MULTICRITERIA CALIBRATION OF HYDROLOGIC MODELS

By

Douglas Patrick Boyle

A Dissertation Submitted to the Faculty of the

DEPARTMENT OF HYDROLOGY AND WATER RESOURCES

In Partial Fulfillment of the Requirements
For the Degree of

DOCTOR OF PHILOSOPHY
WITH A MAJOR IN HYDROLOGY

In the Graduate College

The University of Arizona

2001

UMI Number: 3023521

UMI®

UMI Microform 3023521

Bell & Howell Information and Learning Company
300 North Zeeb Road
P.O. Box 1346
Ann Arbor, MI 48106-1346

THE UNIVERSITY OF ARIZONA ®
GRADUATE COLLEGE

As members of the Final Examination Committee, we certify that we have read the dissertation prepared by Douglas Patrick Boyle

entitled Multicriteria Calibration of Hydrologic Models

and recommend that it be accepted as fulfilling the dissertation requirement for the Degree of Doctor of Philosophy

Dr. Soroosh Sorooshian	Date 6/8/01
Dr. Hoshin Vijai Gupta	Date 6/8/01
Dr. Juan B. Valdes	Date 6/8/01
Dr. Ferenc Szidarovszky	Date 6/8/01
Dr. Suvrajeet Sen	Date 6/8/01

Final approval and acceptance of this dissertation is contingent upon the candidate's submission of the final copy of the dissertation to the Graduate College.

I hereby certify that I have read this dissertation prepared under my direction and recommend that it be accepted as fulfilling the dissertation requirement.

Dissertation Director
Dr. Soroosh Sorooshian

Date 6/8/01

STATEMENT BY THE AUTHOR

This dissertation has been submitted in partial fulfillment of requirements for an advanced degree at The University of Arizona and is deposited in the University Library to be made available to borrowers under rules of the Library.

Brief quotations from this dissertation are allowable without special permission, provided that accurate acknowledgment of source is made. Requests for permission for extended quotation from or reproduction of this manuscript in whole or in part may be granted by the head of the major department or the Dean of the Graduate College when in his or her judgment the proposed use of the material is in the interests of scholarship. In all other instances, however, permission must be obtained from the author.

SIGNED: *D.L. P. B.*

ACKNOWLEDGMENTS

First I would like to thank my parents, Nancy Anne Hafner and Laverne Calvin Hafner for the opportunity and encouragement to study as an undergraduate and graduate student. All of this could not have been possible without their constant love and support. I wish to thank my wife, Mary Horvath, for her patience, support, love, and encouragement throughout the last six years.

I would like to thank Jim Broermann, Corrie Thies, Rannie Fox, and Dan Braithwaite for all of their support over the last five years. I also wish to thank Dr. Juan Valdes, Dr. Ferenc Szidarovszky, and Dr. Suvrajeet Sen, for serving on my committee and providing suggestions that improved the research presented in this dissertation. I wish to give special thanks to Dr. Hoshin Vijay Gupta for his guidance, patience, and support as both the director of my research and a friend. Finally, I wish to thank Dr. Soroosh Sorooshian for providing me with the opportunity, support, and freedom that made this work possible. To all these individuals I am deeply indebted.

TABLE OF CONTENTS

TABLE OF CONTENTS – *Continued*

LIST OF ILLUSTRATIONS

LIST OF ILLUSTRATIONS – *Continued*

LIST OF ILLUSTRATIONS – *Continued*

LIST OF TABLES

ABSTRACT

The level of spatial and vertical detail of important hydrologic processes within a watershed that needs to be represented by a conceptual rainfall-runoff (CRR) model in order to accurately simulate the streamflow is not well understood. The paucity of high-resolution hydrologic information in the past guided the direction of model development to more accurately represent processes directly related to the vertical movement of moisture within the watershed rather than the spatial variability of these processes. As a result, many of the CRR models currently available are so complex (vertically), that expert knowledge of the model and watershed system is required to successfully estimate values for model parameters using manual methods. Automatic parameter estimation procedures, developed to reduce the time and effort required with manual methods, do not provide parameter estimates and hydrograph simulations that are considered acceptable by the hydrologists responsible for operational forecasting. Newly available, high-resolution hydrologic information may provide insight to the spatial variability of important rainfall-runoff processes. However, effective and efficient methods to incorporate the data into the current modeling strategies need to be developed.

This work describes a new hybrid multicriteria calibration approach that combines the strength of automatic and manual calibration methods. The new approach was used to investigate the benefits of different levels of spatial and vertical representation of important watershed hydrologic variables with conceptual rainfall runoff models.

CHAPTER ONE

INTRODUCTION AND SCOPE

1.1 INTRODUCTION

Conceptual rainfall-runoff (CRR) models have become widely used for streamflow forecasting as the demand for timely and accurate forecasts has increases. The level of spatial and vertical detail of important hydrologic processes within a watershed that needs to be represented by a CRR model in order to accurately simulate the streamflow, however, is not well understood. The desire to improve model performance has lead to the development of CRR models with highly elaborate functional structures to represent the real watershed system as accurately as possible. This has resulted in a mismatch between model complexity, problem complexity, and the needs and abilities of the user.

The hydrologic modeling problem can be partitioned into three main components; hydrologic model structure, hydrologic data, and parameter estimation procedures (see Table 1.1). Successful development and application of any hydrologic model requires careful consideration of each component and its relevance to the overall modeling problem. In the next section, the important aspects of each component of the hydrologic modeling problem are described in the context of the development and application of CRR models.

Table 1.1: Hydrologic modeling problem overview

Hydrologic Data	Parameter Estimation	Hydrologic Model
Precipitation	Consistency	Spatial Complexity
Soils Properties	Performance	a) Excess
Vegetation		b) Routing
DEM	Behaviors	c) Parameters
Energy Forcings		
Streamflow	Measures	Vertical Complexity
a) Outlet	Procedures	a) Infiltration
b) Interior		b) Storage
		c) PET

1.2 BACKGROUND AND LITERATURE REVIEW

1.2.1 Hydrologic Models

Conceptual rainfall-runoff (CRR) models provide an approximate, lumped description of the dominant sub-watershed scale processes that contribute to the overall watershed scale hydrologic response of the watershed system. In their most basic form, CRR models transform rainfall into runoff with two main components, precipitation excess generation and flow routing. Precipitation excess is generated as a function of the vertical movement of moisture (precipitation, evaporation, transpiration, and losses to the system) into and out of the watershed. The flow routing component involves the movement of the excess precipitation over the land surface and along stream and channel networks to the outlet of the watershed.

The variability of the excess generation process within the watershed is related to the level of spatial variability of the soil properties, vegetation type, and precipitation rates throughout the watershed. In the past, high-resolution information describing these characteristics was not readily available. As a result, performance improvements of CRR models were primarily focused on improving the representation of processes directly related to the vertical movement of moisture within the watershed rather than the spatial variability of these processes.

The level of vertical detail describing the movement of moisture through the soil (vertical model structure complexity) can vary significantly among different CRR models. The Antecedent Precipitation Index method (Linsley et al., 1949; Sittner et al.,

1969) is an example of a simple CRR model in which an antecedent soil moisture index is combined with a set of linear equations to estimate runoff from storm precipitation. Examples of CRR models with complex vertical structures include the Stanford Watershed Model (Crawford, 1966), the Sacramento Soil Moisture Accounting (SAC-SMA) Model (Burnash et al., 1973), the KINEROS model (Woolhiser et al., 1990), and TOPMODEL (Beven and Kirkby, 1979). These models all use large numbers of parameters and complex relationships to represent the vertical movement of moisture through the soil.

In general one might expect that the more detail or complexity included in the structure of the model, the better the representation of the real watershed behavior (streamflow). A number of studies have compared the performance of simple versus complex models [Naef, 1981; Loague and Freeze, 1985; Wilcox et al, 1990; Mayeux, 1995; Michaud and Sorooshian, 1994; Refsgaard and Knudesn (1996); among others]. Results from these studies have generally suggested that performance of simple models are approximately as accurate as complex models. The development and application of complex models, however, continues despite the results of these studies.

Several researchers have recognized the importance of selecting a model with a level of complexity that is appropriate for a given problem [Wheater et al. 1986, Beck et al., 1990; Jakeman and Hornberger, 1993; Grayson and Chiew, 1994; among others]. Jakeman and Hornberger (1993) explored the constraints that observed data place on the allowable complexity of rainfall-runoff models. They used time series techniques to determine how many parameters are appropriate to describe the relationship between precipitation and streamflow for several watersheds of various sizes using available precipitation, air temperature and streamflow data. Their results suggest that the information content in a rainfall-runoff record is sufficient to support models of only very limited complexity.

1.2.2 Hydrologic Data

In the past, areal averaged inputs to the model, such as precipitation and potential evapotranspiration, have been estimated from point measurements made at gauges within or near the watershed. Remotely sensed, high-resolution, hydrologic data are now becoming available in the United States through a variety of different sources. The NWS Next Generation Radar system (NEXRAD) stage III product offers gridded precipitation estimates spatially averaged over 4 km by 4 km grid cells and temporally averaged over 1 hour. Other high-resolution hydrologic information such as topographic, vegetation, and soils characteristics are widely available from digital databases (e.g., USGS 3-arc second digital elevation models, state soils geographic (STATSGO) 1-km gridded soils data [U.S. Department of Agriculture, 1994], U.S. Forest Service 1-km gridded vegetation type and density data [U.S. Department of Agriculture, 1992], and USGS 1-km gridded land use/land cover data [Anderson et al., 1997]). The incorporation of these high-resolution data sets, in particular, the NEXRAD stage III data, into current modeling procedures is considered highly desirable by hydrologists. The development of efficient and effective methods for doing so is an active area of research.

A simple method to incorporate the high-resolution data into the modeling process is to average the information over the entire area of the watershed and proceed with the current lumped model application. The main advantage of this approach is that the existing modeling structure does not need to be modified to use the new data. The main disadvantage is clearly the loss of the spatial distribution of information as well as the potential for further modeling improvement and understanding. Another, strategy is the "semi-distributed" approach in which the watershed is partitioned into a network of hydrologic units based on the spatial variability of the precipitation. The main disadvantage of this strategy is the increase in model complexity and parameters parallel to the increase in partitioning. For complex, highly parameterized models, as the number of hydrologic units is increased, the calibration procedure quickly becomes intractable. Further, many of the parameters may not be supported (identifiable) by the information contained within the observed data, remotely sensed or otherwise.

Regardless of the model application procedure, successful estimation of model parameters will be limited by the information content and length of the high-resolution data sets. While it is generally accepted that the high-resolution data sets contain more spatial information than point measurements, it is not clear that the short period of time for which these data sets are available is adequate to estimate model parameters with current calibration procedures. The results of previous studies (Yapo et al., 1998) indicate that, with current automatic calibration procedures, approximately eight years of streamflow and precipitation data are required to adequately parameterize the SAC-SMA model in lumped applications. An additional 8 years of historical data is recommended for model evaluation. However, the length of NEXRAD stage III data sets for most locations in the United States is less than 6 years.

Additional information available in long-term gauge precipitation and streamflow data may help to identify certain model parameters, thus reducing the negative effect of the short NEXRAD stage III data records. Studies of the mean areal precipitation estimates derived from NEXRAD versus those obtained from precipitation gauge networks have revealed differences between the long term accumulations, individual storm totals, and the timing of the events (Finnerty and Johnson, 1997). The impact of these differences on model parameterization is not fully understood and is the subject of current research efforts.

1.2.3 Parameter Estimation

Model parameter estimation (calibration) involves the selection of values for the parameters so that the model matches the behavior of the watershed system as closely as possible. CRR models are often difficult to calibrate due to the large number of parameters and complex relationships. For example, the SAC-SMA model has 17 parameters whose values must be specified. While a few of these parameters might be estimated by relating them to observable characteristics of the watershed, most are abstract conceptual representations of non-measurable watershed characteristics that must be estimated through a calibration procedure.

Calibration approaches generally involve two components: (1) evaluation of the "closeness" between the model outputs and the corresponding measurement data, and (2) adjustment of the values of the parameters to improve the closeness. The important characteristics of any calibration approach are consistency (the results should be repeatable) and performance (the approach should find the optimal solution(s) in an efficient manner) – the approach should provide the user with an "optimal" solution(s) irrespective of the data, model, or user's ability.

Manual calibration is the approach most widely used to calibrate hydrologic models. In this approach, the "closeness" is typically evaluated in terms of several measures, and a semi-intuitive trial-and-error process is used to perform the parameter adjustment. Usually, the exact number and nature of these measures is not specified clearly. While some aspects of "closeness" may be evaluated in terms of objective measures (i.e., one or more mathematical criteria), most are in fact evaluated subjectively based on visual comparison of the model outputs and the data. Although the performance evaluation and parameter adjustment procedures are usually influenced by guidelines established through previous experiences of model calibration, the actual sequence of parameter adjustments will vary from person to person based on their experience and training, their understanding of the model structure, the properties of the data, and the characteristics of the watershed system.

The NWS has developed a sophisticated, highly interactive manual procedure to estimate parameter values for the SAC-SMA model [Anderson, 1997]. This process has been developed and refined over the years to result in excellent model calibrations. However, the process is complicated, difficult to learn, and highly labor-intensive, requiring a substantial commitment of human resources. As a result, the expertise acquired by one individual through extensive hands-on training and experience with a specific model is not easily transferred to another person (or another model).

Over the past two and a half decades, much research has been devoted to the development of automatic methods for model calibration that take advantage of the speed and power of digital computers. Such methods seek to be objective and relatively easy to implement. However, the model calibrations provided by such methods do not tend to provide parameter estimates and hydrograph simulations that are considered acceptable by the hydrologists responsible for operational forecasting and have therefore not entered into widespread use. Gupta et al. [1998] acknowledged some of the more serious limitations with the classical automatic approach and presented a more general multicriteria framework for model calibration that recognizes the multiobjective nature of the problem. Yapo et al. [1998] developed the MOCOM-UA algorithm, an effective and efficient methodology for solving the multiobjective global optimization problem. The new multicriteria framework provides an opportunity to reexamine the methods in which automatic calibration procedures evaluate the "closeness" between the model outputs and the corresponding measurement data.

Traditional automated calibration methods rely on mathematical measures to compare the computed output to the observed data. These mathematical measures are primarily based on the computation of the differences, or errors, between the model output value and the observed value at each time step. The errors are then aggregated over a large range of hydrologic behaviors to estimate an "overall" measure of model performance. This aspect of the automatic approach is significantly different than the manual approach and potentially the primary reason for generating solutions that are not considered desirable by experts. In addition, the use of overall measures provides little, if any, information to the user about the performance of specific structural components of the model, potentially hindering future improvements of the model structure.

1.3 STATEMENT OF THE PROBLEM

The current approach to development and application of hydrologic models for the purposes of streamflow forecasting is in need of serious review. The level of spatial and vertical detail of important hydrologic processes within a watershed that needs to be represented by a hydrologic model in order to accurately simulate the streamflow is not well understood. It is clear, however, that there is a mismatch between model structural complexity, hydrologic problem complexity, and the needs and abilities of the user. Specifically, the following issues need examination and investigation:

i) The paucity of high-resolution hydrologic information in the past guided the direction of model development to more accurately represent processes directly related to the vertical movement of moisture within the watershed rather than the spatial variability of these processes.

ii) Many of the CRR models currently available are so complex (vertically), that expert knowledge of the model and watershed system is required to successfully estimate values for model parameters using manual methods.

iii) Automatic parameter estimation procedures, developed to reduce the time and effort required with manual methods, do not provide parameter estimates and hydrograph simulations that are considered acceptable by the hydrologists responsible for operational forecasting.

iv) Newly available, high-resolution hydrologic information may provide insight to the spatial variability of important rainfall-runoff processes. However, effective and efficient methods to incorporate the data into the current modeling strategies need to be developed.

v) A redistribution of model structural complexity may be necessary to accurately reflect the vertical and spatial complexity of the watershed system.

1.4 RESEARCH TO BE UNDERTAKEN

1.4.1 Objectives

The primary objective of this work is to develop and test a strategy to improve the calibration of CRR models that combines the strengths of manual and automated approaches. This strategy will facilitate the investigation of parsimonious model structures that balance the level of spatial and vertical model complexity with the information available in the data, and the needs and abilities of the user. Specific objectives are as follows:

i. Develop a new "hybrid" approach to model calibration that combines the strengths of existing manual and automatic multicriteria methods.

ii. Demonstrate the power of the approach by means of a case study using the SAC-SMA model.

iii. Compare the results obtained with the new approach to those obtained with existing manual and single-criteria automatic approaches.

iv. Use the approach to investigate the performance improvements associated with spatial variability of the excess generation process through a variety of different semi-distributed applications of the SAC-SMA model using high-resolution hydrologic data.

v. Use the approach to investigate the performance improvements associated with different levels of vertical detail describing the movement of moisture through the soil.

1.4.2 Methods

In this work, a thorough analysis of current manual and automatic calibration approaches performed. From this analysis, a hybrid approach that combines the strengths of each is developed and tested. A multi-criteria formulation is used to "model" the evaluation techniques and strategies used in manual calibration and the resulting optimization problem is solved by means of a computerized algorithm. The new approach provides a stronger test of model performance than methods that use a single overall statistic to aggregate model errors over a large range of hydrologic behaviors.

The power of the new approach is illustrated by means of case study using the SAC-SMA model in a lumped application.

The new approach is also used to provide an assessment of the potential improvements in CRR model performance that can be achieved by semi-distributed modeling of a watershed using high-resolution remotely sensed hydrologic information. Specifically, performance comparisons are made within an automated multi-criteria calibration framework to evaluate the benefit of "spatial distribution" of the model input (precipitation), structural components (soil moisture and streamflow routing computations), and surface characteristics (parameters). A comparison of these results is made with those obtained through manual calibration.

Finally, the new approach is used to investigate the value of different levels of vertical detail in the model structure to describe the movement of moisture through the soil. To accomplish this, a variety of different model structures are constructed with different combinations of important excess generation components (e.g., interception storage, tension storage, free storage, percolation, etc.). Each model configuration is calibrated using the new hybrid approach and evaluated in terms of multiple behavioral measures.

1.4.3 Model and Data

This research focuses on applications of the SAC-SMA model for the purposes of streamflow forecasting. In particular, a research version of the C language source code, maintained by the Department of Hydrology and Water Resources at the University of

Arizona, is used in all of the SAC-SMA model applications described within this document. In addition, a simple hydrologic model, HYMOD [Boyle et al. 1997; Wagener et al., 2001], is used to investigate the value of different levels of vertical detail describing the movement of moisture through the soil. HYMOD consists of a variety of different excess generation (interception storage, tension storage, free storage, etc.), percolation, and streamflow routing functions that can be put together in different combinations to describe the different hydrologic behaviors of the watershed system.

Two data sets are used in this work. The first is the historical data from the Leaf River watershed (1950 km^2) located north of Collins, Mississippi. Forty consecutive years of data (WY 1948-88) are available for this watershed, representing a wide variety of hydrologic conditions. The Leaf River data have been used extensively in previous studies [e.g., Brazil and Hudlow, 1981; Sorooshian and Gupta, 1983]. The second data set is for the Blue River watershed located in southern Oklahoma near Blue, Oklahoma. A NEXRAD stage III precipitation data set for the Blue River watershed is available for the period 1 June 1993 to 30 April 1999. The NEXRAD product offers gridded precipitation estimates, spatially averaged over 4 km by 4 km grid cells and temporally averaged over 1 hour. For the same time period, hourly estimates of instantaneous discharge are available from a USGS surface-water discharge station. NWS-derived hourly values of Potential Evapotranspiration (PET) based on the atlas of NOAA Free Water Evaporation [NOAA, 1982] are also available. It should be noted that the PET values vary seasonally but not annually (i.e., an estimate of PET for one year has been made and is repeated for each year of record). Soil property information from the USDA

State Soil Geographic Database (STATSGO) are available as well as a digital elevation model (DEM) of the Blue River watershed created by the staff at the NWS hydrology lab from 30 meter elevation data. The Blue River data set is used in Chapter 3 to explore the value of representing spatial variability of important rainfall-runoff processes.

1.5 ORGANIZATION

This dissertation is organized into five chapters as follows:

Chapter two presents an analysis of the similarities and differences between the automatic and manual approaches to calibration and proposes a new "hybrid" approach that combines the strengths of each. In Chapter three, the approach is used to provide an assessment of the potential improvements in rainfall-runoff model performance that can be achieved by semi-distributed modeling of a watershed using radar-based (NEXRAD) remotely sensed precipitation data. The relative benefits of spatially distributing the model input (precipitation), structural components (soil moisture and streamflow routing), and surface characteristics (parameters) are examined. In Chapter four, the approach is used to investigate the performance improvements associated with different levels of vertical detail describing the movement of moisture through the soil. Finally, in Chapter five, the conclusions of this research are given and directions for future extensions of this work are suggested.

1.6 PUBLICATION OF RESEARCH

The automatic multi-criteria approach described in Chapter 2 has been published in a professional journal [Boyle et al., 2000]. A manuscript describing the semi-distributed modeling study presented in Chapter 3 is in press [Boyle, et al., 2001]. A manuscript describing the investigation of the performance improvements associated with different levels of vertical detail describing the movement of moisture through the soil presented in Chapter 4 is in preparation. This manuscript will be submitted to an appropriate professional journal in the near future. Finally, an initial model complexity study (results not specifically presented in this dissertation) conducted in collaboration with researchers from the Department of Civil and Environmental Engineering, Imperial College, London, UK is currently in press [Wagener, et al., 2001].

CHAPTER TWO

COMBINING THE STRENGTHS OF AUTOMATIC AND MANUAL METHODS

2.1 INTRODUCTION AND SCOPE

Conceptual Rainfall-Runoff (CRR) models have become widely used for flood forecasting as the demand for timely and accurate forecasts has increased. Such models provide an approximate, lumped description of the dominant subwatershed scale processes that contribute to the overall watershed scale hydrologic response of the system. Most operational CRR models have on the order of 10 or more parameters, whose values must be selected so that the modeled response to rainfall closely simulates the actual behavior of the watershed of interest. For example, the SACramento Soil Moisture Accounting (SAC-SMA) model [Burnash et al., 1973] is used by the National Weather Service (NWS) for flood forecasting throughout the United States (Figure 2.1). The model has 17 parameters whose values must be specified (Table 2.1). While a few of these parameters might be estimated by relating them to observable characteristics of the watershed, most are abstract conceptual representations of non-measurable watershed characteristics that must be estimated through a calibration procedure.

The NWS has developed a sophisticated, highly interactive manual procedure to estimate parameter values for the SAC-SMA model [Anderson, 1997]. This process has been developed and refined over the years to result in excellent model calibrations.

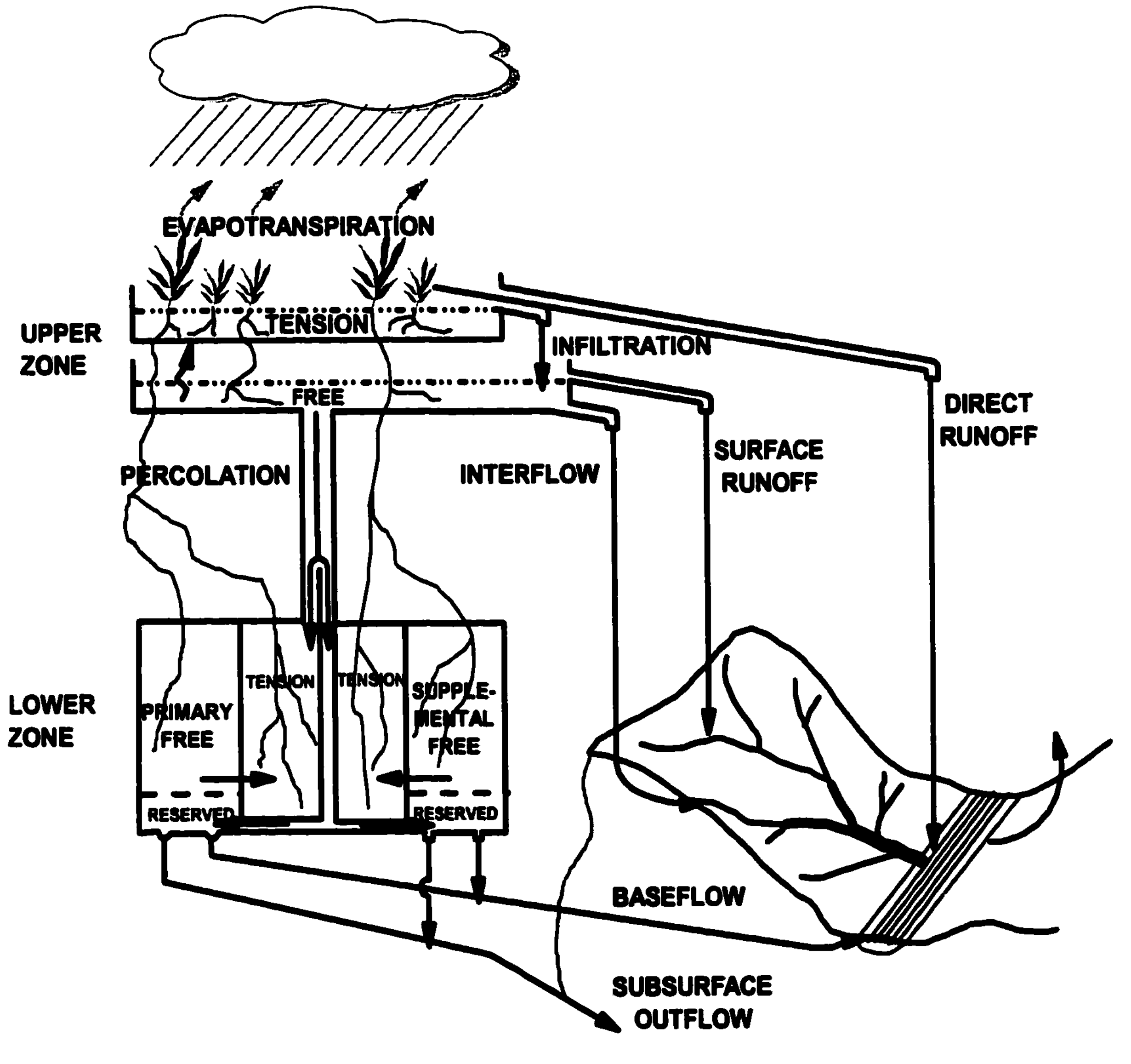

Figure 2.1 National Weather Service Sacramento Soil Moisture Accounting Model (SAC-SMA).

