<td>3.41</td>
</tr>
<tr>
<td>3</td>
<td>1.43</td>
<td>0.81</td>
<td>1.61</td>
</tr>
<tr>
<td>4</td>
<td>4.29</td>
<td>4.27</td>
<td>4.34</td>
</tr>
<tr>
<td>5</td>
<td>1.05</td>
<td>2.72</td>
<td>1.06</td>
</tr>
<tr>
<td>6</td>
<td>1.01</td>
<td>2.54</td>
<td>0.83</td>
</tr>
<tr>
<td>7</td>
<td>1.88</td>
<td>1.87</td>
<td>2.30</td>
</tr>
<tr>
<td>8</td>
<td>2.96</td>
<td>3.91</td>
<td>4.74</td>
</tr>
<tr>
<td>9</td>
<td>1.39</td>
<td>2.92</td>
<td>1.48</td>
</tr>
<tr>
<td>10</td>
<td>1.25</td>
<td>1.27</td>
<td>1.84</td>
</tr>
<tr>
<td>11</td>
<td>4.53</td>
<td>3.16</td>
<td>4.54</td>
</tr>
<tr>
<td>12</td>
<td>3.44</td>
<td>7.89</td>
<td>3.04</td>
</tr>
<tr>
<td>13</td>
<td>1.38</td>
<td>2.19</td>
<td>1.43</td>
</tr>
<tr>
<td>14</td>
<td>5.00</td>
<td>8.37</td>
<td>4.88</td>
</tr>
<tr>
<td>15</td>
<td>4.08</td>
<td>1.84</td>
<td>3.99</td>
</tr>
<tr>
<td>16</td>
<td>0.93</td>
<td>0.32</td>
<td>0.55</td>
</tr>
<tr>
<td>17</td>
<td>2.06</td>
<td>2.18</td>
<td>2.08</td>
</tr>
<tr>
<td>18</td>
<td>1.31</td>
<td>2.30</td>
<td>1.44</td>
</tr>
<tr>
<td>19</td>
<td>2.38</td>
<td>4.48</td>
<td>3.35</td>
</tr>
<tr>
<td>20</td>
<td>1.55</td>
<td>1.27</td>
<td>1.54</td>
</tr>
<tr>
<td>21</td>
<td>1.00</td>
<td>1.65</td>
<td>0.97</td>
</tr>
<tr>
<td>22</td>
<td>4.22</td>
<td>4.64</td>
<td>4.53</td>
</tr>
<tr>
<td>23</td>
<td>1.12</td>
<td>3.34</td>
<td>1.16</td>
</tr>
<tr>
<td>24</td>
<td>0.89</td>
<td>1.56</td>
<td>1.13</td>
</tr>
</tbody>
</table>

Results presented in Fig. 4.9 and Table 4.5 show that the MODHMS-3D model has the most consistent predictions (lowest RMSE values). We note, however, that the fit to the observed drainage data for each of the models is fairly good for most drainage units, especially in the light of the measurement uncertainties associated with the surface boundary conditions and drainage data at these spatial scales.

Since each drainage unit was treated as a separate simulation and parameter domain, we obtained 24 HDP plots similar to Fig. 4.5, however, for illustrative purposes we only show the results of Drainage Unit 2 in Figure 4.10. However, similar results were found for the other drainage units. Disaggregation of the water district into individual drainage units does not reduce
the amount of uncertainty associated with most of the parameters (compare Fig. 4.5 with Fig. 4.10) and in some cases even appears to increase parameter uncertainty.

![Normalized Parameter Range](image)

**Figure 4.10.** Case II-calibration: Normalized uncertainly plots for each of the parameters of the (a) BUCKET, (b) MODHMS-1D, and (c) MODHMS-3D models, constructed using the results for drainage unit 2. Each line across the graph denotes a single parameter set: gray is a member of the HPD region; the solid line presents the most likely parameter set.

The results for each drainage unit, suggest that the only parameters that are well identifiable and, as such, are warranted by the spatially distributed drainage data are the parameters $b$ and $f_b$ of the BUCKET model, and the drain conductance, $D_s$, and bypass fraction, $f_o$, of the MODHMS models. The optimized values for the bypass flow parameter $f_o$ for each of the different drainage units results in a contribution of quick-flow to total drainage flow that ranges between 5 and 30%.

Physical interpretation of the optimized hydraulic parameters of the MODHMS models is difficult since these represent effective properties that are not only a function of soil type, but also depend on the boundary conditions [Blöschl et al., 1995]. The large size of the HPD region of the scaling factor and saturated water content within the prior defined feasible parameter space and the resulting prediction uncertainty ranges of the soil water retention and unsaturated soil
hydraulic conductivity functions, confirm our earlier conclusion that the effective soil hydraulic properties are poorly identifiable. To explore whether the non-identifiability of the parameters in each of the models is caused by the poor quality of the models, the SCEM-UA derived parameter ranges for the scaling factors of both MODHMS models for all drainage units are presented in Figure 4.11.

Regardless of the quality of the fit, the uncertainty associated with the scaling factor is considerable as for most drainage units the HPD region of the parameter space occupies almost the entire prior defined parameter range.

Poorly constrained parameters may be caused by parameter correlation, poor model quality, and lack of parameter sensitivity as a result of insufficient information content [Yapo et al., 1996]. For example, the latter could be the case if the spatially distributed drainage data do not contain the hydrologic conditions required to properly identify the soil hydraulic parameters. None of the parameter interactions in the MODHMS models were large enough to support parameter interaction as a major reason for poor identifiability of the soil hydraulic properties. This is illustrated in Table 4.6, which presents the generally small-valued correlation coefficients between the SCEM-UA generated parameter samples of the HPD region of the parameter space for both the MODHMS-1D and 3-D models.
Table 4.6. Correlation coefficients between the SCEM-UA generated parameters in the HPD region of the parameter space for drainage unit 2 using the MODHMS-1D and MODHMS-3D model.

<table>
<thead>
<tr>
<th></th>
<th>MODHMS-1D</th>
<th></th>
<th>MODHMS-3D</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>χ</td>
<td>D&lt;sub&gt;1&lt;/sub&gt;</td>
<td>D&lt;sub&gt;2&lt;/sub&gt;</td>
</tr>
<tr>
<td>χ</td>
<td>1.00</td>
<td>0.47</td>
<td>0.46</td>
</tr>
<tr>
<td>D&lt;sub&gt;1&lt;/sub&gt;</td>
<td>1.00</td>
<td>0.14</td>
<td>-0.17</td>
</tr>
<tr>
<td>D&lt;sub&gt;2&lt;/sub&gt;</td>
<td>1.00</td>
<td>1.00</td>
<td>-0.45</td>
</tr>
<tr>
<td>θ&lt;sub&gt;in&lt;/sub&gt;</td>
<td>1.00</td>
<td>-0.17</td>
<td>-0.45</td>
</tr>
<tr>
<td>θ&lt;sub&gt;j&lt;/sub&gt;</td>
<td>1.00</td>
<td>-0.46</td>
<td>-0.27</td>
</tr>
<tr>
<td>f&lt;sub&gt;1&lt;/sub&gt;</td>
<td>1.00</td>
<td>-0.52</td>
<td>-0.42</td>
</tr>
<tr>
<td>f&lt;sub&gt;2&lt;/sub&gt;</td>
<td>1.00</td>
<td>-0.24</td>
<td>-0.48</td>
</tr>
</tbody>
</table>

While the listed results correspond to Drainage Unit 2 only, similar correlation matrices were found for the other drainage units.

Independent numerical optimizations using spatially distributed drainage data, generated from known parameter sets, confirmed that the measured spatially distributed drainage data contain only limited information content for the identification of the soil hydraulic properties. These results are not very surprising, as similar conclusions were drawn for the core-scale laboratory outflow experiments by Toorman et al. [1992] and Ebing and Hopmans [1993]. From these studies it also became clear that additional measurements within the soil compartment are needed to improve the identifiability of the soil hydraulic parameters. We note that limited parameter identifiability is also caused by the potential uncertainties of the boundary conditions.

To evaluate the consistency and reliability of the calibration results, the performance for each of the parameter sets in the HPD region for each of the three models was evaluated for the 1996 validation year. The most important results of this analysis are presented in Table 4.5 and in Figures 4.12, 4.13, and 4.14. Since most agricultural fields had different crops in 1996 (Table 4.1), the actual transpiration rates for the evaluation period were estimated by calibrating the crop coefficient adjustment factor, f<sub>r</sub>, during 1996 using the SCEM-UA algorithm, while keeping the other calibration parameters equal to their most likely values as estimated from the calibration. Adjusting the f<sub>r</sub> parameter slightly (typically between 5 and 20%) improved the results. The
drainage unit-average initial water content profiles for 1996 were obtained using the simulated water content distribution of the most likely parameter set at the end of the 1995 year.

While the predictive capabilities of the conceptual and physically-based models were similar for the 1995 calibration period, the 1996 validation results are slightly different, with the MODHMS-3D model consistently generating better forecasts (lowest RSME values) of measured drainage flow in most of the drainage units (Table 4.5 and Figs. 4.12 and 4.13).

![Figure 4.12](image)

Figure 4.12. Case II-validation: Drainage rates (mm/week) for a selection of 4 representative drainage units (validation, Case II) using the most likely parameter set for the BUCKET (solid line), MODHMS-1D (dashed line), and MODHMS-3D models (dotted line).

The good quality fit of the MODHMS models to the measured 1996 drainage flows demonstrates that the information content of a single year is sufficient to obtain reliable estimates for most model parameters. In contrast, the model performance of the BUCKET model is significantly less, when compared with the 1995 calibration period (Figs. 4.8 and 4.9), suggesting that a longer calibration period may be needed to obtain reliable estimates of the model parameters. It therefore appears advantageous to use physically-based hydrologic models at the temporal (week) and spatial (field) scales of this study.
APPLICATION OF THE SCEM-UA ALGORITHM

BUCKET MODEL

MODHMS-1D MODEL

MODHMS-3D MODEL

Simulated drainflow [mm week⁻¹]

Figure 4.13. Case II-validation: Scatterplot of simulated versus observed drain flow values for all drainage units, when assigning different parameter sets to each unit using the (a) BUCKET, (b) MODHMS-1D, and (c) MODHMS-3D models.

To further evaluate the predictive capability of the three calibrated models, we compared drainage unit average simulated groundwater table depths with their corresponding observed values (Figure 4.14). These latter values were obtained by arithmetically averaging groundwater table observations conducted within the same drainage unit.

Figure 4.14. Case II-cross validation: Scatterplot of drainage unit average simulated groundwater table depths versus observed values using the (a) MODHMS-1D and (b) MODHMS-3D models.

As these spatially-distributed water table depths have not been used during calibration, their agreement is a much stronger test of the internal consistency of the calibrated model, than when...
evaluating the performance of the model on measured drainage dynamics only. While considerable scatter around the 1:1 line is apparent (RMSE values are 59.11 and 79.89 cm for the MODHMS-1D and MODHMS-3D model, respectively), there is no systematic bias: The arithmetic mean of predicted water table depths is reasonably close to the arithmetic mean of the measured water table depths, suggesting that the calibrated model is suited to predict the various components of the water balance at the scale of drainage units. The RMSE values are large but may be explained by significant spatial variability of water table elevations within fields, especially when field drains are present and fields are irrigated in sections.

4.4. Summary and conclusions

The aim of the present Chapter was to explore the usefulness and applicability of the inverse method to estimate vadose zone properties at the water district scale by using spatially distributed drainage data from the Broadview Water District located in the San Joaquin Valley of California. Results demonstrate that measured spatially distributed patterns of drainage flux data contain only limited information for the specification of the vadose zone model parameter calibration parameters, and are particularly inadequate for the soil hydraulic properties at those scales. In part, the inability of the inverse modeling approach to identify soil hydraulic properties at large spatial and temporal scales is a consequence of estimation errors in the boundary conditions and their uncertainty. The study showed that the only parameters with relatively small uncertainties are related to drain conductance and bypass flow. This result indicates that these are the critical parameters controlling the drainage flow processes at the considered space (field/drainage system) and time (week) scales. Despite the large uncertainty of most calibration parameters, the fit to the observed drainage data for each of the models was fairly good for most drainage units. This is particularly true when considering the uncertainty in the drainage data and boundary conditions for these large time and space scales. We note especially that the predictive ability of the simple BUCKET model is about equal to that of the three-dimensional MODHMS model. Notwithstanding the significant CPU time needed for model calibration, there are advantages of using physically-based hydrologic models to study spatial and temporal patterns of water flow at larger spatial scales. These mechanistic models not only generate consistent forecasts of spatially-distributed drainage during both the calibration and validation periods, but simultaneously possess unbiased predictive capabilities of groundwater table depths. In a later paper we will extensively discuss the various trade-offs in the fitting of drainage flow, drainage salinity, and groundwater table depths, by posing the optimization problem into a multi-criteria
framework and solving for the Pareto set of solutions using the recently developed Multi-objective Shuffled Complex Evolution Metropolis algorithm (MOSCEM-UA) algorithm in Chapter 6.
Part III

Recursive Parameter Estimation
CHAPTER 5

Toward Improved Identifiability of Hydrologic Model Parameters:
The Information Content of Experimental Data

Abstract

We have developed a sequential optimization methodology, entitled the Parameter Identification Method based on the Localization of Information (PIMLI) that increases information retrieval from the data by inferring the location and type of measurements that are most informative for the model parameters. The PIMLI approach merges the strengths of the Generalized Sensitivity Analysis (GSA) method [Spears and Hornberger, 1980], the Bayesian Recursive Estimation (BaRE) algorithm [Thiemann et al., 2001], and the Metropolis algorithm [Metropolis et al., 1953]. Three case studies with increasing complexity are used to illustrate the usefulness and applicability of the PIMLI methodology. The first two case studies consider the identification of soil hydraulic parameters using soil water retention data and a transient multi-step outflow experiment (MSO), whereas the third study involves the calibration of a conceptual rainfall-runoff model.

5.1. Introduction and scope

Hydrologic models often contain parameters that cannot be measured directly, but can only be meaningfully inferred by calibration to a historical record of input-output data. In its most elementary form, calibration is performed manually by visually inspecting the agreement between observations and model predictions [Janssen and Henberger, 1995; Boyle et al., 2000]. More sophisticated approaches express the agreement or misfit between model and measurements quantitatively in terms of misfit measures and use an optimization algorithm to minimize this measure. The definition of these misfits should reflect the intended use of the model and should concern the model quantities, which are deemed important [Gupta et al., 1998]. The aim of these model calibration procedures is to reduce the uncertainty in the correct choice of the parameter values (parameter uncertainty) while simultaneously accounting for structural inadequacies in the model and uncertainties in the values of the measured input-output time series [Thiemann et al., 2001]. However, because hydrological models will, by nature, only render an approximate description of reality and because the data used for calibration contain measurement errors, estimates of parameters and dependent variables from these models are generally error-prone. That is, in most cases, there is not a single point in the parameter space associated with good simulations; in some cases, there may not even exist a well-defined region, in the sense of a compact region interior to the prior defined parameter space.

One of the serious limitations of classical automated optimization strategies, such as population-evolution-based search strategies [Brazíl and Krajewski, 1987; Brazíl, 1988; Wang, 1991; Duan et al., 1992; Sorooshian et al., 1993], the GLUE procedure [Beven and Binley, 1992; Freer et al., 1996], Monte Carlo membership procedure [Keesman, 1990; van Straten and Keesman, 1991], or the prediction uncertainty method [Klepper et al., 1991], is that they do not provide information about which sets of measurements are most informative for specific model parameters. Currently, such expertise is usually acquired by an individual only through extensive hands-on training (manual calibration) and experience with a specific model. As a trivial example, when using a manual calibration approach in rainfall runoff modeling, periods dominated by baseflow are used to estimate the baseflow recession rate parameter of the model. Clearly, increased information retrieval from the data about the type and location of the most informative measurements provides useful information for optimal experimental design or monitoring strategies. This serves as an attempt to obtain unique parameter estimates, being a prerequisite for finding pedotransfer functions [Schaap et al., 1998; Duan et al., 2001].
The purpose of this Chapter is to develop a sequential optimization algorithm, entitled the Parameter Identification Method based on the Localization of Information (PIMLI), to increase information retrieval from the data. The PIMLI algorithm is a hybrid approach that merges the strengths of the Generalized Sensitivity Analysis (GSA) method [Spear and Hornberger, 1980], the Bayesian inference scheme used in the Bayesian Recursive Estimation (BaRE) algorithm [Thiemann et al., 2001], and the sampling efficiency of the Metropolis algorithm [Metropolis et al., 1953] to select those sets of measurements that contain the most information for the identification of specific model parameters. As measurements with the highest information content for the various model parameters are recursively assimilated by the PIMLI algorithm, the uncertainty associated with the parameter estimates reduces.

This Chapter is organized as follows. Section 5.2 presents a general outline of the GSA method, the BaRE algorithm, and the Metropolis algorithm as foundations for the PIMLI algorithm. In section 5.3, three case studies are used to illustrate the usefulness and applicability of the PIMLI methodology. The first case study considers estimating hydraulic parameters in the soil water retention curve of van Genuchten [1980], whereas the second case considers a multi-step transient outflow experiment. In this section, we are especially concerned with the influence of the experimental range of measurements on the identifiability of the soil hydraulic parameters. Moreover, we discuss the weaknesses of the current parametric models of soil hydraulic properties. Finally, the third case study involves the calibration of a conceptual rainfall-runoff model. Section 5.4 presents a summary of the material presented in this Chapter.

5.2. Model calibration

The fundamental problem with which we are concerned is to estimate parameter values and their uncertainty from observed hydrological data (inputs) using a specified mathematical model that simulates actual input-output relations. Given a model structure, the uncertainty in the parameter estimates can be reduced by successively assimilating measurement data. Traditional methods of calibration lump all of the differences between model predictions and measurements in a residual vector $E(\mathbf{\beta}) = \{e_1(\mathbf{\beta}),...,e_k(\mathbf{\beta})\}$, without explicitly recognizing differences in model sensitivity of the model parameters to the various measurements. The purpose of this Chapter is to explicate a methodology entitled PIMLI to increase information retrieval from the data by identifying subsets of data that contain most information on the specific model parameters.
CHAPTER 5

The object of much experimentation is to study the relationship between a response or output variable $y$ subject to error and input variables. Following Thiemann et al. [2001], the hydrologic model $\eta$ can be written as:

$$j = \eta(\xi | \beta) + e$$  \hspace{1cm} (5.1)

where $j = (j_1, j_2, ..., j_s)$ is an $s \times 1$ vector of model predictions, $\xi = (\xi_1, \xi_2, ..., \xi_p)$ is an $n \times p$ matrix of input variables, $\beta = (\beta_1, \beta_2, ..., \beta_p)$ is the vector of $p$ unknown parameters, and $e$ is a vector of statistically independent errors with zero expectation and constant variance $\sigma^2$.

The aim of model calibration procedures is to reduce the uncertainty in the parameter values $\beta$ while simultaneously accounting for uncertainties in the measured input-output time series (data uncertainty) and uncertainties in the structural ability of the model, $\eta(\xi | \beta)$, to simulate the processes of interest (model uncertainty). We assume that the mathematical structure of the model is essentially predetermined and fixed. We begin by introducing a prior probability density distribution on the possible parameter sets and denoting this density $\pi(\beta)$. The purpose of the prior distribution is to quantify the initial uncertainty about the model parameters before any input-output data are collected. In order to avoid favoring any initial value, a uniform prior over the range of parameters may often seem reasonable [Beven and Binley, 1992].

5.2.1. The Generalized Sensitivity Analyses as first level in the PIMLI algorithm

The PIMLI procedure is schematically summarized in Figure 5.1. In view of the inevitably complicated nature of the hydrological model $\eta(\xi | \beta)$, it is evident that an explicit expression for the statistical distribution of the parameters is not possible. Instead, we use the power of Monte Carlo simulations to estimate measures of central tendency and dispersion for the various densities. For a prescribed number of Monte Carlo simulations $m$, we randomly sample a parameter set $\beta_j = (\beta_{1j}, \beta_{2j}, ..., \beta_{pj})$ from the prior probability distribution $\beta_j \leftarrow \pi(\beta)$ and generate the corresponding output $\hat{y}_j$. 
This results in an ensemble of $m$ models, each with structure $\eta(\xi | \beta)$ and with a parameter vector $\beta$, which is a random member of the distribution $p(\beta)$. The next step is to ascertain which elements of the parameter vector are able to mimic the important characteristics of the system being studied as reflected in the observations $y$. For this, a criterion function is needed (cf. 5.2.2) that classifies any $\beta_j$ for each single element of the vector $y$ as being a "good" or "bad" simulation. The partitioning of the parameter space into "good" or productive and "bad" or non-productive regions allows the application of a large variety of multivariate statistical analyses in exploring differences in the posterior distributions associated with good and bad simulations, or in exploring the structure induced into the joint distributions of parameters associated with good simulations. This part of the PIMLI algorithm is directly related to the GSA method [Spear and Hornberger, 1980; Spear et al., 1994].
5.2.2. Second level: the Bayesian Recursive Estimation algorithm

The PIMLI algorithm implements the Bayesian inference scheme (see section B of Fig. 5.1), recently used in recursive mode in the BaRE algorithm, as a criterion function to distinguish between “good” and “bad” simulations [Thiemann et al., 2001]. The Bayesian framework for statistical inference is suited to such problems, because it allows for the use of probability distributions to quantify the uncertainty in the parameters. Moreover, as Bayesian estimators properly represent measurement errors, the confidence intervals of the parameters can be evaluated formally [Hollenbeck and Jensen, 1998]. Assuming that measurement errors are mutually independent, the likelihood of parameter set \( \beta, L(\beta, \sigma, \gamma | r) \) for describing \( y \), can be calculated according to [Box and Tiao, 1973]:

\[
L(\beta, \sigma, \gamma | r) = \frac{\mu(\gamma)}{\sigma} \exp \left[ -c(\gamma) \left| \frac{p}{\sigma} \right|^{1+\gamma} \right] \tag{5.2}
\]

where

\[
r = G(j) - G(y) \tag{5.3}
\]

and

\[
c(\gamma) = \left\{ \frac{\Gamma[3(1+\gamma)/2]}{\Gamma[(1+\gamma)/2]} \right\}^{1/\gamma} \quad \mu(\gamma) = \frac{\left\{ \Gamma[3(1+\gamma)/2] \right\}^{1/2}}{(1+\gamma) \left\{ \Gamma[(1+\gamma)/2] \right\}^{3/2}} \tag{5.4}
\]

where \( y \) is the vector with observations, and \( \gamma \in (-1,1] \) is a fixed “shape parameter” that can be regarded as a measure of kurtosis, indicating the extent of the non-normality of the error density distribution. The density is normally distributed when \( \gamma = 0 \), double exponential when \( \gamma = 1 \), and tends to a uniform distribution as \( \gamma \to -1 \). The transformation \( G(\cdot) \) in Eq. (5.3) allows us to handle heteroscedastic and autocorrelated error cases in the residuals [e.g., Sorooshian and Dracup, 1980; Kuczera, 1988].

The PIMLI algorithm proceeds by computing the posterior density for each single parameter set \( \beta \) for each single measurement \( y \).
According to Eq. (5.5) the “best” parameter set depends on the set of measurements, because \( L(\beta_j, \sigma, y | v_i) \) will always emphasize the ability of a certain parameter set to reproduce the selected observations. If a formerly superior parameter set, indicated by its prior density \( p(\beta) \), cannot simulate the desired measurement \( y \), it will receive lower weight in the computation of the posterior density, hence allowing a shift in the estimated elements of \( \beta \). The posterior density, \( p(\beta, \sigma, y | v) \), associated with each parameter set \( \beta_j \), is then sorted in ascending order and used to compute the cumulative distribution function. Based on an appropriate percentile (e.g., 95% interval), the model and classification criteria now provide a means of dividing the hypercube for each single measurement into two regions, one associated with “good” simulations and the other one with “bad”. Finally, for the “good” simulations, the information content (IC) of each successive observation for each member \( \{\beta_1, \beta_2, \ldots, \beta_p\} \) of parameter set \( \beta \) is computed according to [see Vrugt et al., 2001a]:

\[
\text{IC}_{i,q} = 1 - \frac{\sigma_{\beta_i,q}^\text{post}}{\sigma_{\beta_i,q}^\text{prior}} \quad i = 1 \ldots n, \ q = 1 \ldots p
\]  

(5.6)

where \( \sigma_{\beta_i,q}^\text{prior} \) and \( \sigma_{\beta_i,q}^\text{post} \) denote the standard deviation of member \( q \) of parameter set \( \beta \) in the prior and posterior probability distribution, respectively. If, for each successive member of \( \beta \), the IC diagnostic is close to zero after processing measurement \( y \), this implies a lack of information in that particular observation for identification purposes. In contrast, if one of the parameters in the “good” parameter sets occupies a single well-defined region internally of its prior distribution, measurement \( y \) is informative for this parameter, thereby resulting in an IC value close to one.

5.2.3. Third level: The Metropolis algorithm for assessing parameter uncertainty

In the third step of the PIMLI procedure (part C of Fig. 5.1), the measurement with the highest information content is added within a Metropolis sampling framework to progressively resample the parameter space in the most promising region, while relinquishing occupations in the
nonproductive portions of the parameter space. This prevents the algorithm from collapsing to a single best parameter set by recursively assimilating more informative measurements, thus addressing a deficiency of BaRE developed by Thiemann et al. [2001]. The Metropolis algorithm has received considerable attention in the last decade in the Bayesian statistics literature and is closely related to the probabilistic optimization technique called simulated annealing [Metropolis et al., 1953; Kirkpatrick et al., 1983].

A Markov Chain Monte Carlo (MCMC) method for assessing parameter confidence intervals in nonlinear models is based on the idea that, rather than compute a probability density, $p(\beta)$, it is sufficient to have a large random sample from $p(\beta)$ that approximates the form of the density. Intuitively, if the sample were large enough, diagnostic statistical measures of the probability density function can be computed using the mean and standard deviation of the large sample. The most general and earliest MCMC algorithm, known as the Metropolis algorithm [Metropolis et al, 1953], is given as follows:

1. Randomly start at a location in the feasible parameter space, $\beta^{(0)} = \beta^0$, and compute the posterior density of this parameter set, $p(\beta | y)$, based on the set of measurements included in $y$ using Eqs. (2), (3), and (4).
2. Sample a new parameter set $\beta^{(t+1)}$ using the multi-normal distribution as proposal distribution:

$$\beta^{(t+1)} \leftarrow N(\beta^0, \Sigma_\beta)$$

(5.7)

where $\beta^0$ is the mean, $\Sigma_\beta$ is the covariance matrix of $\beta$, and $\epsilon$ is a scaling factor, typically in the range of 1-3 to ensure that one can sample from regions of $p(\beta | y)$ which are not adequately approximated by the multi-normal distribution in Eq. (5.7).
3. Calculate $p(\beta^{(t+1)} | y)$ and compute the ratio $\Omega = p(\beta^{(t+1)} | y) / p(\beta^0 | y)$.
4. Randomly sample a uniform label $Z$ over the interval 0 to 1.
5. If $Z \leq \Omega$, then accept the new configuration. However, if $Z > \Omega$, then remain at the current position, that is, $\beta^{(t+1)} = \beta^0$.
6. Increment $t$. If $t$ is less than a prespecified number of draws $N$, then return to step 2.

The Metropolis algorithm will always accept candidate points (jumps) into a region of higher posterior probability, but will also explore regions with lower posterior probability with
probability $\Omega$. This algorithm is a Markov Chain Monte Carlo sampler generating a sequence of parameter sets, $\{\beta^{(0)}, \beta^{(1)}, \ldots, \beta^{(t)}\}$, that converges to the stationary distribution, $p(\beta | y)$ for large $t$ [Gelman et al., 1997]. Thus, if the algorithm is run sufficiently long, the samples generated can be used to estimate statistical measures of the posterior distribution, such as mean, variance, etc. To speed up the convergence rate of the Metropolis sampler to the posterior target distribution, the covariance matrix of the proposal distribution, $\Sigma_\beta$, was periodically updated using a sample of the $\beta$s generated in the Markov Chain [Kucera and Parent, 1998]. A heuristic strategy based on running multiple sequences generated in parallel was used to test whether convergence of the Metropolis sampler to a stationary posterior target distribution has been achieved [Gelman and Rubin, 1992]. Moreover, the initial proposal distribution in Eq. (5.7) was approximated using the mean value and covariance structure induced in the “good” parameter sets derived from level 2.

