

CONCEPTUAL HYDROLOGIC MODELING: INSIGHTS INTO
BAYESIAN ANALYSIS, MODEL DEVELOPMENT, AND
PREDICTIONS IN UNGAUGED BASINS

by

Tyler Jon Smith

A dissertation submitted in partial fulfillment
of the requirements for the degree

of

Doctor of Philosophy

in

Ecology and Environmental Sciences

MONTANA STATE UNIVERSITY
Bozeman, Montana

April 2012

©COPYRIGHT

by

Tyler Jon Smith

2012

All Rights Reserved

APPROVAL

of a dissertation submitted by

Tyler Jon Smith

This dissertation has been read by each member of the dissertation committee and has been found to be satisfactory regarding content, English usage, format, citation, bibliographic style, and consistency and is ready for submission to The Graduate School.

Dr. Lucy A. Marshall

Approved for the Department of Land Resources and Environmental Sciences

Dr. Tracy M. Sterling

Approved for The Graduate School

Dr. Carl A. Fox

STATEMENT OF PERMISSION TO USE

In presenting this dissertation in partial fulfillment of the requirements for a doctoral degree at Montana State University, I agree that the Library shall make it available to borrowers under rules of the Library. I further agree that copying of this dissertation is allowable only for scholarly purposes, consistent with "fair use" as prescribed in the U.S. Copyright Law. Requests for extensive copying or reproduction of this dissertation should be referred to ProQuest Information and Learning, 300 North Zeeb Road, Ann Arbor, Michigan 48106, to whom I have granted "the exclusive right to reproduce and distribute my dissertation in and from microform along with the non-exclusive right to reproduce and distribute my abstract in any format in whole or in part."

Tyler Jon Smith

April 2012

ACKNOWLEDGEMENTS

I would like to offer my appreciation to all of the members of my graduate committee, Dr. Lucy Marshall, Dr. Ashish Sharma, Dr. Brian McGlynn and Dr. Paul Stoy for their endless assistance to me in this endeavor. I wish to especially thank Dr. Lucy Marshall for her continuous support and encouragement toward the completion of this research project. Additionally, I would like to thank Dr. Ashish Sharma for hosting me at the University of New South Wales for a year. This research was made possible thanks to funding from the National Science Foundation (EAR-0943640), the Australian Research Council (DP0879218), and a fellowship from the Montana Institute on Ecosystems. I would like to express my appreciation to the members of the Watershed Hydrology Lab at MSU, and especially Kelsey Jencso and Fabian Nippgen, for providing much of the data used in this study. Finally, I would like to thank my wife Kelly and son Parker for their support, encouragement, and patience.

TABLE OF CONTENTS

1. INTRODUCTION	1
Scope and Context.....	1
Research and Motivation.....	4
Dissertation Outline.....	6
Chapter 2.....	7
Chapter 3.....	7
Chapter 4.....	8
Chapter 5.....	9
Chapter 6.....	10
Chapter 7.....	10
Chapter 8.....	11
Figures.....	12
References	14
2. DEVELOPMENT OF A FORMAL LIKELIHOOD FUNCTION FOR IMPROVED BAYESIAN INFERENCE OF EPHEMERAL CATCHMENTS	19
Contribution of Authors	19
Manuscript Information.....	20
Abstract	21
Introduction	21
Formal Likelihood Function Development	22
Test Applications.....	24
Implementation of Bayesian Inference	24
Hydrologic Model and Forcing Data	24
Synthetic Case Study	24
Real Case Study: Wanalta Creek	26
Discussions and Conclusions	29
Acknowledgements	30
References	31
3. A BAYESIAN LIKELIHOOD FUNCTION SPECIFICATION METHODOLOGY FOR CONCEPTUAL HYDROLOGIC MODELING.....	32
Contribution of Authors	32
Manuscript Information.....	33
Abstract	34
Introduction	34
Selection of a Formal Likelihood Function.....	37

TABLE OF CONTENTS – CONTINUED

Case Studies	42
Hydrologic Model	43
Parameter Estimation	43
A Synthetic Application	45
A Real Data Application	50
Discussion	55
Conclusions	62
Acknowledgements	63
Tables	64
Figures	68
References	74
 4. IMPROVING HYDROLOGIC MODEL CALIBRATION USING A FLOW-CORRECTED TIME TRANSFORMATION	 78
Contribution of Authors	78
Manuscript Information	79
Abstract	80
Introduction	80
Flow-Corrected Time Transformation	82
Case Studies	84
Application to a Snow-Dominated Mountainous Catchment	85
Application to a Dynamically Responsive Australian Catchment	87
Discussion & Conclusions	89
Acknowledgements	91
Tables	91
Figures	92
References	99
 5. USING FIELD DATA TO INFORM AND EVALUATE A NEW MODEL OF CATCHMENT HYDROLOGIC CONNECTIVITY	 101
Contribution of Authors	101
Manuscript Information	102
Abstract	103
Introduction	103
The Role of Landscape Hydrologic Connectivity in Total Catchment Response	106
Study Site Description	107
Overview of Foundational Work	108
The Catchment Connectivity Model (CCM)	109
Test Case: Stringer Creek Catchment	113

TABLE OF CONTENTS – CONTINUED

Data	113
Model Calibration	115
Model Results	116
Model Parameterization via Bayesian Inference	116
Model Simulations	116
Model Evaluation and Consistency with Field Observations	117
Discussion of Results	118
How Appropriate is the CCM from a Predictive Perspective?	118
How Appropriate is the CCM in Representing Internal Catchment Functioning?	120
Can Templates of Hydrologic Behavior be Developed Through Analysis of Landscape Structure Alone?	121
How Transferrable is the CCM Structure to Other Catchments?	122
Conclusions	123
Acknowledgements	124
Tables	125
Figures	126
References	132
 6. A CROSS-CATCHMENT COMPARISON AND SENSITIVITY ANALYSIS OF THE CATCHMENT CONNECTIVITY MODEL	 135
Contribution of Authors	135
Manuscript Information	136
Abstract	137
Introduction	137
Methods	140
Study Site Description	140
Data	140
Hydrologic Model	141
Model Calibration & Uncertainty Analysis Framework	143
Applications	144
Spatial Variability and Transferability: A Cross-Catchment Analysis	144
Input Sensitivity	147
Model Structural Variability	148
Discussion	150
Conclusions	156
Acknowledgements	157
Tables	158
Figures	160
References	171

TABLE OF CONTENTS – CONTINUED

7. PREDICTING HYDROLOGIC RESPONSE THROUGH A POOLED CATCHMENT KNOWLEDGEBASE	175
Contribution of Authors	175
Manuscript Information.....	176
Abstract	177
Introduction	177
Bayes Empirical Bayes.....	179
Methods and Materials	182
Model Description	183
Parameter Estimation	184
Test Case Applications	185
Test Case 1: An Application to Synthetic Data	186
Test Case 2: An Application to Real Data	188
Discussion	192
Conclusions	197
Acknowledgements	198
Tables	199
Figures	201
References	209
8. CONCLUSION.....	212
Summary	212
Limitations and Recommendations for Future Research	216
References	220
REFERENCES	221

LIST OF TABLES

Table	Page
2.1. Mathematical Formulas and Parameters for Each of the Likelihood Functions Considered in This Study	24
2.2. Summary of the Reliability and Sharpness Results for the Synthetic Case Study for Each of the Likelihood Functions	26
2.3. Summary of the Reliability and Sharpness Results for the Wanalta Creek Case Study During Calibration and Validation Periods for Each of the Likelihood Functions C and D	28
3.1. The log-Likelihood Functions, Assumptions, and Calibrated Parameters Used in This Study	64
3.2. Summary of Performance Ranks for Synthetic Data Case E.....	66
3.3. Summary of Efficiency Criteria for the Real Case Study	66
3.4. Summary of Uncertainty Criteria for the Real Case Study.....	67
4.1. Description of Parameters of the Probability Distributed Model	91
5.1. Names and Descriptions of the CCM Parameters and Formal Likelihood Function Parameters Requiring Calibration.....	125
5.2. The 10 th , Median, and 90 th Percentile Values for the Calibrated Parameters (CCM and Likelihood Function)	125
6.1. Catchment Characteristics for Each Catchment Modeled in this Study	158
6.2. Names and Descriptions of the CCM Parameters Requiring Calibration.....	159
7.1. A Description of Global Priors Used Under the Hierarchical Bayesian Modeling Framework	199

LIST OF TABLES – CONTINUED

Table		Page
7.2.	A Description of the Catchments Selected for Analysis in Test Case 2	200

LIST OF FIGURES

Figure		Page
1.1.	Location of the 240 Australian catchments featuring basic hydrologic model forcing data and calibration data. These catchments exhibit a wide range of characteristics (geology, topography, vegetation, climate, etc.) suitable for examining issues of model development, transferability, and uncertainty analysis.....	12
1.2	The Tenderfoot Creek Experimental Forest (TCEF) is located in the Little Belt Mountains of Montana, USA, and consists of eight (non-nested) catchments. The catchments of TCEF show variability in terms of geology, topography, and vegetation, but are consistent in terms of climate inputs	13
2.1.	Visual representation of the multilevel mixture logic at the foundation of the proposed likelihood architecture for use in ephemeral, zero-inflated catchments	23
2.2.	Simplified version of the Australian Water Balance Model structure with parameters S, BFI, and K.....	25
2.3.	Box plots of storage parameter derived from calibrated posterior distributions for each likelihood function of the synthetic case study with true value shown	25
2.4.	Quantile-quantile plots of the errors for each likelihood function of the synthetic case study. The Filliben r statistic is given on each subplot, representing the degree of similarity (coefficient of correlation) between the two quantiles. Likelihoods B and D include only the nonzero errors, as these likelihoods assume the zero errors do not belong to the same distribution as the nonzero errors.....	27
2.5.	Observed discharge at Wanalta Creek from 1961 through 1996.....	27
2.6.	Box plots of each calibrated parameter derived from the posterior distributions for each likelihood function of the Wanalta Creek case study	28

LIST OF FIGURES – CONTINUED

Figure		Page
2.7.	Quantile-quantile plots of the errors for both likelihood functions of the Wanalta Creek case study. The left panels show the results for the 25 year calibration period, and the right panels show the results for the 10 year validation period. The Filliben r statistic is given on each subplot, representing the degree of similarity between the two samples. Likelihood D includes only the nonzero errors, as these likelihoods assume the zero errors do not belong to the same distribution as the nonzero errors.....	29
3.1.	The likelihood function selection logic. A basic assumption of the form of the residuals is made and followed by checks on typical areas of deficiency at each level of the logic	68
3.2.	The simplified version of the Australian Water Balance Model structure with calibrated parameters S, BFI, and K.....	69
3.3.	Graphical representation of likelihood function performance for each synthetic data case and likelihood function. Scores are rated on a scale of (0-5), where a value of (5) represents success in each of the considered assessment criteria	69
3.4.	Step-by-step application of the logic presented in Figure 1 for the synthetic data case corresponding to Likelihood H. The labels for the residual scatterplots are observed discharge on the x-axis and model residuals on the y-axis with units of mm/day for Step 2 and units of log(mm/day) for Steps 3, 4. The labels for the ACFs are lags on the x-axis and autocorrelations on the y-axis.....	70
3.5.	Box plots for the storage parameter (S, mm/day) of the AWBM model, as calibrated for each of the eight likelihood functions (A-H) of the real case study	71
3.6.	Quantile-quantile plots with Filliben r statistic for each of the eight likelihood functions (A-H) examined in the real case study	72
3.7.	Hydrograph (black line) and uncertainty bounds for a	

LIST OF FIGURES – CONTINUED

Figure		Page
	200-day snapshot from the 23-year simulation period for two likelihood functions (C and E) examined under the conditions of the real case study, highlighting the under-capture of the uncertainty bounds for Likelihood C on higher streamflows.....	73
4.1.	Illustration of the flow-corrected time transform used to account for positive skewness present in a hypothetical streamflow hydrograph. The top panel shows the untransformed hydrograph and the bottom panel shows the FCT transformed hydrograph. Lines between the two panels map streamflow events from the untransformed to the transformed hydrograph.....	92
4.2.	Histograms of the streamflow data shown in Figure 1 under the untransformed, Box-Cox transformed, and flow-corrected time transformed conditions. Note that the Box-Cox transformed data has been shifted for clarity of presentation (due to negatives occurring for flows original flows less than 1) and the flow-corrected time transformed histogram is in terms of flow time and not clock time (achieved by performing a linear interpolation of the FCT weights). The skewness present in the untransformed data and even the Box-Cox transformed data highlights the effectiveness of the FCT transformation at removing such skewness in the data (in the “flow” time domain)	93
4.3.	Box plots of the untransformed and flow-corrected time transformed streamflow distributions (for hydrograph shown in Figure 1). The skewness present in the untransformed data and the effectiveness of the FCT transform at removing such skewness in the data are highlighted, noting that the mean streamflow represents an “outlier” value for the untransformed data but falls very near to the median of the transformed data.....	94
4.4.	Schematic diagram of the probability distributed model (PDM)	94

LIST OF FIGURES – CONTINUED

Figure		Page
4.5.	The Box-Cox transformed, the untransformed, and the FCT transformed streamflow hydrograph for the Stringer Creek (snow-dominated) catchment application. Results are shown for each transform in both “clock” time and “flow-corrected” time. The Nash-Sutcliffe efficiency (NSE) is also given for each case (applied to the appropriate transformed space), with an optimal value equal to one.....	95
4.6.	Residual scatterplots and quantile-quantile plots of the model residuals for the Box-Cox transformed, untransformed, and flow-corrected time transformed Stringer Creek catchment streamflows. The Filliben statistic is also given for each case, with an optimal value equal to one. Note the units on the x-axis of the scatterplots are the same as those on the y-axis.....	96
4.7.	The Box-Cox transformed, the untransformed, and the FCT transformed streamflow hydrograph for the Currency Creek (dynamically responsive) catchment application. Results are shown for each transform in both “clock” time and “flow-corrected” time. The Nash-Sutcliffe efficiency (NSE) is also given for each case (applied to the appropriate transformed space), with an optimal value equal to one.....	97
4.8.	Residual scatterplots and quantile-quantile plots of the model residuals for the Box-Cox transformed, untransformed, and flow-corrected time transformed Currency Creek catchment streamflows. The Filliben statistic is also given for each case, with an optimal value equal to one. Note the units on the x-axis of the scatterplots are the same as those on the y-axis.....	98
5.1.	The Tenderfoot Creek Experimental Forest, with Stringer Creek catchment (used in modeling study) highlighted. Model input data was obtained from two sources: (1) the precipitation inputs (as rain plus snowmelt) were derived from the two SNOTEL stations located in TCEF (Onion Park and Stringer Creek) and (2) the evapotranspiration data was collected from the collocated eddy-covariance	

LIST OF FIGURES – CONTINUED

Figure		Page
	towers. The well transects measure shallow groundwater connectivity across the hillslope-riparian-stream continuum (on each side of the stream)	126
5.2.	The conceptual form of the Catchment Connectivity Model as defined by the interactions of the drivers of the model structure: landscape analysis, fluxes and stores, and hydrologic connectivity and thresholds	127
5.3.	A graphic representation of the Catchment Connectivity Model structure that highlights the relationship between the size of landscape units (as upslope accumulated area) and threshold function mediated hydrologic connectivity.....	128
5.4.	The marginal parameter posterior distributions (shown as boxplots) of the CCM and likelihood function. The posteriors for each parameter shown are generally well defined.....	129
5.5.	The observed and predicted streamflow hydrograph is shown, along with the simulated hydrologic connectivity for the warm-up year and the two subsequent analysis years. An active landscape unit (binary) connection to the stream is indicated by the presence of a gray circle marker. The blue shaded regions highlight the periods of data used for parameter estimation.....	130
5.6.	The storage content dynamics (simulated by CCM) over time for each of the 50 landscape unit storages. Note the log scale on the y-axis.....	130
5.7.	The predicted and observed connectivity duration curves (bottom) and the residual error (top) for the study analysis period. The plot indicates slight underestimation of connectivity for the model's predictions, most notable at intermediate discharge	131
6.1.	The Tenderfoot Creek Experimental Forest, with each of the eight catchments used in this study shown. Model forcing	

LIST OF FIGURES – CONTINUED

Figure		Page
	data was retrieved from two sources: (1) the precipitation inputs (as rain plus snowmelt) were derived from SNOTEL stations located within TCEF (the Onion Park station in the UTEN catchment and the Stringer Creek station near the outlet of the LSC catchment) and (2) the evapotranspiration data was gathered from the collocated eddy-covariance towers. The well transects measure shallow groundwater connectivity across the hillslope-riparian-stream continuum (on each side of the stream)	160
6.2.	Cumulative precipitation (as rain plus snowmelt) over the course of water year 2008 using three different data sources for the Stringer Creek catchment: (1) the Onion Park SNOTEL station; (2) the Stringer Creek SNOTEL station; and (3) and elevation-weighting of the two SNOTEL stations	161
6.3.	A schematic depiction of the Catchment Connectivity Model, highlighting the computational implementation for a single landscape unit	162
6.4.	A comparison of the streamflow hydrographs and model predicted binary connections for each catchment. Note that the first year of data was used for model warm-up (to allow for initialization of the model storages) and that the model was calibrated over the peak streamflow season (April-September) for the remaining two years of data (periods shown without shading). The NSE value shown is for the model calibration period only and is based on the Box-Cox transformation used in the model calibration. Refer to Figure 1 and Table 1 for catchment details.....	163
6.5.	Flow duration curves (both observed and predicted) for each of the catchments used in this study for the entirety of the second and third years of data.....	164
6.6.	Cross-catchment comparison of the connectivity duration curve for each of the eight catchments included in the study	

LIST OF FIGURES – CONTINUED

Figure		Page
	for the entirety of the second and third years of data. Note that the “observed” curve is based on an experimentally-derived relationship that has been extrapolated to each of the individual catchments	165
6.7.	Boxplots of the posterior distributions for each of the calibrated CCM parameters and each of the eight catchments investigated	166
6.8.	Comparison of streamflow hydrographs for the three different precipitation input time series. In each case, the first year was used for warm-up and the parameters were calibrated over the April-September period for each of the subsequent years (no shading)	167
6.9.	A flow duration curve (computed for years two and three of data) highlighting the differences in performance of the CCM in response to varying precipitation inputs at the Stringer Creek catchment.....	167
6.10.	A connectivity duration curve (computed for years two and three of data) highlighting the differences in the internal consistency of the CCM in response to varying precipitation inputs at the Stringer Creek catchment.....	168
6.11.	Boxplots of the posterior distributions for each of the calibrated CCM parameters at the Lower Stringer Creek (LSC) catchment for each of the three precipitation input data sets. Recall that the Onion Park SNOTEL has the highest measured precipitation followed by the elevation-weighted series and Stringer Creek SNOTEL	168
6.12.	Comparison of streamflow hydrographs for the basic CCM structure, the CCM with a riparian area-limited overland flow component, and the CCM with an unrestricted overland flow component. In each case, the first year was used for warm-up and the parameters were calibrated over the April-September period for each of the subsequent years (no shading)	169

LIST OF FIGURES – CONTINUED

Figure		Page
6.13.	A comparison of the flow duration curves for the CCM, the CCM with riparian area-limited overland flow, and CCM with unrestricted overland flow. The first year of data was used for model warm-up and was excluded from the computation of flow durations	169
6.14.	A comparison of the connectivity duration curves for the CCM, the CCM with riparian area-limited overland flow, and CCM with unrestricted overland flow. The first year of data was used for model warm-up and was excluded from the computation of connectivity durations.....	170
7.1.	Diagram outlining the hierarchical architecture of the Bayes Empirical Bayes methodology	201
7.2.	Schematic diagram of the probability distributed model (PDM)	201
7.3.	Location of the catchments used in this study. (Left) Each of the 240 catchments in the dataset is shown. (Right) The reduced set of 118 catchments are shown (non-ephemeral with at least ten years of continuous data), along with the sites used as donors for the ungauged target catchment based on hierarchical cluster analysis	202
7.4.	Plot displaying the posterior mode for each parameter and each case, superimposed on the global prior distribution estimated based on the composite of the cases for the synthetic test application.....	203
7.5.	Box plots of the global prior distributions (based on “optimal” hyperparameters) with the true parameter values for the “ungauged” catchment shown with the target symbol. Note that the true parameter value is captured within the whiskers ($\pm 2.7\sigma$; 99.3% coverage) in all cases and within the 25 th and 75 th percentiles for parameters <i>C</i> _{MAX} , <i>B</i> , <i>TR1</i> , and <i>TR2</i>	204

LIST OF FIGURES – CONTINUED

Figure		Page
7.6.	Results of the BEB analysis for the parameter transfer to the “ ungauged ” catchment for the synthetic data application. The plot shows the observed hydrograph of the “ ungauged ” catchment, the 90% confidence interval width shown using the corresponding color gradient, and the interval reliability (96.6% observations captured) represented as a binary plot (present = interval captured observation; absent = interval failed to capture observation).....	205
7.7.	A dendrogram illustrating the results of the hierarchical clustering analysis based on the optimum parameter sets determined with the DDS algorithm. Note the x-axis indicates nodes (1-30) that contain all of the 240 catchments in the data set and the y-axis represents the relative distance between nodes (based on linkage height). For Test Case 2, the ungauged catchment and the catchments corresponding to the five and ten cluster selected scenarios were contained within node 6	206
7.8.	Box plots of the global prior distributions (based on “ optimal ” hyperparameters) for each of the scenarios (5 randomly selected donors; 5 cluster selected donors; 10 cluster selected donors) examined in the real data test application, with the optimal parameter values for the “ ungauged ” catchment (obtained with DDS algorithm) shown by the vertical dashed line in each subplot.....	207
7.9.	Plots of the 90% confidence intervals and observed streamflow for the “ ungauged ” catchment for a subset of the real data used in the study. The blue triangles represent the observations captured by the 90% interval and the inverted red triangles represent the observations missed by the 90% interval	208

ABSTRACT

Water is a fundamental resource, essential for human activities and underpinning economic and environmental sustainability. As such, management of water resources is a critical task, even more so in this time of climate uncertainty and change. Though hydrologic models have long been used to address water resource management, there is a growing desire to develop more reliable, transferable, and physically meaningful hydrologic models in light of data scarcity and uncertainty. In this dissertation, the three primary topics of (1) statistical implementation of models under uncertainty, (2) hydrologic model development, and (3) model applicability/transferability under data scarcity are considered. As a result of this research, we have: (1) developed methods for improved statistical implementation of hydrologic models through the development of a formal likelihood function for ephemeral catchments, the creation of a framework for identifying adequate likelihood functions across catchment conditions, and the introduction of a time domain transformation that was not previously used in hydrologic modeling; (2) developed a new model of catchment hydrologic connectivity based on extensive empirical observations that was shown to be consistent with both external and internal hydrologic variables, as well as transferable; and (3) introduced a new hierarchical Bayesian approach that was shown to accurately quantify uncertainty at an ungauged catchment based on pooled information from similar gauged catchments. This dissertation is divided into six key chapters/manuscripts that each address various aspects involved in improving hydrologic model performance, whose cumulative contribution results in a better understanding of how to construct, implement, and apply hydrologic models to lead to improved water resource management and planning problems of societal importance.

INTRODUCTION

Scope and Context

Water is a fundamental resource, essential for human activities and underpinning economic and environmental sustainability. Management of available water resources is a critical task, even more so in this era of a changing climate beset with uncertainty and non-stationarity [*Blöschl and Montanari, 2010; Milly et al., 2008*]. There exists an outstanding need for improved forecasting of available water resources at the catchment scale. Hydrologic models are conceptually well suited to this task, allowing simulation of hydrologic processes using climatic inputs.

However, despite the widespread usage of hydrologic models, the ability of these models to accurately simulate known hydrologic processes is still generally limited. This is particularly highlighted in the necessity for model calibration to the observed record of streamflow [*Madsen et al., 2002*]. However, the difficulty in models simulating observed processes, particularly those internal to the model calibration, is somewhat surprising given the extensive field research that has been focused on understand streamflow generation processes [e.g., *Kirchner, 2006; Seibert and McDonnell, 2002; Wagener, 2003*].

Dating back to the early 1930's, research focused on identifying and understanding streamflow generation processes has been an important part of the science of hydrology. A number of benchmark studies have shaped the view of hydrology both as a science and as a practice (e.g., design, planning, forecasting, etc.) from research into

infiltration processes [e.g., *Horton*, 1933] to variable source areas contributing to streamflow [e.g., *Dunne and Black*, 1970; *Hewlett and Hibbert*, 1967; *Ragan*, 1968] to the movement of water through the subsurface [e.g., *Anderson and Burt*, 1978; *Beasley*, 1976; *Freeze*, 1972; *Sklash and Farvolden*, 1979; *Weyman*, 1970]. At the same time hydrologic models have proliferated, some in response to new understandings of streamflow generating mechanisms [e.g., *Beven and Kirkby*, 1979; *Engman and Rogowski*, 1974; *Freeze*, 1972; 1972]. And while hydrologic modeling has become more and more capable, coinciding with the advancement of computing capabilities, there remains a difficulty in keeping pace with the multitude of newly identified dominant streamflow generating mechanisms at specific locations across a wide range of conditions [*Seibert and McDonnell*, 2002]. This fact coupled with a general inability to reliably identify catchments that should behave similarly to one another [e.g., *Blöschl*, 2006; *McDonnell and Woods*, 2004; *McDonnell et al.*, 2007; *Wagener et al.*, 2007], has led to the necessary usage of calibration techniques to tune existing hydrologic models for a specific catchment of interest.

Model calibration has long been a necessary element in hydrologic modeling studies due to model parameters that lack physical meaning [e.g., black-box or conceptual models; *Wagener and Gupta*, 2005] or are difficult to measure at the appropriate scale [e.g., conceptual models or physics-based models; *Blöschl and Sivapalan*, 1995]. At the same time, there has been an increasing interest in understanding and quantifying the uncertainty of model predictions that are caused by the inability to limit a number of factors including: randomness in weather processes,

mathematical description of model catchment behavior, true values of model parameters, and true measurements of observational data [*Montanari et al.*, 2009].

The popularity of Bayesian statistical approaches has increased in hydrologic applications due to its convenient framework within which uncertainty can be formally addressed through the use of probability theory [*Gelman et al.*, 2004]. Bayesian methods have been extensively used to quantify uncertainty due to a number of factors, including that arising from the parameterization [e.g., *Bates and Campbell*, 2001; *Kuczera*, 1983; *Kuczera and Parent*, 1998; *Marshall et al.*, 2004; *Smith and Marshall*, 2008], the model structure [e.g., *Butts et al.*, 2004; *Clark et al.*, 2008; *Marshall et al.*, 2005; *Smith and Marshall*, 2010], and the input data [e.g., *Kavetski et al.*, 2006; *Kuczera et al.*, 2006; *Renard et al.*, 2010].

Despite the difficulties in reconciling these sources of uncertainty there is a light at the end of the tunnel; hydrologists are beginning to focus more on both inter- and intra-disciplinary research that engages scientists interested in identifying dominant mechanisms of streamflow with those interested in building models that effectively utilize such information [e.g., *Bonell et al.*, 2006; *Clark et al.*, 2011; *Dunn et al.*, 2008; *Fenicia et al.*, 2008; *Gupta et al.*, 2008; *Kuraś et al.*, 2011; *Seibert and McDonnell*, 2002; *Son and Sivapalan*, 2007; *Vaché et al.*, 2004; *Wagener and Montanari*, 2011; *Weiler and McDonnell*, 2004]. Much of this hydrological engagement has arisen out of the International Association of Hydrological Sciences' (IAHS) decadal initiative on Predictions in Ungauged Basins (PUB) [*Sivapalan et al.*, 2003]. The PUB initiative serves as an ideal template upon which such research can be undertaken. The primary

objective of the PUB initiative is to improve upon the current ability to make predictions in catchments that lack calibration data implicitly requires (i) improved understanding of streamflow mechanisms, (ii) improved catchment classification techniques, and (iii) improved models and modeling approaches.

Research Motivation

Hydrologic modeling is extremely complicated in nature due to difficulty in adequately representing complex, dynamic catchment processes mathematically [Montanari *et al.*, 2009]. Inherently modeling studies are built upon and subjected to a number of interacting components, which often have multiple competing objectives or goals. These include but are not limited to: (1) development of model structures that are flexible enough to be applicable across a range of catchments (with varying climate, streamflow mechanisms, etc.) and conceptually sound enough to represent internal catchment variables adequately; (2) implementation of model calibration techniques to estimate unknown model parameters with an understanding of their associated uncertainty; and (3) an ability to make predictions for catchments that lack data to support (1) and (2). In light of these diverse goals, a broad yet integrated study has been undertaken here to address the interplay between hydrologic model, development, calibration, and predictive performance.

We aimed to address the following objectives in an iterative manner through several studies and applications:

1. Investigate the importance of statistical assumptions arising from the use of Bayesian analysis and other calibration approaches in hydrologic modeling applications.
 - a. Develop a formal likelihood function suitable for continuous modeling of ephemeral (non-continuous streamflow) catchments to be used within the Bayesian framework.
 - b. Formalize a methodology for selecting statistically appropriate likelihood functions to be used in Bayesian analysis.
 - c. Explore the applicability of time-domain transformations for improving the ability of model residuals to satisfy the statistical assumptions inherent to any model calibration technique (Bayesian or otherwise).
2. Develop a new model conceptualization of catchment generation processes based on new hydrologic theory.
 - a. Utilize extensive field observations from an experimental catchment to form the conceptual underpinnings of a new hydrologic model.
 - b. Assess the ability of the developed model to simulate both internal and external model variables across multiple catchments.
 - c. Test the sensitivity of the model structure to variation in model input data sources and alternate representations of streamflow generating mechanisms.
3. Explore the application of hydrologic models to ungauged catchments.