Table 2.1: Parameters and state variables of the SAC-SMA model

Parameter	Level Zero	Level One	Level Two: Multiple Criteria: Range	Level Two: Multiple Criteria: "Best"	Level Two: Single Criterion: RMSE	Level Two: Single Criterion: LOG-RMSE
(a) Maximum Capacity Thresholds						
UZTWM	1.0 - 150.0	-	20.8 - 96.3	58.0	10.1	127.6
UZFWM	1.0 - 150.0	-	26.9 - 49.3	45.8	31.2	21.0
LZTWM	1.0 - 500.0	-	196.5 - 250.3	247.5	257.7	210.0
LZFPM	1.0 - 1,000.0	120.0 - 140.0	122.4 - 128.3	124.4	120.0	120.1
LZFSM	1.0 - 1,000.0	40.0 - 60.0	40.0 - 40.9	40.5	40.1	40.1
ADIMP	0.0 - 0.4	-	0.38 - 0.40	0.40	0.37	0.28
(b) Recession Parameters						
UZK	0.1 - 0.5	-	0.13 - 0.31	0.18	0.19	0.34
LZPK	0.0001 - 0.025	0.006 - 0.01	0.006 - 0.008	0.006	0.010	0.007
LZSK	0.01 - 0.25	0.15 - 0.20	0.17 - 0.19	0.18	0.20	0.15
(c) Percolation and Other Parameters						
ZPERC	1.0 - 250.0	-	204.6 - 250.0	230.2	249.2	247.5
REXP	0.0 - 5.0	-	2.5 - 3.8	3.2	2.5	3.5
PCTIM	0.0 - 0.1	0.0 - 0.02	0.0033 - 0.0088	0.005	0.00002	0.0178
PFREE	0.0 - 0.6	-	0.002 - 0.13	0.04	0.0002	0.25
(d) Parameters Not Optimized						
RIVA	0.00					
SIDE	0.00					
RSERV	0.30					
PXMLT	1.00					
(e) State Variables						
UZTWC						
UZFWC						
LZTWC						
LZFPC						
LZFSC						
ADIMC						

However, the process is complicated, difficult to learn, and highly labor-intensive, requiring a substantial commitment of human resources. As a result, the expertise acquired by one individual through extensive hands-on training and experience with a specific model is not easily transferred to another person (or another model).

Over the past two and a half decades, much research has been devoted to the development of automatic methods for model calibration that take advantage of the speed and power of digital computers. Such methods seek to be objective and relatively easy to implement. However, the model calibrations provided by such methods do not tend to provide parameter estimates and hydrograph simulations that are considered acceptable by the hydrologists responsible for operational forecasting. As a result, automatic methods have not yet entered into widespread use.

In this chapter, an analysis of the similarities and differences between the manual and automatic approaches to hydrologic model calibration is performed and a hybrid approach that combines the strengths of each is proposed. The approach "models" the evaluation techniques and strategies used in manual calibration by using several objective measures to reflect different observable characteristics of watershed behavior. The resulting optimization problem is posed within the multi-criterion framework presented by Gupta et al. [1998] and solved by means of a computerized optimization algorithm [Yapo et al., 1998; Bastidas et al., 1999]. The following sections describe and discuss the rationale for this methodology and illustrate the power of the approach through a case study using the SAC-SMA model.

2.2 STRATEGIES FOR PARAMETER ESTIMATION

2.2.1 Three Level Classification Scheme

Model parameter estimation involves the selection of valucs for the parameters so that the model matches the behavior of the watershed system as closely as possible. For the purpose of this discussion, we shall classify the parameter estimation process into three levels of increasing sophistication.

In Level Zero, approximate ranges for the parameter estimates are specified by examining lookup tables or by borrowing values from similar watersheds that have been previously calibrated. Note that these ranges reflect our uncertainty in the values of the parameters and might be termed prior estimates, in the sense that they are not conditioned on any input-output time series data collected for the watershed of interest. If a point estimate is desired, it can be selected from within this range. This approach is often applied in the case of ungaged watersheds or when historical data records are insufficient for the application of Level One and Two calibration techniques (see below). It is also used typically with complex "physics-based" models whose parameters are intended to represent measurable physical quantities (e.g., distributed hydrologic models such as SHE and land-surface energy and water balance models such as BATS).

Levels One and Two involve the use of progressively more sophisticated procedures for refining the parameter estimates by the use of information available in input-output time series data collected for the watershed of interest. They might, therefore, be termed posterior estimates. In Level One, the parameter ranges are refined by identifying periods in the output time series where the effects of individual parameters (or small group of parameters) are relatively dominant and can therefore be isolated. A defining characteristic of Level One is that the effects of parameter interactions are essentially ignored. Such procedures are commonly described in introductory level textbooks on hydrology [e.g., see Chapter 7 of Linsley et al., 1982]. The U.S. National Weather Service follows an explicit set of procedures and guidelines for Level One refinement of individual parameters (discussed further in Section 2.2.2).

Finally, Level Two consists of further refinement (narrowing) of the parameter ranges by means of a detailed analysis of parameter interactions and model performance tradeoffs. This is the most difficult step in the parameter estimation process, because it requires a great deal of understanding of the model and typically involves complex decisions that weigh the multiple effects of adjusting several parameters at a time.

2.2.2 Approaches to Calibration

The methods employed in both Levels One and Two involve two components: (a) evaluation of the "closeness" between the model outputs and the corresponding measurement data, and (b) adjustment of the values of the parameters to improve the closeness. Manual (sometimes called expert) and automatic calibration techniques can be compared and contrasted in terms of how each of these components is implemented.

Manual Calibration

Manual calibration is the approach most widely used to calibrate hydrologic models. In this approach, the "closeness" is typically evaluated in terms of several (more than three!) measures, and a semi-intuitive trial-and-error process is used to perform the parameter adjustment. Usually, the exact number and nature of these measures is not specified clearly. While some aspects of "closeness" may be evaluated in terms of objective measures (i.e., one or more mathematical criteria), most are in fact evaluated subjectively based on visual comparison of the model outputs and the data. Although the performance evaluation and parameter adjustment procedures are usually influenced by guidelines established through previous experiences of model calibration, the actual sequence of parameter adjustments will vary from person to person based on their experience and training, their understanding of the model structure, the properties of the data, and the characteristics of the watershed system. In the following discussion, references are made mainly to the parameter estimation procedures used by the NWS during manual calibration of the SAC-SMA flood forecast model.

The manual calibration process employed by the NWS begins with Level Zero parameter estimates obtained (primarily) by examining the range of parameter values for previously calibrated watersheds in the local forecast group (a group of neighboring watersheds having similar geology, hydrology, and climatology). Next, a systematic sequence of steps is followed to develop Level One parameter estimates by examination of the hydrometeorological data base for the watershed [Peck, 1976]. Periods in the observed time series data are identified where specific hydrologic processes are dominant (e.g., baseflow, interflow, surface flow, evaporation, transpiration, abstraction, infiltration, etc). For each of these periods, the closeness of the model and the data is evaluated visually, and values for the relevant parameters are estimated by heuristic methods. As a trivial example, periods dominated by baseflow are used to estimate the baseflow recession rate parameter of the model (for each period) as the slope of the log of the hydrograph time series (Figure 2.2). The range of recession values identified over all the relevant periods then represents the new (refined) uncertainty in the value of the baseflow recession parameter; a point estimate is typically computed by averaging the values. For the SAC-SMA model, Peck [1976] listed 5 parameters (lzfpm, lzpk, lzfsm, lzsk, and pctim) for which Level One estimates can be easily obtained from the observed hydrograph and precipitation data. Level One estimates for another 6 parameters (lztwm, uztwm, uzk, ssout, uzfwm, and pfree) are considered possible but more difficult to obtain from the data. The estimates for the remaining 6 parameters (sarva, zperc, rexp, side, adimp, and rserv) cannot be refined by Level One procedures.

A graphical user interface called the Interactive Calibration Program (ICP) [NWS, 1997] was developed by the NWS to facilitate the identification of different hydrometeorologic periods within the data for Level One analysis. The ICP program will also be an integral part of the new Advanced Hydrologic Prediction System (AHPS) initiative recently funded by Congress starting Fiscal year 2000. However, the ICP code also enables more detailed empirical and statistical analyses of the observed data, state values, and outputs simulated by the model and is, therefore, the primary tool used to obtain Level Two parameter estimates via manual calibration. The interactive graphical display allows the user to visually observe the impact of parameter adjustments on the internal behavior of the model (state variables) and on the "closeness" of the model-simulated streamflows to the observed hydrograph.

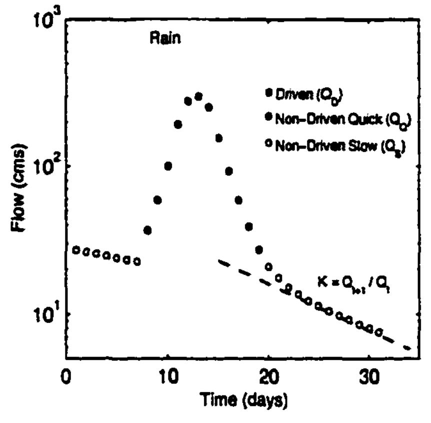

Figure 2.2 Partitioning of the observed hydrograph into three components: Q_D, Q_Q, and Q_S. The dashed line shows how the observed hydrograph can be used to estimate the recession constant, K

The complex effects of parameter interactions on the model responses make Level Two calibration a difficult process, requiring training and a great deal of practice to master. The hydrologist must simultaneously evaluate a number of subjective and objective criteria while iteratively adjusting the parameter values, some of which influence the overall behavior of the model while others have a significant impact only during certain special hydrologic events. While there are experts skilled at manual methods for Level Two parameter estimation, the procedures they use are not explicitly defined (some very general guidelines can be found in the ICP manual [NWS, 1997]).

A comprehensive understanding of the model, the real watershed system, and the data is required to produce consistent and reliable results with manual calibration. Unfortunately, this knowledge is not easily transferred from person to person or model to model and must come from extensive hands-on training and experience with the specific model. Even when performed by an expert, the process is highly labor-intensive and requires a substantial commitment of human resources. Manual calibration of the SAC-SMA model to a single watershed may require several hundred hours of effort. Further, the subjectivity of the evaluation procedure and the complex trade-offs in the model's abilities make it very likely that many different solutions (sets of parameter values) may seem to produce equally "good" results. The NWS conducts several workshops annually to train hydrologists in calibration procedures for their models. Development of the skills needed to conduct a competent calibration for the SAC-SMA model can require several months of practice [Michael Smith, Hydrologic Research Lab, NWS personal communication, 1999].

Automatic Calibration

Automatic calibration procedures for hydrologic models have been under development for at least three and a half decades, with the degree of sophistication generally paralleling the increases in computing power. The goal has been to develop an objective strategy for parameter estimation that provides consistent performance by eliminating the kinds of subjective human judgements involved in the manual approach. In the classical approach, borrowed from systems and operations research theory, the "closeness" is typically evaluated in terms of a single objective measure (a mathematical criterion) of the overall difference between the simulated and observed hydrographs, and the parameter adjustment is performed using an optimization algorithm. Considerable research has been devoted to trying to identify a "best" criterion [e.g., Sorooshian and Dracup, 1980; James and Burges, 1982; Kuczera 1983a,b; among others] and the "best" optimization algorithms [Brazil and Krajewski, 1987; Brazil, 1988; Wang, 1991; Duan et al., 1992, 1993; Sorooshian et al., 1993; among others]. Recent work has suggested that automatic calibration procedures also can be extended to handle multiple criteria [Gupta et al., 1998, 1999; Yapo et al., 1998].

In the automatic approach, the performance evaluation and parameter adjustment procedures are objective in the sense that they establish explicit rules by which the actual sequence of parameter adjustments is made. The power of the procedure, therefore, depends on how well it has been designed to reflect the factors important to a successful calibration, and much effort has been devoted to establishing what those factors are. The following discussion is based primarily on experiences with the development and testing of automatic parameter estimation procedures for the SAC-SMA flood forecast model. In this discussion, the classical single-criterion method and the new multi-criteria approach are explicitly distinguished.

In general, implementation of an automatic calibration process requires the user to specify a region of the parameter space which is considered to contain feasible values for the parameters; this is given typically as upper and lower bounds on each parameter. It is common practice for these bounds to be established via the Level Zero and Level One parameter estimation procedures described above for manual calibration. The single-criterion and multi-criteria automatic calibration methods, therefore, differ only in their approach to Level Two calibration.

The single-criterion approach searches the feasible parameter space for a single point that optimizes the mathematical criterion selected to measure the closeness of the model output and the data. The OPT3 automatic calibration code used by the NWS allows the user to select from among several choices for the criterion and the optimization algorithm. In general, the criterion most commonly used in the literature has been the root mean squared error (RMSE) evaluated on either the streamflows or the log of the streamflows:

$$RMSE = \sqrt{\frac{1}{N}\sum_{t=1}^{N}\left(z_t^{sim} - z_t^{obs}\right)^2} \tag{2.1}$$

where z_t^{sim} is either the simulated streamflow or log-streamflow, and z_t^{obs} is the corresponding observed value. Of the several optimization algorithms that have been tested, the SCE-UA method has generally proved to be both robust and efficient [e.g., Duan et al, 1992; Gan and Biftu, 1996; Kuzcera, 1997; Thyer et al., 1999].

It has been suggested that, although the single-criterion automatic approach is able to provide good parameter estimates, it can also "degenerate into pure curve-fitting and produce a set of parameters that fit the calibration reasonably well, but are hydrologically unrealistic" [Peck, 1976]. For this reason, and because automatic approaches require some mathematical background and familiarization with more complicated computational tools, the approach has not become popular for routine calibration of hydrologic models. It is not difficult to see that the approach has both a major strength and a major weakness. Its strength is its use of a robust optimization procedure to rapidly make successful parameter adjustments in the presence of strong parameter interaction. Its weakness, however, is its "complete dependency on one error function" [Peck, 1976].

The main problem with the single-criterion automatic approach is that the criterion must be chosen carefully to measure the difference between the simulated and observed hydrographs in a manner that does not place undue emphasis on matching one aspect of the hydrograph at the expense of another. While an expert can give due consideration to this fact in the course of manual calibration, the selection of a single criterion before initiating the process of calibration can predispose the automatic procedure towards an inappropriate (even hydrologically unrealistic) solution. For example, it has been the experience of researches at the University of Arizona that use of the RMSE criterion evaluated on the streamflows tends to overemphasize fitting of the flood peaks and leads to strongly biased simulations of the recessions (and hence incorrect tracking of soil-moisture storages).

The multi-criteria approach addresses this problem through a two-step process. In the first step, an automatic search of the feasible parameter space is used to find the set of solutions (the so-called "Pareto optimal" region) which simultaneously optimizes several user-selected criteria that measure different aspects of the closeness of the model output and the data. This quickly results in several viable solutions, reflecting the range of different ways in which the hydrograph can be simulated with different kinds of "minimal" error [Yapo et al., 1997; Gupta et al., 1998]. In the second step, the solutions having unacceptable trade-offs in fitting of the different parts of the hydrograph are rejected, and additional criteria (both objective and subjective) are employed to narrow the solution space. The major objective of this paper is to demonstrate how the two-step multi-criteria automatic approach can be used to develop a hybrid strategy that combines the strengths of the manual and automatic calibration, resulting in efficient, yet acceptable, estimates for the parameters of a conceptual hydrologic model.

2.3 COMBINING THE STRENGTHS OF MANUAL AND AUTOMATIC CALIBRATION

Manual calibration uses a subjective process of visual inspection and comparison of the model output and the observed data to implicitly evaluate (measure) the ability of the model to simulate specific aspects of the hydrologic behavior. Examples of these behaviors include the magnitude and timing of the peak flows and the shapes of the rising and falling limbs of the hydrograph. In addition, statistical criteria such as monthly and seasonal biases are typically used to provide an objective evaluation of the overall (long-term) behavior of the model. In principle, it could be said that the hydrologists doing manual calibration keep track of a number of different "criteria", most of which are not explicitly defined, while adjusting the model parameters. This allows them to balance the trade-offs in the ability of the model to simulate various aspects of the hydrograph (i.e., to consider the characteristics of the model structural error) with recognition of any potential errors in the observed data. Emulation of this process via automatic calibration requires the use of explicit mathematical criteria to approximate (model) the implicit measures used in the manual approach. This development involves identification of several characteristic features of the observed streamflow hydrograph, each representing a distinct (preferably unique) aspect of the behavior of the watershed. The "closeness" of the model output to the observed data for each of these features is measured objectively by means of a mathematical criterion.

There have been several recent reports on the exploration of hybrid methodologies that emulate manual procedures through stepwise application of optimization techniques. Sugawara et al. (1984) used a rule-based procedure to identify subperiods in the simulated hydrograph for automated adjustment of parameter subgroups of the Tank model. Brazil (1988) used an interactive procedure to perform Level One estimates of parameter values of the SAC-SMA model, followed by automated global random search and iterative parameter optimization steps. A post-simulation multi-criteria evaluation (based on up to 10 criteria) was used to obtain Level Two parameter estimates. Harlin (1991) and later Zhang and Lindstrom (1997) developed automated approaches that emulate the manual calibration procedures for the HVB model (Bergstrom, 1976). Their procedures involve partitioning the runoff time series into several subperiods associated with specific, dominant hydrologic processes (e.g., rain flood, snowmelt flood, and baseflow). Each subperiod is used to calibrate a different parameter (or subgroup). An iterative automated procedure cycles through the different periods to search for an "optimal" parameter set.

While the previous methods have been shown to be effective, they avoid dealing explicitly with the effects of interactions among the model parameters, and in particular provide little insight into the performance trade-offs associated with model structural errors. In addition, they are largely model-dependent and must be adapted for use with different hydrologic models. The automatic multi-criteria calibration strategy, however, provides a more versatile approach to emulating the evaluation procedures used in the manual approach to Level Two calibration, while taking advantage of the power and efficiency of an optimization algorithm to search the parameter space. In this approach, a multi-criteria optimization problem is defined in terms of simultaneous minimization of all the model performance criteria representing the "closeness" of the model output to the observed streamflow hydrograph. A multi-criteria optimization methodology, such as the Multi-Objective COMplex evolution (MOCOM-UA) algorithm [Yapo et al., 1998], is used to search for the Pareto solution space containing the "good" solutions to the problem. Finally, one or more "acceptable" parameter estimates are selected from within the Pareto solution space by rejecting solutions with unacceptable trade-offs and/or poor overall (long-term) statistical characteristics (e.g., annual, monthly, and flow group biases).

The MOCOM-UA is a general purpose global multi-objective optimization algorithm that provides an effective and efficient estimate of the Pareto space with a single optimization run and is based on an extension of the SCE-UA population evolution method [Duan et al., 1993]. A detailed description and explanation of the method are given in Yapo et al. [1997, 1998] and so will not be repeated here. In brief, the MOCOM-UA method involves the initial selection of a population of p points distributed randomly throughout the s-dimensional feasible parameter space. In the absence of prior information about the location of the Pareto optimum a uniform sampling distribution is used. For each point the multi-objective vector F is computed, and the population is ranked and sorted using a Parteo-ranking procedure suggested by Goldberg [1989]. Simplexes of s + 1 points are then selected from the population according to a robust rank-based selection method [Whitley, 1989]. A multi-objective extension of the downhill simplex method is used to evolve each simplex in a multi-objective improvement direction. Iterative application of the ranking and evolution procedures causes the entire population to converge toward the Pareto optimum. The procedure terminates automatically when all points in the population become nondominated. Experiments conducted using standard synthetic multi-objective test problems have shown that the final population provides a fairly uniform approximation of the Pareto solution space [Yapo et al., 1997, 1998]."

In this chapter, an automated multi-criteria scheme that emulates the manual approach with a simple two-step procedure is investigated. In the first step, the hydrograph is partitioned into three components based on the reasonable assumption that the behavior of the watershed is inherently different during periods "driven" by rainfall and periods without rain. Further, the periods immediately following the cessation of rainfall and dominated by interflow can be distinguished from the later periods that are dominated by baseflow. The streamflow hydrograph can, therefore, be partitioned into three components (Figure 2.2), which we call "driven" (QD), "non-driven-quick" (QQ), and "non-driven-slow" (QS). The time steps corresponding to each of these components are identified through an analysis of the precipitation data and the time of concentration for the watershed. The time steps with non-zero rainfalls, lagged by the time of concentration for the watershed, are classified as driven. Of the remaining (non-driven) time steps, those with streamflows lower than a certain threshold value (e.g., mean of the logarithms of the flows) are classified as non-driven-slow, and the rest are classified as non-driven-quick. In this chapter, a simple empirical method is used for partitioning the non-driven flows; note that there are a wide range of different techniques based on physical reasoning [e.g., Rutledge, 1993; Nathan and McMahon, 1990; among others] available for base flow separation that could be used to improve the partitioning of the non-driven component. For each of the components, the closeness between the model outputs and the corresponding observed values is estimated separately with one or more statistical criteria. Finally, MOCOM-UA is used to search for the Pareto solution space containing the "good" solutions to the problem.

In the second step, one or more "acceptable" parameter estimates are selected from within the Pareto solution space by rejecting solutions with unacceptable trade-offs and/or poor overall (long-term) statistical characteristics (e.g., annual, monthly, and flow group biases). This approach is illustrated in the next section with a case study involving parameter estimation for the SAC-SMA model.

2.4 CASE STUDY

2.4.1 Introduction

In this study, the multi-criteria approach outlined above was used to estimate values for the parameters of the SAC-SMA flood forecast model using historical data from the Leaf River watershed (1950 km^2) located north of Collins, Mississippi. Forty consecutive years of data (WY 1948-88) are available for this watershed, representing a wide variety of hydrologic conditions. The details of the SAC-SMA model and the Leaf River data have been discussed previously in the literature [e.g., Burnash et al., 1973; Peck, 1976; Brazil and Hudlow, 1981; Sorooshian and Gupta, 1983; etc.]. The Level Zero and Level One estimates of parameter uncertainty (Table 2.1), the unit hydrograph ordinates, and potential evapotranspiration demand curve for this watershed have been determined previously by Brazil [1988]. Of the 17 model parameters, only 13 are usually estimated via Level Two parameter estimation procedures. An 11-year period (WY 1952–1962 inclusive) was selected to be used for this purpose.

2.4.2 Level Two Parameter Estimation

As described above, the hydrograph was partitioned into three components -- driven (QD), non-driven-quick (QQ), and non-driven-slow (QS). For each of the components, the simulation error was measured using the RMSE statistic, resulting in three evaluation criteria, designated as FD (driven), FQ (non-driven quick), and FS (non-driven slow) respectively. The Pareto optimal solution space for the three criteria was estimated using the MOCOM-UA multi-criteria optimization algorithm [Yapo et al., 1997]; the algorithm used 15,533 trials and approximately 1 hour of computer time (on a Sun workstation) to converge to a set of 500 Pareto optimal solutions.