After Metropolis sampling, the information content of the remaining $n-1$ measurements is computed, using the Bayesian inference scheme. The PIMLI procedure continues until all observations are selected for identification purposes and included in the objective function. We would like to emphasize that the PIMLI algorithm presented here differs from our previous methodology [Vrugt et al., 2001a] in the sense that simply one objective function is used to simultaneously identify all of the parameters. During the course of our investigations, we became aware that the explicit presence of strong parameter interdependence and model errors significantly lowers the chance of finding disjointive subsets of data each containing explicit information for just one of the model parameters.

5.3. Case studies

We illustrate the power and applicability of the PIMLI algorithm by means of three case studies. The first is a case study in which the four-parameter water retention function of van Genuchten [1980] is fitted to synthetic measurements. This illustrates the insights, which the PIMLI algorithm can offer with respect to parameter identifiability and optimal experimental design strategies. The second case study explores the utility of the PIMLI algorithm for identifying measurement sets that are most informative for the unsaturated soil hydraulic parameters, using a laboratory multi-step outflow experiment. To explore the applicability of the PIMLI algorithm in the presence of measurement and model errors, the last case study involves the calibration of a conceptual rainfall-runoff model using hydrologic data from the Leaf River basin near Collins, Mississippi. In all case studies, we assume a Gaussian error model (i.e., $\gamma = 0$) and a uniform
prior distribution $p(\beta)$ for each of the parameters. Based on our previous work, the jump rate was fixed to 0.5 [Vrugt and Bouten, 2002].

### 5.3.1. Case study I: Water retention model

One of the most commonly used models of capillary pressure-saturation is the water retention function of *van Genuchten* [1980]:

$$
\theta(\psi) = \theta_r + (\theta_f - \theta_r)[1 + (\alpha |\psi|)^n]^{-\frac{1}{n-1}}
$$

(5.8)

in which $\theta$ ($L^3$) denotes water content, $\theta_f$ and $\theta_r$ ($L^3$) are the water contents at full and residual saturation, respectively, $\psi$ ($L$) is the soil water pressure head, $\alpha$ ($L^{-1}$) is related to the inverse of the air-entry value, and $n$ is a unitless pore-size distribution index. Because, $\theta_f$, $\theta_r$, $\alpha$, and $n$ are unknown model parameters, their values must be estimated through calibration. Synthetic $(\psi, \theta)$ measurements were generated by using the parameters $\theta_f$, $\theta_r$, $\alpha$, and $n$ given in Table 5.1 for a sandy soil and by evaluating Eq. (5.8) for a given set of pressure heads.

<table>
<thead>
<tr>
<th>Par.</th>
<th>Unit</th>
<th>Sand†</th>
<th>Min.</th>
<th>Max.</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\theta_r$</td>
<td>[m$^3$ m$^{-3}$]</td>
<td>0.353</td>
<td>0.300</td>
<td>0.600</td>
</tr>
<tr>
<td>$\theta_f$</td>
<td>[m$^3$ m$^{-3}$]</td>
<td>0.003</td>
<td>0.000</td>
<td>0.000</td>
</tr>
<tr>
<td>$\alpha$</td>
<td>[m$^{-1}$]</td>
<td>2.122</td>
<td>0.100</td>
<td>5.000</td>
</tr>
<tr>
<td>$n$</td>
<td>[-]</td>
<td>5.665</td>
<td>1.050</td>
<td>10.00</td>
</tr>
<tr>
<td>$\theta_f$</td>
<td>[m d$^{-1}$]</td>
<td>0.500</td>
<td>0.01</td>
<td>1.00</td>
</tr>
<tr>
<td>$l$</td>
<td>[-]</td>
<td>0.039</td>
<td>-2.00</td>
<td>2.00</td>
</tr>
</tbody>
</table>

† Adapted from *Vrugt et al.* [2001a]

For this, the logarithmic pressure head was discretized into 50 equidistant points between $\psi = -0.01$ and $\psi = -10^5$ m. Because the exact results are known, this set of water retention observations serves as a way of demonstrating the usefulness and applicability of the PIMLI algorithm. In addition, the synthetic "true" water content measurements were corrupted with error to investigate the effects of data error on the selection procedure. Data error may be regarded as representing the combined effects of measurement error and soil nonhomogeneity, which are not represented in the model. We assume here that this data error can be represented
as a normally distributed $N(0, \sigma^2)$ error term, which is added to the exact water contents [Kool and Parker, 1988]. The error standard deviation for water content is taken to be 0.005 ($L^3 L^{-3}$). Figure 5.2 shows the original generated water content measurements as well as the effect of the data error on these synthetic water content measurements for the sandy soil in the range $0 \leq \text{pF} \leq 4$.

![Figure 5.2. Water retention curve of the sandy soil and the effect of data error.](image)

The feasible parameter space for the parameters $\theta_s$, $\theta_i$, $\alpha$, and $n$ was defined by the bounds given in Table 5.1. The results derived with the PIMLI algorithm for both the uncorrupted and corrupted error case are discussed below.

5.3.1.1. Uncorrupted water retention measurements

In the case of no data error, the locations of the four most informative water content measurements identified and processed using the PIMLI algorithm, as being associated with each parameter, are indicated in Figure 5.3a. The sequential order in which the measurements are recursively identified and assimilated with the algorithm is indicated in the graph. Additionally, Figure 5.3b illustrates the general behavior of parameter sensitivity for the sandy soil in the VG model. To allow comparison of sensitivity coefficients between different parameters, the Jacobian elements were scaled to the parameter's prior uncertainty ranges defined in Table 5.1.
Figure 5.3. (a) Uncorrupted error case: Location of the 4 most informative water retention measurements along the curve identified using the PIMLI algorithm, (b) Behavior of sensitivity of water content to the water retention parameters $\theta_s$, $\theta_r$, $\alpha$, and $\gamma$ over the prior defined range of pressure head values for the sandy soil in the VG model.

Moving from full to residual saturation, Fig. 5.3a demonstrates that the most informative water content measurements are first found for $\theta$, close to saturation, then for $\alpha$ close to the air-entry value of the soil, for $\gamma$ at the inflection point in the low water content range, and finally for $\theta$, close in the low water content range, respectively. The location of these most informative measurements matches very well with the general behavior of parameter sensitivity, as illustrated in Fig. 5.3b. Seemingly, some small deviations are found between the exact location of the most informative measurements along the curve, depicted in Fig. 5.3a, and judgments about these locations based on the marginal sensitivity coefficients only. For instance, the PIMLI algorithm identifies a moisture content measurements at $pF = 1.5$ as most informative for the parameter $\alpha$, whereas Fig. 5.3b suggests that the sensitivity to this parameter is larger for moisture content measurements at $pF = 1.75$. This discrepancy is attributed to parameter interdependence, as closer inspection of Fig. 5.3b demonstrates that, at $pF = 1.5$, the overall sensitivity of the VG model to $\alpha$, being $|\partial \theta / \partial \alpha| - |\partial \theta / \partial n|$ is at its maximum.

Besides the exact location of most informative measurements, the hierarchical selection order of the measurements is also strongly determined by the interdependency among the retention parameters. For instance, using the first-order sensitivity coefficients, it is expected that the PIMLI algorithm will first identify measurements close to $pF = 1.75$ as most informative for the parameter $\alpha$ because the absolute magnitude of the sensitivity coefficients is largest at this pressure head and associated with the retention parameter $\alpha$. Again, the explicit presence of parameter interdependence in the range of $pF = 1.3$ increases the extent of the uncertainty.
associated with the parameter $\alpha$. Consequently, because volumetric water content measurements close to saturation ($pF \leq 1$) and in the low water content range ($pF > 2.5$) are almost solely sensitive to either the saturated or residual water content, these measurements are first selected to constrain the saturated and residual water content in the sampling. Once the first two measurements are successively assimilated and the uncertainty associated with the parameters $\theta_s$ and $\theta_r$ diminishes, the information for the retention parameters $\alpha$ and $n$ appears from the water content measurements in the intermediate water content ranges.

In Figure 5.4, we present the relative size of the high probability density (HPD) region, including the information content diagnostic for each of the retention parameters as functions of the number of water content measurements used for identification purposes.

**Figure 5.4.** Evolution of the Bayesian confidence intervals (---) and information content diagnostic (•) of the VG retention parameters for the sandy soil.

Because only one measurement is recursively added in the likelihood function (Eq. (5.2)) at a time, the value of the iteration number is equivalent to the number of measurements used for parameter identification. Fig. 5.4 demonstrates that, as more and more measurements are assimilated with the PIMLI algorithm, the HPD region tends to concentrate on the most likely parameter values. Note, however, that the size of the HPD region will remain of finite size because we included the predefined water content data error term in Eq. (5.2). When processed
efficiently, only a relatively small amount of measurements are needed for a reliable identification for each of the model parameters. The PIMLI algorithm pinpoints the true parameter values for the perfect data case after processing approximately five water content measurements. Clearly, there is only a marginal improvement in the size of the 95% Bayesian confidence interval, centered on the “most likely” parameter set, between using the five most informative measurements and all 50 measurements. Seemingly, approximately 90% of the data contains information that is largely redundant for parameter identification purposes. It is not the “amount” of data that discriminates between parameter sets, but the information content of the data and the efficiency with which that information is extracted during model calibration [e.g., Kuczera, 1982; Sorooshian et al., 1983; Gupta and Sorooshian, 1985]. The reason for the jumping behavior of the confidence limits for the parameter \( \theta \), shown in Fig. 5.4d, is that, to maintain computational efficiency, the Metropolis sampler was not allowed to densely resample the parameter space. Consequently, the exact Bayesian confidence limits are, to a certain extent, subject to the stochastic properties of the Metropolis sampler.

The results presented in this case study illustrate two important findings. The first is that the newly developed sequential optimization methodology correctly infers the location and hierarchical selection order of the most informative measurements, without having apriori knowledge of the system properties (i.e., most optimal parameter set). The latter is a necessity to be able to draw these conclusions utilizing the first-order approximation. Another finding is that, in the case of perfect data, only a very limited amount of water content measurements is needed for a reliable estimation of the retention parameters.

### 5.3.1.2. Corrupted water retention measurements and experimental design

A more realistic case study is one in which the water retention observations are corrupted with data error. For this case, the four most informative water content measurements, identified using the PIMLI algorithm as being associated with each parameter, are indicated in Figure 5.5a. For illustrative purposes, we have switched the \( x \)- and \( y \)-axes of the retention characteristic. The striking resemblance between the Figs. 5.3 and 5.5a demonstrates that the PIMLI algorithm is able to successfully infer the location of the most informative water content measurements in the presence of data error. Although not shown here, the data error resulted in slower convergence of the HPD region into the vicinity of the true parameter values used to generate the synthetic observations of the sandy soil in the forward mode.
The Bayesian confidence intervals of the parameters reached a constant width after processing eight water content measurements.

To further investigate the influence of the experimental range of pressure heads on the final uncertainty of the retention parameters, we performed the following experiment. Starting at $\psi = -10^4$ m (location I in Fig. 5.5), two subsequent retention observations were omitted at each step from the original set of corrupted water content measurements, resulting in 25 data sets with varying numbers of observations \{50, 48, 46, ..., 2\} thereby, varying experimental pressure head range before arriving at full saturation. These data sets were used as input for the PIMLI algorithm to assess the corresponding uncertainty of the retention parameters. Fig. 5.5a illustrates how the size of the HPD region for each of the parameters depends on the experimental range of pressure heads. To allow comparison between uncertainties of different parameters, the HPD region was scaled according to the prior uncertainty bounds of the parameters, defined in Table 5.1 to yield normalized ranges between 0 and 1. Omitting retention measurements in the dry water content range between $-10^2 \leq \psi [m] \leq -10^3$ does not significantly affect the size of the confidence intervals of the parameters. Thereafter, the uncertainty

Figure 5.5. (a) Corrupted error case: Location of the 4 most informative water retention measurements along the curve identified using the PIMLI algorithm, (b) Normalized range of each of the retention parameters as function of the experimental range of pressure heads. For more explanation, see text.
associated with the parameters increases, reflecting the fact that the information for the parameters $\theta$, $n$, and $\alpha$ (in following order) is then starting to appear outside of the experimental range of pressure heads. Notice the striking resemblance in the shape of uncertainty curves depicted in Fig. 5.5b and the functional shape of the water retention characteristic. Finally, when arriving at full saturation, the saturated water content remains well identifiable, while the normalized ranges of the other parameters then reflect their prior uncertainty bounds. This indicates the lack of information in retention measurements close to saturation for the residual water content and curve shape parameters $\alpha$ and $n$. In the interest of brevity, we have only presented graphical results for the sandy soil. For more fine-textured soils, the curves depicted in Fig. 5.5a exhibit a more sigmoid shape reflecting the shape of the water retention characteristics for those kinds of soils. In addition, with increasing water-holding capacity of the soils, the different curves shifted towards lower pressure head values.

The results presented here illustrate that the PIML algorithm provides valuable information about the location and hierarchical selection order of the most informative measurements and can also be used to evaluate different experimental design strategies for their suitability for identifying the hydraulic parameters. The method, therefore, supports the identification of unique parameter values (preferably having a small variance) as a prerequisite for finding pedotransfer functions.

5.3.2. Case study II: Synthetic multi-step outflow experiment

A more demanding test of the PIML algorithm was devised by applying the algorithm to a synthetic, laboratory Multi-Step Outflow experiment. The two basic soil hydraulic characteristics controlling flow in unsaturated porous media are the retention characteristic, $\Theta(\psi)$ in Eq. (5.8), and the unsaturated hydraulic conductivity characteristic, $K(\psi)$:

$$K(\psi) = K_s \left[ 1 - \frac{(\alpha|\psi|)^{1-\gamma}}{1 + (\alpha|\psi|)^{\gamma}} \left( 1 + (\alpha|\psi|)^{\gamma} \right)^{\gamma^{-1}} \right]^{-2}$$

(5.9)

where $K_s$ (L.T$^{-1}$) is the saturated hydraulic conductivity, and $\gamma$ is an additional parameter related to the exponent parameter in Mualem [1976]. Because direct methods for determination of the $\Theta(\psi)$ and $K(\psi)$ curve require rather restrictive initial and boundary conditions and are therefore
RECURSIVE PARAMETER ESTIMATION

relatively tedious, numerical inversion is an attractive alternative for determining both curves from a single experiment. In such an approach, the optimal soil hydraulic parameters are found by fitting a numerical solution of Richards’ equation to observations of measured variables during the experiment. When a joint parametric description of retention and unsaturated hydraulic conductivity is assumed, inversion of the Richards equation will yield parameter estimates that apply to both characteristics simultaneously.

Synthetic outflow measurements were generated for a soil core with a height of 5 cm. In keeping up with the previous case study, hydraulic properties were used that correspond to the sandy soil previously defined in Table 5.1. The HYDRUS-1D software package [Simunek et al., 1998b] was used to numerically solve Richards’ equation in one dimension using a Galerkin-type linear finite element scheme. As the initial condition, hydraulic equilibrium was assumed with a pressure head \( \psi = -0.01 \) m at the bottom of the soil core, and \( \psi = -0.06 \) m at the top. The following pressure head steps and time periods (in brackets) were applied in the numerical experiment, \( \psi_1 = -0.0030 \) m \((0 \leq 0.50 \) d), \( \psi_2 = -0.15 \) m \((0.50 \leq 1.50 \) d), \( \psi_3 = -0.50 \) m \((1.50 \leq 3.50 \) d), \( \psi_4 = -1.00 \) m \((3.50 \leq 5.50 \) d), \( \psi_5 = -3.00 \) m \((5.50 \leq 12.50 \) d), and \( \psi_6 = -5.00 \) m \((12.50 \leq 20.00 \) d). Subsequently, the unsaturated hydraulic properties were inversely estimated using the PIMLI algorithm in combination with the HYDRUS-1D code. The soil hydraulic parameters \( \theta, \theta_r, \alpha, n, K_r, \) and \( l \) were assumed to be unknown and having the prior uncertainty ranges defined in Table 5.1.

Based on our earlier work [Vrugt et al., 2001a], we treated measured cumulative outflow, its first derivative (flux density), and the average water content in the soil core, deduced from observed cumulative outflow dynamics and the water content at the beginning of the experiment, as three separate measurement sets. The measurement error of the outflow was set to 0.05 cm\(^3\), which is identical to the accuracy of pressure transducers that are used for automated monitoring of the outflow dynamics in the laboratory. Subsequently, the measurement errors of the water content and flux density data sets were derived according to [Vrugt et al., 2001a] assuming an error in initial soil water content of 0.01 [m\(^3\) m\(^{-3}\)]. The results obtained with the PIMLI algorithm using the synthetic outflow observations are discussed below. Moving from full to residual saturation during the outflow experiment, the location of the six most informative outflow observations for the various model parameters within the different “artificial” measurement sets is illustrated in Figure 5.6.
CHAPTER 5

Figure 5.6. Synthetic outflow experiment: Location of the 6 most informative outflow observations for the various model parameters within the different measurement sets.

The sequential selection order in which the measurements are recursively identified and assimilated with the PIMLI algorithm is indicated in the graph. Fig. 5.6 illustrates that water content measurements at hydraulic equilibrium close to saturation are most informative for the saturated water content and therefore included in Eq. (5.5) to identify $\theta_s$. Additionally, cumulative outflow measurements at hydraulic equilibrium close to the air-entry value of the soil in the intermediate water content ranges and close to residual saturation are most informative for the curve shape parameters $\alpha$ and $n$ and the residual water content, respectively. Finally, flux density measurements immediately after passing the air-entry value of the soil are most informative for the saturated hydraulic conductivity. No informative measurements in this experiment can be found for the pore connectivity parameter $\lambda$ because the information content diagnostic for this parameter exhibits relatively low values, implying a lack of information. Instead, measurements close to saturation, being associated with $\theta_s$, are recursively assimilated and processed with the PIMLI algorithm.

There is a striking similarity in the location and hierarchical selection order of the most informative measurements for the parameters $\theta_s$, $\theta_r$, $\alpha$, and $n$ found under transient conditions and similar results presented for the static retention characteristic (case study #1). In both cases,
the most informative measurements for the saturated and residual water content are found at hydraulic equilibrium close to full and residual saturation, respectively. Additionally, the most information for α and n is located during hydraulic equilibrium conditions close to the air-entry value of the soil and in the intermediate water content range, respectively. Seemingly, the joint appearance of the parameter n in the retention and unsaturated soil hydraulic characteristic hardly affects the location of the most informative measurements being associated with this parameter during transient conditions.

Although not presented for this case study, the evolution of the Bayesian confidence intervals of the parameters demonstrated that the PIMLI algorithm was able to exactly pinpoint the parameter values of the sandy soil used to generate the synthetic measurements after processing only 15 outflow measurements. There was only a marginal improvement in the size of the confidence intervals between using the 15 most informative measurements and 225 synthetic measurements. Also under transient conditions, approximately 95% of the “synthetic” measurements contained redundant information for parameter identification purposes.

5.3.3. Discussion

An important finding, illustrated by Figs. 5.3, 5.5, and 5.6, is that for more fine-textured soils, problems occur with the simultaneous identification of the parameters θ, and n, because conventional laboratory experiments generally yield a pressure head range between 0.1 and 10 m (1 < pF < 3). Unless augmented with laboratory measurements at relatively low water contents, the most informative measurements for the parameters θ, and n are then beyond the experimental range and estimation of these parameters is then based primarily on extrapolation. This leads to high correlation between estimates of θ, and n, which is often graphically found in response surfaces. Consequently, the inverse problem is ill-posed, especially in the case of measurement and model errors. There have been numerous reports in the literature of problems with the identification of the parameters θ, and n on the basis of water content or pressure head data [see e.g., van Genuchten, 1980; Parker et al., 1985; Šimůnek et al., 1998a; Vrugt et al., 2001a]. To avoid identifiability problems for θ, and n, it is important to have some independent data or procedure for estimating the residual water content.

Although the inclusion of a residual water content in the description of a water retention curve is physically correct [Brooks and Corey, 1964; van Genuchten, 1980; Kosugi, 1996], its value is controversial [Luckner et al., 1989] and not supported by experimental data [Rossi and Nimmo, 1994] and its value is ill-determined from a inverse point of view as the most informative
measurements for this parameter are, for most soils, usually far outside the measurement range. If in such cases \( \theta \) is treated as an unknown parameter it strongly enhances the likelihood of non-uniqueness in the inverse problem. Lack of data at the dry end will in any case, cause the optimized value of \( \theta \), to be that of a fitting parameter without any physical significance [Vereecken et al., 1997]. Identifiability problems for \( \theta \), can be eliminated if one would independently measure \( \theta_a \) or if one would use a water retention model, which does not explicitly contain a residual water content in the model formulation [Rossi and Nimmo, 1994]. This is an attractive alternative, as its suitability for characterizing the \( \theta(\psi) \) and \( K(\psi) \) curves is acceptable [Rossi and Nimmo, 1994], as compared to the models which explicitly contain \( \theta \), in their model formulation [e.g. Brooks and Corey, 1964; van Genuchten, 1980; Russo, 1988; Kosugi, 1996].

5.3.4. Case study III: Application to conceptual rainfall-runoff model

The third study presented in this Chapter explores the usefulness of the PIMLI algorithm to the identification of parameters in a conceptual rainfall-runoff (CRR) model. The illustrative study here involves calibration of the HYMOD conceptual watershed model using approximately 1 year of historical data (28 July 1952-30 September 1953) from the Leaf River watershed near Collins, Mississippi. The HYMOD model (see Figure 5.7), first introduced by Boyle [2001], consists of a relatively simple rainfall excess model, described in detail by Moore [1985], connected with two series of linear reservoirs (three identical quick and a single reservoir for the slow response) and requires the optimization of five parameters to observed streamflow data: the maximum storage capacity in the catchment, \( C_{\text{max}} \) (L), the degree of spatial variability of the soil moisture capacity within the catchment, \( b_{\text{exp}} \) (-), the factor distributing the flow between the two series of reservoirs, \( \text{Alpha} \) (-), and the residence time of the linear quick and slow reservoirs, \( R_q \) (T) and \( R_s \) (T), respectively.

Figure 5.7. Schematic overview of the HYMOD conceptual watershed model.
Because the HYMOD model and Leaf River data have been discussed extensively in previous work [Sorooshian et al., 1993; Duan et al., 1993, 1994; Boyle, 2000; Wagener et al., 2001], the details will not be presented here.

To investigate whether the location of the most informative streamflow measurements in the hydrograph is influenced by the presence of errors (measurement and model) in the observed hydrological data two experiments were performed. In the first experiment, the PIMLI algorithm was applied to synthetic daily streamflow data (in m$^3$/s), generated for the period 28 July 1952 to 30 September 1953, by driving HYMOD with mean areal boundary conditions for the selected 1-year period (rainfall and potential evapotranspiration) and fixed values for the parameters. The parameter values were set identical to their most optimal value derived when separately fitting the HYMOD model to the observed streamflows for the selected period using the SCE-UA global optimization algorithm [Duan et al., 1992, 1993] and the RMSE criterion. In the second experiment the PIMLI algorithm was applied to the observed streamflows of the Leaf River watershed for the approximately one-year period under consideration. In both experiments, the measurement error of the streamflow data was fixed to the RMSE value of the most optimal parameter set obtained using the SCE-UA global optimization algorithm. Further, it was assumed that the output errors have a heteroscedastic variance that is related to flow level and which can be stabilized by the Box-Cox transformation using $\lambda = 0.3$ [e.g., Sorooshian and Dracup, 1980; Thiemann et al., 2001]. To reduce sensitivity to state value initialization, a 65-day warm-up period was used during which no updating of the posterior density functions was performed. The results for each of the experiments are illustrated in Figure 5.8 and discussed below.

Fig. 5.8a presents the synthetic streamflow observations in the transformed space, while Fig. 5.8b shows the lumped dynamics of soil-moisture storage within the watershed corresponding to the true parameter set for a portion of WY 1952. The open circles correspond to the location of the most informative streamflow measurements, whereas the number between parentheses refers to the hierarchical selection order. The results presented in the Figs. 5.8a and b illustrate that the most informative streamflow measurements for the parameters $R_0$ and $R_v$ are located after the cessation of rainfall in the non-driven quick and non-driven slow parts of the hydrograph, respectively.
Closer inspection of the information content diagnostic for these parameters demonstrated that the information for these parameters is typically independent of the storage in the slow, quick, and soil-moisture tank. On the contrary, Fig. 5.8b demonstrates that the location of the most informative streamflow measurements for the parameter $C_{\text{max}}$ in the hydrograph is essentially dependent on the current state of the soil-moisture. When the conceptual soil compartment is saturated with water, then the identifiability of the $C_{\text{max}}$ parameter is at its maximum. Additionally, streamflows occurring at intermediate soil-moisture storages in the catchment contain the most information for the $b_{\exp}$ parameter. Unfortunately, even after processing the five most informative measurements, no direct information was found for the parameter $\text{Alpha}$ that partitions the streamflow into a quick and slow-flow component. Although not presented here, the HPD region converged readily into the vicinity of the true parameter values, after recursively assimilating and processing the 10 most informative streamflow measurements. The results for the observed streamflow data (HYMOD results indicated with solid line), presented in Figs. 5.8c-
demonstrate that there is a close correspondence in the location and hierarchical selection order of the HYMOD parameters found for the synthetic and observed streamflow measurements. The explicit presence of various sources of errors in the observed streamflows does not seem to significantly affect the location of the most informative streamflow measurement in the hydrograph for the various model parameters.

Finally in Figure 5.9, we present the evolution of the HPD region (light gray) and most likely parameter set (dark line) as functions of the number of measurements processed using the PIMLI algorithm.

Figure 5.9. Evolution of the 95% Bayesian confidence intervals (gray area) of the different HYMOD model parameters as a function of the number of streamflow measurements assimilated and processed with the PIMLI algorithm. The dashed lines indicate the most likely parameter values, whereas the asterisks denote the most likely parameter set derived when separately fitting the HYMOD model to the entire 1-year period of observed streamflow data using the SCE-UA algorithm.

The parameters were scaled according to their prior uncertainty ranges to yield normalized ranges. The asterisks correspond to the most likely parameter values derived using the SCE-UA global optimization algorithm. Without any calibration, the parameter ranges reflect their initial uncertainty, not conditioned on any input-output time series of measured streamflows. After processing the first 6 most informative streamflow measurements with the PIMLI algorithm, the
HPD region narrows down rather quickly for most of the parameters. The characteristic jumping behavior of the HPD region throughout the feasible parameter space is caused by the presence of structural inadequacies in the HYMOD model and errors in the hydrological data. After recursively assimilating a sufficient amount of streamflow measurements with the PIMLI algorithm (approximately 15), the HPD region of the parameters finally settles down in the parameter space. The results in Fig. 5.9 illustrate that the HYMOD model parameters are reasonably well determined by calibration to streamflow data. Note also the excellent correspondence between the most optimal parameter values identified using a conventional batch calibration approach (SCE-UA) for the entire 1-year period and the location of the HPD region derived with the PIMLI algorithm after processing only the 20 most informative streamflow measurements. It appears that the use of disjunctive data regions that contain the highest information content for the parameters results in a reliable calibration of the HYMOD model, using only a very limited amount of streamflow measurements.

