A TOP-DOWN FRAMEWORK FOR WATERSHED MODEL EVALUATION AND SELECTION UNDER UNCERTAINTY

A Thesis in
Civil Engineering
by
Yaoling Bai

© 2008 Yaoling Bai

Submitted in Partial Fulfillment of the Requirements for the Degree of Master of Science

August 2008
The thesis of Yaoling Bai was reviewed and approved* by the following:

Thorsten Wagener
Assistant Professor of Civil Engineering
Thesis Co-Advisor

Patrick M. Reed
Associate Professor of Civil Engineering
Thesis Co-advisor

Michael Gooseff
Assistant Professor of Civil Engineering

Peggy Johnson
Professor of Civil Engineering
Head of the Department of Civil and Environmental Engineering

*Signatures are on file in the Graduate School
Abstract

This study presents a novel top-down strategy for model evaluation and selection under uncertainty. It extends the top-down approach suggested by Klemes and formalized by Sivapalan and colleagues through a Monte Carlo framework in which model ensembles are tested for four different watershed response signatures across time scales. Lumped watershed model structures of increasing complexity have been applied to twelve watersheds across a range of hydro-climatic conditions within the US. The necessary minimum complexity and the related model assumptions provide indicators of the dominant controls on the watershed response at each temporal scale represented by different water balance signatures. Probabilistic measures of model performance with respect to reliability (Is the model ensemble capturing the observed signature?) and with respect to shape (Is the model structure capable of representing the signature variability?) have been developed to distinguish the ability of the models to represent watershed response behavior for each time-scale. The probabilistic measures are combined in a fuzzy rule system to guide model selection. Results suggest that the framework can be tuned to function as a screening tool that formalizes our model selection process. This fuzzy model selection framework therefore enhances our ability to objectively and automatically select parsimonious model structures for large databases of watersheds.
# TABLE OF CONTENTS

LIST OF FIGURES ............................................................................................................. v  
LIST OF TABLES ........................................................................................................... viii  
LIST OF NOTATION ....................................................................................................... ix  

ACKNOWLEDGEMENT ................................................................................................. xi  

Chapter 1. INTRODUCTION .............................................................................................. 1  

Chapter 2. METHODOLOGY ............................................................................................. 5  
  2.1 Water balance and model components ................................................................. 5  
  2.2 Modeling framework ............................................................................................. 6  
    2.2.1 Previous studies ............................................................................................ 6  
    2.2.2 Monte Carlo framework ................................................................................ 6  
  2.3 Modular modeling structure ................................................................................... 8  
  2.4 Parameter estimation ............................................................................................ 13  
  2.5 Signatures ............................................................................................................. 15  
  2.6 Measures of acceptability .................................................................................... 15  
  2.7 Fuzzy rules ........................................................................................................... 17  

Chapter 3. WATERSHEDS AND DATA ......................................................................... 19  
  3.1 Watersheds and data ............................................................................................ 19  

Chapter 4. RESULTS AND DISCUSSION ...................................................................... 22  
  4.1 Model performance with respect to reliability ................................................... 22  
  4.2 Model performance with respect to shape ...................................................... 29  
  4.3 Fuzzy evaluation of model performance ............................................................ 32  
  4.4 Discussion ........................................................................................................... 38  

Chapter 5. CONCLUSIONS .............................................................................................. 43  

Chapter 6. RECOMMENDATIONS FOR FUTURE ........................................................ 44  

REFERENCES .................................................................................................................. 46  

APPENDIX A: Model structures and model equations ..................................................... 50  
APPENDIX B: Feasible range and \textit{a priori} range of parameters ..................................... 58  
APPENDIX C: Estimation of \textit{a priori} parameters ........................................................ 59  
APPENDIX D: Further Results ......................................................................................... 61
LIST OF FIGURES

Figure 1. Flow chart of proposed model selection framework ..................................................8

Figure 2. Modules with increasing complexity used in the model structures investigated ...........................................................................................................................................................................9

Figure 3. Eight models were derived using different combinations of the available SMA, ET and Routing modules ..................................................................................................................................................................13

Figure 4. Membership functions (schematic) for reliability and shape measures ............18

Figure 5. Locations of the 12 MOPEX watersheds in the US (a), and on the Budyko curve (b). AE is actual evapotranspiration, P is precipitation and PE is potential evapotranspiration ..............................................................................................................................................20

Figure 6. (a) & (b) Daily streamflow hydrographs of 12 watersheds in 1961. (c) Flow duration curves of 12 watersheds using daily streamflow data for period 1960 to 1998.....................................................................................................................................21

Figure 7. Ensemble prediction ranges of inter-annual variability. Ensembles produced by parameter sets drawn from feasible and a priori ranges for the (a) and (b) Guadalupe watershed, (c) and (d) East Fork White watershed, and (e) and (f) French Broad watershed respectively. The vertical axis represents annual runoff over average annual runoff ........................................................................................................................................23

Figure 8. Ensemble prediction ranges of intra-annual variability of streamflow. Ensembles produced by parameter sets drawn from feasible and a priori ranges for the (a) and (b) Guadalupe watershed, (c) and (d) East Fork White watershed, and (e) and (f) French Broad watershed respectively. The vertical axis represents average monthly runoff over average annual runoff ........................................................................................................................................25

Figure 9. Ensemble prediction ranges for the flow duration curve. Ensembles produced by parameter sets drawn from feasible and a priori ranges for the (a) and (b) Guadalupe watershed, (c) and (d) East Fork White watershed, and (e) and (f) French Broad watershed respectively ........................................................................................................................................26

Figure 10. Ensemble prediction ranges for daily streamflow in the year 1961 on logarithmic scale. Ensembles produced by parameter sets drawn from feasible and a priori ranges for the (a) and (b) Guadalupe watershed, (c) and (d) East Fork White watershed, and (e) and (f) French Broad watershed respectively ........................................................................................................................................28

Figure 11. Simulations of the best shape reproduction by models S1 to S4 with respect to the observed flow duration curve. Ensembles produced by parameter sets drawn from feasible and a priori ranges for the (a) and (b) Guadalupe watershed, (c) and (d) East Fork White watershed, and (e) and (f) French Broad watershed respectively ........................................................................................................................................31
Figure 12. Fuzzy evaluation of model performance for the 12 watersheds based on model ensembles of signatures derived from their feasible parameter ranges. ............................36

Figure 13. Fuzzy evaluation of model performance for the 12 watersheds based on model ensembles of signatures derived from their \textit{a priori} parameter ranges. ............................37

Figure A1. Schematic describing model S1 .......................................................................50

Figure A2. Schematic describing model S2 .......................................................................51

Figure A3. Schematic describing model S3 .......................................................................52

Figure A4. Schematic describing model S4 .......................................................................54

Figure A5. Schematic describing the multiple-bucket model ............................................58

Figure D1. Simulations of the best shape reproduction by models S1 to S4 with respect to the observed inter-annual variability. Ensembles produced by parameter sets drawn from feasible and \textit{a priori} ranges for the (a) and (b) Guadalupe watershed, (c) and (d) East Fork White watershed, and (e) and (f) French Broad watershed respectively. The vertical axis represents annual runoff over average annual runoff. ......................................61

Figure D2. Simulations of the best shape reproduction by models S1 to S4 with respect to the observed intra-annual variability. Ensembles produced by parameter sets drawn from feasible and \textit{a priori} ranges for the (a) and (b) Guadalupe watershed, (c) and (d) East Fork White watershed, and (e) and (f) French Broad watershed respectively. The vertical axis represents average monthly runoff over average annual runoff. ........................................62

Figure D3. Simulations of the best shape reproduction by models S1 to S4 with respect to the observed daily streamflow in the year 1961. Ensembles produced by parameter sets drawn from feasible and \textit{a priori} ranges for the (a) and (b) Guadalupe watershed, (c) and (d) East Fork White watershed, and (e) and (f) French Broad watershed respectively. ..............................63

Figure D4. Simulations of the best shape reproduction by models S1 to S4 with respect to the observed daily streamflow in the year 1961 on logarithmic scale. Ensembles produced by parameter sets drawn from feasible and \textit{a priori} ranges for the (a) and (b) Guadalupe watershed, (c) and (d) East Fork White watershed, and (e) and (f) French Broad watershed respectively ..........................................................64

Figure D5. Simulations of the best shape reproduction by models M1 to M4 with respect to the observed inter-annual variability. Ensembles produced by parameter sets drawn from feasible and \textit{a priori} ranges for the (a) and (b) Guadalupe watershed, (c) and (d) East Fork White watershed, and (e) and (f) French Broad watershed respectively. The vertical axis represents annual runoff over average annual runoff .............................................65
Figure D6. Simulations of the best shape reproduction by models M1 to M4 with respect to the observed intra-annual variability. Ensembles produced by parameter sets drawn from feasible and a priori ranges for the (a) and (b) Guadalupe watershed, (c) and (d) East Fork White watershed, and (e) and (f) French Broad watershed respectively. The vertical axis represents average monthly runoff over average annual runoff.

Figure D7. Simulations of the best shape reproduction by models M1 to M4 with respect to the observed flow duration curve. Ensembles produced by parameter sets drawn from feasible and a priori ranges for the (a) and (b) Guadalupe watershed, (c) and (d) East Fork White watershed, and (e) and (f) French Broad watershed respectively.

Figure D8. Simulations of the best shape reproduction by models M1 to M4 with respect to the observed daily streamflow in the year 1961. Ensembles produced by parameter sets drawn from feasible and a priori ranges for the (a) and (b) Guadalupe watershed, (c) and (d) East Fork White watershed, and (e) and (f) French Broad watershed respectively.

Figure D9. Simulations of the best shape reproduction by models M1 to M4 with respect to the observed daily streamflow in the year 1961 on logarithmic scale. Ensembles produced by parameter sets drawn from feasible and a priori ranges for the (a) and (b) Guadalupe watershed, (c) and (d) East Fork White watershed, and (e) and (f) French Broad watershed respectively.

Figure D10. Ensemble prediction ranges for inter-annual variability produced by model M1 to M4. Ensembles produced by parameter sets drawn from feasible and a priori ranges for the (a) and (b) Guadalupe watershed, (c) and (d) East Fork White watershed, and (e) and (f) French Broad watershed respectively. The vertical axis represents annual runoff over average annual runoff.

Figure D11. Ensemble prediction ranges for intra-annual variability produced by model M1 to M4. Ensembles produced by parameter sets drawn from feasible and a priori ranges for the (a) and (b) Guadalupe watershed, (c) and (d) East Fork White watershed, and (e) and (f) French Broad watershed respectively. The vertical axis represents average monthly runoff over average annual runoff.

Figure D12. Ensemble prediction ranges for flow duration curve produced by model M1 to M4. Ensembles produced by parameter sets drawn from feasible and a priori ranges for the (a) and (b) Guadalupe watershed, (c) and (d) East Fork White watershed, and (e) and (f) French Broad watershed respectively.