The results of the multi-criteria automatic calibration run are shown in Figures 2.3 and 2.4. Figures 2.3a-c present two-dimensional projections of the three-criterion trade-off surface represented by the 500 Pareto optimal parameter sets (indicated by the light gray dots). These figures clearly illustrate the inability of the model to simultaneously match all three aspects of the hydrograph. For example, Figure 2.3b indicates a smoothly varying trade-off between the models' ability to match the driven (QD) and the non-driven-slow (QS) portions of the hydrograph (similarly see Figure 2.3c). Figure 2.3a, however, shows that the FD and FQ criteria are very highly correlated, indicating that these two portions of the hydrograph contain very similar information about the parameters of this model, and that one of these criteria can be considered to be redundant in this case study. As a result, equivalent solutions can be expected (for this model structure and data set) using only two criteria, either a combination of FD and FQ with FS, or FD with FS, or FQ with FS. The behavior of the model within FD and FQ is strongly influenced by the shape of the unit hydrograph, which was fixed (not optimized) in this case study. Optimization of the unit hydrograph shape, or replacement and optimization of a different routing model, may increase the models ability to fit the FD and FQ components and, as a result, reduce some of the correlation. Note also that the solutions corresponding to the best fits to the QD, QQ, and QS portions of the hydrograph (i.e., solutions corresponding to minimal FD, FQ and FS) correspond to the extreme points of the set of light gray dots on each plot.

Figures 2.3a-c also reveal that the trade-offs in fitting the three hydrograph components are quite significant. For example, the RMSE error in fitting the slow-flow component QS increases from 2.5 cms to 4 cms as the RMSE error in fitting the driven (rising limb) component QD decreases from 25 cms to 22 cms. The 1.5 cms variation in QS error is a fairly large fraction (33%) of the average slow-flow level (4.5 cms), while the 3 cms variation in QD error is only 9% of the average rising limb flow level (35 cms).

The normalized parameter plot presented in Figure 2.3d shows the variability in the parameter values for the 500 Pareto optimal solutions (indicated by the light gray lines). Each line across the graph represents one of the parameter sets. The maximum range for each parameter is the parameter uncertainty remaining after the Level One analysis (Table 2.1). Notice that the parameter uncertainty has been reduced significantly

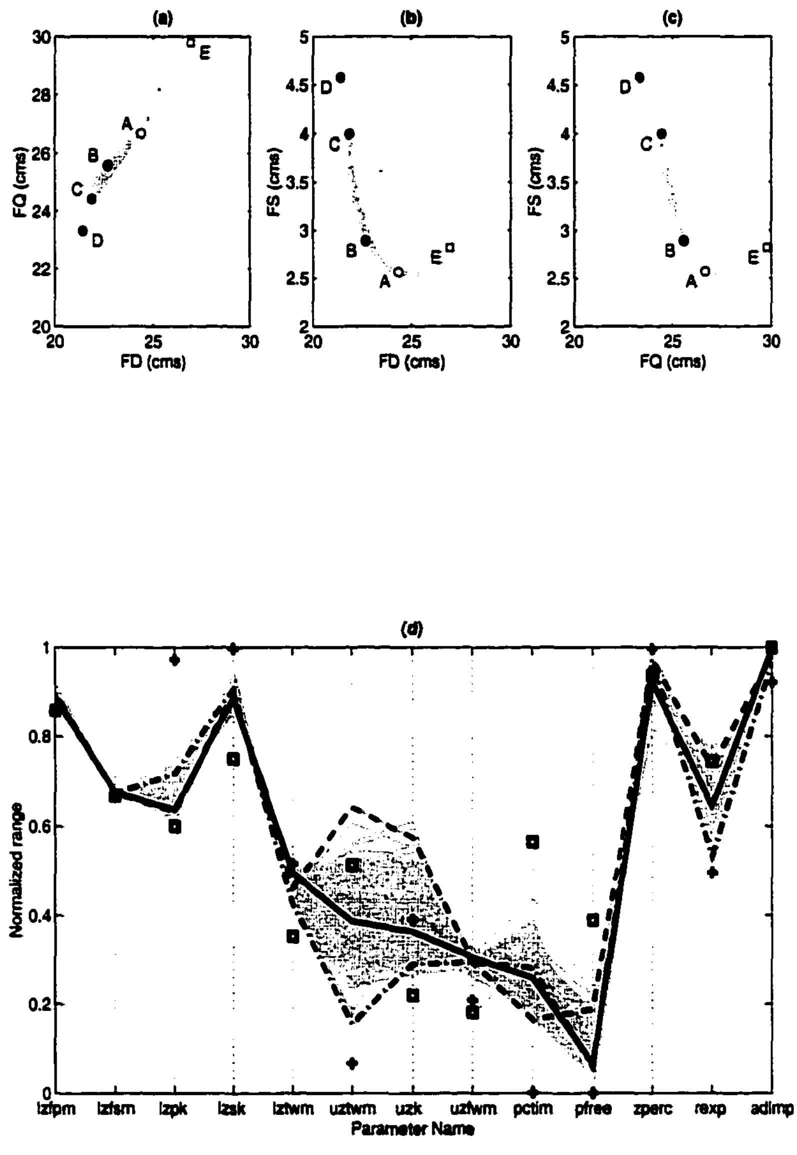

Figure 2.3 Pareto solutions obtained with the automatic multi-criteria approach to calibrate the SAC-SMA model: (a-c) two-dimensional projections of objective space, and (d) normalized parameter space. Marked points correspond to, respectively, 500 Pareto solutions (gray points), extreme points in RMSE and bias trade-off analysis (A and C), "best" point selected from trade-off analysis (B), SCE-RMSE solution (D), and SCE-Log-RMSE solution (E). Lines correspond to, respectively, 500 Pareto solutions (gray solid), solution A (dashed black), solution B (solid black), and solution C (dash-dot black)

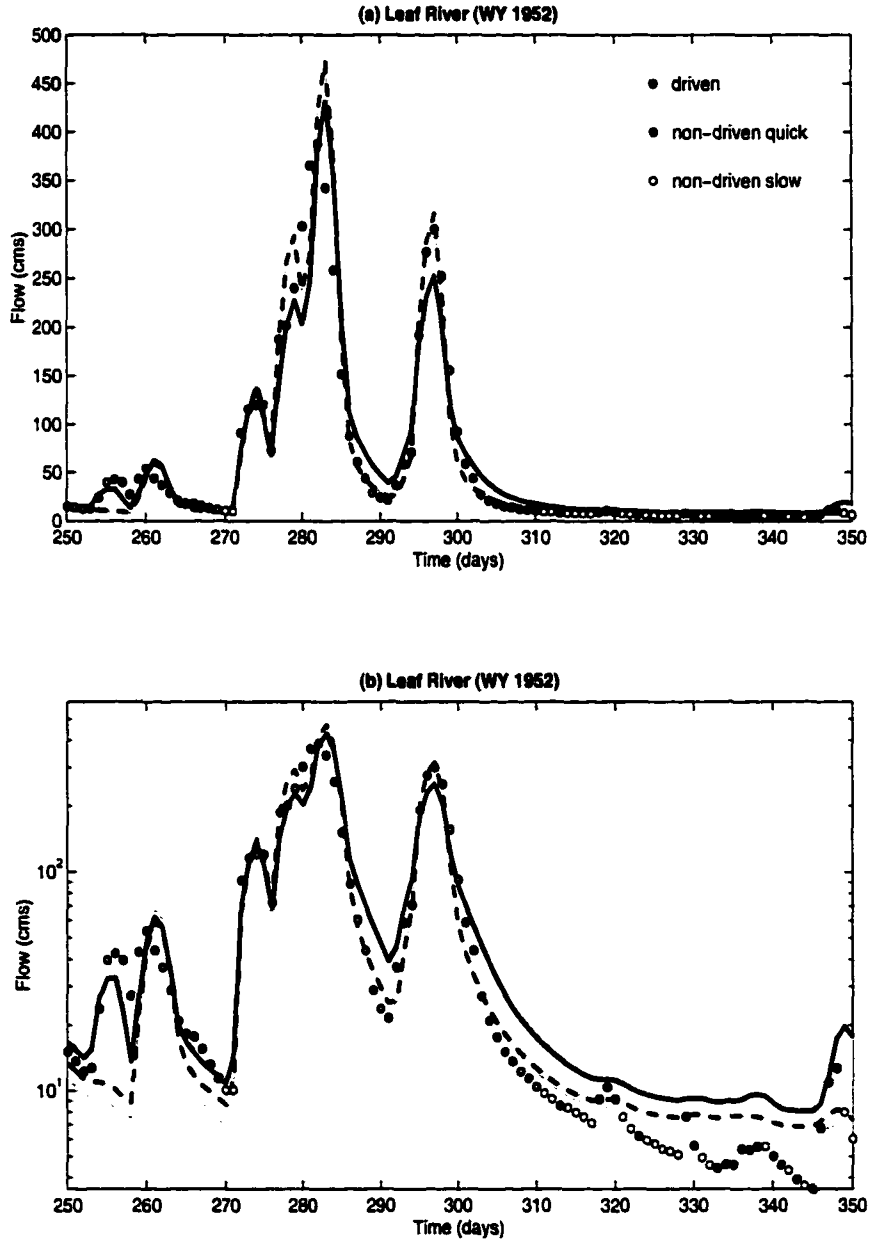

Figure 2.4 Hydrograph range associated with the Pareto solution set for: (a) arithmetic scale and (b) logarithmic scale. The filled circles correspond to the bserved streamflow data, the light-gray region corresponds to the 500 Pareto solutions, the solid line corresponds to the minimal FD solution, and the dashed line corresponds to the minimal FS solution

by the multi-criteria optimization. In particular, the parameters lzfpm and lzfsm show virtually no sensitivity to the trade-off analysis, indicating that these values are very precisely determined. In addition, the parameters lzpk and lzsk show only small amounts of variability. These results provide strong support for the claim by Peck [1976] that estimates for these 4 lower zone parameters (which are primarily associated with the shape of the slow-flow component of the hydrograph) can be easily obtained from the observed data (see Section 2.2.2).

In general, there is larger variability in the estimates for pctim, uztwm, uzk, zperc, rexp, and pfree. These parameters are associated with partitioning of the hyetograph into quick-flow (overland and interflow) and infiltration (called percolation in the SAC-SMA model). The variability in these parameters is consistent with the intuitive notion that the structural error in the model is associated primarily with difficulties in correctly modeling the non-linear and spatially variable infiltration processes at the surface.

Taken together, these results suggest (not surprisingly) that the lumped conceptual representation of lower zone recession processes in the SAC-SMA model is actually quite adequate, and that further attempts to improve the model would be most productive if focused on upper zone processes. These results also imply that the long sequences of systematic errors commonly observed in the slow-flow portions of the simulated hydrograph are caused primarily by incorrect estimation of the volume and timing of percolation.

Figure 2.4a presents the results in the hydrograph space for a 100-day portion of the calibration period. Figure 2.4b shows the same information in the log-transformed flow space so the slow-flow behavior can be clearly seen. The filled circles correspond to the observed data, the light-gray region corresponds to the hydrograph trade-off uncertainty associated with the 500 Pareto optimal solutions, the solid line corresponds to the minimal FD (driven) solution, and the dashed line corresponds to the minimal FS (non-driven-slow) solution. Notice that the minimal FD solution tends to fit the peaks better at the expense of overestimating the recessions, while the minimal FS solution fits the recessions better at the expense of overestimating the peaks (the minimal FQ solution is almost identical to the minimal FD solution and is not plotted). More interesting, however, is that the uncertainty region as a whole tends to overestimate the hydrograph recessions, again supporting the inference that the percolation estimates are incorrect (biased).

2.4.3 Selecting a "Best" Parameter Set

The analysis presented above illustrates how the automatic multi-criteria approach generates a set of Pareto optimal solutions which provide useful information about the characteristic behaviors of the model and its strengths and limitations. However, to use the model for on-line streamflow forecasting, it is desirable to select a single representative parameter set that provides an acceptable trade-off in fitting of the different parts of the hydrograph. For this purpose, we draw again on the standard methods employed in manual calibration by NWS hydrologists. Two important considerations are that the long-term bias of the model simulations is as close to zero as possible (i.e., the model preserves the observed water balance) and the overall residual variance is relatively small. A third important consideration is that the bias by flow level also tends to be close to zero. Figure 2.5 shows a bi-criterion plot of the overall 11-year percent bias against residual variance for the 500 potential solutions generated by the multi-criteria optimization run. The pattern of the points clearly indicates a trade-off between the bias and variance -- the points with minimal variance tend to have strong positive bias, whereas the points having close to zero bias have somewhat larger error variances. This plot shows that the classical approach, which uses a single overall RMSE criterion for model calibration, may give small error variances but at the expense of significant model bias (leading to parameter estimates that "fit the data" but are unacceptable to manual calibration experts). Lindstrom (1997) also acknowledged this fact and proposed a single-criterion approach to correct for it based on a weighted combination of variance and bias measures.

Figure 2.5 also shows that only 24 of the original 500 points fall on the Pareto frontier (i.e., these 24 points are superior to the remaining 476 points in a multi-criteria sense). The three points marked A, B, and C are of special interest -- A and C represent the extreme points in the trade-off analysis, while B has the smallest overall bias (essentially zero). Figure 2.6 shows the percent bias by flow group for each of the 500 solutions, with points A, B, and C highlighted. Notice that the minimum variance point C (dash-dot line) tends to severely overestimate the low and medium flow portions of the hydrograph, while point A (dashed line) tends to underestimate the low flow region. However, point B (solid line) seems to be relatively balanced across the full range of flows. Based on this analysis, the point B would be selected as the parameter set to be used for streamflow forecasting.

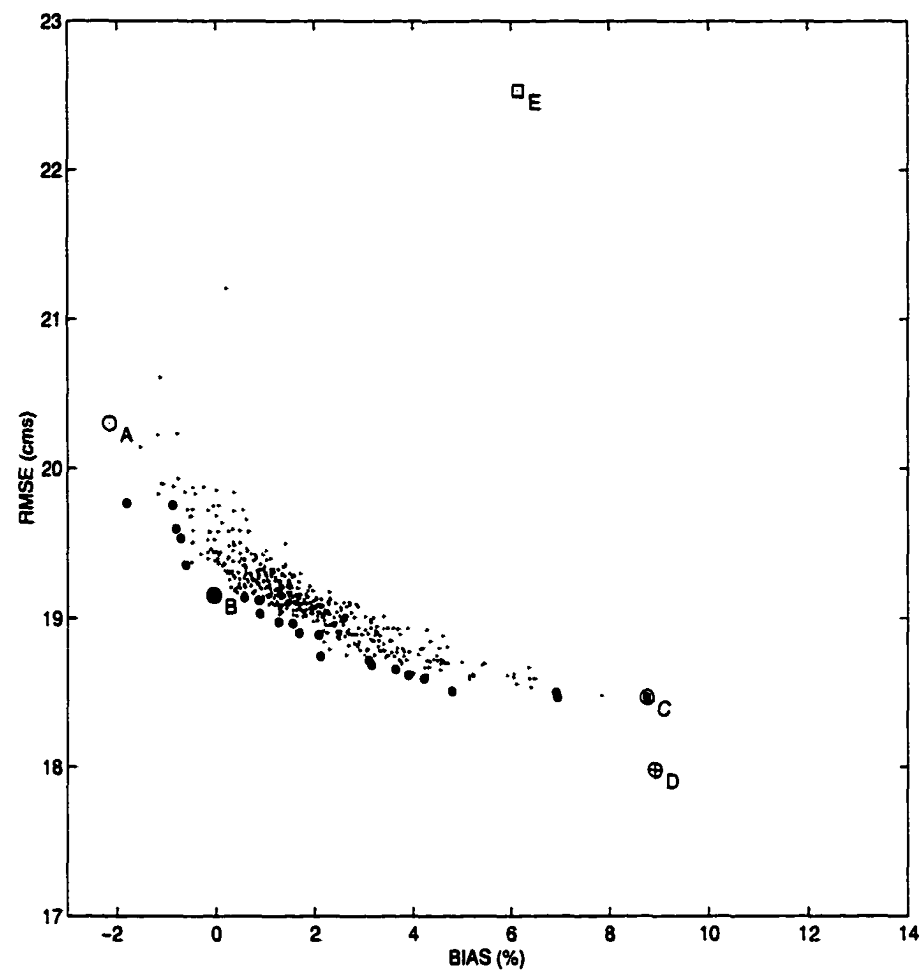

Figure 2.5 Trade-off analysis of 500 Pareto solutions evaluated over entire calibration period. Marked points correspond to, respectively, 24 Pareto frontier solutions (solid black points), extreme points in trade-off analysis (A and C), "best" point selected from trade-off analysis (B), SCE-RMSE solution (D), and SCE-Log-RMSE solution (E)

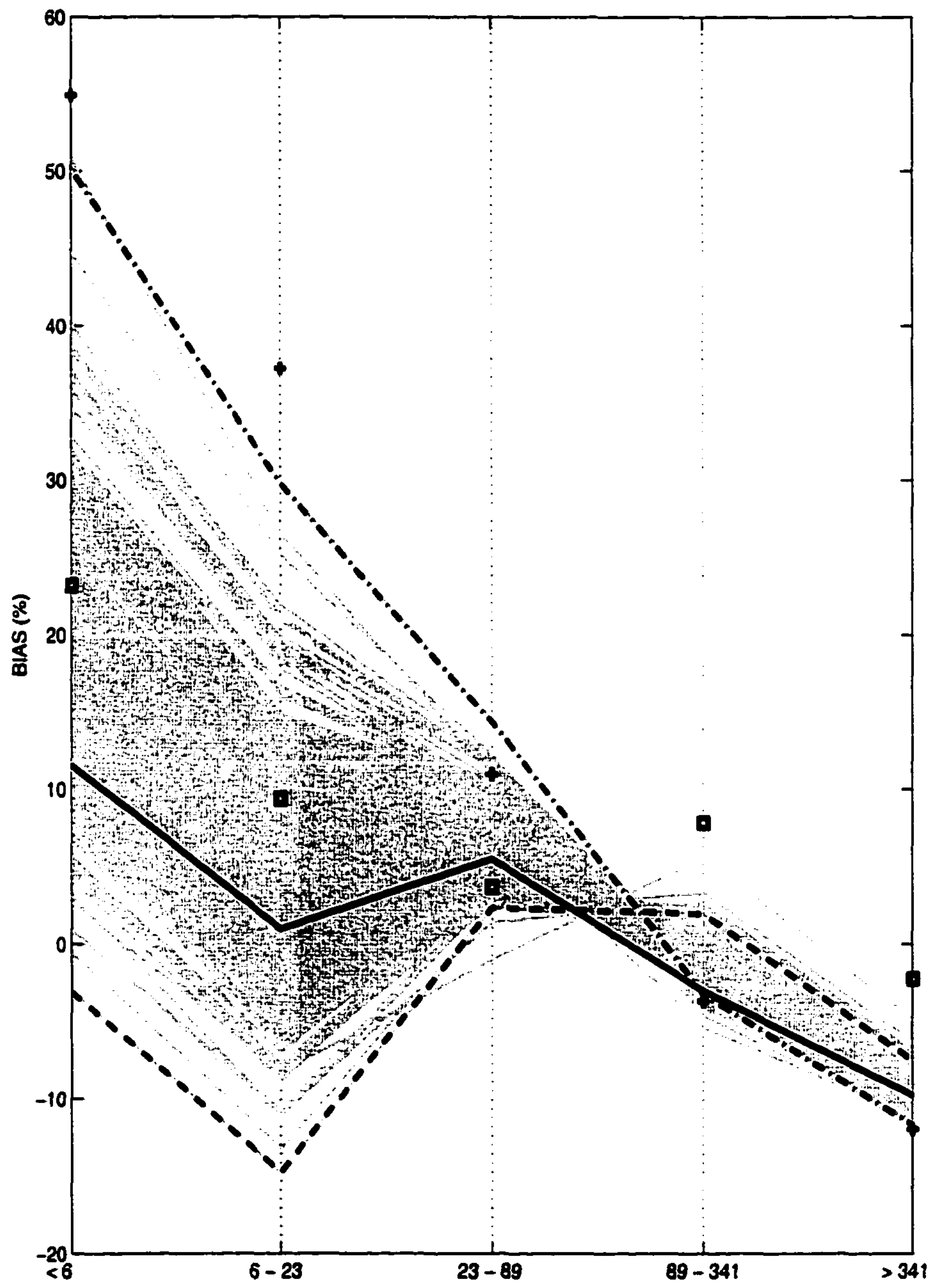

Figure 2.6 Evaluation of model performance over calibration period in terms of flow groups. Marked points and lines correspond to, respectively, 500 Pareto solutions (gray lines), multi-criteria extreme point solution A (dashed black line), best multi-criteria solution B (solid black line), multi-criteria extreme point solution C (dash-dot black), SCE-RMSE solution D (plus sign), and SCE-Log-RMSE solution E (squares)

The relative locations of the points A, B, and C in the original multi-criteria (FD, FQ, FS) space are shown in Figure 2.3. Points A and C fall near opposite ends of the trade-off surface. However, point B, selected to have minimal overall bias, seems to represent a reasonable compromise between matching of the three criteria (FD(B) = 22.7, FQ(B) = 25.6, FS(B) = 2.9). For completeness, Figure 2.3d shows the relative position of the three points in the parameter space.

2.4.4 Comparison with Conventional Single-Criterion Calibration

To demonstrate the advantages of the Level Two multi-criteria approach, we show a comparison with the results obtained by conventional single-criterion automatic calibration (e.g., Sorooshian and Gupta, 1993). The SAC-SMA model was calibrated by separately fitting to the RMSE criterion and the Log-RMSE criterion (RMSE evaluated after log-transformation of the flows) using the SCE-UA global optimization algorithm (Duan et al., 1993). The SCE-RMSE and SCE-Log-RMSE calibration results have been plotted as points D (plus symbol) and E (square symbol), respectively, in Figures 2.3, 2.5, and 2.6 for easy comparison with the multi-criteria results. From Figures 2.3a-c and Figure 2.6, we see (as expected) that the SCE-RMSE solution is biased towards fitting the high flow portions of the hydrograph (particularly the non-driven-quick (FQ) portion) at the expense of poor fitting of the non-driven-slow portion, while the SCE-Log-RMSE solution is biased towards fitting the low-flow portions of the hydrograph at the expense of poor fitting of the driven and non-driven-quick portions. In general, Figures 2.3, 2.5, and 2.6 show that the SCE-RMSE (point D) parameter values are similar (and provide similar performance) to those for point C (Pareto set solution having minimum overall variance) and tend to provide hydrographs with larger overall bias.

To verify the consistency and reliability of the results, the performance of the multi-criteria estimate B and the single-criterion estimates D and E were evaluated over a 36 water-year period of the available data. Table 2.2 shows that the overall percent bias and variance statistics for the parameter estimates change (increase) in a similar way from calibration to evaluation period. However, the multi-criteria estimate B provides smaller overall bias. Figure 2.7 shows the percent bias and variance statistics broken out for each of the 36 individual water years. The three columns show the statistics plotted against the mean annual flow (as a measure of "wetness") for the three parameter sets B (Figure 2.7a-b), D (Figure 2.7c-d), and E (Figure 2.7e-f). Figure 2.8 shows the frequency plots for each statistic. On the average, the percent bias for wet years tends to be similar for all three parameter sets, while the percent bias for dry and average years tends to be larger (more positively biased) for the single-criterion estimate D (Figures 2.7a-c). This difference in the statistical tendencies can be seen clearly in the percent bias frequency plot (Figure 2.8). Further, the annual RMSE statistic tends to increase with annual wetness for all three parameter sets (Figures 2.7d-e), but the statistical tendency is similar (on average) for the B and D parameter sets and significantly better than the E parameter set. In general, we can conclude that the multi-criteria estimate B provides statistically similar distributions of annual error variance as estimate D but with smaller overall bias.

Table 2.2: Overall statistics for selected parameter sets

	% Bias		RMSE	
	Calibration	Evaluation	Calibration	Evaluation
Multi-Criteria (B)	-0.05	6.5	19.2	19.7
Single Criterion (D)	8.9	14.4	18.0	18.8
Single Criterion (E)	6.1	11.7	22.5	24.6

2.4.5 Comments about "Equifinality"

It has been suggested (e.g., see Beven and Binley, 1992) that many models are overparameterized resulting in "equifinality" of model performance associated with widely different values for the model parameters (where we understand equifinality to mean essentially indistinguishable model behaviors). Figure 2.9 shows a frequency plot of the overall calibration period RMSE for all 500 parameter estimates belonging to the Pareto solution set. The plot shows that about 90% of the points have very similar overall RMSE values, seemingly supporting the argument for equifinality. However, Figures 2.3a-c clearly show that these same points do not appear to be similar when examined in terms of their abilities to simulate different portions of the hydrograph, clearly indicating that the points cannot be considered to be equifinal.

It can be argued, therefore, that any single overall statistic, which aggregates model performance errors over a large range of hydrologic behaviors, is a relatively weak test of model performance. One cannot, therefore, conclude equifinality by recourse to such a test. Furthermore, one should be particularly careful not to infer erroneous conclusions about parameter identifiability without recourse to examination of a number of different measures, each emphasizing a different important aspect of model behavior.