5.4. Summary and conclusions

This Chapter has discussed a practical methodology entitled the PIMLI algorithm that merges the strengths of the Generalized Sensitivity Analyses, BaRE algorithm, and Metropolis algorithm to identify sets of measurements that contain the most information for the identification of the model parameters. Three case studies with increasing complexity were used to illustrate the applicability and usefulness of the PIMLI methodology. The first two case studies considered estimating soil hydraulic parameters using soil water retention data and a more complex multi-step transient outflow experiment. Finally, the third case study involved the calibration of a simple conceptual rainfall-runoff model using synthetic and observed streamflow data from the Leaf River watershed near Collins, Mississippi. The most obvious advantage of the PIMLI algorithm is that it infers the type and location of the most informative measurements being associated with the different hydrologic parameters without having apriori knowledge of the system properties or most likely parameter set. The case study results suggest that the algorithm is well suited to identify which parameters control what part of model behavior, while providing insight into the importance of different kinds of data in experimental procedures. Finally, we believe that the PIMLI algorithm can be a valuable tool to help find complementary non-commensurable measures for multi-criteria parameter estimation schemes, with each criterion being sensitive to a different parameter or aspect of model behavior.
Part IV

Multi-objective Parameter Estimation
CHAPTER 6

Effective and Efficient Algorithm for Multi-Objective Optimization
of Hydrologic Models

Abstract

Practical experience with the calibration of hydrologic models suggests that any single objective function, no matter how carefully chosen, is often inadequate to properly measure all of the characteristics of the observed data deemed to be important. One strategy to circumvent this problem is to define several optimization criteria (objective functions) that measure different (complementary) aspects of the system behavior and to use multi-criteria optimization to identify the set of non-dominated, efficient, or Pareto optimal solutions. In this Chapter, we present an efficient and effective Markov Chain Monte Carlo sampler, entitled the Multi-Objective Shuffled Complex Evolution Metropolis (MOSCEM-UA) algorithm, which is capable of solving the multi-objective optimization problem for hydrologic models. MOSCEM-UA is an improvement over the Shuffled Complex Evolution Metropolis (SCEM-UA) global optimization algorithm, using the concept of Pareto dominance (rather than direct single objective function evaluation) to evolve the initial population of points toward a set of solutions stemming from a stable distribution (Pareto set). The efficacy of the MOSCEM-UA algorithm is compared with the original MOCOM-UA algorithm for three hydrologic modeling case studies of increasing complexity.

\[1\text{This Chapter has been published by: Jasper A. Vrugt, Hoshin V. Gupta, Luis A. Bastidas, Willem Bouten, and Soroosh Sorooshian, Effective and efficient algorithm for multi-objective optimization of hydrologic models, Water Resources Research, 39(8), 1214, doi:10.1029/2002WR001746, 2003. © Reprinted with permission of the American Geophysical Union.}\]
6.1. Introduction and scope

Many hydrologic models contain parameters that cannot be measured directly, but must be estimated through indirect methods such as calibration. In the process of calibration, the hydrologist adjusts the values of the model parameters such that the model is able to closely match the behavior of the real system it is intended to represent. In its most elementary form, this calibration is performed by manually adjusting the parameters while visually inspecting the agreement between observations and model predictions [Janssen and Heuberger, 1995]. In this approach, the "closeness" of the model with the measurements is typically evaluated in terms of several subjective visual measures, and a semi-intuitive trial-and-error process is used to perform the parameter adjustments [Boyle et al., 2000]. Because of the subjectivity and time-consuming nature of manual trial-and-error calibration, there has been a great deal of research into the development of automatic methods for calibration of hydrologic models [e.g., Gupta and Sorooshian, 1994]. Automatic methods seek to take advantage of the speed and power of digital computers, while being objective and relatively easy to implement.

Research into the development of automatic calibration methods has focused mainly on the selection of a single objective measure and the selection of an automatic optimization strategy that can reliably optimize (maximize or minimize, as appropriate) that measure. In this regard, the Shuffled Complex Evolution (SCE-UA) global optimization algorithm developed by Duan et al. [1992, 1993] has proved to be consistent, effective, and efficient in locating the parameter values of a hydrologic model that optimize a given objective function. However, practical experience with the calibration of hydrologic models suggests that single objective functions, no matter how carefully chosen, are often inadequate to properly measure all of the characteristics of the observed data deemed to be important. Consequently, single objective calibration methods do not usually provide parameter estimates that are considered acceptable by practicing hydrologists. Another emerging problem is that many of the latest hydrologic watershed or land-surface models simulate several output fluxes (e.g., water, energy, chemical constituents, etc.) for which measurement data are available, and all these data must be correctly utilized to ensure proper model calibration [Beven and Kirkby, 1979; De Grosois et al., 1988; Kurczer, 1982, 1983; Woolhisser et al., 1990; Kuczer and Mroczkowski, 1988; Gupta et al., 1998]. One strategy to explicitly recognize the multi-objective nature of the calibration problem is to define several optimization criteria (objective functions) that measure different (complementary) aspects of the system behavior and to use a multi-criteria optimization method to identify the set of non-dominated, efficient, or Pareto optimal solutions [Gupta et al., 1998; Yapo et al., 1998; Boyle et al.,
MULTI-OBJECTIVE PARAMETER ESTIMATION

The Pareto solutions represent tradeoffs among the different incommensurable and often conflicting objectives, having the property that moving from one solution to another, results in the improvement of one objective while causing deterioration in one or more others.

A simple way to obtain a crude approximation of the Pareto solution set is to weigh the different criteria into a single aggregated scalar and to run a large number of independent single criteria optimization runs using different values for the weights. Popular aggregation methods include the weighted-sum approach, target vector optimization, and the method of goal attainment [Srinivas and Deb, 1994; Fonseca and Fleming, 1995]. Recently, Madsen [2000] used an aggregation approach, in combination with the single-criterion SCE-UA global optimization algorithm, to construct an estimate of the Pareto front for a rainfall-runoff model application. Although the aggregation method is (in principle) simple to implement, a complete single-objective optimization must be solved to obtain each discrete Pareto solution, making this approach inefficient, cumbersome, and time-consuming. Moreover, there is arguably a significant advantage to maintaining the independence of the various criteria, because a full multi-criteria optimization will allow an analysis of the tradeoffs among the different criteria and enable hydrologists to better understand the limitations of the current hydrologic model structure, thereby gaining insights into possible model improvements [Gupta et al., 1998].

Recently, a variety of evolutionary algorithms have become available which are designed to evolve multiple non-dominated solutions concurrently in a single optimization run, thereby guiding the search in several directions at the same time. These algorithms use the concepts of Pareto dominance, rather than single objective function evaluations, to construct a uniform estimate of the Pareto solution set. In the context of hydrologic modeling, Yapo et al. [1998] developed the Multi-Objective COMplex evolution (MOCOM-UA) global optimization method, which solves the multi-objective calibration problem by combining the strengths of the complex shuffling strategy and downhill Simplex evolution (adapted from the SCE-UA global optimization algorithm) with the concepts of Pareto dominance. Various hydrologic and hydrometeorologic calibration and evaluation studies have demonstrated that the MOCOM-UA algorithm can provide an efficient estimate of the Pareto set of solutions [Gupta et al., 1998, 1999; Yapo et al., 1998; Bastidas et al., 1999; Boyle et al., 2000, 2001; Wagener et al., 2001; Xia et al., 2002; Leplat et al., 2002; among others]. However, during the course of these investigations, we became aware that the current procedure has several weaknesses [Gupta et al., 2002], including the tendency of the MOCOM-UA algorithm to cluster the Pareto solutions into a central compromise region of the Pareto set, and a tendency to converge prematurely for case studies involving larger numbers of parameters and strongly correlated performance criteria.
In this Chapter, we present an effective and efficient Markov Chain Monte Carlo (MCMC) sampler, entitled the Multi-Objective Shuffled Complex Evolution Metropolis (MOSCEM-UA) algorithm, which is capable of generating a fairly uniform approximation of the “true” Pareto frontier within a single optimization run. The algorithm is closely related to the SCEM-UA algorithm [Vrugt et al., in Chapter 3 of this thesis], recently developed to infer the probabilistic uncertainty associated with the use of a single objective function, and uses a newly developed, improved concept of Pareto dominance to generate a fairly uniform approximation of the “true” Pareto frontier (thereby also containing the single criteria solutions at the extremes of the Pareto solution set). The features and capabilities of the MOSCEM-UA algorithm are illustrated using three hydrologic modeling case studies of increasing complexity, and the results are compared with the original MOCOM-UA algorithm.

6.2. Multi-Objective Optimization

To facilitate the description of the multi-objective optimization approach, let us write the hydrologic model \( \eta \) as follows:

\[
\hat{y} = \eta(\xi | \theta)
\]  

(6.1)

where \( \hat{y} \) denotes the model output, \( \xi \) is the input data, and \( \theta \) is a vector with \( n \) unknown parameters. We assume that the model structure specified by Eq. (6.1) is predetermined and fixed, and that realistic upper and lower bounds for each of the model parameters can be specified, thereby defining \( \Theta \), the feasible parameter space:

\[
\theta \in \Theta \subseteq \mathbb{R}^n
\]  

(6.2)

Here \( \mathbb{R}^n \) denotes the \( n \)-dimensional Euclidean space. If \( \Theta \) is not the entire domain space \( \mathbb{R}^n \), the identification problem is said to be constrained.

We first consider the situation in which the hydrologic model is required to simulate only one aspect of the system and for which \( \xi \) directly observed output values, \( \{y_i, i = 1, \ldots, i\} \), are available. The difference between the model-simulated output and the observed data can be represented by the residual vector:
where the function $G(\cdot)$ allows for linear or non-linear transformations of the simulated and observed data. The development of a measure $F(\theta)$, hereafter referred to as objective function, that mathematically measures the "size" of $E(\theta)$ is typically based on assumptions regarding the distributions of the measurement errors presented in the data. By far the most popular objective function is the Simple Least Square (SLS), which is also the maximum likelihood estimator when the measurement errors are known to be Gaussian, homoscedastic, and uncorrelated.

For single objective problems (where $F(\theta)$ is a scalar), the Shuffled Complex Evolution (SCE-UA) global optimization algorithm developed by Duan et al. [1992, 1993] has proved to be consistent, effective, and efficient in locating the values of the hydrologic model parameters that minimize (or maximize) the objective function. However, to quote Kuczera and Parent [1998] “...no hydrologist should be naive enough to rely on a uniquely determined value for each of the model parameters $\theta$, whatever the skill and imagination of the modeler might be”. In fact, it is typical that the vicinity of the global optimum contains several behavioral parameter sets with similar performance in reproducing the observed data.

The classical single objective optimization approach operates under the central assumption that a single objective function is able to properly extract all of the information contained in the time series of observations. However, practical experience with the calibration of hydrologic models suggests that the magnitude of structural error in the model for some portions of the model response may, in general, be equivalent to or even substantially larger than the measurement error and that these structural or model errors do not necessarily have any inherent probabilistic property that can be exploited in the construction of an objective function [Gupta et al., 1998]. Due to the presence of these structural inadequacies in the hydrologic model, any single (scalar) objective function, no matter how carefully chosen, is inadequate to properly measure all of the characteristics of the observed data deemed to be important.

These considerations imply the design of a calibration strategy that has the ability to simultaneously incorporate several objective functions. A strategy that can address this challenge
is multi-objective optimization, which has its roots in late-19th century welfare economics, in the work of Edgeworth [1881], and can be stated as follows:

\[
\min_{\theta \in \Theta} F(\theta) = \begin{bmatrix}
F_1(\theta) \\
F_2(\theta) \\
\vdots \\
F_M(\theta)
\end{bmatrix} \tag{6.4}
\]

where \( F_i(\theta) \) is the \( i \)th of \( M \) objective functions. The solution to this problem will in general, no longer be a single “best” parameter set but will consist of a Pareto set \( P(\Theta) \) of solutions in the feasible parameter space \( \Theta \) corresponding to various trade-offs among the objectives. The Pareto set of solutions defines the minimum uncertainty in the parameters that can be achieved without stating a subjective relative preference for minimizing one specific component of \( F(\theta) \) at the expense of another. Figure 6.1 illustrates the Pareto solution set for a simple problem where the aim is to simultaneously minimize two objectives \( (F_1, F_2) \) with respect to two parameters \((\theta_1, \theta_2)\).

**Figure 6.1.** Illustration of the concept of Pareto optimality for a problem having two parameters \((\theta_1, \theta_2)\) and two criteria \((F_1, F_2)\), in the parameter (a) and objective (b) space. The points A and B indicate the solutions that minimize each of the individual criteria \( F_1 \) and \( F_2 \). The thick line joining A and B corresponds to the Pareto set of solutions; \( \gamma \) is an element of the solution set, which is superior in the multi-criteria sense to any other point in \( \delta \).

The individual points A and B minimize objectives \( F_1 \) and \( F_2 \) respectively, whereas the solid line joining A and B represents the theoretical Pareto set of solutions. The black dots indicate an initial set of parameter estimates, while the number in subscript denotes their corresponding
Pareto rank. Moving along the line from $A$ to $B$ results in the improvement of $F_2$ while successively causing deterioration in $F_1$. The points falling on the line $AB$ represent trade-offs between the objectives and are called non-dominated, non-inferior, or efficient solutions. Put simply, the feasible parameter space (shaded region) can be partitioned into “good” or Pareto solutions and “bad” or “inferior” solutions. In the absence of additional information, it is impossible to distinguish any of the Pareto solutions (rank 1 points) as being objectively better than any of the other Pareto solutions. Furthermore, every member of the Pareto set will match some characteristic of the observed data better than any other member of the Pareto set, but the trade-off will be that some other characteristic of the observed data will not be as well-matched [Yapo et al., 1998]. Because of errors in the model structure (and other possible sources), it is usually not possible to find a single point $\theta$ at which all of the criteria have their minima. Note that this multi-objective equivalence of parameter sets is different from the probabilistic representation of parameter uncertainty, estimated using the SCEM-UA algorithm.

6.3. Effective and Efficient Algorithm to Solve the Multi-Objective Optimization Problem

While it may be relatively simple to pose the optimization problem into a multi-criteria framework, solving this problem to identify the Pareto set of solutions is not easy and has been the subject of much research. Ideally, the multi-objective optimization algorithm should find the set of all non-dominated solutions, which will constitute the global trade-off surface. However, because computational resources are finite, multi-objective solution algorithms typically approximate the Pareto set using a number of representative solutions.

The Vector Evaluated Genetic Algorithm (VEGA), developed by Schaffer [1985], was the first algorithm able to cope with multiple objectives simultaneously, without resorting to a strategy of scalarization by aggregation in order to solve a single objective surrogate problem instead. However, case studies have demonstrated that the VEGA algorithm has a tendency to ignore the most compromising points among the objectives on the trade-off curve, and that the algorithm eventually converges to a single point on the Pareto set. The first problem is an artifact of the VEGA strategy in which the evolution tends to favor individuals with extremely good performance among one of the objectives. To avoid these problems, Ritzel et al. [1994] modified the evolution process in VEGA by selecting parents based on a non-dominance ranking procedure called “Pareto ranking” [Goldberg, 1989]. Performance testing demonstrated that the resulting Pareto Genetic Algorithm (GA) was superior to VEGA. However, the Pareto GA has
an inherent tendency to converge too quickly, especially when the algorithmic parameters are not set properly, thereby yielding indistinguishable solutions that do not necessarily belong to the Pareto set.

In response to these issues, Yapo et al. [1998] developed the Multi-Objective COMplex Evolution (MOCOM-UA) method, a general-purpose multi-objective global optimization algorithm designed to efficiently generate a fairly uniform approximation of the Pareto set for a broad class of problems. The MOCOM-UA algorithm is an extension of the successful SCE-UA single objective global optimization algorithm developed by Duan et al. [1992, 1993] and merges the strengths of controlled random search [Price, 1987] with a competitive evolution [Holland, 1975], Pareto ranking [Goldberg, 1989], and multi-objective downhill Simplex strategy. For a detailed description and explanation of the algorithm, please refer to Yapo et al. [1996]. Various applications of the MOCOM-UA algorithm in hydrologic and hydrometeorologic calibration and evaluation studies [Gupta et al., 1998, 1999; Yapo et al., 1998; Bastidas et al., 1999; Boyle et al., 2000, 2001; Wagener et al., 2001; Xia et al., 2002; Lapastrier et al., 2001; among others] have demonstrated the usefulness of the MOCOM-UA algorithm. However, during the course of these investigations, it has become apparent that the current methodology has some serious weaknesses that need to be resolved. In further investigations, we discovered that these weaknesses are typical of the evolutionary algorithms, which are currently available for solving the multi-objective optimization problem.

The first failing of the MOCOM-UA algorithm is that it does not consistently generate a uniform approximation to the Pareto front, but tends to cluster the solutions in the compromise region among the objectives (e.g., see points on the Pareto set of Fig. 6.1b), thereby leaving the ends of the Pareto frontier unrepresented. Consequently, the Pareto set of solutions does not contain the individual single criterion (SCE-UA) solutions, which represent the theoretical extreme ends of the Pareto frontier. The second, perhaps more important, failure is the inability of the evolution strategy in the MOCOM-UA algorithm to converge to solutions within the “true” Pareto set for case studies involving large numbers of parameters and highly correlated performance criteria (e.g., typical of Soil-Vegetation-Atmosphere Transfer Scheme (SVATS) models, also known as Land-Surface Models (LSMs)). The algorithm tends, instead, to converge to a fuzzy region surrounding the Pareto set and, in some cases, does not converge at all. Note, that the phenomenon of genetic drift, where the members of the population drift to a single solution, is a characteristic typical of many evolutionary search algorithms. To prevent the collapse of the algorithm to a single region of highest attraction, the evolutionary algorithm incorporates a strategy that preserves the diversity of the sampled population.
6.3.1. Preservation of diversity in population

To find a set of non-dominated solutions, rather than a single-point solution, a multi-objective evolutionary algorithm must perform a multi-modal search that samples the Pareto-optimal set uniformly [Zitzler and Thiele, 1999]. We believe that there are two main reasons why current strategies for solving the multi-objective calibration problem do not preserve the diversity in the population, thereby tending to converge toward a compromise solution among the objectives instead:

1. **Replacement strategy**: Replacement of a member of the existing population occurs only if the generated offspring has a higher fitness than its parent. Although this evolution strategy essentially causes the algorithm to converge to a set of Pareto solutions, in the case of complex-shaped response surfaces with different regions of attraction (i.e., in the case of hydrologic models), this might cause the algorithm to prematurely converge to a single region of highest attraction surrounding the Pareto set.

2. **Fitness assignment**: In a multi-objective problem, several objectives are to be considered simultaneously, and ordered ranking of the population by conventional scalar sorting is therefore not possible. Fortunately, Goldberg [1989] suggested an elegant superiority-inferiority method for the ranking of a population of criteria, based on their mutual dominance relations. However, this widely used concept of Pareto ranking for fitness assignment in the case of multiple objectives does not distinguish between members having an identical rank. The solutions at the extreme ends of the Pareto frontier are assigned an identical “fitness” as the members of the Pareto set located in the most compromised region; however a greater number of solutions are usually found in this compromise region or “niche” of the parameter space.

With regard to the “replacement strategy” problem, the clumping tendency might be overcome by using a large initial population size, but this will significantly increase the number of required function evaluations and adversely affect the efficiency of the search algorithm. In examining ways to preserve the diversity in the sampled population and therefore avoid clumping of the solutions and premature convergence of the MOCOM-UA algorithm, it seems natural to consider the evolution strategy employed in the SCEM-UA algorithm [Vrugt et al., in Chapter 3 of this thesis], which is also designed to converge to a distribution of points, rather than a single “best” parameter set. The stochastic nature of the Metropolis–annealing scheme in the SCEM-
UA algorithm counters any tendency to collapse to a single region of attraction, thereby making possible the simultaneous identification of the "best" parameter set as well as its underlying posterior distribution. With regard to the "fitness assignment" problem, we propose a new (improved) concept of Pareto dominance that enables an evolutionary algorithm to preserve the diversity in the population.

6.3.2. Fitness assignment based on the number of external non-dominated points

The rank fitness assignment procedure begins by identifying all of the non-dominated individuals in the population and assigning them rank "one". While the original Pareto ranking concept now proceeds by peeling off these points and identifies the non-dominated points of the remaining population (assigned rank "two"), the proposed fitness assignment by Zitzler and Thiele [1999] proceeds as follows:

1. Store all of the rank "one" points in an external non-dominated set $P^*$ and the remaining dominated points of the population in a set entitled $P$.
2. Each solution $i \in P$ is assigned a real value $r_i \in [0,1)$, called strength. The strength is proportional to the number of population members $j \in P$ for which $i \succeq j$. Let $N$ be the number of individuals in $P$ that are covered by $i$ and $s$ is the population size ($P + P^*$). The strength is now defined as, $r_i = \frac{N}{s}$. For each member $i$ of $P^*$, the fitness ($f_i$) is identical to its computed strength ($r_i$).
3. The fitness of the remaining dominated individuals $j \in P$ is calculated by summing the strengths of all external non-dominated solutions $i \in P^*$ that cover $j$:

$$f_j = 1 + \sum_{i=1}^{i \succeq j} r_i \quad \text{where } f_j \in [1,s) \quad (6.5)$$

To ensure that the members of $P$ have a lower fitness than the members of $P^*$, the number one is added to the total sum. The closer the computed $f$ value (in Eq. (6.5)) is to zero, the higher the fitness of the sampled point.
To illustrate the proposed fitness assignment method and the difference with the conventional ranking method for a two-objective ($F_1,F_2$) problem, consider Figure 6.2 which presents computed fitness values for each ranking method at two different stages during the optimization. In the case of conventional Pareto ranking (Figs. 6.2a-b), points having an identical rank number are not distinguishable, even though the solutions at the extreme end are in some sense much more unique than other solutions having the same rank. In the case of the proposed fitness assignment (Figs. 6.2c-d), non-dominated individuals at the extreme end of the Pareto cluster are preferred, and individuals having many neighbors in their niche are penalized due to the high strength value of the associated non-dominated point (see circled area in Fig. 6.2c).

![Conventional Pareto ranking vs. Proposed fitness assignment](image)

**Figure 6.2.** Illustration of the conventional Pareto ranking concept (a,b) and the proposed fitness assignment concept (c,d) for a two objective ($F_1,F_2$) problem. For more explanation, please refer to the text.

Both of these principles of the proposed fitness assignment method preserve the diversity of the population and therefore favor uniform spacing of the solutions along the Pareto frontier (see...
Fig. 6.2c), thereby further reducing the chances of clumping of the solutions in the most compromised region (e.g., Fig. 6.2b) and of premature convergence.

Although the fitness assignment method by Zitzler and Thiele [1999] aims to preserve the diversity in the population, our attempts to apply the method to hydrologic models containing large numbers of parameters and with correlated objectives has revealed a major drawback of the method. When the non-dominated external set ($P$) contains only one member, which tends to occur when evaluating a population of points having strongly correlated objectives, the total set of dominated points will then appear to have identical fitness, thereby making it difficult for the MO algorithm to find a direction of improvement. To circumvent this problem, we modified the fitness assignment of the members of the dominated set, previously defined in Step 3, by adding the Pareto rank of each of the members of $P$ computed using the traditional ranking concept of the dominated set [Goldberg, 1989] to the sum of strengths calculated in Eq. (6.5). We found that this modification further improves the convergence properties of the MOSCEM-UA algorithm.

6.3.3. The Multi-Objective Shuffled Complex Evolution Metropolis (MOSCEM-UA) algorithm

In this section, we present the newly developed Multi-Objective Shuffled Complex Evolution Metropolis (MOSCEM-UA developed in collaboration between the University of Amsterdam and University of Arizona) algorithm. The evolution strategy employed in the MOSCEM-UA algorithm is identical to the strategy utilized in the SCEM-UA algorithm [Vrugt et al., in Chapter 3 of this thesis], with the exception that, to evolve the initial population of points toward a set of solutions stemming from a stable distribution, the probability ratio concept in the SCEM-UA algorithm is replaced with a multi-objective fitness assignment concept. The MOSCEM-UA algorithm is presented below and illustrated in the Figures 6.3 and 6.4.

**SETUP** (see Fig. 6.3)

1. To initialize the process, choose the population size $s$ and the number of complexes $q$.
2. Generate $s$ samples $\{\theta_1, \theta_2, \ldots, \theta_s\}$ from the prior distribution and compute the multi-objective vector $F(\theta)$ at each point $\theta$.
3. Compute the fitness $f_i$ for each individual of the sample (Section 6.3.2), sort the $s$ individuals by increasing fitness value, and store them in an array $D[1:s;1:n+M+1]$, where $n$ is the number of parameters, so that the first row of $D$ represents the point
with the “best” fitness. The extra columns in $D$ are used to store the multi-objective vector and the fitness values.

4. Initialize the starting points of the parallel sequences, $S_1^k, S_2^k, \ldots, S_q^k$, such that $S_k^k$ is $D[k, 1:n+1]$, where $k = 1, 2, \ldots, q$.

5. Partition $D$ into $q$ complexes $C_1^k, C_2^k, \ldots, C_q^k$, each containing $m$ points, such that the first complex contains every $q(j-1)+1$ sorted point of $D$, the second complex contains every $q(j-1)+2$ sorted point, and so on, where $j = 1, 2, \ldots, m$.

---

**Figure 6.3.** Flow chart of the MOSCEM-UA algorithm.
SEQUENCE EVOLUTION (see Fig. 6.4): FOR $k = 1$ to $k = q$ DO BEGIN

FOR $\beta = 1$ to $\beta = L$ DO BEGIN

Compute the covariance structure $\Sigma^k$ of the parameters of $C^k$.

Randomly draw a uniform label $Z$ over interval $[0,1]$.

WHILE NOT DRAWN FEASIBLE OFFSPRING DO BEGIN

Generate offspring according to:
$$\theta^{(r+1)} = N(\theta^k, \Sigma^k)$$

where $\theta^k$ is the current draw of $S^k$.

END

Begin Metropolis Step

[I] Compute $f_{r+1}$ using the points in $C^k$ and the current draw of $S^k$.

[II] Compute the ratio, $\alpha = \left( \frac{f_r}{f_{r+1}} \right)^{\beta/f_{r+1}}$ where $\beta$ is a scaling factor

and $f_r$ is the fitness associated with the current draw of $S^k$.

[III] If $\alpha \geq Z$, then accept the offspring. However, if $\alpha < Z$, then

reject the offspring and remain at the current position, that is,

$$\theta^{(r+1)} = \theta^k$$

[IV] Add the point $\theta^{(r+1)}$ to the sequence $S^k$.

[V] Replace the worst point of $C^k$ with $\theta^{(r+1)}$.

End Metropolis Step

END

END

6. Unpack all complexes $C$ back into $D$ and sort the points in order of increasing fitness value.

7. Check convergence statistic. If convergence criteria are satisfied, stop; otherwise, return to Step 4.

The MOSCEM-UA algorithm combines the strengths of (a) the complex shuffling employed in the SCE-UA algorithm [Duan et al., 1992, 1993], (b) the probabilistic covariance-annealing search procedure of the SCEM-UA algorithm [Vrugt et al., in Chapter 3 of this thesis] and, (c) our improved version of the fitness assignment concept of Zitzler and Thiele [1999] to construct an efficient and uniform estimate of the Pareto solution set.
To summarize, the MOSCEM-UA algorithm takes an initial population of points, randomly spread out in the feasible parameter space. For each individual of the population the multi-objective vector $F$ is computed and the population is ranked and sorted using an improved version of the fitness assignment concept developed by Zitzler and Thiele [1999]. The population is partitioned into several complexes and, in each complex $k$ ($k=1,2,\ldots,q$), a parallel sequence is
launched starting from the point that exhibits the highest fitness. A new candidate point in each
sequence \( k \) is generated using a multi-variate normal distribution centered around the current
draw of sequence \( (k) \) augmented with the covariance structure induced between the points in
complex \( k \). A Metropolis-type of acceptance rule is used to test whether the offspring should be
added to the current sequence or not. If the offspring (candidate point) is accepted, it replaces
the worst member of the current complex \( k \). Finally, after a prescribed number of iterations, the
complexes are replaced into the fixed population of points and new complexes are formed
through a process of shuffling. Iterative application of the various algorithmic steps causes the
population to converge toward the Pareto set of solutions.