Figure D13. Ensemble prediction ranges for daily streamflow in the year 1961 on logarithmic scale produced by model M1 to M4. Ensembles produced by parameter sets drawn from feasible and a priori ranges for the (a) and (b) Guadalupe watershed, (c) and (d) East Fork White watershed, and (e) and (f) French Broad watershed respectively.
LIST OF TABLES

Table 1: Characteristics of twelve MOPEX watersheds used ...........................................20

Table 2: Reliability of four signatures over both feasible ranges and a priori ranges for the Guadalupe, East Fork White and French Broad watersheds. All reliability values show as percents (%) ..............................................................................................29

Table 3: The shape measures of four signatures over both feasible ranges and a priori ranges for Guadalupe, East Fork White and French Broad. The ideal shape value is 1.00 that indicates simulations are perfectly parallel to observations. The shape value of the best simulations with respect to the shape measure is chosen to represent the model’s ability to reproduce the dynamics of signatures .................................................................32

Table 4: Membership function thresholds for the shape measure at each time scale .......35
LIST OF NOTATION

$\theta_{fc}$  field capacity (dimensionless)
$\theta_{wlt}$ permanent wilting point (dimensionless)
$\phi$ porosity (dimensionless)
$f_c$ threshold storage parameter (dimensionless) ($0 < f_c < 1$)
$S_b$ maximum storage of the bucket model (mm)
$S_{fc}$ threshold storage (mm).
$P$ precipitation (mm/d)
$E_p$ potential evapotranspiration (mm/d)
$E$ actual evapotranspiration (mm/d)
$E_i$ interception (mm/d)
$E_v$ vegetation transpiration (mm/d)
$E_{bs}$ bare soil evaporation (mm/d)
$E_{v us}$ transpiration from unsaturated zone (mm/d)
$E_{v sat}$ transpiration from saturated zone (mm/d)
$E_{bs us}$ evaporation from unsaturated zone (mm/d)
$E_{bs sat}$ evaporation from saturated zone (mm/d)
$S_{sat}$ soil water storage in saturated zone (mm)
$S_{us}$ soil water storage in unsaturated zone (mm)
$S_{usfc}$ field capacity of current unsaturated zone (mm)
$S_{deep}$ soil water storage in deep store (mm)
$S_t$ total soil water storage at current time $t$ (mm)
$S_{t-1}$ total soil water storage at last time step $t-1$ (mm)
$S_{sat \ t-1}$ soil water storage of saturated zone at last time step $t-1$ (mm)
$r_p$ daily recharge to saturated zone from unsaturated zone in which water storage exceeds field capacity (mm)
$r_g$ daily recharge from upper saturated zone to deeper store (mm)
$Q$ total runoff (mm/d)
$Q_{se}$ surface runoff generated by saturation excess (mm/d)
$Q_{ss}$ subsurface flow originating from saturated zone (mm/d)
$Q_{bf}$ base flow originating from deep store (mm/d)
$M$ fraction of catchment area covered by deep rooted vegetation (dimensionless)
$K_v$ vegetation transpiration efficiency (dimensionless)
$\alpha_{ss}$ recession coefficient for subsurface flow from saturated zone store in the linear storage-outflow model (day$^{-1}$)
\( \alpha_{bf} \) recession coefficient for subsurface flow from deep store in the linear storage-outflow model (day\(^{-1}\))

\( k_d \) deep recharge coefficient from the upper saturated zone to the deep store (day\(^{-1}\))

\( b \) shape parameter for spatial soil water storage distribution in the multi-bucket model

\( F \) cumulative probabilities at which soil water storages of 10 buckets fit spatial soil water storage distribution

\( S_{\text{max}} \) maximum soil water storage in the watershed

\( S_{b,f} \) soil water storage capacities in the 10 buckets
ACKNOWLEDGEMENT

I would like thank my thesis advisors, Dr. Thorsten Wagener and Dr. Patrick Reed, for providing me with valuable guidance and assistance throughout my studies. I also appreciate Dr. Michael Gooseff for taking the time to serve on my thesis committee and contribute his valuable input and suggestions.

I appreciate Tang Yong, Shuangcai, Zhenxing, Katie, Geoffrey, Rashi and the other fellow students for their help with my studies.

I am thankful for the love and support I have received from my parents, sister and brother. They give me my most precious and important possession, their lifelong unconditional love and acceptance. I wish to express my deepest appreciation to my family.
Chapter 1: INTRODUCTION

Hydrologic models are important tools for testing hypotheses about watershed behavior and for predicting the response of hydrologic systems. Choosing the appropriate model structure is a crucial step in hydrologic modeling in order to accurately predict streamflow or other variables, and to understand the dominant physical controls on watershed responses. In the context of hypothesis testing, we want to select the model structure with the minimum level of complexity that is capable of reproducing the observed watershed response.

Bottom-up and top-down approaches are two different strategies used for model identification. In the bottom-up approach, modelers formulate and combine the model components based on an *a priori* physically-based formulation of continuum-scale processes. The top-down model approximates hydrologic behavior at the lumped watershed-scale and model structures are evaluated against observed watershed streamflow characteristics. Bottom-up models generally require detailed information on the physical characteristics of the watershed under study, and, due to the need for spatially detailed modeling, often yield the potential for over-parameterization when predicting streamflow.

Klemes (1983) suggested a systematic approach to identify the appropriate model structure for a given case through a process of hypothesis testing, termed the top-down or downward modeling strategy. The basic idea is to capture the hydrologic response at a given temporal scale with the minimum level of model complexity necessary and to introduce additional state variables and parameters when moving to finer scales.

Sivapalan and colleagues (Jothityangkoon et al., 2001; Atkinson et al., 2002, 2003;
Farmer et al., 2003; Eder et al., 2003; Son and Sivapalan, 2007) in a series of studies, developed a collection of lumped models with increasing complexity to simulate water balance signatures across temporal scales as an implementation of the top-down strategy. The lumped model structures represent different hypotheses about how watersheds control streamflow response. Prior studies have employed visual evaluations of the model performance, using a priori parameter estimates based on physical watershed characteristics to choose the most parsimonious structure capable of representing the watershed behavior at each scale. These studies resulted in qualitative relationships between model complexity (and thus hypotheses of watershed function), climate conditions (the watersheds analyzed were located in different climatic regimes) and time scales (annual, monthly and daily). This ultimately provided some insight into climate, soil and landscape controls at different time scales. Hickel and Zhang (2006) used the top-down approach to estimate the impact of rainfall seasonality on the watershed-scale mean annual water balance. The results showed that the effects of seasonal change of water storage on the mean annual evapotranspiration needs to be considered to improve evapotranspiration predictions for watersheds with winter-dominant rainfall. Young (2003, 1998) presented a data-based mechanistic modeling methodology as an example of how the top-down modeling philosophy at the watershed scale can be implemented. The data-based mechanistic modeling strategy identifies the appropriate routing component, for a chosen soil moisture accounting component, using an instrumental variable strategy. While this limits the routing component to combinations of linear transfer functions, it avoids over-parameterization by trading off model performance and parameter uncertainty by calculating Young’s information criterion (Young, 2003; 1998).
Additionally, a routing component is only acceptable if it is physically interpretable by the modeler. Littlewood et al. (2003) compared top-down and bottom-up approaches and showed the advantages that the top-down modeling approach can reduce data requirements and limit model complexity.

Manual top-down approaches that require visual examination of simulation performance to select model structures are limited, by their lack of transferability and their time demands if large numbers of watersheds must be analyzed. Some studies have relied on statistical measures that comparing simulated responses with observed response (Atkinson et al., 2002). However, these approaches typically ignore parametric uncertainties and instead focus on deterministic simulations.

In this study we contribute a new framework for top-down model evaluation and selection that facilitates a more objective quantitative classification of model performance that considers each model’s given parametric uncertainty. We separately evaluate each model’s ability to capture the range of streamflow observed and the temporal characteristics of a watershed’s response. These independent performance measures are combined in a multi-objective fuzzy classification of model performance to achieve consistent decisions in selecting the appropriate model complexity for a given time scale. This study contributes an analysis of the impacts of \textit{a priori} data on the two measures and elucidates the limitations associated with a range of model structures. We apply our proposed fuzzy top-down model evaluation and selection method to twelve US watersheds that compose a strong hydro-climatic gradient (Van Werkhoven et al., 2008). By sampling watersheds across a range of climatic conditions our results demonstrate the ability of fuzzy top-down modeling to clarify the dominant controls on hydrologic
behavior for a range of time scales. Recent paper by Clark (2008) on designing model evaluation calls for the need to evaluate large ensembles of watersheds. This study is the first step towards large scale efforts to classify model performance for hundreds or thousands of watersheds at very large scales. Ultimately this approach will allow us to choose a model structure based on similarity in watershed physical and climatic conditions, even before observations of the watershed response are available.
Chapter 2: METHODOLOGY

2.1 Water balance and model components

Assuming that the water balance can be closed at the watershed scale, i.e. net groundwater flow that’s diverted away from the surface outlet is negligible, it can be described as follows:

\[ S_{t+1} = S_t + P_t - Q_t - E_t \]  \hspace{1cm} (1)

where \( S \) is the average moisture state, \( P \) is precipitation, \( Q \) is streamflow and \( E \) is actual evapotranspiration. In a watershed, water balance equation can represent changes of the water storage per unit surface area over a given period of time.

Based on the water balance equation, we have formulated a range of lumped watershed models with different process and parametric complexities. Each model has a soil moisture accounting component with three elements, i.e. interception, actual evapotranspiration and runoff production. Interception is controlled by vegetation type and the fraction of vegetation cover. It is a function of the precipitation with one parameter (interception coefficient) to estimate the loss of rainfall via interception. Actual evapotranspiration is controlled by water availability and potential evapotranspiration (PE). Actual evapotranspiration (AE) is estimated by modeling bare soil evaporation and transpiration separately. Transpiration is controlled by a threshold for water storage, by the fraction of vegetation covered area and the vegetation type. Bare soil evaporation is controlled by the soil water storage capacity and the area of bare soil. The detailed formulations of all the models are shown in Appendix A. We assume a saturation excess runoff mechanism for runoff production for all models. Saturation runoff excess is generated if the storage capacity is satisfied, and subsurface flow is generated if the water
storage reaches field capacity. To add delay to the subsurface runoff, we use linear reservoirs as the routing components throughout this study.

2.2 Modeling framework

2.2.1 Previous studies

Previous studies (e.g. Atkinson, 2002; Farmer et al., 2003) using the classic bucket model (Manabe, 1963) take advantage of a simple conceptualization of the hydrologic system and require the specification of a minimal set of parameters derived from landscape properties. Similar models have been adopted in this study with only minimal changes. In a manner similar to prior studies we start with our simplest model to capture the water balance at the annual timescale. Failures to predict runoff for finer time scales (intra-annual, monthly, daily, etc.) using the simplest model, are assumed to represent model structure deficiencies and justify the introduction of additional variables and processes. By progressively increasing the model complexity, the predictions for finer scales are satisfied. Our extensions to this approach are discussed in the next sections.

2.2.2 Monte Carlo framework

We propose a fuzzy top-down model evaluation and selection framework that explicitly accounts for parametric uncertainties using Monte Carlo ensemble simulation (Figure 1). Prior top-down studies have been limited in their consideration of uncertainty in a priori model parameters. Son and Sivapalan (2007) recently presented a methodology to use auxiliary data such as deuterium concentrations in the streamflow to help improve the model structure and reduce the parameter uncertainty at the same time.
They used statistical measures that minimize the residuals between simulated and observed response to evaluate the model ability to reproduce observations. In our study we separately evaluate a model’s ability to reproduce the magnitude of watershed responses and a model’s ability to reproduce the dynamics of watershed responses in the framework of considering parameter uncertainty in models.

In this study we are assuming independent, identically distributed parameters of all the models. We use uniform prior distributions for all parameters. Samples are drawn over both the feasible parameter space and the \textit{a priori} parameter space by implementing uniform random sampling (URS) to calculate ensemble simulations with respect to these two ranges. By implementing Monte Carlo simulation, ensemble predictions of watershed responses.

Appendix B provides details on how the lower limit $\theta_1$ and upper limit $\theta_2$ of the feasible range of each parameter were determined according to assumptions about the physical meaning of the parameters. The feasible ranges of parameters are estimated either by using values derived from equations relating watershed physical characteristics to model parameters, or by referring to empirical techniques from previous studies (Van Werkhoven et al., 2008). The \textit{a priori} parameter space refers to reduced parameter ranges attained using watershed specific information. \textit{A priori} parameter ranges constrain feasible parameter ranges using the \textit{a priori} parameter estimations to give sampling space over which best parameter sets can be chosen with higher probability.

We use estimates of the \textit{a priori} parameters $\theta_a$ and their ranges $\Delta\theta_f$ (difference between upper limit $\theta_2$ and lower limit $\theta_1$ of the feasible range of each parameter) to calculate \textit{a priori} ranges according to the following equation $[\theta_a - \Delta\theta_f * 0.15, \theta_a + \Delta\theta_f * 0.15]$
(see Appendix B). The \textit{a priori} parameters in the study watersheds are estimated from vegetation data, soil data and runoff observations. Since data are hard to collect for some parameters such as $k_d$ (deep recharge coefficient from upper saturated zone to deeper store) and $b$ (shape parameter for spatial soil water storage distribution in the multi-bucket model), these parameter values were estimated using manual calibration. The methods used to estimate parameters and their estimates for all study watersheds are presented in Appendix C.

\begin{figure}
\centering
\includegraphics[width=\textwidth]{flowchart.png}
\caption{Flow chart of proposed model selection framework.}
\end{figure}

2.3 Modular modeling structure
All of the model structures used in this study can be separated into three modules: (i) soil moisture accounting (SMA), (ii) actual evapotranspiration (ET) and (iii) routing (R). The resulting modules are shown schematically in Figure 2 as a function of increasing complexity.