2.5. CONCLUSIONS

There is an increase in the level of effort involved in obtaining progressively more refined estimates for the parameters of hydrologic models. Level Two parameter estimates (that account for parameter interactions and model performance trade-offs) are

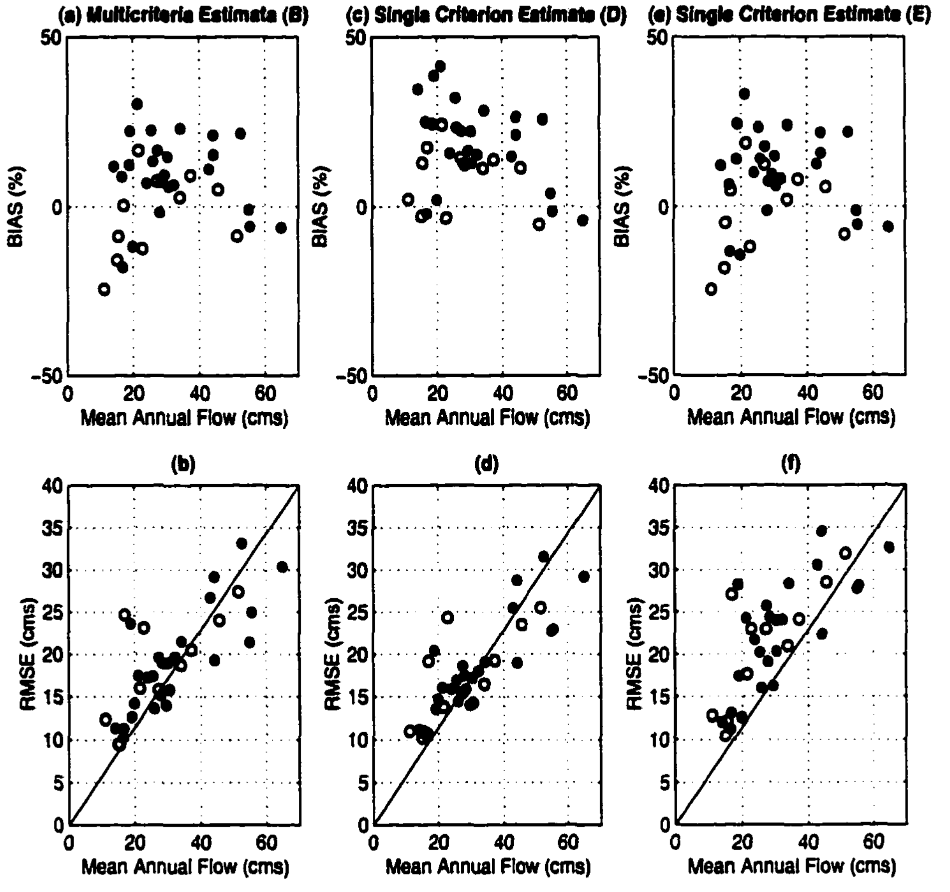

Figure 2.7 Evaluation of model performance over each of the 36 water years: (a-b) best multi-criteria solution B, (c-d) SCE-RMSE solution D, and (e-f) SCE-Log-RMSE solution E. Marked points correspond to, respectively, years used for calibration (white-filled circles) and years not used for calibration (black-filled circles)

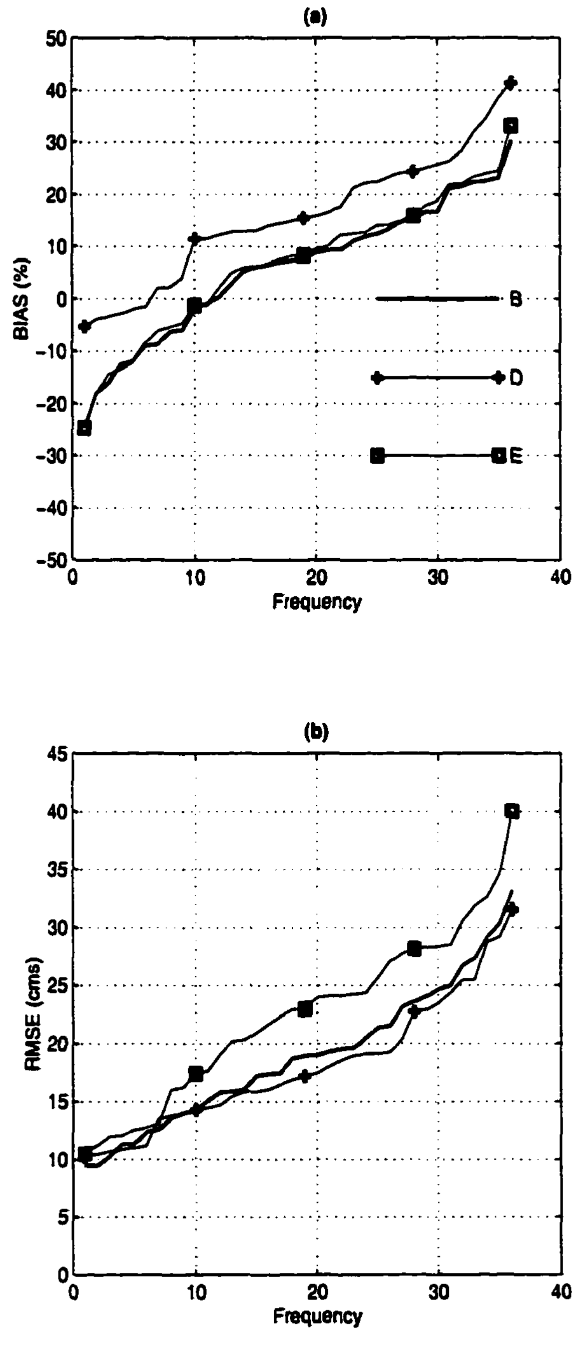

Figure 2.8 Frequency plots for each statistic shown in Figure 7. Lines correspond to, respectively, best multi-criteria solution B (heavy black line), SCE-RMSE solution D (light black line with plus symbols), and SCE-Log-RMSE solution E (light black line with squares)

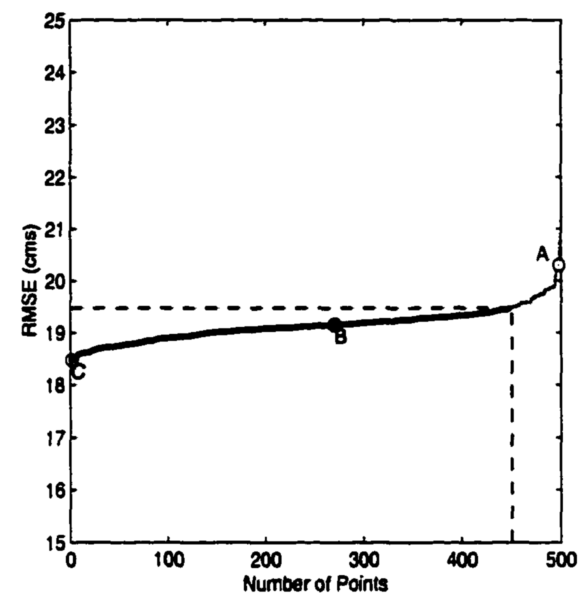

Figure 2.9 Frequency plot of RMSE over calibration period for all 500 Pareto solutions. Marked points correspond to, respectively, extreme points in trade-off analysis (solutions A and C) and the "best" multi-criteria solution B. The dashed line highlights the 90% of the 500 Pareto solutions which have very similar RMSE values

particularly difficult to obtain by manual methods, requiring extensive training as well as considerable experience with the specific model. This expertise is difficult to transfer from person to person and model to model.

This chapter explores the relationship between the manual and automatic procedures for model calibration and presents a two-step automatic multi-criteria approach for obtaining Level Two parameter estimates. The approach combines the strengths of manual and automatic methods by emulating the evaluation techniques and strategies used in manual calibration. By solving the resulting multi-criteria optimization problem with a computerized global search procedure algorithm, the time and effort required to estimate the parameter range representing the trade-offs in the performance of the model are dramatically reduced. As a result, the attention of the hydrologist can be redirected from the tedious effort of manually searching for the "good" region to the more productive task of evaluating solutions from within the region found with the use of the automatic search algorithm. Finally, a simple strategy, modeled on manual procedures, is suggested as an aid to the hydrologist in analysis and elimination of the solutions within the trade-off range.

A case study using the SAC-SMA model was used to demonstrate the performance of the new approach. The results indicate that parameter sets selected from within the Pareto region tend to provide consistent and reliable model forecasts. Further, the properties of the Pareto region provide information useful for evaluating the limitations of the various components of the watershed model, thereby pointing towards potential structural improvements. The approach provides a stronger test of model performance than methods that use a single overall statistic to aggregate model errors over a large range of hydrologic behaviors.

CHAPTER THREE

INCORPORATING SPATIAL INFORMATION

3.1 INTRODUCTION, MOTIVATION, AND SCOPE

Conceptual Rainfall-Runoff (CRR) models are often difficult to calibrate due to the large number of functional parameters and complex relationships. Systematic manual calibration techniques such as those developed by the U.S. National Weather Service (NWS) for calibration of the Sacramento Soil Moisture Accounting (SAC-SMA) model can result in good model calibrations but are complicated and highly labor-intensive. Traditional automatic calibration procedures take advantage of the speed and power of digital computers, while being objective and relatively easy to implement. However, they do not provide parameter estimates and hydrograph simulations that are considered acceptable by the hydrologists responsible for operational forecasting and have therefore not entered into widespread use. Gupta et al. [1998] acknowledged some of the more serious limitations with the classical automatic approach and presented a more general multi-criteria framework for model calibration that recognizes the multi-objective nature of the problem. Yapo et al. [1998] developed the MOCOM-UA algorithm, an effective and efficient methodology for solving the multi-objective global optimization problem. In Chapter 2, an analysis of the similarities and differences between the automatic and manual calibration approaches was performed and a new hybrid multi-criteria approach that combines the strengths of each was developed. In Chapter 2, it was demonstrated that the new approach could be used to emulate some of the important aspects of the manual approach pursued by an expert. Further, the approach provides information that is useful for evaluating the limitations of the various structural components of the model, thereby pointing toward potential structural improvements.

The majority of CRR model calibration studies have typically been concerned with "lumped" applications (i.e., averaging of the dominant subwatershed scale processes that contribute to the overall watershed scale response). However, remotely sensed, high-resolution hydrologic information (e.g., NEXRAD Stage III precipitation data [Klazura, G.E., and Imy, D.A., 1993]) is now becoming widely available (at least in Europe and the U.S.), and hydrologists have begun to incorporate this information into their modeling procedures [e.g., Smith et al., 1999; Michaud and Sorooshian, 1994; Ogden and Julien, 1993, 1994; Winchell et al., 1998; Krajewski, et al., 1991; among others]. This has led to the development of relatively complex "fully-distributed" models that allow the user to construct a very detailed representation of the spatial variability of the hydrologic processes within the watershed (e.g., SHE [abbott et al., 1986a], among others.). The main premise is that the spatial detail will lead to an improved understanding of the watershed behavior while improving simulations. However, the degree to which the spatial variability of each process needs to be represented (to provide the maximum improvement in the simulations) is not well understood. Further, these models typically have a very large number of parameters for which values must be estimated, either through a calibration procedure or from maps of watershed properties.

In the case of streamflow (flood) forecasting, the primary concern is to generate estimates of discharge at a limited number of fixed points along the river channel, and there is typically little need for spatially detailed information about the various internal states of the watershed. An attractive alternative to the lumped and fully distributed approaches is the so-called "semi-distributed" approach in which the watershed is conceptualized as a network of functionally distinguishable land segments (sub-watersheds), each represented by a lumped CRR model. The runoff response from each segment is routed to the outlet of the watershed to compute the total watershed response. In this approach, the spatial variability of the hydrologic processes within the watershed is represented through the number and location of segments, which determine the degree of spatial distribution of the model input (precipitation), structural components (soil moisture and streamflow routing computations), and surface characteristics (parameters).

The added complexity in going from a lumped to a semi-distributed representation of a watershed results in a significantly more complex model calibration problem. CRR models tend to use complex, highly parameterized, vertical representations (of the movement of moisture through the soil), and the overall number of parameters to be estimated for each unit can be quite large. Further, many of the parameters may not be supported (identifiable) by the information contained within the observed data, remotely sensed or otherwise [Jakeman and Hornberger, 1993; Wagener et al., 1999; Wheater et al., 1993; among others]. As the number of hydrologic segments is increased, the calibration procedure (manual or automatic) can quickly become intractable. Operational use of semi-distributed rainfall-runoff model applications to produce flood forecasts has therefore been limited.

The focus of this chapter is to provide an assessment of the potential improvements in rainfall-runoff model performance that can be achieved by semi-distributed modeling of a watershed using radar-based (NEXRAD) remotely sensed precipitation data. The relative benefits of spatially distributing the model input (precipitation), structural components (soil moisture and streamflow routing), and surface characteristics (parameters) are examined. The CRR model used is the NWS Sacramento Soil Moisture Accounting (SAC-SMA) model [Burnash et al., 1973]. The multi-criteria framework developed in Chapter 2 for application to lumped hydrologic models is used to calibrate the semi-distributed model in terms of three objective measures designed to reflect the different observable characteristics of watershed behavior (peak flow and timing, quick recession, and baseflow). Multi-criteria performance comparisons among the different model applications are used to evaluate the benefits of various types and degrees of spatial complexity. Results from an independent NWS manual (expert) calibration study are used as a basis to evaluate the approach.

This chapter is organized as follows: the background and context for the work is discussed in Chapter 3.2. The theoretical and practical basis for applying the multi-criteria methodology to investigate spatially distributed modeling approach on a study watershed is presented in Chapter 3.3. The results of the model applications are presented in Chapter 3.4, and Chapter 3.5 discusses the results and future extensions of the study results.

3.2 BACKGROUND

Since the early 1990s, the NWS has been investigating methods to assess the benefit (improvement in streamflow forecasts) of representing the spatial variability of the high-resolution precipitation, soil, and vegetation properties within the NWS modeling system [Smith et al., 1999, 2000]. One primary focus has been the incorporation of high-resolution, remotely sensed, Next Generation Radar (NEXRAD) Stage III precipitation data into standard NWS procedures for flood forecasting using the SAC-SMA model (SAC-SMA model described in Chapter 2).

In a recent study, the NWS applied the SAC-SMA model to five different watersheds in both lumped and semi-distributed applications and compared the resulting streamflow simulations to observed data [Smith et al., 1999, 2000; Zhang et al., 2000]. The results indicated that for four of the watersheds characterized as having a deep soil layer (more than 150 cm) the accuracy of the semi-distributed simulations "did not show significant improvement in accuracy" when compared to the lumped simulations. However, the results for the remaining watershed (Blue River at Blue, OK), characterized as having much shallower soils, "were improved significantly" by use of the semi-distributed model. Based on these findings, the Blue River watershed was selected for the studies reported in this chapter.

3.2.1 Description of Blue River Watershed and Data

The Blue River watershed is located in southern Oklahoma near Blue, Oklahoma (Figure 3.1). The watershed is a long narrow valley with a contributing area of 1227 km^2 1999]. The United States Geological Survey (USGS) surface-water discharge station #07332500, Blue River near Blue, Oklahoma, is located at the outlet of the watershed and has been in operation since 1948.

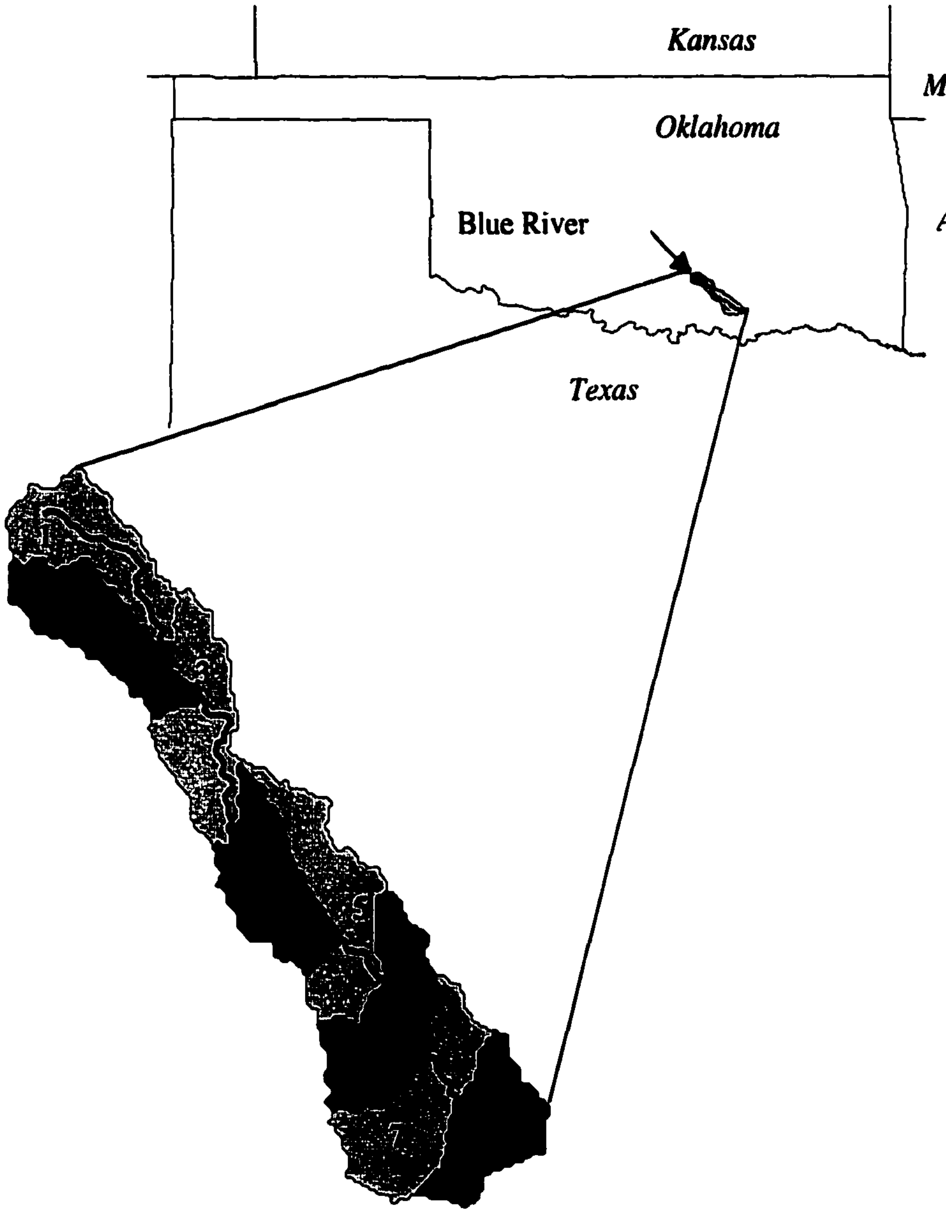

Figure 3.1 Location of Blue River watershed

A NEXRAD stage III precipitation data set for the Blue River watershed was obtained for the period 1 June 1993 to 30 April 1999. The NEXRAD product offers gridded precipitation estimates, spatially averaged over 4 km by 4 km grid cells and temporally averaged over 1 hour. For the same time period, hourly estimates of instantaneous discharge are available from the USGS surface-water discharge station. The NWS-derived hourly values of Potential Evapotranspiration (PET) are based on long-term average values obtained from the atlas of NOAA Free Water Evaporation [NOAA, 1982]. It should be noted that the PET values vary seasonally but not annually (i.e., an estimate of PET for one year has been made and is repeated for each year of record).

The NWS created a digital elevation model (DEM) of the Blue River watershed from 100 meter (cell size) elevation data. The watershed was partitioned into eight subwatersheds (Figure 3.1) based on an analysis of DEM stream connectivity data (stream channel structure), and the variability of the high-resolution soil property information available from the USDA State Soil Geographic Database (STATSGO) for the resulting subwatersheds [Victor Koren, Hydrologic Lab, NWS, personal communication, 2000]. The average soil depth determined from the STATSGO information and the contributing area of each subwatershed are listed in Table 3.1.

Mean Areal Precipitation (MAP) values for each of the eight subwatersheds were estimated from the 4 x 4 km NEXRAD Stage III hourly precipitation data. Unit hydrographs for each subwatershed were developed in conjunction with the DEM, using

Table 3.1: Subwatershed attributes

Subwatershed	Contributing Area (km^2)	Ave. Soil Depth (m)
1	153.1	1.19
2	150.0	0.88
3	153.1	0.83
4	144.0	1.70
5	162.9	1.71
6	165.0	1.75
7	169.9	1.75
8	129.0	1.75

the methodology described by Smith et al. [1999], to route the simulated channel inflow to the outlet of the watershed. For the lumped conceptualization, the unit hydrograph was derived from the subwatershed unit hydrographs.

3.2.2 NWS Modeling Strategy

The NWS applied the SAC-SMA model in both lumped and semi-distributed (eight subwatersheds) forms to the Blue River watershed. In the lumped case, the channel inflow was computed at each time step for the entire watershed and then routed to the outlet with a single unit hydrograph. In the semi-distributed case, the soil-moisture computations were made separately for each subwatershed, and the resulting simulated channel inflows were then routed independently to the outlet of the watershed and combined to compute the total simulated streamflow for the entire watershed.

The NWS first used a sophisticated, highly interactive manual procedure to estimate values for 13 of the SAC-SMA parameters (four were set to default values) [Anderson, 1997] for the lumped watershed case. In this procedure, initial values for the calibration parameters are estimated from analytical relationships derived by Koren et al. [2000] based on soil property data (STATSGO soil texture data). While these estimated values are not "optimum" values, they are considered (by the NWS) to be "very reasonable initial approximations" [Koren et al., 2000]. Next, a systematic sequence of steps is followed to develop parameter estimates based on an examination of the hydrological database (precipitation, PET, and streamflow) of the watershed. Periods in the observed time series data are identified where specific hydrologic processes are dominant (e.g., baseflow, interflow, surface flow, evaporation, transpiration, abstraction, infiltration, etc). For each of these periods, the closeness of the model and the data is evaluated visually, and values for the relevant parameters are estimated by heuristic methods. In the last (and most difficult) step of the procedure, the hydrologist must deal with the effects of parameter interactions on the model responses by simultaneously evaluating a number of subjective and objective criteria while iteratively adjusting the parameter values so that the model matches the behavior of the watershed system as closely as possible.

For the semi-distributed case, there are 104 parameters to be estimated (13 for each of the eight subwatersheds). For each subwatershed, initial parameter values were estimated from the soils data using the analytical relationships developed by Koren et al. [2000]. Next, the calibration parameter values were adjusted manually while maintaining the ratios between the initial parameter values among the subwatersheds. Finally, a "fine tuning" of the parameter estimates for each subwatershed was done through a complicated and subjective examination of the hydrological database of the watershed and through consideration of the soil-moisture state and simulated hydrograph (at the outlet of the watershed) for each subwatershed. This step requires the hydrologist to search the entire simulation period for consistent deviations between the model and watershed behavior and to make appropriate parameter adjustments (e.g., precipitation events that occur primarily over a single subwatershed can be isolated and the corresponding parameters adjusted to improve the associated streamflow simulations).

Clearly, the large number of model parameters and limited number of isolated events for each subwatershed limit the effectiveness of this approach.

3.3 METHODS

The manual calibration conducted by the NWS was used in this study as the basis for an evaluation of the strengths and weaknesses of the automatic multi-criteria calibration approach [Gupta et al., 1998; Boyle et al., 2000]. In addition, a number of different model runs were conducted to investigate the benefit of different levels of spatial representation of model input (precipitation), structural components (soil moisture and streamflow routing computations), and surface characteristics (parameters) of the SAC-SMA model applied to the Blue River watershed. The study was designed to complement the NWS semi-distributed studies on the Blue River by expanding our understanding of the specific benefits associated with different levels of spatial representation of the model.

3.3.1 Multi-criteria Parameter Estimation Methodology

The multi-criteria approach to calibration presented in detail in Chapter 2 and by Boyle et al. [2000] combines the strengths of both automated and manual calibration methods. The approach involves the identification of several characteristic features of the observed streamflow hydrograph, each representing a distinct (preferably unique) aspect of the behavior of the watershed ("driven" (Q_D), "non-driven-quick" (Q_Q), and "non-driven-slow" (Q_S) – see Chapter 2 for detail). For each of the components, the closeness between the model outputs and the corresponding observed values is estimated separately using the RMSE statistic, resulting in three evaluation criteria, designated as FD (driven), FQ (non-driven quick), and FS (non-driven slow), respectively. The Multi-Objective COMplex evolution (MOCOM-UA) algorithm [Yapo et al., 1998; Bastidas et al., 1999] was used to solve the multi-criteria optimization problem.