The newly developed MOSCEM-UA algorithm differs from the original MOCOM-UA
algorithm in three essential ways. These modifications prevent premature convergence of the
algorithm to an indistinguishable region surrounding the Pareto set and help to avoid clustering
of solutions in the most compromised region among the objectives. In the first place, the
MOSCEM-UA algorithm uses an improved fitness assignment method, which preserves the
diversity of the population, whereas the MOCOM-UA algorithm uses the standard Pareto
ranking concept introduced by Goldberg [1989]. In the second place, the multi-objective downhill
simplex method used by the MOCOM-UA algorithm is replaced with a probabilistic covariance-
annealing search method, which is well-suited to deal with the strong correlation structures
between the parameters in the Pareto set that are typically encountered in hydrologic modeling.
Moreover, the stochastic nature of the annealing scheme prevents the collapse of the MOSCEM-
UA algorithm into a relatively small region of some single "best" parameter set, thereby further
preserving diversity of the sampled population and enabling the algorithm to generate a fairly
uniform approximation of the Pareto front. Finally, the MOSCEM-UA algorithm uses the
strengths of the shuffling procedure and complex partitioning employed in the single objective
SCE-UA global optimization algorithm [Duan et al., 1992, 1993] to conduct an efficient search of
the parameter space.

The MOSCEM-UA algorithm has four algorithmic parameters that must be specified by
the user: the population size \( s \), the number of complexes-sequences \( q \), which in turn also
determine the number of points within each complex \( m = s / q \), the number of evolutionary
steps in each complex before reshuffling \( L \), and the scaling factor \( \beta \) that directly determines
the acceptance probability of the generated offspring. The version of the MOSCEM-UA
algorithm used for the optimizations reported in this Chapter used the values of \( L = (m/4) \) and \( \beta
= (1/2) \). Therefore, the only variables that need to be specified by the user are the population
size \( s \) and the number of complexes \( p \). In the first case study, we will pay special attention to the
sensitivity of the performance of the MOSCEM-UA algorithm to the algorithmic parameters $s$ and $p$.

### 6.3.4. Performance criteria

In multi-objective optimization, the definition of performance is substantially more complex than for single-objective optimization problems, because the optimization goal itself consists of various subgoals:

1. The distance of the non-dominated solution set to the Pareto-optimal front should be minimized.
2. A uniform distribution of the solutions along the Pareto front is desirable.
3. The Pareto solution set should cover the full trade-off range of the various objectives, thereby including the single-objective solutions, which represent the theoretical ends of the Pareto frontier.

In the literature, attempts can be found that try to formalize the above objectives by means of quantitative measures. However, in this study, we used a visual comparison of the results obtained using the MOCOM-UA and MOSCEM-UA algorithms to evaluate performance in terms of the three factors mentioned above.

### 6.3.5. Initial sampling distribution

The purpose of the initial (prior) sampling distribution is to quantify the knowledge which is available before collecting and processing any data about the location of the Pareto solution set in the parameter space. If the initial sampling distribution is selected in order to closely approximate the true joint distribution of the parameters associated with the “true” Pareto set of solutions, the MOSCEM-UA algorithm will very rapidly generate a set of non-dominated solutions that closely approximates this “true” Pareto set. However, in the case of hydrologic models, usually very little a priori knowledge is available about the location of the Pareto set in the parameter space. Consequently, if the initial sampling distribution is chosen to express this high level of initial uncertainty (for example, Beven and Binley [1992] suggested a uniform distribution over a large rectangle of parameter values), the rate of convergence of the algorithm to the final Pareto set of solutions will tend to be slow.
To explicitly address the influence of the prior sampling distribution on the effectiveness and computational efficiency of the MOSCEM-UA algorithm for constructing an estimate of the Pareto solution set, we conducted two different experiments (case studies 2 and 3). In the first experiment, we assumed that there was no prior information available about the location of the Pareto solution set in the parameter space. Accordingly, a uniform prior distribution over the pre-specified upper and lower bounds for each of the model parameters was used to initialize the MOSCEM-UA algorithm (Step 2). In the second experiment, we initialized the MOSCEM-UA algorithm using approximate prior information about the location and structure induced in the joint distribution of the parameters in the Pareto set of solutions. Such a prior sampling distribution was approximated as follows:

**Step 1:** Use each of the $M$ objective functions, $F_i$, involved in the multi-criteria framework ($i = 1, 2, \ldots, M$) to separately locate the best attainable parameter values ($\Theta_{\text{opt}}$) using the SCE-UA global optimization algorithm [Duan et al., 1992, 1993].

**Step 2:** Use a traditional first-order approximation to estimate the multi-variate posterior joint probability density function, $p(\Theta | y)$, at each of the solutions $i$:

$$p(\Theta | y) \propto \exp \left[ -\frac{1}{2\sigma^2} (\Theta - \Theta_{\text{opt}})^T X^T X (\Theta - \Theta_{\text{opt}}) \right]$$

(6.7)

where $\sigma$ is the Root Mean Square Error (RMSE) of the fit at the final solution, and $X$ is the Jacobian or sensitivity matrix evaluated at $\Theta_{\text{opt}}$.

**Step 3:** For each of the objectives under consideration, generate $s/M$ points using the multi-variate posterior joint probability distribution specified in Eq. (6.7). The initial population of points for the MOSCEM-UA algorithm now constitutes the total of $s/M$ generated samples corresponding to each objective $i$.

Empirical investigations reported in this Chapter reveal that the latter approach, of first approximating the structure induced in the joint distributions of the Pareto solutions followed by initializing the MOSCEM-UA algorithm with this distribution, is computationally very efficient and, because the theoretical ends of the Pareto frontier are reflected in the prior, helps to reduce the typical MO algorithm problems reported above.
6.4. Case Studies

We compare the power and applicability of the original MOCOM-UA and MOSCEM-UA algorithm for three case studies with increasing complexity. The first is a standard mathematical case study using a simple two-dimensional Multi-Objective (MO) test problem. This illustrates the concepts of indistinguishability and demonstrates the ability of each algorithm to infer the known location of the Pareto optimal set. The second and third case studies explore the relative effectiveness and efficiency of the MOCOM-UA and MOSCEM-UA algorithms for two multi-criteria calibration hydrologic model applications; the Sacramento Soil Moisture Accounting model (SAC-SMA) conceptual watershed model, and the Biosphere Atmosphere Transfer Scheme (BATS) land surface model. In these case studies, we are especially concerned with the robustness of the MOCOM-UA and MOSCEM-UA algorithms by comparing the results of the multi-objective optimization with individual single criterion solutions obtained using the SCE-UA global optimization algorithm developed by Duan et al. [1992, 1993]. We also explicitly examine the influence of the initial sampling distribution on the effectiveness and efficiency of the MOSCEM-UA algorithm.

6.4.1. Case Study I: A Simple Two-Dimensional MO Problem

The applicability of the MOCOM-UA and MOSCEM-UA algorithms for generating an approximation of the Pareto set was examined using a simple mathematical test problem for which the exact location and shape of the Pareto set is known and can be easily computed using geometry. Consider the following two-dimensional MO test problem:

\[
\min_{\theta_1, \theta_2} F(\theta) \left\{ \theta_1^2 + \theta_2^2, (\theta_1 - 1)^2 + \theta_2^2, \theta_1^2 + (\theta_2 - 1)^2 \right\}
\] (6.8)

The Pareto solution set corresponding to Eq. (6.8) consists of a triangular-shaped area in the parameter space, having the corner points (0,0) – (0,1) and (1,0) for \(\theta_1\) and \(\theta_2\), respectively.

Figure 6.5 presents scatterplots of the final rank 1 points after performing 5,000 function evaluations with the MOCOM-UA (Figs. 6.5a-d) and MOSCEM-UA (Figs. 6.5e-g) algorithms using population sizes (\(\ell\)) of 10, 20, 50, and 100 points, respectively.
Figure 6.5. Scatterplots of the final set of rank 1 points after performing 5,000 function evaluations with the MOCOM-UA (5a-d) and MOSCEM-UA (5e-g) algorithm using a population size \( j \) of 10, 20, 50, and 100 points for a two-dimensional mathematical test problem having three objectives.

In general, the MOCOM-UA and MOSCEM-UA-generated points are consistent with the Pareto target distribution of points defined in Eq. (6.8) and are indicated by the dashed gray lines in Fig. 6.5. Note, however, that the MOCOM-UA algorithm does not generate a uniform sampling of the Pareto set of solutions, but has the tendency to cluster the sampled points around a single mode, even with increasing population size. On the contrary, the MOSCEM-UA algorithm maintains a uniform sampling density within the Pareto set of solutions, and the results are relatively insensitive to the specified population size and number of complexes (see also Table 6.1).

Table 6.1. Total number of rank 1 points after 5,000 function evaluations with the MOSCEM-UA algorithm as a function of the population size \( j \) and the number of complexes \( q \).

<table>
<thead>
<tr>
<th>Number of Complexes ( q )</th>
<th>10</th>
<th>20</th>
<th>50</th>
<th>100</th>
<th>250</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1789</td>
<td>1970</td>
<td>2282</td>
<td>2728</td>
<td>2515</td>
</tr>
<tr>
<td>2</td>
<td>2931</td>
<td>1934</td>
<td>2047</td>
<td>2327</td>
<td>2484</td>
</tr>
<tr>
<td>5</td>
<td>1915</td>
<td>1740</td>
<td>1846</td>
<td>2085</td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>1582</td>
<td>1856</td>
<td>1885</td>
<td></td>
<td></td>
</tr>
<tr>
<td>20</td>
<td></td>
<td>1572</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>50</td>
<td></td>
<td></td>
<td>1384</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Hence, a more diverse population yields better estimates of the final statistical moments of the Pareto distribution of points and as such is an additional advantage of the MOSCEM-UA algorithm over the MOCOM-UA algorithm.

6.4.2. Case Study II: The Sacramento Soil Moisture Accounting Model

We compare the effectiveness and efficiency of the MOCOM-UA and MOSCEM-UA algorithms by means of a case study involving calibration of the Sacramento Soil Moisture Accounting (SAC-SMA) model using data from the Leaf River watershed (1950 km²) near Collins, Mississippi. The SAC-SMA model is used by the National Weather Service (NWS) for flood forecasting throughout the United States and has 16 parameters that need to be specified by the user (see Table 6.2). While a few of these parameters might be estimated by relating them to observable watershed characteristics, most of the parameters are abstract conceptual representations of the watershed and must be estimated through calibration. Based on a recommendation by Peck [1976], the parameters SIDE, RIVA, and RSERV were fixed at prespecified values. The remaining 13 parameters were selected for the multi-criteria optimization, and the feasible parameter space was defined by fixing the upper and lower bounds at their “level zero” estimates presented by Boyle et al. [2000].

The data, obtained from the National Weather Service Hydrology Laboratory (HL), consist of mean areal precipitation (mm/day), potential evapotranspiration (mm/day), and streamflow (m³/s). Because the SAC-SMA and Leaf River data have been discussed extensively in previous work [see e.g., Sorooshian et al., 1993; Duan et al., 1993, 1994; Yapo et al., 1998], the details of these will not be reported here. In keeping with previous multi-criteria studies [Boyle et al., 2000], approximately 10 years (28 July 1952 to 30 September 1962) of historical hydrologic data were used for model calibration. To reduce sensitivity to state-value initialization errors, a 65-day warm-up period was used.

Because any conceptual watershed model will, in general, be unable to match all of the different aspects of the watershed’s behavior observed in the measured hydrograph, we follow a method similar to Boyle et al. [2000] and partition the hydrograph into a driven (D) and non-driven (ND) part, based on information from the measured hyetograph. A pair of Root Mean Square Error (RMSE) objective functions were computed, $F_D$ to measure the ability of the model to simulate the driven portion of the hydrograph, and $F_{ND}$ to measure the ability of the model to simulate the non-driven part of the hydrograph.
### Table 6.2: Parameters of the SAC-SMA model, with their initial uncertainty and multi-criteria calibrated range using the MOCOM-UA and MOSCEM-UA algorithm.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Description</th>
<th>Unit</th>
</tr>
</thead>
<tbody>
<tr>
<td>$C_v$</td>
<td>Fraction lower zone free water not transmissible to tension water</td>
<td>[-]</td>
</tr>
<tr>
<td>$L_n$</td>
<td>Injection recession from upper to lower zone free water store</td>
<td>[-]</td>
</tr>
<tr>
<td>$T$</td>
<td>Depression of the water table</td>
<td>[-]</td>
</tr>
<tr>
<td>$S_2$</td>
<td>Upper zone free water head depression rate</td>
<td>[-]</td>
</tr>
<tr>
<td>$S_1$</td>
<td>Lower zone free water head depression rate</td>
<td>[-]</td>
</tr>
<tr>
<td>$S_3$</td>
<td>Additional recharge area</td>
<td>[-]</td>
</tr>
<tr>
<td>$Q_{LZM}$</td>
<td>Lowerton flow from upper zone to lower zone</td>
<td>[-]</td>
</tr>
<tr>
<td>$Q_{LZD}$</td>
<td>Lower zone flow from upper zone</td>
<td>[-]</td>
</tr>
<tr>
<td>$Q_{LZM}$</td>
<td>Lower zone actuarial maximum storage</td>
<td>[-]</td>
</tr>
<tr>
<td>$Q_{LZM}$</td>
<td>Upper zone actuarial maximum storage</td>
<td>[-]</td>
</tr>
<tr>
<td>$Q_{LZM}$</td>
<td>Upper zone actuarial maximum storage</td>
<td>[-]</td>
</tr>
<tr>
<td>$Q_{LZM}$</td>
<td>Upper zone actuarial maximum storage</td>
<td>[-]</td>
</tr>
<tr>
<td>$Q_{LZM}$</td>
<td>Upper zone actuarial maximum storage</td>
<td>[-]</td>
</tr>
<tr>
<td>$Q_{LZM}$</td>
<td>Upper zone actuarial maximum storage</td>
<td>[-]</td>
</tr>
<tr>
<td>$Q_{LZM}$</td>
<td>Upper zone actuarial maximum storage</td>
<td>[-]</td>
</tr>
<tr>
<td>$Q_{LZM}$</td>
<td>Upper zone actuarial maximum storage</td>
<td>[-]</td>
</tr>
<tr>
<td>$Q_{LZM}$</td>
<td>Upper zone actuarial maximum storage</td>
<td>[-]</td>
</tr>
<tr>
<td>$Q_{LZM}$</td>
<td>Upper zone actuarial maximum storage</td>
<td>[-]</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Initial Range</th>
<th>MOCOM-UA</th>
<th>MOSCEM-UA</th>
</tr>
</thead>
<tbody>
<tr>
<td>$C_v$</td>
<td>0.0 - 0.02</td>
<td>0.0 - 0.02</td>
<td>0.0 - 0.02</td>
</tr>
<tr>
<td>$L_n$</td>
<td>0.00 - 0.00</td>
<td>0.00 - 0.00</td>
<td>0.00 - 0.00</td>
</tr>
<tr>
<td>$T$</td>
<td>0.00 - 1.00</td>
<td>0.00 - 1.00</td>
<td>0.00 - 1.00</td>
</tr>
<tr>
<td>$S_2$</td>
<td>1.00 - 0.00</td>
<td>1.00 - 0.00</td>
<td>1.00 - 0.00</td>
</tr>
<tr>
<td>$S_1$</td>
<td>0.00 - 0.00</td>
<td>0.00 - 0.00</td>
<td>0.00 - 0.00</td>
</tr>
<tr>
<td>$S_3$</td>
<td>0.00 - 0.00</td>
<td>0.00 - 0.00</td>
<td>0.00 - 0.00</td>
</tr>
<tr>
<td>$Q_{LZM}$</td>
<td>0.00 - 0.00</td>
<td>0.00 - 0.00</td>
<td>0.00 - 0.00</td>
</tr>
<tr>
<td>$Q_{LZM}$</td>
<td>0.00 - 0.00</td>
<td>0.00 - 0.00</td>
<td>0.00 - 0.00</td>
</tr>
<tr>
<td>$Q_{LZM}$</td>
<td>0.00 - 0.00</td>
<td>0.00 - 0.00</td>
<td>0.00 - 0.00</td>
</tr>
<tr>
<td>$Q_{LZM}$</td>
<td>0.00 - 0.00</td>
<td>0.00 - 0.00</td>
<td>0.00 - 0.00</td>
</tr>
<tr>
<td>$Q_{LZM}$</td>
<td>0.00 - 0.00</td>
<td>0.00 - 0.00</td>
<td>0.00 - 0.00</td>
</tr>
<tr>
<td>$Q_{LZM}$</td>
<td>0.00 - 0.00</td>
<td>0.00 - 0.00</td>
<td>0.00 - 0.00</td>
</tr>
<tr>
<td>$Q_{LZM}$</td>
<td>0.00 - 0.00</td>
<td>0.00 - 0.00</td>
<td>0.00 - 0.00</td>
</tr>
<tr>
<td>$Q_{LZM}$</td>
<td>0.00 - 0.00</td>
<td>0.00 - 0.00</td>
<td>0.00 - 0.00</td>
</tr>
<tr>
<td>$Q_{LZM}$</td>
<td>0.00 - 0.00</td>
<td>0.00 - 0.00</td>
<td>0.00 - 0.00</td>
</tr>
</tbody>
</table>
The Pareto optimal solution space for the two criteria was estimated using a population size of 500 points and 100,000 trials with the MOCOM-UA and MOSCEM-UA algorithms. The results of this two-criteria \(\{F_{\text{ND}}, F_D\}\) calibration are summarized in Figures 6.6, 6.7, and 6.8.

Figure 6.6 presents normalized parameter plots for each of the parameters of the SAC-SMA model using either MOCOM-UA algorithm (Fig. 6.6a), the MOSCEM-UA algorithm with uniform prior sampling on the feasible parameter space (Fig. 6.6b), or the MOSCEM-UA algorithm utilizing prior sampling information (Fig. 6.6c).

![Normalized parameter plots for each of the SAC-SMA model parameters using a two-criteria \(\{F_{\text{ND}}, F_D\}\) calibration with the MOCOM-UA (A), MOSCEM - no prior information (B), and MOSCEM - prior information (C) algorithm. Each line across the graph denotes a single parameter set, gray = Pareto solution set, solid and dashed black lines are single criterion solutions of \(F_D\) and \(F_{\text{ND}}\), respectively. The squared plots at the righthand side are two-dimensional projections of the objective space of the Pareto set of solutions.](image)

The 13 SAC-SMA model parameters are listed along the x-axis, while the y-axis corresponds to the parameter values scaled according to their prior uncertainty ranges (defined in Table 6.2) to
yield normalized ranges. Each line across the graph represents one parameter set. The solid and
dashed black lines going from left to right across the plots correspond to the single objective
solutions of $F_D$ and $F_{ND}$ obtained by separately fitting to each criterion using the SCE-UA global
optimization algorithm [Duan et al., 1992], while the gray lines denote members of the Pareto set
of solutions. The objective function plots on the righthand side in Fig. 6.6 depict two-
dimensional projections of the bi-criterion trade-off surfaces represented by the Pareto set of
solutions. Additionally, Table 6.2 lists the Pareto uncertainty intervals of the parameters
estimated with the MOCOM-UA and MOSCEM-UA algorithms in the non-transformed
parameter space. Fig. 6.6a clearly illustrates that the MOCOM-UA algorithm has generated a
fairly uniform approximation of the Pareto frontier only in the most compromised region among
the objectives and does not represent the extreme points of the Pareto frontier well. This
clustering of solutions in the middle region of the Pareto frontier is also demonstrated in the
estimated Pareto uncertainty intervals of the parameters, which for seven of the 13 SAC-SMA
parameters (UZFW, UZK, ADIMP, ZPERC, REXP, LZWWM, and LZPK) does not bracket
the single objective solutions (Fig. 6.6a). On the contrary, the MOSCEM-UA algorithm
generates a fairly uniform approximation of the “true” Pareto solution set, which contains the
single criterion solutions at the extreme ends of the Pareto frontier. This is especially true when
using prior information about the location of the Pareto solution set in the parameter space in
the sampling strategy (Fig. 6.6c). As a consequence, the estimated Pareto uncertainty intervals of
the SAC-SMA parameters contain the single criterion SCE solutions. Notice that, for most of the
parameters, the Pareto solution set tends to cluster closely in the parameter space for the two
objectives. However, there is considerable uncertainty associated with the percolation parameter
ZPERC and the recession parameters UZK and LZPK in the SAC-SMA model, which play a
major role in determining the shape of the hydrograph during recession periods. Also notice the
close match between the Pareto uncertainty intervals estimated with the MOSCEM-UA
algorithm as illustrated in Figs. 6.6b-c.

To further illustrate the advantage of using prior information about the location of the
Pareto solution set in the parameter space in the initial sampling with the MOSCEM-UA
algorithm, consider Figure 6.7 that presents the evolution of the bi-criterion trade-off surface in
the two-dimensional objective space as a function of the number of SAC-SMA model
evaluations. When utilizing a uniform initial sampling of the feasible parameter space (no prior
information) with the MOSCEM-UA algorithm, typically 30,000 SAC-SMA model evaluations
are needed to construct an estimate of the Pareto solution set.
The benefit of going from 30,000 to 50,000 model evaluations can be considered to be marginal given the extra cost in terms of model runs. When prior information is used, 10,000 independent model evaluations with the SCE-UA algorithm are first needed (separately) for each of the two objectives to identify the single criterion ends of the Pareto frontier. Notice that the initial population of points, stemming from this approach (outlined in Section 6.3.5) using the single criterion ends of the Pareto cluster, directly approximates the Pareto solution set as depicted with the upper triangular symbols in the two-dimensional bi-criterion plot in Fig. 6.7b. This suggests that the structure induced in the joint distribution of the parameters in the Pareto solution set can be well approximated using information from the single criterion ends of the Pareto frontier. Clearly, the use of prior information in the initial sampling with the MOSCEM-UA algorithm is not only computationally more efficient, but also generates a more uniform estimate of the Pareto frontier which includes the single criterion SCE solutions.

The hydrograph uncertainty ranges (shaded area) associated with the Pareto solution set estimated with the MOSCEM-UA algorithm for a 500-day portion of the calibration period are displayed in Figure 6.8 using a logarithmic transformation of the streamflows. The observed streamflows are indicated with dots, while the single criterion solutions for the driven and non-driven portions of the hydrograph are indicated with the solid and dashed black lines, respectively.
Figure 6.8. Hydrograph prediction uncertainty ranges (shaded area) associated with the Pareto solution set estimated with the MOSCEM-UA algorithm for a 500-day portion of the calibration period. The solid circles correspond to the observed streamflow data; the solid line corresponds to the minimal F_D solution, and the dashed curve corresponds to the minimal F_ND solution.

Note that the streamflow prediction uncertainty ranges match the medium- and high-flow events very well, but do not bracket the observations and display bias (systematic error) on the long recessions, suggesting that the model structure may be in need of further improvement. Notice that the relatively large uncertainty found during low flow and recession periods is consistent with the relatively large uncertainty in the UZK and LZPK parameters.

6.4.3. Case Study III: The Biosphere Atmosphere Transfer Scheme (BATS) Model

The third case study illustrates the power of the MOSCEM-UA algorithm to perform a multi-criteria \( \{H, \lambda E\} \) calibration of the Biosphere-Atmosphere Transfer Scheme (BATS) land-surface model [Dickinson et al., 1993] using measured sensible \( (H) \) and latent heat \( (\lambda E) \) fluxes from the Oklahoma ARM-CART site. BATS is a conceptual parameterization of the ecohydrologic processes at the scale of individual plots of vegetation (50-1000 m). The model consists of six interacting hydrometeorological components (three layers of soil, a canopy air component, a canopy leaf-stem component, and a snow-covered portion) and has 27 parameters to be estimated, including 16 related to vegetation properties and eight related to soil properties, together with three initial soil-moisture conditions (see Table 6.3).
Table 6.3. Parameters of the BATS model, including their initial uncertainty and final multi-criteria calibrated range obtained with the MOSCEM-UA algorithm.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Description</th>
<th>Initial Reasonable Range</th>
<th>Multi-criteria Range</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>veg. rough</td>
<td>0.40 - 0.05</td>
<td>0.40 - 0.10</td>
</tr>
<tr>
<td>2.</td>
<td>veg. alp</td>
<td>0.05 - 0.01</td>
<td>0.01 - 0.35</td>
</tr>
<tr>
<td>3.</td>
<td>soil cover</td>
<td>1.00 - 0.20</td>
<td>0.20 - 0.40</td>
</tr>
<tr>
<td>4.</td>
<td>min. area</td>
<td>0.01 - 0.50</td>
<td>0.01 - 0.50</td>
</tr>
<tr>
<td>5.</td>
<td>transp.</td>
<td>0.35 - 0.06</td>
<td>0.35 - 0.06</td>
</tr>
<tr>
<td>6.</td>
<td>rough depth</td>
<td>0.50 - 0.10</td>
<td>0.10 - 0.50</td>
</tr>
<tr>
<td>7.</td>
<td>light dep.</td>
<td>0.05 - 0.01</td>
<td>0.05 - 0.01</td>
</tr>
<tr>
<td>8.</td>
<td>min. leaf</td>
<td>0.01 - 0.50</td>
<td>0.01 - 0.50</td>
</tr>
<tr>
<td>9.</td>
<td>dim. height</td>
<td>0.06 - 0.01</td>
<td>0.06 - 0.01</td>
</tr>
<tr>
<td>10.</td>
<td>min. stom</td>
<td>0.01 - 0.50</td>
<td>0.01 - 0.50</td>
</tr>
<tr>
<td>11.</td>
<td>max. stom</td>
<td>0.06 - 0.01</td>
<td>0.06 - 0.01</td>
</tr>
<tr>
<td>12.</td>
<td>depth</td>
<td>0.01 - 0.50</td>
<td>0.01 - 0.50</td>
</tr>
<tr>
<td>13.</td>
<td>depth</td>
<td>0.01 - 0.50</td>
<td>0.01 - 0.50</td>
</tr>
<tr>
<td>14.</td>
<td>depth</td>
<td>0.01 - 0.50</td>
<td>0.01 - 0.50</td>
</tr>
<tr>
<td>15.</td>
<td>depth</td>
<td>0.01 - 0.50</td>
<td>0.01 - 0.50</td>
</tr>
<tr>
<td>16.</td>
<td>depth</td>
<td>0.01 - 0.50</td>
<td>0.01 - 0.50</td>
</tr>
</tbody>
</table>

Note: The parameters are listed in the order of their initial uncertainty and final multi-criteria calibrated range.
Two of the parameters, \( x_{\text{mowil}} \) (the wilting point) and \( x_{\text{mofc}} \) (the ratio of field capacity to the saturated water content) are actually not independent parameters. The parameter \( x_{\text{mowil}} \) is computed as a function of the hydraulic conductivity \( x_{\text{mohyd}} \) and the minimum soil suction \( x_{\text{mosuc}} \), while \( x_{\text{mofc}} \) is used only when the land cover is assigned to be semi-desert [Dickinson et al., 1993; Gupta et al., 1999].

Land Surface Models (LSMs), like BATS, differ from hydrologic watershed models, such as the SAC-SMA model used in the previous case study, in that they are concerned with both water and energy balance (and more recently carbon and other fluxes): they are driven by multiple input variables (precipitation, short-wave and long-wave radiation, wind speed, air temperature, humidity, etc.), and they present the evolution of several observable state variables (soil temperature, surface soil-moisture content, etc.) and output fluxes (latent heat, sensible heat, runoff, etc.) [Bastidas et al., 1999]. The hydrometeorological data set used in this study correspond to station E13 of the Atmospheric Radiation Measurement Cloud and Radiation Testbeds (ARM-CART) program in the Southern Great Plains site (SGP) in Oklahoma. The data cover the 5-month period from April 1 to August 25, 1995, with a sampling interval of 30 minutes, and include all of the necessary atmospheric forcing for the model and observational information on sensible heat \( (H \text{ in } \text{W/m}^2) \) and latent heat fluxes \( (\lambda E \text{ in } \text{W/m}^2) \). It has been previously established that the MOCOM-UA algorithm fails to converge for this data set with this particular combination \( \{H, \lambda E\} \) of objectives [Bastidas, 1998]. For each of the two criteria, the simulation error was measured using the RMSE statistic. For more information about the BATS model, the hydrometeorological data, and multi-criteria calibration approaches applied to the BATS model, please refer to Dickinson et al. [1993], Bastidas [1998], Bastidas et al. [1999], and Gupta et al. [1999]. The Pareto optimal solution space for the two criteria was estimated using a population size of 2,000 points in combination with 100,000 trials with the MOSCEM-UA algorithm. The results of the two-criteria \( \{H, \lambda E\} \) calibration are summarized in Figures 6.9 and 6.10 and discussed below.