Figure 2. Modules with increasing complexity used in the model structures investigated.

The SMA module estimates the soil water storage in different soil layers based on water distribution by interception, evapotranspiration, infiltration, runoff and percolation. Starting with a single bucket as the representation of a watershed, the simplest SMA module S1 only produces surface runoff via the saturation excess mechanism. The total
storage capacity controls the runoff generation. By increasing the complexity of the SMA module, we obtain 4 SMA modules. S2 adds subsurface flow produced from the excess water generated after field capacity has been satisfied. The threshold parameter (field capacity) and the total storage capacity control runoff generation. In S3, the soil profile is divided into the upper unsaturated zone and lower saturated zone and includes subsurface flow from the saturated zone as well as surface runoff resulting from saturation excess. No runoff is generated from the unsaturated zone in S3. The storage condition of the unsaturated zone is determined by water availability and its field capacity controls the recharge to the saturated zone. S4 extends S3 by incorporating a deep storage recharged by percolation from the saturated store. The percolation rate determines runoff generation from the deep zone.

Building on S1-S4, we have also formulated four multiple-bucket models with the same modular components as the single-bucket SMA modules. M1 M2 M3 M4 are the multi-bucket representations of S1 S2 S3 S4, respectively. We use 10 buckets to consider the spatial variability of the watershed response as another control factor for runoff generation. The 10 buckets of variable soil moisture distribution that fit the Xinanjiang distribution (Zhao et al., 1980, Son and Sivapalan, 2007). The 10 buckets are combined in parallel. The maximum storage of the Xinanjiang distribution $S_{\text{max}}$ and the maximum water storages of the 10 buckets $S_{b\_f}$ are calculated through the following equations:

$$ S_{\text{max}} = (1 + b)S_b $$  \hspace{1cm} (2)

$$ F = [0.05 \ 0.15 \ 0.25 \ 0.35 \ 0.45 \ 0.55 \ 0.65 \ 0.75 \ 0.85 \ 0.95] $$  \hspace{1cm} (3)

$$ S_{b\_f} = S_{\text{max}} \left[ 1 - (1 - F)^{\frac{1}{10}} \right] $$  \hspace{1cm} (4)
where $S_{\text{max}}$ is maximum soil water storage in the watershed, $S_b$ is maximum storage of the bucket model (mm), $b$ is shape parameter for spatial soil water storage distribution in the multi-bucket model, $F$ is cumulative probabilities at which soil water storages of 10 buckets fit spatial soil water storage distribution, $S_{b-f}$ is soil water storage capacities in the 10 buckets.

The ET module estimates the actual evapotranspiration by modeling the bare soil evaporation and vegetation transpiration separately. We assume that the root system reaches from the surface through the saturated zone. ET is estimated from different soil zones divided in each SMA module. ET$_1$ describes the evaporation from the moisture storage as one zone. ET$_2$ describes the evaporation from the unsaturated zone and saturated zone, respectively. It assumes that there is no ET loss from deep storage. ET from the saturated zone is controlled by energy supply through PE. Water availability determines the ET from the unsaturated zone. If the soil water storage satisfies the threshold water storage (field capacity), transpiration reaches the potential transpiration rate. When soil water storage falls below field capacity, transpiration is a function of soil water storage. Bare soil evaporation is a function of soil water storage. The equations for calculating transpiration $E_v$ and bare soil evaporation $E_{bs}$ are:

$$E_v = \begin{cases} ME_p & S_t \geq S_{fc} \\ M \frac{S}{S_{fc}} E_p & S_t < S_{fc} \end{cases}$$

(5)

$$E_{bs} = (1 - M) \frac{S}{S_b} E_p$$

(6)
where $E_{bs}$ is bare soil evaporation (mm/d), $E_v$ is vegetation transpiration (mm/d), $E_p$ is potential evapotranspiration, $S_{fc}$ is field capacity (mm), $S_t$ is total soil water storage at current time $t$ (mm), $M$ is the fraction of catchment area covered by deep rooted vegetation (dimensionless), $K_v$ is vegetation transpiration efficiency (dimensionless).

The routing module describes the flow release from different storages. The subsurface flow is delayed through a linear storage-discharge relationship. Quick subsurface flow is produced from the saturated zone and slow subsurface flow is produced from deep storage. The runoff from the deep storage is delayed by deep percolation as well as the deep soil layer that causes slower recession than the saturated zone. There is no runoff generated from the unsaturated zone until its storage exceeds the field capacity and no routing is considered for the excess saturation runoff produced in this way.

We therefore consider a total of eight models with increasing complexity by combining these 3 components (Figure 3). For descriptions of all parameters see the List of Notations and model equations are provided in Appendix A.
Figure 3. Eight models were derived using different combinations of the available SMA, ET and Routing modules.

2.4 Parameter estimation

Methods and data used for *a priori* estimation are shown in detail Appendix B and are summarized in this section. Most methods are straightforward except those used to estimate the recession coefficients $\alpha_{ss}$ and $\alpha_{bf}$. We use the linear reservoir model to describe the subsurface flow (Wittenberg, 1999). Its exponential response is as follows:

$$Q_t = Q_0 \exp(-\alpha t)$$  \hfill (7)

A storage-outflow relationship based on this model is used in our models to produce the subsurface flow:

$$Q = \alpha S$$  \hfill (8)

After adjusting the form of equation (7), the recession coefficient can be calculated using the following equation:
To estimate the recession coefficient from the saturated zone and deep storage using the observed flow data, we need to separate the falling limb of the streamflow time series to get reasonable subsurface flow partitioning of quick subsurface flow and slow subsurface flow. Linsley et al. (1958) developed an empirical relationship to identify the point along the falling limb where the quick flow ends and all runoff is considered base flow:

\[ D = 0.827 A^{0.2} \]  \hspace{1cm} (10)

where \( D \) is the number of days between the flood peak and the end of quick flow, and \( A \) is the watershed area (km\(^2\)). We use this method to decide the point in time when all runoff is quick subsurface flow and overland flow ceases. For all the watersheds in this study, \( D \) is close to 4 days. We assume slow subsurface starts at the point of 3 days after the point where all runoff is quick subsurface flow. So we calculate the recession coefficient for quick subsurface flow using flow data in the period from 5 to 7 days after the peak appears. Also we assume quick subsurface flow ceases at the point of 15 days after the point where quick subsurface flow starts and all runoff is the slow subsurface flow. So the flow data in the period 20 days after the peak appears are used to calculate recession coefficient for slow subsurface flow.

The falling limbs are chosen only from flood events whose peaks are greater than the 75% quantile of the streamflow time series. We used these recession segments separated by the above methods as quick subsurface flow and slow subsurface flow data. Using equation (9) we calculate the recession coefficients of each quick recession segment and slow recession segment. We take the averages of the recession coefficients.

\[
\alpha = \frac{\ln Q_a - \ln Q_i}{t} \hspace{1cm} (9)
\]
estimated in this manner for both quick subsurface runoff and slow subsurface runoff as the estimated $\alpha_{ss}$ and $\alpha_{bf}$.

2.5 Signatures

A signature is an index or time-series of the response behavior of a watershed at a given time-scale, which is reflective of the watershed functional behavior. Signature plots can be used to represent the characteristics of the watershed response at a given time scale. This study uses similar signature plots as some of the aforementioned previous studies (Jothityangkoon et al., 2001; Atkinson et al., 2002; Farmer et al., 2003). Signatures enable us to compare the model predictions at increasingly finer time scales. They are obtained by aggregating daily streamflow simulations to the appropriate coarser time scale. The four signatures measure runoff variability with deceasing time scales of interest: inter-annual variability, intra-annual variability, the flow duration curve as well as the daily streamflow time series. Inter-annual variability predicts variation of the annual runoff yield that is indicated by long term climate variability. Intra-annual variability predicts seasonality of runoff. The flow duration curve represents the flow regime and its steepness reflects the speed of watershed drainage. Daily streamflow captures the timing and magnitude of daily runoff.

2.6 Measures of acceptability

To reduce the subjectivity challenges and limitations of visually examining hydrograph fit for the large number of combinations of watersheds and model types considered in this study, we measure model performance in an uncertainty framework
without reliance on measures of closeness of fit. We consider two measures jointly to using fuzzy analysis to classify model performance.

We use the reliability measure (Yadav et al., 2007; Zhang et al., 2008) to tests a model’s capability to reproduce the observed magnitude of the four signatures. Reliability is formulated as the ratio of the number of observations captured by the model ensemble to the total number of observations:

\[
Rel = \frac{Num_{in}}{Num}
\]

where \( Rel \) is the reliability measure, \( Num_{in} \) is the number of observations falling inside the ensemble, and \( Num \) is the total number of observations. We determine a certain value of \( Rel \) as a threshold of model acceptability. When reliability is below this value, the model cannot capture the magnitude range of the signature. The points that fall out of the ensemble indicate the parts of the flow regime that the model cannot capture and provide potential insight into the deficiencies of the model structure.

Additionally, we introduce shape to measure the model’s capability to reproduce the signatures dynamics. The shape measure is calculated as the magnitude of the difference of slope between observations and simulations datasets. Its value is normalized by the variance of the observed slope as follows:

\[
Slopdiff = 1 - \frac{\sum_{i=1}^{T} (go_i - gs_i)^2}{\sum_{i=1}^{T} (go_i - \overline{go})^2}
\]

where \( Slopdiff \) is the shape measure, \( go \) is the slope of observations, and \( gs \) the slope of simulations, \( \overline{go} \) is the mean slope of observations during the period \( T \). The shape value is between negative infinity and 1. When simulations are perfectly paralleled to the
observation, $Slopediff$ equals 1. When simulations and observations are orthogonal, the shape value is negative infinity. Using a threshold value of $Splodiff$, which varies with time scale, we can assess whether the model is capable of capturing the signature dynamics.

**2.7 Fuzzy rules**

Fuzzy theory provides an alternative method to quantify the model performance (Bardossy, 2005). We use fuzzy rules to combine the two different measures and obtain a joined measure to evaluate the overall model performance.

In classical mathematics, a set is a deterministic set whose membership function is binary. For instance, the membership of a variable only has two values, 0 and 1 represents the variable does not belong to the set and 1 represents the variable belong to the set. In contrast, a fuzzy set is a vaguely defined set whose membership function describes the degree with which a certain element is a member of a group (e.g. the group of acceptable models). The range of the membership function is [0 1]. Here, 0 represents the degree of membership is 0 and 1 represents the degree of membership is 100%. For example, for the measure of Reliability (R) in this study, the membership function is defined as $M(R)$ (Figure 4). When $M(R)=m_1$, the model is accepted with a membership degree of $m_1$ with respect to reliability. Similarly, the membership of the measure shape($S$) is $M(S)=m_2$. The model is accepted with a membership degree of $m_2$ with respect to shape. The model is rejected if $S<=x_2$ and it is 100% accepted at $S>=y_2$ with respect to shape.
We use a fuzzy multiple objective function (FMOF) to obtain a consistent measure considering both reliability and shape. Previous studies (Yu and Yang, 2000; Yang and Yu, 2006; Nasir and Huang, 2007) used different formulations of FMOF. We define FMOF as

$$FMOF = \min(m_1, m_2)$$  \hspace{1cm} (13)

where $m_1$ and $m_2$ are the membership degrees of reliability and shape, respectively. The value of FMOF varies from 0 to 1 and gives the degree of acceptability of the model.