3.3.2 Description of this Study

In this study, a series of lumped and semi-distributed applications of the SAC-SMA model to the Blue River watershed was made to investigate the improvements in model performance associated with various levels of spatial representation of model input (precipitation), structural components (soil moisture and streamflow routing computations), and surface characteristics (parameters). Each model application was designed to isolate the effects of the different levels of spatial representation in terms of specific desirable watershed behaviors (driven flow–"peaks and timing", non-driven quick flow–"quick recession" responses, and non-driven slow–"baseflow" responses). The calibration data set (precipitation, PET, and streamflow) used in this study was the same as that used in the NWS study. Model calibration and evaluation of the performance improvements for each application were performed using the multi-criteria approach described above. For each case, the Pareto optimal solution space for the three criteria (FD, FQ, and FS) was estimated by 500 solutions generated using the MOCOM algorithm.

Five separate cases of spatial distribution of the SAC-SMA model were investigated (see Table 3.2). In Case 1 (LUMP-ALL), the SAC-SMA model was applied in a lumped configuration (precipitation P, soil moisture computations S, and streamflow routing computations R, were all lumped) to the Blue River watershed. This case served as a benchmark for performance comparisons with Cases 2-5, in which the SAC-SMA model was applied in varying levels of spatial distribution to the 8-subwatershed configuration used by the NWS. Note that, in Cases 2-4, the model parameters were treated as lumped (all the subwatersheds were assigned the same values of the 13 calibration parameters - see Table 3.3) and only the spatial distribution of the model input and structural components was investigated.

In Case 2 (DIST-PS), the precipitation and soil-moisture computations were spatially distributed among the subwatersheds, but the routing was treated as lumped. In this application, soil-moisture computations were performed separately to compute separate channel inflow sequences for each subwatershed, but these were combined into a total channel inflow for the entire watershed before routing to the outlet of the watershed using a single unit hydrograph. In Case 3 (DIST-PSR), the precipitation, soil-moisture computations, and streamflow routing computations were spatially distributed among the subwatersheds to assess the additional benefit of distributed routing. In this application, the channel inflow computed for each subwatershed was independently routed to the outlet of the watershed with separate unit hydrographs and then combined to estimate the total runoff from the watershed. In Case 4 (DIST-SR), the precipitation was treated as

Table 3.2: Table of different modeling cases

Case #	Case Name	Precip. (P)	Soil moisture. Comp. (S)	Routing Comp. (R)	Params (θ)
1	LUMP-ALL	L	L	L	L
2	DIST-PS	D	D	L	L
3	DIST-PSR	D	D	D	L
4	DIST-SR	L	D	D	L
5	DIST-PSRθ	D	D	D	D

L = Lumped
D = Distributed

Table 3.3: Parameters and state variables of the SAC-SMA model

Parameter	Description	Multi-criteria Calibration Range	NWS Lumped
UZTWM	Upper zone tension water max. storage (mm)	1.0 - 200.0	40
UZFWM	Upper zone free water max. storage (mm)	1.0 - 100.0	37
LZTWM	Lower zone tension water max. storage (mm)	90 - 200.0	160
LZFPM	Lower zone free water primary max. storage (mm)	50 - 400.0	120
LZFSM	Lower zone free water supplemental max. storage (mm)	1.0 - 90.0	75
ADIMP	Additional impervious area (decimal fraction)	0.0 - 0.4	0.0
UZK	Upper zone free water lateral depletion rate (day^{-1})	0.1 - 0.9	0.5
LZPK	Lower zone primary free water depletion rate (day^{-1})	0.0008 - 0.01	0.002
LZSK	Lower zone supplemental free water depletion rate (day^{-1})	0.01 - 0.2	0.03
ZPERC	Maximum percolation rate (dimensionless)	1.0 - 250.0	180
REXP	Exponent of the percolation equation (dimensionless)	1.0 - 3.0	1.8
PCTIM	Impervious fraction of the watershed area (decimal fract.)	0.0 - 0.01	0.008
PFREE	Fraction of water percolating from upper zone directly to lower zone free water storage (decimal fract.)	0.0 - 0.6	0.04
RIVA	Riparian vegetation area (decimal fract.)	0.00	0.00
SIDE	Ratio of deep recharge to channel baseflow (dimensionless)	0.00	0.00
RSERV	Fraction of lower zone free water not transferable to lower zone tension water (decimal fract.)	0.30	0.30
PXMLT	Precipitation multiplication factor (dimensionless)	1.00	1.00
UZTWC	Upper zone tension water storage content (mm)		
UZFWC	Upper zone free water storage content (mm)		
LZTWC	Lower zone tension water storage content (mm)		
LZFPC	Lower zone free primary water storage content (mm)		
LZFSC	Lower zone free secondary water storage content (mm)		
ADIMC	Additional impervious area content (mm)		

lumped over the entire watershed, but the soil moisture and streamflow routing computations computations were spatially distributed among the subwatersheds. This configuration was designed to investigate the value of the spatially distributed precipitation through comparison with Cases 1-3. In Case 5 (DIST-PSRθ), the additional value of spatially distributing the watershed soil properties was investigated by allowing some of the model parameters to vary among the subwatersheds. The five parameters allowed to vary were LZTWM, REXP, UZTWM, UZFWM, AND UZK, selected based on the results of previous studies by the NWS [Koren et al., 2000] indicating that these parameters exhibit emperical relationships with different soil-property data.

Finally, to further investigate the effects of spatial representation, Cases 2-5 were repeated using a smaller number of subwatersheds (i.e., the entire watershed was partitioned into a 3-subwatershed configuration). In this new configuration (also provided to us by the NWS), the original subwatersheds 1, 2, and part of 3 were combined to form the new subwatershed 1 of the 3-subwatershed configuration. Similarly, 4, 5, and parts of 3 and 6 were combined to form the new subwatershed 2, while 7, 8, and part of 6 were combined to form the new subwatershed 3. The mean areal precipitation and PET for each of the three new subwatersheds were estimated by the NWS using the same methods mentioned previously.

3.4 RESULTS

3.4.1 NWS Manual Calibration Results

The NWS manual calibration studies were used as benchmarks for evaluation of the automatic calibration studies described above. The manual calibration results are shown in the multicriteria format in Figure 3.2a-c. Figures 3.2a-c present the results for each case using two-dimensional projections of the three-criteria solution space (NWS lumped case = large open square and NWS semi-distributed case = large open circle). Clearly, the semi-distributed application results in an improvement in the model's ability to simulate the observed flow in terms of FQ and FS, as compared with the lumped application. There is a slight decrease, however, in the model's ability to simulate the driven flows measured by FD.

3.4.2 Automatic Multi-criteria Calibration Results

a) Case 1: Lumped Model

The results of the multi-criteria automatic calibration of the lumped case (LUMP-ALL) are also shown in Figures 3.2a-c, as a three-criteria trade-off surface represented by the set of 500 Pareto optimal solutions (indicated by the light-gray dots). The inability of the model to simultaneously match all three aspects of the hydrograph is clearly illustrated. For example, Figure 3.2b illustrates the smoothly varying trade-off between the model's ability to match the driven (Q_D) and the non-driven-slow (Q_S) portions of the hydrograph (similarly see Figure 3.2c and, to a lesser extent, Figure 3.2a).

A visual comparison of the 500 Pareto solutions with the NWS lumped solution (open square) in Figures 3.2a-c shows that the automatic approach provides a closer fit of the baseflow responses (FS) and, to a lesser extent, the quick recession responses (FQ). In

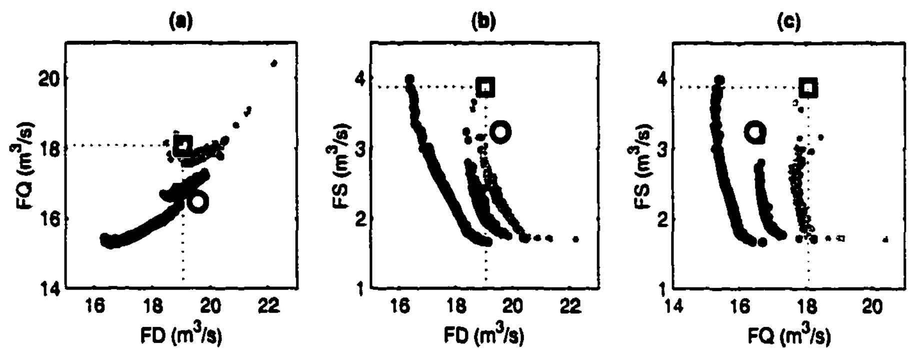

Figure 3.2 Pareto solutions obtained with the automatic multi-criteria approach to calibrate the SAC-SMA model: (a-c) two-dimensional projections of objective space. Marked points correspond to, respectively, 500 Pareto solutions for Case 1 (light-gray points), Case 2 (dark-gray points), and Case 3 (black points). NWS manual calibration results for lumped (square) and semi-distributed (open circle)

terms of the peaks and timing (FD), however, most of the 500 Pareto solutions are inferior to the NWS lumped solution.

The variability in the parameter values across the 500 Pareto optimal solutions for the case LUMP-ALL is shown in Figure 3.3a (gray lines) using a normalized parameter plot. Each line across the graph represents one of the parameter sets. The maximum range for each parameter represents the range over which the multi-criteria calibration procedure was performed (see Table 3.3). Notice that the multi-criteria optimization has resulted in a significant reduction in the parameter range. Notice also that the seven parameters LZFPM, LZFSM, LZPK, LZSK, PCTIM, ZPERC, and REXP show only small amounts of variability while corresponding well with the NWS lumped solution (black line) obtained by an expert via manual calibration. The relative lack of variability in these parameters indicates that they are not primarily responsible for the performance trade-offs associated with the inability of the model to simultaneously match the three different components of the hydrograph (indicated by FD, FQ, and FS). Of these parameters, the first five (LZFPM, LZFSM, LZPK, LZSK, and PCTIM) are considered to be the ones that are "most identifiable" and relatively easy to estimate via careful off-line examination of the observed hydrograph and precipitation data [Peck, 1976]. The fact that the automatic multi-criteria approach gives comparable values for these parameters is a good indication of the reasonableness of the automatic approach. Note, however, that the two parameters, ZPERC and REXP, which control the behavior of the percolation (infiltration) component of the model, are not easy to estimate directly from the observed data.

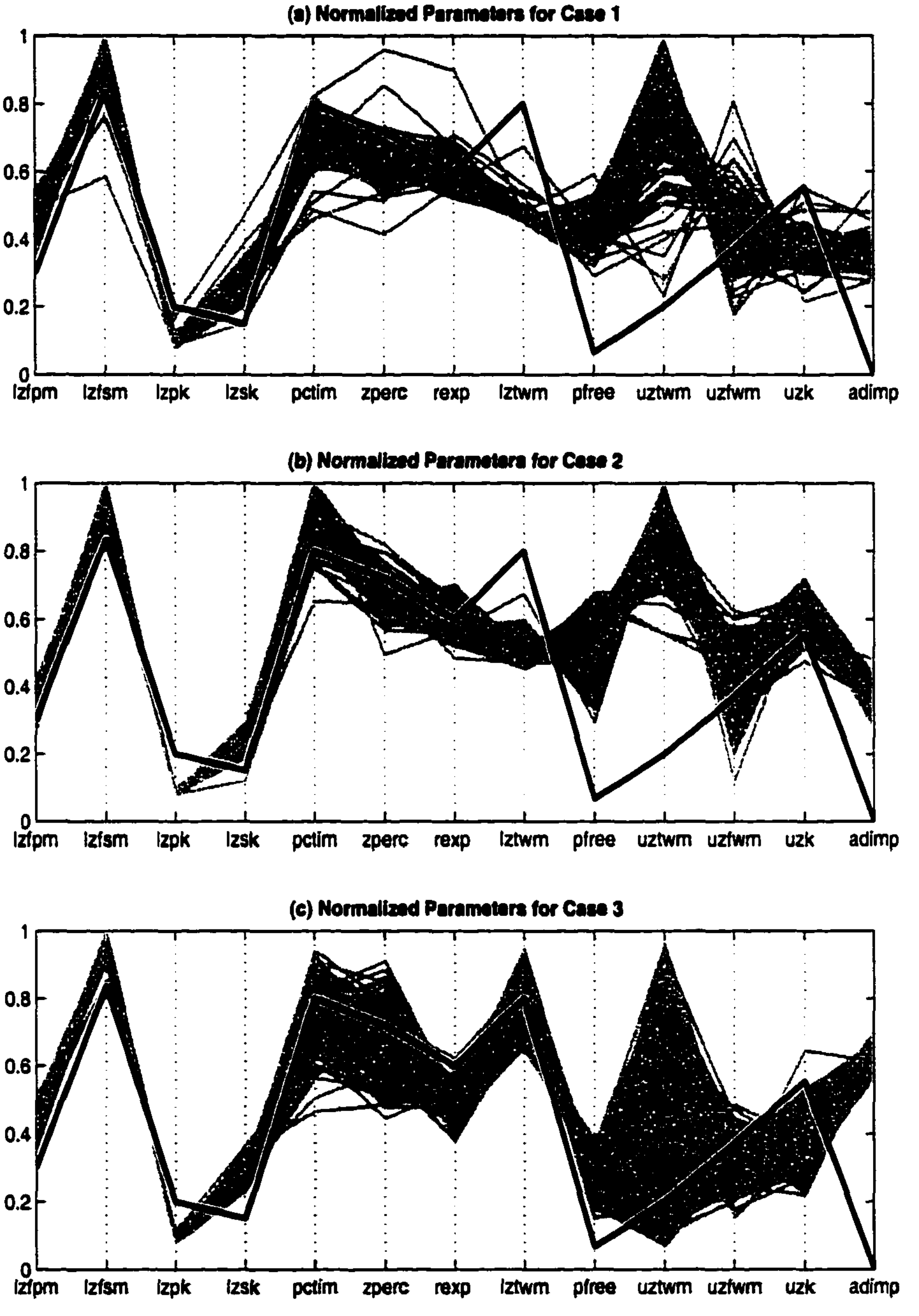

Figure 3.3 Normalized parameters for Pareto solutions obtained with the automatic multi-criteria approach to calibrate the SAC-SMA model: (a) Case 1, (b) Case 2, and (c) Case 3

There is, in general, larger variability in the estimates for the remaining six model parameters (and in particular UZTWM and UZFWM), which are primarily related to the properties of the upper soil layer and control the partitioning of the hydrograph into different quick flow components (overland and interflow). This suggests that the model performance tradeoffs indicated in Figures 3.2a-c are associated primarily with inadequacies in the representation of the near-surface soil processes.

b) Case 2: Distributed Input and Soil Moisture

The results of the multi-criteria automatic calibration of Case 2 (DIST-PS) are shown in Figures 3.2a-c. Note that, in this case, the channel inflows for all the sub-watersheds are lumped together and routed to the outlet using a single unit hydrograph. The results for the 8-subwatershed configuration did not give better results than the 3-subwatershed configuration. Therefore, the results presented here will draw primarily from the results of the 3-subwatershed study. Comparison of the solutions for this case (dark-gray dots) with the lumped case (Case 1, LUMP-ALL, light-gray dots) indicates a significant benefit to allowing the precipitation input and the soil-moisture computations to be distributed. In particular, the ability of the model to simulate the quick recession responses (FQ) and, to a lesser extent, the peaks/timing (FD) has been improved. However, there appears to be no additional impact on the model's ability to simulate the baseflow responses (FS).

A visual comparison of the DIST-PS results with the NWS lumped solution in Figures 3.2a-c clearly shows that the automatic approach provides a closer fit to the observed data in terms of all three criteria FD, FQ, and FS. Further, comparison of the

DIST-PS results with the NWS semi-distributed solution shows that most of the 500 Pareto solutions provide a better fit to the baseflow (FS) and peaks/timing (FD), while providing a comparable fit to the quick recession (FQ).

The variability in the parameter values across the 500 Pareto optimal solutions for Case DIST-PS is shown in Figure 3.3b. Notice that the variability in the parameters has generally decreased when compared with Case 1 (LUMP-ALL). As before, the ranges for the seven parameters, LZFPM, LZFSM, LZPK, LZSK, PCTIM, ZPERC, and REXP, are close to the values for the lumped NWS solution, further supporting the notion that the automatic multi-criteria approach is finding appropriate values for these parameters.

c) Case 3: Distributed Input, Soil Moisture and Routing

The results of the multi-criteria automatic calibration of Case 3 (DIST-PSR) are also shown in Figures 3.2a-c. In this case, the precipitation, soil-moisture computations, and channel routing are all treated separately for each subwatershed. Again, the results for the 8-subwatershed configuration did not give better results than the 3-subwatershed configuration, and results are therefore only presented for the latter configuration. The 500 Pareto optimal parameter sets (black dots) show that routing the channel inflow independently from each subwatershed to the outlet of the watershed improves the model's ability to simulate both the quick recession responses (FQ) and the peaks/timing (FD). Once again, there is no additional improvement in the model's ability to simulate the baseflow responses (FS). A visual comparison of the 500 Pareto solutions for this case with the NWS lumped and semi-distributed solutions (Figures 3.2a-c) clearly shows that the automatically calibrated semi-distributed model DIST-PSR provides a much

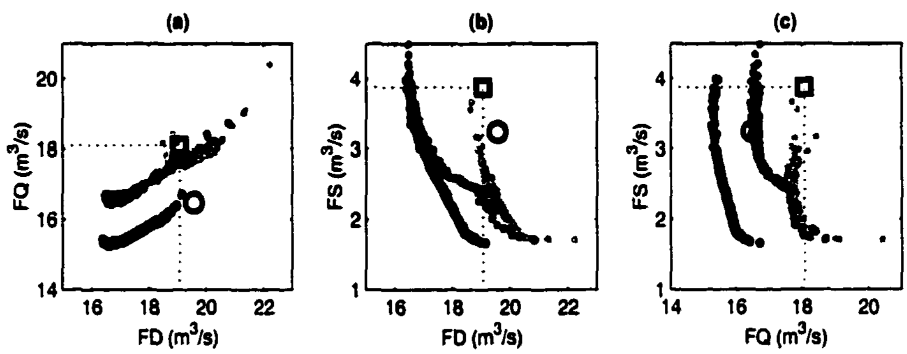

Figure 3.4 Pareto solutions obtained with the automatic multi-criteria approach to calibrate the SAC-SMA model: (a-c) two-dimensional projections of objective space. Marked points correspond to, respectively, 500 Pareto solutions for Case 1 (light-gray points), Case 4 (dark-gray points), and Case 3 (black points). NWS manual calibration results for lumped (square) and semi-distributed (open circle)

better reproduction of the watershed response, in terms of all three criteria FD, FQ, and FS.

The normalized parameter values for the 500 Pareto solutions are shown in Figure 3.3c. Again, the variability in the parameters is generally less than in Case 1 (LUMP-ALL), and the parameters LZFPM, LZFSM, LZPK, LZSK, PCTIM, ZPERC, and REXP are close to the lumped NWS solution. However, the variability of parameter UZTWM (upper zone tension water maximum capacity) significantly increased when compared with the results from Cases 1-2, suggesting that this parameter may need to take different values in each subwatershed.

d) Case 4: Distributed Soil Moisture and Routing, but Lumped Input

The results of the multi-criteria automatic calibration of Case 4 (DIST-SR) are shown in Figures 3.4a-c. In this case, the input was treated as lumped (averaged over the watershed), thereby simulating the *lack of availability* of radar-based precipitation data, while the model computations were run in semi-distributed mode. Only the 3-sub-watershed case is shown. The 500 Pareto optimal parameter sets obtained for this case (dark-gray dots) are shown overlaid on the results from Case 1 (LUMP-ALL, light-gray dots), and Case 3 (DIST-PSR, black dots). Note that FQ for most of the Case 4 solutions is improved over Case 1, but not as "good" as Case 3, indicating that spatially distributed routing and input (precipitation) both contribute to improvements in simulating the quick recession (FQ). However, the Case 4 solutions (lumped input, distributed routing) that provide the best match to FD (peaks/timing) are comparable to the Case 3 solutions (distributed input, distributed routing), indicating that the simulation of the peaks is

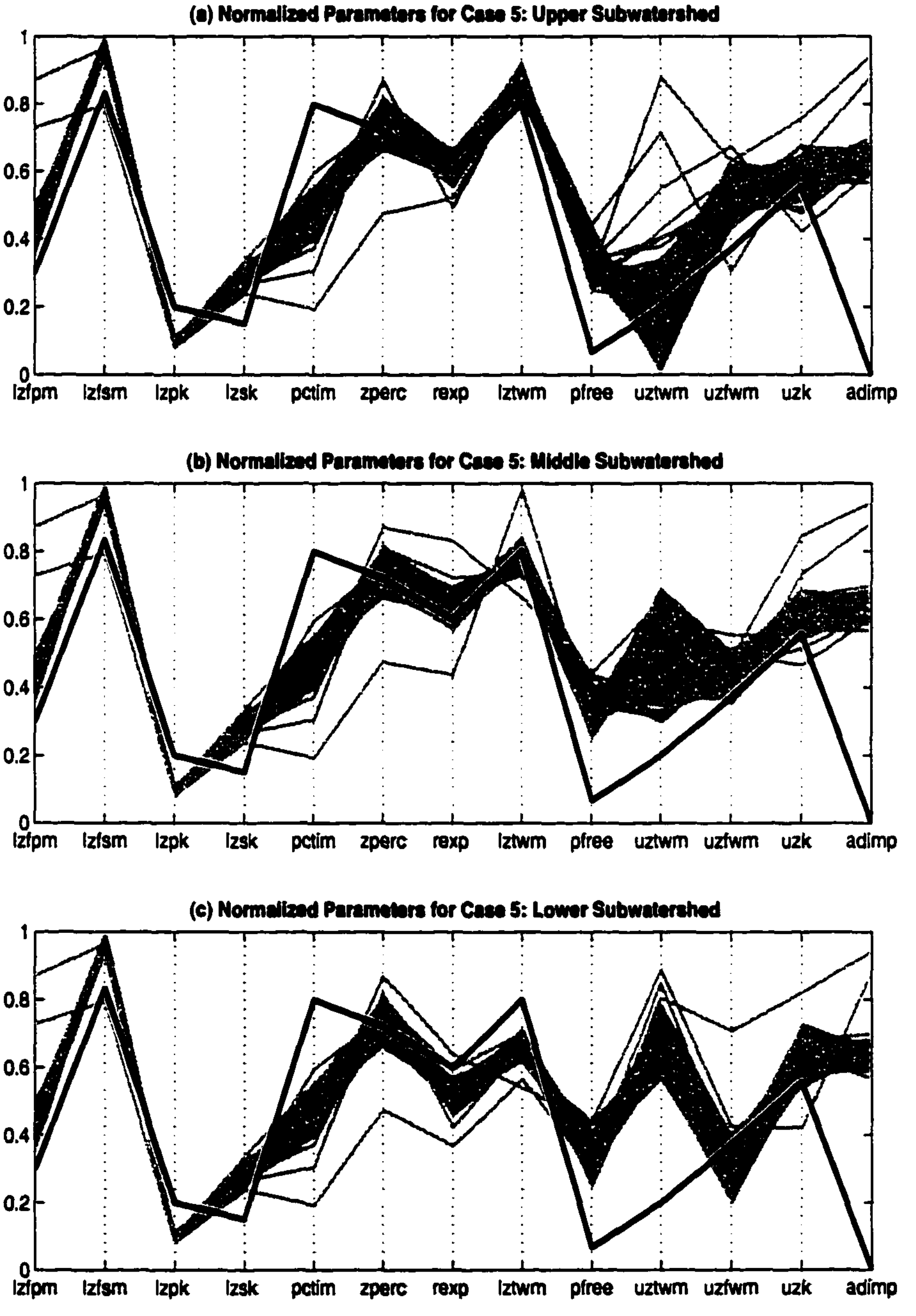

Figure 3.5 Normalized parameters for Pareto solutions obtained with the automatic multi-criteria approach to calibrate the SAC-SMA model for Case 5: (a) upper sub-watershed, (b) middle subwatershed, and (c) lower subwatershed

controlled more strongly by the spatial representation of routing than by the spatial representation of the inputs (precipitation).

e) Case 5: Distributed Input, Soil Moisture, Routing and Parameters

In Case 5 (DIST-PSRθ), the additional value of spatially distributing the watershed soil properties was investigated by allowing five of the model parameters (LZTWM, REXP, UZTWM, UZFWM, and UZK) to vary among the subwatersheds. Unexpectedly, the calibration results showed no noticeable criterion value improvements over the solutions obtained for the lumped parameter Case 3 (DIST-PSR). This is surprising, given the larger number of calibration parameters (23) compared with the number of calibration parameters (13) in Case 3 (DIST-PSR). Additional runs made with other combinations of parameters also provided no noticeable gains. However, the parameter plots shown in Figures 3.5a-c (each subplot represents a different subwatershed) indicate that the calibration has converged to significantly different values for the parameter UZTWM (upper zone tension water maximum capacity) in each sub-watershed: low value for the upper sub-watershed, medium value for the middle sub-watershed and high value for the lower subwatershed. This variability is consistent with results obtained under Case 3, where forcing this parameter to be lumped results in a large trade-off range (i.e., the parameter must take different values to fit different portions of the data record). Further, it seems reasonable that the increasing trend for UZTWM from the upper to lower sections of the watershed is hydrologically realistic and is consistent with soil depth maps for the watershed.