- a. Examine the utility of hierarchical Bayesian analysis in constraining and quantifying predictive uncertainty.
- b. Understand the role of donor catchment quality (i.e., hydrologic similarity) and quantity on predictive uncertainty for ungauged catchments.

This dissertation has made use of two primary data resources: (1) an extensive dataset of 200⁺ catchments located across Australia (primarily the southeast; Figure 1) that includes basic daily hydroclimatic data required to run and calibrate simple hydrologic models (precipitation, evapotranspiration, and observed streamflow) and (2) a detailed dataset for the Tenderfoot Creek Experimental Forest located in central Montana, USA, that includes both basic hydrologic modeling data and additional detailed data about streamflow generation mechanisms (Figure 2). Through the use of these unique data sets, there was great potential to implement diverse modeling studies with the goal of developing a better understanding of hydrologic model applications across the data availability continuum.

Dissertation Outline

The following chapters of this dissertation address conceptual hydrologic modeling, through insights into the usage of Bayesian analysis, the development of hydrologic models, and applications to predictions in ungauged basins. The chapters have been organized to begin with a focus on the details of Bayesian analysis to build a solid foundation for the techniques that are utilized for subsequent chapters focusing on model development and application.

Chapter 2

Bayesian analysis has been frequently used in hydrologic modeling studies over the past 15 years. Despite the extensive usage of such approaches to model parameter estimation and uncertainty analysis, the assumptions that drive the subsequent analysis are often overlooked or ignored. Chapter 2 focuses on developing a better understanding of the role that statistical assumptions associated with the likelihood function of Bayesian analysis play in terms of model performance and uncertainty analysis. The study concentrates in particular on ephemeral catchments that pose specific issues for continuous modeling due to distinct states of streamflow and no streamflow. We developed a formal likelihood function that allows the Bayesian approach to be used for ephemeral catchments to address the following questions:

1. What is the impact of failing to satisfy the statistical assumptions made in Bayesian analysis (via the likelihood function), when applied to ephemeral catchments, on parameter identification?
2. How does application of the newly developed likelihood function affect model performance and uncertainty analysis?

Chapter 3

Chapter 3 generalizes the findings regarding the application of Bayesian inferential approaches to ephemeral catchments (Chapter 2) to all catchments. In particular, this chapter sought to formalize a procedure for selecting appropriate likelihood functions (i.e., one's whose statistical assumptions are adequately satisfied)

across multiple catchment conditions through the development of a likelihood function selection framework. As such, we introduced a simple, conceptual framework that addressed the most commonly violated assumptions in hydrologic modeling applications. The framework included both visual and statistical checks to ensure that issues such as zero-inflated, heteroscedastic, and/or autocorrelated residuals were identified and accounted for under a top-down (simple to complex) approach. The following questions were addressed in this study:

1. Does an iterative approach to likelihood specification improve selection of appropriate likelihood functions?
2. Does the approach result in the selection of the simplest likelihood function that satisfies the assumption criteria?

Chapter 4

Transformations are common tools in hydrologic modeling that can often be useful in terms of satisfying important statistical assumptions. For example, log transforms are often used to normalize residuals. In Chapter 4 we propose a transformation on the time domain rather than the data domain. The transformation, which weights streamflow at each time step according to its proportion relative to mean streamflow, was applied to two catchments of contrasting streamflow regimes to examine the usefulness of the transformation in answering the following questions:

1. Is a time-domain transformation useful at de-skewing positively skewed streamflow data?

2. How does this transformation compare to other commonly used transformations for model calibration?
3. What are the limitations of using this transformation for real-time modeling applications?

Chapter 5

A vital aspect to any modeling study is selecting an appropriate model structure. In Chapter 5, we shift focus from the details of how to appropriately apply Bayesian inferential approaches to the development of appropriate hydrologic models. In this chapter we focus on building a conceptually simple (i.e., small number of unknown parameters) and internally verifiable model structure, called the Catchment Connectivity Model. To do so, we made use of extensive field observations at the Tenderfoot Creek Experimental Forest that relate landscape structure to streamflow generation. We used the concept of the hydrologic connectivity between hillslopes, riparian areas, and the stream as the underlying premise of our model to address the following questions:

1. Will a model based on the frequency of hydrologic connections (rather than the magnitudes of the hillslopes connected to the stream) be able to represent the observed streamflow hydrograph following model calibration?
2. Will the model show good agreement with observations of hydrologic connectivity without calibrating model parameters to these data?

Chapter 6

The utility of any model is highly dependent upon its ability to perform well across a number of catchments (i.e., to be transferable). In a follow-up to study conducted in Chapter 5, in Chapter 6 we investigate the application of the Catchment Connectivity Model across eight neighboring catchments (including the catchment used in the Chapter 5 application). We investigate the performance of the model both externally (streamflow) and internally (hydrologic connectivity) in response to the following research questions:

1. How transferable is the Catchment Connectivity Model to other catchment of similar physical processes, in terms of both model performance and parameterization?
2. How sensitive are the model outputs (streamflow, hydrologic connectivity) to the model input data (precipitation)?
3. How sensitive are the model outputs to the inclusion of alternate process (overland flow) representations?

Chapter 7

The ability to make predictions at locations where streamflow data is unavailable for model calibration relies on using gauged proxy catchments as donors for the ungauged catchment. This is a well-studied problem for which many approaches have been developed and applied with varying degrees of success. However, to date there has been little research into developing methods that can formalize the uncertainty in the predictions at the ungauged locations. In Chapter 7 we explore the idea of using a hierarchical Bayesian approach to improve our ability to quantify the uncertainty of

predictions made in ungauged basins. The hierarchical Bayesian approach was applied to both synthetic and real data sets to explore the following questions:

1. Is a hierarchical Bayesian analysis able to produce predictions that quantify (and constrain) uncertainty accurately?
2. What is the sensitivity of the method to the quality (hydrologic similarity to the ungauged catchment) of donor catchments?
3. What is the sensitivity of the method to the quantity of donor catchments?

Chapter 8

In Chapter 8, the main findings of each of the chapters of this dissertation are summarized. The collection of chapters presented in this dissertation reflects an integration of the key aspects of hydrologic modeling including their development, statistical implementation, and application to real world problems of societal importance. Based on the conclusions of each of the chapters of this dissertation, recommendations are made for potential future studies that could utilize and extend the techniques presented herein.

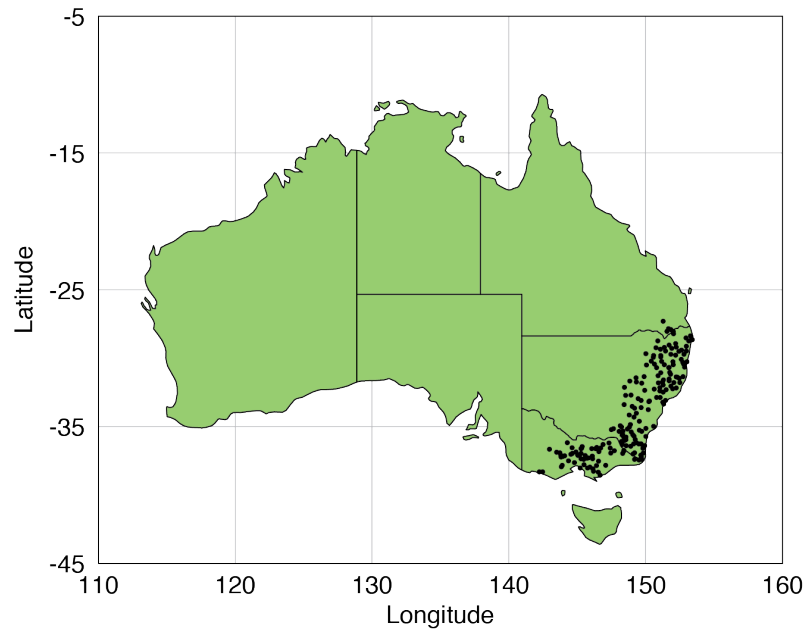
Figures

Figure 1. Location of the 240 Australian catchments featuring basic hydrologic model forcing data and calibration data. These catchments exhibit a wide range of characteristics (geology, topography, vegetation, climate, etc.) suitable for examining issues of model development, transferability, and uncertainty analysis.

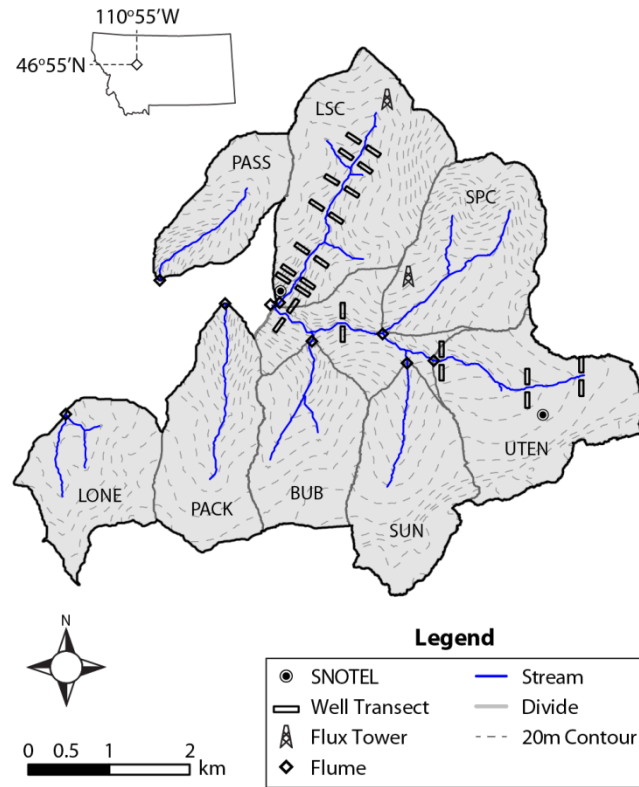


Figure 2. The Tenderfoot Creek Experimental Forest (TCEF) is located in the Little Belt Mountains of Montana, USA, and consists of eight (non-nested) catchments. The catchments of TCEF show variability in terms of geology, topography, and vegetation, but are consistent in terms of climate inputs.

References

- Anderson, M. G., and T. P. Burt (1978), The role of topography in controlling throughflow generation, *Earth Surface Processes*, 3(4), 331-344.
- Bates, B. C., and E. P. Campbell (2001), A Markov chain Monte Carlo scheme for parameter estimation and inference in conceptual rainfall-runoff modeling, *Water Resour. Res.*, 37(4), 937-947.
- Beasley, R. S. (1976), Contribution of Subsurface Flow from the Upper Slopes of Forested Watersheds to Channel Flow, *Soil Sci. Soc. Am. J.*, 40(6), 955-957.
- Beven, K. J., and M. J. Kirkby (1979), A physically based, variable contributing area model of basin hydrology, *Hydrol. Sci. Bull.*, 24(1), 43-69.
- Blöschl, G., and M. Sivapalan (1995), Scale issues in hydrological modeling: A review, *Hydrol. Processes*, 9(3-4), 251-290.
- Blöschl, G. (2006), Hydrologic synthesis: Across processes, places, and scales, *Water Resour. Res.*, 42, W03S02, doi:10.1029/2005WR004319.
- Blöschl, G., and A. Montanari (2010), Climate change impacts - throwing the dice?, *Hydrol. Processes*, 24(3), 374-381.
- Bonell, M., J. J. McDonnell, F. N. Scatena, J. Seibert, S. Uhlenbrook, and H. A. J. van Lanen (2006), HELPing FRIENDs in PUBs: charting a course for synergies within international water research programmes in gauged and ungauged basins, *Hydrol. Processes*, 20(8), 1867-1874.
- Butts, M. B., J. T. Payne, M. Kristensen, and H. Madsen (2004), An evaluation of the impact of model structure on hydrological modelling uncertainty for streamflow simulation, *J. Hydrol.*, 298(1-4), 242-266.
- Clark, M. P., A. G. Slater, D. E. Rupp, R. A. Woods, J. A. Vrugt, H. V. Gupta, T. Wagener, and L. E. Hay (2008), Framework for Understanding Structural Errors (FUSE): A modular framework to diagnose differences between hydrological models, *Water Resour. Res.*, 44, W00B02, doi:10.1029/2007WR006735.
- Clark, M. P., H. K. McMillan, D. B. G. Collins, D. Kavetski, and R. A. Woods (2011), Hydrological field data from a modeller's perspective: Part 2: process-based evaluation of model hypotheses, *Hydrol. Processes*, 25(4), 523-543.
- Dunn, S. M., J. Freer, M. Weiler, M. J. Kirkby, J. Seibert, P. F. Quinn, G. Lischied, D. Tetzlaff, and C. Soulsby (2008), Conceptualization in catchment modelling: simply learning?, *Hydrol. Processes*, 22(13), 2389-2393.

- Dunne, T., and R. D. Black (1970), Partial Area Contributions to Storm Runoff in a Small New England Watershed, *Water Resour. Res.*, 6(5), 1296-1311.
- Engman, E. T., and A. S. Rogowski (1974), A Partial Area Model for Storm Flow Synthesis, *Water Resour. Res.*, 10(3), 464-472.
- Fenicia, F., J. J. McDonnell, and H. H. G. Savenije (2008), Learning from model improvement: On the contribution of complementary data to process understanding, *Water Resour. Res.*, 44, W06419, doi:10.1029/2007WR006386.
- Freeze, R. A. (1972), Role of Subsurface Flow in Generating Surface Runoff 2. Upstream Source Areas, *Water Resour. Res.*, 8(5), 1272-1283.
- Freeze, R. A. (1972), Role of subsurface flow in generating surface runoff: 1. Base flow contributions to channel flow, *Water Resour. Res.*, 8(3), 609-623.
- Gelman, A., J. B. Carlin, H. S. Stern, and D. B. Rubin (2004), *Bayesian Data Analysis*, 2nd ed., Chapman & Hall/CRC, New York.
- Gupta, H. V., T. Wagener, and Y. Liu (2008), Reconciling theory with observations: elements of a diagnostic approach to model evaluation, *Hydrol. Processes*, 22(18), 3802-3813.
- Hewlett, J. D., and A. R. Hibbert (1967), Factors affecting the response of small watersheds to precipitation in humid areas, in *International Symposium on Forest Hydrology*, edited by W. E. Sopper and H. W. Lull, pp. 275-290, Pergamon, Oxford, UK.
- Horton, R. E. (1933), The role of infiltration in the hydrologic cycle, *Trans. Am. Geophys. Un.*, 14, 446-460.
- Kavetski, D., G. Kuczera, and S. W. Franks (2006), Bayesian analysis of input uncertainty in hydrological modeling: Theory, *Water Resour. Res.*, 42, W03407, doi:10.1029/2005WR004368.
- Kirchner, J. W. (2006), Getting the right answers for the right reasons: Linking measurements, analyses, and models to advance the science of hydrology, *Water Resour. Res.*, 42, W03S04, doi:10.1029/2005WR004362.
- Kuczera, G. (1983), Improved Parameter Inference in Catchment Models 1. Evaluating Parameter Uncertainty, *Water Resour. Res.*, 19(5), 1151-1162.
- Kuczera, G., and E. Parent (1998), Monte Carlo assessment of parameter uncertainty in conceptual catchment models: the Metropolis algorithm, *J. Hydrol.*, 211(1-4), 69-85.

- Kuczera, G., D. Kavetski, S. Franks, and M. Thyer (2006), Towards a Bayesian total error analysis of conceptual rainfall-runoff models: Characterising model error using storm-dependent parameters, *J. Hydrol.*, 331(1-2), 161-177.
- Kuraś, P. K., Y. Alila, M. Weiler, D. Spittlehouse, and R. Winkler (2011), Internal catchment process simulation in a snow-dominated basin: Performance evaluation with spatiotemporally variable runoff generation and groundwater dynamics, *Hydrol. Processes*, doi:10.1002/hyp.8037, in press.
- Madsen, H., G. Wilson, and H. C. Ammentorp (2002), Comparison of different automated strategies for calibration of rainfall-runoff models, *J. Hydrol.*, 261(1-4), 48-59.
- Marshall, L., D. Nott, and A. Sharma (2004), A comparative study of Markov chain Monte Carlo methods for conceptual rainfall-runoff modeling, *Water Resour. Res.*, 40, W02501, doi:10.1029/2003WR002378.
- Marshall, L., D. Nott, and A. Sharma (2005), Hydrological model selection: A Bayesian alternative, *Water Resour. Res.*, 41, W10422, doi:10.1029/2004WR003719.
- McDonnell, J. J., and R. Woods (2004), On the need for catchment classification, *J. Hydrol.*, 299(1-2), 2-3.
- McDonnell, J. J., M. Sivapalan, K. Vaché, S. Dunn, G. Grant, R. Haggerty, C. Hinz, R. Hooper, J. Kirchner, M. L. Roderick, J. Selker, and M. Weiler (2007), Moving beyond heterogeneity and process complexity: A new vision for watershed hydrology, *Water Resour. Res.*, 43, W07301, doi:10.1029/2006WR005467.
- Milly, P. C. D., J. Betancourt, M. Falkenmark, R. M. Hirsch, Z. W. Kundzewicz, D. P. Lettenmaier, and R. J. Stouffer (2008), Stationarity Is Dead: Whither Water Management?, *Science*, 319(5863), 573-574.
- Montanari, A., C. A. Shoemaker, and N. van de Giesen (2009), Introduction to special section on Uncertainty Assessment in Surface and Subsurface Hydrology: An overview of issues and challenges, *Water Resour. Res.*, 45, W00B00, doi:10.1029/2009WR008471.
- Ragan, R. M. (1968), An experimental investigation of partial area contributions, in *Hydrological Aspects of the Utilization of Water, Reports and Discussions (Proc. IAHS Assembly at Bern)*, edited, pp. 241–251.
- Renard, B., D. Kavetski, G. Kuczera, M. Thyer, and S. W. Franks (2010), Understanding predictive uncertainty in hydrologic modeling: The challenge of identifying input and structural errors, *Water Resour. Res.*, 46, W05521, doi:10.1029/2009WR008328.

- Seibert, J., and J. J. McDonnell (2002), On the dialog between experimentalist and modeler in catchment hydrology: Use of soft data for multicriteria model calibration, *Water Resour. Res.*, 38, 1241, doi:10.1029/2001WR000978.
- Sivapalan, M., K. Takeuchi, S. W. Franks, V. K. Gupta, H. Karambiri, V. Lakshmi, X. Liang, J. J. McDonnell, E. M. Mondono, P. E. O'Connell, T. Oki, J. W. Pomeroy, D. Schertzer, S. Uhlenbrook, and E. Zehe (2003), IAHS Decade on Predictions in Ungauged Basins (PUB), 2003-2012: Shaping an exciting future for the hydrological sciences, *Hydrol. Sci. J.*, 48(6), 857-880.
- Sklash, M. G., and R. N. Farvolden (1979), The role of groundwater in storm runoff, *J. Hydrol.*, 43(1-4), 45-65.
- Smith, T. J., and L. A. Marshall (2008), Bayesian methods in hydrology: A study of recent advancements in Markov chain Monte Carlo techniques, *Water Resour. Res.*, 44, W00B05, doi:10.1029/2007WR006705.
- Smith, T. J., and L. A. Marshall (2010), Exploring uncertainty and model predictive performance concepts via a modular snowmelt-runoff modeling framework, *Environ. Modell. Software*, 25(6), 691-701.
- Son, K., and M. Sivapalan (2007), Improving model structure and reducing parameter uncertainty in conceptual water balance models through the use of auxiliary data, *Water Resour. Res.*, 43, W01415, doi:10.1029/2006WR005032.
- Vaché, K. B., J. J. McDonnell, and J. Bolte (2004), On the use of multiple criteria for a posteriori model rejection: Soft data to characterize model performance, *Geophys. Res. Lett.*, 31, L21504, doi:10.1029/2004GL021577.
- Wagener, T. (2003), Evaluation of catchment models, *Hydrol. Processes*, 17(16), 3375-3378.
- Wagener, T., and H. V. Gupta (2005), Model identification for hydrological forecasting under uncertainty, *Stochast. Environ. Res. Risk Assess.*, 19(6), 378-387.
- Wagener, T., M. Sivapalan, P. Troch, and R. Woods (2007), Catchment Classification and Hydrologic Similarity, *Geography Compass*, 1(4), 901-931.
- Wagener, T., and A. Montanari (2011), Convergence of approaches toward reducing uncertainty in predictions in ungauged basins, *Water Resour. Res.*, 47, W06301, doi:10.1029/2010WR009469.
- Weiler, M., and J. McDonnell (2004), Virtual experiments: a new approach for improving process conceptualization in hillslope hydrology, *J. Hydrol.*, 285(1-4), 3-18.

Weyman, D. R. (1970), Throughflow on hillslopes and its relation to the stream hydrograph, *Bull. Int. Assoc. Sci. Hydrol.*, 15(2), 25-33.

CHAPTER TWO

DEVELOPMENT OF A FORMAL LIKELIHOOD FUNCTION FOR IMPROVED
BAYESIAN INFERENCE OF EPHEMERAL CATCHMENTSContribution of Author and Co-Authors

Manuscript in Chapter 2

Author:	Tyler Smith
Contributions:	Responsible for the development and analysis of technique introduced, the majority of the writing, and figure development.
Co-Author:	Ashish Sharma
Contributions:	Provided extensive discussions of the development of the methods used in the study and contributed to the editing of the manuscript throughout its development.
Co-Author:	Lucy Marshall
Contributions:	Provided extensive discussions of the development of the methods used in the study and contributed to the editing of the manuscript throughout its development.
Co-Author:	Raj Mehrotra
Contributions:	Performed initial data quality control, was involved in technical discussions throughout the manuscript's development, and edited the final version of the paper.
Co-Author:	Scott Sisson
Contributions:	Provided technical discussions throughout the manuscript's development and edited the final version of the paper.

Manuscript Information

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

- Journal: Water Resources Research
- Status of manuscript (check one)
 - ☐ Prepared for submission to a peer-reviewed journal
 - ☐ Officially submitted to a peer-reviewed journal
 - ☐ Accepted by a peer-reviewed journal
 - ☒ Published in a peer-reviewed journal
- Publisher: American Geophysical Union
- Date of submission: May 3, 2010
- Date manuscript was published: September 16, 2010
- Issue in which manuscript appeared: Volume 46, W12551

Reproduced by permission of American Geophysical Union.

Development of a formal likelihood function for improved Bayesian inference of ephemeral catchments

Tyler Smith,^{1,2} Ashish Sharma,² Lucy Marshall,¹ Raj Mehrotra,² and Scott Sisson³

Received 3 May 2010; revised 3 September 2010; accepted 15 September 2010; published 28 December 2010.

[1] The application of formal Bayesian inferential approaches in hydrologic modeling is often criticized for requiring explicit assumptions to be made about the distribution of the errors via the likelihood function. These assumptions can be adequate in some situations, but often little attention is paid to the selection of an appropriate likelihood function. This paper investigates the application of Bayesian methods in modeling ephemeral catchments. We consider two modeling case studies, including a synthetically generated data set and a real data set of an arid Australian catchment. The case studies investigate some typical forms of the likelihood function which have been applied widely by previous researchers, as well as introducing new implementations aimed at better addressing likelihood function assumptions in dry catchments. The results of the case studies indicate the importance of explicitly accounting for model residuals that are highly positively skewed due to the presence of many zeros (zero inflation) arising from the dry spells experienced by the catchment. Specifically, the form of the likelihood function was found to significantly impact the calibrated parameter posterior distributions, which in turn have the potential to greatly affect the uncertainty estimates. In each application, the likelihood function that explicitly accounted for the nonconstant variance of the errors and the zero inflation of the errors resulted in (1) equivalent or better fits to the observed discharge in both timing and volume, (2) superior estimation of the uncertainty as measured by the reliability and sharpness metrics, and (3) more linear quantile-quantile plots indicating the errors were more closely matched to the assumptions of this form of the likelihood function.

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

1. Introduction

[2] The use of conceptual hydrologic models necessitates some type of calibration procedure to determine optimum values for the models' parameters such that the predicted outcome most closely matches the observed. Typically this will be done either manually or automatically, with automatic methods becoming the primary method of use recently [Madsen *et al.*, 2002]. Automatic calibration can be performed by a variety of methods such as global optimization algorithms [e.g., Duan *et al.*, 1992], Monte Carlo methods [e.g., Beven and Binley, 1992], and Bayesian inferential approaches [e.g., Kuczera and Parent, 1998], and many techniques have the ability to incorporate multiple objective measures into their calibration logic [e.g., Vrugt *et al.*, 2003].

[3] As interest in quantifying the uncertainty in model predictions has grown, the use of Monte Carlo (such as the generalized likelihood uncertainty estimation (GLUE)

methodology) and Bayesian inferential approaches have gained interest, as well. Much debate has centered on the differences between these methods [see Beven *et al.*, 2007; Mantovan and Todini, 2006; Mantovan *et al.*, 2007]; in actuality these methods can be looked at as special cases of one another (i.e., GLUE is a less statistically strict form of the formal Bayesian approach).

[4] Bayesian inference is based on the application of Bayes' theorem, which states that

$$P(\theta|Q) \propto P(\theta)P(Q|\theta), \quad (1)$$

where Q is the data, θ is the parameter set, $P(\theta|Q)$ is the posterior distribution, $P(\theta)$ is the prior distribution, and $P(Q|\theta)$ is the likelihood function summarizing the model (for the input data) given the parameters. Even a cursory examination of equation (1) reveals that the choice of the likelihood function will play an important role. Because of its importance to the resultant posterior distribution, the likelihood function must make explicit assumptions about the form of the model residuals [Stedinger *et al.*, 2008].

[5] It is in the likelihood function where the GLUE method largely deviates from the formal Bayesian approach. The GLUE methodology typically opts for informal likelihood (objective) functions rather than risking violating the strong assumptions that arise from the use of formal likelihood functions, with critics of the Bayesian approach

¹Department of Land Resources and Environmental Sciences, Montana State University, Bozeman, Montana, USA.

²School of Civil and Environmental Engineering, University of New South Wales, Sydney, New South Wales, Australia.

³School of Mathematics and Statistics, University of New South Wales, Sydney, New South Wales, Australia.

holding the overriding view that the appropriateness of such assumptions is inherently difficult to satisfy in hydrologic problems, and thus the use of Bayesian methods is problematic [Beven, 2006a].

[6] Despite this criticism, the use of Bayesian methods has become increasingly common [e.g., Kuczera and Parent, 1998; Marshall et al., 2004; Samanta et al., 2007; Smith and Marshall, 2008, 2010; Vrugt et al., 2009]. However, the predominant usage of Bayesian methods in hydrology has been under the assumption of uncorrelated, Gaussian errors, as evidenced by a plethora of studies [e.g., Ajami et al., 2007; Campbell et al., 1999; Duan et al., 2007; Hsu et al., 2009; Marshall et al., 2004; Samanta et al., 2008; Smith and Marshall, 2010; Vrugt et al., 2006]. Although there are undoubtedly situations (perhaps catchments with a very simple hydrograph) in which the errors are adequately represented by such a likelihood function, these scenarios are more likely to be the exception than the rule in hydrology.

[7] Because the posterior distribution is proportional to the prior distribution multiplied by the likelihood function (equation (1)), it is important to consider the form of the errors (assumed by the likelihood) when implementing the Bayesian method. Given the tendency for errors arising from hydrologic models to be (at least) heteroscedastic [e.g., Kuczera, 1983; Sorooshian and Dracup, 1980], the application of likelihood functions should not be chosen on the basis of computational simplicity. The ultimate effectiveness of Bayesian inference relies on properly characterizing the form of the errors via the formal likelihood function and holds untapped potential for improvement in uncertainty estimation.

[8] Xu [2001, p. 77] points out that “in the field of hydrological modeling, few writers examine and describe any properties of residuals given by their models when fitted to the data”. This provides a potential limiting constraint on more appropriate and widespread use of Bayesian inference. The need to address the lack of normality in hydrologic modeling errors is not new, but there is little guidance on exactly how to go about selecting an appropriate likelihood for a particular data-model combination.

[9] Although the assumption of normality is common to hydrological modeling studies, several alternatives aimed at addressing typical characteristics of rainfall-runoff modeling errors have been made. For example, Bates and Campbell [2001] considered multiple likelihood functions that aimed to address the issues of nonconstant variance (via a Box-Cox transformation) and correlation (via an autoregressive model) in the modeled errors. Marshall et al. [2006] considered a Student's t distribution with four degrees of freedom due to its higher peaks and heavier tails. More recently, Schaeffli et al. [2007] attempted to address the nonconstant variance problem through the use of a mixture likelihood corresponding to high and low flow states. In arid regions, there is further difficulty in appropriately capturing the form of the errors caused by extended periods of no-flow conditions. This can lead to errors that are severely zero inflated (i.e., a significant proportion of the residuals are zero) due to the ability of the model to correctly predict a zero flow at a zero flow observation.