![Membership functions (schematic) for reliability and shape measures.](image)

**Figure 4.** Membership functions (schematic) for reliability and shape measures.
Chapter 3: WATERSHEDS AND DATA

3.1 Watersheds and data

Twelve MOPEX basins from US were chosen for this study. The MOPEX data sets are provided by NOAA including the hydrometeorologic data sets and vegetation and soil data used in this study. Previous studies [Duan et al, 2006, Gan and Burges, 2006] presented some of the physical properties and hydroclimatic information for the 12 watersheds. The daily data of potential evapotranspiration (PE), precipitation and observed streamflow data for 15 years (1961-1975) from MOPEX data sets are used for this study. We use the data in 1960 and run each model for each watershed to get the initial soil moisture. The MOPEX data sets also include the vegetation adjustment factors for PE to get the monthly PE.

These basins, ranging from dry conditions to wet conditions, represent different hydroclimatic conditions. The locations, relevant climatic and topographic information of 12 watersheds are presented in Table 1 and Figure 1. We use ID in Table 1 as the symbol for each watershed. As illustrated in Figure 1 and Table1, area of these MOPEX basins range from about 1000 km$^2$ to 4500 km$^2$ and the wetness index(P/PE) range from 0.50 to 1.68. A further summary of watershed hydrologic characteristics is presented in Figure 2. The flow duration curve (FDC) plot (Figure 2c) demonstrates that FDC of the English River basin has the steepest slope and FDC of the French Broad River basin has the flattest slope.

The source of basin characteristic data and methods for $a$ priori parameter estimation are presented in Appendix B.
Table 1. Characteristics of the twelve MOPEX watersheds used.

<table>
<thead>
<tr>
<th>ID</th>
<th>Basin name</th>
<th>Area (km²)</th>
<th>Average elevation (m)</th>
<th>Mean annual P (mm)</th>
<th>Mean annual Q (mm)</th>
<th>Mean annual PE (mm)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Guadalupe River near Spring Branch, TX</td>
<td>3406</td>
<td>289</td>
<td>765</td>
<td>116</td>
<td>1528</td>
</tr>
<tr>
<td>2</td>
<td>San Marcos River near Luling, TX</td>
<td>2170</td>
<td>98</td>
<td>827</td>
<td>179</td>
<td>1449</td>
</tr>
<tr>
<td>3</td>
<td>English River near Kalona, IA</td>
<td>1484</td>
<td>193</td>
<td>893</td>
<td>270</td>
<td>994</td>
</tr>
<tr>
<td>4</td>
<td>Spring River near Waco, MO</td>
<td>3015</td>
<td>254</td>
<td>1076</td>
<td>299</td>
<td>1094</td>
</tr>
<tr>
<td>5</td>
<td>Rappahannock River near Fredericksburg, VA</td>
<td>4134</td>
<td>17</td>
<td>1030</td>
<td>378</td>
<td>920</td>
</tr>
<tr>
<td>6</td>
<td>Monocacy River near Frederick, MD</td>
<td>2116</td>
<td>71</td>
<td>1041</td>
<td>421</td>
<td>896</td>
</tr>
<tr>
<td>7</td>
<td>East Fork White River near Columbus, IN</td>
<td>4421</td>
<td>184</td>
<td>1015</td>
<td>378</td>
<td>855</td>
</tr>
<tr>
<td>8</td>
<td>South Branch Potomac River, Springville, WV</td>
<td>3810</td>
<td>171</td>
<td>1042</td>
<td>341</td>
<td>761</td>
</tr>
<tr>
<td>9</td>
<td>Bluestone River near Pipestem, WV</td>
<td>1021</td>
<td>465</td>
<td>1018</td>
<td>417</td>
<td>741</td>
</tr>
<tr>
<td>10</td>
<td>Amite River near Denham Springs, LA</td>
<td>3315</td>
<td>56</td>
<td>1564</td>
<td>610</td>
<td>1073</td>
</tr>
<tr>
<td>11</td>
<td>Tygart Valley River near Pipestem, WV</td>
<td>2372</td>
<td>390</td>
<td>1166</td>
<td>736</td>
<td>711</td>
</tr>
<tr>
<td>12</td>
<td>French Broad River near Asheville, NC</td>
<td>2448</td>
<td>594</td>
<td>1383</td>
<td>800</td>
<td>819</td>
</tr>
</tbody>
</table>

Figure 5. Locations of the 12 MOPEX watersheds in the US (a), and on the Budyko curve (b). AE is actual evapotranspiration, P is precipitation and PE is potential evapotranspiration.
Figure 6. (a) & (b) Daily streamflow hydrographs of 12 watersheds in 1961. (c) Flow duration curves of 12 watersheds using daily streamflow data for period 1960 to 1998.
Chapter 4: RESULTS AND DISCUSSION

4.1 Model performance with respect to reliability

Model reliability is estimated by calculating how many of the observations are contained by the model ensemble. A reliability value of 100% indicates that the model ensemble is capturing all of the observations. The analysis of reliability is supported by visualizing the simulations of the four different signatures. The question is, how does reliability vary from models S1 to M4, from annual to daily time scales, and from dry watersheds to wet watersheds? Figures 7-10 show the ensemble predictions of models S1 to S4 (plots for models M1 to M4 are shown in the Appendix) for the four signatures over both feasible ranges (left column) and a priori ranges (right column) for the three test watersheds shown in the figures from top to bottom [Guadalupe (dry), East Fork White (medium) and French Broad (wet)]. Table 2 shows the calculated reliability values.

Figure 7 shows the ensemble predictions of models S1, S2, S3 and S4 for the inter-annual variability for the feasible parameter ranges (left column) and the a priori ranges (right column). For the feasible parameter range all of the observations fall inside the ensembles of all models and reliability values of all models are therefore equal to 1. The uncertainty ranges are quite wide though, in particular for the dry watershed (Guadalupe). Plots of the ensembles of the inter-annual variability signature using a priori parameter ranges in the right column show that as expected site specific information significantly narrows the uncertainty bands relative to the results for the feasible ranges. The a priori ensembles of S1 do not capture any observation for all the 3 watersheds, since the runoff produced is too low in all cases. Apart from S1, the other 3 models can capture all the observations for the Guadalupe and East Fork White watersheds. For the French Broad
watershed, the increased complexity of models S3 and S4 is required to achieve a 100% reliability. The ensemble of S2 misses approximately 27 percent of the observations.

Figure 7. Ensemble prediction ranges of inter-annual variability. Ensembles produced by parameter sets drawn from feasible and a priori ranges for the (a) and (b) Guadalupe watershed, (c) and (d) East Fork White watershed, and (e) and (f) French Broad watershed respectively. The vertical axis represents annual runoff over average observed annual runoff.

The intra-annual variability results in Figure 8 show a perfect reliability for the feasible parameter ranges for the Guadalupe and East Fork White watersheds. For the French Broad watershed (the wettest of the three), a couple of observations fall just outside of the ensemble of S1 (reliability of 83%), while the other 3 models capture all the observations. For the Guadalupe watershed, the ensemble of S1 for intra-annual
variability for the \textit{a priori} parameter ranges does not include any observation whereas the other 3 models have 100\% reliability. When constrained to the \textit{a priori} parameter ranges model S1 performs very poorly for both the East Fork White and the French Broad watersheds, missing 75\%-percent of the observations for the intra-annual signature. For the East Fork White watershed, using \textit{a priori} ranges, S3 and S4 can capture all observations and only a couple of observations are located just outside of the ensemble of S2 (reliability of 83\%). The French Broad watershed required the increased complexity of the S4 model to attain a perfect reliability. Models S2 and S3 failed to capture the spring and summer observations yielding reduced reliability values of 58\% and 75\%, respectively.

Figure 9 shows the ensemble simulations of the 4 models for the flow duration curve over both the feasible and \textit{a priori} parameter ranges. Ensemble predictions using the feasible parameter ranges for all the 3 watersheds, show that S1 can only capture the high flows and therefore suffers from low reliability values ranging between 11 and 20\%. For the Guadalupe watershed, the increased complexity of models S3 and S4 is required to capture the full range of the flow regime for the feasible range results in the left column of Figure 9. The wetter East Fork White and French Broad watersheds required less model complexity to capture the full range of the flow regime, where model S2 was sufficient given the feasible parameter range. The \textit{a priori} parameter ranges significantly differentiated model performance for all three watersheds. For all 3 watersheds, the models cover an increasingly wider range of the flow regime as the model complexity increases when moving from models S1 to S4. With increasing model complexity, the
models enhance their ability to capture medium and low flow conditions by providing additional routing mechanisms.

**Figure 8.** Ensemble prediction ranges of intra-annual variability of streamflow. Ensembles produced by parameter sets drawn from feasible and *a priori* ranges for the (a) and (b) Guadalupe watershed, (c) and (d) East Fork White watershed, and (e) and (f) French Broad watershed respectively. The vertical axis represents average monthly runoff over average observed annual runoff.
Figure 9. Ensemble prediction ranges for the flow duration curve. Ensembles produced by parameter sets drawn from feasible and *a priori* ranges for the (a) and (b) Guadalupe watershed, (c) and (d) East Fork White watershed, and (e) and (f) French Broad watershed respectively.

Figure 10 shows the necessity of increasing model complexity when simulating daily streamflow. Plots are shown at the log-transformed scale to emphasize recession periods. One year of daily streamflow (1961) is selected for visualization purposes, while all 10 years are used to calculate the reliability values (Table 2). For the dry Guadalupe watershed, none of the models can reproduce the continuous low flow recession during long dry periods since the models drain too quickly. This is different for the two wetter watersheds in which more frequent rainfall events prevent the lower stores from drying out most of the time. Model S1 cannot capture low flows because of a lack of subsurface
flow components. The additional slow flow routing enables models S2, S3 and S4 to reproduce low flows. Their reliability values are therefore all close to 1 for all the 3 watersheds. Plots in the right column show that the \textit{a priori} parameter ranges make it even more difficult for the models to capture the daily streamflow recessions and only S4 is capable of doing so even in the wet watersheds now.

Table 2 shows the values of reliability of all 8 models for the 3 watersheds (Guadalupe, East Fork White and French Broad). Reliability values using feasible ranges are very close between models at inter-annual and intra-annual scales for all 3 watersheds. For FDC and daily streamflow reliability values of models S2, S3 and S4 are significantly higher compared to model S1. In general, reliability values using \textit{a priori} parameter ranges increase significantly with increasing model complexity for FDC and daily streamflow.
Figure 10. Ensemble prediction ranges for daily streamflow in the year 1961 on logarithmic scale. Ensembles produced by parameter sets drawn from feasible and \textit{a priori} ranges for the (a) and (b) Guadalupe watershed, (c) and (d) East Fork White watershed, and (e) and (f) French Broad watershed respectively.
Table 2: Reliability of four signatures over both feasible ranges and *a priori* ranges for the Guadalupe, East Fork White and French Broad watersheds. All reliability values show as percents (%).