3.4.3 Comparison of Automatic and Manual Hydrographs

Examples of the simulated hydrographs for a 120-hour portion of the calibration period are shown in Figure 3.6. The black dots correspond to the observed data, the light-gray region corresponds to the hydrograph trade-off uncertainty associated with the Case 3 (DIST-PSR) multi-criteria solution, and the dashed and solid black lines correspond to the NWS lumped and semi-distributed results, respectively. Because the ordinates of the unit hydrographs were kept fixed (not dynamically adjusted) during the calibration, the hydrograph recessions are not well-simulated. Nevertheless, the multi-criteria solutions are able to reproduce the timing and magnitude of the flood peak extremely well.

3.5 SUMMARY AND CONCLUSIONS

The "semi-distributed" approach to modeling the spatial variability of important hydrologic processes is becoming popular as the availability of high-resolution hydrologic information continues to increase. However, the spatial detail with which this variability needs to be represented to provide accurate streamflow simulations is not well understood. Further, the increased complexity of the semi-distributed approach (and in particular the large number of parameters) results in a calibration problem of considerable difficulty. Traditional manual and automatic calibration approaches typically provide a single sub-optimal solution, with little or no information about the uncertainty in the estimated parameters or the model performance.

The applicability of multi-criteria automatic calibration methods to the calibration of semi-distributed watershed models has been demonstrated in this study. The

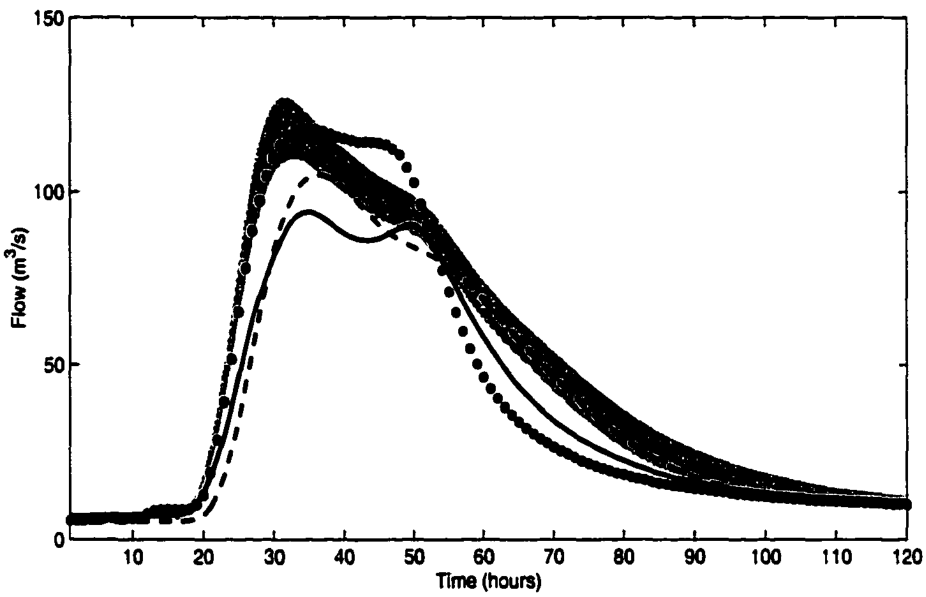

Figure 3.6 Hydrograph range associated with the Pareto solution set for Case 3. The dots correspond to the observed streamflow data, the light-gray region corresponds to the 500 Pareto solutions, the dashed line corresponds to the NWS lumped solution, and the solid line corresponds to the NWS semi-distributed solution

automatic approach was shown to be able to effectively handle the increased model complexity and provide parameter estimates (and hence model performance) that are comparable or superior to those obtained by the manual/expert approach. The improved results are also achieved at a considerable savings in "calibration time"–after problem setup, the automatic approach requires about 2 to 4 hours of computing time on a Sun workstation, compared with several days to weeks of person-time for the manual approach [M. Smith, Hydrology Laboratory, NWS, personal communication, 2000]. In addition, the study has demonstrated how multi-criteria methods provide a useful framework for the systematic investigation of appropriate model complexity.

The study was specifically motivated by the needs of the U.S. National Weather Service for improved procedures for calibration of watershed models in the context of the Advanced Hydrologic Prediction System (AHPS) modernization initiative. In particular, it was designed to complement ongoing NWS studies into the development of semi-distributed modeling strategies that exploit the increasing availability of spatial hydrologic information, including NEXRAD precipitation. The effectiveness and efficiency of the automatic approach enabled us to rapidly investigate the specific benefits associated with different levels of spatial representation of various model components, including the model input (precipitation), structural components (soil moisture and streamflow routing computations), and surface characteristics (parameters). In particular, it was found that (a) the semi-distributed model provided significant performance improvements over the lumped model; (b) there was a limit to the performance improvements associated with increasing representation of spatial hydrologic variability in the model; (c) the main improvements were provided by spatial representation of the precipitation (inputs) and structural components (soil moisture and streamflow routing computations); (d) little or no improvements were provided by spatial representation of the soil properties (model parameters); (e) the spatial variability in hydrologic information contributed mainly to improved simulation of the flood peaks and the quick recessions; and (f) semi-distributed modeling provided no improvements in representation of the baseflow. These results are, of course, specific to the SAC-SMA model for the Blue River watershed, although conclusions a, b, and f may prove to be more generally applicable.

Conclusion "d", relating to the lack of improvement in model performance when the model parameters are allowed to be distributed spatially, is both unexpected and interesting. This seems contrary to the general belief among hydrologists that the spatial variability in soil properties exerts a significant control on the hydrologic response of a watershed. For the Blue River watershed at least, our results indicate that the dominant control on the hydrologic response is the watershed topography–which strongly determines the routing characteristics and (perhaps less strongly) the spatial distribution of precipitation–and that the impacts of variations in the soil and vegetation properties are averaged out by the time the streamflow reaches the watershed outlet. Further studies to investigate the conditions under which the various components of hydrologic variability do (or do not) contribute to variations in total watershed response would provide significant improvements to our understanding of watershed behavior.

CHAPTER FOUR

INCORPORATING VERTICAL INFORMATION

4.1 INTRODUCTION AND SCOPE

Conceptual rainfall-runoff (CRR) models provide an approximate, lumped description of the dominant sub-watershed scale processes that contribute to the overall watershed scale hydrologic response of the watershed system. In their most basic form, CRR models transform rainfall into runoff with two main components, precipitation excess generation and flow routing. Precipitation excess is generated as a function of the vertical movement of moisture (precipitation, evaporation, transpiration, and losses to the system) into and out of the watershed. The variability of the excess generation process within the watershed is related to the level of spatial variability of the soil properties, vegetation type, and precipitation rates throughout the watershed.

In the past, high-resolution information describing the spatial variability of these watershed characteristics was not readily available. As a result, performance improvements of CRR models were primarily achieved through improvements in the representation of the hydrologic processes directly related to the vertical movement of moisture within the watershed rather than the spatial variability of these processes. The level of vertical detail describing the movement of moisture through the soil (vertical model structure complexity) that needs to be represented by a conceptual rainfall-runoff (CRR) model in order to accurately simulate the streamflow is not well understood.

The degree of vertical model structure complexity can vary significantly among different CRR models. The Antecedent Precipitation Index method (Linsley et al., 1949;

Sittner et al., 1969) is an example of a simple CRR model in which an antecedent soil moisture index is combined with a set of linear equations to estimate runoff from storm precipitation. Examples of CRR models with complex vertical structures include the Stanford Watershed Model (Crawford, 1966), the Sacramento Soil Moisture Accounting (SAC-SMA) Model (Burnash et al., 1973), the KINEROS model (Woolhiser et al., 1990), and TOPMODEL (Beven and Kirkby, 1979). These models all use large numbers of parameters and complex relationships to represent the vertical movement of moisture through the soil.

In general one might expect that the more detail or complexity included in the structure of the model, the better the representation of the real watershed behavior (streamflow). A number of studies have compared the performance of simple versus complex models [Naef, 1981; Loague and Freeze, 1985; Wilcox et al, 1990; Michaud and Sorooshian, 1994; Refsgaard and Knudesn (1996); among others]. Results from these studies have generally suggested that the performance of simple models is approximately as accurate as complex models. The development and application of complex models, however, continues despite the results of these studies.

Several researchers have recognized the importance of selecting a model with a level of complexity that is appropriate for a given problem [Wheater et al. 1986, Beck et al., 1990; Jakeman and Hornberger, 1993; Grayson and Chiew, 1994; among others]. Jakeman and Hornberger (1993) explored the constraints that observed data place on the allowable complexity of rainfall-runoff models. They used time series techniques to determine how many parameters are appropriate to describe the relationship between precipitation and streamflow for several watersheds of various sizes using available precipitation, air temperature and streamflow data. Their results suggest that the information content in a rainfall-runoff record is sufficient to support models of only very limited complexity.

In Chapter 3, the impact of "spatially distributing" the model input (precipitation), structural components (soil moisture and streamflow routing computations) and surface characteristics (parameters) was explored through lumped and semi-distributed applications of the SAC-SMA model. The focus of this chapter is to provide an assessment of the potential improvements in streamflow simulation that can be achieved through various levels of representation of the vertical movement of moisture through the watershed using CRR models in lumped applications. The relative benefits of different levels of vertical model structure (direct runoff, upper soil moisture storage, and the percolation process) are examined with a simple hydrologic model, HYMOD [Boyle et al. 1997; Wagener et al., 2001]. HYMOD consists of a variety of different excess generation (interception storage, tension storage, free storage, etc.), percolation, and streamflow routing functions that can be put together in different combinations to describe the different hydrologic behaviors of the watershed system. The multi-criteria framework developed by Boyle et al. [2000] (also described in detail in Chapter 2) for application to hydrologic models is used to calibrate each CRR model in terms of three objective measures designed to reflect the different observable characteristics of watershed behavior (peak flow and timing, quick recession, and baseflow). Multi-criteria performance comparisons among the different model applications are used to evaluate the benefits of various types and degrees of vertical model complexity. Results obtained from a lumped application of the SAC-SMA model in Chapter 2 are used as a benchmark for comparison with results from this study.

This chapter is organized as follows: the methodology to investigate different levels of vertical model complexity on a study watershed is presented in Section 4.2. The results of the model applications are presented in Section 4.3, and Section 4.4 discusses the results of the study.

4.2 METHODS

4.2.1 Overview

The CRR model, HYMOD [Boyle et al. 1997], was used in this study to investigate the value of different levels of vertical detail describing the movement of moisture through the soil. HYMOD consists of a variety of different excess generation (interception storage, tension storage, free storage, etc.), percolation, and streamflow routing functions that can be put together in different combinations to describe the different hydrologic behaviors of the watershed system. Extensive applications of different configurations of HYMOD were made on the Leaf River Watershed (see description of Leaf River Watershed in Chapter 2). The results obtained with the application of the SAC-SMA model on the Leaf River Watershed (described in Chapter 2) are used as a benchmark for comparison with results obtained in this study. The study was designed to expand our understanding of the specific benefits associated with different levels of representation of the important vertical processes in a watershed.

4.2.2 Description of HYMOD

The hydrologic model HYMOD [Boyle et al. 1997; Wagener et al., 2001], was developed to represent the basic spatial and vertical hydrologic processes within a watershed with a minimum number of modeling units and functional parameters. HYMOD has a flexible model structure (see Figure 4.1) that allows the user to assemble and test a variety of different excess generation (impervious area, tension and free water storages, etc.), percolation, and streamflow routing functions in different combinations to describe the different hydrologic behaviors of the watershed system.

The general configuration of HYMOD for the purposes of this study is shown in Figure 4.1. The watershed is partitioned into two areas, pervious and impervious, by means of a single parameter, percent impervious area (*PCTIM*). Precipitation (rainfall) falling on the impervious portion of the watershed becomes direct runoff available for routing along with the surface runoff component, to the outlet of the watershed. Precipitation falling on the pervious portion of the watershed enters the upper soil moisture zone (UZ). The UZ consists of two components, tension water storage and free water storage. Tension water storage must be completely satisfied before moisture can move to the free water portion of the UZ. Soil moisture within the UZ tension and free water storages is available to satisfy the potential evapotranspiration (PET) demand.

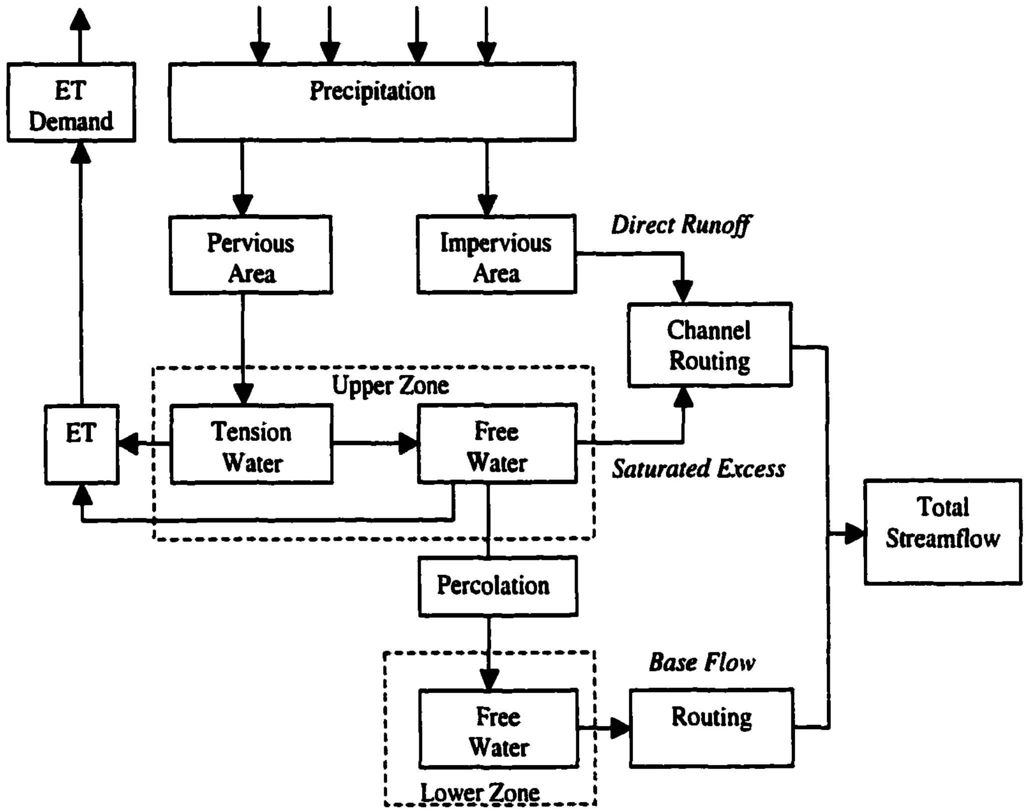

Figure 4.1 HYMOD watershed model

Saturated excess is generated from the free water storage and then combined with the direct runoff to estimate the quick (or surface) runoff. The surface runoff is then routed through a series of *NUMQ* linear reservoirs, each with the same recession coefficient, *KQ*, to the outlet of the watershed. Soil moisture percolates from the UZ free water storage to the lower soil moisture zone (LZ) free water storage. The moisture in the LZ free water is routed through a single linear reservoir, with recession coefficient KS, to estimate the slow (or baseflow) runoff at the outlet of the watershed. The quick and slow flows are then combined to estimate the total streamflow at the outlet of the watershed.

4.2.3 Description of this Study

In this study, a series of 60 different applications (or cases) of HYMOD to the Leaf River watershed were made to investigate the improvements in model performance associated with various levels of vertical detail describing the movement of moisture through the soil (UZ tension and free water storages, percolation process, and pervious area). Each model application was designed to isolate the effects of different levels of vertical model complexity in terms of specific desirable watershed behaviors (driven flow – "peaks and timing", non-driven quick flow – "quick recession" responses, and non-driven slow – "baseflow" responses). The multi-criteria framework described in Chapter 2 is used to calibrate each of the different applications of HYMOD in terms of the three objective measures driven flow (FD), non-driven quick flow (FQ), and non-driven slow flow (FS). For each modeling case, the Pareto optimal solution space for the three criteria (FD, FQ, FS) was estimated by 500 solutions generated using the MOCOM algorithm.

In Cases 1-30 and 31-60, two different functional relationships describing the UZ free water storage were tested. In Cases 1-30, a simple bucket loss (BL) model (see Figure 4.2a) is used to generate saturated excess when the soil moisture level in the tank is greater than parameter *FMAX*. In Cases 31-60, the Moore loss (ML) model (see Figure 4.2b), described in detail by Moore [1985 and 1999], assumes that the soil moisture capacity varies across the watershed and therefore the proportion of the watershed with saturated soils varies over time. The spatial variability of the soil moisture capacity is described by the following distribution function:

$$F(c) = 1 - \left[1 - \left(\frac{c(t)}{CMAX}\right)\right]^{BEXP} \qquad 0 \le c(t) \le CMAX$$

where, c is the storage capacity (mm), $CMAX$ is the maximum storage capacity in the watershed (mm), $F(c)$ is the fraction of the watershed that is saturated (unit less), and $BEXP$ is the degree of spatial variability of the soil moisture capacity within the watershed (unit less).

For each of the UZ free water functions, ten different functional relationships describing the percolation between the UZ and LZ free water storages were also tested in this study. The ten functions are described by the following five equations:

$$PERC(t)_{1\text{-}2} = min(\,contents(t),\, PMAX)$$

$$PERC(t)_{3\text{-}4} = PFRAC{*}contents(t)$$

$$PERC(t)_{5\text{-}6} = min(\,PFRAC{*}contents(t),\, PMAX)$$

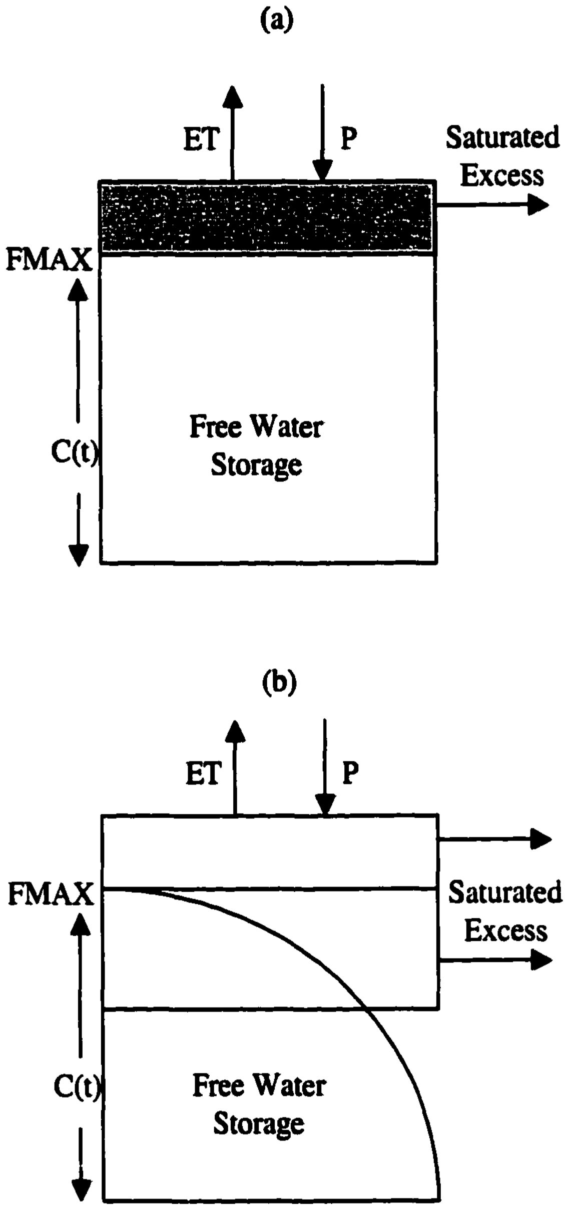

Figure 4.2 Representation of Upper Zone free water with (a) simple bucket model, and (b) the Moore loss model

$$PERC(t)_{7\text{-}8} = min(\,BETA(1 + ZETA(LZDR(t))^{REXP},\ 1.0)\,)*contents(t)$$

$$PERC(t)_{9\text{-}10} = min(\,BETA(X\_FREE(t)/FMAX)\ (1 + ZETA(LZDR(t))^{REXP},\ 1.0)\,)*contents(t)$$

where *contents(t)* is the UZ saturated excess at time *t* ($PERC(t)_{1,3,5,7,9}$) or the free water storage at time *t* ($PERC_{2,4,6,8,10}$), *LZDR(t)* is the lower zone deficiency ratio (1- contents of lower zone free water storage/maximum contents of LZ free water storage) at time *t*, *X_FREE(t)* is the UZ free water contents, *FMAX* is the maximum contents of UZ free water storage, and *PMAX*, *PFRAC*, *BETA*, *ZETA*, and *REXP* are calibration parameters.

Each of these ten functions provides a unique conceptualization of the relationship between the UZ and LZ free water contents and the percolation process. In general, the complexity of the relationships ranges from low in $PERC(t)_{1\text{-}2}$ and $PERC(t)_{3\text{-}4}$ to high in $PERC(t)_{7\text{-}8}$ and $PERC(t)_{9\text{-}10}$. The functional relationships in $PERC(t)_{1\text{-}2}$ and $PERC(t)_{3\text{-}4}$ describe the amount of UZ free water storage that can be percolated to LZ free water storage as a maximum amount and fraction, respectively, of the contents. $PERC(t)_{5\text{-}6}$ describes the percolation process as a fraction of the UZ free water storage that can be percolated to LZ free water storage with a maximum value for a given time step. $PERC(t)_{7\text{-}8}$ and $PERC(t)_{9\text{-}10}$ approximate the complex percolation process used in the SAC-SMA model. The primary difference between the latter two being that the function $PERC(t)_{9\text{-}10}$ allows the UZ free water storage contents (*X_FREE(t)/FMAX*) and the lower zone (*LZDR(t)*) to influence the percolation rate while $PERC(t)_{7\text{-}8}$ is influenced by the lower zone (*LZDR(t)*).

Table 4.1: Table of modeling cases – simple bucket loss (BL)

Case #	Case Name	UZ Free (BL/ML)	Percolation (1-10)	UZ Tension (T)	Impervious Area. (I)
1	BL-1	BL	1	N	N
2	BL-2	BL	2	N	N
3	BL-3	BL	3	N	N
4	BL-4	BL	4	N	N
5	BL-5	BL	5	N	N
6	BL-6	BL	6	N	N
7	BL-7	BL	7	N	N
8	BL-8	BL	8	N	N
9	BL-9	BL	9	N	N
10	BL-10	BL	10	N	N
11	BL-1T	BL	1	Y	N
12	BL-2T	BL	2	Y	N
13	BL-3T	BL	3	Y	N
14	BL-4T	BL	4	Y	N
15	BL-5T	BL	5	Y	N
16	BL-6T	BL	6	Y	N
17	BL-7T	BL	7	Y	N
18	BL-8T	BL	8	Y	N
19	BL-9T	BL	9	Y	N
20	BL-10T	BL	10	Y	N
21	BL-1TI	BL	1	Y	Y
22	BL-2TI	BL	2	Y	Y
23	BL-3TI	BL	3	Y	Y
24	BL-4TI	BL	4	Y	Y
25	BL-5TI	BL	5	Y	Y
26	BL-6TI	BL	6	Y	Y
27	BL-7TI	BL	7	Y	Y
28	BL-8TI	BL	8	Y	Y
29	BL-9TI	BL	9	Y	Y
30	BL-10TI	BL	10	Y	Y

Y = Yes
N = No

The influence of the UZ tension water storage and the percent impervious area of the watershed were also investigated. In Cases 1-10, the bucket loss representation of the UZ free water storage was combined, separately, with each of the ten percolation functions (see Table 4.1). In each of these cases, the size of the tension water storage tank and the percent impervious area of the watershed were set to zero (no UZ tension water storage and no impervious area in the watershed). In Cases 11-20, Cases 1-10 were repeated with the UZ tension water storage component but no impervious area. Finally, in Cases 21-30, the UZ tension water storage and the impervious area components of HYMOD were considered. This was repeated for Cases 31-60 with the Moore loss representation of the UZ free water storage instead of the simple bucket loss representation (see Table 4.2).

4.3 RESULTS

4.3.1 SAC-SMA Model Results

The results obtained in Chapter 2 with the SAC-SMA model were used as a benchmark for comparison with the studies described above. The ranges of the SAC-SMA results, in terms of the three criteria FD, FQ, and FS, are shown in gray in the multi-criteria format in Figure 4.3a-c.