[10] In spite of the variations in the likelihood function outlined above, the issue of residuals originating from distinct flow states, as is the case with ephemeral catchments

worldwide, has not been addressed at length in the literature. In this paper, we focus on the development and application of a likelihood function that specifically focuses on addressing the potential problems uniquely caused by zero-inflated errors under a Bayesian inferential approach. This paper is divided into the following sections: section 2 introduces a formal likelihood function that addresses the zero-inflation problem common in arid catchments, section 3 introduces two test applications for the formal likelihood function to be analyzed, and section 4 provides a discussion of the results and the important conclusions of the study.

2. Formal Likelihood Function Development

[11] Conceptual hydrologic models can be thought of, in a generic sense, as being of the form

$$Q_t = f(x_t; \theta) + \varepsilon_t, \quad (2)$$

where t indexes time, Q_t is the observed discharge, $f(x_t; \theta)$ is the model predicted discharge, x_t is the model forcing data (typically, rainfall and evapotranspiration), θ is the set of unknown model parameters, and ε_t is the error.

[12] Proper understanding of the form of the errors (ε_t) is vital to the success of Bayesian inference and must be properly modeled via the formal likelihood function. Because arid catchments can have extended periods of no streamflow, the entire streamflow record can be thought of as being composed of a mixture of two dominant streamflow states (zero discharge and nonzero discharge). Conceptualization of the observed streamflow record into multiple states or regimes is typical of operational rainfall-runoff modeling where the design of an engineered feature depends on a corresponding flow regime [see Wagener and McIntyre, 2005]. We propose the use of a formal likelihood function that is formulated as a mixture model with components corresponding to the dominant discharge states. The standard form of the finite mixture distribution [e.g., Robert, 1996] can be defined as

$$g(\varepsilon_t) \approx \hat{g}(\varepsilon_t) = \sum_{m=1}^M w_m f_m(\varepsilon_t | \phi_{t_m}), \quad (3a)$$

where $g(\varepsilon_t)$, the true error distribution, is approximated by a mixture of components such that $w_1 + \dots + w_M = 1$, $f_m(\varepsilon_t | \phi_{t_m})$, are the components of the mixture (probability density functions with errors ε_t and parameters ϕ_{t_m}), and M is the number of mixture components. For a mixture where the error-generating component is known, the likelihood function can be written as [Robert, 1996]

$$P(Q|\theta) = \prod_{t=1}^n g(\varepsilon_t) = \prod_{t,m=1} f(\varepsilon_t | \phi_1) \cdots \prod_{t,m=M} f(\varepsilon_t | \phi_M), \quad (3b)$$

where n is the number of observed values, ε_t are the errors generated by the component m , and all other terms are as previously defined. Mixture models have not been used extensively in hydrology, although Schaeffli et al. [2007] recently introduced a mixture of two normal distributions (corresponding to low and high discharge states) likelihood function that attempts to address the nonconstant variance problem common to residuals in hydrological modeling.

[13] The mixture likelihood proposed here for arid catchments is conditioned on the discharge state and is

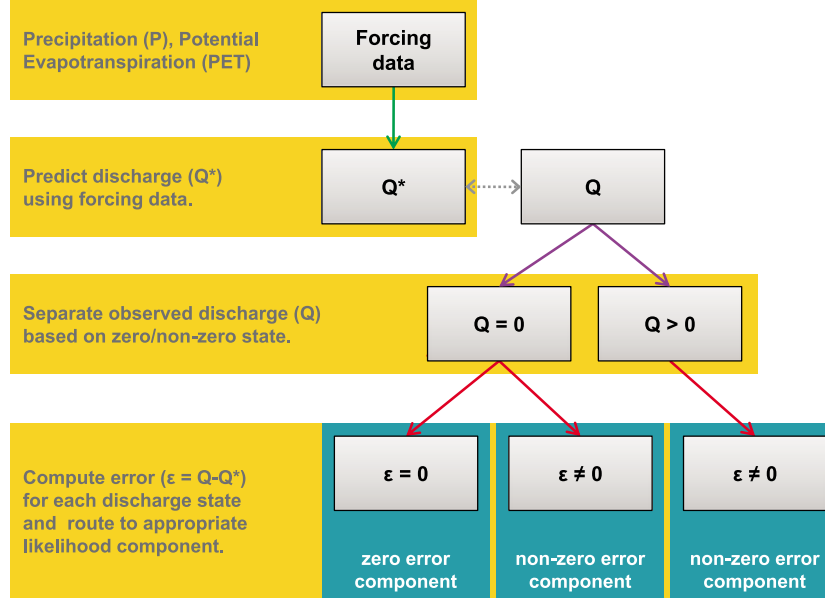


Figure 1. Visual representation of the multilevel mixture logic at the foundation of the proposed likelihood architecture for use in ephemeral, zero-inflated catchments.

formed using two primary components: (1) a component corresponding to the zero discharge state and (2) a component corresponding to the nonzero discharge state. The component addressing the zero discharge state is further separated into a secondary mixture based on the zero/nonzero residual state, modeled as a mixture of a binomial distribution for the zero residual state and a Gaussian (or transformed Gaussian) distribution for the nonzero residual state. The component corresponding to the nonzero discharge state is modeled as a simple Gaussian (or transformed Gaussian) distribution. Figure 1 represents this mixture likelihood graphically. The use of a multilevel mixture model is necessary due to the fact that only zero discharges have a realistic probability of being modeled exactly ($\epsilon_i = 0$) due to double precision accuracy of the machine. Given this knowledge, the multilevel mixture likelihood is able to properly assign uncertainty to the different discharge states. The use of a zero/nonzero mixture model is not entirely new in environmental applications [see Tu, 2002], dating back to the 1950s and the delta model introduced by Aitchison [1955], but it has been used more recently in medical cost data applications [see Tu and Zhou, 1999; Zhou and Tu, 1999, 2000].

[14] Mathematically, the likelihood function can then be expressed as the product of three components such that equation (3b) becomes

$$P(Q|\theta) = \{[\rho^{n_1}][1 - \rho]^{n_2} L_0|_{n_2}\} \{L_1|_{n_3}\}, \quad (4)$$

where ρ is the probability of a zero residual and is computed as $n_1/(n_1 + n_2)$, n_1 is the number of zero discharge observations modeled with zero error, n_2 is the number of zero discharge observations modeled with nonzero error, n_3 is the number of nonzero discharge observations (modeled with

nonzero error), the total number of observations is $N = n_1 + n_2 + n_3$, L_0 is the likelihood function for the zero discharge observations with nonzero errors (evaluated over n_2), and L_1 is the likelihood function for the nonzero discharge observations with nonzero errors (evaluated over n_3).

[15] In this study we compare the appropriateness of this mixture likelihood to several others. Table 1 introduces the four likelihood functions to be examined. For simplicity the likelihood functions have been labeled (A–D). Likelihood A represents the typical likelihood function used in hydrologic modeling studies; it assumes the errors are Gaussian and uncorrelated and does not account for zero inflation in the data. Likelihood B holds the same assumptions of normality that are made in likelihood A; however, the zero-inflation problem is considered explicitly under the form of equation (4). Likelihood C again ignores the zero inflation (as with likelihood A), but assumes the errors are heteroscedastic in normal space. A Box-Cox transformation [Box and Cox, 1964] was applied to the data such that

$$Q^{(\lambda)} = \log(Q + \lambda), \text{ with } \lambda = 0.5. \quad (5)$$

The value of the transformation parameter (λ) was selected following the recommendation of Bates and Campbell [2001] for medium-sized Australian catchments that are highly positively skewed. Likelihood C assumes that the errors are homoscedastic in transformed space. Finally, likelihood D makes use of the conditional mixed likelihood function logic presented in equation (4) and the Box-Cox transformation presented in equation (5).

[16] Each of the likelihood functions has a single calibrated parameter (σ^2) that represents the variance of the errors. Likelihoods B and D could be constructed such that there are two variance parameters that correspond to the zero

Table 1. Mathematical Formulas and Parameters for Each of the Likelihood Functions Considered in This Study

Likelihood Function	Assumption	Mathematical Formula ^a	Calibrated Parameter
A	Independent, homoscedastic errors	$L_A = (2\pi\sigma^2)^{-N/2} \prod_N \exp\left(-\frac{\varepsilon_i^2}{2\sigma^2}\right)$	σ^2
B	Zero-inflated, independent, homoscedastic errors	$L_B = \rho^{n_1} \left\{ (1 - \rho)^{n_2} (2\pi\sigma^2)^{-n_2/2} \prod_{n_2} \exp\left[-\frac{\varepsilon_i^2}{2\sigma^2}\right] \right\} \times \left\{ (2\pi\sigma^2)^{-n_3/2} \prod_{n_3} \exp\left[-\frac{\varepsilon_i^2}{2\sigma^2}\right] \right\}$	σ^2
C	Independent, heteroscedastic errors transformed via Box-Cox transformation	$L_C = (2\pi\sigma^2)^{-N/2} \prod_N \exp\left[-\frac{(\varepsilon_i^*)^2}{2\sigma^2}\right] (Q_{\text{obs}} + \lambda)^{-1}$	σ^2
D	Zero-inflated, independent, heteroscedastic errors transformed via Box-Cox transformation	$L_D = \rho^{n_1} \left\{ (1 - \rho)^{n_2} (2\pi\sigma^2)^{-n_2/2} \prod_{n_2} \exp\left[-\frac{(\varepsilon_i^*)^2}{2\sigma^2}\right] (Q_{\text{obs}} + \lambda)^{-1} \right\} \times \left\{ (2\pi\sigma^2)^{-n_3/2} \prod_{n_3} \exp\left[-\frac{(\varepsilon_i^*)^2}{2\sigma^2}\right] (Q_{\text{obs}} + \lambda)^{-1} \right\}$	σ^2

^aWhere N is the number of observations in the data set, n_1 is the number of zero observations with zero error, n_2 is the number of zero observations with nonzero error, n_3 is the number of nonzero observations, ρ is the probability of a zero error given a zero observation, $\varepsilon = Q_{\text{obs}} - Q_p$, $\varepsilon^* = \log(Q_{\text{obs}} + \lambda) - \log(Q_p + \lambda)$, $\lambda = 0.5$, and σ^2 is the variance of the errors.

discharge with nonzero error state and the nonzero discharge with nonzero error state; however, owing to the small number of zero discharge observations with nonzero error, this was avoided to reduce the computational complexity of the method. The values of n_1 and n_2 (and ρ) are dependent upon the model simulations, but they are not calibrated directly as n_1 is the number of zero observations with zero error and n_2 is the number of zero observations with nonzero error. These values are computed directly from the model simulations generated by each calibrated parameter set. As such, these variables are functions of the calibration but not calibrated. The Box-Cox transformation parameter used in likelihoods C and D was set to a fixed value following standard convention, but could be calibrated as well. Under the multilevel mixture likelihood architecture, likelihood B will reduce to likelihood A (and likelihood D to likelihood C) as the number of zero discharge observations approaches zero.

3. Test Applications

[17] In this section, we present two case studies to explore the differences between the likelihood functions (Table 1) described in the previous section in terms of their impact on model performance and predictive uncertainty assessment. First, a synthetic example is developed to provide a controlled situation in which the likelihood functions can be compared with the true solution known a priori. A second application is provided using real streamflow data from Wanalta Creek (an ephemeral catchment) in Victoria, Australia, to analyze the likelihood functions in a more realistic setting.

3.1. Implementation of Bayesian Inference

[18] The use of Bayesian inference in hydrology typically requires the use of numerical techniques to estimate the posterior distributions due to analytical intractabilities that arise from nonlinearities common in even simple hydrologic models. Markov chain Monte Carlo (MCMC) sampling

methods numerically approximate the posterior distributions and have gained attention for their increasing use in hydrologic applications [e.g., Kuczera and Parent, 1998; Marshall et al., 2004; Smith and Marshall, 2008; Vrugi et al., 2008]. For this study, the adaptive Metropolis (AM) algorithm [Haario et al., 2001] was selected as the MCMC method for implementing the Bayesian approach. This algorithm has been shown to perform well in hydrologic problems [Marshall et al., 2004] and contains a logic that is very simple to apply. For further detail on the implementation of the AM algorithm to hydrologic modeling studies, refer to the study reported by Marshall et al. [2004].

3.2. Hydrologic Model and Forcing Data

[19] The hydrologic model selected for analysis and synthetic data generation was the simplified version of the Australian water balance model (AWBM) developed by Boughton [2004]. The model was selected for its conceptual simplicity and common usage in Australia. The AWBM (see Figure 2) requires daily rainfall and potential evapotranspiration data to produce estimates of stream discharge. The simplified version has three parameters: surface storage capacity (S), recession constant (K), and base flow index (BFI).

[20] The model forcing data used here included catchment-average daily rainfall and mean monthly areal potential evapotranspiration. Stream discharge is used to calibrate the AWBM's parameters to optimize the selected likelihood function. The catchments and data selected come from a larger collection of 331 unimpaired catchments that were part of an Australian Land and Water Resources Audit project [Peel et al., 2000].

3.3. Synthetic Case Study

[21] In order to maintain an unambiguous understanding of cause and effect, a synthetic case study was devised. A time series of synthetic discharge was created using observed rainfall and evapotranspiration data from Warrambine Creek

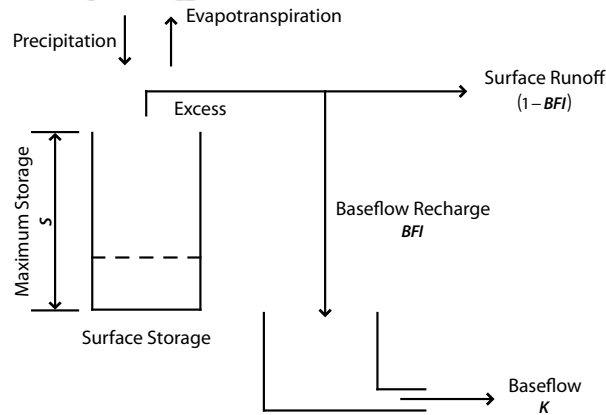


Figure 2. Simplified version of the Australian water balance model structure with parameters S , BFI , and K .

(Victoria, Australia) to force the AWBM with a set of fixed parameter values. This synthetic data series was then corrupted with noise consistent with the characteristics of the true data; the zeros in the discharge data were preserved to ensure the catchment remained ephemeral and the corrupting noise was formulated to exhibit heteroscedasticity (i.e., noise was added to Box-Cox transformed flows such that in untransformed space small discharges had less noise than large discharges). The design of the corrupting noise that was applied to the synthetic discharge followed the assumptions made in the mixture likelihood formulation presented in equation (4).

[22] In this application, for simplicity we fixed two of the three AWBM parameters (BFI and K) to their known, true values. The remaining calibration problem involved estimating the AWBM storage parameter (S) and the variance parameter (σ^2) associated with each of the likelihood functions of interest (refer to Table 1) using the AM algorithm described in section 3.1. Diffuse priors were used for both S and σ^2 to reflect the lack of prior knowledge about the parameters. The AM algorithm was implemented for each of the likelihood functions separately and run until the convergence of the posterior distributions was achieved. Convergence was determined by both visual assessment of the posterior chains and the *Gelman and Rubin* [1992] R statistic which considers convergence in terms of the variance within a single chain and the variance between multiple parallel chains.

[23] Figure 3 shows box plots derived from the posterior distributions of the AWBM storage parameter as a result of the model calibration corresponding to each of the four likelihood functions. The box indicates the interquartile range (25th–75th percentile), the central red line represents the median value, and the whiskers show the extremes of the posterior distribution. It is clear that likelihoods A and B (which assume the errors have constant variance) fail to identify the true value of the storage parameter (indicated by the green line), while likelihoods C and D (which assume the errors are heteroscedastic) are able to obtain optimal values very similar to the true value (optimal value deviates 0.48% from true value).

[24] From this case study, it is clear that likelihoods C and D provide better optimal values for the calibrated parameter and therefore better performance than was achieved with likelihoods A and B. This result is indicative of the importance of properly characterizing the error distribution in Bayesian approaches. However, while model performance is the most commonly cited factor in determining the fitness of one model versus another, predictive uncertainty is another aspect that should be addressed.

[25] A useful way of quantifying the uncertainty estimates obtained from the posterior distributions is through the reliability and sharpness summary measures [Smith and Marshall, 2010; Yadav et al., 2007]. The reliability measures the percentage of the discharge observations that are captured by the prediction interval, and the sharpness measures the mean width of the prediction interval. Ideally, the

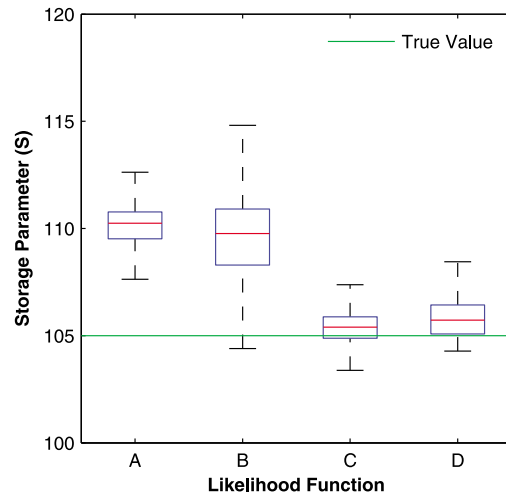


Figure 3. Box plots of storage parameter derived from calibrated posterior distributions for each likelihood function of the synthetic case study with true value shown.

Table 2. Summary of the Reliability and Sharpness Results for the Synthetic Case Study for Each of the Likelihood Functions

	Category	Likelihood			
		A	B	C	D
Reliability ^a (%)	All	96.3	98.2	89.5	96.6
	Nonzero ^b	87.7	93.9	65.3	88.9
	Over/under ^c	9.7/2.6	5.2/0.9	23.6/11.1	8.3/2.8
Sharpness (mm/day)	All	4.01	2.21	0.99	1.17
	Nonzero ^b	4.01	7.31	1.89	3.87

^aThe optimal value of reliability is equal to the value of the desired interval, 90% in this case.

^bThe nonzero category values represent the collection of data observations which are modeled with nonzero error by likelihood functions B and D (all nonzero observations).

^cThe over/under values indicate whether the observations that were not captured by the interval lie above or below the interval bounds.

reliability should be equal to the desired interval percentage (i.e., 90% of observations should be captured by a 90% interval) with the smallest possible value of the sharpness.

[26] Table 2 contains the reliability and sharpness values corresponding to each of the likelihood functions, based on a 90% prediction interval. Note that values are presented for two cases: (1) for residuals on the entire simulation (i.e., zero and nonzero states) and (2) for residuals with nonzero values only. Because likelihoods B and D are conditioned on the zero/nonzero state of the discharge, the intervals over the zero observations that were modeled with zero error are not directly interpretable (interval width equal to zero). In the cases of both likelihoods B and D, all the zero discharge observations were modeled with zero error.

[27] From Table 2, it can be seen that likelihood C has a reliability value for the entire data set of 89.52%; however, much of that is accounted for by capturing the errors on the zero streamflows and missing the peak values. Only 65.3% of the nonzero observations are captured by the interval. On the other hand, likelihood D captures 88.9% of the nonzero observations and models all zero observations with zero error. Similar results are seen in comparing likelihoods A and B, and interval accuracy tends to be worse (reliability deviates further from the target) due to the false assumption of homoscedasticity. It should be noted that an exact comparison between the reliability and sharpness between likelihoods C and D (or likelihoods A and B) is not directly applicable because they are not conditioned in the same manner (i.e., the likelihood functions use “different” data). Note that the model simulations are comparable for the different assumptions of the likelihood functions, just not the likelihood function values themselves.

[28] Figure 4 introduces quantile-quantile (QQ) plots of the residuals to provide a visual estimate of the accuracy of the distributional assumptions that underpin each of the likelihood functions tested. The use of QQ plots as an assessment tool for residual assumptions has been utilized in recent studies including those of *Thyer et al.* [2009] and *Schoups and Vrugt* [2010]. If the residuals belong to the assumed distribution, the QQ plot should be close to linear. A quantitative measure of the linearity of the QQ plot is given by the Filliben r statistic [Filliben, 1975] and is provided for each plot. The Filliben r statistic takes on values near unity as the plot becomes increasingly linear. In Figure 4, the residuals corresponding to likelihoods B and D are only those that are nonzero (based on the assumption of the

mixture likelihood) and the residuals corresponding to likelihoods C and D have been Box-Cox transformed (based on the assumption of normality only in transformed space).

[29] Comparing the QQ plots for likelihoods A and B, it is clear that there is a tangible benefit in accounting for zero inflation as measured by the increase in the Filliben statistic. Likewise, comparing the QQ plots for likelihoods A and C shows the benefit of accounting for the heteroscedastic nature of the residuals. By accounting for both zero inflation and heteroscedasticity (likelihood D), the QQ plot becomes very linear (Filliben statistic near 1) and indicates the errors come from the same distribution as assumed in the likelihood function.

[30] This synthetic test application provides a relatively objective setting for the four likelihood functions considered to be compared, where the true form of the errors and values of the AWBM parameters are known in advance. The results illustrate the potential troubles caused by selecting an incorrect likelihood function in a Bayesian setting. All points of comparison (calibration performance, uncertainty quantification, distributional correctness) identified the likelihood function (likelihood D) whose assumptions coincided with the form of the noise used to corrupt the synthetic discharge time series as the “best” likelihood, as expected. The application to real data that follows will implement the same general analysis procedure, but focuses only on likelihoods C and D given the observed heteroscedasticity in the errors.

3.4. Real Case Study: Wanalta Creek

[31] The second application focuses on the assessment of likelihood functions C and D, which were compared with regard to model performance and uncertainty quantification. As in the synthetic case study, the simplified version of the AWBM was used to generate predictions of discharge; however, all three of the parameters (S , BFI , K) along with the variance parameter associated with the likelihood functions (σ^2) were estimated with the AM algorithm. Again, diffuse priors were used on S and σ^2 , while the priors used for BFI and K were very similar to those suggested by *Bates and Campbell* [2001] for dry Australian catchments:

$$BFI \sim m\beta(\tau = 0.41, \alpha_1 = 1.50, \delta_1 = 5.19, \alpha_2 = 9.74, \delta_2 = 8.24), \quad (6)$$

$$K \sim \beta(\alpha = 150, \delta = 30), \quad (7)$$

where $m\beta(\tau, \alpha_1, \delta_1, \alpha_2, \delta_2)$ denotes a mixed beta distribution having mixing parameter τ and shape parameters (α, δ) corresponding to the first and second components of the mixture.

[32] Thirty-six years of rainfall, evapotranspiration, and discharge data were available for Wanalta Creek. The mean runoff ratio of the entire data set was 0.074 with 69.5% of the observed discharges being recorded as zeros; however, there were regular discharge events occurring throughout the record of data (see Figure 5). To initialize the conceptual storage of the AWBM structure, the first year of the record was used as a warm-up period to the calibration phase. Of the remaining 35 years, 25 years were used for calibration of the model parameters and the final 10 years of record were reserved for model validation testing.

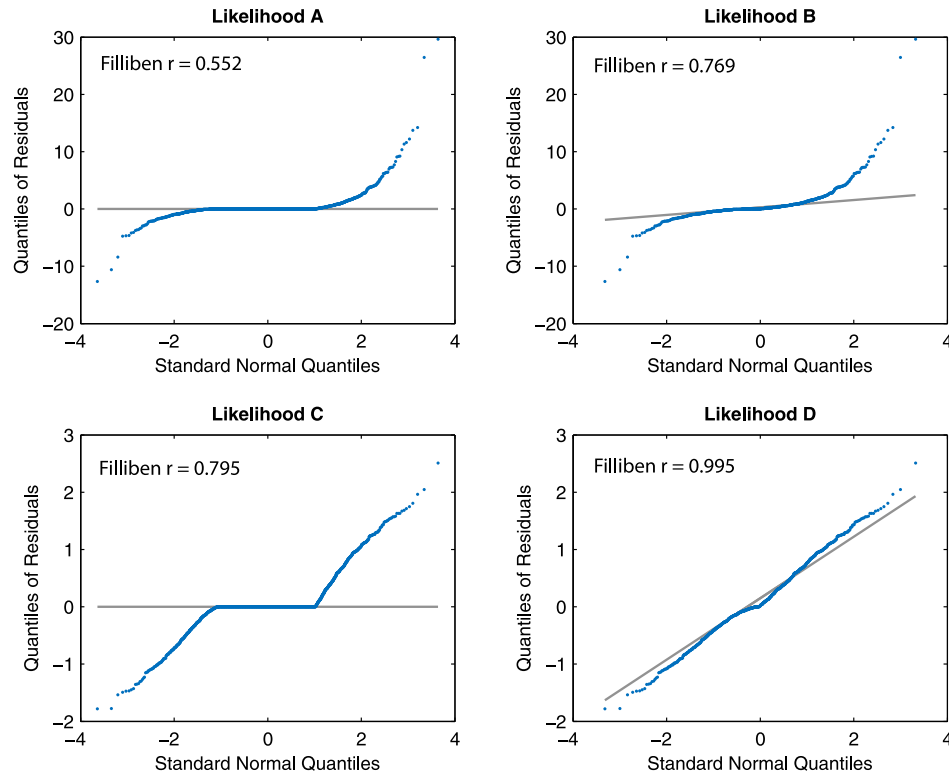


Figure 4. Quantile-quantile plots of the errors for each likelihood function of the synthetic case study. The Filliben r statistic is given on each subplot, representing the degree of similarity (coefficient of correlation) between the two quantiles. Likelihoods B and D include only the nonzero errors, as these likelihoods assume the zero errors do not belong to the same distribution as the nonzero errors.

[33] Figure 6 presents the results of the calibrated posterior distributions graphically as box plots. For the storage parameter (S), similar optimal values are obtained under both likelihood functions but the posterior distribution found with likelihood D is less peaked as seen by the wider interquartile range on the box plot. Similar results for posterior distribution width characteristics were found for the BFI parameter, where the posterior from likelihood D is less peaked than the posterior from likelihood C. For the recession constant parameter (K) it is clear that the choice of likelihood function is impacting the optimal value, where the posteriors do not intersect whatsoever. The variance parameter (VARP) shows a much smaller optimal value for likelihood C than for likelihood D due to the differences in how the data is conditioned; because likelihood C assumes that the errors corresponding to each of the observations arise from a single distribution, the estimated variance is impacted by the residuals that occur at the zero discharge observations which tend to be small.

[34] However, despite the differences in the calibrated optimal parameter sets (S , BFI, K), the overall performances associated with each of the likelihood functions remain very similar (during both calibration and validation phases) for measures emphasizing errors in timing and volume. To assess the fit of the parameter sets, the root-mean-square error (RMSE; applied to the Box-Cox transformed dis-

charges) was selected to indicate errors in timing and the deviation of runoff volume (DRV; as the ratio of the sums of predicted and observed discharges) was selected to indicate errors in total volume. Table 3 provides a summary of these results for likelihood functions C and D over the calibration and validation periods.

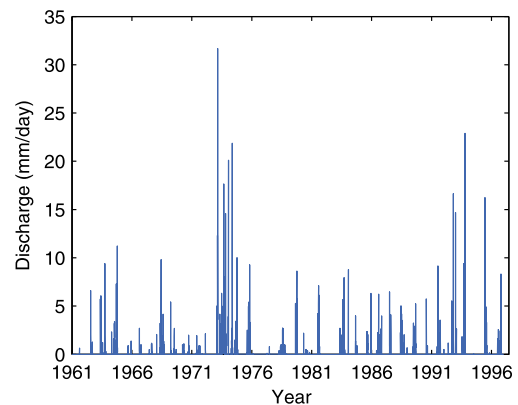


Figure 5. Observed discharge at Wanalta Creek from 1961 through 1996.

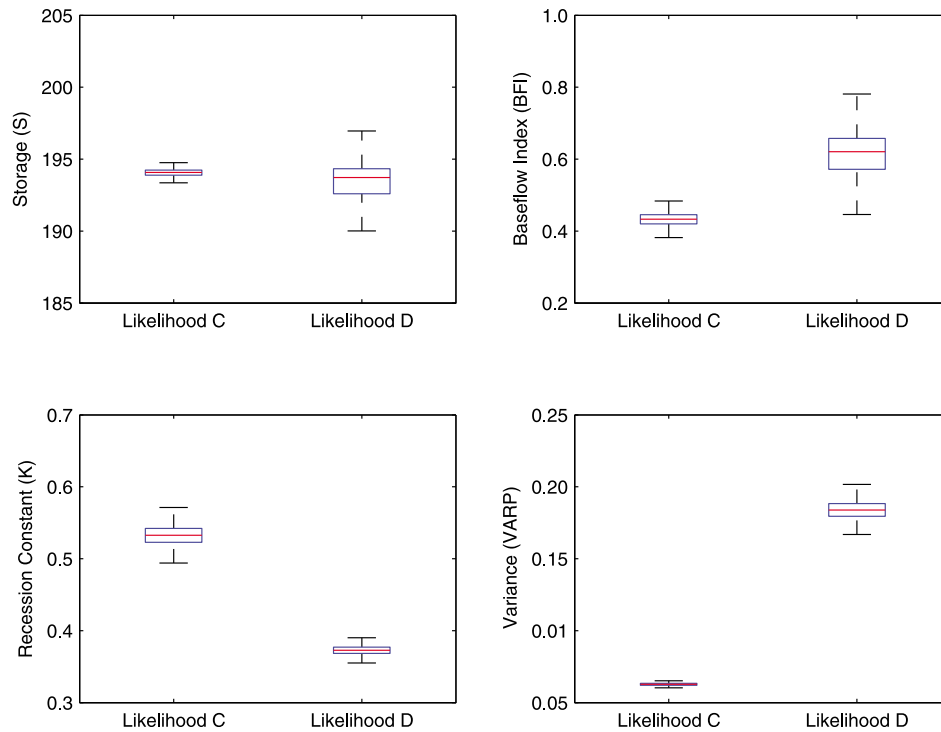


Figure 6. Box plots of each calibrated parameter derived from the posterior distributions for each likelihood function of the Wanalta Creek case study.