Figure 6.9 presents the results for the two-criteria \( \{H, \lambda E\} \) calibration with the MOSCEM-UA algorithm in the normalized parameter and objective space for three cases. In the first case (Fig. 6.9a), a uniform prior sampling of the feasible parameter space was used, whereas in the second and third cases, each end (\( \{H\} \) and \( \{\lambda E\} \)) of the Pareto frontier was first identified by single objective optimization using the SCE-UA algorithm [Duan et al., 1992] (second case, Fig. 6.9b) or the SCEM-UA algorithm [Vrugt et al., in Chapter 3 of this thesis] (third case, Fig. 6.9c), and used to initialize the prior distribution for the MOSCEM-UA algorithm.
Figure 6.9. Normalized parameter plots for the ARM-CART site using a two-criteria \{\(H, \lambda E\}\) calibration with the MOSCEM – no prior information (a), MOSCEM - prior SCE information (b), and MOSCEM - prior SCEM information (c) algorithm. Each line across the graph denotes a single parameter set, solid and dashed black lines are single criterion SCEM solutions of \(\lambda E\) and \(H\), respectively, and gray = Pareto solution set, squares, and triangles denote single criterion SCE-solutions of \(\lambda E\) and \(H\), respectively. The squared plots at the righthand side denote two-dimensional projections of the objective space of the Pareto set of solutions. In these plots, rank one solutions are indicated by black dots.
Each line going from left to right across the normalized parameter plots corresponds to a different parameter set (the solid and dashed black line denote the \{\lambda E\} and \{H\} single criterion SCEM solutions, respectively); each gray line denotes a member of the \{H,\lambda E\} Pareto set of solutions, and the square and triangular symbols denote the SCE solutions for \{H\} and \{\lambda E\}, respectively. The 24 BATS parameters and three initial soil-moisture parameters are listed along the x-axis, and the y-axis corresponds to the parameter values, normalized by their initial uncertainty ranges, as defined in Table 6.3. The linear-linear squared shaped plots at the righthand side in Fig. 6.9 depict two-dimensional projections of the bi-criterion trade-off surface represented by the total set of points sampled with the MOSCEM-UA algorithm. The Pareto rank one solutions in these plots are indicated by the black dots.

The results presented in Fig. 6.9 emphasize several important observations. In the first place, notice that the SCEM and MOSCEM-UA single criterion (\{H\} and \{\lambda E\}) solutions are significantly better in terms of their RMSE performance measures (typically 10%) and occupy a different part of the parameter space than their counterparts obtained with the SCE-UA algorithm. To verify the consistency of the results with the SCE-UA algorithm, we ran the algorithm ten different times with an increasing number of complexes. Indeed, the algorithm consistently converged to the same region in the parameter space, well removed from the SCEM-identified global minimum. Consequently, when initialized with prior information obtained with the SCE-UA algorithm, the MOSCEM-UA algorithm fails to converge to the "true" Pareto set as depicted in Fig. 6.9b. In the other cases, however, the MOSCEM-UA algorithm generated a fairly uniform approximation of the Pareto frontier, thereby containing the best attainable single objective \{H\} and \{\lambda E\} SCEM solutions, denoted with the solid and dashed lines in the squared plots.

The second interesting observation is that the Pareto solution set is discontinuous in the objective space with clusters of solutions close to the single criterion ends of the Pareto frontier, but with no solutions in the most compromised region among these objectives. This discontinuity is also observed in the normalized parameter plots where, for some of the BATS model parameters (VEGC, RSMIN, XLA, SAI, DEPTV, and SKRAT), two separate well-defined clusters of Pareto solutions can be found close to the single criterion \{H\} and \{\lambda E\} SCE solutions, while no Pareto solutions are found in the parameter space in between these extreme ends.

Although beyond the scope of this Chapter, we believe that the convergence problems of the SCE-UA algorithm are caused by the large number of interacting parameters in the BATS model and the highly complex, non-convex shape of the response surface mapped out in the
parameter space. At one end of the spectrum, deterministic search methods, like the Simplex algorithm implemented in the SCE-UA complex evolution strategy, are especially designed for response surfaces that exhibit a well-defined global minimum. At the opposite end, however, probabilistic search methods do not impose constraints on the shape of the response surface and are especially suited to deal with a high degree of randomness in the response surface. We posit that the calibration of the parameters in the BATS model involves a high degree of randomness and non-convexity in the response surface, thereby causing problems in the identification of the global minimum for classical deterministic search algorithms.

This explanation is also supported by Figure 6.10, which presents the evolution of the best RMSE values for the $\{H\}$ and $\{LE\}$ criteria, as functions of the number of BATS model evaluations with the SCE-UA [Duan et al., 1992], SCEM-UA [Vrugt et al., in Chapter 3 of this thesis], and MOSCEM-UA [this Chapter] optimization algorithms.

![Figure 6.10](image)

**Figure 6.10.** Evolution of the best Root Mean Square Error values for sensible (a) and latent heat fluxes (b) as a function of the number of BATS model evaluations with the SCE-UA, SCEM-UA, and MOSCEM-UA optimization algorithms.

The results depicted in Fig. 6.10 show that the SCEM-UA and MOSCEM-UA algorithms converge more quickly and to smaller function values, indicating that the probabilistic covariance-based search method has superior search capabilities over the Simplex search strategy implemented in the SCE-UA global optimization algorithm. While the SCE-UA algorithm requires 75,000 model evaluations to converge to a sub-optimal solution, only approximately
15,000 trials with the BATS model are needed with the SCEM-UA and MOSCEM-UA algorithms to identify the minimal \{H\} and \{\lambda E\} objective RMSE solutions. Although not explicitly illustrated here, using this prior information, the pattern of the resulting population evaluated in the two-dimensional \{H,\lambda E\} objective space showed a striking similarity to the "true" Pareto set of solutions illustrated in Fig 6.9c. This suggests again that the joint distribution of the parameters in the Pareto solution set can be estimated using information obtained from the single criterion ends of the Pareto frontier. This approach to estimating the prior distribution seems to be both robust and efficient and, because the theoretical ends of the Pareto frontier are computed beforehand, should help to minimize pitfalls that may arise in a wide variety of multi-objective calibration problem applications.

A third significant and interesting observation (demonstrated in Fig. 6.10) is that the multi-criteria calibration approach seems to provide superior convergence speed compared to the single objective (compare the dashed lines in Figs. 6.10a-b) approaches. This suggests that minimizing several objectives simultaneously can increase the identifiability of the global minimum in the parameter space, an observation that deserves further investigation in future work.

Finally, Figure 6.11 shows time series plots of modeled sensible heat flux (W/m²), latent heat flux (W/m²), ground temperature (°K), and soil moisture content (mm) with the BATS model against the observed data (denoted with circles) for a representative 10-day period of the calibration period, during which rainfall occurred. Each of the Pareto set of solutions corresponding to the \{H,\lambda E\} two-criteria calibration is indicated by a gray line, while the single criterion \{H\} and \{\lambda E\} SCEM solutions are indicated by the solid and dotted black lines, respectively. The Pareto prediction uncertainty ranges bracket the sensible and latent heat fluxes during most of the time, but do not match the ground temperature (not included in the multi-criteria calibration) very well. It is possible that a better match to this state variable could be obtained by including the ground temperature observations in the multi-criteria optimization framework. Notice that the BATS model response to the observed sensible and latent heat fluxes is quite similar for each of the two single criterion \{H\} and \{\lambda E\} SCEM solutions (indicated with the solid and dashed lines), while two disconnected regions of model response are found for the ground temperature, which was not included in the calibration. This clearly implies that, although there are two disconnected Pareto solution set clusters in the parameter space associated with each of the two calibration criteria (see Fig. 6.9), both clusters generate similar model responses in terms of sensible and latent heat fluxes, indicating an interesting model structural issue that deserves further investigation.
Figure 6.11. Time series plots of modeled sensible heat flux (W/m²), latent heat flux (W/m²), ground temperature (°K) and soil-moisture content (mm), with the BATS model for a representative 10-day period of the calibration period. The solid circles denote observed data, the gray lines represent the Pareto \( \{H, \lambda, E\} \) set of solutions, and the single criterion \( \{H\} \) and \( \{\lambda, E\} \) SCEM solutions are indicated by the solid and dotted black lines, respectively.
6.5. Summary and conclusions

This Chapter has presented a Markov Chain Monte Carlo sampler, which is well suited for solving the multi-criteria optimization problem for hydrologic models. The sampler, entitled the Multi-Objective Shuffled Complex Evolution Metropolis (MOSCEM-UA developed in collaboration between the University of Amsterdam and the University of Arizona), merges the strengths of complex shuffling employed in the SCE-UA algorithm [Duan et al., 1992, 1993] with the probabilistic covariance-based search methodology of the Metropolis algorithm [Metropolis, 1953] and an improved fitness assignment concept of Zitzler and Thiele [1999] to construct an efficient and uniform estimate of the Pareto solution set. The MOSCEM-UA algorithm is a multi-objective relative of the SCEM-UA algorithm [Vrugt et al., in Chapter 3 of this thesis], originally developed to infer the probabilistic uncertainty associated with the use of a single objective function, but uses an innovative concept of Pareto dominance rather than direct objective function evaluations to generate a fairly uniform approximation of the “true” Pareto frontier which includes the single criteria end points of the Pareto solution set.

The efficiency and effectiveness of the newly developed MOSCEM-UA algorithm for constructing an estimate of the Pareto solution set was compared with the MOCOM-UA algorithm developed by Yapo et al. [1998] for three case studies of increasing complexity. The first case study considered a simple two-dimensional mathematical test problem, while the second and third case studies explored the effectiveness and efficiency of the MOSCEM-UA algorithm for a two-criteria calibration of the Sacramento Soil Moisture Accounting (SAC-SMA) conceptual watershed model and the Biosphere Atmosphere Transfer scheme (BATS) land-surface model. The three case studies clearly demonstrated that the MOCOM-UA algorithm has the tendency to cluster the Pareto solutions in the most compromised region among the objectives, in the third case study, the MOCOM-UA failed to converge. In contrast, the MOSCEM-UA algorithm generates a fairly uniform approximation of the entire Pareto front, which includes the single criterion end points in the estimated Pareto uncertainty intervals of the parameters. Furthermore, empirical investigations reported in this Chapter revealed that a strategy of first locating the single criterion end points of the Pareto frontier, and using this information as a prior estimate of the structure induced in the Pareto solution set of the parameters, is computationally more efficient than imposing a uniform prior distribution on the model parameters during the initialization of the MOSCEM-UA algorithm.
Part V

Simultaneous Parameter and State Estimation
CHAPTER 7

Improved Treatment of Uncertainty in Hydrologic Modeling: Combining the Strengths of Global Optimization and Data Assimilation

Abstract

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. In practice, however, because of errors in the model structure and the input (forcing) and output data, this has proven to be difficult, leading to considerable uncertainty in the model predictions. This Chapter surveys the limitations of current model calibration methodologies, which treat the uncertainty in the input–output relationship as being primarily attributable to uncertainty in the parameters, and presents a Simultaneous Optimization and Data Assimilation method, entitled SODA, which improves the treatment of uncertainty in hydrologic modeling. The usefulness and applicability of SODA is demonstrated by means of a pilot study using data from the Leaf River Watershed in Mississippi and a simple hydrologic model with typical conceptual components.

1This Chapter has been written by Jasper A. Vrugt, Cees G.H. Diks, Hoshin V. Gupta, Willem Bouten, and Jacobus M. Verstraten, and was submitted to Water Resources Research.
CHAPTER 7

7.1. Introduction and scope

Hydrologic models often contain parameters that cannot be measured directly, but can only be meaningfully inferred by calibration to a historical record of input-output data. During model calibration, the parameters are adjusted in such a way 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. Because of the time consuming nature of manual trial-and-error model calibration, there has been a great deal of research into the development of automated (computer based) calibration methods [see e.g., Gupta and Sorooshian, 1994; Yapo et al., 1998; Boyle et al., 2000]. Automatic methods for model calibration seek to take advantage of the speed and power of computers, while being relatively objective and easier to implement than manual methods.

While considerable progress has been made in the development and application of automated optimization methods, major weakness of these approaches include their underlying treatment of the uncertainty in the input-output representation of the model as being primarily (and explicitly) attributed to uncertainty in the parameter estimates, without explicit treatment of the input, output and model structural uncertainties. However, uncertainties in the modeling procedure stem not only from uncertainties in the parameter estimates, but also from measurement errors associated with the system input (forcing) and output, and from model structural errors arising from the aggregation of spatially distributed real-world processes into a mathematical model. Not properly accounting for these errors during model calibration, can result in model simulations and their associated prediction uncertainty bounds, which do not consistently represent and bracket the measured system behavior. This is usually evidenced by residuals, which exhibit considerable variations in bias (non-stationarity), variance (heteroscedasticity), and correlation structures under different hydrologic conditions. Several contributions to the hydrologic literature have therefore brought into question the continued usefulness of the classical paradigm for estimating model parameters [Beven and Binley, 1992; Gupta et al., 1998; Kavetski et al., 2003], especially in the face of the emerging generation of spatially distributed multi-input-output hydrologic models for which multiple (and often conflicting) sources of information are available for model calibration.

Some interesting methods for addressing these problems, particularly in the context of estimating reasonable confidence bands on the model simulations, have begun to appear in the literature. These methods include the use of classical Bayesian [Kuczera and Parent, 1998; Kavetski et al., 2003; Thiemann et al., 2001; Vrugt et al., in Chapter 3 of this thesis], pseudo-Bayesian [Beven and
SIMULTANEOUS PARAMETER AND STATE ESTIMATION

Binley, 1992; Freer et al., 1996], set-theoretic [Keesman, 1990; Klepper et al., 1991; van Straten and Keesman, 1991; Vrugt et al., 2001a], and multiple-criteria [Gupta et al., 1998; Yapo et al., 1998; Boyle et al., 2000; Vrugt et al., in Chapter 6 of this thesis] methods to represent model parameter and prediction uncertainty. Each of these methods, however, uses one or more aggregate statistics of model performance over a large range of hydrologic behaviors, an action that is now understood to result in considerable loss of important information that can be used to distinguish between competing parameter sets. To increase the discriminative power of model calibration strategies, interest has therefore recently begun to switch towards recursive model identification strategies [Thieman et al., 2001; Young, 2001; Vrugt et al., in Chapter 5 of this thesis; Wagener et al., 2003], which sequentially move through a time series of discharge data and can provide estimates of parameter uncertainty. These approaches can not only be used to localize the most informative measurements for parameter estimation [see Vrugt et al., in Chapter 5 of this thesis], but perhaps more importantly, to provide a method for checking for violations of the underlying assumption that parameters are constant [Wagener, 2003; Misirli, 2003]. Despite progress made, such methods still assign the uncertainty in the input–output representation primarily to a combination of uncertainty in the parameter estimates plus a residual "model" error term [see Thieman et al., 2001, Gupta et al., 2003], and therefore lack the conceptual rigor needed to distinguish between all the important sources of uncertainty.

In a separate line of research, considerable progress has been made in the development and application of sequential data assimilation (SDA) techniques. Such methods provide a general framework for explicitly dealing with input, output and model structural uncertainty, and for optimal merging of uncertain model predictions with observations. In contrast to classical model calibration strategies, SDA methods continuously update the states in the model when new measurements become available to improve the model forecast and evaluate the forecast accuracy. The prototype of the SDA methods, the Kalman Filter (KF), was developed in the 1960s for optimal control of systems governed by linear equations and was first introduced into hydrology during the early 80s to improve real-time forecasting of river discharges [Kitanidis and Bras, 1980a,b], and to recursively estimate model parameters [Beck, 1987]. For nonlinear dynamics, the extended Kalman Filter (EKF) can be used, which linearizes the error covariance equation using a tangent linear operator. However, because third- and higher order moments in the closure scheme are discarded, this linearization is notoriously unstable if the nonlinearities are strong [Evensen, 1992; Miller et al., 1994].

Although the KF and EKF offer a very general framework for accounting for all sources of uncertainty, the filter typically assumes that the optimal values of the model parameters are
known prior to the modeling procedure. This is particularly difficult in hydrologic modeling, where parameters often represent conceptual properties whose values can generally not be independently measured or assessed from lookup tables. This uncertainty in the correct choice of the parameter values, results in uncertainty in the estimated state values and therefore the model output prediction. To improve the usefulness and applicability of SDA methods to hydrologic modeling, it is important to explicate a formal methodology which recursively estimates the states in the model, while simultaneously also identifying the most likely values of the model parameters. The major difference between such a framework and the aforementioned classical model calibration strategies is that this approach no longer assigns the primary uncertainty in the input-output representation to uncertainty in the parameters, but explicitly accounts for input, output and model structural errors during model calibration.

This Chapter presents a Simultaneous parameter Optimization and Data Assimilation method entitled, SODA, which combines the strengths of the parameter search efficiency and explorative capabilities of the Shuffled Complex Evolution Metropolis (SCEM-UA) algorithm [Vrugt et al., in Chapter 3 of this thesis], with the power and computational efficiency of the Ensemble Kalman Filter [Evensen, 1994] to provide a better treatment of the input, output, parameter and model structural uncertainties in hydrologic modeling. The implementation of SODA requires a change in philosophy because it means that the model calibration procedure can no longer be considered as finding a single set of parameter values which generate the best possible long-term model forecasts. Instead, the optimal set of parameters identified with SODA only has meaning in combination with a SDA method, and consequently, will (per definition) not generate the best possible forecasts when no state adjustments are allowed, which is typically the case when evaluating the predictive capabilities of the model during an independent evaluation period. We hypothesize that the implementation of SODA will result in, (i) meaningful prediction uncertainty bounds on the model simulations, (ii) parameter estimates which better represent system properties that are less corrupted by modeling errors, thereby increasing the prospects of finding useful regionalization relationships, and (iii) a time series of recursive state and output adjustments, whose interpretation will generate inspiration to improve our model concepts and as such our understanding of the functioning of hydrologic systems.

This Chapter is organized as follows. Section 7.2 surveys the limitations of current model calibration strategies, which do not explicitly treat input, output and model structural uncertainty, but assign the uncertainty in the input-output representation primarily to uncertainty in the parameter estimates. In section 7.3 and 7.4 we discuss the rationale and architecture of SDA and global optimization methods, and demonstrate how the strengths of these methods can be
combined to derive the hybrid SODA framework. In section 7.5 we illustrate the power and applicability of SODA by means of a standard mathematical study using the three-parameter highly nonlinear Lorenz model \cite{Lorenz1963}, and a classical rainfall-runoff modeling example. In this section we also introduce a nonparametric variance estimator, which is especially designed to estimate the measurement error of output data. Finally, in section 7.6 we summarize the results.

7.2. Basic Inverse Formulation

The fundamental problem with which we are concerned is to simultaneously estimate parameter values and state variables in a hydrologic model using a historical record of input – output data. The formulation of this resulting inverse problem can be expressed in a generic form if we assemble the state variables in the nonlinear hydrologic model at time \( t \) into the state vector \( \psi \). The evolution of this state vector is described with:

\[
\psi_{t+1} = \eta(\psi_t, \bar{x}_t, \theta) \tag{7.1}
\]

where \( \psi \) is a vector of \( m \) unknown state variables, \( \eta(\cdot\) represents the nonlinear hydrologic model used to simulate the state evolution, \( \bar{x} \) is an observed forcing field, \( \theta \) is a set of \( p \) model parameters, and \( t \) denotes time. We further assume that:

\[
\theta \in \Theta \in \mathbb{R}^p \tag{7.2}
\]

where \( \mathbb{R}^p \) denotes the \( p \)-dimensional Euclidean space. If \( \Theta \) is not the entire domain space \( \mathbb{R}^p \), the inverse problem is said to be constrained. In hydrologic modeling, the feasible parameter space \( \Theta \) can usually be properly constrained using realistic upper and lower bounds on each of the model parameters. Let \( \tilde{Y} = \{ \tilde{y}_1, ..., \tilde{y}_n \} \) denote the vector of streamflow measurement data available at time steps 1, \( ..., n \) and let \( Y(\theta) = \{ y_1(\theta), ..., y_n(\theta) \} \) represent the corresponding vector of model output predictions using the parameter values \( \theta \). These model output predictions are related to the model state according to:

\[
y_i = H(\psi_i) \tag{7.3}
\]
where \( H(\cdot) \) is the measurement operator, which maps the state space into the measurement or model output space. The tilde over \( X \) and \( Y \) emphasizes that these quantities are measured and hence subject to sampling and measurement error. The differences between the model-simulated output and measured data can be represented by the residual vector:

\[
E(\theta) = G[Y(\theta)] - G[\hat{Y}] = \{e_1(\theta), ..., e_{\eta}(\theta)\}
\]

(7.4)

where the function \( G(\cdot) \) allows for various user-selected linear or nonlinear transformations.

The classical approach to estimating the parameters in Eq. (7.1) is to ignore input uncertainty \( (X = \hat{X}) \) and to assume that the predictive model \( \eta \) is a correct, or at least accurate, representation of the underlying physical data-generating process. In line with classical statistical estimation theory, the residuals in Eq. (7.4) are then assumed to be mutually independent (uncorrelated), Gaussian distributed, with a constant variance. Under these circumstances, the traditional “best” parameter set in Eq. (7.1) can be found by minimizing the following lumped simple least square (SLS) objective function with respect to \( \theta \):

\[
F_{\text{SLS}}(\theta) = \sum_{i=1}^{n} e_{i}^2
\]

(7.5)

The limitations of neglecting input errors and lumping observed response and model structural error into one single white noise term, becomes immediately apparent when inspecting the time series of residuals after model calibration. To illustrate this, consider Figure 7.1, which displays a 5-month portion of the measured and simulated hydrographs for the Leaf River basin in Mississippi. The HYMOD conceptual watershed model was calibrated using the Shuffled Complex Evolution (SCE-UA) global optimization algorithm developed by Duan et al. [1992] using the SLS criterion and 1 year of calibration data. Notice, that the fit to the observed data can be considered quite good, but that the residuals exhibit considerable variation in bias (non-stationarity), variance (heteroscedasticity) and correlation structure under different hydrologic conditions. Indeed, the common approach of dealing with model structural and data errors, as being “small” or somehow “absorbed” into the output error residual needs serious reconsideration. Moreover, given the presence of these errors, overconditioning of the model to a single parameter set is unreasonable, and cannot be justified.
One set of responses set forth to directly address the problem of overconditioning is to abandon the Frequentists approach of believing that the model parameters in Eq. (7.1) are fixed but unknown, and to adopt a Bayesian viewpoint which allows the identification of a plausible set of values for the parameters of the model given the available data. The Bayesian approach treats the model parameters in Eq. (7.1) as probabilistic variables having a joint posterior probability density function (pdf), which summarizes our probabilistic belief about the parameters $\theta$ in the light of the observed data $\tilde{y}$. Examples of Bayesian approaches to hydrology include the Generalized Likelihood Uncertainty Estimation (GLUE) framework of Beven and Binley [1992] and the Bayesian Recursive Estimation (BaRE) approach of Thiemann et al [2001] for representing model parameter and prediction uncertainty within the context of Monte Carlo analysis, and the Shuffled Complex Evolution Metropolis (SCEM-UA) global optimization algorithm of Vrugt et al [Chapter 3 of this thesis] for simultaneously estimating the traditional "best" parameter set along with a sample set of parameter values describing the probabilistic representation of the remaining parameter uncertainty (which is then used to generate probabilistic forecasts). We note that the SCEM-UA and BaRE methods operate within the
context of the classical Bayesian framework (which makes implicit assumptions about the correctness or adequacy of the model of the data generating process), while GLUE requires the user to make other kinds of subjective decisions including the selection of a cutoff threshold that separates behavioral from non-behavioral parameter sets.

An alternative set of responses set forth to identify model structural inadequacies, is to use Recursive Parameter Estimation algorithms, which provide better extraction of the information in the data, because the temporal aggregation associated with traditional batch processing is reduced. Examples of recursive algorithms which can be applied to hydrologic models, include the PIMLI and recursive SCEM-UA algorithms [Vrugt et al., in Chapter 3 of this thesis 2002; Vrugt et al., 2003b], the DYNIA approach [Wagener et al., 2003], the BaRE algorithm [Thiessen et al., 2001], and the application of Transfer Functions with time-varying parameters identified using Instrumental Variable techniques [Young, 2001]. While such methods introduce some much needed creative thinking into the field of hydrologic model identification, the use of such methods to help diagnose and quantify model structural errors has yet to be satisfactorily demonstrated.

A third set of responses set forth to more directly confront the problem of model structural errors is to pose the model calibration problem into a multi-objective framework, as was advocated by Gupta et al. [1998]. A major consequence of model structural imperfection is that the model is incapable of reproducing all aspects and portions of the output data with a single parameter set. By employing a number of complementary criteria in the optimization procedure, and analyzing the tradeoffs in the fitting of these criteria, the hydrologist is able to better understand the limitations of current hydrologic model structures, and gain insights into possible model improvements. The resulting Pareto solution set of parameters defines a kind of parameter uncertainty attributable to model structural errors. Yapo et al. [1998] and later Vrugt et al. [Chapter 5 of this thesis] presented the MOCOM-UA and MOSCEM-UA algorithms, respectively, capable of solving, in a single optimization run, the multi-objective problem posed by this approach. Unlike the previous methods, the multi-objective approach explicitly acknowledges the existence of model structural errors during the parameter estimation procedure. However, because of subjectivity in the selection of the measures [Gupta et al., 1998] and data subsets [see Boyle et al., 2000] to be included in the multi-objective vector, considerable fuzziness in the specification of the Pareto optimal parameter space and thus related model structural errors exists. Moreover, multi-objective calibration is typically response focused, and does not explicitly consider the influence of input errors [see Kavetski et al., 2003]. Like the
previous methods all sources of uncertainty are treated primarily as parameter uncertainty, giving rise to the concept of Pareto optimality.

Finally, the last set of responses set forth to more directly confront the problem of input errors is to define more realistic error models. One such approach is the recently proposed BAyesian Total Error Analysis (BATEA) of Kavetski et al. [2003], which imposes a hitherto missing rigor in environmental modeling by requiring the specification of realistic input uncertainty models. From the surveyed methods in this section, BATEA is the only method that explicitly takes into account input errors in the development of the likelihood function, and as such no longer assigns all the uncertainty in the input–output representation to uncertainty in the parameters. However, it is not yet clear how this likelihood function can be easily extended to include for model structural and output errors as well.

In summary, current model calibration strategies are poorly suited to the task of explicitly treating the various important sources of uncertainty associated with the application of hydrologic models. In the next two sections of this Chapter, we will discuss the challenging question of how to develop a model calibration strategy that explicitly accounts for input, output and model structural errors.