Table 4.2: Table of modeling cases – Moore loss (ML)

Case #	Case Name	UZ Free (BL/ML)	Percolation (1-10)	UZ Tension (T)	Impervious Area. (I)
31	ML-1	ML	1	N	N
32	ML-2	ML	2	N	N
33	ML-3	ML	3	N	N
34	ML-4	ML	4	N	N
35	ML-5	ML	5	N	N
36	ML-6	ML	6	N	N
37	ML-7	ML	7	N	N
38	ML-8	ML	8	N	N
39	ML-9	ML	9	N	N
40	ML-10	ML	10	N	N
41	ML-1T	ML	1	Y	N
42	ML-2T	ML	2	Y	N
43	ML-3T	ML	3	Y	N
44	ML-4T	ML	4	Y	N
45	ML-5T	ML	5	Y	N
46	ML-6T	ML	6	Y	N
47	ML-7T	ML	7	Y	N
48	ML-8T	ML	8	Y	N
49	ML-9T	ML	9	Y	N
50	ML-10T	ML	10	Y	N
51	ML-1TI	ML	1	Y	Y
52	ML-2TI	ML	2	Y	Y
53	ML-3TI	ML	3	Y	Y
54	ML-4TI	ML	4	Y	Y
55	ML-5TI	ML	5	Y	Y
56	ML-6TI	ML	6	Y	Y
57	ML-7TI	ML	7	Y	Y
58	ML-8TI	ML	8	Y	Y
59	ML-9TI	ML	9	Y	Y
60	ML-10TI	ML	10	Y	Y

Y = Yes
N = No

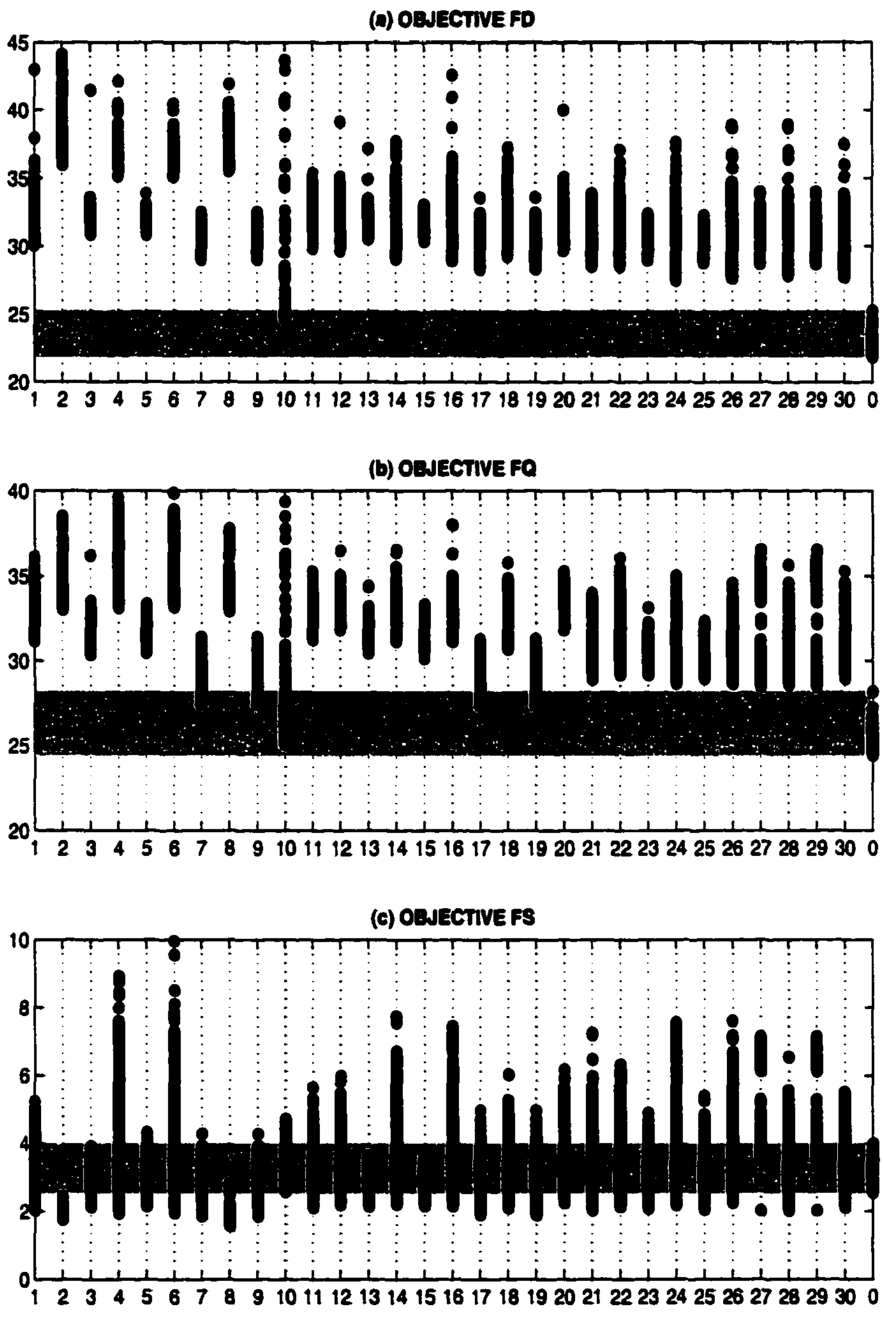

Figure 4.3 500 Pareto solutions obtained with the automatic multi-criteria approach (black dots) to calibrate Cases 1-30 using the simple bucket model: (a) objective function FD, (b) objective function FQ, (c) objective function FS. The gray shading represents the SAC-SMA results.

4.3.2 Bucket Loss Model Results

General Results

Figure 4.3a-c presents the results for Cases 1-30 (application of HYMOD using the simple bucket representation of the UZ free water storage). Notice that in terms of the criteria FD, the results for all the Cases (except Case 10) are inferior to all of the solutions obtained with the SAC-SMA model. In addition, the results for many of the cases are inferior to all of the solutions obtained with the SAC-SMA model in terms of FQ. These results indicate that the model structures in Cases 1-30 are not representing hydrologic processes important to the simulation of the larger flows in the same way that the SAC-SMA model does. On the other hand, all of the thirty of the cases contain solutions that are superior to those obtained with the SAC-SMA model in terms of the FS criterion. This result indicates that the LZ representation used by each of the Cases 1-30 may be adequate to represent the important hydrologic processes required to simulate the lower flows.

Cases 1-10

The results for Cases 1-10 (simple bucket loss representation of UZ free water storage with no UZ tension water storage or impervious area) are also shown in Figures 4.3a-c. A visual comparison of the results shows that Cases 1,3, 5, and 7 are superior to Cases 2, 4, 6, and 8 in terms of both FD and FQ indicating that the model performs better (at least in this configuration) with the percolation source as the saturated excess rather than the UZ free water storage contents. Notice that the results for Cases 7 and 9 (as well as 17 and 19, and 27 and 29) appear to be identical for all three criteria, FD, FQ, and FS.

The fact that these results are not unique is a consequence of combining the simple bucket model with percolation relation $PERC(t)_{9\text{-}10}$. When the *contents(t)* variable is set to be the saturated excess (as it is in Cases 7, 9, 17, 19, 27, and 29) the only time there can be percolation is at times when there is saturated excess (when $X\_FREE(t)/FMAX = 1.0$). This effectively makes $PERC(t)_{7\text{-}8} = PERC(t)_{9\text{-}10}$ for all thirty of the simple bucket applications. This is not a relevant issue in Cases 31-60 since saturated excess can occur in the Moore loss representation without $X\_FREE(t)/FMAX = 1.0$.

From visual inspection of Figure 4.3a-c it can be clearly seen that the results for Case 10 are superior in terms of fitting the FD and FQ criteria than the results for any other of the Cases. Figure 4.4a-c presents the results for Cases 9 and 10 and the SAC-SMA model in two-dimensional projections of the three-criteria solution space. The SAC-SMA solutions (black dots) are clearly superior in terms of FD to most of the solutions in Case 10 (dark gray dots) and all of the solutions in Case 9 (light gray dots). Many of the solutions for Case 10, however, are at least as good as the SAC-SMA solutions in terms of the FQ and FS criteria (the lower values of FS criteria for Case 10 are out of the plotting range indicating a very large trade-off between the fitting of the FD and FS criteria). Many of the Case 9 solutions are superior to both the SAC-SMA and Case 10 solutions in terms of the FS criteria. Also note that the solutions for Case 9 have much less trade-off in fitting the FD and FS, and FQ and FS criteria, although the fit to both FD and FQ is inferior compared with Case 10 and the SAC-SMA model.

The variability in the parameter values across the 500 Pareto optimal solutions for Cases 9 and 10 is shown in Figures 4.5a-b (gray lines) using a normalized parameter plot.

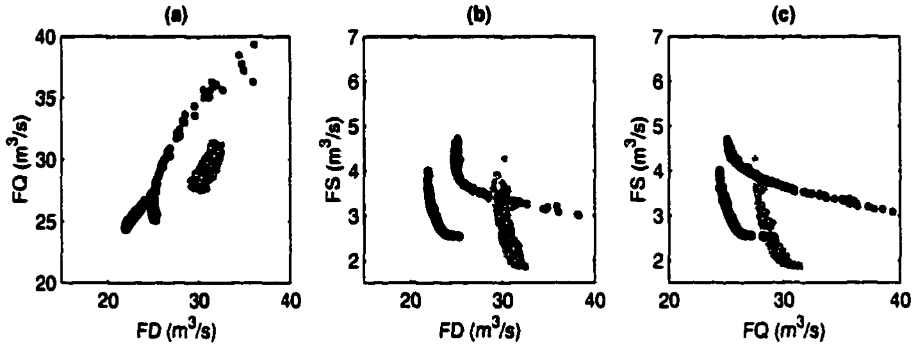

Figure 4.4 Pareto solutions obtained with the automatic multi-criteria approach to calibrate the HYMOD model: (a-c) two-dimensional projections of objective space. Marked points correspond to, respectively, 500 Pareto solutions for Case 9 (light-gray points), Case 10 (dark-gray points), and SAC-SMA model (black points)

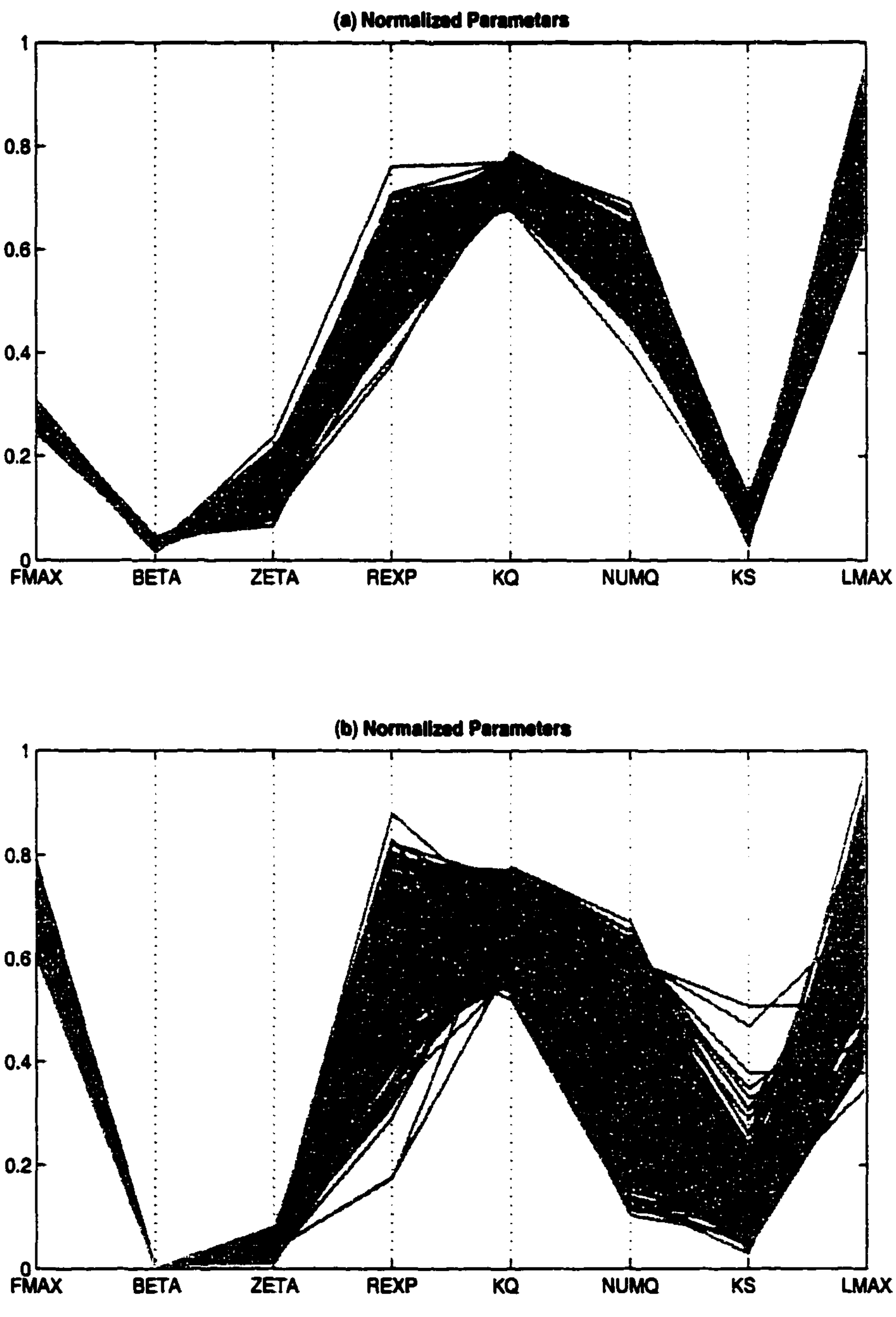

Figure 4.5 Normalized parameters for Pareto solutions obtained with the automatic multi-criteria approach to calibrate HYMOD model for (a) Case 9, and (b) Case 10

The maximum range for each parameter represents the range over which the multi-criteria calibration was performed. Notice that the multi-criteria optimization has resulted in a significant reduction of the parameter range. Notice also that ranges for the parameters *LMAX*, *KS*, *NUMQ*, *KQ*, and *REXP* for Case 9 show smaller amounts of variability compared to those for Case 10. The only parameter that appears to be in a significantly different range is *FMAX* (normalized value = 0.2-0.4, and 0.6-0.8 for Cases 9 and 10, respectively). This is most likely related to the different sources of moisture (excess saturation and UZ free water storage) used by the two models.

Cases 11-30

A visual inspection of Figures 4.3a-c indicates that the addition of the tension water storage in Cases 11-20 has generally improved the results slightly in terms of FD and FQ. In addition, the differences between Cases 1,3, 5, and 7 and Cases 2, 4, 6, and 8, in terms of both FD and FQ, have been significantly reduced. Further improvement is made in Cases 21-30, in terms of FQ criteria, when the pervious area component is added to the model. In none of these cases, however, did the addition of the UZ tension water storage or the impervious area component substantially improve the fitting of the FS criterion.

The three cases that appear to have the "best" solutions for each of the three main model configurations are Case 10 (Cases 1-10), Case 19 (Cases 11-20), and Case 30 (Cases 21-30). Figures 4.6a-c present the results for Case 10 (light gray dots), Case 19 (dark gray dots) and Case 30 (black dots) in two-dimensional projections of the three-criteria solution space. Notice that the solutions for these cases have very little trade-off

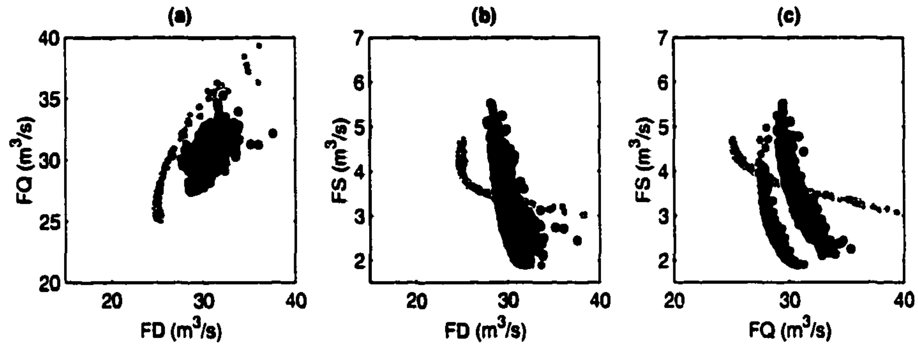

Figure 4.6 Pareto solutions obtained with the automatic multi-criteria approach to calibrate the HYMOD model: (a-c) two-dimensional projections of objective space. Marked points correspond to, respectively, 500 Pareto solutions for Case 10 (light-gray points), Case 19 (dark-gray points), and Case 30 (black points)

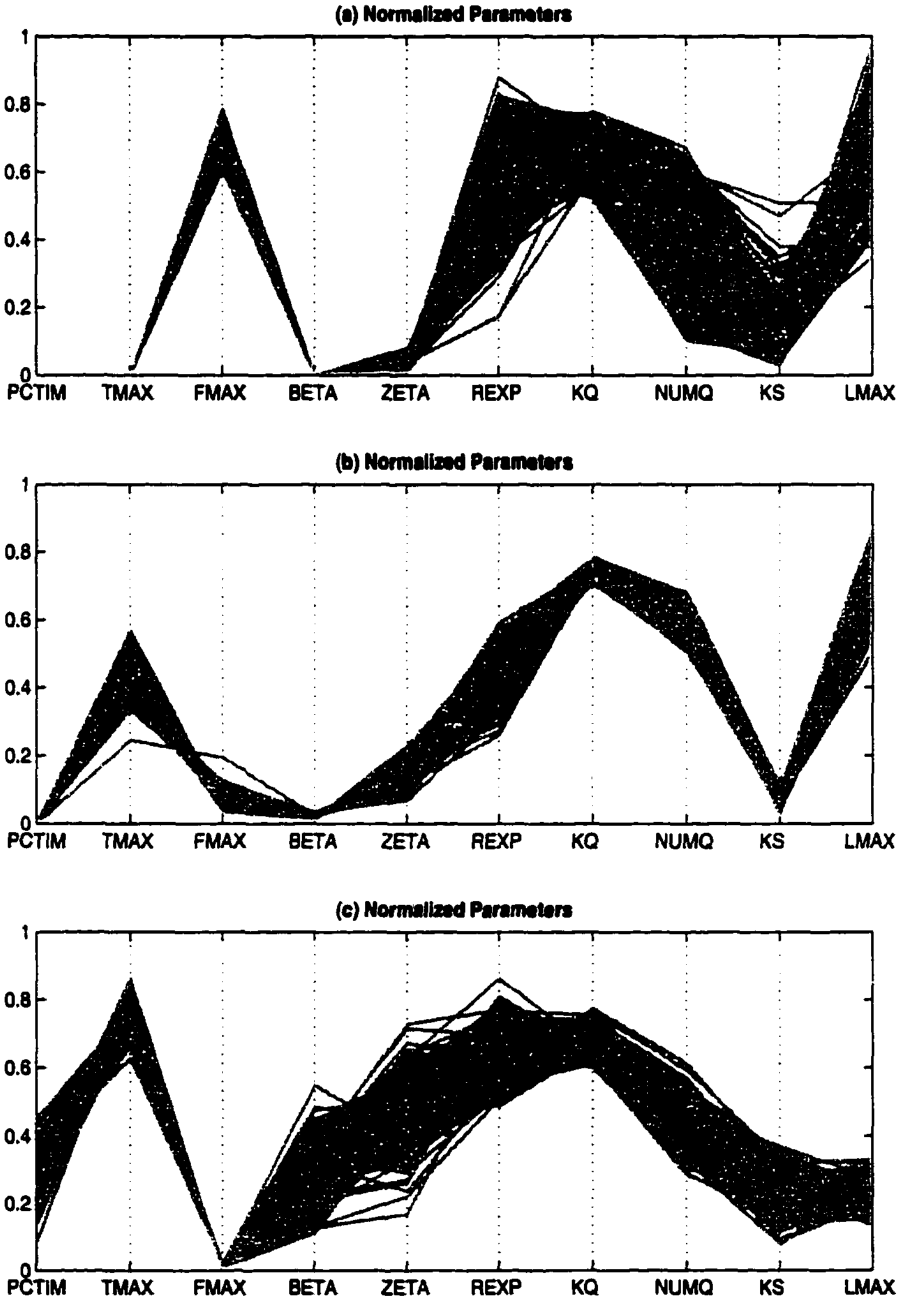

Figure 4.7 Normalized parameters for Pareto solutions obtained with the automatic multi-criteria approach to calibrate HYMOD model for (a) Case 10, (b) Case 19, and (c) Case 30

in terms of the FD and FQ criteria. From the figure it can clearly be seen that many of the solutions for Case 10 are superior to those in Cases 19 and 30. The trade-off range, however, is dramatically different for Cases 19 and 30 compared to that of Case 10 (improvement is FS is possible with very little cost in terms of FD and FQ).

The variability in the parameter values across the 500 Pareto optimal solutions for Cases 10, 19, and 30 is shown in Figures 4.7a-c (gray lines) using a normalized parameter plot. The maximum range for each parameter represents the range over which the multi-criteria calibration was performed. Notice that the multi-criteria optimization has resulted in a significant reduction of the parameter range. Notice also that the parameter *FMAX* (controls the size of the UZ free water storage) is significantly reduced in Case 19 compared to Case 10. This is appears to be directly related to the addition of the UZ tension water storage component. *FMAX* is further reduced (*TMAX* becomes larger) in Case 30 as the impervious area component is added to the model. There is not much of a noticeable change in the remaining parameters, except for *LMAX*, among the three cases. The values for *LMAX* are significantly reduced in Case 30 compared with Cases 10 and 19. This may be attributed to the model's ability to produce a significant portion of the driven "quick" flow with the impervious area component and, as a result, the UZ free water storage component's ability to produce more of the non-driven quick flow component. The LZ free water storage, therefore, needs less water (lower *LMAX*) to simulate the non-driven slow flow.

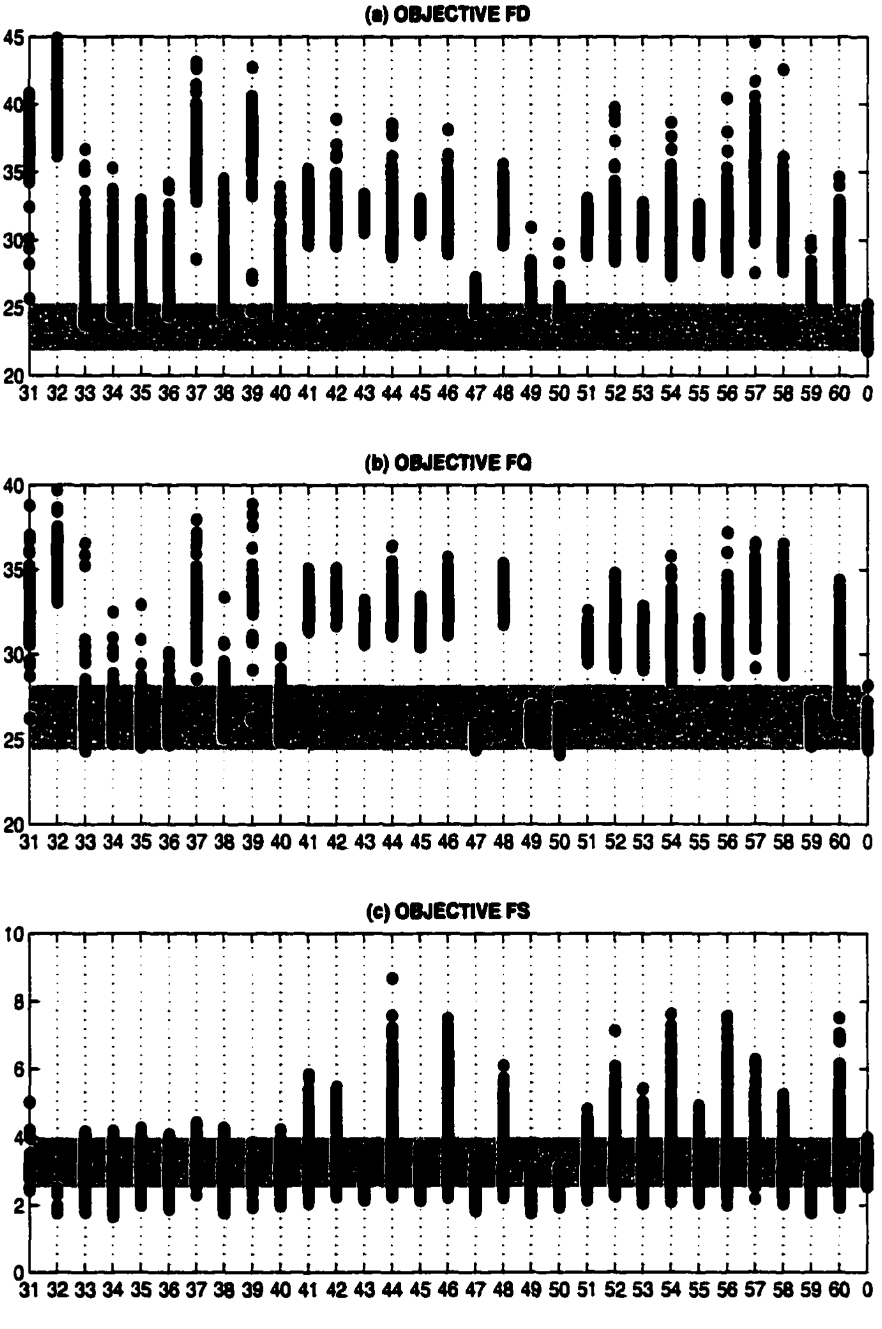

Figure 4.8 500 Pareto solutions obtained with the automatic multi-criteria approach (black dots) to calibrate Cases 31-60 using the Moore loss model: (a) objective function FD, (b) objective function FQ, (c) objective function FS. The gray shading represents the SAC-SMA results

4.3.3 Moore Loss Model Results

General Results

Figure 4.8a-c present the results for Cases 31-60 (application of HYMOD using the Moore loss representation of the UZ free water storage). Notice that none of the cases have solutions that are able to match both the FD and FQ criteria as well as the SAC-SMA model. In all cases, however, at least some of the solutions (and in some cases many or almost all of the solutions) are superior to all of the SAC-SMA solutions in terms of the FS criteria. Also notice that, in general, the addition of the tension and impervious area components (Cases 41-50 and 51-60, respectively) actually degrade the model's ability to fit the FD and FQ criteria. This suggests that the Moore loss model may have the functionality of the tension and impervious components already incorporated into its structure. An exception is Case 47 (UZ tension water storage component), which appears to have the "best" overall fit to the three criteria.