[35] As with the synthetic application, 90% uncertainty intervals were generated and the reliability and sharpness measures were computed. Table 3 shows these measures for the entire data set (under the heading “All”) and for the portion of the observations which were simulated with nonzero errors (as identified by likelihood D; under the heading “Nonzero”). Because likelihood D is conditioned on the data that produce nonzero errors, comparing reliability and sharpness criteria across the entire data set is problematic. To address this complication, the uncertainty-based criteria were computed for likelihood C on the “nonzero” portion of the data alone as well. Focusing on the reliability of likelihood C, it can be seen that the 90% interval actually captures 95% of the entire set of observations, indicating that the interval is too wide. When considering only the observations that likelihood D models with nonzero error, the reliability is reduced to 85.5%, highlighting the fact that the interval constructed using the posterior information from likelihood C is being decreased by the zero observations and is not capturing the nonzero observations properly. Similar patterns exist for the validation period as well, but with even greater impacts. Likelihood D, on the other hand, shows better performance in the quantification of uncertainty, capturing 91% of the nonzero observations for the calibration period and 88% for the validation period. These results indicate likelihood D’s superior ability to quantify the uncertainty associated with the nonzero discharge observations, while at the same time perfectly modeling 5992 of the 5993 (99.98%) zero dis-

charge observations during the calibration phase and all 2802 zero discharge observations during the validation phase.

[36] An examination into the appropriateness of the assumptions of the likelihood functions are plotted visually as quantile-quantile plots in Figure 7. The difficulty in modeling severely zero-inflated and, in general, low-yielding catchments is clearly seen in these plots. There is obvious improvement in the linearity (improvement of the Filliben statistic) of these plots when moving from likelihood C to

Table 3. Summary of the Reliability and Sharpness Results for the Wanalta Creek Case Study During Calibration and Validation Periods for Each of the Likelihood Functions C and D

		Calibration		Validation	
		C	D	C	D
RMSE		0.2501	0.2512	0.2484	0.2481
DRV ^a		1.0332	1.0399	1.0312	1.0381
Reliability ^b (%)	All	95.0	96.9	95.5	97.2
	Nonzero ^c	85.5	91.0	80.6	88.0
	Over/under ^d	8.7/5.8	5.3/3.7	10.7/8.7	7.6/4.4
Sharpness	All	0.54	0.48	0.54	0.40
(mm/day)	Nonzero ^c	0.77	1.38	0.94	1.71

^aDRV, deviation of runoff volume.

^bThe optimal value of reliability is equal to the value of the desired interval, 90% in this case.

^cThe nonzero category values represent the collection of data observations which are modeled with nonzero error by likelihood D.

^dThe over/under values indicate whether the observations that were not captured by the interval lie above or below the interval bounds.

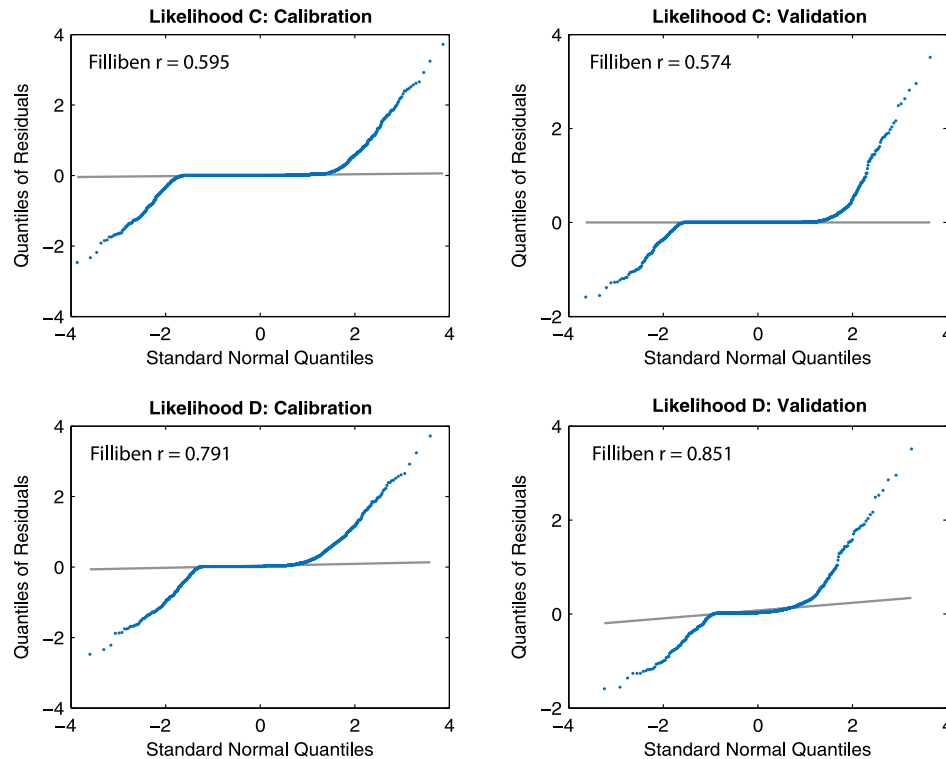


Figure 7. Quantile-quantile plots of the errors for both likelihood functions of the Wanalta Creek case study. The left panels show the results for the 25 year calibration period, and the right panels show the results for the 10 year validation period. The Filliben r statistic is given on each subplot, representing the degree of similarity between the two samples. Likelihood D includes only the nonzero errors, as these likelihoods assume the zero errors do not belong to the same distribution as the nonzero errors.

likelihood D, as a consequence of the way in which likelihood D treats zero streamflow observations. However, while likelihood D is capable of removing zero inflation from the residuals, the results are still impacted by a number of predominantly small model residuals associated with very low discharge observations. Despite this, likelihood D maintains an advantage in its ability to properly characterize the uncertainty of the predictions while not suffering any noticeable negative impacts on its predictive performance.

[37] This test application to the discharge data recorded at Wanalta Creek in Victoria, Australia, has illustrated some of the difficulties that come about when modeling low-yield catchments. When considering the two alternative likelihood functions (C and D), it was found that the posterior distributions estimated by the AM algorithm were impacted. Although the posterior distributions were different (and for the recession constant parameter nonintersecting), similar performances were obtained in both timing (RMSE) and volume (DRV) measures (consistent with the equifinality principle). Nonetheless, the posterior distributions form the basis of uncertainty analysis procedures in Bayesian statistics and the differences in these posteriors translate to the observed differences in uncertainty quantification (as measured by the reliability and sharpness metrics). A better fit to the distributional assumptions of the likelihood functions was found for likelihood D as well, but the QQ plot revealed

that further refinement of the likelihood function's form should be considered.

4. Discussions and Conclusions

[38] The increasing interest in uncertainty analysis that has gained momentum in hydrologic modeling studies recently has led to prevalent use of Bayesian methods. Such approaches, however, require the modeler to make formal assumptions about the form of the modeled errors via the likelihood function. Critics of the Bayesian approach point to these assumptions as the main weakness behind such methods and advocate for less statistically formal methods to avoid false enumeration of the associated uncertainty.

[39] The importance of selecting an appropriate likelihood function was investigated through two applications: (1) a synthetically produced time series of streamflow that was corrupted with zero-inflated, heteroscedastic noise and (2) a real time series of streamflow from Wanalta Creek in Victoria, Australia. In both applications the simplified, three-parameter Australian water balance model was selected as the hydrologic model and parameter estimation and uncertainty analysis was carried out by the adaptive Metropolis algorithm under a Bayesian framework.

[40] In the synthetic case study we examined four unique likelihood functions, each of which made different assump-

tions about the form of the errors. Of the four likelihoods, likelihood D was the true error model which the form of the errors used to corrupt the synthetic discharge record was designed to match. From our analyses it was apparent that the selection of a misinformed likelihood function (such as likelihood A) can impact the ability of the AM algorithm to identify the true parameter values (Figure 3) and properly quantify the associated uncertainty (Table 2 and Figure 4). Likelihood D, which assumed the errors to be heteroscedastic and zero inflated, was best able to achieve uncertainty intervals consistent with expectations (i.e., reliability close to the interval value of 90% and the smallest possible value of sharpness).

[41] In the Wanalta Creek case study we considered only two (C and D) of the four likelihood functions tested in the synthetic application based on the heteroscedastic nature of the residuals. Again, we considered both of the likelihood functions in terms of their ability to fit the observed data, properly quantify the uncertainty, and correctly match the assumed form of the errors. On the basis of the composite results of these criteria, likelihood D was again shown to be preferred due to its ability to account for the zero inflation of the residuals. Although each likelihood function's optimal parameter set resulted in similar fits to the observed streamflow during both calibration and validation phases (Table 3), the calibrated posterior distributions deviated significantly from one another (Figure 6). The differences in posterior distributions resulted in distinct characterization of the uncertainty intervals for each of the likelihood functions, as quantified by the reliability and sharpness metrics, which indicated more appropriate assignment of uncertainty by likelihood D (Table 3). The potential hazards of using an inappropriate likelihood function were also illustrated by the QQ plots (Figure 7), which showed a more linear relationship for likelihood D than for likelihood C but also indicated that likelihood D still has much room for potential improvement.

[42] Although the zero-inflation adaptation was shown to effectively improve the simulations for the synthetic case study (as quantified by the QQ plots, uncertainty metrics, and identification of true parameter values), the S-shaped nature of the QQ plots (Figure 7) for the real case study indicate that other factors are complicating the error structure (although the uncertainty metrics indicate decent performance). Among these are the potential for autocorrelated residuals, better transformations of the data to remove heteroscedasticity (potentially by further separation of the nonzero discharge likelihood function component into low flow and high flow states), input data uncertainty, and model structural uncertainty.

[43] While an investigation into each of these potential complicating factors is well outside the scope of this study's objective, many of these areas are topics of ongoing research into a generic likelihood selection framework. In the specific areas of input data uncertainty and model structural uncertainty, a wealth of recent research from the University of Newcastle research group [e.g., Kavetski *et al.*, 2006; Renard *et al.*, 2010; Thyer *et al.*, 2009] has indicated the importance of addressing these components for improved quantification of total uncertainty. Further, Schoups and Vrugt [2010] have recently introduced a paper on the development of a flexible formal likelihood function that aims to address many of the issues caused by heteroscedasticity and auto-

correlation. Although the results of the study offer an intriguing and valuable tool for future work, their flexible likelihood function still suffered many of the issues outlined here. In particular, their method was found to work well for a wet catchment, but experienced a breakdown in performance for a semiarid basin (nonephemeral, continuous streamflow).

[44] The test applications detailed in this study highlight the importance of checking the assumptions made in the likelihood function, while also demonstrating the potential consequences of not. Modeling studies in hydrology have been increasingly focusing on incorporating uncertainty assessment into the presented results, yet too few studies perform simple checks on the uncertainty estimates themselves. In this paper, we have shown that false assumptions in the likelihood function might not result in noticeable differences in the ability of the optimum parameter set to the fit the observed data. This is not overly surprising given the attention and support that has been given to the equifinality thesis [Beven, 2006b]. However, the unique benefit of the Bayesian approach is that it results in a proper probability distribution for each of the calibrated parameters (the posterior distribution). It is in the posterior distributions where the false assumptions of the likelihood function persist and can be propagated into false estimates of the associated uncertainty. In both the synthetic and real data case studies, likelihood D was found to perform well in terms of the uncertainty metrics (reliability and sharpness). This result indicates the importance of a multifaceted assessment strategy that considers multiple aspects of "performance". Despite this we also acknowledge that further refinement is necessary to improve the linearity of the QQ plots.

[45] The formal likelihood function introduced here sought to address the problem of severely zero-inflated residuals. The problem of zero inflation has not been addressed in previous hydrologic modeling studies performed under a Bayesian inferential approach. The results indicate improved composite performance (fit, uncertainty, distributional correctness) in both synthetic and real data settings by explicitly accounting for this zero inflation through the application of the multilevel mixture logic summarized in Figure 1 and equation (4) when compared to the more traditionally applied likelihood functions. This type of conditional mixing logic is advantageous due to its computational simplicity (same number of parameters as the simpler approaches) and its extendibility to other settings. In fact as the catchment moves away from zero inflation, likelihoods C and D become more and more similar and eventually (when there are no zero discharge observations) become identical to one another. Future work is currently under way to extend and incorporate the likelihood function adaptation implemented here (designed to accommodate zero-inflated errors) with other previously used adaptations (autocorrelation, heteroscedasticity, distinct discharge state mixtures, etc.) into a generic framework that assists modelers in selecting an appropriate likelihood function based on the given data-model coupling to be used in Bayesian studies; this work aims to fill a void in the advancement of such methods in hydrology.

[46] **Acknowledgments.** This work was made possible by funding from the Australian Research Council. We thank Francis Chiew for providing the data used in this study.

References

- Aitchison, J. (1955), On the distribution of a positive random variable having a discrete probability mass at the origin, *J. Am. Stat. Assoc.*, 50(271), 901–908.
- Ajami, N. K., Q. Duan, and S. Sorooshian (2007), An integrated hydrologic Bayesian multimodel combination framework: Confronting input, parameter, and model structural uncertainty in hydrologic prediction, *Water Resour. Res.*, 43, W01403, doi:10.1029/2005WR004745.
- Bates, B. C., and E. P. Campbell (2001), A Markov chain Monte Carlo scheme for parameter estimation and inference in conceptual rainfall-runoff modeling, *Water Resour. Res.*, 37(4), 937–947, doi:10.1029/2000WR900363.
- Beven, K. (2006a), On undermining the science? *Hydrol. Processes*, 20(14), 3141–3146.
- Beven, K. (2006b), A manifesto for the equifinality thesis, *J. Hydrol.*, 320(1–2), 18–36.
- Beven, K., and A. Binley (1992), The future of distributed models: Model calibration and uncertainty prediction, *Hydrol. Processes*, 6(3), 279–298.
- Beven, K., P. Smith, and J. Freer (2007), Comment on “Hydrological forecasting uncertainty assessment: Incoherence of the GLUE methodology” by Pietro Mantovan and Ezio Todini, *J. Hydrol.*, 338(3–4), 315–318.
- Boughton, W. (2004), The Australian water balance model, *Environ. Modell. Software*, 19(10), 943–956.
- Box, G. E. P., and D. R. Cox (1964), An analysis of transformations, *J. R. Stat. Soc., Ser. B*, 26(2), 211–252.
- Campbell, E. P., D. R. Fox, and B. C. Bates (1999), A Bayesian approach to parameter estimation and pooling in nonlinear flood event models, *Water Resour. Res.*, 35(1), 211–220, doi:10.1029/1998WR900043.
- Duan, Q., S. Sorooshian, and V. Gupta (1992), Effective and efficient global optimization for conceptual rainfall-runoff models, *Water Resour. Res.*, 28(4), 1015–1031, doi:10.1029/91WR02985.
- Duan, Q., N. K. Ajami, X. Gao, and S. Sorooshian (2007), Multi-model ensemble hydrologic prediction using Bayesian model averaging, *Adv. Water Resour.*, 30(5), 1371–1386.
- Filliben, J. J. (1975), The probability plot correlation coefficient test for normality, *Technometrics*, 17(1), 111–117.
- Gelman, A., and D. B. Rubin (1992), Inference from iterative simulation using multiple sequences, *Stat. Sci.*, 7(4), 457–472.
- Haario, H., E. Saksman, and J. Tamminen (2001), An adaptive Metropolis algorithm, *Bernoulli*, 7(2), 223–242.
- Hsu, K.-I., H. Moradkhani, and S. Sorooshian (2009), A sequential Bayesian approach for hydrologic model selection and prediction, *Water Resour. Res.*, 45, W00B12, doi:10.1029/2008WR006824.
- Kavetski, D., G. Kuczera, and S. W. Franks (2006), Bayesian analysis of input uncertainty in hydrological modeling: Theory, *Water Resour. Res.*, 42, W03407, doi:10.1029/2005WR004368.
- Kuczera, G. (1983), Improved parameter inference in catchment models. 1. Evaluating parameter uncertainty, *Water Resour. Res.*, 19(5), 1151–1162, doi:10.1029/WR019i005p01151.
- Kuczera, G., and E. Parent (1998), Monte Carlo assessment of parameter uncertainty in conceptual catchment models: The Metropolis algorithm, *J. Hydrol.*, 211(1–4), 69–85.
- Madsen, H., G. Wilson, and H. C. Ammentorp (2002), Comparison of different automated strategies for calibration of rainfall-runoff models, *J. Hydrol.*, 261(1–4), 48–59.
- Mantovan, P., and E. Todini (2006), Hydrological forecasting uncertainty assessment: Incoherence of the GLUE methodology, *J. Hydrol.*, 330(1–2), 368–381.
- Mantovan, P., E. Todini, and M. L. V. Martina (2007), Reply to comment by Keith Beven, Paul Smith, and Jim Freer on “Hydrological forecasting uncertainty assessment: Incoherence of the GLUE methodology,” *J. Hydrol.*, 338(3–4), 319–324.
- Marshall, L., D. Nott, and A. Sharma (2004), A comparative study of Markov chain Monte Carlo methods for conceptual rainfall-runoff modeling, *Water Resour. Res.*, 40, W02501, doi:10.1029/2003WR002378.
- Marshall, L., A. Sharma, and D. Nott (2006), Modeling the catchment via mixtures: Issues of model specification and validation, *Water Resour. Res.*, 42, W11409, doi:10.1029/2005WR004613.
- Peel, M. C., F. H. S. Chiew, A. W. Western, and T. A. McMahon (2000), Extension of unimpaired monthly streamflow data and regionalisation of parameter values to estimate streamflow in ungauged catchments, in *National Land and Water Resources Audit*, CRC for Catchment Hydrology, Melbourne.
- Renard, B., D. Kavetski, G. Kuczera, M. Thyer, and S. W. Franks (2010), Understanding predictive uncertainty in hydrologic modeling: The challenge of identifying input and structural errors, *Water Resour. Res.*, 46, W05521, doi:10.1029/2009WR008328.
- Robert, C. P. (1996), Mixtures of distributions: inference and estimation, in *Markov Chain Monte Carlo in Practice*, edited by W. R. Gilks, et al., pp. 441–464, Chapman and Hall, London.
- Samanta, S., D. S. Mackay, M. K. Clayton, E. L. Kruger, and B. E. Ewers (2007), Bayesian analysis for uncertainty estimation of a canopy transpiration model, *Water Resour. Res.*, 43, W04424, doi:10.1029/2006WR005028.
- Samanta, S., M. K. Clayton, D. S. Mackay, E. L. Kruger, and B. E. Ewers (2008), Quantitative comparison of canopy conductance models using a Bayesian approach, *Water Resour. Res.*, 44, W09431, doi:10.1029/2007WR006761.
- Schaefli, B., D. B. Talamba, and A. Musy (2007), Quantifying hydrological modeling errors through a mixture of normal distributions, *J. Hydrol.*, 332(3–4), 303–315.
- Schoups, G., and J. A. Vrugt (2010), A formal likelihood function for parameter and predictive inference of hydrologic models with correlated, heteroscedastic and non-Gaussian errors, *Water Resour. Res.*, 46, W10531, doi:10.1029/2009WR008933.
- Smith, T. J., and L. A. Marshall (2008), Bayesian methods in hydrology: A study of recent advancements in Markov chain Monte Carlo techniques, *Water Resour. Res.*, 44, W00B05, doi:10.1029/2007WR006705.
- Smith, T. J., and L. A. Marshall (2010), Exploring uncertainty and model predictive performance concepts via a modular snowmelt-runoff modeling framework, *Environ. Modell. Software*, 25(6), 691–701.
- Sorooshian, S., and J. A. Dracup (1980), Stochastic parameter estimation procedures for hydrologic rainfall-runoff models: Correlated and heteroscedastic error cases, *Water Resour. Res.*, 16(2), 430–442, doi:10.1029/WR016i002p00430.
- Stedinger, J. R., R. M. Vogel, S. U. Lee, and R. Batchelder (2008), Appraisal of the generalized likelihood uncertainty estimation (GLUE) method, *Water Resour. Res.*, 44, W00B06, doi:10.1029/2008WR006822.
- Thyer, M., B. Renard, D. Kavetski, G. Kuczera, S. W. Franks, and S. Srikanthan (2009), Critical evaluation of parameter consistency and predictive uncertainty in hydrological modeling: A case study using Bayesian total error analysis, *Water Resour. Res.*, 45, W00B14, doi:10.1029/2008WR006825.
- Tu, W. (2002), Zero-inflated data, in *Encyclopedia of Environmetrics*, edited by A. H. El-Shaarawi and W. W. Piegorsch, pp. 2387–2391, John Wiley, Chichester, U.K.
- Tu, W., and X.-H. Zhou (1999), A Wald test comparing medical costs based on log-normal distributions with zero valued costs, *Stat. Med.*, 18(20), 2749–2761.
- Vrugt, J., C. ter Braak, H. Gupta, and B. Robinson (2009), Equifinality of formal (DREAM) and informal (GLUE) Bayesian approaches in hydrologic modeling?, *Stochastic Environ. Res. Risk Assess.*, 23(7), 1011–1026.
- Vrugt, J. A., H. V. Gupta, W. Bouten, and S. Sorooshian (2003), A shuffled complex evolution Metropolis algorithm for optimization and uncertainty assessment of hydrologic model parameters, *Water Resour. Res.*, 39(8), 1201, doi:10.1029/2002WR001642.
- Vrugt, J. A., H. V. Gupta, S. C. Dekker, S. Sorooshian, T. Wagener, and W. Bouten (2006), Application of stochastic parameter optimization to the Sacramento soil moisture accounting model, *J. Hydrol.*, 325(1–4), 288–307.
- Vrugt, J. A., C. J. F. ter Braak, M. P. Clark, J. M. Hyman, and B. A. Robinson (2008), Treatment of input uncertainty in hydrologic modeling: Doing hydrology backwards with Markov chain Monte Carlo simulation, *Water Resour. Res.*, 44, W00B09, doi:10.1029/2007WR006720.
- Wagener, T., and N. McIntyre (2005), Identification of rainfall-runoff models for operational applications, *Hydrol. Sci. J.*, 50(5), 735–751.
- Xu, C.-Y. (2001), Statistical analysis of parameters and residuals of a conceptual water balance model: Methodology and case study, *Water Resour. Manage.*, 15(2), 75–92.
- Yadav, M., T. Wagener, and H. Gupta (2007), Regionalization of constraints on expected watershed response behavior for improved predictions in ungauged basins, *Adv. Water Resour.*, 30(8), 1756–1774.
- Zhou, X.-H., and W. Tu (1999), Comparison of several independent population means when their samples contain log-normal and possibly zero observations, *Biometrics*, 55(2), 645–651.
- Zhou, X.-H., and W. Tu (2000), Confidence intervals for the mean of diagnostic test charge data containing zeros, *Biometrics*, 56(4), 1118–1125.

L. Marshall and T. Smith, Department of Land Resources and Environmental Sciences, Montana State University, Bozeman, MT 59717, USA.

R. Mehrotra and A. Sharma, School of Civil and Environmental Engineering, University of New South Wales, Sydney, NSW 2052, Australia. (a.sharma@unsw.edu.au)

S. Sisson, School of Mathematics and Statistics, University of New South Wales, Sydney, NSW 2052, Australia.

CHAPTER THREE

A BAYESIAN LIKELIHOOD FUNCTION SPECIFICATION METHODOLOGY
FOR CONCEPTUAL HYDROLOGIC MODELINGContribution of Author and Co-Authors

Manuscript in Chapter 3

Author: Tyler Smith

Contributions: Responsible for the development and analysis of technique introduced, the majority of the writing, and figure development.

Co-Author: Lucy Marshall

Contributions: Provided extensive discussions of the development of the methods used in the study and contributed to the editing of the manuscript throughout its development.

Co-Author: Ashish Sharma

Contributions: Provided extensive discussions of the development of the methods used in the study and contributed to the editing of the manuscript throughout its development.

Manuscript Information

Smith, T., L. Marshall, and A. Sharma. A Bayesian likelihood function specification methodology for conceptual hydrologic modeling, *Water Resources Research*, submitted.

- Journal: Water Resources Research
- Status of manuscript (check one)
 - ☐ Prepared for submission to a peer-reviewed journal
 - ☒ Officially submitted to a peer-reviewed journal
 - ☐ Accepted by a peer-reviewed journal
 - ☐ Published in a peer-reviewed journal
- Publisher: American Geophysical Union
- Date of submission: February 9, 2012

A BAYESIAN LIKELIHOOD SPECIFICATION METHODOLOGY FOR CONCEPTUAL HYDROLOGIC MODELING

Abstract

Hydrologic modelers are confronted with the challenge of producing estimates of the uncertainty associated with model predictions across an array of catchments and hydrologic flow regimes. In order to obtain these uncertainty estimates, a numerical method must be selected for hydrologic model parameter estimation. Formal Bayesian approaches are often criticized for making strong assumptions about the nature of model residuals via the likelihood function that are not well satisfied, leading to inappropriate uncertainty estimates. This paper outlines a likelihood function specification framework that aims to address these issues through the development of a simple and straightforward procedure for building appropriate likelihood functions that is both flexible and extendible. The design addresses the problems brought about by variability in model residuals as a function of hydrologic regime, including specific acknowledgement of the difficulties posed by continuous modeling of ephemeral catchments (e.g., zero-inflated residuals). In doing so, many of the likelihood function adaptations that have previously been used in application-specific settings were assembled within the selection framework. The framework was applied to four synthetic datasets of differing error structure such that each of the likelihood functions was investigated and assessed based on its ability to properly characterize the errors. Further, the framework was applied to a real dataset, the Black River catchment in Queensland, Australia. The results of these test applications indicated that a multifaceted assessment strategy is necessary to determine the adequacy of an individual likelihood function, and that under such a strategy the most appropriate likelihood function considered is identifiable.

Introduction

The use of Bayesian inferential approaches are predicated on the formal likelihood function adequately characterizing the form of the true model errors [Mantovan and Todini, 2006]. In hydrology this can be a difficult task, and has led to criticism of the appropriateness of formal Bayesian methods in “real data” cases [Beven *et al.*, 2007]. Regardless of this, Bayesian methods have become a common part of the hydrologic modeling literature [e.g., Kuczera and Parent, 1998; Marshall *et al.*, 2004;

Smith and Marshall, 2008]. *Box and Tiao [1973]* provide some insight into the development and application of formal likelihood functions within a Bayesian inferential approach, highlighting the importance of iterative improvement, stating “to be on firm ground we must do more than merely postulate a model; we must build and test a tentative model at each stage of investigation” (p. 7). However, general guidance on how to appropriately select a formal likelihood function to be used in hydrologic modeling studies has yet to be formally addressed beyond application-specific problems. Even application-specific examples of appropriate likelihood function selection are relatively rare [e.g., *Bates and Campbell, 2001; Kuczera, 1983; Schaefli et al., 2007; Schoups and Vrugt, 2010; Smith et al., 2010; Sorooshian and Dracup, 1980*]; especially when compared to the widespread usage of the basic assumption of normality [e.g., *Ajami et al., 2007; Campbell et al., 1999; Duan et al., 2007; Hsu et al., 2009; Samanta et al., 2008; Vrugt et al., 2006*].

Xu [2001, p. 77] points out that “in the field of hydrological modeling, few writers examine and describe any properties of residuals given by their models when fitted to the data”; however, recently there have been a few notable exceptions to this [e.g., *Reichert and Mieleitner, 2009; Schoups and Vrugt, 2010; Smith et al., 2010; Thyer et al., 2009*]. This frequent lack of concern for finding an appropriate likelihood function provides a potential limiting constraint on more appropriate and widespread use of Bayesian inference. The need to address the lack of normality in hydrologic modeling errors is not new, but there is little advice on exactly how to go about selecting an appropriate likelihood for a particular data-model combination.

The intention of this paper is to introduce a broad framework to guide the selection of a formal likelihood function to be used within a Bayesian context for modeling residual errors in a hydrologic context. The significance and contribution of this work does not necessarily lie in the idea of adding complexity to formal likelihood functions (when warranted), but rather in a formalization and illustration of this process for use in conceptual hydrological modeling. Further, the likelihood functions employed here are in no way meant to represent all possibilities or even necessarily advocated as the best formulations. Rather, they represent many of the most frequently used forms and have been used here to illustrate the framework we are proposing. The framework presented herein is tested with regard to rainfall-runoff modeling scenarios, but is independent of the specific problem setting. Within this framework, a series of decisions are used to guide the modeler to an appropriate likelihood function based on a number of tests and checks on the residuals. By identifying the most appropriate likelihood function, the estimates of parameter uncertainty that are obtained in the calibration process are improved. Obtaining the most precise estimates of uncertainty has great value and is especially important to regionalization methods used to transfer knowledge from a gauged catchment to an ungauged catchment [Zhang *et al.*, 2008].