<table>
<thead>
<tr>
<th>Model</th>
<th>S1</th>
<th>M1</th>
<th>S2</th>
<th>M2</th>
<th>S3</th>
<th>M3</th>
<th>S4</th>
<th>M4</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Feasible range</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Guadalupe</td>
<td>100%</td>
<td>100%</td>
<td>100%</td>
<td>100%</td>
<td>100%</td>
<td>100%</td>
<td>100%</td>
<td>100%</td>
</tr>
<tr>
<td>Inter-annual</td>
<td>100%</td>
<td>100%</td>
<td>100%</td>
<td>100%</td>
<td>100%</td>
<td>100%</td>
<td>100%</td>
<td>100%</td>
</tr>
<tr>
<td>Intra-annual</td>
<td>11%</td>
<td>11%</td>
<td>98%</td>
<td>98%</td>
<td>100%</td>
<td>100%</td>
<td>100%</td>
<td>100%</td>
</tr>
<tr>
<td>FDC</td>
<td>12%</td>
<td>12%</td>
<td>93%</td>
<td>92%</td>
<td>96%</td>
<td>96%</td>
<td>100%</td>
<td>100%</td>
</tr>
<tr>
<td>Daily</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>East Fork White</td>
<td>100%</td>
<td>100%</td>
<td>100%</td>
<td>100%</td>
<td>100%</td>
<td>100%</td>
<td>100%</td>
<td>100%</td>
</tr>
<tr>
<td>Inter-annual</td>
<td>100%</td>
<td>100%</td>
<td>100%</td>
<td>100%</td>
<td>100%</td>
<td>100%</td>
<td>100%</td>
<td>100%</td>
</tr>
<tr>
<td>Intra-annual</td>
<td>20%</td>
<td>20%</td>
<td>100%</td>
<td>100%</td>
<td>100%</td>
<td>100%</td>
<td>100%</td>
<td>100%</td>
</tr>
<tr>
<td>FDC</td>
<td>22%</td>
<td>22%</td>
<td>98%</td>
<td>98%</td>
<td>99%</td>
<td>99%</td>
<td>99%</td>
<td>99%</td>
</tr>
<tr>
<td>Daily</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>French Broad</td>
<td>100%</td>
<td>100%</td>
<td>100%</td>
<td>100%</td>
<td>100%</td>
<td>100%</td>
<td>100%</td>
<td>100%</td>
</tr>
<tr>
<td>Inter-annual</td>
<td>100%</td>
<td>100%</td>
<td>100%</td>
<td>100%</td>
<td>100%</td>
<td>100%</td>
<td>100%</td>
<td>100%</td>
</tr>
<tr>
<td>Intra-annual</td>
<td>83%</td>
<td>83%</td>
<td>100%</td>
<td>100%</td>
<td>100%</td>
<td>100%</td>
<td>100%</td>
<td>100%</td>
</tr>
<tr>
<td>FDC</td>
<td>20%</td>
<td>20%</td>
<td>100%</td>
<td>100%</td>
<td>100%</td>
<td>100%</td>
<td>100%</td>
<td>100%</td>
</tr>
<tr>
<td>Daily</td>
<td>21%</td>
<td>21%</td>
<td>97%</td>
<td>97%</td>
<td>98%</td>
<td>97%</td>
<td>100%</td>
<td>100%</td>
</tr>
<tr>
<td><strong>A priori range</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Guadalupe</td>
<td>0%</td>
<td>0%</td>
<td>100%</td>
<td>100%</td>
<td>100%</td>
<td>100%</td>
<td>100%</td>
<td>100%</td>
</tr>
<tr>
<td>Inter-annual</td>
<td>0%</td>
<td>25%</td>
<td>100%</td>
<td>100%</td>
<td>100%</td>
<td>100%</td>
<td>100%</td>
<td>100%</td>
</tr>
<tr>
<td>Intra-annual</td>
<td>2%</td>
<td>4%</td>
<td>39%</td>
<td>39%</td>
<td>65%</td>
<td>66%</td>
<td>100%</td>
<td>99%</td>
</tr>
<tr>
<td>FDC</td>
<td>2%</td>
<td>5%</td>
<td>37%</td>
<td>38%</td>
<td>60%</td>
<td>60%</td>
<td>98%</td>
<td>98%</td>
</tr>
<tr>
<td>Daily</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>East Fork White</td>
<td>0%</td>
<td>0%</td>
<td>100%</td>
<td>100%</td>
<td>100%</td>
<td>100%</td>
<td>100%</td>
<td>100%</td>
</tr>
<tr>
<td>Inter-annual</td>
<td>17%</td>
<td>25%</td>
<td>83%</td>
<td>75%</td>
<td>92%</td>
<td>83%</td>
<td>100%</td>
<td>100%</td>
</tr>
<tr>
<td>Intra-annual</td>
<td>4%</td>
<td>4%</td>
<td>79%</td>
<td>85%</td>
<td>97%</td>
<td>98%</td>
<td>99%</td>
<td>99%</td>
</tr>
<tr>
<td>FDC</td>
<td>5%</td>
<td>5%</td>
<td>70%</td>
<td>67%</td>
<td>75%</td>
<td>72%</td>
<td>86%</td>
<td>82%</td>
</tr>
<tr>
<td>Daily</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>French Broad</td>
<td>0%</td>
<td>0%</td>
<td>73%</td>
<td>87%</td>
<td>100%</td>
<td>100%</td>
<td>100%</td>
<td>100%</td>
</tr>
<tr>
<td>Inter-annual</td>
<td>25%</td>
<td>25%</td>
<td>58%</td>
<td>67%</td>
<td>75%</td>
<td>75%</td>
<td>100%</td>
<td>100%</td>
</tr>
<tr>
<td>Intra-annual</td>
<td>4%</td>
<td>4%</td>
<td>48%</td>
<td>46%</td>
<td>55%</td>
<td>52%</td>
<td>98%</td>
<td>98%</td>
</tr>
<tr>
<td>FDC</td>
<td>4%</td>
<td>4%</td>
<td>38%</td>
<td>39%</td>
<td>44%</td>
<td>44%</td>
<td>85%</td>
<td>85%</td>
</tr>
</tbody>
</table>

4.2 Model performance with respect to shape

The shape measure is used to evaluate the models’ ability to capture the dynamics of the water balance signatures rather than their magnitude as was the case with the reliability measure. The value of the shape measure equals 1 when the simulated signature is absolutely parallel to the observed signature, which indicates that the model can perfectly reproduce the dynamics of the signature at the given time scale. Similar to the previous section, we test the ability of models S1 to M4 for reproducing the observed
system dynamics by selecting the best parameter set from 10,000 uniformly sampled
draws from feasible and a priori parameter ranges respectively. Visually, only the results
for the FDC signature are shown, while all shape values are listed in Table 3. The
remaining plots are shown in the Appendix.

Plots in the left column of Figure 11 show the best simulations of models S1, S2, S3
and S4 with parameter sets drawn from their feasible range with respect to the shape
measure. Watersheds are from top to bottom moving from driest to wettest (Guadalupe,
East Fork White and French Broad). For all 3 watersheds, model S1 cannot reproduce
the observed shape for medium and low flow percentiles. For the Guadalupe watershed,
only model S4 can reproduce the shape of the flow regime. For the East Fork White and
French Broad watersheds all models except S1 can reproduce the shape of the FDC quite
well. Though all models except S4 have problems with the very low flows. Plots in the
right column show the best simulations with parameter sets drawn from a priori ranges
for the 3 watersheds. With increasing model complexity, the models improve their ability
to reproduce the shape of the FDC. For all 3 watersheds, only model S4 is capable of
reproducing the full flow regime.

Table 3 lists the shape measure values of all 8 models using both the feasible ranges
and a priori parameter ranges for the 3 watersheds. Shape values at inter-annual scale do
not always increase as the model complexity increases. In some cases, simpler models
have higher shape value and single-bucket models have higher shape values than the
corresponding multi-bucket models for inter-annual variability. For intra-annual
variability, FDC and daily streamflow shape values increase in general as the model
complexity increases.
Figure 11. Simulations of the best shape reproduction by models S1 to S4 with respect to the observed flow duration curve. Ensembles produced by parameter sets drawn from feasible and a priori ranges for the (a) and (b) Guadalupe watershed, (c) and (d) East Fork White watershed, and (e) and (f) French Broad watershed respectively.
Table 3: The shape measures of four signatures over both feasible ranges and *a priori* ranges for Guadalupe, East Fork White and French Broad. The ideal shape value is 1.00 that indicates simulations are perfectly parallel to observations. The shape value of the best simulations with respect to the shape measure is chosen to represent the model’s ability to reproduce the dynamics of signatures.

<table>
<thead>
<tr>
<th>Model</th>
<th>Feasible range</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>S1</td>
<td>M1</td>
<td>S2</td>
<td>M2</td>
<td>S3</td>
<td>M3</td>
<td>S4</td>
</tr>
<tr>
<td>Guadalupe</td>
<td>Inter-annual</td>
<td>-0.07</td>
<td>-0.07</td>
<td>0.29</td>
<td>0.19</td>
<td>0.14</td>
<td>0.14</td>
<td>0.21</td>
</tr>
<tr>
<td></td>
<td>Intra-annual</td>
<td>0.43</td>
<td>0.45</td>
<td>0.55</td>
<td>0.56</td>
<td>0.59</td>
<td>0.61</td>
<td>0.73</td>
</tr>
<tr>
<td></td>
<td>FDC</td>
<td>-*</td>
<td>-*</td>
<td>0.22</td>
<td>0.26</td>
<td>0.33</td>
<td>0.44</td>
<td>0.50</td>
</tr>
<tr>
<td></td>
<td>Daily</td>
<td>0.00</td>
<td>0.04</td>
<td>0.27</td>
<td>0.12</td>
<td>0.07</td>
<td>0.13</td>
<td>0.12</td>
</tr>
<tr>
<td>East Fork White</td>
<td>Inter-annual</td>
<td>0.21</td>
<td>0.15</td>
<td>0.12</td>
<td>0.14</td>
<td>0.35</td>
<td>0.38</td>
<td>0.27</td>
</tr>
<tr>
<td></td>
<td>Intra-annual</td>
<td>0.41</td>
<td>0.39</td>
<td>0.55</td>
<td>0.55</td>
<td>0.64</td>
<td>0.63</td>
<td>0.65</td>
</tr>
<tr>
<td></td>
<td>FDC</td>
<td>-0.62</td>
<td>0.12</td>
<td>0.49</td>
<td>0.48</td>
<td>0.49</td>
<td>0.49</td>
<td>0.50</td>
</tr>
<tr>
<td></td>
<td>Daily</td>
<td>0.00</td>
<td>0.00</td>
<td>0.26</td>
<td>0.16</td>
<td>0.19</td>
<td>0.22</td>
<td>0.24</td>
</tr>
<tr>
<td>French Broad</td>
<td>Inter-annual</td>
<td>0.09</td>
<td>0.06</td>
<td>0.01</td>
<td>-0.02</td>
<td>0.36</td>
<td>0.23</td>
<td>0.26</td>
</tr>
<tr>
<td></td>
<td>Intra-annual</td>
<td>0.51</td>
<td>0.51</td>
<td>0.57</td>
<td>0.58</td>
<td>0.58</td>
<td>0.58</td>
<td>0.78</td>
</tr>
<tr>
<td></td>
<td>FDC</td>
<td>-1.46</td>
<td>0.18</td>
<td>0.49</td>
<td>0.49</td>
<td>0.49</td>
<td>0.50</td>
<td>0.50</td>
</tr>
<tr>
<td></td>
<td>Daily</td>
<td>0.00</td>
<td>0.00</td>
<td>0.34</td>
<td>0.33</td>
<td>0.33</td>
<td>0.33</td>
<td>0.41</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Model</th>
<th>A priori range</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>S1</td>
<td>M1</td>
<td>S2</td>
<td>M2</td>
<td>S3</td>
<td>M3</td>
<td>S4</td>
</tr>
<tr>
<td>Guadalupe</td>
<td>Inter-annual</td>
<td>-0.16</td>
<td>-0.08</td>
<td>0.38</td>
<td>0.35</td>
<td>0.14</td>
<td>0.11</td>
<td>0.29</td>
</tr>
<tr>
<td></td>
<td>Intra-annual</td>
<td>0.22</td>
<td>0.36</td>
<td>0.44</td>
<td>0.43</td>
<td>0.25</td>
<td>0.25</td>
<td>0.38</td>
</tr>
<tr>
<td></td>
<td>FDC</td>
<td>-*</td>
<td>-*</td>
<td>-0.22</td>
<td>0.28</td>
<td>0.28</td>
<td>0.41</td>
<td>0.49</td>
</tr>
<tr>
<td></td>
<td>Daily</td>
<td>0.00</td>
<td>0.02</td>
<td>0.06</td>
<td>-0.01</td>
<td>0.03</td>
<td>0.12</td>
<td>0.12</td>
</tr>
<tr>
<td>East Fork White</td>
<td>Inter-annual</td>
<td>0.20</td>
<td>0.13</td>
<td>-0.27</td>
<td>-0.24</td>
<td>0.04</td>
<td>-0.13</td>
<td>-0.07</td>
</tr>
<tr>
<td></td>
<td>Intra-annual</td>
<td>0.33</td>
<td>0.35</td>
<td>0.55</td>
<td>0.55</td>
<td>0.61</td>
<td>0.60</td>
<td>0.68</td>
</tr>
<tr>
<td></td>
<td>FDC</td>
<td>-*</td>
<td>-*</td>
<td>0.45</td>
<td>0.46</td>
<td>0.47</td>
<td>0.47</td>
<td>0.50</td>
</tr>
<tr>
<td></td>
<td>Daily</td>
<td>-2.21</td>
<td>-4.40</td>
<td>-1.52</td>
<td>-2.57</td>
<td>-0.99</td>
<td>-1.54</td>
<td>-0.52</td>
</tr>
<tr>
<td>French Broad</td>
<td>Inter-annual</td>
<td>-0.09</td>
<td>-0.05</td>
<td>-0.40</td>
<td>-0.39</td>
<td>0.32</td>
<td>0.22</td>
<td>0.12</td>
</tr>
<tr>
<td></td>
<td>Intra-annual</td>
<td>0.12</td>
<td>0.12</td>
<td>0.36</td>
<td>0.36</td>
<td>0.37</td>
<td>0.38</td>
<td>0.67</td>
</tr>
<tr>
<td></td>
<td>FDC</td>
<td>-*</td>
<td>-*</td>
<td>0.39</td>
<td>0.39</td>
<td>0.45</td>
<td>0.46</td>
<td>0.50</td>
</tr>
<tr>
<td></td>
<td>Daily</td>
<td>-4.35</td>
<td>-4.39</td>
<td>-0.13</td>
<td>-0.10</td>
<td>-0.05</td>
<td>-0.01</td>
<td>0.41</td>
</tr>
</tbody>
</table>

* The full range of simulated flow duration curve does not exceed 25% time.