Comparison with Bucket and SAC-SMA Results

Figure 4.9a-c presents the results for Cases 10 (best case for the bucket representation), Case 47 (best cases for the Moore loss representation), and the SAC-SMA model in two-dimensional projections of the three-criteria solution space. The SAC-SMA solutions (black dots) are clearly superior (but not much) in terms of FD to most of the solutions in Case 47 (dark gray dots) and all of the solutions in Case 10 (light gray dots). Many of the solutions for Cases 47 and some of the solutions for Case 10, however, are at least as good as the SAC-SMA solutions in terms of the FQ criterion.

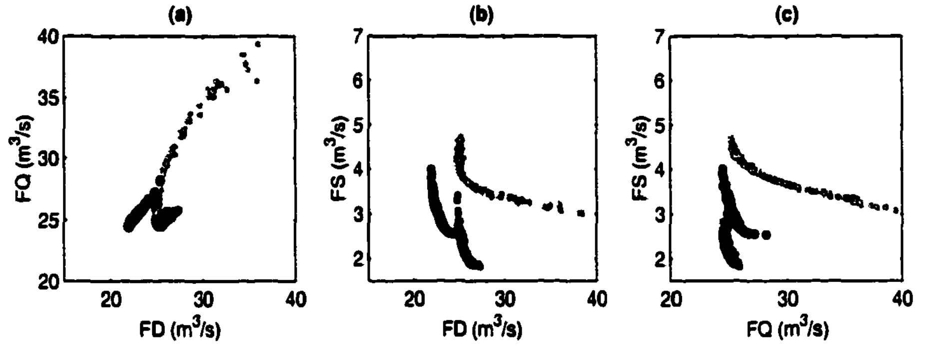

Figure 4.9 Pareto solutions obtained with the automatic multi-criteria approach to calibrate the HYMOD model: (a-c) two-dimensional projections of objective space. Marked points correspond to, respectively, 500 Pareto solutions for Case 10 light-gray points), Case 47 (dark-gray points), and SAC-SMA model (black points).

Many of the Case 47 solutions are superior to both the SAC-SMA and Case 10 solutions in terms of the FS criteria.

4.4 SUMMARY AND CONCLUSIONS

Many of the CRR models used to simulate streamflow consist of highly complex functional relationships to describe the movement of moisture vertically through the soil. These models are often very difficult to calibrate due to the large number of parameters and complex relationships within the model. Further, the large number of parameters may limit the manner in which newly available high-resolution (spatial) hydrologic information can be incorporated into the hydrologic modeling process, thereby limiting the actual benefit(s) of the potential added information contained within the new data. As a result, there is a real need to understand the specific benefits associated with increased representation of the movement of moisture vertically through the soil. With this new understanding, new models can be developed from new and existing modeling concepts, with parsimonious model structures that represent only those response modes that are identifiable within the available data.

This chapter explores the specific improvements in streamflow simulation that can be achieved through various levels of vertical model structure (direct runoff, upper soil moisture storage, and the percolation process). This was accomplished through application and calibration of a CRR model, HYMOD, with a variety of different combinations of excess generation (interception storage, tension storage, free storage, etc.), percolation, and streamflow routing functions. In this study, it has been demonstrated how multi-criteria methods provide a useful framework for the systematic investigation of appropriate model complexity. In addition, the applicability of the multi-criteria automatic calibration methods to the calibration of CRR models with increased model complexity has been demonstrated in this study. Manual calibration of the different model structures would have required a significant effort since little was known (no "expert" knowledge) about the behavior and performance of many of the model structures prior to testing. Instead, the automatic approach allowed efficient and "consistent" estimation of parameters (and hence model performance) with a minimal amount of effort (5-10 minutes for each case on a Sun workstation).

The effectiveness and efficiency of the automatic approach allowed rapid investigation of the specific benefits associated with different levels of vertical model structural complexity, including impervious area contribution, UZ tension and free water storage, and percolation computations. In particular it was found that (a) none of the model configurations were simultaneously able to match all three criteria as well as the SAC-SMA model; (b) in all cases, each model had at least some solutions that were superior to all of the SAC-SMA model in terms of the FS criterion; (c) in general, the tension and impervious components improved the performance of the simple bucket model and decreased the performance of the Moore loss model; (d) the best model combination for the simple bucket and Moore loss models were able to fit the non-driven quick and non-driven slow flows at least as good (better for the Moore loss model in terms of the non-driven slow flows) as the SAC-SMA model; (e) the performance of the simple bucket was superior when percolation was taken from the saturated excess rather than the UZ free water storage unless UZ tension water storage and impervious area components were added – the Moore loss model did not exhibit an obvious preference. These results are specific to the model configurations tested on the Leaf River data set. Based on the results from this study and those presented in Chapter 3, the next logical step would be to perform a study in which the spatial resolution and vertical structural complexity of the CRR model application is investigated simultaneously.

CHAPTER FIVE

CONCLUSIONS AND FURTHER RESEARCH

5.1 SUMMARY

The primary objective of this research was to develop and test a new hybrid multi-criteria calibration approach that combines the strength of automatic and manual calibration methods. The new approach was used to investigate the benefits of different levels of spatial and vertical representation of important watershed hydrologic variables with conceptual rainfall runoff models.

The first step was to understand the similarities and differences between the manual "expert" approach and the automatic "systems" approach to calibration of CRR models. The most important aspect of the manual approach is the consideration of specific aspects of the hydrologic behavior in the evaluation of the model performance during the parameter estimation procedure. The automatic procedures rely on objective mathematical measures (usually only one) in the performance evaluation and parameter adjustment procedures. The second step was to develop objective measures that emulate the manual approach by partitioning the streamflow hydrograph into three components, "driven" (QD), "non-driven-quick" (QQ), and "non-driven-slow" (QS), and estimating the closeness between the model outputs and the corresponding observed values separately with mathematical measures. A computer algorithm, MOCOM-UA, is used to search for the Pareto solution space containing the "good" solutions to the problem.

Finally, the new approach was used in two separate studies to investigate the level of spatial and vertical detail of important hydrologic processes within a watershed that needs to be represented by a conceptual rainfall-runoff (CRR) model in order to accurately simulate the streamflow. Results from each of the studies provide insights into the complicated relationships between model complexity, problem complexity, and information content of the data. Specifically the following important findings were made:

- The time and effort required to estimate the parameter range representing the trade-offs in the performance of the model are dramatically reduced from the manual approach.
- The new approach may serve as an aid to the hydrologist in analysis and elimination of the solutions within the trade-off range.
- Results from the studies indicate that parameter sets selected from within the Pareto region tend to provide consistent and reliable model forecasts.
- The properties of the Pareto region provide information useful for evaluating the limitations of the various components of the watershed model, thereby pointing towards potential structural improvements.
- The new approach provides a stronger test of model performance than methods that use a single overall statistic to aggregate model errors over a large range of hydrologic behaviors.
- Two case studies helped to demonstrate how multi-criteria methods provide a useful framework for the systematic investigation of appropriate spatial and vertical model complexity. The effectiveness and efficiency of the automatic approach enabled rapid investigate the specific benefits

associated with different levels of spatial and vertical representation of important watershed processes.

- The automatic approach was shown to be able to effectively handle the increased model complexity and provide parameter estimates (and hence model performance) that are comparable or superior to those obtained by the manual/expert approach.
- The improved results are also achieved at a considerable savings in "calibration time"–after problem setup.

5.2 RECOMMENDATIONS FOR FUTURE RESEARCH

The issues of "which measures" and "how many measures" are appropriate for use in the evaluation of CRR model performance are important. In this study, three characteristic watershed behaviors were selected to emulate the visual inspection of the results performed by an "expert" in manual calibration. The use of these three simple measures was intended to serve as a "proof of concept" that the manual process can be successfully emulated with a computerized multi-criteria approach. Certainly, however, it is reasonable to expect that more sophisticated partitioning schemes could improve the approach and the results; note that there are a wide range of different techniques based on physical reasoning available for base flow separation that could be used to improve the partitioning of the non-driven component. Further investigation of different approaches to partitioning the observed hydrograph into the three behaviors should be performed.

Further efforts to understand the sensitivity of the results to the number and type of criteria would also be valuable. In this research only three criteria were used in the calibration procedure, however, at least two (overall RMSE and BIAS) were used to select from among the 500 Pareto solutions. A popular criticism, expressed by several reviewers, of this work is that the overall BIAS measure was not included (either in place of one of the three measures or in addition to the three measures) during the optimization portion of the approach. A reasonable response to this criticism is that the approach was specifically developed to emulate the expert manual procedures developed by the NWS which do not rely on overall measures during the Level Two portion of the procedures (overall measures are only used to evaluate potential solutions determined in the Level Two procedure). A more philosophical response, however, is that there may not be much additional information gained from the use of the overall BIAS measure during the optimization procedure. While it is clearly desirable to have a solution with an overall BIAS that is very low (as close to zero as possible) it may not be possible to distinguish between two very different solutions that have the same value of overall BIAS (e.g., one that is %50 greater and one that is %50 less than every observed flow value) and, hence, there may not be valuable information gained through the use of the overall BIAS measure during the calibration procedure. These responses, however, certainly do not put the issue to rest and, as a result, further investigation of the use of overall measures should be a focus of future research.

With respect to the number of measures that are appropriate for use in the evaluation of CRR model performance it is clear from the results in this study that a single objective measure does not provide an adequate measure of model performance (at least in the cases tested in this study). In Chapter Two it was demonstrated that the use of a single measure appears to support the concept of equifinality. However, the use of multiple measures (in this case three) clearly highlights the model's inability to simultaneously simulate different characteristics of the observed hydrograph, hence undermining the validity of the equifinality assumption based on a single measure. Further, it was demonstrated that the use of a single criterion (e.g., RMSE) may predispose the "optimal" solution to significant bias while providing little if any information on the performance trade-offs of the model. Despite these findings, there is still much to be learned about the number of measures needed to adequately assess the performance of CRR models. It would be very interesting to conduct an experiment aimed at developing a better understanding of the relationship between the number of performance measures and the information gained in terms of the performance of the model.

There is an increasing awareness in the hydrologic modeling community that the information content in the observed data to identify model structure and associated parameter values is limited. On the other hand, new and innovative methods to measure the spatial distribution of important hydrologic variables are continually being developed. As a result, there is a need for a framework to guide model development and application in a way that maximizes the use of information contained within the data while matching the model structure to the available data. The insight gained from the investigation of spatial and vertical model structural complexity in this study should be used to develop a framework for development and application of parsimonious hydrologic models. This would require additional studies that focused simultaneously on spatial and vertical model complexity on a large number of watersheds under a variety of different hydrologic conditions. The Distributed Model Intercomparison Project (DMIP), proposed by the Hydrology Laboratory (HL) of the National Weather Service (NWS) is an excellent setting for the development of such a framework. The intent of DMIP is to invite the academic community and other researchers to help guide the NWS's distributed modeling research by participating in a comparison of distributed models applied to many different test data sets.

REFERENCES

Anderson, E.A. (1997). "Hydrologic model calibration using the Interactive Calibration Program (ICP)." Hydrologic Research Laboratory, U.S. National Weather Service, Silver Spring, Maryland.

Bastidas, L.A., H.V. Gupta, and S. Sorooshian (1999). "The Multi-Objective COMplex evolution algorithm, MOCOM-UA, User's Guide." Department of Hydrology and Water Resources, The University of Arizona, Tucson Arizona.

Bergstrom, L. (1976) "Development and application of a conceptual runoff model for Scandinavian catchments." *SMHI Report, RHO No. 7*, Swedish Meteorological and Hydrological Institute, Norrkoping, Sweden.

Beck, M.B., F.M. Kleissen, and H.S. Wheater (1990). "Identifying flow paths in models of surface water acidification." *Reviews of Geophysics*, 28(2), 207-230.

Beven, K.J. and M.J. Kirkby (1979). "A physically-based variable contributing area model of basin hydrology." *Hydrol. Sci. J.*, 24(1), 43-69.

Beven, K.J. and A.M. Binley (1992). "The future of distributed models: model calibration and predictive uncertainty." *Hydrological Process*, 6, 279-298.

Boyle, D.P., H.V. Gupta, and S. Sorooshian, "Comparison of the SAC-SMA and HyMod Conceptual Flood Forecasting Models." Hydrologic Research Lab, U.S. National Weather Service, Silver Spring, MD, November 17-19, 1997.

Boyle, D.P., H.V. Gupta, and S. Sorooshian, (2000). "Towards improved calibration of hydrologic models: Combining the strengths of manual and automatic methods." *Water Resources Research*, 36(12), 3663-3674.

Boyle, D.P., H.V. Gupta, S. Sorooshian, V. Koren, Z. Zhang, and M. Smith, (2001) "Towards improved streamflow forecasts: The value of semi-distributed modeling," *Water Resources Research*, AGU, (in press).

Brazil, L.E. (1988). " Multilevel calibration strategy for complex hydrologic simulation models." Ph.D. Dissertation, 217p., Colorado State University, Fort Collins.

Brazil, L.E. and M.D. Hudlow (1981). " Calibration procedures used with the National Weather Service Forecast System." in Water and Related Land Resource Systems, Edited by Y.Y. Haimes and J. Kindler, pp. 457-466, Pergamon Press, New York.

Brazil, L.E. and W.F. Krajewski (1987). " Optimization of complex hydrologic models using random search methods." Paper presented at Conference on Engineering Hydrology, Hydraulics Division, American Society of Civil Engineering, Williamsburg, Virginia, Aug. 3-7.

Burnash, R. J.C., R.L. Ferral, and R.A. McGuire (1973). " A generalized streamflow simulation system: conceptual modeling for digital computers." U.S. Department of Commerce, National Weather Service and State of California, Department of Water Resources. Sacramento, California.

Duan, Q., V.K. Gupta, and S. Sorooshian (1992). "Effective and efficient global optimization for conceptual rainfall-runoff models." *Water Resources Research*, 28(4), 1015-1031.

Duan, Q., V.K. Gupta, and S. Sorooshian (1993). "A shuffled complex evolution approach for effective and efficient global minimization." *Journal of Optimization Theory and Applications*, 76(3), 501-521.

Finnerty, B. and D. Johnson (1997). "Comparison of mean areal precipitation estimates derived from NEXRAD radar vs. rain gage networks." Presented at ASCE 27th Congress of the Int'l Assoc. for Hydraulic Research, San Francisco, CA.

Gan, T.Y. and G.F. Biftu (1996). "Automatic calibration of conceptual rainfall-runoff models: Optimization algorithms, catchment conditions, and model structure." *Water Resources Research*, 32(12), 3513-3524.

Goldberg, D.E., (1989) <u>Genetic Algorithms in search, optimization, and machine learning</u>. Addison-Wesley, Reading, Mass., 412 pp.

Grayson, R.B. and F.H.S. Chiew (1994). "An approach to model selection." Presented at Hydrology and Water Resources Symposium, Institute of Engineering, Austrailia.

Gupta, H.V., S. Sorooshian, and P.O. Yapo (1998) "Towards improved calibration of hydrologic models: multiple and non-commensurable measures of information." *Water Resources Research*, 34(4), 751-763.

Gupta, H.V., S. Sorooshian, and P.O. Yapo (1999) "Status of automatic calibration for hydrologic models: comparison with multilevel expert calibration." *Journal of Hydrologic Engineering*, 4(2), 135-143.

Harlin, J. (1991) "Development of a process oriented calibration scheme for the HBV hydrologic model." *Nordic Hydrology*, 22, 15-36.

Jakeman, A.J. and G.M. Hornberger (1993). "How much complexity is warranted in a rainfall-runoff model?" *Water Resources Research*, 29(8), 2637-2649.

James, L.D. and S.J. Burges (1982). "Selection, calibration, and testing of hydrologic models." Hydrologic Testing of Small Watersheds, C.T. Hann, H.P. Johnson, and D.L. Brakensiek, Editors, American Society of Agricultural Engineers, St. Joseph, Michigan, 435-472.

Koren, V.I., M. Smith, D. Wang, and Z. Zhang (2000). "Use of soil property data in the derivation of conceptual rainfall-runoff model parameters." Preprints, 15th Conference on Hydrology, American Meteorological Society, Long Beach, California, 103-106.

Kuczera, G. (1983a). "Improved parameter inference in catchment models: 1. evaluating parameter uncertainty." *Water Resources Research*, 19, 1151-1162.

Kuczera, G. (1983b). "Improved parameter inference in catchment models: 2. combining different kinds of hydrologic data and testing their compatibility." *Water Resources Research*, 18, 1163-1172.

Kuczera, G. (1997). "Efficient subspace probabilistic parameter optimization for catchment models." *Water Resources Research*, 33, 177-185.

Lindstrom, G., (1997). "A simple automatic calibration routine for the HBV model". *Nordic Hydrology*, 28(3), 153-168.

Linsley, R.K., M.A. Kohler, and J.L.H. Paulhus (1949). *Aplied Hydrology*. McGraw-Hill, New York.

Mayeux, B.C., (1995). "The relative importance of rainfall intensity versus saturated hydraulic conductivity for hydrologic modeling of semi-arid watersheds", Unpublished Master of Science Thesis, *Department of Hydrology and Water Resources, University of Arizona*.

Michaud, J. and S. Sorooshian (1994). "Comparison of simple versus complex distributed runoff models on a midsized semiarid watershed." *Water Resources Research*, AGU, 30(3), 593-605.

Naef, F., (1981) "Can we model the rainfall-runoff process today? *Hydrol. Sci. Bull.*, 26(3), 281-289.

Nathan, R.J. and T.A. McMahon (1990). "Evaluation of automated techniques for base flow and recession analyses." *Water Resources Research*, 26(7), 1465-1473.

National Weather Service, (1997). National Weather Service River Forecast System User's Manual. Part IV.4.1, NWS, Silver Spring, MD.

Niadas, I.A. (1999) "Rainfall-runoff model regionalisation: An application to Midwest USA." M.Eng. Thesis, Department of Civil and Environmental Engineering, Imperial College, United Kingdom.

NOAA, (1982). Atlas of NOAA Free Water Evaporation. National Oceanic and Atmospheric Administration (NOAA) Technical Report NWS 33.

Ogden, F.L. and P.Y. Julien (1993). "Runoff sensitivity to temporal and spatial rainfall variability at runoff plane and small basin scales." *Water Resources Research*, 29(8), 2589-2597.

Ogden, F.L. and P.Y. Julien (1994). "Runoff model sensitivity to radar rainfall resolution." *Journal of Hydrology*, 158, 1-18.

Peck, E. L., (1976) "Catchment Modeling and Initial Parameter Estimation for the National Weather Service River Forecast System." NOAA Technical Memorandum NWS HYDRO-31. Office of Hydrology. Washington, D.C. National Weather Service. Silver Spring, Maryland.

Refsgaard, J.C. and J. Knudsen (1996). " Operational validation and intercomparison of different types of hydrologic models." *Water Resources Research*, AGU, 32(7), 2189-2202.

Rutledge, A.T. (1993). Computer Programs for Describing the Recession of Ground-water Discharge and for Estimating Mean Ground-water Recharge and Discharge from Streamflow Records. Water Resources Investigations Report 93-4121, U.S. Geological Survey, 45 pp.

Crawford, N.H. and R.K. Linsley (1966). *Digital simulation in hydrology: Stanford watershed model IV. Technical Report 39*. Stanford, CA.

Sittner, W.T., C.E. Schauss, and J.C. Monro (1969). "Continuous hydrograph synthesis with an API-type hydrologic model." *Water Resources Research*, AGU, 5(5), 1007-1022.

Smith, M.B., V. Koren, D. Johnson, B.D. Finnerty, and D.J. Seo (1999). "Distributed modeling: Phase 1 results." NOAA Technical Report NWS 44, National Hydrologic Research Lab, 210p.

Smith, M.B., V. Koren, Z. Zhang, D. Wang, and S. Reed (2000). "Semi-distributed vs lumped model simulations: Comparisons using observed data for RFC scale basins." *EOS, Transactions*, American Geophysical Union, Spring Meeting Supplement, 81(19). S214.

Sorooshian, S. and J.A. Dracup (1980). "Stochastic parameter estimation procedures for hydrologic rainfall-runoff models: correlated and heteroscedastic error cases." *Water Resources Research*, 16(2), 430-442.

Sorooshian, S. and V.K. Gupta (1983). "Automatic calibration of conceptual rainfall-runoff models: the question of parameter observability and uniqueness." *Water Resources Research*, 19(1), 251-259.

Sorooshian, S., Q. Duan, and V.K. Gupta (1993). "Calibration of rainfall-runoff models: application of global optimization to the Sacramento soil moisture accounting model." *Water Resources Research*, 29, 1185-1194.

Sugawara, M., I. Watanabe, E. Ozaki, and Y. Katsuyama (1984). Tank Model with Snow Component. Research notes of the National Research Center for Disaster Prevention, 293 p., National Research Center for Disaster Prevention , Japan.

Thyer, M., G. Kuczera, and B.C. Bates (1999). "Probabilistic optimization for conceptual rainfall-runoff models: a comparison of the shuffled complex evolution and simulated annealing algorithms." *Water Resources Research*, 35, 767-773.

Wang, Q.J. (1991). "The genetic algorithm and its application to calibrating conceptual rainfall-runoff models." *Water Resources Research*, 27(9), 2467-2471.

Whitley, D. (1989). "The genitor algorithm and selection pressure: Why rank-based allocation of reproductive trials is best." Paper presented at Third International Conference on Genetic Algorithms.

Yapo, P.O., H.V. Gupta, and S. Sorooshian (1997). *A multiobjective global optimization algorithm with application to calibration of hydrologic models, HWR Technical Report No. 97-050.* Department of Hydrology and Water Resources, The University of Arizona, Tucson, Arizona.

Yapo, P.O., H.V. Gupta, and S. Sorooshian (1998). "Multi-objective global optimization for hydrologic model." *Journal of Hydrology*, 204, 83-97.

Wagener, T., M.J. Lees, and H.S. Wheater (1999). "A generic rainfall-runoff modeling toolbox." *EOS, Transactions*, American Geophysical Union, Fall Meeting Supplement, 80(17). F203.

Wagener, T., D.P. Boyle, M.J. Lees, H.S. Wheater, H.V. Gupta, and S. Sorooshian, "A framework for development and application of hydrological models," *Hydrology and Earth System Science*, European Geophysical Society, (in press).

Wheater, H.S., K.H. Bishop, and M.B. Beck (1986). "The identification of conceptual hydrological models for surface water acidification." *Hydrological Process*, 1, 89-109.

Wheater, H.S., A.J. Jakeman, and K.J. Beven (1993). "Chapter 5–Progress and directions in rainfall-runoff modeling." In Modeling change in environmental systems (Eds. A.J. Jakeman, M.B. Beck, and M.J. McAleer), Wiley, Chichester, 101-132.

Woolhiser, D.A., R.E. Smith, and D.C. Goodrich (1990). A kinematic runoff and errosion manual: Documentation and user manual. ARS 77. U.S. Dep. Of Agric., Washington, D.C.,130pp

Zhang, X. and G. Lindstrom, (1997) "Development of an automatic calibration scheme for the HBV hydrological model." *Hydrological Processes*, 11, 1671-1682.