This paper is divided into the following sections: section 2 introduces the basic framework devised for proper selection of a formal likelihood function, section 3 offers two test applications on which the method was utilized, section 4 presents a discussion of the results, and section 5 highlights the key conclusions of the study.

Selection of a Formal Likelihood Function

Conceptual hydrologic models can be thought of, in a generic sense, as being of the form

$$Q_t = f(x_t; \theta) + \varepsilon_t \quad (1)$$

where t indexes time, Q_t is the observed discharge, $f(x_t; \theta)$ is the model simulated discharge, x_t is the model forcing data (typically, rainfall and evapotranspiration), θ is the set of unknown model parameters, and ε_t is the error. Proper understanding of the form of the errors (ε_t) is important for using the model for predictions and vital to the success of Bayesian inference. In statistical inference the errors are modeled via the formal likelihood function.

Debate over the selection and appropriate use of formal likelihood functions has dominated recent hydrologic modeling literature [e.g., *Beven et al.*, 2007; *Beven et al.*, 2008; *Mantovan and Todini*, 2006; *Mantovan et al.*, 2007; *Stedinger et al.*, 2008]. Given the debate over formal likelihood functions, a set of general guidelines illustrating some simple tests of the likelihood function's appropriateness are needed to improve the application of Bayesian methods to hydrologic modeling problems and address the concerns raised by opponents of the formal Bayesian approach.

Our likelihood selection framework, introduced in Figure 1, consolidates many of the previously developed approaches that aimed to produce more appropriate formal likelihood functions; innovations that account for error characteristics such as autocorrelation [e.g., *Sorooshian and Dracup*, 1980], non-constant variance [e.g., *Bates and Campbell*, 2001], and zero-inflation caused by extended periods of no observed

streamflow [Smith *et al.*, 2010]. To address heteroscedasticity, a Box-Cox transformation is applied (in two possible forms)

$$y^{\text{BC}} = \frac{(y + \lambda_2)^{\lambda_1} - 1}{\lambda_1} \quad (2a)$$

$$y^{\text{BC}} = \log(y + \lambda_2) \quad (2b)$$

where y^{BC} is the transformed data, y is the untransformed data, and λ_1 and λ_2 are the Box-Cox parameters with the restraint that $y > -\lambda_2$. Autocorrelation is accounted for with an autoregressive model

$$y_t^{\text{AR}} = \sum_{j=1}^p \phi_j y_{t-j} \quad (3)$$

where y^{AR} is the data with time index t , p is the order of the autoregressive process, and ϕ are the autoregression coefficients. Zero-inflation is dealt with using the mixture likelihood approach of Smith *et al.* [2010] and is given in Table 1. While these likelihood functions are not inclusive of every possible error model, they do broadly capture the types of model errors that typically occur for hydrologic models. The equations describing each of the mentioned likelihood functions are presented in Table 1.

Mirroring a top-down model development approach [Sivapalan *et al.*, 2003], the likelihood selection framework presented herein advocates starting with a simple likelihood function (such as a likelihood based on the assumption of independent errors distributed normally with zero mean and constant variance) and only adding complexity when merited by analysis of the resultant residuals [refer to Box and Tiao, 1973]. This type of approach is advantageous in that it forces the modeler to actively investigate the error structure, as well as limiting the computational complexity of the method. The

alternative bottom-up approach would start with a very complex likelihood and allow the algorithm to choose which components are necessary (under a nested likelihood model frame). Although both methods have the potential for providing the same result, the bottom-up approach may provide computational challenges, as the MCMC algorithm must calibrate a higher dimensional model. For example, the bottom-up approach might potentially result in autoregressive parameter values very close to zero, suggesting the errors are not autocorrelated; however, the MCMC algorithm still has to calibrate these parameters. We believe that a top-down approach provides good scope for model diagnostics as it enables successive assessment of the most evident error characteristics. Note that the computational advantages of a top-down approach are expected to vary with the specific problem; i.e., very complex problems may require several iterations in the pursuit of an appropriate likelihood function leading to increased computational demand.

The formal likelihood functions contained within the overall framework were, however, selected in a nested manner to allow the most complex likelihood to be reduced to less complex likelihood functions given certain conditions. In this way, the eight formal likelihood functions (in Table 1) can be reduced to one when using the two-parameter Box-Cox transformation (with λ_1 and λ_2 equal to one) or to two when using the one-parameter Box-Cox transformation (as this form of the transformation is irreducible). Although more common in a bottom-up approach, nested likelihood functions reduce the complexity required to implement the advocated top-down approach, allowing the user to “turn on” additional components where necessary.

To determine the appropriateness of the additional components, several statistical tests and diagnostic plots are available to guide the modeler. Zero-inflation testing can be performed directly from the observed data, as the observed discharge data controls the percentage of zero residuals that are possible; only zero discharges have a realistic probability of being represented exactly within the hydrologic model due to double precision accuracy of the computer running the model code. In this study, a general guideline of 10% zero discharge observations (based on previous experience with modeling of ephemeral catchments in Australia; see *Smith et al.* [2010]) was adopted as a threshold value in determining whether or not zero-inflation is a potential problem. The 10% guideline can be used as an initial cursory check, prior to any modeling, to indicate whether zero-inflation is a likely issue; however, once model simulations have been performed zero-inflation of the residuals should be checked as indicated in Figure 1. To investigate the heteroscedasticity in the residuals, the *Filliben* [1975] r statistic is used. The Filliben statistic quantifies the linearity of quantile-quantile (QQ) plots; as the quantile-quantile plots become more linear, the Filliben statistic approaches unity. Further, the Filliben statistic can be extended to cases where the base assumption is not a Gaussian distribution (e.g., Student's t -distribution), a benefit for further expansion of the framework. The use of QQ plots as an assessment tool for residual assumptions has been utilized in recent studies including those of *Thyer et al.* [2009] and *Schoups and Vrugt* [2010]. Scatter plots of the residuals are also used as a supplementary visual check for heteroscedasticity. Autocorrelation is assessed using the test statistic of *Ljung and Box* [1978], as well as visual assessment of correlograms calculated at a specified number of

time lags. The Ljung-Box test statistic tests the “overall” randomness of the data based on a number of time lags rather than at each distinct lag.

Following the logic of Figure 1, the initial modeling attempt is performed, applying the simplest likelihood function (here based on the assumption of independent, homoscedastic, Gaussian errors). The residuals are then analyzed for zero-inflation, based on the 10% rule. A quantile-quantile plot of the residuals will show zero-inflation graphically with a strong horizontal “step”. If the residuals are zero-inflated, this component is added into the likelihood function and the parameters are re-calibrated. Otherwise, the modeler moves to the next step. The third step is a check for residual heteroscedasticity. Again, the use of scatter plots of the errors and the Filliben statistic will be invoked as basic checks. As before, the heteroscedasticity component is added to the likelihood if the condition is present by applying a Box-Cox transformation. The final level of the logic is to check the errors for autocorrelation using correlograms and the Ljung-Box statistic to assess the overall randomness of the data. We believe that the iterative nature of this framework allows for successive assessment of common hydrologic model error characteristics. We recognize that in most cases hydrologic model errors will take complex forms, exhibiting different degrees of heteroscedasticity, autocorrelation, and non-stationarity. The procedure outlined in Figure 1 advocates for testing of non-stationarities (specifically zero-inflation) as a first step as it may mask other error behavior, such as normality. Tests for heteroscedasticity are then applied, as error transformations can illuminate potential autocorrelations in model residuals. It should be noted that at the inclusion of each additional component, the model parameters

must be re-calibrated and it is suggested that the modeler repeat the checks on all components to ensure that the added component actually addresses the problem originally diagnosed in the model errors.

Case Studies

In this section, we apply the formal likelihood function selection framework to both synthetic and real data case studies. In the synthetic study, we consider the four most complex forms (likelihood functions E, F, G, and H) of the eight potential likelihood functions to generate synthetic datasets with corresponding error structures. In each synthetic data case, all eight likelihood function forms are tested and compared via a multifaceted model assessment exercise that examines: (1) the ability to capture the true model parameter values used to generate the data, (2) the normality of the resulting residuals, (3) the accurate modeling of “zero discharge” observations, (4) the correlation/randomness of the residuals, and (5) the appropriate representation of the model predictive uncertainty (model "reliability"). For the real-world example, the likelihood function selection framework will be used to find the most appropriate likelihood function of the eight available in the framework. The same multifaceted assessment strategy as was prescribed for the synthetic study will be applied here, with the exception of the first test as the true parameter values are unknown for the real data study.

Hydrologic Model

The hydrologic model selected for analysis and synthetic data generation was a simplified version of the Australian Water Balance Model (AWBM) developed by *Boughton* [2004]. The model was selected for its conceptual simplicity and extensive past use. The AWBM (see Figure 2) requires daily rainfall and potential evapotranspiration data to produce estimates of stream discharge. The simplified version has three parameters: surface storage capacity (S), recession constant (K), and baseflow index (BFI). The model forcing data used here included catchment-average daily rainfall and mean monthly areal potential evapotranspiration. Stream discharge is used to calibrate the AWBM's parameters to optimize the selected likelihood function. The catchments and data selected come from a larger collection of 331 unimpaired catchments that were part of an Australian Land and Water Resources Audit project [*Peel et al.*, 2000].

Parameter Estimation

Due to the complex nature of hydrologic modeling, the use of Bayesian inference is dependent upon numerical techniques to estimate the parameter posterior distributions. The most common method of doing this in hydrology is to use Markov chain Monte Carlo (MCMC) sampling [e.g., *Kuczera and Parent*, 1998; *Marshall et al.*, 2004; *Smith and Marshall*, 2008; *Vrugt et al.*, 2008]. For this study, the Adaptive Metropolis (AM) algorithm [*Haario et al.*, 2001] was selected as the MCMC method for implementing the Bayesian approach. This algorithm has been shown to perform well in hydrologic problems [*Marshall et al.*, 2004] and contains a logic that is very simple to apply. For

further detail on the implementation of the AM algorithm to hydrologic modeling studies, readers are requested to refer to *Marshall et al.* [2004].

In this study, we calibrated each of the AWBM parameters (S, K, BFI), as well as the likelihood parameters including the variance parameter (σ^2) and autoregressive component parameter(s) (ϕ , when applicable) (refer to Figure 1, Table 1) using the AM algorithm. The Box-Cox transformation (Eq. 2b) parameter used to account for heteroscedasticity (likelihoods C, E, G, and H) was set to a fixed value ($\lambda_2 = 0.5$) following the recommendation of *Bates and Campbell* [2001] for medium-sized, ephemeral Australian catchments, but could be calibrated as well. In our experience, numerical complications do arise for zero-inflated catchments when trying to calibrate the Box-Cox transformation parameter (i.e., the likelihood approaches infinity for very small values of the transformation under zero-inflation conditions). Diffuse priors were used for S, σ^2 , and ϕ to reflect the lack of prior knowledge about the parameters. Based on the suggested priors for the AWBM model when applied to dry Australian catchments [*Bates and Campbell*, 2001], the priors used for BFI and K were:

$$BFI \sim m\beta(\tau = 0.41, \alpha_1 = 1.50, \delta_1 = 5.19, \alpha_2 = 9.74, \delta_2 = 8.24) \quad (4)$$

$$K \sim \beta(\alpha = 150, \delta = 30) \quad (5)$$

where $m\beta(\tau, \alpha_1, \delta_1, \alpha_2, \delta_2)$ denotes a mixed beta distribution having mixing parameter τ and shape parameters (α, δ) corresponding to the first and second components of the mixture. The AM algorithm was implemented for each of the likelihood functions separately and run until the convergence of the posterior distributions was achieved. Convergence was determined by both visual assessment of the posterior chains and the

Gelman and Rubin [1992] *R* statistic which considers convergence in terms of the variance within a single chain and the variance between multiple parallel chains.

Additional runs using alternative MCMC algorithms (including the differential evolution adaptive Metropolis algorithm [*Vrugt et al.*, 2008]) were performed to confirm the results of parameter estimation and convergence found using the AM algorithm.

A Synthetic Application

The forcing data from two unique catchments were utilized in this synthetic case study. The Black River catchment (Queensland, Australia; runoff ratio = 0.26; 24 years of daily data) was used to test the ability of the likelihood functions under ephemeral conditions (presence of zero-inflation). The Wadbilliga River catchment (New South Wales, Australia; runoff ratio = 0.35; 24 years of daily data) was used to test the selection framework on non-ephemeral streams (absence of zero-inflation). In each case, the observed rainfall and evapotranspiration data were used to force the AWBM with a set of fixed parameter values (for ephemeral catchment: $S = 110$, $K = 0.75$, $BFI = 0.5$; for perennial catchment: $S = 80$, $K = 0.9$, $BFI = 0.65$). The resultant stream discharges were corrupted with noise to create the synthetic time series. The form of the corrupting noise was designed to match the assumptions of likelihood functions E (zero-inflated, heteroscedastic, independent errors), F (zero-inflated, homoscedastic, correlated errors), G (heteroscedastic, correlated errors), and H (zero-inflated, heteroscedastic, correlated errors), resulting in four synthetic datasets on which to test the framework. The Black River catchment was used for cases E, F, and H, while the Wadbilliga River catchment was used for case G. For heteroscedastic noise, the one-parameter version of the Box-

Cox transform was used with $\lambda = 0.5$, and for correlated noise an AR(1) model was used with $\phi = 0.8$. Zero-inflated signals were generated by adding corrupting noise (of the specified form) exclusively to the non-zero observations. In all cases the corrupting noise had a mean of zero and a standard deviation of one-tenth.

The results were analyzed using five criteria (efficiency, normality, degree of zero-inflation, independence, and uncertainty) to account for differing aspects of performance. Efficiency was measured as the ability of the likelihood function to identify the true values of the AWBM parameters (S, BFI, K); normality was assessed with the Filliben statistic [Filliben, 1975]; zero-inflation was evaluated in terms of the percentage of zero observations that were modeled with zero error; independence was investigated with the Ljung-Box statistic [Ljung and Box, 1978]; and uncertainty was judged based on the reliability statistic (a measure of the percentage of observations captured by prediction intervals) [Yadav *et al.*, 2007].

For the synthetic application (due to its simplicity and our a priori knowledge of the true likelihood), each of these measures was given equal weight in our assessment and was scored as either a success (1) or a failure (0); summing the score for each, a total score (out of five) was computed. Success or failure for the criteria were determined based on simple rules for each metric: (1) if the posterior captured the true value for each parameter, the efficiency metric was scored a success; (2) if the Filliben statistic was greater than 0.9, the normality metric was scored a success; (3) if all zero observations were predicted with zero error, the zero-inflation metric was scored a success; (4) if the Ljung-Box statistic was less than the 5% critical value, the independence metric was

scored a success; and (5) if the percent of observations captured by the uncertainty interval was (approximately) equal to the expected percent capture, the uncertainty metric was scored a success. Note that although five assessment criteria are being considered in the assessment strategy for the synthetic case, they are not independent of one another. As such, instead of quantitatively weighting the different assessment criteria, a more purpose-specific qualitative approach is advocated here for applications to real data.

Figure 3 shows the summary scores for each of the likelihood functions (A-H) corresponding to each of the four synthetic datasets tested (synthetic error structures matching likelihood function E-H's assumptions). For each of the synthetic datasets, the true error structure is correctly identified (perfect rank summary score). Synthetic datasets E and G are also correctly identified by likelihood function H. This is a result of H being the parent equation to E and G, when there is no autocorrelation H collapses to E and when there is no zero-inflation H collapses to G. In the stepwise approach advocated here, the modeler would stop at the simplest correct model (with a score of five in this example). This approach would preclude the testing of more complicated models because simplicity is favored over complexity given equal performance (i.e., H would not be tested because E and G were previously identified as the true model and are less complex).

Focusing more closely on the synthetic dataset for case E, the strengths of the selection framework and assessment criteria become clear. Table 2 indicates the individual and total scores for each of the formal likelihood functions examined. As noted previously, the true likelihood function is selected based on the assessment criteria. A

more detailed look at the results reveals that only four of the eight likelihood functions (C, E, G, and H) resulted in calibrated parameter posterior distributions that contained the true values of the AWBM parameters used to generate the dataset (the "efficiency" criteria for this application). All likelihood functions correctly identified no residual autocorrelation. Only likelihood functions E and H obtained reasonable results for the Filliben r statistic (> 0.9), indicative of Gaussian residuals. The zero-inflation criterion, which considers the percentage of zero observations that are perfectly modeled, identified the four likelihood functions that explicitly account for this component (B, E, F, and H); however, all eight likelihoods resulted in at least 95% success. Finally, the reliability metric was analyzed to ensure that the uncertainty bounds are capturing the correct percentage of data (i.e., 90% interval captures 90% of the observations). In this case, likelihood functions C, E, G, and H were successful, while the others featured uncertainty intervals that captured too much of the data (excessive interval width).

Note that while likelihood E and its parent model, likelihood H, both resulted in the same "answer", likelihood H was more computationally expensive to execute due to the necessary calibration of additional parameters contained in the autoregressive component (which had an optimal value of zero). The next best performing likelihood functions, C and its parent G, failed to properly account for the ephemeral nature of the catchment, which resulted in non-normal residuals (as measured by the Filliben statistic) due to the high zero-inflation. Similar results can be shown for each of the synthetic case studies.

As a brief example of how the logic presented in Figure 1 is carried out in practice, let us examine our most complex synthetic example (that consistent with the form of Likelihood H). While we have presented results for testing all eight of the likelihood functions for each of our synthetic data cases, it can be shown that the methodology outlined in Figure 1 will lead the modeler to the correct likelihood function. Figure 4 outlines the procedure for doing so for this synthetic example. In the first step, the observed discharge record was examined for the percentage of zero values. In this case the number of zero observed streamflows made up more than 75% of the 24 years of record. Following Figure 1 and the 10% threshold rule discussed previously, the initial likelihood function to try will include a zero-inflation component (Likelihood B). The model is then calibrated with Likelihood B and the residuals are checked for heteroscedasticity (Figures 1 and 4). Because of the prevalence of heteroscedastic residuals, the methodology advocates the addition of a Box-Cox transform to the likelihood function. Continuing through the flow of Figure 1, this suggests we move to Likelihood E. The model is then calibrated again, this time using Likelihood E, and the residuals are re-checked for heteroscedasticity (to ensure the transform worked) and checked for autocorrelation. Based on the ACF shown in “Step 3” of Figure 4, it is apparent that autocorrelation is also an issue that must be addressed. Again, following along with Figure 1, this suggests we add an autoregressive component to the likelihood function (and thus select Likelihood H). Finally, we calibrate the model one more time, using Likelihood H, and re-check the residuals for heteroscedasticity and autocorrelation.

As is apparent in Figure 4, these simple rules for residual checks result in the true likelihood function being selected for this synthetic example.

A Real Data Application

Continuing with the Black River catchment (ephemeral; Queensland, Australia) dataset that was used to generate synthetic data for the synthetic application, we test here the ability of the likelihood function specification framework to perform under real conditions. In this application, each of the eight likelihood functions contained in the selection framework were sequentially applied and analyzed, as with the synthetic application. The same assessment criteria were used in the analysis; however, to determine “efficiency”, the Nash-Sutcliffe efficiency was used rather than proper identification of the true parameter values (as these are unknown). The “uncertainty” metric of reliability was selected as a criterion of particular interest, in line with a forecasting modeling scenario. This criterion should reflect a general overall appropriateness of the likelihood function as it propagates the assumptions of the likelihood function through its computations.

For this study, of the 24 years of available data the first year was used as a warm-up period and the remaining 23 years of data were used for model calibration. As noted previously, the Box-Cox transformation parameter was set to a fixed value of 0.5 (section 3.2). Autoregressive models of order one through four were evaluated in the likelihood function featuring the autoregressive component. Based on these results an AR(2) model was selected; higher order AR parameters were found to be insignificant.

Table 3 shows the efficiency scores resulting from the optimum parameter set from calibration of each of the eight likelihood functions. Because different likelihood functions and efficiency scores will favor different components of the hydrograph (low/high discharges, timing/volume errors) the use of an efficiency score to determine the best likelihood function is problematic. As Table 3 highlights, likelihood functions that apply the Box-Cox transform have higher values of Nash-Sutcliffe efficiency, but maintain similar values to the Gaussian likelihood functions in terms of absolute bias/deviation of runoff volume that characterize the volumetric errors produced by the predictions. The closeness of the efficiency values makes it difficult to convincingly discriminate between the performance of competing likelihood functions.

Notwithstanding the similarities found in many of the model efficiencies, the parameter posterior distributions were quite variable across the eight different likelihood functions. Figure 5 shows box plots for the storage parameter (S) of the AWBM model. The box indicates the interquartile range, the central line represents the median value, and the whiskers show the extremes of the posterior distribution. The top row, consisting of the likelihood functions that do not account for heteroscedasticity, shows considerably wider interquartile ranges for the parameter when compared with the bottom row. This spread indicates the untransformed likelihoods result in more uncertainty in the parameter values (less peaked posterior distributions) than the transformed likelihoods.

The randomness of the data was tested quantitatively with the Ljung-Box test statistic. As previously mentioned, the likelihood functions that incorporated the autoregressive component were of the second order. Despite this inclusion in likelihood

functions D, F, G, and H, each of the eight likelihood functions considered failed to remove all the autocorrelation from the residuals at a 5% significance level; however, qualitative analysis of the correlograms indicate that for the residuals arising under likelihoods D, F, G, and H, much of the autocorrelation was removed. These results suggest that a component to address correlation that is more complex than a simple linear autoregressive model is needed to fully address the autocorrelation issues. Another reason for the inability of the autoregressive component to remove autocorrelation from the errors could be the use of the rather simple rainfall-runoff modeling structure in our 3-parameter conceptual model (one-storage version of the AWBM).

Figure 6 shows the quantile-quantile plots of the residuals for each of the eight likelihood functions considered. In addition, the Filliben r statistic is displayed on each subplot; as the QQ plot becomes more linear the Filliben r value will approach its optimum value of one. The top row of the figure shows the results for the untransformed likelihood functions, while the bottom row shows the results for the transformed likelihoods. Note that the scales on the y-axes are dramatically different between the top and bottom rows due to the data transformation. Moving from left to right, the likelihood functions are increasing in complexity (number of parameters/included components). From this figure, it becomes clear that the untransformed residuals do not satisfy the assumptions that the likelihood function is imposing. However, for the transformed residuals the QQ plots become much more linear and the values of the Filliben r statistic begin to approach its optimum. Given the results of Figure 6, it appears that the strongest candidate likelihood function is that which accounts for heteroscedasticity and zero-

inflation (E). The results further indicate that additional consideration should be given to likelihood functions that consider only heteroscedasticity (C) and to a lesser extent that which accounts for heteroscedasticity, zero-inflation, and autocorrelation (H).

Moving to the fourth aspect of the assessment (zero-inflation performance) reveals that the likelihood function that was best able to simulate zero discharge observations with zero error was likelihood E, with each properly simulating over 98.5% of all zero discharge observations. This likelihood function has explicitly incorporated the zero-inflation logic into its formulation. The results obtained from this piece of the assessment support the information provided by the QQ plots (Figure 6).

Finally, each likelihood function's ability to produce reasonable uncertainty estimates was considered. Table 4 shows the reliability (percentage of observations captured by the interval) and sharpness (mean width of the interval) statistics of the computed 90% prediction intervals. The reliability should be equal to the desired interval percentage (i.e., 90% of observations should be captured by a 90% interval) with the smallest possible value of the sharpness. Focusing on the best performing likelihood functions from the previous criteria (likelihoods C, E, and H), it is obvious that good performance as quantified by the QQ plots and Filliben r statistic result in good performance via the reliability metric as these three likelihood functions are again the best performing. Each of the "untransformed" likelihoods (A, B, D, F) inappropriately quantify the uncertainty intervals, capturing too many of the observations. The best performing of these untransformed likelihood functions (D) captures 7.9% too much of

the data; for the 23 years of data being used that translates to 1.8 years of data that were captured by the interval that should not have been.

Examining the results associated with likelihoods C, E, and H more closely provides one final check of the dependability of the uncertainty assessment. Although the reliability is similar for likelihoods C, E, and H, likelihood C does not account for zero-inflation and likelihood H fails to successfully simulate zero-inflation (despite its inclusion in the likelihood function). Given this, more of the observations being captured are zero observations for likelihoods C and H, leading to an artificially narrow mean interval widths (sharpness; to a much greater extent for C than H). On the other hand, as Figure 7 highlights, likelihood E accounts for the zero-inflation successfully and, because of this, captures near 90% of the non-zero observations, which provides a more meaningful interval. This is directly reflected in the wider mean interval widths for these likelihood functions.

Aggregating the results of the five-level assessment into a single “optimal solution” is difficult to do given the interdependencies of the various measures used; however, likelihood functions C, E, and H were consistently among the best performing. In this case, it appears likelihood function E (which includes components for heteroscedasticity and zero-inflation) most closely matches the error structure of the model residuals as quantified by the efficiency, zero-inflation, normality, and uncertainty assessment criteria. We presented results from each likelihood function to show a comprehensive analysis and comparison of model predictions. However, a stepwise approach would begin by diagnosing zero-inflated errors, then heteroscedasticity in the

model residuals, and finally potential autocorrelation. This would then identify likelihood E as the most appropriate approach. Despite the increased difficulty caused by using real data, similar results held for the real case study as were found for the ideal synthetic case study.

Discussion

As the interest in uncertainty analysis for hydrologic modeling studies has increased, so too has the usage of statistical methods necessary to appropriately quantify such uncertainty. Although a debate still persists about which statistical method is most appropriate (informal or formal), both types of approach are being applied widely. As pointed out by *Mantovan and Todini* [2006] and *Stedinger et al.* [2008] informal uncertainty approaches are informal due to the selection of likelihood function, and this informality based on the use of informal likelihood functions (objective functions) leads to confidence/prediction intervals that lack statistical meaning. This fact clearly highlights the critical importance of the adequacy of the underlying assumptions of the likelihood function (formal or informal). Guidance on how to improve our ability to properly form likelihood functions is greatly needed. However, to date, there has been little structured direction offered to this pursuit in hydrologic modeling despite general guidelines proposed in statistical textbooks [e.g., *Box and Tiao*, 1973] and application-specific studies highlighting the importance of such [e.g., *Kuczera*, 1983; *Schoups and Vrugt*, 2010; *Sorooshian and Dracup*, 1980]. Here we introduced a framework to assist in formal likelihood function selection under an iterative improvement scheme following a

top-down modeling paradigm. The framework is accompanied by a toolkit containing all the visual and statistical residual tests described previously, as well as the log likelihood functions. The toolkit consists of a collection of Matlab® functions requiring only basic inputs (such as observed and predicted stream discharge). The functions are available for download at <http://watershed.montana.edu/analysis/smith/software.html>.

Recently there has been a focus on the types of errors that arise in hydrologic modeling applications. In particular, *Beven and Westerberg* [2011] and *Beven et al.* [2011] point out the issue of informative and disinformative hydrologic data in terms of their value in hydrologic modeling, arguing that formal statistical approaches make the assumption that the errors are aleatory whereas periods of disinformation represent epistemic errors. Although the framework as posed here considers the entire record of data to be informative, it should be noted that the zero-inflation component in the likelihood framework does formally address the non-stationarities that arise in the model residuals for this type of catchment. Under such circumstances, the outlined framework addresses the assumptions of aleatory errors (additive, stationary, random) in a formal manner. In particular, the treatment of non-stationarity can be seen in comparing the results from likelihood C with the results from likelihood E for the real data example.

The assumptions made by the likelihood function have wide-ranging effects on the results of a hydrologic model calibration, foremost among them the quantification of uncertainty. While a handful of studies have attempted to ensure the assumptions that are made by the likelihood function fit more closely with the actual data, most are done so in an application-specific manner. The general framework proposed here gathers many of

the previously applied approaches and consolidates them. Using a set of simple conventions the framework guides the modeler toward a more appropriate likelihood function via a top-down perspective (from simple to complex). However, it should be noted that the framework is not intended to be used blindly (with hard and fast rules), but rather as a construct that helps the modeler traverse the difficulties common to finding an appropriate likelihood function for hydrologic modeling applications.

The proposed likelihood function selection toolkit offers a broad collection of forms that encompass many of the most commonly cited forms that hydrologic errors take (i.e., autocorrelation, heteroscedasticity, and zero-inflation), while allowing the flexibility to incorporate other forms if desired. These error forms can be linked to the hydrologic model structure and data directly. For example, the "bucket storage" accounting common to most simple hydrologic models often gives rise to autocorrelated errors, discussed in further detail by *Kavetski et al.* [2003]. Ephemeral streams tend to result in a high number of zero discharges that are modeled without error leading to zero-inflated residuals. Heteroscedasticity is often brought about by multiple states of discharge (low/high streamflows). The knowledge of these connections between error form and model function provide the motivation for more thoughtful inclusion of such components into the likelihood function when justified by appropriate analysis of the model residuals.