### 4.3 Fuzzy evaluation of model performance

We use fuzzy rules discussed in section 2.7 to combine reliability and shape measures and to obtain a consistent measure to evaluate model suitability across the twelve watersheds analyzed in this study. According to the examination of visual model performance using Figures 7 to 11 (and the remaining Figures shown in the Appendix), we tune the membership function thresholds of reliability and shape measures to reflect
our subjective opinion on model performance based on the visual examination of best simulation of each model. While this is not making the general approach more objective, the fuzzy membership functions and their thresholds model our subjective preference related to judging a model’s structure and therefore allows us to use this approach as a screening tool and apply it to a potentially much larger selection of watersheds and model choices, as long as our selection to tune the rule parameters is sufficiently representative. Infeasible models are therefore eliminated early and we can focus a more detailed analysis on the remaining more promising system representations. After determining the thresholds values, we can calculate the fuzzy values of model performance using the fuzzy measure we defined, i.e. FMOF. Table 4 provides these threshold values of membership function for the shape measure at each time scale. Our visual analysis did make clear that different thresholds are required depending on the observed variability of the signature. The thresholds values of the membership function for the reliability measure are 95% (model whose reliability is not less than 95% will be 100% accepted) and 75% (model whose reliability is no greater than 75% will be rejected) at all time scales.

Figure 12 shows model acceptance values based on the fuzzy evaluation according to FMOF derived from an assessment of the model ensembles of all signatures using the feasible parameter ranges. Watersheds are sorted from driest (Guadalupe, watershed 1) to the wettest (French Broad, watershed 12).

At the inter-annual time scale model S1 already satisfies our thresholds for the simulation of inter-annual variability in the medium and wet watersheds, while model S2 satisfies that the thresholds in the dry watersheds. For the Amite watershed (10), models
S1 and M1 have higher acceptability than models S2, M2 and S3, while model M3 is unacceptable. For the San Marcos (2) and English (3) watersheds, single-bucket models have slightly higher acceptability than the corresponding multi-bucket models.

For the intra-annual variability, model S2 and more complex models are acceptable for the Guadalupe (1), San Marcos (2), Spring (4), East Fork White (7), Tygart (11) and French Broad (12) watersheds. All models are acceptable for Amite (10) at the intra-annual scale. For the remaining watersheds, models S4 and M4 are acceptable except for the English (3) watershed, where none of the models is acceptable. Model S1 has a higher acceptability than S2 for the Monocacy (6) Watershed. For the Rappahannock (5), Monocacy (6) and Bluestone (9) watersheds, we can see that the single-bucket models have a slightly higher acceptability than the corresponding multi-bucket models.

When simulating the FDC, only model S4 is acceptable in reproducing the observed curve with respect to both reliability and shape for the dry watersheds (1, 2, 3 and 4). For the other watersheds, model S2 already satisfies the simulation criteria for the FDC.

At the daily time scale, none of the models provides acceptable simulations of the observed streamflow except models S4 and M4, which can be accepted for the French Broad (12), Rappahannock (5) and Monocacy (6) watersheds. For the FDC and daily streamflow signatures, single-bucket models have slightly lower acceptability than the corresponding multi-bucket models.

Figure 13 visualizes the model acceptance based on the fuzzy evaluation with respect to reliability and shape derived from an assessment of ensemble signature simulations using \textit{a priori} parameter ranges. At the inter-annual time scale, models S2 and M2 can already satisfy the simulation for most of watersheds. For the other watersheds, models
S3 and M3 can capture the inter-annual variability for the Tygart (11) and French Broad (12) watersheds. Models S4 and M4 can capture this signature for the Amite (10) watershed. For the East Fork White (7) watershed, none of models but S3 can capture the inter-annual variability.

At the intra-annual scale, models S4 and M4 can capture intra-annual signature variability in the Guadalupe (1), San Marcos (2), Amite (10) and French Broad (12) watersheds. Model S2 also captures the intra-annual variability for the Spring (4) and East Fork White (7) watersheds. None of the models can reproduce the intra-annual variability for the remaining watersheds. For the FDC signature, models S4 and M4 are acceptable representations for all watersheds. When simulating daily streamflow, all of the models unacceptable except that models S4 and M4 work in the French Broad (12) watershed.

Table 4: Membership function thresholds for the shape measure at each time scale.

<table>
<thead>
<tr>
<th>Membership function thresholds</th>
<th>Unacceptable</th>
<th>100% acceptable</th>
</tr>
</thead>
<tbody>
<tr>
<td>Inter-annual</td>
<td>-0.2</td>
<td>0.2</td>
</tr>
<tr>
<td>Intra-annual</td>
<td>0.35</td>
<td>0.55</td>
</tr>
<tr>
<td>FDC</td>
<td>0.45</td>
<td>0.5</td>
</tr>
<tr>
<td>Daily</td>
<td>0.2</td>
<td>0.4</td>
</tr>
</tbody>
</table>
Figure 12. Fuzzy evaluation of model performance for the 12 watersheds based on model ensembles of signatures derived from their feasible parameter ranges.
Figure 13. Fuzzy evaluation of model performance for the 12 watersheds based on model ensembles of signatures derived from their \textit{a priori} parameter ranges.
4.4 Discussion

The results derived through the fuzzy evaluation of model performance combine the reliability and shape measures to quantify the degree of acceptability of each model for each watershed and provide a screening tool for the necessary model complexity at a given time scale.

At the inter-annual time scale, the simplest model, S1, can capture the inter-annual variability for medium and wet watersheds. For dry watersheds, model S2 is needed to capture the inter-annual variability, which indicates that the addition of subsurface flow and the delay of runoff are required at the annual scale. S1 only produces saturation excess flow and lacks any routing. When using S1 to model dry watersheds, all water stored evapotranspires between rainfall events due to the water limit conditions in the dry watershed since there is no subsurface flow draining from the model yielding an underestimation of runoff. The introduction of subsurface drainage in model S2 reduces this problem and yields acceptable simulations of the inter-annual variability for the drier watersheds.

At the intra-annual time scale, the required model complexity varies with watershed. Basically, model S4, the most complex one, is always acceptable. But for the English watershed (3), which is likely the most impacted by snow since it is in the furthest north, none of the models seems to be capable of reproducing an observed double-peak in March and May due to lack of snow-accumulation-melt component. The models can only simulate the peak in May, which degrades the intra-annual shape metric’s results for watershed 3. Watersheds (1, 2, 4, 10 and 12), located further south, are less likely to be impacted by snow. Model S1 can reproduce intra-annual runoff variability for watersheds
4 and 10. The fact that a simple bucket model is sufficient suggests that for these two watersheds, climate is the dominant control for the seasonal runoff dynamics. Model S2 can satisfy the simulation of inter-annual variability for watersheds 1, 2 and 12. For the dry watersheds (1 and 2), explicit simulation of the recessions is needed for sustaining runoff during long dry periods. For the wettest watershed (12), the carryover effect between sequential months is significant because of frequent precipitation events and thus both antecedent soil moisture and subsurface contributions to runoff are important for the simulation of monthly flows. For the other watersheds, model structural limitations complicate the classification of model performance because of the influence of snowmelt on runoff production (particularly since the peak of intra-annual variability in March cannot be simulated well) making it difficult to distinguish models S2, S3 or S4 at the intra-annual time scale.

For the flow duration curve (FDC), model S2 is acceptable for medium and wet watersheds. Because the FDC reflects the watershed regime and flow magnitudes without consideration of timing, this signature can be captured with a simple model for wetter systems. For the dry watersheds, model S4 is needed to simulate the full FDC regime since slow subsurface flow in the deep storage is important for low flow generation that is necessary to sustain the streamflow in the dry season when evapotranspiration is high and rainfall events are infrequent.

The recession part of the daily streamflow simulation is difficult to reproduce for the drier watersheds for which consequently none of the models is acceptable. Models S4 and M4 are acceptable for a few medium and wet watersheds. The fuzzy-performance classification at the daily timescale did successfully detect a known structural limitation
in all of the models. The study watersheds are relatively large (all larger than 1000 km$^2$) which requires more extensive routing of the quick flow at the daily prediction scale. This structural deficiency was purposefully used in the design of this study to ensure that known model limitations could be detected using the fuzzy model performance criteria.

Overall the inter-annual variability and FDC signatures are more easily reproduced since they do not require the model to simulate the correct timing. While for intra-annual variability and daily streamflow simulations, both correct timing and magnitude are needed. The lack of adequate routing was correctly identified in the rejection of the full suite of model structures.

The shape measure controls model selection when model performance is evaluated based on sampling from the feasible parameter range. The reliabilities of most models across all signatures are generally greater than 0.9 and the membership with respect to reliability is subsequently close to 1. Therefore, according to the FMOF we define, the shape measure always determines the degree of acceptability of a model structure when using feasible parameter ranges. When using $a$ priori parameter ranges, however, both reliability and shape measures impact the classifications of model acceptability during fuzzy evaluation.

The analysis using feasible parameter ranges predominantly focuses on evaluating the model structures, without strong constraints on the actual parameter ranges. To also test our assumptions about what the model parameters represent and how they are related to physical watershed characteristics we further constrained them to $a$ priori ranges using available soils and vegetation data. If the model is still acceptable with these narrower
parameter ranges, then this is a confirmation of assumptions for the parameters, in addition to the model structure.

For the FDC and daily streamflow signatures, the multi-bucket models have a slightly higher degree of acceptability than the corresponding single-bucket models. The single-bucket models often simulate a peak that is too low and recessions that fall too quickly. The spatial variability included in the multi-bucket models provides a more realistic fit of the variability of the watershed response observed for daily dynamics and results in a better fit of peaks and recessions. However, for watersheds 6 and 8, the multi-bucket model M1 shows a slightly lower degree of acceptability than the corresponding simple-bucket model S1 at the intra-annual timescale. Based on an examination of the simulations using multi-bucket models, we find that model M1 produces much higher runoff in winter than observations because the snowfall produces runoff directly in winter due to no snow-accumulation-melt components and the sub-buckets with small storage capacity contained in M1 cause that M1 produces more runoff than S1 in winter, which cause their simulations to be not parallel to the observations and therefore yield lower shape values.