### 7.3. Sequential Data Assimilation

Sequential data assimilation methods, at least in principle, specify the uncertainty in the system states that arises from imperfect process representation and from input and output data uncertainty. To facilitate the description of the KF, we start by writing the model dynamics in Eq. (7.1) as a stochastic differential equation:

\[
\frac{d\psi}{dt} = \eta(\psi_{t-1}, \tilde{X}_{t-1}, \theta)dt + dq_{t-1}
\]

(7.6)

where \( dq \in \mathbb{R}^n \) is a dynamical noise term representing errors in the model formulation. This stochastic forcing term tends to flatten the probability density function of the model states during the integration. In section 7.3.2 of this Chapter we further elaborate on the simulation of \( q \). The observation equation, Eq. (7.3), also has a random additive error \( \varepsilon \), called the measurement error:

\[
y = H(\psi) + \varepsilon, \quad \varepsilon \sim N(0, \sigma^2) \]

(7.7)
where \( \sigma^2 \) signifies the error deviation of the measurements. At each measurement time \( t = 1, \ldots, n \) when an observation becomes available compute the output forecast error, \( z_t \):

\[
z_t = \tilde{y}_t - H(y'_t)
\]  

(7.8)

and update the forecasted states, \( y'_t \) using the standard analysis equation:

\[
y'_t = y'_t + K_t \left[ \tilde{y}_t - H(y'_t) \right]
\]  

(7.9)

where \( y'_t \) is the updated or analyzed state, and \( K_t \) denotes a matrix of weights (Kalman gain), which is computed as:

\[
K_t = \Sigma_t H^T \left[ H \Sigma_t H^T + \Sigma^t \right]^{-1}
\]  

(7.10)

where \( \Sigma_t \) and \( \Sigma^t \) denote the covariance matrices of the stochastic model error term (= \( q_q^T \)) and observations (= \( e_e^T \)), respectively. The difference between the forecasted and updated state is commonly referred to in the literature as the state innovation, \( \psi'_t \):

\[
\psi'_t = y'_t - y'_t
\]  

(7.11)

which mapped in the output space is called the output innovation, \( I'_t \):

\[
I'_t = H(y'_t) - H(y'_t)
\]  

(7.12)

The analyzed state \( y'_t \) then recursively feeds the next state propagation step in the model:

\[
y'_{t+1} = \eta(y'_t, \tilde{y}_t, \theta)
\]  

(7.13)

Since, uncertainty in the model structure and output data can be specified through the stochastic forcing term (\( q \)) and output measurement error (\( e \)), respectively, and uncertainty in the input data
can be taken into account through stochastic perturbations of the elements of the input \( \tilde{X} \), the KF offers a very general framework for dealing with all sources of uncertainty.

Even though the KF or EKF implement a more sensible model calibration strategy than traditional model calibration methods, as these filters explicitly specify the uncertainty in the system states that arises from imperfect process representation and from input and output data uncertainty, their widespread application has been limited by the strong non-linearity of hydrologic models, and the computational requirements associated with the storage and forward integration of the error covariance matrix. To resolve these two major problems, Evensen [1994] proposed the Ensemble Kalman Filter (EnKF), which 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. The EnKF avoids many of the problems associated with the traditional EKF method, e.g., 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 easy of implementation and computational efficiency of the EnKF make the method an attractive option for data assimilation in the meteorologic, oceanographic and hydrologic sciences [Evensen and van Leeuwen, 1996; Houtekamer and Mitchell, 1998; Lermusiaux, 1999; Madsen and Cañizares, 1999; Keppenne, 2000; Reichle et al., 2002; among others].

### 7.3.1. The Ensemble Kalman Filter

The EnKF uses an ensemble of model trajectories to solve for Eqs. (7.6) to (7.12). The description of the basic algorithm of the EnKF below follows Evensen [1994] and is further illustrated in Figure 7.2.

1. **Generate initial ensemble.** Sample \( N \) combinations of \( m \) model states, \( \psi^j, j = 1, ..., N \) randomly from the prior distribution and store them in a matrix \( A[1:m,1:N] \):

   \[
   A = (\psi^1, ..., \psi^N) \in \mathbb{R}^{m \times N}
   \]  

2. **Forecast step.** Propagate each of the \( N \) ensemble members of \( A \) forward in time with the nonlinear model \( \eta(\cdot) \), using a fixed set of model parameters \( \theta \) and a corresponding ensemble of \( N \) stochastic forcing fields:

   \[
   \tilde{X} = \eta(\tilde{X}, \theta, \psi^j, u^j) \]
Explicit details on how to simulate \( q_{t+1} \) will be given in the next section.

\[ \psi_{t+1} = \eta(\psi_{t+1}, \tilde{X}_{t+1}, \theta) + q_{t+1} \]  

(7.15)

**Figure 7.2.** Flowchart of the Ensemble Kalman Filter (EnKF) used to recursively estimate state variables.
[3] Compute the covariance matrix of the forecasted states. At each time step \( t \) compute the covariance matrix, \( \Sigma^* \in \mathbb{R}^{m \times m} \) of the ensemble forecasted states in matrix \( A \):

\[
\Sigma^* = \frac{A'(A')^T}{N-1}
\]

(7.16)

where \( A' = A - \bar{A} \) and \( \bar{A} \) denotes the ensemble mean of the forecasted states.

[4] Compute the mean ensemble forecast error. At time \( t \) when a measurement becomes available, compute the mean forecast error of the ensemble:

\[
\bar{e}_t = H(A) - \tilde{y}_t
\]

(7.17)

[5] Generate vector of observations. At each measurement time, generate an ensemble of \( N \) observations from a distribution with mean equal to the observation, \( \tilde{y}_t \), and covariance equal to \( \Sigma^* \):

\[
\tilde{y}_j' = \tilde{y}_j + \epsilon_j'
\]

\[
\epsilon_j' \sim N(0, \sigma_j^2) \quad j = 1, \ldots, N
\]

\[
\Sigma^* = \epsilon \epsilon_j'^T
\]

(7.18)

and store them in a matrix \( D[1:1:N] \):

\[
D = (\tilde{y}_1', \ldots, \tilde{y}_N') \in \mathbb{R}^{1 \times N}
\]

(7.19)

This particular step ensures that the spread of the updated ensemble is consistent with the true estimation error covariance [Burgers et al., 1998].

[6] State update or analysis step. At each measurement time, update each ensemble member of \( A \) using the following analysis equation:

\[
A = A + \Sigma^* H^T (H \Sigma^* H^T + \Sigma^*_t)^{-1} [D - H(A)]
\]

(7.20)
where the product $\Sigma^w H^T (H \Sigma^w H^T + \Sigma^\nu)^{-1}$ is a numerical approximation of the Kalman gain, $K$, in Eq. (7.9).

[7] **Check stop criterion.** If $t$ is less than the total number of time steps, $n$, return to step 2, otherwise stop.

In summary, the EnKF uses a large ensemble of model trajectories to approximate the probability density of the model states at each time step. The mean of this ensemble represents the "best" state estimate, whereas the variance provides a measure of the spread of the ensemble members (uncertainty). For more information about the derivation of the EnKF and its practical implementation, please refer to Evensen [1994].

### 7.3.2. Simulation of the Stochastic Model Error Term

To successfully implement the EnKF it is particularly important to specify accurate values of the output measurement error and stochastic forcing term used to simulate the time evolution of dynamical model noise. Hence, it is not difficult to see that the size of these terms directly controls the spread among the ensemble members, and as such the mean ensemble model forecasts and associated prediction uncertainty bounds. Fortunately, in most situations a reasonable a-priori estimate can be made of the measurement error, $\sigma^m$, of the observations. A more delicate issue however, is how to retrieve a reasonable estimate of the stochastic forcing term. This term reflects many sources of uncertainty including model discretization errors and neglected dynamical processes.

Based on recommendations in the classic treatise of the EnKF in Evensen [1994] we used the following equation to simulate the time evolution of the stochastic forcing:

$$s_t = \phi s_{t-1} + \sqrt{1 - \phi^2} w_{t-1}$$

(7.21)

where $w_{t-1}$ is an $m \times 1$ vector of draws from a normal distribution, and $\phi \in [0,1)$ represents the first-order decorrelation of the stochastic forcing. The form of Eq. (7.21) ensures that, with an increasing number of elements, $s$ will approach a distribution having mean equal to zero and variance equal to one. A value for $\phi = 0$ generates a sequence of white noise terms, while $\phi = 1$
SIMULTANEOUS PARAMETER AND STATE ESTIMATION

will remove the stochastic forcing. Based on random walk theory, the nonlinear model \( \eta(\cdot) \) can then be written as:

\[
\psi_i = \eta(\psi_{i-1}, \hat{X}_{t-1}, \theta) + \tau_{t-1} \sqrt{\Delta t} \sigma
\]  

(7.22)

in which \( \Delta t \) denotes the time step, \( \sigma \) is an \( m \times 1 \) vector of error deviations of individual states and \( \tau \) is a factor which ensures that the forcing error over time becomes independent of \( \phi \) and \( \Delta t \) and is computed as [Evensen, 1994]:

\[
\tau = \frac{1}{\sqrt{\Delta t / -2\phi \phi_j + 2\phi_j^2}}
\]  

(7.23)

where \( j \) denotes the number of steps in each time unit \( \Delta t (\Delta t = 1) \).

While this equation is frequently used in meteorology and oceanography to simulate the time evolution of the stochastic forcing, our attempt to apply this equation to conceptual hydrologic modeling has revealed one major drawback. When the states in the model do not have a clear physical meaning, as is typically the case with conceptual hydrologic models, it is difficult to specify accurate values of the error deviations of individual states. To circumvent this problem, we decided to specify \( \sigma \) in the output space of the model and to use this output perturbation to compute the corresponding corrupted states in the model:

\[
\psi_i = \eta(\psi_{i-1}, \hat{X}_{t-1}, \theta) + \tau_{i-1,1} \sum_{\eta} H[\sum_{\eta} \Sigma H']^{-1} \sqrt{\Delta t} \sigma_{\eta}
\]  

(7.24)

where \( \Sigma_{\eta} \) represents the covariance matrix of the forecasted states and \( \sigma_{\eta} \) denotes the error deviation of the stochastic forcing. Among various alternatives, this implementation was shown to work well for simulating the time evolution of the stochastic forcing for a variety of test problems.

7.4. Combining Global Optimization and Sequential Data Assimilation

Even though the EnKF provides a general framework to account for input, output, and model structural uncertainty, a defining characteristic of the filter is its focus on state estimation alone,
without recourse to parameter estimation. When applying the EnKF to recursively estimate state values, it is common practice to fix the model parameters to some predefined values, thereby explicitly ignoring the effects of parameter uncertainty and interaction. This section discusses how the strengths of global optimization and the EnKF can be merged to derive a hybrid framework, which recursively tracks the model states, while simultaneously also estimating the values of the model parameters (and their associated uncertainty).

To extend the applicability of the EnKF to simultaneous state and parameter estimation we must do two things. First, we must specify a mathematical criterion that measures the “closeness” between the EnKF derived mean ensemble model forecast and the corresponding measurement, which can be used to extract the information content in the data and transform it into estimates for the model parameters. This issue will be addressed in this paragraph. Second, we must implement a method that can effectively and efficiently minimize this measure, thereby preferably also generating an estimate of parameter uncertainty. We will tackle this problem in the next paragraph. To estimate the model parameters, we adopt a classical Bayesian approach using the following posterior density criterion, \( p(\theta | \tilde{Y}) \), [Box and Tiao, 1973]:

\[
p(\theta | \tilde{Y}) \propto \left( \sum_{i=1}^{n} |\tilde{z}(\theta)|^2 \right)^{-\frac{1}{2}q}
\]

in which \( \tilde{z}(\theta) \) denotes the time series of mean ensemble forecast errors corresponding to the parameter set \( \theta \). Using this criterion, the best set of model parameters is defined as that set which is on average associated with the best one-observation-ahead model forecasts.

To generate samples from Eq. (7.25), and to summarize the posterior parameter pdf using statistical moments and histograms, we use an implementation of the Shuffled Complex Evolution Metropolis (SCEM-UA) algorithm. The SCEM-UA algorithm is a general-purpose global optimization algorithm that provides an efficient estimate of the most likely parameter set and its underlying posterior probability distribution within a single optimization run [see Vrugt et al., in Chapter 3 of this thesis]. The algorithm is a Markov Chain Monte Carlo (MCMC) sampler, which generates a sequence of parameter sets \( \{\theta^{(0)};\theta^{(2)};\ldots;\theta^{(k+1)}\} \) that converges to the stationary posterior distribution for a large enough number of simulations \( k \). The SCEM-UA algorithm is related to the successful SCE-UA global optimization method, but uses the Metropolis-Hastings (MH) search strategy [Metropolis et al, 1953; Hastings, 1970] instead of the Downhill Simplex method for population evolution, and is therefore able to simultaneously infer
both the most likely parameter set and its underlying posterior probability distribution within a single optimization run. A detailed description and explanation of the method appears in Vrugt et al. [Chapter 3 of this thesis], and so will not be repeated here.

In brief, the SCEM-UA method involves the initial selection of a population of points distributed randomly throughout the $p$-dimensional feasible parameter space. In the absence of prior information about the location of the maximum likelihood value a uniform sampling distribution is used. For each point the complete system of equations of the EnKF, as outlined in steps [1] to [7] in section 7.3.1, is executed and the time series of mean ensemble forecast errors returned to the SCEM-UA algorithm. This time series is subsequently inserted in Eq. (7.25) to calculate the posterior density for each point. The population of parameter sets is subsequently partitioned into a number of complexes, and in each complex a parallel sequence is launched from the point that exhibits the highest posterior density. A new candidate point in each sequence is generated using a multivariate normal distribution either centered on the current draw of the sequence or the mean of the points in the complex augmented with the covariance structure induced between the points in the complex. The Metropolis-annealing [Metropolis et al., 1953] criterion is used to test whether the candidate point should be added to the current sequence. Finally, the new candidate point is shuffled into the original population of complexes. The evolution and shuffling procedures are repeated until the Gelman – Rubin convergence diagnostic for each of the parameters demonstrates convergence to a stationary posterior target distribution [Gelman and Rubin, 1992]. Experiments conducted using standard mathematical test problems have shown that the SCEM-UA derived posterior distribution closely approximates the target distribution [Vrugt et al., in Chapter 3 of this thesis]. Throughout the remainder of this Chapter, we refer to this Simultaneous Optimization and Data Assimilation procedure, as SODA.

The overriding characteristic of the SODA methodology is the improved treatment of input, output, parameter and model structural uncertainties, during model calibration. Theoretically, if the model structure would be correct, and the input and output data are observed without error, no state adjustments are needed during model calibration, and the set of parameters that maximizes the posterior density in Eq. (7.25), will also be associated with, on average, the smallest output residuals, when performing traditional batch calibration without state adjustments (as done in section 7.2). However, because of errors in the model structure and input (forcing) and output data this will generally not be the case. The major objective of this Chapter is to demonstrate how the hybrid SODA strategy can be used to simultaneously estimate model parameters and state variables, resulting in reliable model prediction uncertainty
bounds and a time series of state and output innovations whose interpretation will generate inspiration to improve our model concepts and as such our understanding of the functioning of hydrologic systems.

7.5. Case Studies

We illustrate the power and applicability of SODA by means of two different case studies. The first is a classical mathematical study using the three-parameter highly nonlinear Lorenz model \cite{Lorenz1963}, to demonstrates the ability of SODA to simultaneously estimate state variables and parameter values when confronted with highly nonlinear model dynamics. The second case study explores the usefulness of SODA by application to hydrologic modeling using the simple HYMOD conceptual watershed model and historical streamflow data from the Leaf River Watershed in Mississippi. In this study we are especially concerned with the estimation of model structural errors and present a novel nonparametric method, which is especially designed to estimate the measurement error of streamflow data.

7.5.1. Case Study I: A Highly Nonlinear Case: The Lorenz Equations

To demonstrate that SODA is indeed able to simultaneously estimate state variables and model parameters when confronted with highly nonlinear model dynamics, the first case study considers the three-parameter Lorenz model \cite{Lorenz1963}. This model consists of a system of three nonlinear and coupled ordinary differential equations:

\[
\begin{align*}
\frac{dx}{dt} &= \gamma(y - x) \\
\frac{dy}{dt} &= \rho x - y - xz \\
\frac{dz}{dt} &= x y - \beta z
\end{align*}
\] (7.26)

where \(x(t), y(t),\) and \(z(t)\) denote the dependent variables and \(\gamma, \rho,\) and \(\beta\) represent model parameters. Because of its highly nonlinear nature, this model has served as a testbed in the field of data assimilation for examining the properties for various data assimilation methods. A reference solution of \((x(t), y(t), z(t))\) with output print step \(\delta t = 0.25\) was computed for \(t \in [0, 40]\) by solving Eq. (7.26) starting from the initial condition \((x_0, y_0, z_0) = (1.508870, -1.531271, 25.46091)\)
and fixed values for the parameters of $\gamma=10$, $\rho=28$, and $\beta=8/3$. Synthetic observations were subsequently computed by imposing normally distributed noise, with mean zero and variance equal to 2.0, to the $x$, $y$ and $z$-values in this reference solution.

The error covariance of the stochastic forcing is defined to be diagonal with variances equal to 2.00, 12.13, and 12.31 for the three equations in (7.26) respectively, and is included in the model through the stochastic forcing term in Eq. (7.22). The error variances of the initial conditions and observations were taken to be the same values as those used for generating the synthetic observations. The values for the parameters, initial conditions, and error variances used in the setup of this case study are identical to those used in similar studies reported in the literature [see e.g. Evensen, 1994; Miller et al., 1994; Evensen, 1997; Evensen and van Leeuwen, 2000; among others].

The SODA algorithm was used to simultaneously estimate the state variables and model parameters using the synthetically generated observations and the density criterion in Eq. (7.25). A search population of 50 points in combination with 5 parallel sequences was selected based on recommendations in our previous work. The feasible parameter space was taken to be a uniform hypercube between $[0-30]$, $[0-50]$ and $[0-10]$ for the parameters $\gamma$, $\rho$, and $\beta$ respectively. Furthermore, an ensemble of 100 members was used to compute the error statistics with the EnKF, on the basis of experimental evidence that larger ensemble sizes gave only marginal improvements in the approximation of the probability density of the model states. The estimated ensemble prediction uncertainty ranges associated with the computed $x$-values in Eq. (7.26) and corresponding to the most likely SODA identified parameter set is shown in Figure 7.3a.
Figure 7.3. Results obtained with SODA for the three-parameter highly nonlinear Lorenz model: (a) Prediction uncertainty ranges corresponding to the best parameter set identified with SODA and 100 ensemble members. The solid dots denote the observations, (b-d) Sampled $\gamma$, $\beta$, and $\rho$ parameters in five different sequences (coded with different symbols) during the evolution of SODA to the posterior target distribution. The asterisks at the right hand side of these figures denote the “true” parameter values used to generate the synthetic observations.

The solid circles correspond to the observed $x$-values, whereas the dashed black line represents the evolution of the mean ensemble prediction. Note, that the prediction uncertainty ranges generally bracket the observations very well, indicating that the EnKF does a good job in tracking the state transitions. Further, the width of the ensemble prediction uncertainty bounds is consistent with the specified error variance of the model and observations.

The transitions of the parameters $\gamma$, $\rho$, and $\beta$ in each of the five sequences (Markov Chains) during the evolution of SODA to the stationary posterior distribution is illustrated in the Figures 7.3b-d respectively. For clarity, the three different parallel sequences are coded with different symbols. The asterisks at the right hand side indicate the “true” values of the parameters. The 1-D scatterplots of the sampled parameter space demonstrate that at early stages during the evolution, the individual sequences tend to occupy different regions of the parameter space. At a later stage during the evolution, however, all of the individual sequences have been able to fully explore the feasible parameter space, and indeed they assign the highest probability
to the “true” parameter values used for generating the synthetic observations. All in all, we can conclude that SODA is able to simultaneously estimate state variables and model parameters when confronted with highly nonlinear model dynamics, so that we can proceed with the next case study.

7.5.2. Case Study II: The HYMOD Conceptual Watershed Model

We illustrate the power and applicability of the SODA algorithm to hydrologic modeling by application to the HYMOD conceptual watershed model using historical data from the Leaf River watershed (1950 km$^2$) located North of Collins, Mississippi. The data, obtained from the Hydrologic Research Laboratory (HL), consists of mean areal precipitation (mm/day), potential evapotranspiration (mm/day), and streamflow (m$^3$/s). Forty consecutive years of data (1948-1988) are available for this watershed, representing a wide variety of hydrologic conditions. The HYMOD model, consisting of a relatively simple rainfall excess model, described in detail by Moore [1985], connected with two-series of linear reservoirs (three-identical quick and a single reservoir for the slow response). A schematic overview of HYMOD appears in Figure 7.4.

![Figure 7.4](image)

**Figure 7.4.** Schematic representation of the HYMOD conceptual watershed model.

The model has five parameters: the maximum storage capacity in the catchment, $C_{\text{max}}$ (L), the degree of spatial variability of the soil moisture capacity within the catchment, $h$ (-), the factor distributing the flow between the two series of reservoirs, $\alpha$ (-), and the residence time of the linear quick and slow reservoirs, $R_q$ (T) and $R_s$ (T), respectively. The upper and lower bounds that define the prior uncertainty ranges of these parameters appear in Table 7.1.
Table 7.1. Prior uncertainty ranges and description of the HYMOD model parameters.

<table>
<thead>
<tr>
<th>Par.</th>
<th>Unit</th>
<th>Description</th>
<th>Minimum</th>
<th>Maximum</th>
</tr>
</thead>
<tbody>
<tr>
<td>$C_{\text{max}}$</td>
<td>[mm]</td>
<td>Maximum storage watershed</td>
<td>200.00</td>
<td>500.00</td>
</tr>
<tr>
<td>$b_{sp}$</td>
<td>[-]</td>
<td>Spatial variability soil moisture capacity</td>
<td>0.10</td>
<td>2.00</td>
</tr>
<tr>
<td>$\alpha$</td>
<td>[-]</td>
<td>Distribution factor between two reservoirs</td>
<td>0.50</td>
<td>0.99</td>
</tr>
<tr>
<td>$R_s$</td>
<td>[d]</td>
<td>Residence time slow flow reservoir</td>
<td>0.00</td>
<td>0.10</td>
</tr>
<tr>
<td>$R_q$</td>
<td>[d]</td>
<td>Residence time quick flow reservoir</td>
<td>0.30</td>
<td>0.70</td>
</tr>
</tbody>
</table>

Because the HYMOD model and the Leaf River data have been discussed extensively in previous work [e.g., Yapo et al., 1996; Vrugt et al., in Chapters 3 and 5 of this thesis; Boyle et al., 2002; Misirli, 2003], we will not describe the details of these here. To implement the SODA framework outlined above, it is further necessary to specify an appropriate value for the parameter $\phi$ in Eq. (7.21). As our knowledge about a correct value for this first-order decorrelation coefficient is quite poor, we added this parameter to the optimization problem, using a feasible range between [0-1]. Based on our experience with the model we used 3 years of streamflow data (28 July 1952 – 28 July 1955) to illustrate that the SODA method is relatively simple to implement and helps to distinguish between the various sources of uncertainty in hydrologic modeling.

7.5.2.1. Derivation of the Measurement Error of Streamflow Data

To solve the model calibration problem using SODA, we must do two things. First, a realistic estimate of the error deviation of the output (streamflow) data must be retrieved. This issue will be addressed in this section. Second, we must find a way to obtain a reasonable estimate of input and model structural uncertainty. We will tackle this problem in the next section.

Unfortunately, the problem of estimating the measurement error of streamflow data has not proved to be simple. To further elaborate on this problem, consider observations obtained through:

$$\tilde{y}_t = h(t) + e_t, \quad e_t \sim N(0, \sigma_e^2), \quad (t = 1, ..., n)$$

(7.27)

where $h(t)$ is the actual streamflow at time $t$ and the errors $e_t$ are independent random variables with zero mean and unknown variance, $\sigma_e^2$. Without any doubt, most attention in hydrologic modeling has been given to a correct estimation of the function $h$. However, estimation of the
error deviation is nearly as important as the estimation of $h$ itself, to obtain reasonable confidence intervals on the model predictions.

During the last two decades several nonparametric methods have been proposed which estimate the error deviation in Eq. (7.27) [see e.g. Rice, 1984; Hall et al., 1990; Seifert et al., 1993; Dette et al., 1998]. All these methods involve differencing the original time series, $y_i$, and estimating the error deviation as:

$$\hat{\sigma}_i = \sqrt{\frac{1}{2(n-1)} \sum_{j=2}^{n} (\hat{y}_i - \hat{y}_{i-j})^2}$$  \hspace{1cm} (7.28)

The underlying assumptions of this approach include that (1) $h(t)$ as a function of $t$ is sufficiently smooth, (2) the sampling interval is high compared to the typical timescale of $h(t)$, and (3) the error terms, $e_i$, exhibit a constant variance (homoscedastic). While the first two assumptions can be considered quite reasonable for a time series of daily streamflow data, the assumption of homoscedasticity of the streamflow error terms is not very realistic. For instance, Sorooshian and Dracup [1980] commented that large flows tend to have larger error variances compared to smaller flows, partially because of the non-linear nature of the rating curve used to transform stage measurements to flow volume estimates. To enable the identification of these non-homogeneous (heteroscedastic) streamflow errors, we decided not to compute an average standard error estimate of the total time series, as suggested in Eq. (7.28), but instead, to apply the following nonparametric error deviation estimator locally in the time series:

$$\hat{\sigma}_i = \sqrt{\frac{2n}{n} \left( \Delta^u \hat{y}_i \right)^2}$$  \hspace{1cm} (7.29)

where $\Delta^u$ denotes the difference operator applied $n$ times. It can be readily verified that the estimator in Eq. (7.29) is insensitive to polynomial trends in $h(t)$ up to order $u$. In the literature more sophisticated higher-order differencing procedures have been proposed [Hall et al., 1990]. However, investigations with numerically generated streamflow data showed that Eq. (7.29) with the choice $u = 3$ works well in practice.

To demonstrate the validity of the proposed approach for estimating the error deviation of streamflow data, we synthetically generated two different test problems. Synthetic daily streamflow data, $\{y_1, \ldots, y_n\}$ were first generated for the period June 28, 1952 to September 30,
1961, by driving the HYMOD model with observed mean areal rainfall of the Leaf River watershed and values for the parameters identical to those used in Fig. 7.1. This output time series was subsequently corrupted by adding one of the two following error models to the data to yield two different time series of 'observed' streamflow data; (1) a homoscedastic error model: \( \varepsilon_i \sim N(0,10) \); and (2) a heteroscedastic error model: \( \varepsilon_i \sim N(y,0.1y) \). Both these time series were subsequently used in conjunction with Eq. (7.29) to estimate the error deviation of the streamflow measurements. The results of this analysis are summarized in Figure 7.5, which presents scatterplots of 'observed' streamflow data versus estimated error deviations for the (a) homoscedastic, and (b) heteroscedastic error case.

![Figure 7.5](image)

**Figure 7.5.** Synthetic case studies to demonstrate the validity of the nonparametric estimator to derive the error deviation of streamflow measurements; (a) The homoscedastic error case, and (b) Heteroscedastic error case. The solid and dashed black lines in each of the figures denote the estimated and actual error model, respectively.

To further facilitate graphic interpretation of the results, the dark black line in each of the Figures represents the optimal fit of a spline function through the data, whereas the dashed line corresponds to the original error model used to corrupt the streamflow data. Notice, that the agreement between the predicted and actual error model can be considered quite excellent. This is especially true for the homoscedastic error case. Although not further reported here, we performed a variety of synthetic experiments with different error distributions, and found that the nonparametric estimator in Eq. (7.29) provides a close approximation to the actual error model, when the time series of streamflow data is sufficiently long (> 2 years).

To unravel the relationship between flow level and measurement error for the Leaf River watershed, Figure 7.6, presents a similar scatterplot as Fig. 7.5 for this real world data set. Again, the black line denotes the optimal fit of a spline function through the scattered data. The pattern
of the scatter plot and slope of the regression function clearly indicate the presence of heteroscedastic errors in the streamflow data.

![Figure 7.6. Scatterplot of observed streamflow data versus estimated error deviation of the measurements using 10 years of data from the Leaf River watershed. The solid line represents the fit of a spline function through the data.](image)

In general, the size of the streamflow error almost log-linearly increases from 0.01 to about 80 m³/s in the flow range of 2 to 1000 m³/s respectively. These results provide strong support for the claim by Sorooshian and Dracup [1980], that streamflow data exhibit non-homogeneous (heteroscedastic) errors. To optimally exploit the information contained in the scatter plot of Fig. 7.6, it would seem most productive to first cluster the data into various flow levels, and then to approximate the corresponding error probabilities with Gaussian mixture models. However, because the purpose of this case study is illustrative, we decided not to follow such an approach, but rather to implement the fitted nonlinear regression function in SODA, to relate the error deviation of the streamflow data to the flow level.