There are a number of ways to account for these typical error characteristics, including many forms not tested within the framework presented here. *Schaefli et al.* [2007] addressed the issue of heteroscedasticity using a mixture of two normal

distributions. *Schoups and Vrugt* [2010] introduced a flexible likelihood function (the “generalized likelihood” or GL) that addresses the issues of heteroscedasticity, correlation, bias, and non-Gaussian characteristics. In their study, *Schoups and Vrugt* [2010] utilize the Skew Exponential Power Distribution as the foundation for their likelihood function and provide a thorough study that exploits many of the tests/checks advocated here to check the validity of the assumptions made by their likelihood function. Although such likelihood innovations are not included in the decision framework applied here, the concept of the framework was to be a flexible tool that has the ability to be adapted to different likelihood functions. In doing so, the framework maintains its usefulness in terms of guiding the modeler through a set of universal checks/tests to ensure that the validity of assumptions is properly considered.

Ultimately, the intent of this manuscript was not to prove one likelihood function (or likelihood function modification) superior to another. Rather, we wished to investigate how the case study could be instructive on the importance of the iterative process from a diagnostic perspective. At the same time, we also wished to highlight the importance of an iterative framework within such checks are made. Although complex, reducible likelihood functions are convenient in that their usage has the potential to simplify the iterative process, there are pitfalls that come with such an approach. For instance, the inclusion of multiple components (e.g., heteroscedasticity, autocorrelation, etc.) in a single step without intermediate checks has the potential to mask characteristics of the residuals or unnecessarily account for a certain characteristic (i.e., the inclusion of an autoregressive component when the errors are not correlated is not guaranteed to “fall

out” of the likelihood – falsely signifying autocorrelation). Because the systems being modeled are complex and widely variable, a fully nested approach can offer a false sense of security to the modeler that the likelihood function’s assumptions are properly characterizing the structure of the errors.

To fully understand the role these likelihood function components play in the model predictions, robust model checking is essential. These checks ensure that the resultant model predictions are appropriate (or at the very least more appropriate), both physically and conceptually. As such, model checking is very commonly done, however, it is usually done using simple efficiency criteria such as the Nash-Sutcliffe efficiency [*Nash and Sutcliffe*, 1970]. Recently, there has been a push toward considering performance with regard to multiple criteria that highlight the different aspects of the model simulation. While these multi-objective approaches offer improvement over a single efficiency statistic approach, they focus primarily on the appropriateness of the model’s fit; there is still a lack of robust checking on the underlying statistical assumptions used to make the predictions themselves in hydrologic modeling studies. As was highlighted by the synthetic test case, it is not unlikely for an inappropriate likelihood function to yield similar (and often indistinguishable) Nash-Sutcliffe efficiency results to the true likelihood function. Whereas, a more robust model checking strategy that considers both the physical efficiency and the statistical appropriateness has greater power to decipher the true error structure (likelihood function) and yield "the right answers for the right reasons" [*Kirchner*, 2006].

This enhanced error structure interpretability has the potential to have a major impact in regionalization studies. As was evident from Figure 5, the calibration of the AWBM structure with each of the different likelihood functions led to significant variations in the mean of the storage parameter. By identifying a more appropriate likelihood function (one that is most similar to the error structure of the residuals) the chance of identifying biased parameter estimates is reduced and the chance of successfully transferring parameter sets to additional catchments is improved.

The ability of the framework to guide the modeler toward the correct likelihood function was exhibited with a synthetic application featuring four artificial datasets with varying error structures, as well as a real-world application to the ephemeral Black River catchment in Queensland, Australia. In all applications, the simplified version of the Australian Water Balance Model was used to produce predictions of stream discharge given daily inputs of rainfall and potential evapotranspiration. Parameter estimation and uncertainty analysis was performed with the Adaptive Metropolis algorithm under a formal Bayesian statistical approach (all results were confirmed using alternative MCMC algorithms). The performance of the likelihood functions were assessed based on multiple criteria, including efficiency, zero-inflation, correlation, normality, and uncertainty.

Analysis of the results clearly indicated that a multifaceted assessment strategy is of extreme importance in the determination of an appropriate likelihood function because incorrect likelihoods have the ability to satisfy one or more criteria well while failing at others; this characteristic was clearly illustrated by Figure 3 of the synthetic application. It was further shown that when the structure of the likelihood function was in line with

the structure of the errors, the true likelihood would be successfully identified in the assessment strategy. Translating this procedure to the Black River catchment dataset was straightforward; however, the error structure of this (and most) real data was necessarily more complicated. The real data application highlighted the flaws in a catch-all approach to formal likelihood function selection, as the most complex likelihood tested within the framework (likelihood H) was not selected as the optimal choice. It was shown that the interdependencies of the assessment criteria pose some difficulties in identifying a definitive “best” likelihood function. We advocate an alternative solution that lends more weight to certain criteria depending on the modeler’s purpose; for example, the reliability metric quantifying the performance of the uncertainty predictions may be more important in a forecasting application, whereas the normality of the residuals might be more important in a regionalization application. Despite the difficulties posed by working with real data, the application of the likelihood function selection toolkit proved valuable in emphasizing the importance of the assumptions made via the likelihood function.

Regardless of the application of the broad collection of likelihood functions, the results are not perfect (i.e., the assumptions of the likelihood function are not identical to the structural form of the errors). Other factors that may be considered to improve performance further include an alternate representation of autocorrelation (such as an autoregressive-moving average approach), a better transformation of the data to remove heteroscedasticity (potentially by mixture modeling; see *Schaepli et al.* [2007]), alternative error distribution assumptions [e.g., *Schoups and Vrugt*, 2010; *Thiemann et*

al., 2001], an accounting of input data uncertainty, or consideration of model structural uncertainty.

While an investigation into each of these potential complicating factors is outside the scope of this study's objective, many of these areas are topics of ongoing hydrologic research and combine to further complicate the hydrologic modeling problem. In the specific areas of input data uncertainty and model structural uncertainty, a wealth of recent research [for example, *Kavetski et al.*, 2006; *Renard et al.*, 2010; *Thyer et al.*, 2009] has indicated the importance of addressing these components for improved quantification of total uncertainty. At the same time, model structural uncertainty is known to affect the results of model calibration. Many authors have investigated this issue [e.g., *Clark et al.*, 2008; *Smith and Marshall*, 2010] as well as related issues such as model time step [e.g., *Clark and Kavetski*, 2010; *Kavetski and Clark*, 2010]. Further, the recent investigation of *Schoups and Vrugt* [2010] into distributional forms that move beyond the basic Gaussian distributional underpinning of most approaches shows promise for the future.

Conclusions

The framework for selecting an appropriate likelihood function presented here has the aim of increasing the awareness of the consequences of using an inappropriate likelihood function within a formal statistical approach, while providing guidance on a general procedure for determining a more reasonable set of assumptions. By providing a structure that requires checking and re-checking of model residuals, the modeler is

necessarily more aware of the potential deficiencies in their methods and allowed to work toward correcting them.

As presented here, the framework was arranged to consider zero-inflation, heteroscedasticity, and autocorrelation. Other components (or alternate representations for the already included components) could easily be added to the framework. The utility of the method lies in its simplicity and flexibility. It is flexible enough to be used in a top-down manner (as advocated here) or from a bottom-up approach due to the nested design of the likelihood functions themselves. At the same time, the framework relies on simple checks on the residuals that can be executed quickly (given the calibrated parameter posterior distributions) and assessed under a multifaceted strategy that allows for a more comprehensive understanding the performance.

Acknowledgements

This work was made possible by funding from the Australian Research Council. We thank Francis Chiew for providing the data used in this study, Jasper Vrugt for providing the code for the differential evolution adaptive Metropolis algorithm, and Gerrit Schoups for providing the code for the Generalized Likelihood function. Finally, we wish to thank Keith Beven and two anonymous reviewers for their constructive suggestions that greatly improved the quality of this manuscript.

Table 1. The log likelihood functions, assumptions, and calibrated parameters used in this study.

Name	Assumptions	log Likelihood Function	Calibration Parameters
A	Gaussian, homoscedastic, independent errors	$\ell_A = -\frac{n}{2} \cdot \log(2\pi\sigma^2) + \sum_t -\frac{\varepsilon_t^2}{2\sigma^2}$	σ^2
B	Gaussian, homoscedastic, independent, zero-inflated errors	$\ell_B = n_1 \cdot \log(\rho) + n_2 \cdot \log(1-\rho) - \frac{n_2}{2} \cdot \log(2\pi\sigma^2) + \sum_{t_2=1}^{n_2} -\frac{\varepsilon_{t_2}^2}{2\sigma^2} - \frac{n_3}{2} \cdot \log(2\pi\sigma^2) + \sum_{t_3=1}^{n_3} -\frac{\varepsilon_{t_3}^2}{2\sigma^2}$	σ^2
C ^{b,c}	Gaussian, heteroscedastic, independent errors	$\ell_C = -\frac{n}{2} \cdot \log(2\pi\sigma^2) + \sum_t -\frac{(\varepsilon_t^*)^2}{2\lambda_1^2\sigma^2} + \kappa$	$\sigma^2, \lambda_1, \lambda_2$
D	Gaussian, homoscedastic, autocorrelated errors	$\ell_D = -\frac{n-p}{2} \cdot \log(2\pi\sigma^2) + \sum_t -\frac{\left(\varepsilon_t - \sum_{j=1}^p \phi_j \cdot \varepsilon_{t-j}\right)^2}{2\sigma^2}$	σ^2, ϕ
E ^{b,c}	Gaussian, heteroscedastic, independent, zero-inflated errors	$\ell_E = n_1 \cdot \log(\rho) + n_2 \cdot \log(1-\rho) - \frac{n_2}{2} \cdot \log(2\pi\sigma^2) + \dots$ $\sum_{t_2=1}^{n_2} -\frac{(\varepsilon_{t_2}^*)^2}{2\lambda_1^2\sigma^2} + \kappa_{n_2} - \frac{n_3}{2} \cdot \log(2\pi\sigma^2) + \sum_{t_3=1}^{n_3} -\frac{(\varepsilon_{t_3}^*)^2}{2\lambda_1^2\sigma^2} + \kappa_{n_3}$	$\sigma^2, \lambda_1, \lambda_2$

Table 1 (continued). The log likelihood functions, assumptions, and calibrated parameters used in this study.

F	Gaussian, homoscedastic, autocorrelated, zero-inflated errors	$\ell_F = n_1 \cdot \log(\rho) + n_2 \cdot \log(1 - \rho) - \frac{n_2 - p}{2} \cdot \log(2\pi\sigma^2) + \dots$ $\sum_{t_2=1}^{n_2} - \frac{\left(\varepsilon_{t_2} - \sum_{j=1}^p \phi_j \cdot \varepsilon_{t_2-j} \right)^2}{2\sigma^2} - \frac{n_3 - p}{2} \cdot \log(2\pi\sigma^2) + \sum_{t_3=1}^{n_3} - \frac{\left(\varepsilon_{t_3} - \sum_{j=1}^p \phi_j \cdot \varepsilon_{t_3-j} \right)^2}{2\sigma^2}$	σ^2, ϕ
G ^{b,c}	Gaussian, heteroscedastic, autocorrelated errors	$\ell_G = -\frac{n-p}{2} \cdot \log(2\pi\sigma^2) + \sum_t - \frac{\left(\varepsilon_t^* - \sum_{j=1}^p \phi_j^* \cdot \varepsilon_{t-j}^* \right)^2}{2\lambda_1^2 \sigma^2} + \kappa$	$\sigma^2, \lambda_1, \lambda_2, \phi$
H ^{b,c}	Gaussian, heteroscedastic, autocorrelated, zero-inflated errors	$\ell_H = n_1 \cdot \log(\rho) + n_2 \cdot \log(1 - \rho) - \frac{n_2 - p}{2} \cdot \log(2\pi\sigma^2) + \dots$ $\sum_{t_2=1}^{n_2} - \frac{\left(\varepsilon_{t_2}^* - \sum_{j=1}^p \phi_j^* \cdot \varepsilon_{t_2-j}^* \right)^2}{2\lambda_1^2 \sigma^2} + \kappa_{n_2} - \frac{n_3 - p}{2} \cdot \log(2\pi\sigma^2) + \sum_{t_3=1}^{n_3} - \frac{\left(\varepsilon_{t_3}^* - \sum_{j=1}^p \phi_j^* \cdot \varepsilon_{t_3-j}^* \right)^2}{2\lambda_1^2 \sigma^2} + \kappa_{n_3}$	$\sigma^2, \lambda_1, \lambda_2, \phi$

^aWhere n is the total number observations, n_1 is the number of zero discharge observations with zero error, n_2 is the number of zero discharge observations with non-zero error, n_3 is the number of non-zero observations, the untransformed errors are defined as $\varepsilon = Obs - Pred$, the one-parameter Box-Cox transformed errors are defined as $\varepsilon^* = \log(Obs + \lambda_2) - \log(Pred + \lambda_2)$, and the two-parameter Box-Cox transformed errors are defined as $\varepsilon^* = (Obs + \lambda_2)^{\lambda_1} - (Pred + \lambda_2)^{\lambda_1}$.

^bFor one-parameter Box-Cox transformed data, set $\lambda_1 = 1$ in given log likelihood function equations; set $\kappa = -\log(Obs + \lambda_2)$.

^cFor two-parameter Box-Cox transformed data, set $\kappa = (\lambda_1 - 1) \cdot \log(Obs + \lambda_2)$.

Table 2. Summary of Performance Ranks for Synthetic Data Case E

Likelihood	Efficiency	Correlation	Normality	Zero-Inflation	Uncertainty	Total
A	0	1	0	0	0	1
B	0	1	0	1	0	2
C	1	1	0	0	1	3
D	0	1	0	0	0	1
E	1	1	1	1	1	5
F	0	1	0	1	0	2
G	1	1	0	0	1	3
H	1	1	1	1	1	5

^aScores of 0 (failure) or 1 (success) were given for each assessment metric and likelihood function. Success was defined as capturing the true parameter values within the calibrated posterior for efficiency, a value of the Ljung-Box statistic less than the 5% critical value, a Filliben r statistic greater than 0.9 for normality, 100% accurate modeling of zero discharge observations for zero-inflation, and a value for reliability within $\pm 2.5\%$ of the prescribed interval width (90%).

Table 3. Summary of efficiency criteria for the real case study

Likelihood	NSE	ABIAS	DRV
A	0.5455	0.4662	1.4662
B	0.5345	0.4984	1.4984
C	0.6640	-0.0258	0.9742
D	0.5880	0.3152	1.5360
E	0.5647	-0.2023	1.2023
F	0.5856	0.3725	1.6505
G	0.8706	-0.0420	-0.9356
H	0.8490	-0.0523	-0.2638

^aWhere NSE is the Nash-Sutcliffe efficiency, ABIAS is the absolute bias, and DRV is the deviation of runoff volume.

Table 4. Summary of uncertainty criteria for the real case study

Likelihood	Reliability (90%)	Sharpness (mm/day)
A	97.92%	11.17
B ^b	99.07%	18.49
C	92.00%	1.87
D	97.94%	11.29
E ^b	87.43%	7.45
F ^b	99.13%	18.77
G	99.23%	8.44
H ^b	88.90%	5.19

^aWhere reliability is the percentage of observations captured by the uncertainty interval and sharpness is the mean interval width.

^bFor likelihood functions that assumed zero-inflation for the model residuals, statistics were only calculated on non-zero predictions.

Figures

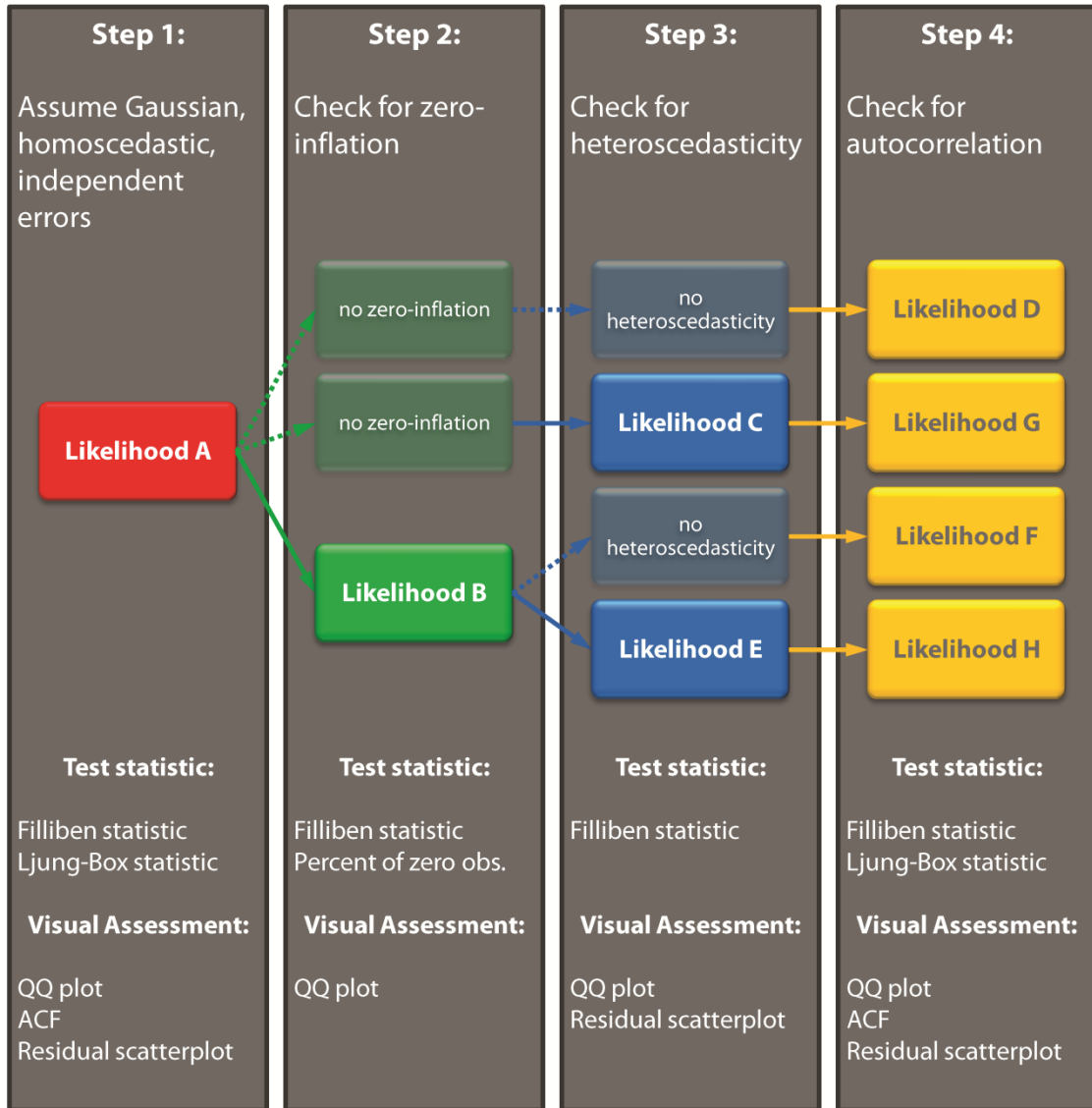


Figure 1. The likelihood function selection logic. A basic assumption of the form of the residuals is made and followed by checks on typical areas of deficiency at each level of the logic.

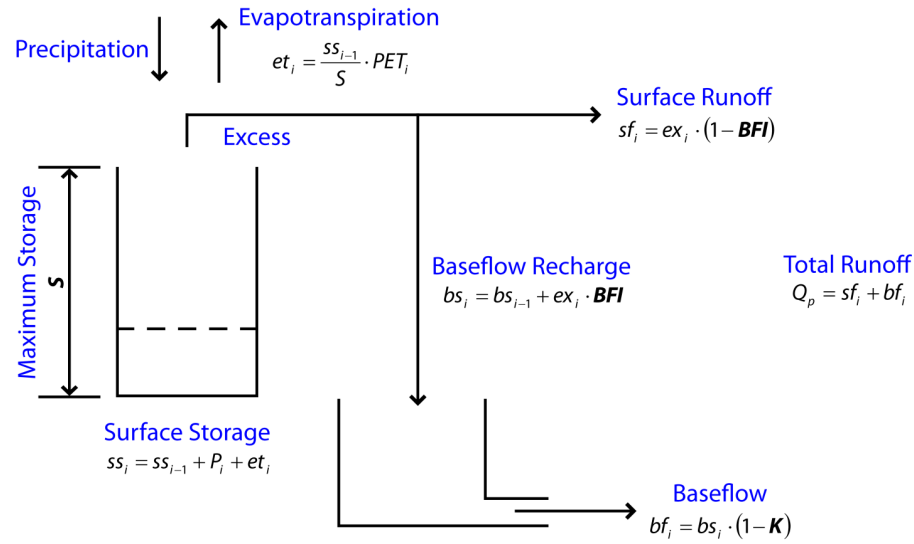


Figure 2. The simplified version of the Australian Water Balance Model structure with calibrated parameters S , BFI , and K .

Synthetic Noise Form					
	E	F	G	H	
Likelihood Function	A	1	3	1	0
	B	2	4	1	2
	C	3	1	4	2
	D	1	2	1	1
	E	5	2	4	3
	F	2	5	1	2
	G	3	0	5	3
	H	5	2	5	5

Figure 3. Graphical representation of likelihood function performance for each synthetic data case and likelihood function. Scores are rated on a scale of (0-5), where a value of (5) represents success in each of the considered assessment criteria.

Step 1:

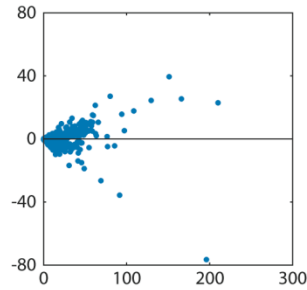
Check for zero-inflation
(>10% zeros in observed
discharge record)

$\approx 75\%$ zero observations

Add zero-inflation
component to likelihood
function

Step 2:

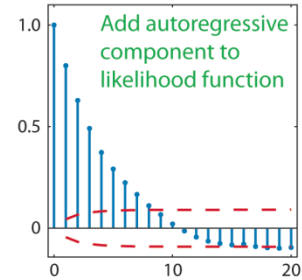
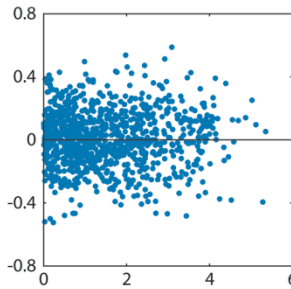
Check for heteroscedastic
residuals



Add Box-Cox transform
component to likelihood
function

Step 3:

Re-check for heterosce-
dastic residuals
Check for autocorrelated
residuals

*Step 4:*

Re-check for heterosce-
dastic residuals
Re-check for autocorrel-
ated residuals

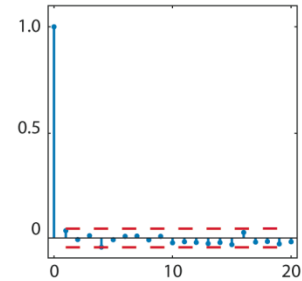
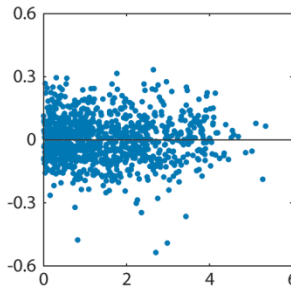


Figure 4. Step-by-step application of the logic presented in Figure 1 for the synthetic data case corresponding to Likelihood H. The labels for the residual scatterplots are observed discharge on the x-axis and model residuals on the y-axis with units of mm/day for Step 2 and units of log(mm/day) for Steps 3, 4. The labels for the ACFs are lags on the x-axis and autocorrelations on the y-axis.

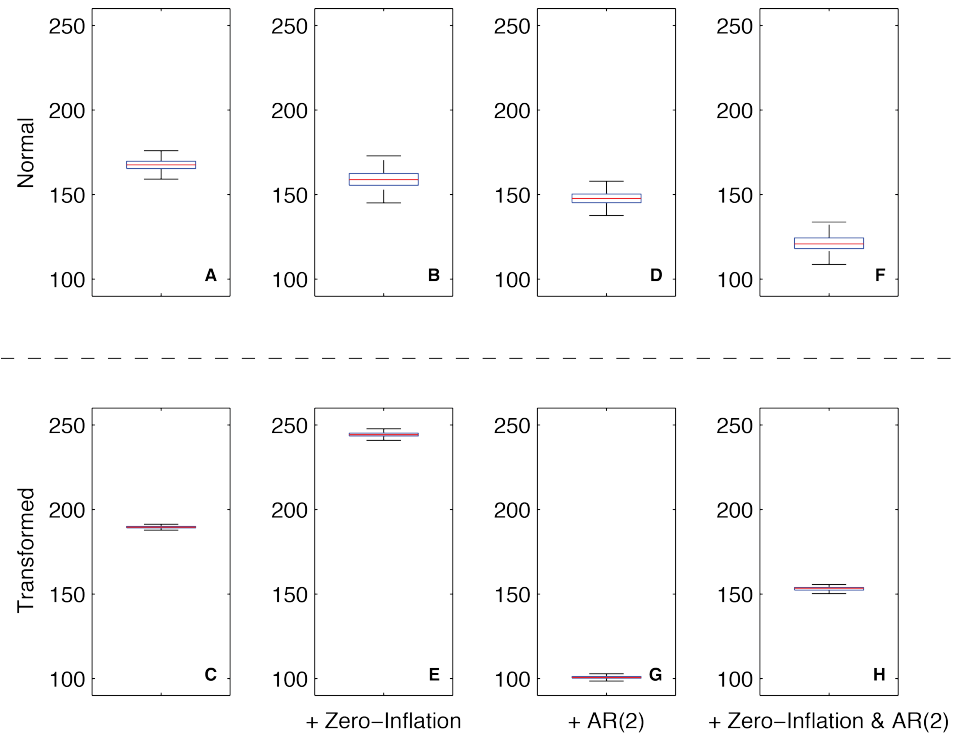


Figure 5. Box plots for the storage parameter (S, mm/day) of the AWBM model, as calibrated for each of the eight likelihood functions (A-H) of the real case study.

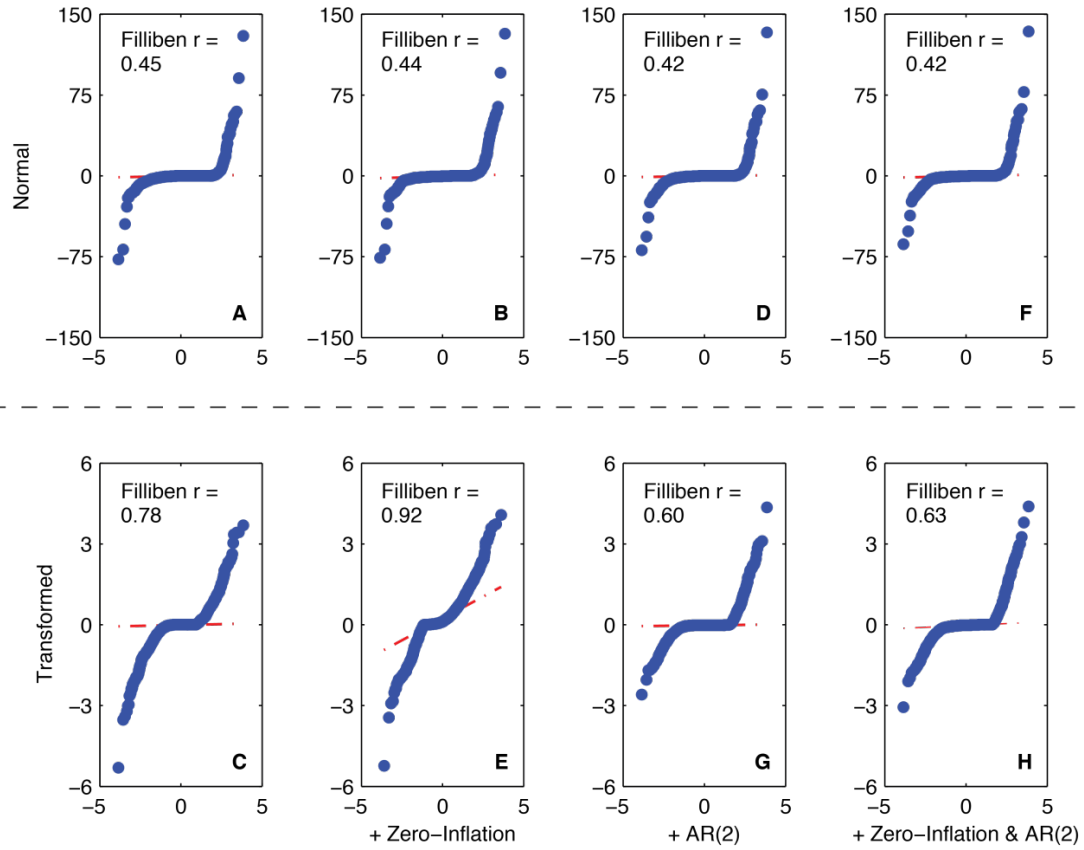


Figure 6. Quantile-quantile plots with Filliben r statistic for each of the eight likelihood functions (A-H) examined in the real case study.