Thresholds for the shape measure membership function at the inter-annual scale are relatively small compared to the other signatures. These thresholds are -0.2 for model rejection and 0.2 for full membership of a model. The reason for this is that the shape measure is normalized by the variance of the observations. The variance of inter-annual variability is relatively small compared to that of the other signatures. The difference between the slope of the simulations and the slope of observations is considerable compared to the variance of the observation, which results in smaller shape values.
Therefore, the range of the shape values is wider. When the shape values between models differ by only a small amount, the best simulations with respect to the shape measure between models will appear nearly equally acceptable. So even though more complicated models provide lower values of the shape measure at inter-annual scale in some cases, they are acceptable during visual evaluation, and thus require a low threshold of acceptability.
Chapter 5: CONCLUSIONS

In this study we present a novel methodology for evaluating the acceptability of models as representations of watershed hydrology in a top-down uncertainty framework. Implementing the methodology in 12 US watersheds with very different physical and climatic characteristics, we identify the necessary model complexity across four different watershed response signatures. This study therefore provides a tool in which model selection can be formalized through acceptability thresholds regarding simulated shape and magnitude in a fuzzy rule system that provides a screening tool for assessing large numbers of models across a wide range of watersheds. The resulting acceptable models provide insight into what processes control the watershed responses at different time scales under different climate conditions. We have shown that a subsurface flow routing component, controlled by the field capacity, is required for the dry watersheds and that climate is the dominant control of the water balance at the inter-annual timescale. At the daily timescale, the models presented in our study are insufficient for the simulation of daily streamflow due to a lack of appropriate routing of the quick flow contribution to the streamflow at the watershed outlet. Further application of this framework to a larger number of models in more watersheds will be needed to add further credibility to this approach.
Chapter 6: RECOMMENDATIONS FOR FUTURE

This study introduced a novel methodology for screening candidate model structures for predictions across time-scales and across a (potentially) large number of watersheds. The model structures that are acceptable after the screening stage can then be analyzed in greater detail regarding their suitability to represent the hydrology at a specific place.

The current study only uses two measures (reliability and shape) for model acceptability evaluation, which are integrated in a fuzzy rule approach, across four selected signatures. In the future, other measures and signatures need to be tested that make it increasingly difficult for a model to pass the acceptability test. Additional signatures have to be based on sound hydrologic theory and could evaluate response characteristics related to a watershed’s residence time or its slow flow contribution.

The model structures currently included in this thesis could be improved by including a snow-accumulation-melt routine and more extensive routing of the quick flow response. In general, the number of model structures considered needs to be expanded to get a better picture of required model complexity across an even more diverse set of watersheds, including semi-arid systems. This testing would add more credibility to our methodology and help to further tune the acceptance threshold values under the different hydro-climate conditions.

Applying this methodology to a large number of watersheds can also contribute to the topic of watershed classification by grouping watersheds with similar hydrologic behavior as suggested by their similarity in required model structures. Based on watershed groups and their representative model structures, we can then select the
suitable model for the watersheds that have similar physical and climatic conditions without this testing process.
REFERENCES


APPENDIX A: Model Structures and Model Equations

The model structures of the single-bucket models are taken from:


The multiple-bucket formulation is from:

The schematic model structures of four single-bucket models are shown in figure A1-A4 and that of the most complicated multi-bucket model is shown in figure A5. In the sections below we provide the equations associated with the five models.

A1. Model S1

Model S1 is a single bucket model with the single store. It only has runoff generation by saturation excess controlled by the maximum soil water storage.

(1) Threshold storage parameter and threshold storage

\[ f_c = \frac{\theta_{fc} - \theta_{wlt}}{\phi - \theta_{wlt}} \]

\[ S_{fc} = S_b f_c \]
(2) Interception and evapotranspiration, here \( S_{t-1} \) is soil water storage of last time step.

\[
E_i = \alpha_{ei} P
\]

\[
S_t = \min \left( S_b, \ S_{t-1} + P - E_i \right)
\]

\[
E_v = \begin{cases} 
M E_p & \text{if } S_t \geq S_{fc} \\
M \frac{S_t - S_{fc}}{S_b} E_p & \text{if } S_t < S_{fc}
\end{cases}
\]

\[
E_{hs} = (1 - M) \frac{S_t}{S_b} E_p
\]

\[
E = E_i + E_v + E_{hs}
\]

(3) Saturated excess runoff and soil water storage at the current step

\[
S_t = S_{t-1} + P - E
\]

\[
Q_{se} = \begin{cases} 
S_t - S_b & \text{if } S_t \geq S_b \\
0 & \text{if } S_t < S_b
\end{cases}
\]

\[S_t = S_t - Q_{se}\]

\[Q = Q_{se}\]

**A2. Model S2**

![Figure A2. Schematic describing model S2.](image)

Model S2 is a single bucket model with the single store. It contains saturation excess runoff and subsurface flow that controlled by threshold storage \( S_{fc} \).

(1) Threshold storage parameter and threshold storage
\[ f_c = \frac{(\theta_{fc} - \theta_{wb})}{(\phi - \theta_{wb})} \]

\[ S_{fc} = S_b f_c \]

(2) Interception and evapotranspiration, here \( S_{t-1} \) is soil water storage of last time step.

\[ E_i = \alpha_{ei} P \]

\[ S_t = \min(S_b, S_{t-1} + P - E_i) \]

\[ E_v = \begin{cases} ME_p & S_t \geq S_{fc} \\ MS_f \frac{E_p}{S_{fc}} & S_t < S_{fc} \end{cases} \]

\[ E_{hs} = (1 - M) \frac{S_t}{S_b} E_p \]

\[ E = E_i + E_v + E_{hs} \]

(3) Saturated excess runoff, subsurface runoff and soil water storage at the current step

\[ S_t = S_{t-1} + P - E \]

\[ Q_{se} = \begin{cases} S_t - S_b & S_t \geq S_b \\ 0 & S_t < S_b \end{cases} \]

\[ S_t = S_t - Q_{se} \]

\[ Q_{ss} = \begin{cases} \alpha_{ss} (S_t - S_b) & S_t \geq S_{fc} \\ 0 & S_t < S_{fc} \end{cases} \]

\[ S_t = S_t - Q_{ss} \]

\[ Q = Q_{se} + Q_{ss} \]

A3. Model S3
Model S3 is a single bucket model with two stores, i.e. unsaturated zone and saturated zone. Evaporation and transpiration occur from unsaturated zone and saturated zone. Flow generation is by saturation excess runoff and subsurface flow from saturated zone.

(1) Threshold storage parameter and threshold storage in unsaturated zone
\[
f_c = \frac{\theta_c - \theta_{wh}}{\theta_{fc} - \theta_{wh}}
\]
\[
S_{usfc} = \left( S_b - S_{sat \; t-1} \right) f_c
\]

(2) Interception and evapotranspiration. Depletion from unsaturated zone and saturated is allocated proportionally according to water storages in them.
\[
E_i = \alpha_e P
\]
\[
S_{us} = S_{t-1} - S_{sat \; t-1}
\]
\[
S_{us}' = S_{us} + P - E_i
\]
\[
r_p = \begin{cases} 
S_{us}' - S_{usfc} & S_{us}' \geq S_{usfc} \\
0 & S_{us}' < S_{usfc}
\end{cases}
\]
where \( r_p \), recharge to the saturated zone, occurs when the field capacity in the unsaturated zone is satisfied.
\[
S_t = \min \left( S_b, \; S_{us}' + S_{sat \; t-1} \right)
\]
\[
S_{sat} = \min \left( S_b, \; S_{sat \; t-1} + r_p \right)
\]
\[
S_{us} = S_t - Q_{sat}
\]
\[ S_{usfc} = f_s(S_b - S_{sat}) \]

\[ E_v\ sat = \frac{S_{sat}}{S_t} ME_p \]

\[ E_{bs\ sat} = \frac{S_{sat}}{S_t} (1 - M) E_p \]

\[ E_{bs\ us} = \frac{S_{us}}{S_t} (1 - M) \frac{S_{us}}{S_b - S_{sat}} E_p \]

\[ E_v\ us = \begin{cases} 
\frac{S_{us}}{S_t} ME_p & S_{us} > S_{usfc} \\
0 & S_{us} = 0 \\
\frac{S_{us}}{S_t} M \frac{S_{us}}{S_{usfc}} E_p & S_{us} < S_{usfc} 
\end{cases} \]

\[ E_{bs} = E_{bs\ us} + E_{bs\ sat} \]

\[ E_v = E_v\ us + E_v\ sat \]

\[ E = E_t + E_v + E_{bs} \]

(3) Saturated excess runoff, subsurface runoff and soil water storage at the current step

\[ S_t = S_{t-1} + P - E \]

\[ Q_{se} = \begin{cases} 
S_i - S_b & S_i \geq S_b \\
0 & S_i < S_b 
\end{cases} \]

\[ S_t = S_t - Q_{se} \]

\[ Q_{ss} = \alpha_{ss} S_{sat} \]

\[ Q = Q_{se} + Q_{ss} \]

\[ S_t = S_t - Q \]

\[ S_{sat} = S_{sat} - Q_{ss} \]

**A4. Model S4**
Figure A4. Schematic describing model S4.

Model S4 is a single bucket model with three stores. The model structure of S4 is basically the same as that of S3 except that it has an additional deep store that is recharged by the deep percolation from unsaturated zone and saturated zone. The deep store only loses water through the base flow, no evapotranspiration in this store.

1) Threshold storage parameter and threshold storage in unsaturated zone

\[ f_c = \frac{(\theta_{fc} - \theta_{wh})}{(\phi - \theta_{wh})} \]

\[ S_{usfc} = (S_b - S_{sat\ t-1})f_c \]

2) Interception and evapotranspiration. Depletion from unsaturated zone and saturated is allocated proportionally according to water storages in them.

\[ E_i = \alpha_{ei}P \]

\[ S_{us} = S_{t-1} - S_{sat\ t-1} \]

\[ S_{us}' = S_{us} + P - E_i \]

\[ r_p = \begin{cases} S_{us}' - S_{usfc} & S_{us}' \geq S_{usfc} \\ 0 & S_{us}' < S_{usfc} \end{cases} \]

\[ S_t = \min\left( S_b, \ S_{us}' + S_{sat\ t-1} \right) \]

\[ S_{sat} = \min\left( S_b, \ S_{sat\ t-1} + r_p \right) \]

\[ S_{us} = S_t - Q_{sat} \]
\[ S_{\text{usfc}} = f_{v}(S_b - S_{\text{sat}}) \]

\[ E_{v \text{ sat}} = \frac{S_{\text{sat}}}{S_t} ME_p \]

\[ E_{bs \text{ sat}} = \frac{S_{\text{sat}}}{S_t} (1 - M) E_p \]

\[ E_{v \text{ us}} = \begin{cases} 
\frac{S_{\text{us}}}{S_t} ME_p & S_{\text{us}} > S_{\text{usfc}} \\
0 & S_{\text{us}} = 0 \\
\frac{S_{\text{us}}}{S_t} M \frac{S_{\text{us}}}{S_{\text{usfc}}} E_p & S_{\text{us}} < S_{\text{usfc}} 
\end{cases} \]

\[ E_{bs \text{ us}} = \frac{S_{\text{us}}}{S_t} (1 - M) \frac{S_{\text{us}}}{S_b - S_{\text{sat}}} E_p \]

\[ E = E_f + E_v + E_{bs} \]

(3) Base flow from the deep store and deep storage.

\[ Q_{\text{bf}} = \alpha_{\text{bf}} S_{\text{deep}} \]

\[ r_g = k_d S_{\text{sat}} \]

\[ S_{\text{deep}} = S_{\text{deep}} - Q_{\text{bf}} + r_g \]

(3) Saturated excess runoff, subsurface runoff and soil water storage at the current step.