### 7.5.2.2. Derivation of Input Error Model and Error Deviation of the Stochastic Forcing

In the absence of a compelling basis for the assignment of a reasonable input error model to stochastically perturb the elements of $X$, we decided to merge input and model structural errors into a single forcing term. The approach that we propose here to estimating the total error deviation of this term, $\sigma^e$ in Eq. (7.24), is to use a classical post-calibration residual examination in combination with a simple error disaggregation method.
If we proceed with classical model calibration the output is a set of residuals which constitute a combination of model structural, input and output errors. The difference between this residual and the associated measurement error of the observation represents the combined effect of model structural and input error. To illustrate this, consider Figure 7.7, which presents a scatter plot of the error deviation of the residuals of the calibrated HYMOD model (see Fig. 7.1) against flow level. This plot was created using 10 years of streamflow data for the Leaf River watershed.

The dashed black line denotes the estimated measurement error model (see Fig. 7.6), whereas the solid line represents the optimal fit of a spline function through the scattered data. According to our statements, the definition of model structural and input error is the area between the two different solid and dashed black lines. Another spline function was fitted through this difference function and subsequently implemented in SODA to relate the error deviation of the stochastic forcing to flow level.

**7.5.3. Use of SODA for Hydrologic Model Calibration: Model Precision and Accuracy**

The SODA algorithm was used with the particular settings discussed in the previous two sections to simultaneously estimate the HYMOD state variables and model parameters using the
SIMULTANEOUS PARAMETER AND STATE ESTIMATION

daily streamflow data of the Leaf River basin and the density criterion specified in Eq. (7.25). A search population of 250 points in combination with 5 parallel sequences and 100 ensemble members was selected. The procedure used about 1,500 parameter evaluations to converge to a reasonable estimate of the optimal model parameters, including their underlying posterior distribution. The results of this analysis are summarized in Table 7.2 and Figures 7.8, 7.9 and 7.10 and discussed below.

The usefulness of the implementation of SODA can be demonstrated in a number of different ways. One of the most straightforward is to compare the hydrograph prediction uncertainty ranges of the HYMOD model using SODA, with those obtained using a classical Bayesian SCEM-UA calibration, which ignores input uncertainty and lumps output and model error into a single white noise term. The size of the prediction uncertainty ranges is a measure of the precision of the model. Figure 7.8 presents the results of this analysis. The hydrograph prediction uncertainty ranges (light-gray region) of the HYMOD model corresponding to the optimal SODA identified parameter set and an ensemble size of 100 members is presented in Fig. 7.8a. The observed streamflows are indicated with solid circles, whereas the mean ensemble prediction is indicated with the dashed black line. Fig. 7.8b presents the prediction uncertainty ranges for the HYMOD simulated streamflows corresponding to the SCEM-UA derived posterior parameter estimates. From visual inspection of the two plots we see (as expected) that the input–output behavior of the HYMOD model when properly accounting for input, output and model structural errors, is more consistent with the observations. Hence, in the case of SODA, the prediction uncertainty ranges are reasonably small and bracket the observations. On the contrary, when applying a traditional Bayesian approach (Fig. 7.8a), without state adjustments, the model seems to be unable to match large portions of the hydrograph; this is indicated by sections where the light-gray region does not bracket the observed streamflow data. Although one might argue that this is a problem of overconditioning, and that better results can be obtained by assigning more relaxed cutoff thresholds to determine what can be considered an acceptable parameter set or not (as is done in GLUE), that in fact leads to considerably larger hydrograph uncertainty bounds than those presented for SODA, particularly during the storm events. Moreover, as argued in the first two sections of this Chapter, it is not realistic to attribute all uncertainty in the modeling exercise to uncertainty in the parameter estimates.
Figure 7.8. Results of SODA by application to hydrologic modeling using the HYMOD conceptual watershed model and historical streamflow data from the Leaf River Watershed in Mississippi; (a) Hydrograph prediction uncertainty ranges corresponding to the best SODA identified parameter set, and 100 ensemble members; The dashed black line denotes the mean ensemble prediction (b) 95% prediction uncertainty ranges of the HYMOD model forecasts associated with the SCEM-UA derived posterior parameter distribution. The dashed black line denotes the evolution of the most likely parameter set.
Another diagnostic measure, which can be used to contrast the results of SODA with those obtained using classical Bayesian model calibration, is to compare the autocorrelation functions of the SCEM-UA (see Fig. 7.1c) and SODA derived time series of forecast errors. Autocorrelation is a measure of the accuracy of the model predictions. Figure 7.9 presents the results of this analysis. To benchmark against a time series of white forecast errors, the dashed lines in these plots represent the theoretical 99% confidence intervals of a white error time series [Box and Jenkins, 1976].

![Figure 7.9. Autocorrelation functions of the time series of forecast errors using, (a) SCEM-UA algorithm, and (b) SODA framework. The dashed lines denote the theoretical upper and lower 99% significance intervals of a time series of white residuals.](image)

When performing a traditional SCEM-UA calibration, there is significant autocorrelation between the residuals at the first few lags, confirming our earlier findings reported in Fig. 7.1. Notice, however, that there is considerably less autocorrelation between the forecast residuals when using the SODA framework, suggesting that recursive state adjustments remove a large part of the bias in the model predictions.

This is further demonstrated in Table 7.2, which presents summary statistics of the one-day-ahead streamflow prediction performance of the SODA and SCEM-UA methods for the 3-year calibration (WY 1952 – 1955) and 5-year evaluation (WY 1955 – 1960) period.
CHAPTER 7

Table 7.2. Summary statistics (RMSE, CORR and BIAS) of the one-day-ahead streamflow forecasts using the SCEM-UA and SODA methods.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>RMSE</td>
<td>CORR</td>
</tr>
<tr>
<td>SCEM-UA</td>
<td>17.25</td>
<td>0.903</td>
</tr>
<tr>
<td>SODA</td>
<td>13.14</td>
<td>0.96</td>
</tr>
</tbody>
</table>

The results in this Table indicate that SODA consistently provides better values of the Root Mean Square Error (RMSE), correlation (CORR), and bias (BIAS) statistics than conventional batch calibration for both time periods. Although not further demonstrated here, the performance of SODA with a simple 5-parameter conceptual model compares favorably well with other modeling approaches, including the 13-parameter Sacramento Soil Moisture Accounting model of Bumazh et al. [1973], and the multi-parameter Self-Organizing Linear Output Map (SOLO), recently developed by Hsu et al. [2003]. Thus, explicitly accounting for uncertainty in the input, output and model structure, during model calibration will lead to both an improved assessment of predictive uncertainty, and a simultaneous improvement in the model forecasts.

7.5.4. Use of SODA for Hydrologic Model Calibration: Parameter Uncertainty

Figure 7.10 presents posterior marginal probability density distributions for each of the HYMOD model parameters estimated for the Leaf River watershed using the SCEM-UA (Fig 7.10. a-e) and SODA (Fig 7.10. f-j) algorithms. The results presented in this Figure highlight two interesting observations. First, notice that for some of the model parameters the SCEM-UA and SODA algorithms have assigned the highest probability to different locations in the parameter space. If the model structure would be correct and the input and output data were observed without error, no state adjustments would be needed during model calibration and the SCEM-UA and SODA method would identify the same location for the mode of the posterior probability distribution in the full parameter space. Not properly accounting for model structural and input errors during model calibration leads to corrupted parameter estimates, which are compensating for these errors.
Figure 7.10. Marginal posterior probability distributions of the HYMOD model parameters derived for the Leaf River Watershed using the (A-E) SCEM-UA algorithm (F-J) and SODA framework.
Based on these propositions, we therefore suggest that the model parameters identified with SODA are less corrupted by model structural and input errors and, therefore, better represent the underlying properties of the physical system. Our view has also been supported by additional numerical experiments with synthetically generated streamflow data, and a corrupted input (rainfall) data during the inverse identification [see e.g., Kavetski et al., 2003 for a similar approach], which demonstrated that model parameters identified with SODA seem to be less sensitive to errors in the input and converge in close proximity of the true parameter values used to generate the synthetic streamflow data. To increase the prospects of finding useful regionalization relationships, it would therefore seem most productive to use the SODA identified mode of the parameter space.

A second significant and interesting observation is that the HYMOD model parameters become less well defined when allowing for recursive state adjustments during model calibration (compare histograms of identical counterparts in Fig. 7.10). For most of the parameters, the dispersion around the mode of the distribution has significantly increased. In the systems theoretic sense, this can be explained as part of the ‘information’ from the streamflow measurements being used in SODA to recursively estimate the state variables.

7.5.5. Use of SODA for Hydrologic Model Calibration: Towards Model Structural Improvements

An interesting byproduct of SODA, which deserves further investigation, is the computed time series of state and output innovations. Any inability of the model to represent the input–output behavior of the underlying hydrologic system, will result in state updates in the model, when new measurements of the system are assimilated and processed. We are therefore left with the intuitively reasonable hypothesis that the computed time series of state updates should contain valuable information about model structural errors. In the last section of this Chapter, we verify the validity of this hypothesis by closer examination of the time series of output innovations.

To facilitate this process, we begin by partitioning the hydrograph into a driven, nondriven quick, and a nondriven slow component, in a method similar to Boyle et al. [2000]. For each of these portions of the hydrograph, we plot the SODA-computed mean ensemble output innovation against the mean ensemble streamflow prediction (prior to the update). The results of this analysis are presented in Figure 7.11. For the driven and nondriven quick part of the hydrograph, these plots do not reveal any simple relationship between the ensemble mean predicted flow level (x-axis) and associated output innovations (y-axis). Nonlinear time series
SIMULTANEOUS PARAMETER AND STATE ESTIMATION

analysis (not presented here) using artificial neural networks (ANN) indicates that some structure is, indeed, present in the time series of output innovations, but the process of how to translate these findings into simple mathematical equations which can be used to improve the conceptual model and as such our understanding of hydrologic processes, remains the topic of future work.

A more easily interpretable result is obtained for the nondriven slow part of the hydrograph (Fig. 7.11c). Here, a clear systematic linear relationship between flow level and associated output innovation is apparent. This relationship can easily be exploited to improve the performance of the HYMOD model during low flows. Indeed, when incorporating a linear relationship between flow level and output innovation into HYMOD, the predictive capabilities of the model during nondriven slow flow increases from an average error of 0.20 to 0.09 m$^3$/s. In contrast, the classical model calibration methodologies, which assign all the uncertainty in the input – output representation to uncertainty in the model parameters, provide little or no such guidance.

**7.6. Summary and conclusions**

The objective of model identification is to obtain a model where the input-state-output behavior is consistent with the measurements, and where the model predictions are accurate and precise. In practice, however, because of errors in the model structure, input (forcing), and output data, this has proven to be quite difficult, leading to considerable uncertainty in the model predictions. Classical model calibration strategies typically ignore input uncertainty, lump model structural and output errors into a single white noise term, and assign the uncertainty in the input-output representation of the model primarily to the parameters, therefore lacking the conceptual rigor required to properly account for all the important sources of uncertainty.
CHAPTER 7

In this Chapter we present a combined global optimization and data assimilation method, which improves the treatment of input, output, parameter, and model structural uncertainty in hydrologic model calibration. The method, entitled Simultaneous Optimization and Data Assimilation (SODA) merges the strengths of the parameter search efficiency and explorative capabilities of the Shuffled Complex Evolution Metropolis algorithm and the power and computational efficiency of the Ensemble Kalman Filter to simultaneously estimate parameters and state variables in hydrologic models.

The usefulness and applicability of SODA has been demonstrated for two preliminary case studies. The first case study considered the highly nonlinear three-parameter Lorenz model, and demonstrated that SODA is indeed successfully able to simultaneously estimate state variables and model parameters when confronted with highly nonlinear model dynamics. The second case study explored the usefulness of SODA by application to hydrologic modeling using the HYMOD conceptual watershed model and historical streamflow data from the Leaf River Watershed in Mississippi. The ability of SODA to deal with input, output, parameter and model structural errors, results in improved estimates of the parameter and model prediction uncertainty ranges. With this basis, a detailed investigation of the state and output innovation time series, can be used to investigate improvements to our model concepts and as such our understanding of the functioning of hydrologic systems.
CHAPTER 8

Epilogue

"Towards Improved Treatment of Parameter Uncertainty in Hydrologic Modeling"

To-date, uncertainty in hydrologic modeling has mostly been neglected, thereby elusively focusing on a single best parameter set that meets the needs and objectives of the modeling procedure. In this thesis, resources have been directed towards the development of stochastic and recursive single- and multi-objective parameter estimation algorithms that, combined with state-of-the-art data assimilation methods, can be used to more completely treat the various sources of uncertainty in hydrologic modeling, with a particular emphasis on parameter uncertainty. The need for parameter estimation and data assimilation tools continues to grow as advancements in remote sensing, communications, computing, and information technologies allow engineers and scientists to work with more complex computer models, spatially distributed data sets and real-time information. With the SCEM-UA, PIMLI, MOSCEM-UA, and SODA parameter and state estimation methods at our disposal we can now forcefully address a variety of research questions, which could not have been easily answered without these algorithms. In this epilogue, some of these most important research questions related to hydrologic modeling are highlighted, and a view on future research in model calibration is explicated.
CHAPTER 8

With the availability of the SCEM-UA, PIMLI, MOSCEM-UA, and SODA parameter and state estimation methods we can now forcefully address the following issues:

(1) Diagnosing systematic errors — Detailed investigations of the model prediction uncertainty ranges associated with the posterior (SCEM-UA) and Pareto (MOSCEM-UA) set of parameters, will reveal important information about systematic (auto-correlated) errors.

Theoretically, if the model structure would be correct, and the input and output data are observed without error, the model output prediction uncertainty ranges, associated with the uncertainty in the parameter estimates, will bracket the observations. Any systematic departure from this is not caused by parameter uncertainty, but due to the combined effect of input, output and model structural error (see for instance Figures 2.11 and 5.9). The insights that were developed with this analyses, led to the development of SODA in Chapter 7 of this thesis.

(2) Parameter stationarity — Detailed investigations of one-dimensional projections of the evolution of the High-Probability Density (HPD) region of the parameter space over a historical record of calibration data, provide a way for checking whether the hydrologic system has undergone changes. A calibrated model can only be reliably used for the simulation and prediction of hydrologic events outside the calibration period, if it can be reasonably assumed that the physical characteristics of the watershed and the hydrologic/climate conditions have remained similar. For instance, alterations in land-use will result in a different response of the watershed to precipitation forcing. Systematic trends in parameter variation can also be used to diagnose and quantify model structural errors, although this has yet to be satisfactorily demonstrated.

As the PIMLI and SCEM-UA algorithmic successfully infer the posterior distribution of the model parameters, these methods are suited to investigate whether model parameters are statistically stationary over a historical record of data or if they are correlated to varying characteristics of the underlying hydrologic system. As an illustrative example, consider Figure 8.1, which presents the evolution of the SCEM-UA derived marginal HPD regions for each of the parameters in the Sacramento Soil Moisture Accounting model of the National Weather Service of the USA (SAC-SMA) using a 36-year historical record of streamflow data for the Leaf River Watershed [taken from Vrugt et al., 2004]. The uncertainty bounds presented in Fig. 8.1 denote averages over a window of 6 years (1953 denotes calibration results over the WY 1953-1959 respectively and so forth).
Normalized parameter space

Figure 8.1. Evolution of the 95% confidence intervals for each of the SAC-SMA parameters over the 36-year historical record of the Leaf River watershed. The gray shaded area in each parameter plot represents the HPD region, whereas the marked squares and dotted line refer to the SCE-UA solution and most likely SCEM-UA solution respectively. Results denote averages over a window of 6 years. Definition of the SAC-SMA parameters appears in Table 6.2.
To allow comparison between uncertainties of different parameters, the HPD region was scaled according to the prior uncertainty bounds of the parameters, as presented in Table 6.2. to yield normalized ranges between 0 and 1. The gray-shaded area in each parameter plot represents the HPD region, whereas the marked squares and dotted line refer to the SCE-UA solution, and most likely SCEM-UA solution within the HPD region, respectively. Definition and explanation of the SAC-SMA model parameters appears in Table 6.2. While some of the SAC-SMA parameters show little variation over the 36-year historical data record, the HPD region for other SAC-SMA parameters traverses through the feasible parameter space. Especially, there is considerable variation and uncertainty associated with the parameters lower zone parameters LZSK, and LZPK, which primarily determine the shape of the hydrograph during the recession periods, and the percolation parameters REXP and PFREE. The apparent systematic variation of the parameters ADIMP, LZSK and LZPK with time, might suggest that the watershed has undergone hydrologic changes. When some of the parameters in the SAC-SMA model are plotted against the mean areal rainfall over the calibration periods, as done in Figure 8.2, a relationship becomes apparent.

**Figure 8.2.** Two-dimensional plot of the most likely parameter value, indicated with a squared symbol, versus the mean areal rainfall for the SAC-SMA parameters ADIMP (a), LZTWM (b), LZSK (c), and LZPK (d). The bars around the most likely parameter value denote the size of the HPD region.
EPILOGUE

To be able to match the observed hydrograph with increasing wetness of the years, the additional impervious fraction is decreased (ADIMP), while the maximum capacity of the lower zone tension water storage (LZTWM) and depletion rate from the lower zone need is to be increased. Seemingly, parameters calibrated for relatively dry years, result in sub-optimal forecasts for the wettest years on record and vice versa. This nonstationarity with increasing wetness of the years for some of the SAC-SMA parameters, point towards aspects of the model structure that needs to be further refined.

(3) Parameter identifiability – Detailed investigations of the SCEM-UA derived posterior mean, standard deviation, coefficient of variation and Pearson correlation coefficients between the samples in the HPD region of the parameter space, (i) facilitates the selection of an adequate model structure, (ii) guides in the development of pedotransfer functions, (iii) helps to assess how much model complexity is warranted by the available calibration data, and (iv) guides the development of optimal experimental design strategies.

(3.1) With respect to issue (i), consider Figure 8.3, which presents two-dimensional scatterplots of the SCEM-UA derived posterior parameter samples for identical counterparts in the soil water retention models of Brooks and Corey [BC, 1964], van Genuchten [VG, 1980], and Kosugi [KS, 1996] [taken from Vrugt et al., 2003a].

Figure 8.3. Scatter plot of 4000 combinations of $\theta_r-a$, $\theta_r-l$, $\theta_r-l$, $\theta_r-x$ (a), $\theta_r-l$, $h_0.5-x$ (h) and $\theta_r-x$ (i) parameters sampled for the clayey soil using the SCEM-UA algorithm for the BC-model (a-c), VG-model (d-i) and KS-model (g-i), respectively.
Each of the retention models contains four parameters, whose values need to be estimated by fitting to observed retention data. While the fit of each of the different parametric models to the observed retention data was generally excellent (not demonstrated here), Fig. 8.3 depicts that the correlation coefficients between the parameters are generally higher in the KS-model (Figs. 8.3g-i) as compared to the BC-model (Figs. 8.3a-c) and VG-model (Figs. 8.3d-f). More specifically, whereas in the BC and VG-model, only the parameters $\theta_r \lambda$ (Fig. 8.3c), and $\alpha \cdot n$ (Fig. 8.3f) are strongly correlated for the clayey soil, in the case of the KS-model three pairs of parameters exhibit considerable correlation, $\theta_r \cdot h_{03}$ (Fig. 8.3g), $h_{05} \cdot \sigma$ (Fig. 8.3h), and $\theta_r \cdot \sigma$ (Fig. 8.3i).

The strong hyperbolic-shaped correlation structure being associated with the parameters $\theta_r$, $h_{03}$ and $\sigma$ in the KS-model not only decreases the identifiability of the parameters, but also enhances non-uniqueness problems of the final optimized parameters. With recourse to the quality of the fit only the competing parametric models of VG and KS appear to fit the experimental data equally well. However, when examining other aspects, such as the multivariate correlation structure as inferred from the joint distributions of the SCEM-UA generated samples (Fig. 8.3), it becomes clear that there are drawbacks associated with the correlation structure of the parameters in the KS-model for fine textured soils. It is clear that when possessing comparable predictive capabilities, the non-linear model that closest approaches linear behavior is in favor. Besides identifiability problems of the parameters, for linear models less iterations are necessary to achieve convergence in parameter estimation, and traditional computational undemanding first-order statistical inferences will be more valid.

(3.2) With respect to issue (ii), an identifiability analysis of the parameters constitutes important information for studies that aim to find pedotransfer functions, relating the calibrated parameters to catchment characteristics. Uniqueness of the parameters is a prerequisite to be able to successfully find pedotransfer functions. To successfully find pedotransfer functions, it seems most productive to concentrate on those model parameters that are well defined after direct estimation (calibration).

(3.3) To demonstrate the usefulness of studying parameter identifiability to assess the information content of calibration data (issue iii), consider Table 8.1, which presents the SCEM-UA derived posterior mean, standard deviation, coefficient of variation (CV), and correlation structure induced between the parameters of a four-parameter single-layer canopy rainfall interception model.
Table 8.1. SCEM-UA derived posterior mean, coefficient of variation (CV [%]), and Pearson correlation coefficients between the parameters in the single-layer interception model obtained when using half-hourly measurements of throughfall or canopy storage at DOY 211. The symbols $a$, $b$, $c$, and $d$ refer to the interception efficiency, drainage, storage capacity, and evaporation efficiency parameter, respectively.

<table>
<thead>
<tr>
<th>Par.</th>
<th>Unit</th>
<th>Mean.</th>
<th>CV [%]</th>
<th>$a$</th>
<th>$b$</th>
<th>$c$</th>
<th>$d$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1.00</td>
<td>-0.12</td>
<td>-0.00</td>
<td>-0.06</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1.00</td>
<td>0.13</td>
<td>0.01</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1.00</td>
<td>0.86</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1.00</td>
</tr>
</tbody>
</table>

These results were derived when assimilating and processing half-hourly throughfall and water storage measurements at the Speulderbos forest in the Netherlands [taken from Vrugt et al., 2003b]. The high coefficients of variations (CV) and standard deviations of the parameters in Table 8.1 demonstrate that measured throughfall dynamics contain only very limited information for the calibration of a canopy interception model. Although there does not exist an exact quantitative threshold to help judge whether a parameter is identifiable or not, the rationales on the identifiability of the interception model parameters that we adopt in this section are based on the extent of the HPD region in the physical plausible or prior parameter space. For instance, it is clear that the storage capacity ($c$), the evaporation efficiency ($d$), and the drainage parameter ($b$) are poorly defined by calibration to measured throughfall dynamics, as for each of these parameters there does not exist a well-defined region, in the sense of a compact region, interior to the physical plausible or prior parameter space. On the contrary, measured water storage dynamics contain sufficient information to be able to identify most of the model parameters with a high degree of confidence. Hence the CV values and standard deviation of the parameters, as reported in Table 8.1, illustrate that for most of the parameters the HPD region occupies only a very small portion of the prior parameter space. Especially the storage capacity and the evaporation efficiency parameter are very well determined by calibration to measured water storage dynamics. Unfortunately, like measured throughfall dynamics, canopy water storage
measurements contain insufficient information to identify the drainage parameter with a satisfying accuracy.

(3.4) With respect to issue (iv), recursive parameter estimation methods, such as PIMLI facilitate the identification of the most informative measurements for model calibration, and as such help in the development of optimal experimental design strategies. Although this has been extensively discussed and demonstrated in Chapter 5, for completeness another illustrative example in hydrometeorology is discussed here [taken from Vrugt et al., 2003b]. Figure 8.4, presents the evolution of the HPD region of the posterior probability density, in the form of one-dimensional projections for each of the parameters in the single-layer canopy interception model, derived when stepwise assimilating and processing the throughfall (Figs. 8.4b–e) or canopy water storage (Figs. 8.4g–j) measurements of Day of Year (DOY) 211 with the PIMLI algorithm. The dark-shaded line marks the evolution of the most likely parameter set at each time step, whereas the asterisks denote the most likely parameter values derived using the SCE-UA global optimization algorithm.

Starting at $t = 211.4$, the size of the HPD region of the parameters, reflects the initial or prior uncertainty of the parameters before any data are collected and processed. Immediately after the first rain event at $t = 211.55$, the uncertainty associated with the interception efficiency parameter $a$ decreases, and remains rather constant thereafter, demonstrating that rain events during the wetting stage of the forest canopy contain the most information for the identification of the $a$-parameter. Although the $a$-parameter is reasonable well determined by calibration to measured throughfall dynamics (see Fig. 8.4b), the additional throughfall measurements between the first rain event and the drying cycle starting at the beginning of DOY 212 contain very limited information for the model parameters. Hence the one-dimensional projections of the evolution of the HPD region of the posterior probability density for $b$, $c$, and $d$ (see Figs. 8.4c–e) suggest that for these parameters there does not exist a well defined region in the sense of a compact region interior to the prior parameter space when calibrating on measured throughfall dynamics.
Figure 8.4. Evolution of the HPD region of the posterior probability density in the form of one-dimensional projections of the parameters (light-gray region) using measured throughfall dynamics (b-e), or canopy water storage observations (g-j). The dotted line denotes the evolution of the most likely parameter set, whereas the asterisks indicate the “best” parameter values obtained using a traditional batch calibration with the SCE-UA global optimization algorithm. The solid circles in the figures a and f denote measured values of throughfall and canopy water storage respectively, whereas the dotted line in these figures denote the model predicted values corresponding to the parameter set with the highest posterior probability.
CHAPTER 8

On the contrary, the evolution of the Bayesian confidence intervals of the parameters depicted in Figs. 8.4g–j illustrate that measured canopy water storage dynamics contain sufficient information to uniquely identify at least three of the interception model parameters \( a, c, \) and \( d \). Moreover, these parameters are identifiable at different stages during the wetting and drying cycles, thereby facilitating the identification of a unique set of parameters. Unfortunately, no information is found for the drainage parameter. Fig. 8.4 demonstrates very well that adding more data does not simply solve the problem of parameter identifiability. Only specific data periods with high information content can reduce the uncertainty associated with the interception model parameters.

Although not explicitly demonstrated here, the use of longer observational time series of throughfall dynamics for calibration purposes, leads to similar findings. While it might seem speculative to generalize the conclusions regarding the identifiability of the interception model parameters using measured throughfall dynamics to other climates, species, or biomasses situations, additional investigations with numerically generated throughfall “measurements” for other situations than what was presented in this Epilogue yielded similar results. We subscribe ourselves, therefore, to the view that model parameters of drainage and evaporation functions, which are obtained by calibration against measured throughfall dynamics must be interpreted with care as these parameters are subject to considerable uncertainty.

(4) Quantifying uncertainty in regionalization studies – A topic, which is currently receiving considerable attention in the hydrologic community, is the simulation of the rainfall-runoff behavior of ungauged watersheds. The International Association of Hydrological Sciences (IAHS) has recently launched a 10-year initiative, called the IAHS Decade for Prediction in Ungauged Basins (PUB), to address this problem.

When we attempt to generate hydrologic predictions for ungauged watersheds, model parameters cannot longer be estimated by calibration to streamflow observations, and need to be estimated from other sources of information, such as neighboring catchments, expert judgment or tabulated literature values. One approach to derive the values of model parameters for ungauged watersheds, which is particularly popular in hydrology, is to use established statistical relationships between model parameters and watershed characteristics for gauged basins. This process of transferring parameters from gauged to ungauged catchments is generally referred to as generalization in surface hydrology, although soil hydrologists prefer the term pedotransfer functions. While significant progress has been made in the development and application of pedotransfer functions to simulate the hydrologic behavior of ungauged watersheds, most of the
reported studies in the literature typically ignore parameter uncertainty and uniqueness in the regionalization process. Usually predictions in ungauged watersheds are given as point estimates. The availability of the SODA framework now enables the quantification of parameter uncertainty for gauged watersheds, while taking into account, input, output and model structural errors. This parameter uncertainty can be explicitly used in the derivation of transfer functions, thereby propagating modeling uncertainty from gauged to ungauged watersheds. Certainly, this will result in a stochastic estimate of the hydrologic response of the ungauged watershed, which is more useful for decision-making since it considers the confidence in the model predictions.

In principle, the summarized points under 2-4 can also be extended to multi-objective parameter estimation problems. The domain of interest is then Pareto uncertainty instead of probabilistic parameter uncertainty. This also provides new and useful ways to define the information content of data, and to evaluate different model structures and their performance.