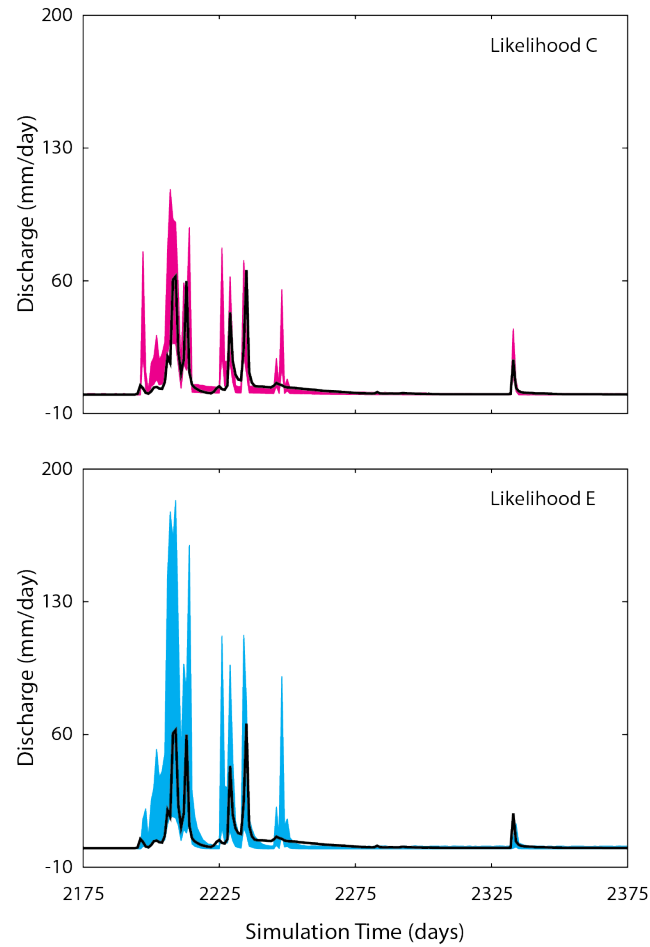


Figure 7. Hydrograph (black line) and uncertainty bounds for a 200-day snapshot from the 23-year simulation period for two likelihood functions (C and E) examined under the conditions of the real case study, highlighting the under-capture of the uncertainty bounds for Likelihood C on higher streamflows.

References

- Ajami, N. K., Q. Duan, and S. Sorooshian (2007), An integrated hydrologic Bayesian multimodel combination framework: Confronting input, parameter, and model structural uncertainty in hydrologic prediction, *Water Resour. Res.*, 43, W01403, doi:10.1029/2005WR004745.
- Bates, B. C., and E. P. Campbell (2001), A Markov chain Monte Carlo scheme for parameter estimation and inference in conceptual rainfall-runoff modeling, *Water Resour. Res.*, 37(4), 937-947.
- Beven, K., P. Smith, and J. Freer (2007), Comment on "Hydrological forecasting uncertainty assessment: Incoherence of the GLUE methodology" by Pietro Mantovan and Ezio Todini, *J. Hydrol.*, 338(3-4), 315-318.
- Beven, K., P. J. Smith, and A. Wood (2011), On the colour and spin of epistemic error (and what we might do about it), *Hydrol. Earth Syst. Sci. Discuss.*, 8(3), 5355-5386.
- Beven, K. J., P. J. Smith, and J. E. Freer (2008), So just why would a modeller choose to be incoherent?, *J. Hydrol.*, 354(1-4), 15-32.
- Boughton, W. (2004), The Australian water balance model, *Environ. Modell. Software*, 19(10), 943-956.
- Box, G. E. P., and G. C. Tiao (1973), *Bayesian Inference in Statistical Analysis*, Addison-Wesley, Reading, MA.
- Campbell, E. P., D. R. Fox, and B. C. Bates (1999), A Bayesian approach to parameter estimation and pooling in nonlinear flood event models, *Water Resour. Res.*, 35(1), 211-220.
- Clark, M. P., A. G. Slater, D. E. Rupp, R. A. Woods, J. A. Vrugt, H. V. Gupta, T. Wagener, and L. E. Hay (2008), Framework for Understanding Structural Errors (FUSE): A modular framework to diagnose differences between hydrological models, *Water Resour. Res.*, 44, W00B02, doi:10.1029/2007WR006735.
- Clark, M. P., and D. Kavetski (2010), Ancient numerical demons of conceptual hydrological modeling: 1. Fidelity and efficiency of time stepping schemes, *Water Resour. Res.*, 46, W10510, doi:10.1029/2009WR008894.
- Duan, Q., N. K. Ajami, X. Gao, and S. Sorooshian (2007), Multi-model ensemble hydrologic prediction using Bayesian model averaging, *Adv. Water Resour.*, 30(5), 1371-1386.

- Filliben, J. J. (1975), The Probability Plot Correlation Coefficient Test for Normality, *Technometrics*, 17(1), 111-117.
- Gelman, A., and D. B. Rubin (1992), Inference from iterative simulation using multiple sequences, *Stat. Sci.*, 7(4), 457-472.
- Haario, H., E. Saksman, and J. Tamminen (2001), An adaptive Metropolis algorithm, *Bernoulli*, 7(2), 223-242.
- Hsu, K.-l., H. Moradkhani, and S. Sorooshian (2009), A sequential Bayesian approach for hydrologic model selection and prediction, *Water Resour. Res.*, 45, W00B12, doi:10.1029/2008WR006824.
- Kavetski, D., S. W. Franks, and G. Kuczera (2003), Confronting input uncertainty in environmental modelling, in *Calibration of Watershed Models*, edited by Q. Duan, et al., pp. 49-68, American Geophysical Union.
- Kavetski, D., G. Kuczera, and S. W. Franks (2006), Bayesian analysis of input uncertainty in hydrological modeling: Theory, *Water Resour. Res.*, 42, W03407, doi:10.1029/2005WR004368.
- Kavetski, D., and M. P. Clark (2010), Ancient numerical demons of conceptual hydrological modeling: 2. Impact of time stepping schemes on model analysis and prediction, *Water Resour. Res.*, 46, W10511, doi:10.1029/2009WR008896.
- Kirchner, J. W. (2006), Getting the right answers for the right reasons: Linking measurements, analyses, and models to advance the science of hydrology, *Water Resour. Res.*, 42, W03S04, doi:10.1029/2005WR004362.
- Kuczera, G. (1983), Improved Parameter Inference in Catchment Models 1. Evaluating Parameter Uncertainty, *Water Resour. Res.*, 19(5), 1151-1162.
- Kuczera, G., and E. Parent (1998), Monte Carlo assessment of parameter uncertainty in conceptual catchment models: the Metropolis algorithm, *J. Hydrol.*, 211(1-4), 69-85.
- Ljung, G. M., and G. E. P. Box (1978), On a Measure of Lack of Fit in Time Series Models, *Biometrika*, 65(2), 297-303.
- Mantovan, P., and E. Todini (2006), Hydrological forecasting uncertainty assessment: Incoherence of the GLUE methodology, *J. Hydrol.*, 330(1-2), 368-381.
- Mantovan, P., E. Todini, and M. L. V. Martina (2007), Reply to comment by Keith Beven, Paul Smith and Jim Freer on "Hydrological forecasting uncertainty assessment: Incoherence of the GLUE methodology", *J. Hydrol.*, 338(3-4), 319-324.

- Marshall, L., D. Nott, and A. Sharma (2004), A comparative study of Markov chain Monte Carlo methods for conceptual rainfall-runoff modeling, *Water Resour. Res.*, 40, W02501, doi:10.1029/2003WR002378.
- Nash, J. E., and J. V. Sutcliffe (1970), River flow forecasting through conceptual models part I - A discussion of principles, *J. Hydrol.*, 10(3), 282-290.
- Peel, M. C., F. H. S. Chiew, A. W. Western, and T. A. McMahon (2000), Extension of unimpaired monthly streamflow data and regionalisation of parameter values to estimate streamflow in ungauged catchments, in *National Land and Water Resources Audit*, edited, CRC for Catchment Hydrology, Melbourne.
- Reichert, P., and J. Mieleitner (2009), Analyzing input and structural uncertainty of nonlinear dynamic models with stochastic, time-dependent parameters, *Water Resour. Res.*, 45, W10402, doi:10.1029/2009wr007814.
- Renard, B., D. Kavetski, G. Kuczera, M. Thyer, and S. W. Franks (2010), Understanding predictive uncertainty in hydrologic modeling: The challenge of identifying input and structural errors, *Water Resour. Res.*, 46, W05521, doi:10.1029/2009WR008328.
- Reusser, D. E., and E. Zehe (2011), Inferring model structural deficits by analyzing temporal dynamics of model performance and parameter sensitivity, *Water Resour. Res.*, 47, W07550, doi:10.1029/2010wr009946.
- Samanta, S., M. K. Clayton, D. S. Mackay, E. L. Kruger, and B. E. Ewers (2008), Quantitative comparison of canopy conductance models using a Bayesian approach, *Water Resour. Res.*, 44, W09431, doi:10.1029/2007WR006761.
- Schaefli, B., D. B. Talamba, and A. Musy (2007), Quantifying hydrological modeling errors through a mixture of normal distributions, *J. Hydrol.*, 332(3-4), 303-315.
- Schoups, G., and J. A. Vrugt (2010), A formal likelihood function for parameter and predictive inference of hydrologic models with correlated, heteroscedastic, and non-Gaussian errors, *Water Resour. Res.*, 46, W10531, doi:10.1029/2009WR008933.
- Sivapalan, M., G. Blöschl, L. Zhang, and R. Vertessy (2003), Downward approach to hydrological prediction, *Hydrol. Processes*, 17(11), 2101-2111.
- Smith, T., A. Sharma, L. Marshall, R. Mehrotra, and S. Sisson (2010), Development of a formal likelihood function for improved Bayesian inference of ephemeral catchments, *Water Resour. Res.*, 46, W12551, doi:10.1029/2010WR009514.

- Smith, T. J., and L. A. Marshall (2008), Bayesian methods in hydrology: A study of recent advancements in Markov chain Monte Carlo techniques, *Water Resour. Res.*, 44, W00B05, doi:10.1029/2007WR006705.
- Smith, T. J., and L. A. Marshall (2010), Exploring uncertainty and model predictive performance concepts via a modular snowmelt-runoff modeling framework, *Environ. Modell. Software*, 25(6), 691-701.
- Sorooshian, S., and J. A. Dracup (1980), Stochastic parameter estimation procedures for hydrologic rainfall-runoff models: Correlated and heteroscedastic error cases, *Water Resour. Res.*, 16(2), 430-442.
- Stedinger, J. R., R. M. Vogel, S. U. Lee, and R. Batchelder (2008), Appraisal of the generalized likelihood uncertainty estimation (GLUE) method, *Water Resour. Res.*, 44, W00B06, doi:10.1029/2008WR006822.
- Thiemann, M., M. Trosset, H. V. Gupta, and S. Sorooshian (2001), Bayesian recursive parameter estimation for hydrologic models, *Water Resour. Res.*, 37(10), 2521-2535.
- Thyer, M., B. Renard, D. Kavetski, G. Kuczera, S. W. Franks, and S. Srikanthan (2009), Critical evaluation of parameter consistency and predictive uncertainty in hydrological modeling: A case study using Bayesian total error analysis, *Water Resour. Res.*, 45, W00B14, doi:10.1029/2008WR006825.
- Vrugt, J. A., H. V. Gupta, S. C. Dekker, S. Sorooshian, T. Wagener, and W. Bouten (2006), Application of stochastic parameter optimization to the Sacramento Soil Moisture Accounting model, *J. Hydrol.*, 325(1-4), 288-307.
- Vrugt, J. A., C. J. F. ter Braak, M. P. Clark, J. M. Hyman, and B. A. Robinson (2008), Treatment of input uncertainty in hydrologic modeling: Doing hydrology backwards with Markov chain Monte Carlo simulation, *Water Resour. Res.*, 44, W00B09, doi:10.1029/2007WR006720.
- Xu, C.-Y. (2001), Statistical Analysis of Parameters and Residuals of a Conceptual Water Balance Model – Methodology and Case Study, *Water Resour. Manag.*, 15(2), 75-92.
- Yadav, M., T. Wagener, and H. Gupta (2007), Regionalization of constraints on expected watershed response behavior for improved predictions in ungauged basins, *Adv. Water Resour.*, 30(8), 1756-1774.
- Zhang, Z., T. Wagener, P. Reed, and R. Bhushan (2008), Reducing uncertainty in predictions in ungauged basins by combining hydrologic indices, regionalization and multiobjective optimization, *Water Resour. Res.*, 44, W00B04, doi:10.1029/2008WR006833.

CHAPTER FOUR

IMPROVING HYDROLOGIC MODEL CALIBRATION USING
A FLOW-CORRECTED TIME TRANSFORMATIONContribution of Author and Co-Authors

Manuscript in Chapter 4

Author: Tyler Smith

Contributions: Responsible for the modeling and analysis performed in this study, the majority of the writing, and figure development.

Co-Author: Lucy Marshall

Contributions: Provided extensive discussions of the development of the methods used in the study and contributed to the editing of the manuscript throughout its development.

Co-Author: Brian McGlynn

Contributions: Provided extensive discussions of the development of the methods used in the study and contributed to the editing of the manuscript throughout its development.

Manuscript Information

Smith, T., L. Marshall, and B. McGlynn. Improving hydrologic model calibration using a flow-corrected time transformation. Manuscript for submittal to *Water Resources Research*.

- Journal: Water Resources Research
- Status of manuscript (check one)
 - ☒ Prepared for submission to a peer-reviewed journal
 - ☐ Officially submitted to a peer-reviewed journal
 - ☐ Accepted by a peer-reviewed journal
 - ☐ Published in a peer-reviewed journal
- Publisher: American Geophysical Union
- Anticipated date of submission: April 2012

IMPROVING HYDROLOGIC MODEL CALIBRATION USING A FLOW-CORRECTED TIME TRANSFORMATION

Abstract

The modeling of streamflow hydrographs can be highly complex, particularly due to problems caused by multiple dominant streamflow states, switching of dominant streamflow generation mechanisms temporally, and dynamic responses to inputs based on antecedent conditions. Because of these complexities and extreme heterogeneity that can exist within a single catchment, model calibration techniques are generally required to obtain reasonable estimates of the model parameters. However, when calibrating a model, the objective function used to determine the goodness of a parameter set plays a large role in determining the “optimal” parameter set. Some objective functions are weighted toward good fits on high streamflows, while others are weighted toward low streamflows. As such, multi-objective calibration schemes have evolved to combine multiple objective functions that consider different parts of the hydrograph (e.g., low, medium, high flows) or data transformations are often applied to provide better overall fits to the observations. In this study, we introduce a transformation that is new to the hydrologic modeling literature that dynamically weights streamflow in the time domain rather than the data domain. The transformation, known as flow-corrected time, is designed to provide greater weight to time periods with larger hydrologic flux. Here the flow-corrected time transformation is compared to the baseline of the untransformed case and the commonly used Box-Cox transform case. We evaluate each approach in terms of both model fit (via Nash-Sutcliffe efficiency) and the ability of the transform to redistribute the model residuals. We considered these criteria across hydrographs arising from two distinct sets of catchment conditions (snow-dominated and subtropical). It was found that the flow-corrected time transformation was better able to remove positive skew from the streamflow time series and resulted in better model fits and subsequent normality of model residuals than the untransformed or Box-Cox transformed data modeling.

Introduction

In hydrology, the use of predictive streamflow models is almost exclusively predicated on generating a good estimate of model parameters via some type of calibration procedure [Madsen *et al.*, 2002]. In general, all approaches to model

calibration in hydrologic settings, whether they be global optimization techniques [e.g., *Duan et al.*, 1992; *Tolson and Shoemaker*, 2007] or uncertainty-based methods [e.g., *Beven and Binley*, 1992; *Kuczera and Parent*, 1998], are driven by the use of cost functions to evaluate the goodness of fit between the observed and model predicted streamflows [generated for a given parameter set; refer to *Gupta and Kling*, 2011]. These functions are commonly known as objective functions in the global optimization literature as well as in informal uncertainty-based approaches [see *Beven and Binley*, 1992]. Formal uncertainty-based approaches (i.e., fully Bayesian approaches) refer to these functions as likelihood functions, which serve the same conceptual purpose as those used in global optimization or informal uncertainty-based methods but take on statistically formal mathematical forms as they describe the distribution of model errors.

In any case, the form of the cost function implies a characteristic form of the object being modeled [*Stedinger et al.*, 2008; implicitly for informal/global optimization methods and explicitly for formal Bayesian methods]. The development of cost functions to address data characteristics (heteroscedasticity, non-normality, correlation, etc.), however, is not new. Indeed, *Sorooshian and Dracup* [1980] developed a maximum likelihood estimator that accounted for heteroscedastic and correlated errors, and many others have successfully applied data transformations that attempt to resolve the issues of heteroscedasticity and non-normality [e.g., *Bates and Campbell*, 2001; *Kuczera and Parent*, 1998; *Thyer et al.*, 2002; *Wagener and McIntyre*, 2005].

In this paper, we present a calibration approach that transforms the hydrologic data in the time domain based on streamflow time rather than the data domain itself (such

as Box-Cox transformations [1964]). The approach is used to better represent the magnitude of streamflow fluxes in model calibration. The remainder of this paper is organized as follows: Section 2 will describe the time transformation and how it fits within the context of the objective function and hydrologic modeling in general, Section 3 will highlight the benefits and limitations of the transform via a simple application, and Section 4 will provide a brief discussion along with the primary conclusions of the study.

Flow-Corrected Time Transformation

In hydrology, properly characterizing model residuals can be difficult because of the often positively skewed distribution of streamflows that can lead to heteroscedastic residuals [*Sorooshian and Dracup*, 1980]. To address this complication here, we introduce a novel objective function based on a flow-corrected time transformation [refer to *Rodhe et al.*, 1996]. The basic concept of flow-corrected time (FCT) was originally used in transit time modeling to address assumptions of time invariance and constant volume [*McGuire and McDonnell*, 2006; *Rodhe et al.*, 1996]. Under the transform, “clock time” is transformed to “flow time” which is defined as

$$fct = \int_{t_0}^t q(t)dt/\bar{q} \quad (1)$$

where q is streamflow, $\bar{q}$ is the mean of streamflow and t is “clock time”.

Applying this transform to a streamflow hydrograph functionally results in the stretching of time during wet periods and the compressing of time during dry periods of the year. This results in more weight being given to times where the streamflow exceeds

the mean streamflow and ultimately favors simulations that accurately predict peak streamflow conditions. Though this has the potential of impacting predictions of low flow, high flows (or total volume of flows) are often a primary concern in water resource management. Figure 1 illustrates the effect of the transform on a simple, hypothetical streamflow hydrograph. The flow-corrected time transformation can be used within any cost function (though use in formal likelihood functions will require additional care) and can be expressed as

$$q^{fct} = w_{fct} \cdot q \quad (2)$$

where q^{fct} are the flow-corrected time weighted data, q are the original data, and w_{fct} are the flow-corrected time weights (which sum to n , the length of q). The flow-corrected time weights (w_{fct}) are simply calculated as

$$w_{fct} = q/\bar{q} \quad (3)$$

where all terms are as previously defined.

It is clear from Figure 1 and Figure 2 that such a transform greatly reduces the skewness of the distribution of observed streamflow (represented by the distinct two-state streamflow regime; baseflow, peak flow). Note that because the FCT transformation effectively leads to discrete observations having “partial flow times” (i.e., not a whole number) the data here are linearly interpolated on the flow time scale to allow estimation of the flow corrected time histogram. To further illustrate the dramatic improvement brought about by the FCT transform, Figure 3 shows box plots for the untransformed and FCT transformed streamflow distributions (along with respective mean flow references). For this sample data set, the mean streamflow value from the untransformed data is

classified as an outlier value (beyond the “whiskers” of the box plot that provide 99.3% coverage of normally distributed data), while the mean streamflow value from the flow-corrected time transformed data lies very near to the median value. Perhaps an interesting result considering some cost functions used in hydrology (e.g., the Nash-Sutcliffe efficiency) use the mean value as a baseline model to compare predictions to despite the fact that it can be representative of an “outlier” value under even typical streamflow regimes. The utility of this transformation in improving streamflow predictions will be investigated in the following section.

Case Studies

In this section, we present two simple applications to highlight the potential benefits and limitations of the flow-corrected time transformation for hydrologic modeling studies. To do this, we begin with the Nash-Sutcliffe efficiency (NSE) objective function

$$NSE = 1 - \frac{\sum (q_{obs} - q_{pred})^2}{\sum (q_{obs} - \bar{q}_{obs})^2} \quad (4)$$

where q_{obs} represents the observed streamflow, $\bar{q}_{obs}$ is the mean of the observed streamflow, and q_{pred} is the model predicted streamflow. Recall that the NSE has an implicit assumption of normality [e.g., *Montanari, 2005; Stedinger et al., 2008*]. Given the Nash-Sutcliffe objective function, we compared the results arising from using no data transformation, the FCT transformation (Eq. 1), and a Box-Cox transformation

$$q^{bc} = \log(q + \lambda) \quad (5)$$

where q^{bc} is the Box-Cox transformed streamflow and λ is a transformation parameter used to prevent $\log(0)$. The value of λ was set to 0.3, a commonly used value in hydrology [e.g., *Wagener et al.*, 2009], though it could be included as an additional parameter in the model calibration.

In each of the applications, we applied the Probability Distributed Model (PDM, Figure 4) [*Moore*, 1985; 2007]. The PDM is a flexible model structure that conceptualizes the catchment as a simple bucket whose depth is controlled by the shape of the storage's probability distribution. The probability distribution concept reflects the spatial variability of the soil's capacity to hold water. Water exceeding the soil's capacity is routed to a surface storage, while all other water drains into a subsurface storage; total streamflow at the catchment outlet is then the combination of water leaving the surface and subsurface storages at any time. The model is represented with six parameters that are described in Table 1.

For each application, we calibrated the parameters of the PDM structure using the Dynamically Dimensioned Search (DDS) introduced by *Tolson and Shoemaker* [2007], a global optimization algorithm. The DDS algorithm was developed and extensively tested on hydrologic model problems and has been shown to perform favorably against other global optimization algorithms in hydrologic applications [*Tolson and Shoemaker*, 2007].

Application to a Snow-Dominated Mountainous Catchment

We investigated the utility of the flow-corrected time transformation in the context of a snowmelt-dominated study catchment that features a fairly simple

hydrograph with one main event occurring during peak snowmelt (refer to Figure 1). The Stringer Creek catchment (5.55 km^2) is situated within the Tenderfoot Creek Experimental Forest (TCEF) of Montana, USA (lat. $46^\circ 55' \text{N}$, long. $110^\circ 52' \text{W}$) and serves as the study location for this application. Tenderfoot Creek Experimental Forest is comprised of a number of gauged sub-catchments that form the headwaters of Tenderfoot Creek (22.8 km^2), which drain into the Smith River (a tributary of the Missouri River). The historical mean annual temperature at TCEF is approximately 0°C and receives approximately 75% of its 850mm annual mean precipitation as snow. For this study, a single year (October 2006 through September 2007) of 6-hourly input time series data (precipitation and evapotranspiration) were utilized, with observed streamflow time series data used for model parameter calibration. Precipitation (rain and snowmelt) observations were derived from the Onion Park SNOTEL station, while evapotranspiration data were obtained from collocated eddy-covariance towers positioned within TCEF [Emanuel *et al.*, 2010].

Streamflow hydrographs (observed in blue; predicted in red) for each of the three NSE transformations (none, flow-corrected time, Box-Cox) and in each time domain (clock and flow-corrected) are shown in Figure 5, highlighting the similarities and differences across each of the methods. Notably, the predicted hydrograph for the untransformed Nash-Sutcliffe objective function (middle) and the Box-Cox transformed NSE (top) were more flashy in their response to precipitation, while at the same time the FCT transformed NSE (bottom) was less sensitive to the inputs (particularly at low streamflows, as expected given flows less than the mean flow are given less weight). In

terms of peak streamflow simulations, the FCT transformed NSE was superior in simulating each of the three primary peaks during maximum streamflow in terms of both timing and magnitude. Here we saw that while each form of the NSE objective function resulted in reasonable values (optimum value equal to one), the FCT transformed NSE clearly outperformed the other two considered in this study with an NSE equal to 0.96.

In terms of the model residuals, Figure 6 (left column) reveals that none of the forms of the NSE considered here were, at least visually, advantageous for maintaining variance constancy. However, the quantile-quantile plots given in the right column of Figure 6 more clearly visualize the satisfaction of the normality assumptions made by each transformation. Additionally, the linearity of the QQ plots are quantified by the *Filliben* [1975] r-statistic, whose optimum value is equal to one. Again, all three transformations perform quite well, but in this case the Box-Cox and FCT transformed NSEs perform similarly to one another and outperformed the untransformed NSE at satisfying the assumptions of normality.

Application to a Dynamically Responsive Australian Catchment

In the second application, we investigated a dynamically responsive catchment with a much more complicated hydrograph featuring numerous peaks over the course of any given year. The Currency Creek catchment (57 km²) is located within the Murray-Darling River basin of South Australia, Australia (lat. 35°27'S, long. 138°42'E). In this test case, ten years (January 1976 through December 1985) of daily input time series data

(precipitation and evapotranspiration) were utilized, with observed streamflow time series data used for model parameter calibration.

As with the previous example, the PDM was calibrated using the Dynamically Dimensioned Search algorithm using three forms of the Nash-Sutcliffe efficiency: (1) untransformed, (2) Box-Cox transformed, and (3) flow-corrected time transformed. The results presented in terms of streamflow hydrographs (with both the clock and flow-corrected time domains are shown to simplify comparisons across the transformations) provide clear evidence of the utility of the FCT transformation. A 25% improvement in model fit (as NSE) was observed, increasing from 0.72 for the untransformed to 0.90 (Figure 7). The Box-Cox transform improved fit to a lesser degree (NSE = 0.77). Similar behavior was seen in this case study as with the previous, the FCT transform limits the flashy behavior during low flows caused by rapid response to precipitation inputs as compared to both the Box-Cox transformed and untransformed NSE model fits.

Overall the results shown in the scatterplots (Figure 8, left column) suggest the residuals may be homoscedastic (particularly the Box-Cox and FCT transformed). The homoscedasticity of the residuals is further reflected in the QQ plots and via the Filliben r statistic (Figure 8, right column). The FCT transformed model fits resulted in more linear QQ plots and a higher Filliben r value indicating that the residuals more closely represent those drawn from a standard normal distribution. In both the Box-Cox and flow-corrected time transformation cases, there was marked improvement in the linearity of the QQ plot compared to that for the untransformed case.

Discussion & Conclusions

As the use of hydrologic models for water resource management increases, the development of cost functions that appropriately reflect the characteristic form of hydrologic data is becoming ever more necessary. Despite the fact that all results are subjected to the satisfaction of assumptions made in the cost function, to date the vast majority of applications (global optimization, informal uncertainty-based, or formal uncertainty-based approaches) continue to rely on simple Gaussian assumptions that commonly go unchecked and unsatisfied.

We introduced and applied a new transformation within the context of hydrologic model calibration that utilizes the flow-corrected time transform, first introduced in transit time modeling [McGuire and McDonnell, 2006; Rodhe *et al.*, 1996]. The results of two case studies, featuring highly dissimilar annual streamflow regimes, highlight the ability of the FCT transform to de-skew positively skewed streamflow data (Figure 2, Figure 3). Consequently, model residuals were less heteroscedastic than those arising from the use of both the untransformed NSE function and the Box-Cox transformed NSE function (Figure 6, Figure 8).

While these results are promising in terms of satisfying the assumptions made under the model calibration framework, and thus its potential for characterizing uncertainty, the use of any data transform comes with caveats. The flow-corrected time transformation method is weighted toward fitting the values of streamflow that exceed the mean streamflow. In catchments that feature a dominant two-state streamflow regime (e.g., single peak hydrographs common to snow-dominated watersheds), the use of the

FCT transform would often make sense in terms of modeling the total volumetric flux of water, which is of most interest to water resource managers. The use of the flow-corrected time transform should be valued under circumstances that are interested in peak streamflow events (e.g., floods), but could be counterproductive (in its current formulation) under circumstances concerned with baseflow events (e.g., droughts). However, using inverse weights would switch the focus of the FCT transform (as formulated here) to instead place priority on low flows rather than peak flows. Lastly, note that as the hydrograph approaches a shape that is symmetric about the mean streamflow the effect of the FCT transform results in identical untransformed and transformed streamflows.

There are real limitations brought about by transforming data for real-time forecasting. Because the methodology down-weights streamflows less than the mean observed streamflow, such streamflows are less important to the model calibration. As a result, the reliability of model predictions for streamflows less than the mean streamflow diminishes with departure from the mean. Modeling is further complicated when there are auxiliary modeled variables of interest that depend on “clock time” (e.g., groundwater storage). Under this scenario, internal model variables that are simulated as the result of threshold exceedences may be negatively impacted at low streamflows.

Despite these potential limitations, the flow-corrected time transform was shown to: (1) more effectively remove positive skew from the observed streamflow record than the alternatives investigated (Figure 2); (2) better satisfy the underlying assumption of normality (following transformation) than the other transformations examined (Figure 6,

Figure 8); and (3) result in a superior model “fit” based on the Nash-Sutcliffe efficiency metric (Figure 5, Figure 7) in both examples considered.

Acknowledgements

This study was funded by the National Science Foundation (EAR-0943640). The first author received additional support as a Graduate Fellow of the Montana Institute on Ecosystems. We thank Francis Chiew and Jai Vaze for sharing the Australian catchment data used in this study.