Here, \( S_t \) do not contain soil water storage in the deep store.

\[ S_t = S_{t-1} + P - E \]

\[ Q_{se} = \begin{cases} 
S_t - S_b & S_t \geq S_b \\
0 & S_t < S_b 
\end{cases} \]

\[ S_{\text{sat}} = S_{\text{sat}} - r_g \]

\[ Q_{ss} = \alpha_{ss} S_{\text{sat}} \]

\[ S_{\text{sat}} = S_{\text{sat}} - Q_{ss} \]

\[ Q = Q_{se} + Q_{ss} + Q_{bf} \]
\[ S_t = S_t - Q_{se} - Q_{ss} - r_c \]

**A5. Multiple-bucket**

The multiple-bucket model use 10 buckets of variable soil moisture distribution that fit the Xinanjiang distribution. The 10 buckets is combined in parallel. The multiple-bucket models M1, M2, M3 and M4 follow the same mechanisms of hydrologic processes as S1, S2, S3 and S4, respectively.

\[
S_{\text{max}} = (1 + b)S_b
\]

\[
F = [0.05 \ 0.15 \ 0.25 \ 0.35 \ 0.45 \ 0.55 \ 0.65 \ 0.75 \ 0.85 \ 0.95]
\]

\[
S_{b-f} = S_{\text{max}} \left[ 1 - \left(1 - F \right)^{\frac{1}{b}} \right]
\]

**Figure A5.** Schematic describing the multiple-bucket model.
APPENDIX B: Feasible Ranges and *A Priori* Ranges of Parameters

### B.1 Feasible range

<table>
<thead>
<tr>
<th>Parameters</th>
<th>Unit</th>
<th>Feasible range</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>( M )</td>
<td>[-]</td>
<td>0-1</td>
<td>fraction of catchment area covered by deep rooted vegetation</td>
</tr>
<tr>
<td>( \alpha_{ci} )</td>
<td>[-]</td>
<td>0-0.49[^a]</td>
<td>interception coefficient</td>
</tr>
<tr>
<td>( S_b )</td>
<td>[mm]</td>
<td>0-1200</td>
<td>maximum soil water storage</td>
</tr>
<tr>
<td>( f_c )</td>
<td>[-]</td>
<td>0-1</td>
<td>field capacity</td>
</tr>
<tr>
<td>( \alpha_{ss} )</td>
<td>[day^{-1}]</td>
<td>0.05-0.5[^b]</td>
<td>recession coefficients for subsurface flow from saturated zone</td>
</tr>
<tr>
<td>( \alpha_{bf} )</td>
<td>[day^{-1}]</td>
<td>0.001-0.05[^b]</td>
<td>recession coefficients for subsurface flow (base flow) from deep store</td>
</tr>
<tr>
<td>( K_d )</td>
<td>[day^{-1}]</td>
<td>0-0.5[^c]</td>
<td>deep recharge coefficient, regulates recharge of deeper store from upper perched zone</td>
</tr>
<tr>
<td>( b )</td>
<td>[-]</td>
<td>0.1-2.5[^d],[^e]</td>
<td>shape parameter for spatial soil water storage distribution</td>
</tr>
</tbody>
</table>

[^a]: Dingman (2002, p.307)  
[^b]: Van Werkhoven et al. (2008)  
[^c]: Farmer et al. (2003)  
[^d]: Yadav et al. (2007)  
[^e]: Moore (2007)

### B.2 A priori range

\( A \text{ priori range} = [\theta_a - \Delta \theta_f \ast 0.15, \theta_a + \Delta \theta_f \ast 0.15] \), where \( \theta_a \) represents the value of \( A \text{ priori} \) parameters. Subscripts \( a \) and \( f \) refer to \( A \text{ priori} \) and feasible parameter sets respectively, while \( \Delta \) stands for the parameter range. It is conditioned by the bounds of feasible range.
APPENDIX C: Estimations of \textit{A Priori} Parameters

C1. Methods to estimate \textit{a priori} parameters

<table>
<thead>
<tr>
<th>Parameters</th>
<th>Unit</th>
<th>Description</th>
<th>Method to Estimate</th>
<th>Information needed</th>
</tr>
</thead>
<tbody>
<tr>
<td>M</td>
<td>[-]</td>
<td>Fraction of catchment covered by deep rooted vegetation</td>
<td>Directly measured</td>
<td>Vegetation coverage\textsuperscript{[1]}</td>
</tr>
<tr>
<td>$\alpha_{ci}$</td>
<td>[-]</td>
<td>Interception coefficient</td>
<td>Weighted by fractional coverage of each vegetation type in the watershed</td>
<td>Fractional coverage of each vegetation type according to UMD vegetation classification system\textsuperscript{[1]}</td>
</tr>
<tr>
<td>D</td>
<td>[mm]</td>
<td>Soil depth</td>
<td>$DTB = \sum_{i}^{n} (0.5(ROCKDEPL + ROCKDEPH) \times COMPPCT)$ \textsuperscript{[a]}</td>
<td>STATSGO Database \textsuperscript{[2]}</td>
</tr>
<tr>
<td>$\Phi$</td>
<td>[-]</td>
<td>Porosity</td>
<td>(1) $\Phi = 1 - BD/DP$ BD=0.5(BDL+BDH) DP=2.65 g cm\textsuperscript{-3}\textsuperscript{[b]} (2) $\Phi = -0.00126F_{\text{sand}} + 0.489$\textsuperscript{[c]}</td>
<td>(1) STATSGO Database\textsuperscript{[2]} (2) Soil type\textsuperscript{[1]}</td>
</tr>
<tr>
<td>$\theta_{fc}$</td>
<td>[-]</td>
<td>Field capacity</td>
<td>$\theta_{fc} = \Phi \left( \psi_{\text{fc}} / \psi_s \right)^{1/b}$ $\psi_{\text{fc}} = -7.74e^{-0.0302F_{\text{sand}}}$ kPa $b=0.0159F_{\text{clay}} + 2.91$\textsuperscript{[d]}</td>
<td>Soil type and porosity\textsuperscript{[1]}</td>
</tr>
<tr>
<td>$\theta_{wlt}$</td>
<td>[-]</td>
<td>Wilting point</td>
<td>$\theta_{wlt} = \Phi \left( \psi_{\text{wlt}} / \psi_s \right)^{1/b}$ \textsuperscript{[e]}</td>
<td>Soil type and porosity\textsuperscript{[1]}</td>
</tr>
<tr>
<td>$S_b$</td>
<td>[mm]</td>
<td>Maximum soil water storage</td>
<td>$S_b = \Phi \times D$ or $S_b = (\Phi - \theta_{wlt}) \times D$\textsuperscript{[f]}</td>
<td>Soil depth\textsuperscript{[2]} and soil porosity and permanent wilting point\textsuperscript{[1]}</td>
</tr>
<tr>
<td>$\alpha_{ss}$</td>
<td>[day$^{-1}$]</td>
<td>Recession coefficients for subsurface flow from saturated zone</td>
<td>Linear recession: [ Q = Q_0 \exp(-\alpha t) ] $\alpha = \frac{\ln Q_0 - \ln Q}{t}$</td>
<td>Observed streamflow data for recession curve analysis</td>
</tr>
<tr>
<td>$\alpha_{bf}$</td>
<td>[day$^{-1}$]</td>
<td>Recession coefficients for subsurface flow (base flow) from deep store</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$K_d$</td>
<td>[day$^{-1}$]</td>
<td>Deep recharge coefficient, regulates recharge of deeper store from upper perched zone</td>
<td>Manual calibration</td>
<td></td>
</tr>
<tr>
<td>b</td>
<td>[-]</td>
<td>Shape parameter for spatial soil water storage distribution</td>
<td>Manual calibration</td>
<td></td>
</tr>
</tbody>
</table>
Source of Method
   ROCKDEPH & ROCKDEPL: maximum and minimum values for the depth-to-bedrock (DTB).
   COMPPCT: percentage of each component within the map unit
   BD: soil bulk density
   BDH & BDL: maximum and minimum values on a layer basis for each component in a STATSGO map unit
   PD: particle density, generally assumed to be 2.65 g cm\(^{-3}\) (Hillel, 1980)
   \(F_{\text{sand}}\): percentages of sand for each soil class.
   \(\psi_s\): the saturation soil matrix potential
   \(\psi_{\text{field}}\): soil matrix potential at the field capacity, assumed to be -10 kPa for 1-3 sandy soil classes and -20 kPa for all other soil classes.
   \(\psi_{\text{wilting point}}\): soil matrix potential at the wilting point, assumed to be -1500 kPa.

Source of Information:
APPENDIX D: Further Results

The following figures are provided for the full view of results as a complementary of section 3.2.

![Inter-annual Variability Graph](image)

**Figure D1.** Simulations of the best shape reproduction by models S1 to S4 with respect to the observed inter-annual variability. Ensembles produced by parameter sets drawn from feasible and *a priori* ranges for the (a) and (b) Guadalupe watershed, (c) and (d) East Fork White watershed, and (e) and (f) French Broad watershed respectively.
Figure D2. Simulations of the best shape reproduction by models S1 to S4 with respect to the observed intra-annual variability. Ensembles produced by parameter sets drawn from feasible and *a priori* ranges for the (a) and (b) Guadalupe watershed, (c) and (d) East Fork White watershed, and (e) and (f) French Broad watershed respectively. The vertical axis represents average monthly runoff over average annual runoff.
**Figure D3.** Simulations of the best shape reproduction by models S1 to S4 with respect to the observed daily streamflow in the year 1961. Ensembles produced by parameter sets drawn from feasible and *a priori* ranges for the (a) and (b) Guadalupe watershed, (c) and (d) East Fork White watershed, and (e) and (f) French Broad watershed respectively.
Figure D4. Simulations of the best shape reproduction by models S1 to S4 with respect to the observed daily streamflow in the year 1961 on logarithmic scale. Ensembles produced by parameter sets drawn from feasible and a priori ranges for the (a) and (b) Guadalupe watershed, (c) and (d) East Fork White watershed, and (e) and (f) French Broad watershed respectively.
Figure D5. Simulations of the best shape reproduction by models M1 to M4 with respect to the observed inter-annual variability. Ensembles produced by parameter sets drawn from feasible and \textit{a priori} ranges for the (a) and (b) Guadalupe watershed, (c) and (d) East Fork White watershed, and (e) and (f) French Broad watershed respectively. The vertical axis represents annual runoff over average annual runoff.
Figure D6. Simulations of the best shape reproduction by models M1 to M4 with respect to the observed intra-annual variability. Ensembles produced by parameter sets drawn from feasible and a priori ranges for the (a) and (b) Guadalupe watershed, (c) and (d) East Fork White watershed, and (e) and (f) French Broad watershed respectively. The vertical axis represents average monthly runoff over average annual runoff.
Figure D7. Simulations of the best shape reproduction by models M1 to M4 with respect to the observed flow duration curve. Ensembles produced by parameter sets drawn from feasible and \textit{a priori} ranges for the (a) and (b) Guadalupe watershed, (c) and (d) East Fork White watershed, and (e) and (f) French Broad watershed respectively.
Figure D8. Simulations of the best shape reproduction by models M1 to M4 with respect to the observed daily streamflow in the year 1961. Ensembles produced by parameter sets drawn from feasible and *a priori* ranges for the (a) and (b) Guadalupe watershed, (c) and (d) East Fork White watershed, and (e) and (f) French Broad watershed respectively.
Figure D9. Simulations of the best shape reproduction by models M1 to M4 with respect to the observed daily streamflow in the year 1961 on logarithmic scale. Ensembles produced by parameter sets drawn from feasible and *a priori* ranges for the (a) and (b) Guadalupe watershed, (c) and (d) East Fork White watershed, and (e) and (f) French Broad watershed respectively.
Figure D10. Ensemble prediction ranges for inter-annual variability produced by model M1 to M4. Ensembles produced by parameter sets drawn from feasible and a priori ranges for the (a) and (b) Guadalupe watershed, (c) and (d) East Fork White watershed, and (e) and (f) French Broad watershed respectively. The vertical axis represents annual runoff over average annual runoff.
Figure D11. Ensemble prediction ranges for intra-annual variability produced by model M1 to M4. Ensembles produced by parameter sets drawn from feasible and \textit{a priori} ranges for the (a) and (b) Guadalupe watershed, (c) and (d) East Fork White watershed, and (e) and (f) French Broad watershed respectively. The vertical axis represents average monthly runoff over average annual runoff.
Figure D12. Ensemble prediction ranges for flow duration curve produced by model M1 to M4. Ensembles produced by parameter sets drawn from feasible and *a priori* ranges for the (a) and (b) Guadalupe watershed, (c) and (d) East Fork White watershed, and (e) and (f) French Broad watershed respectively.
Figure D13. Ensemble prediction ranges for daily streamflow in the year 1961 on logarithmic scale produced by model M1 to M4. Ensembles produced by parameter sets drawn from feasible and a priori ranges for the (a) and (b) Guadalupe watershed, (c) and (d) East Fork White watershed, and (e) and (f) French Broad watershed respectively.