(5) Improving hydrologic models – As mentioned in the introduction section of this thesis, the objectives of the PhD research should not be understood as that of simply developing methods for parameter estimation alone. Instead, the envisaged higher goal of the PhD research is to use the SCEM-UA, MOSCEM-UA, PIMLI and SODA algorithms to contribute to an improved understanding of the various hydrologic processes operating at or near the earth’s surface.

To be able to better understand the limitations of our models and improve our understanding and theory of hydrologic processes, we need to develop strategies, which can distinguish between input, output, parameter, and model structural uncertainty. The SODA framework presented in Chapter 7 of this thesis has been designed to facilitate this task. Because the inversely identified parameters with SODA are less corrupted by modeling errors, detailed interpretation of the time series of output and state innovations will reveal useful information about model structural errors (see Fig. 7.11). These findings can be used to reformulate the model and as such to improve our understanding of hydrologic processes.

The SODA method is an attempt to more completely treat these various sources of uncertainty. However, still progress in this matter is possible. I believe that future research on model calibration would be most productive if focused on the following issues, (1) the derivation and implementation of realistic input error models to further distinguish between input and model structural error, (2) the development and implementation of higher order filters which generate improved estimates of output data errors, (3) the extension to multicriteria problems;
SODA does not exclude the use of different parameter sets to match different portions of hydrologic behavior, (4) the use of recursive parameter estimation methods to further minimize the short term model bias using the state augmentation technique, (5) the development of filtering methods which further relax the Gaussian assumptions of the error distributions in classical KF implementations. With respect to this last issue, Gaussian mixture models are able to represent arbitrarily complex probability density functions. This fact makes them an excellent choice for representing complex likelihood functions for recursive Bayesian and multi-objective parameter estimation or filtering techniques, and (6) the extension of SODA to multiple competing models running in parallel (improved Bayesian Model Averaging).

The work presented in this thesis has primarily focused on the development and application of optimization techniques to quantify parameter uncertainty in hydrologic models. However, as emphasized in Chapter 7, it does not seem reasonable to attribute all the uncertainty in the modeling procedure to uncertainty in the parameter estimates. In the same Chapter we therefore outlined the more general SODA framework to simultaneously deal with input, output, parameter, and model structural uncertainty in hydrologic modeling. It is to be expected that further developments within this framework, in combination with the ever increasing pace of computational power, and the availability of spatially distributed data, will advance our understanding of the functioning of environmental systems.
SUMMARY (ENGLISH)

The research presented in this thesis has focused on the development of stochastic and recursive single- and multi-objective parameter estimation algorithms that combined with state-of-the-art data assimilation methods can be used to more completely treat input, output, parameter, and model structural uncertainty in environmental model calibration. The usefulness and applicability of these algorithms are demonstrated by application to vadose zone and surface hydrology, to improve understanding and predictions of unsaturated soil water flow, water uptake by plant roots, river water discharge, and heat and moisture fluxes at the land surface. What follows here is a short summary of each of the chapters in this thesis.

To state the case and illustrate a typical parameter estimation problem in hydrology, Chapter 2 discussed the inverse estimation of soil physical and root water uptake parameters in a physically-based three-dimensional unsaturated soil water flow model using measured spatial distributions of soil water content around a sprinkler irrigated almond tree. A multi-dimensional approach in root water uptake is needed if uptake is varying in space, thereby allowing a more accurate quantification of spatial variability of the soil water regime, including the water and solute flux densities below the root zone. After calibration of the selected root water uptake model and soil hydraulic parameters, the agreement between simulated and measured spatially distributed water contents during the 16-days calibration period was generally good. To evaluate the benefits of using a sophisticated three-dimensional vadose zone model, the results of this analysis were compared with numerical models describing soil water flow and root water uptake in one and two dimensions. For each of the considered models, independently measured soil water retention data agreed favorably with the optimized retention curves. Moreover, optimized root water uptake distributions between one-, two-, and three-dimensional flow models with corresponding root water uptake models were almost identical. However, major differences occurred for the spatial variation in root water uptake and drainage rates between one-dimensional and multi-dimensional models. This justifies the need for multi-dimensional root water uptake and flow models, especially when the fate and transport of chemicals below the rooting zone for single trees is of concern. Despite the good fit of each of the numerical models to the observed water content data, this chapter also highlighted some of the limitations of current parameter estimation methods: they are computationally very demanding when solving for high-dimensional parameter problems, and they do not provide any information about the probabilistic uncertainty associated with the final parameter estimates. This latter point is of
particular relevance when assessing the identifiability of the model parameters or to diagnose how many parameters are supported by the calibration data.

To overcome the limitations of commonly employed parameter estimation algorithms, Chapter 3 presented a general-purpose code, entitled the Shuffled Complex Evolution Metropolis (SCEM-UA) global optimization algorithm, which is especially designed to provide an effective and efficient estimate of the traditional "best" parameter set and its underlying posterior distribution, within a single optimization run. The algorithm is a Markov Chain Monte Carlo sampler generating a sequence of parameter sets that converge to the stationary posterior distribution, for a large number of model simulations. The SCEM-UA algorithm is related to the successful Shuffled Complex Evolution (SCE-UA) global optimization algorithm, but uses the Metropolis-Hastings (MH) strategy instead of the Downhill Simplex method for population evolution. The stochastic nature of the MH annealing scheme avoids the tendency of the SCE-UA algorithm to collapse into a single region of attraction (local minima), while the information exchange (shuffling) between parallel sequences allows the search to be biased in favor of better regions of the solution space. These adaptive capabilities of the SCEM-UA algorithm significantly reduce the number of model simulations needed to infer the posterior distribution of the parameters when compared with traditional MH samplers. In the same Chapter the identifiability of the parameters in a rather parsimonious conceptual rainfall-runoff model, consisting of a relatively simple rainfall excess model, connected with two series of linear reservoirs was explored. Those results indicated that the entire model structure was well identifiable by calibration to runoff data, thereby supporting statements made in the literature that simple rainfall-runoff models with four to five parameters provide an adequate fit to the streamflow data and that the addition of more model structure and its associated parameters leads to no significant improvement in fit yet introduces poorly identified parameters.

While the capabilities and limitations of the inverse approach for the identification of soil hydraulic properties from laboratory soil cores or small field plots (e.g. Chapter 2) may be considered reasonably well understood, still little is known about the suitability of the inverse approach for the identification of vadose zone properties at larger spatial scales. Fortunately, in the past few years, computational capabilities have evolved to a point, where it is possible to use multi-dimensional physically based watershed models to study spatial and temporal patterns of water flow in the vadose zone. However so far, these models based on complex multi-dimensional governing equations have received only limited attention, in particular because of their computational, distributed input and parameter estimation requirements. In Chapter 4 the usefulness and applicability of the inverse approach for the estimation of spatially-distributed
vadose zone properties was explored, using the solution of a physically-based three-dimensional distributed model combined with spatially distributed measured tile drainage data from the 9700 ha Broadview Water District in the San Joaquin Valley of California. The inverse problem was posed within a single criterion Bayesian framework and solved by means of the computerized SCEM-UA global optimization algorithm, as presented in Chapter 3. To study the benefits of using a spatially distributed three-dimensional vadose zone model, the results of the three-dimensional model were compared with those obtained using a simple conceptual bucket model and a spatial-averaged one-dimensional unsaturated water flow model. Results presented in this Chapter demonstrate that measured spatially-distributed patterns of drainage data contains only limited information for the identification for the vadose zone model parameters, and are particularly inadequate to identify soil hydraulic properties at the larger spatial supports. The dominant hydrologic processes, controlling tile-drain discharge at the watershed scale are preferential flow and the properties of the drain system. Furthermore, results indicate that there are advantages of using physically-based hydrologic models to study spatial and temporal patterns of water flow at larger spatial scales as these mechanistic models not only generate consistent forecasts of spatially-distributed drainage during both the calibration and validation periods, but simultaneously also possess unbiased predictive capabilities of groundwater table depths not included during the calibration.

The primary focus of the Chapters 2-4 has been on the "batch" calibration approach, which aggregates error residuals over a large range of hydrologic behavior. These "batch" processing methods do not explicitly recognize differences in model sensitivity of the model parameters to the various measurements, and as such cannot be used to identify which sets of measurements are most informative for specific model parameters. In Chapter 5 a formal methodology is explicated, entitled the Parameter Identification Method based on the Localization of Information (PIMLI), which increases information retrieval from experimental data, by recursively assimilating and processing the information content in the calibration data. The PIMLI algorithm is a hybrid approach that merges the strengths of the Generalized Sensitivity Analysis (GSA) method, Bayesian recursive estimation, and the sampling efficiency of the Metropolis-Hastings algorithm to select those sets of measurements that contain the most information for the identification of the various model parameters. The usefulness and applicability of the PIMLI algorithm was demonstrated by application to the estimation of soil hydraulic parameters using soil water retention data and a more complex multi-step transient outflow experiment, and the calibration of a conceptual rainfall-runoff model. Each of these studies illustrated that the PIMLI algorithm was well suited to identify which parameters control
SUMMARY (ENGLISH)

what part of model behavior, while providing insights into the importance of different kinds of calibration data in model parameter estimation. Moreover, the method revealed that only a very limited number of streamflow measurements were needed for a reliable calibration of a conceptual rainfall-runoff model.

The research presented in Chapters 2-5 focused on the use of a single-objective measure to extract all the information contained in the calibration data. Practical experience with the calibration of hydrologic models, however, suggests that single-objective functions are often insufficient to properly account for all of the characteristics of the observed data deemed to be important. One strategy to circumvent this problem is to define several optimization criteria (objective functions) that measure complementary aspects of the system behavior and to use a multicriteria optimization method to identify the set of nondominated, efficient or Pareto optimal solutions. While in the past several algorithms have been developed which can solve multi-objective calibration problems, these methods have two characteristic failing problems. In the first place, they tend to cluster the Pareto solutions in the compromise region among the objectives, thereby leaving the ends of the Pareto frontier unexploited. The second, perhaps more important, failure is the inability of these methods to converge to solutions within the “true” Pareto set for hydrologic models involving a large number of parameters and highly correlated performance criteria. In Chapter 6, therefore, the Multi-objective Shuffled Complex Evolution Metropolis (MOSCEM) global optimization algorithm is presented, which is capable of effectively and efficiently solving the multi-objective optimization problem for hydrologic models. The MOSCEM-UA algorithm combines the strengths of the complex shuffling employed in the SCE-UA algorithm, the probabilistic covariance-annealing search procedure of the SCEM-UA algorithm (see Chapter 3) and an improved fitness assignment concept to construct a uniform estimate of the Pareto solution set, thereby containing the single criterion end points. This Pareto set of solutions represent tradeoffs among the different incommensurable and often conflicting objectives, having the property that moving from one solution to another results in the improvement of one objective while causing deterioration in one or another. The MOSCEM-UA algorithm is the multi-objective relative of the SCEM-UA algorithm, but uses an innovative concept of Pareto dominance rather than direct single-objective function evaluations, to evolve the initial population of points towards a set of solutions stemming from a stable distribution (Pareto set). Application of the MOSCEM-UA algorithm to the Sacramento Soil Moisture Accounting Model of the US National Weather Service has demonstrated that there is considerable uncertainty associated with the percolation and recession processes in the model, which play a major role in determining the shape of the
hydrograph during periods without rainfall. A multi-criteria calibration of the bio-sphere atmosphere transfer scheme (BATS) land-surface model using measured heat and moisture fluxes from the Oklahoma ARM-CART site revealed that there are two disconnected regions in the parameter space which generate quite similar model behavior, indicating a interesting model structural issue that deserves further investigation.

With the availability of the stochastic and recursive SCEM-UA, MOSCEM-UA and PIMLI optimization algorithms, developed in Chapters 3, 5 and 6 we are now able to meaningfully and efficiently estimate single and multi-objective parameter uncertainty. However, uncertainties in the modeling procedure not only stem from uncertainties in the parameter estimates, but also from measurement errors associated with the system input (forcing) and output, and from model structural errors arising from the aggregation of spatially distributed real-world processes into a mathematical model. Not properly accounting for these errors during model calibration, results in model simulations and their associated prediction uncertainty bounds, which do not consistently represent and bracket the measured system behavior. This is usually evidenced by residuals, which exhibit considerable variations in bias (non-stationarity), variance (heteroscedasticity), and correlation structures under different hydrologic conditions. To more completely treat input, output, parameter and model structural uncertainty in hydrologic model calibration, Chapter 7 presented a Simultaneous parameter Optimization and Data Assimilation method entitled SODA, which combines the strengths of the parameter search efficiency and explorative capabilities of the SCEM-UA algorithm (see Chapter 3), and the power and computational efficiency of the Ensemble Kalman Filter to simultaneously estimate state variables and model parameters. Additionally, in this Chapter a nonparametric variance estimator is introduced, which is especially designed to estimate the measurement error of output data. The usefulness and applicability of SODA was demonstrated for two preliminary case studies. The first case study considered the highly nonlinear three-parameter Lorenz model, and demonstrated that SODA is indeed successfully able to simultaneously estimate state variables and model parameters when confronted with highly nonlinear model dynamics. The second case study explored the usefulness of SODA by application to hydrologic modeling using a simple conceptual watershed model and historical streamflow data from the Leaf River Watershed in Mississippi. The ability of SODA to properly deal with input, output, parameter and model structural errors, results in honest parameter and model prediction uncertainty ranges. Furthermore, detailed investigation of the computed state and output innovations as function of time, generates useful inspiration to improve our model concepts and as such our understanding of the functioning of hydrologic systems.
Finally, in Chapter 8 an epilogue is presented in which the most important research questions that can now be forcefully addressed with the availability of the SCEM-UA, PIMLI, MOSCEM-UA, and SODA parameter and state estimation are highlighted. Additionally, in this Chapter a short view on future research in model calibration is explicated.
Het onderzoek in dit proefschrift heeft zich gericht op het ontwikkelen van recursieve parameter-optimalisatiemethoden. Deze kunnen, gecombineerd met geavanceerde data-assimilatiemethoden, worden gebruikt om beter met input-, output- en modelfouten om te gaan tijdens de kalibratie van hydrologische modellen. Het nut van de ontwikkelde wiskundige algoritmen is aangetoond door ze toe te passen in bodem- en oppervlaktehydrologie. Dit is succesvol gedaan voor het verbeteren van voorspellingen van waterstoming door onverzadigde bodems, van wateropname door planten, van afvoer van rivieren en van de uitwisseling van warmte en water tussen het aardoppervlak en de atmosfeer. Hier volgt een korte samenvatting van de verschillende Hoofdstukken in dit proefschrift.

Hoofdstuk 2 behandelt een typisch parameter-optimalisatieprobleem in de hydrologie: het schatten van bodemfysische en wortelopname-parameters in een ruimtelijk, gedistribueerd, onverzadigd stromingsmodel op basis van een dataset van ruimtelijk verdeelde metingen van het watergehalte rond een geïrrigeerde amandelboom. Een multi-dimensionale benadering van wateropname door planten is noodzakelijk indien de wortelopname sterk varieert in ruimte en tijd, en nauwkeurige ruimtelijke schattingen van de waterbalans en stofstromingen onder de wortelzone noodzakelijk zijn. Na kalibratie van de wortel en bodemfysische parameters bleek de overeenkomst tussen de gemeten en gesimuleerde, ruimtelijke watergehaltepatronen over het algemeen zeer goed te zijn. Om verder de voordelen van een volledig driedimensionaal model te evalueren, zijn de simulatieresultaten van dit model vergeleken met de resultaten van numerieke modellen die waterstoming en worteloename in de onverzadigde zone in één en twee dimensies beschrijven. Voor elk van de modellen kwamen de geïdentificeerde wortel- en fysische eigenschappen goed overeen. Grote verschillen traden echter op in de voorspellingen van de ruimtelijke verdeling van waterfluxen onder de wortelzone. Deze laatste verschillen rechtvaardigen het gebruik van een volledig fysisch, gedistribueerd, driedimensionaal, onverzadigd stromingsmodel. Ondanks het feit dat de verschillende modellen de gemeten watergehaltepatronen goed kunnen beschrijven, is in dit Hoofdstuk ook duidelijk geworden dat de conventionele parameter-schattingstechnieken niet altijd eenvoudig toepasbaar zijn. Deze methoden vragen nogal wat computationele rekentijd, vooral wanneer ze een groot aantal parameters tegelijkertijd moeten schatten. Additioneel verschaffen ze veelal geen informatie over de uiteindelijke nauwkeurigheid van de parameterschating. Dit laatste is belangrijk indien uitspraken dienen te worden gedaan over hoe goed de modelparameters identificeerbaar zijn, en wat de optimale modelcomplexiteit is gegeven de beschikbare data.
Om de genoemde tekortkomingen van conventionele en veel gebruikte parameterschattingstechnieken op te heffen, wordt in Hoofdstuk 3 een nieuwe parameterzoektechniek gepresenteerd, genaamd het Shuffled Complex Evolution Metropolis (SCEM-UA) globale optimalisatiealgorithm. Deze methode is speciaal ontwikkeld om een snelle en betrouwbare schatting te krijgen van de beste parameterset, en zijn onderliggende stochastische verdeling. Het SCEM-UA-algoritme is een Monte Carlo-zoektechniek, die een sequentie van parametersets genereert die convergeert naar een limiterende verdeling met een toenemend aantal modellsimulaties. Het SCEM-UA-algoritme is nauw gerelateerd aan het Shuffled Complex Evolution (SCE-UA)-algoritme, maar gebruikt in plaats van een deterministische Simplexzoekmethode, de Metropolis-Hastings (MH)-strategie voor populatie-evolutie. De stochastische eigenschappen van het MH-algoritme voorkomen de tendens van het SCE-UA algoritme om te convergeren naar een enkele optimale oplossing, terwijl de informatie-uitwisseling tussen sequenties (shuffling) zorgdraagt voor een optimale exploratie van de parameterruimte. De adaptieve eigenschappen van het SCEM-UA-algoritme resulteren in een significant minder aantal modellsimulaties dan het standaard MH-algoritme om de stochastische verdeling van parameters te schatten. Om de toepasbaarheid van het SCEM-UA algoritme in de hydrologie te illustreren, wordt in dit Hoofdstuk ook de identificeerbaarheid van parameters in een simpel vijfparameter-conceptueel rivierafvoermodel getoetst. De resultaten tonen aan dat alle parameters in dit relatief simpele model goed identificeerbaar zijn door kalibratie met dagelijkse afvoermetingen. Deze resultaten ondersteunen eerder gedane uitspraken in de literatuur dat afvoermodellen met vier tot vijf parameters een goede voorspelling van de gemeten afvoerdata geven. Ook geven de resultaten aan dat de toevoeging van meer modelcomplexiteit de voorspelkracht van het model nauwelijks verbetert, maar over het algemeen wel leidt tot een slechtere identificeerbaarheid van de modelparameters.

De mogelijkheden en tekortkomingen van inverse modelleren voor het verkrijgen van bodemfysische eigenschappen van laboratoriumbodemmonsters of kleine veldopstellingen (zie Hoofdstuk 2) zijn vrij goed bekend. Er is echter nog weinig ervaring met de toepassing van deze methode voor het schatten van bodemhydrologische eigenschappen op de schaal van een stroomgebied. Gelukkig is in de afgelopen jaren de rekenkracht van computers zo ver door ontwikkeld, dat het mogelijk wordt om multi-dimensionale, fysisch-gebaseerde afvoermodellen te gebruiken om ruimtelijke en temporele patronen van waterstroming door bodems te bestuderen. Mede door beperkte rekenkracht en de behoefte aan ruimtelijk variërende input en parameters, hebben dergelijke modellen echter nog weinig aandacht gekregen. In Hoofdstuk 4 wordt het nut en de toepasbaarheid van inverse modelleren voor het schatten van ruimtelijk verdeelde,
bodemhydrologische parameters getoetst, met behulp van een fysisch gebaseerd, ruimtelijk gedistribueerd, driedimensionaal stromingsmodel en ruimtelijke afvoermetingen van drainagebuizen in het Broadview Water District (BWD) in de San Joaquin Vallei in Californië. Het inverseprobleem was opgesteld in een Bayesiaanse context en opgelost met behulp van het SCEM-UA-algoritme, zoals gepresenteerd in Hoofdstuk 3 van dit proefschrift. Om de voordelen van een volledig gedistribueerde aanpak te toetsen, zijn de simulatieresultaten vergeleken met een eendimensionaal, fysisch-gebaseerd stromingsmodel, en een simpel conceptueel model van het BWD. De resultaten in dit Hoofdstuk tonen aan dat ruimtelijk verdeelde metingen van waterafvoer door drainagebuizen slechts weinig informatie bevatten over de bodemhydrologische eigenschappen van het afvoergebied. In het bijzonder zijn deze metingen erg ongeschikt om bodemfysieke eigenschappen vast te stellen. Preferentiële stroming en de doorlatendheid en diepte van het drainagesysteem zijn de dominante hydrologische eigenschappen die de afvoer van het BWD bepalen. Verder toont een vergelijking van de drie afzonderlijke modellen aan, dat het gebruik van fysisch-gebaseerde, driedimensionale stromingsmodellen voordelen heeft. Deze modellen genereren namelijk niet alleen consistentere voorspellingen van gemeten afvoerpatronen gedurende de kalibratie- en evaluatieperiode, maar leveren tegelijkertijd ook zuivere schattingen op van grondwaterstanden die niet voor kalibratiedoeleinden gebruikt zijn.

De Hoofdstukken 2, 3 en 4 in dit proefschrift richten zich vooral op batchkalibratiemethoden, waarin de foutenresiduen tussen model en meting geaggregeerd worden in een enkele doelfunctie. Omdat verschillen in modelgevoeligheid voor de verschillende parameters in een dergelijke benadering niet expliciet worden meegenomen, zijn deze methoden niet geschikt om te achterhalen welke metingen de meeste informatie bevatten voor de verschillende modelparaters. In Hoofdstuk 5 wordt een formele methodologie geëxpliceerd, genaamd de Parameter Identification Method based on the Localization of Information (PIMLI), die meer informatie uit de metingen extrahert door de data recursief te assimileren en te verwerken. Het PIMLI-algoritme is een hybridische methode, die de kracht van de Generalized Sensitivity Analysis (GSA), met Bayesiaanse recursieve schattingen en de zoekefficiëntie van het MH-algoritme combineert om die metingen te lokaliseren die de meeste informatie bevatten voor de modelparaters. Het nut en de toepasbaarheid van het PIMLI-algoritme is aangetoond door toepassing in het schatten van bodemfysieke eigenschappen op basis van waterretentiedata en laboratoriumuitstromingsexperimenten en de kalibratie van een conceptueel afvoermodel. Elk van deze studies toonde aan dat het PIMLI-algoritme een geschikte methode is om te achterhalen welke parameters welk outputgedrag in het model veroorzaken, terwijl tevens
in zichten worden verkregen over hoe belangrijk verschillende datatypen zijn voor het schatten van de modellparameters. Bovendien werd het duidelijk dat slechts een beperkt aantal metingen nodig zijn voor een goede kalibratie van een conceptueel rivierafvoermodel.

discontinu is. Dit duidt op een interessant, modelstructureel probleem dat meer aandacht verdient in toekomstig onderzoek.

Met de beschikbaarheid van de stochastische en recursieve SCEM-UA-, MOSCEM-UA- en PIMLI-optimalisatiealgoritmen, ontwikkeld in de Hoofdstukken 3, 5 en 6 van dit proefschrift, is het nu mogelijk om efficiënt parameters en hun bijbehorende onzekerheden te schatten. Onzekerheid in de modelvoorspellingen komt echter niet alleen voort uit onzekerheid in de parameters, maar wordt ook veroorzaakt door meetfouten in de modelinput en output, en door structurele fouten in het model zelf. Deze laatste fouten ontstaan door simplificatie en aggregatie van ruimtelijk verdeelde patronen in een conceptueel model. Het niet op een juiste wijze omgaan met al deze foutenbronnen resulteert in modellsimulaties en bijbehorende onzekerheidsintervallen, die veelal niet het gemeten systeemgedrag weerspiegelen. Dit is meestal zichtbaar in een tijdserie van foutenresiduen, die aanzienlijke variaties in zuiverheid, variantie en correlatiestructuur vertonen onder verschillende hydrologisch condities. Om beter met input-, output-, parameter- en modelfouten om te gaan gedurende hydrologische modelkalibratie, wordt in Hoofdstuk 7 het Simultaneous Optimization and Data Assimilation (SODA) framework gepresenteerd, die de parameter-zoekefficiëntie van het SCEM-UA-algoritme (Hoofdstuk 3) en de kracht en efficiëntie van de Ensemble Kalman Filter (EnKF) combineert om tegelijkertijd parameters en toestandsvariabelen in hydrologische modellen te schatten. In dit Hoofdstuk wordt ook een niet-parametrische methode besproken, die uitermate geschikt is om de meetfouten van outputdata te schatten. Het nut en de toepasbaarheid van het SODA framework wordt vervolgens geïllustreerd aan de hand van twee verschillende casestudies. De eerste studie betreft het hoog-niet-lineaire, drie-parameter-Lorenz-model, en heeft aangetoond dat het SODA framework inderdaad tegelijkertijd parameters en toestanden kan schatten, wanneer het geconfronteerd wordt met bijna-chaotisch gedrag. In de tweede studie is SODA toegepast op een simpel, conceptueel rivierafvoermodel, gebruikmakend van data van het Leaf River-stroomgebied in Mississippi. Het vermogen van SODA om op een juiste wijze met input-, output-, parameter- en modelfouten om te gaan, resulteert in meer realistische parameter- en modelonzekerheden. Voorts levert een gedetailleerde interpretatie van de berekende recursieve toestandsaanpassingen met SODA waardevolle informatie op, die direct gebruikt kan worden om de proceskennis in hydrologische modellen te verbeteren.

Ter afsluiting van dit proefschrift wordt in Hoofdstuk 8 een epiloog gepresenteerd. Hier worden een aantal belangrijke vraagstukken behandeld, die met behulp van de SCEM-UA, PIMLI, MOSCEM-UA, en SODA algoritmen nu beantwoord kunnen worden. Ook wordt hier een richting geschetst voor mogelijk vervolg onderzoek.
LIST OF PUBLICATIONS

The work presented in this thesis comprises a substantial part of the research that has been conducted during the course of my 4-year PhD appointment at the University of Amsterdam. During this period, I have pleasantly worked with various individuals on a range of different research topics, including the interaction between dissolved organic matter and heavy metals, the interpretation of Time Domain Reflectometry waveforms, and the cycling of nitrogen in forested ecosystems. Because these investigations are beyond the general objectives of the PhD-research, their results are not incorporated in this thesis. For completeness, published papers related to this research are listed below. Papers related to the main Chapters of this thesis are also listed but indicated with an asterisk.

Peer-reviewed papers published


LIST OF PUBLICATIONS


Peer-reviewed papers, in press


Peer-reviewed papers, in review


REFERENCES


REFERENCES


REFERENCES


REFERENCES


Golden software, SURFER, version 6.04c, Golden software, Colorado, 1996.


244


REFERENCES


REFERENCES


REFERENCES


REFERENCES


REFERENCES


REFERENCES


CURRICULUM VITAE

Jasper Alexander Vrugt was born at 8.21 pm on the 28th of February 1976 in the hospital of the Free University (VU) in Amsterdam, The Netherlands. Much of his childhood he spent fishing, playing soccer and chess with his brother, listening to Dire Straits and Neil Diamond, and building/breaking down cars and air planes using advanced Lego Technic construction sets. After graduating from the Copernicus high-school in the wonderful city of Hoorn in the spring of 1994, he started his study at the Department of Physical Geography and Soil Science at the University of Amsterdam. He obtained his MS-degree in August 1999 (cum laude), and continued with his PhD research in January 2000 at the same department. During this period, he developed a strong interest in the modeling of complex systems and natural phenomena, and application and development of systems engineering approaches that merge measurement, numerical modeling and advanced data analysis techniques to increase understanding of cause-effect relationships in environmental systems in general, and in hydrologic sciences specifically. He spent about 50% of his four-year PhD appointment, visiting different research groups in the United States, in order to more effectively and efficiently work with various experts in the hydrologic sciences. Currently, he is applying the various algorithms presented in this thesis to biological problems.