Tables

Table 1. Description of Parameters of the Probability Distributed Model.

<i>Parameter</i>	<i>Description</i>
c_{max}	maximum soil storage capacity [mm]
b	spatial variability within the catchment [-]
kb	rate of drainage into subsurface storage [-]
cf	soil storage threshold for release of subsurface storage to streamflow [mm]
tr_1	fraction of subsurface storage released to streamflow [-]
tr_2	fraction of surface storage released to streamflow [-]

Figures

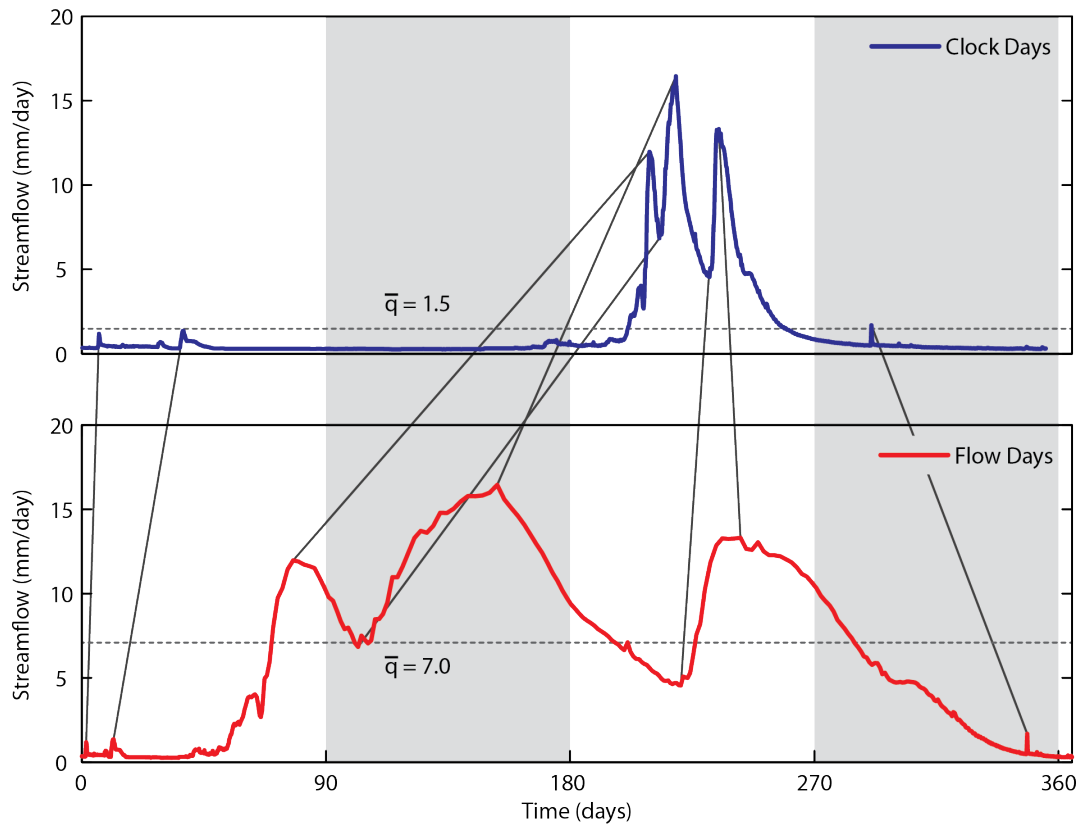


Figure 1. Illustration of the flow-corrected time transform used to account for positive skewness present in a hypothetical streamflow hydrograph. The top panel shows the untransformed hydrograph and the bottom panel shows the FCT transformed hydrograph. Lines between the two panels map streamflow events from the untransformed to the transformed hydrograph.

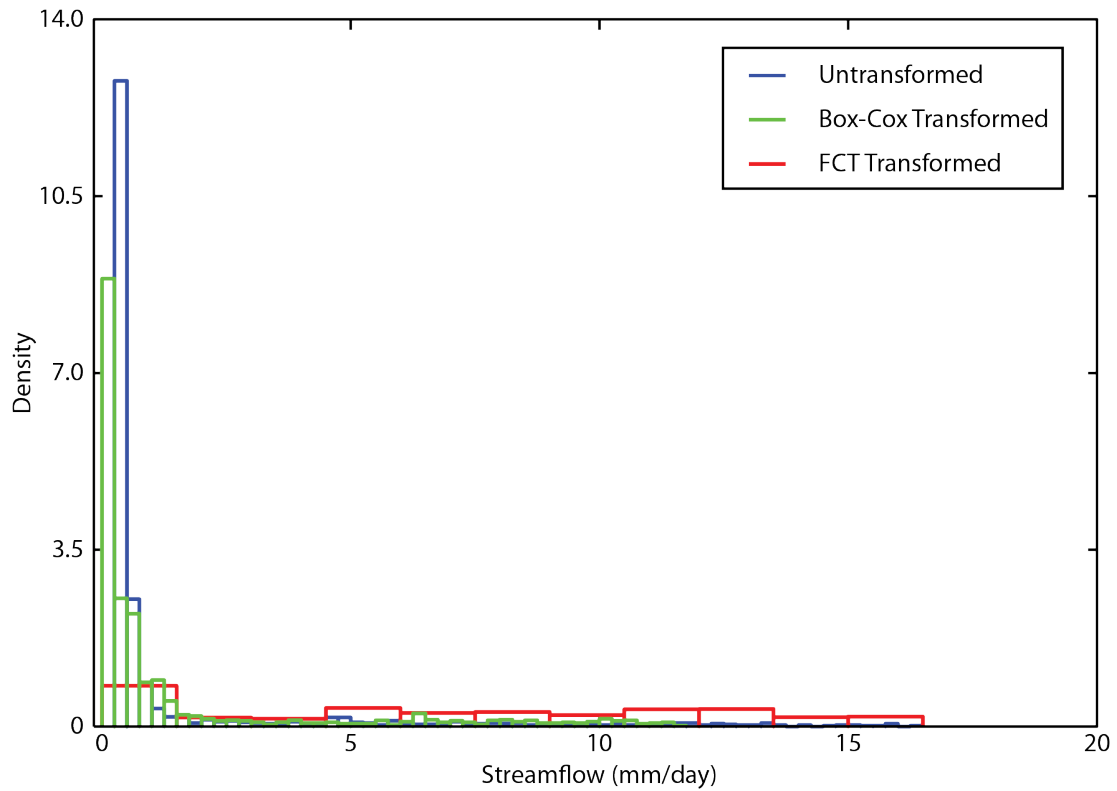


Figure 2. Histograms of the streamflow data shown in Figure 1 under the untransformed, Box-Cox transformed, and flow-corrected time transformed conditions. Note that the Box-Cox transformed data has been shifted for clarity of presentation (due to negatives occurring for flows original flows less than 1) and the flow-corrected time transformed histogram is in terms of flow time and not clock time (achieved by performing a linear interpolation of the FCT weights). The skewness present in the untransformed data and even the Box-Cox transformed data highlights the effectiveness of the FCT transformation at removing such skewness in the data (in the “flow” time domain).

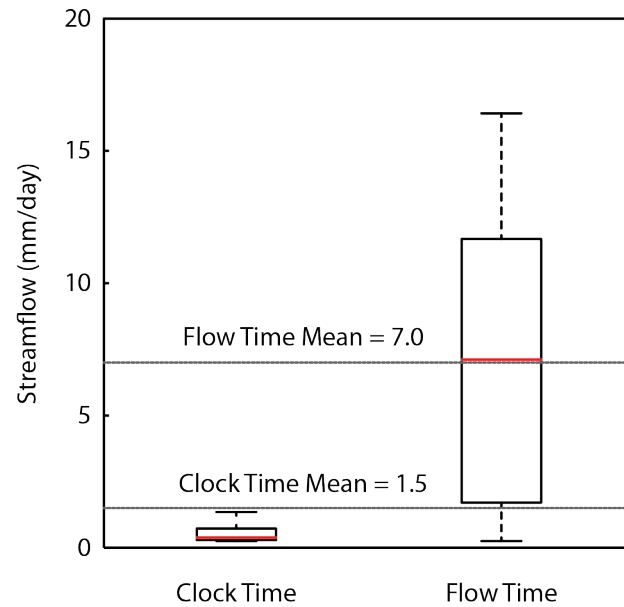


Figure 3. Box plots of the untransformed and flow-corrected time transformed streamflow distributions (for hydrograph shown in Figure 1). The skewness present in the untransformed data and the effectiveness of the FCT transform at removing such skewness in the data are highlighted, noting that the mean streamflow represents an “outlier” value for the untransformed data but falls very near to the median of the transformed data.

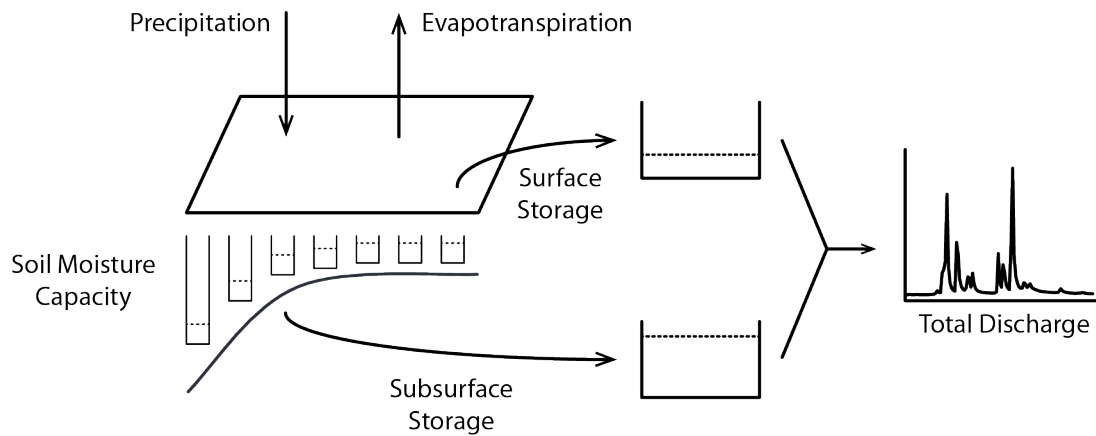


Figure 4. Schematic diagram of the probability distributed model (PDM).

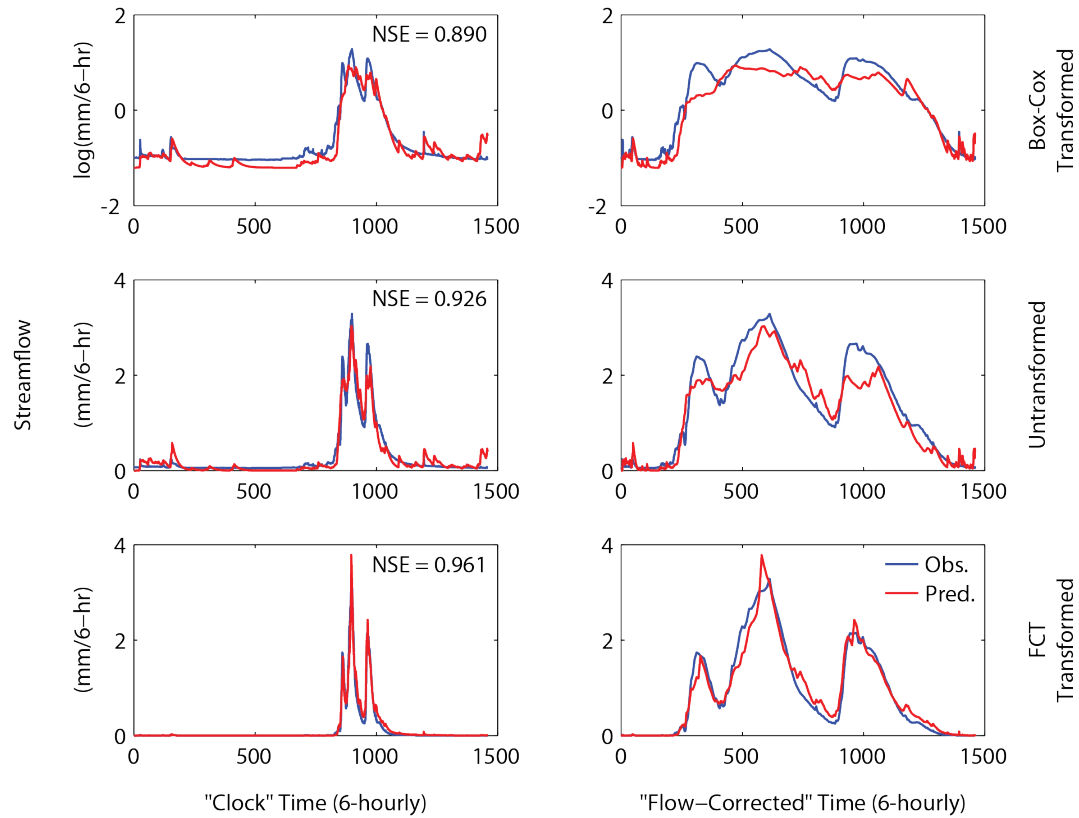


Figure 5. The Box-Cox transformed, the untransformed, and the FCT transformed streamflow hydrograph for the Stringer Creek (snow-dominated) catchment application. Results are shown for each transform in both "clock" time and "flow-corrected" time. The Nash-Sutcliffe efficiency (NSE) is also given for each case (applied to the appropriate transformed space), with an optimal value equal to one.

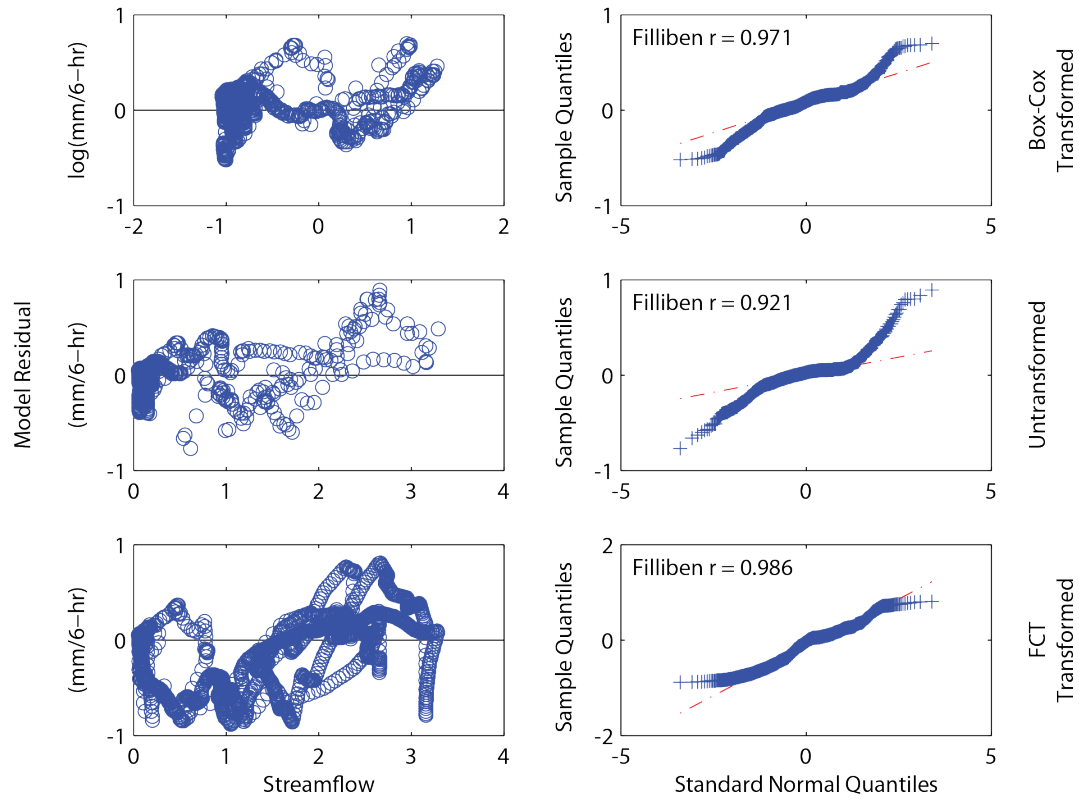


Figure 6. Residual scatterplots and quantile-quantile plots of the model residuals for the Box-Cox transformed, untransformed, and flow-corrected time transformed Stringer Creek catchment streamflows. The Filliben statistic is also given for each case, with an optimal value equal to one. Note the units on the x-axis of the scatterplots are the same as those on the y-axis.

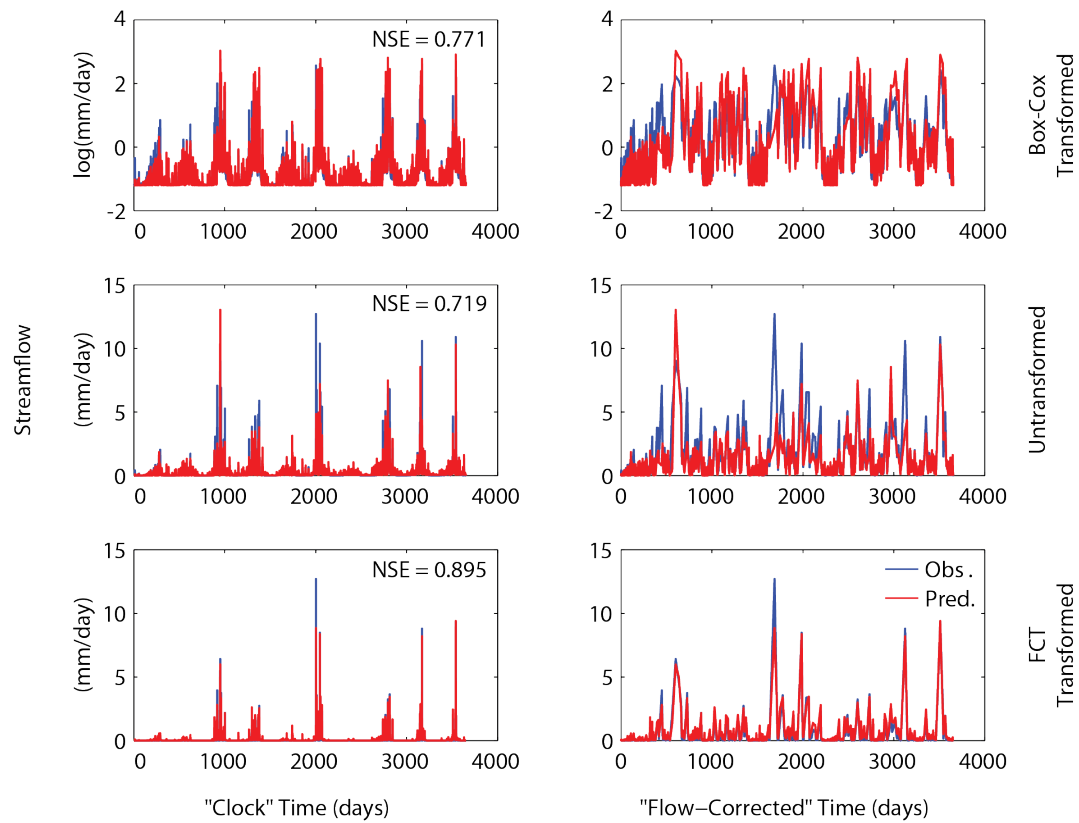


Figure 7. The Box-Cox transformed, the untransformed, and the FCT transformed streamflow hydrograph for the Currency Creek (dynamically responsive) catchment application. Results are shown for each transform in both "clock" time and "flow-corrected" time. The Nash-Sutcliffe efficiency (NSE) is also given for each case (applied to the appropriate transformed space), with an optimal value equal to one.

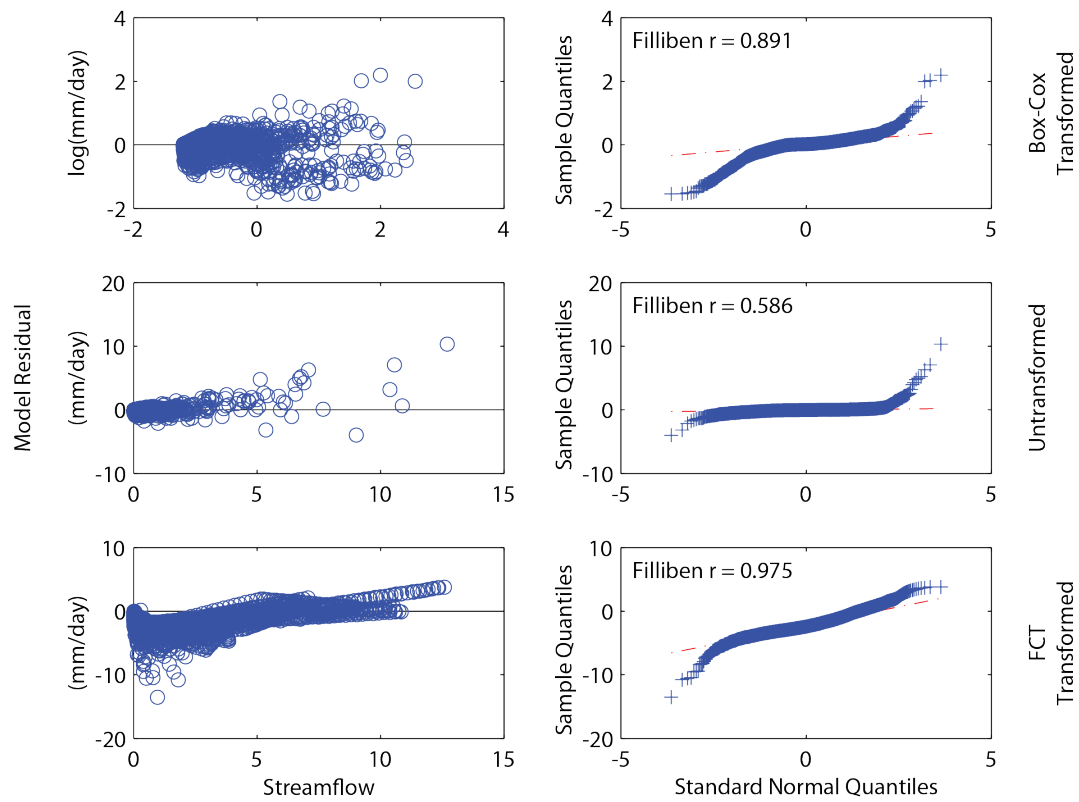


Figure 8. Residual scatterplots and quantile-quantile plots of the model residuals for the Box-Cox transformed, untransformed, and flow-corrected time transformed Currency Creek catchment streamflows. The Filliben statistic is also given for each case, with an optimal value equal to one. Note the units on the x-axis of the scatterplots are the same as those on the y-axis.

References

- Bates, B. C., and E. P. Campbell (2001), A Markov chain Monte Carlo scheme for parameter estimation and inference in conceptual rainfall-runoff modeling, *Water Resour. Res.*, 37(4), 937-947.
- Beven, K., and A. Binley (1992), The future of distributed models: Model calibration and uncertainty prediction, *Hydrol. Processes*, 6(3), 279-298.
- Box, G. E. P., and D. R. Cox (1964), An Analysis of Transformations, *J. Roy. Stat. Soc. B Meth.*, 26(2), 211-252.
- Duan, Q., S. Sorooshian, and V. Gupta (1992), Effective and efficient global optimization for conceptual rainfall-runoff models, *Water Resour. Res.*, 28(4), 1015-1031.
- Emanuel, R. E., H. E. Epstein, B. L. McGlynn, D. L. Welsch, D. J. Muth, and P. D'Odorico (2010), Spatial and temporal controls on watershed ecohydrology in the northern Rocky Mountains, *Water Resour. Res.*, 46, W11553, doi:10.1029/2009WR008890.
- Filliben, J. J. (1975), The Probability Plot Correlation Coefficient Test for Normality, *Technometrics*, 17(1), 111-117.
- Gupta, H. V., and H. Kling (2011), On typical range, sensitivity, and normalization of Mean Squared Error and Nash-Sutcliffe Efficiency type metrics, *Water Resour. Res.*, 47, W10601, doi:10.1029/2011wr010962.
- Kuczera, G., and E. Parent (1998), Monte Carlo assessment of parameter uncertainty in conceptual catchment models: the Metropolis algorithm, *J. Hydrol.*, 211(1-4), 69-85.
- Madsen, H., G. Wilson, and H. C. Ammentorp (2002), Comparison of different automated strategies for calibration of rainfall-runoff models, *J. Hydrol.*, 261(1-4), 48-59.
- McGuire, K. J., and J. J. McDonnell (2006), A review and evaluation of catchment transit time modeling, *J. Hydrol.*, 330(3-4), 543-563.
- Montanari, A. (2005), Large sample behaviors of the generalized likelihood uncertainty estimation (GLUE) in assessing the uncertainty of rainfall-runoff simulations, *Water Resour. Res.*, 41, W08406, doi:10.1029/2004WR003826.
- Moore, R. J. (1985), The probability-distributed principle and runoff production at point and basin scales, *Hydrol. Sci. J.*, 30(2), 273-297.

- Moore, R. J. (2007), The PDM rainfall-runoff model, *Hydrol. Earth Syst. Sci.*, 11(1), 483-499.
- Rodhe, A., L. Nyberg, and K. Bishop (1996), Transit Times for Water in a Small Till Catchment from a Step Shift in the Oxygen 18 Content of the Water Input, *Water Resour. Res.*, 32(12), 3497-3511.
- Sorooshian, S., and J. A. Dracup (1980), Stochastic parameter estimation procedures for hydrologic rainfall-runoff models: Correlated and heteroscedastic error cases, *Water Resour. Res.*, 16(2), 430-442.
- Stedinger, J. R., R. M. Vogel, S. U. Lee, and R. Batchelder (2008), Appraisal of the generalized likelihood uncertainty estimation (GLUE) method, *Water Resour. Res.*, 44, W00B06, doi:10.1029/2008WR006822.
- Thyer, M., G. Kuczera, and Q. J. Wang (2002), Quantifying parameter uncertainty in stochastic models using the Box-Cox transformation, *J. Hydrol.*, 265(1-4), 246-257.
- Tolson, B. A., and C. A. Shoemaker (2007), Dynamically dimensioned search algorithm for computationally efficient watershed model calibration, *Water Resour. Res.*, 43, W01413, doi:10.1029/2005WR004723.
- Wagener, T., and N. McIntyre (2005), Identification of rainfall-runoff models for operational applications, *Hydrol. Sci. J.*, 50(5), 735-751.
- Wagener, T., K. van Werkhoven, P. Reed, and Y. Tang (2009), Multiobjective sensitivity analysis to understand the information content in streamflow observations for distributed watershed modeling, *Water Resour. Res.*, 45, W02501, doi:10.1029/2008WR007347.

CHAPTER FIVE

USING FIELD DATA TO INFORM AND EVALUATE A NEW MODEL
OF CATCHMENT HYDROLOGIC CONNECTIVITYContribution of Author and Co-Authors

Manuscript in Chapter 5

Author: Tyler Smith

Contributions: Responsible for the development and analysis of model that is introduced, the majority of the writing, and figure development.

Co-Author: Lucy Marshall

Contributions: Provided extensive discussions of the development of the methods used in the study and contributed to the writing and editing of the manuscript throughout its development.

Co-Author: Brian McGlynn

Contributions: Provided extensive discussions of the development of the methods used in the study and valuable critique and editing of the manuscript.

Co-Author: Kelsey Jencso

Contributions: Provided extensive discussions of the development of the methods used in the study and valuable critique and editing of the manuscript.

Manuscript Information

Smith, T., L. Marshall, B. McGlynn, and K. Jencso. Using field data to inform and evaluate a new model of catchment hydrologic connectivity, *Water Resources Research*, submitted.

- Journal: Water Resources Research
- Status of manuscript (check one)
 - ☐ Prepared for submission to a peer-reviewed journal
 - ☒ Officially submitted to a peer-reviewed journal
 - ☐ Accepted by a peer-reviewed journal
 - ☐ Published in a peer-reviewed journal
- Publisher: American Geophysical Union
- Date of submission: November 18, 2011

USING FIELD DATA TO INFORM AND EVALUATE A NEW MODEL OF CATCHMENT HYDROLOGIC CONNECTIVITY

Abstract

We present a new hydrologic model based on the frequency distribution of hillslope landscape elements along the stream network as a basis for simulating landscape-scale hydrologic connectivity and catchment runoff. Hydrologic connectivity describes shallow water table continuity between upland and stream elements of the catchment and is important for the movement of water and solutes to streams. This concept has gained traction in physical hydrology but has received less attention in rainfall-runoff modeling. Our model is based on the empirical findings of *Jencso et al.* [2009; 2010], who found strong correlation between the duration of shallow groundwater connectivity across hillslope, riparian, and stream zones and upslope accumulated area. *Jencso et al.* [2010] theorized that this could be used to predict catchment runoff dynamics. We explored the relationship between catchment form and function by testing the extent to which hydrologic functioning (streamflow production) could be predicted by a model based on catchment topographic form (distribution of landscape elements). We applied the model to the Stringer Creek catchment of the Tenderfoot Creek Experimental Forest, located in Montana, USA. Detailed field observations collected by *Jencso et al.* [2009] were used to inform the model's underpinnings and to corroborate internal consistency of the model's simulations. The model demonstrated good agreement between the observed and predicted streamflow and connectivity duration curves. The ability of the model to simulate internal dynamics without conditioning the parameters on these data suggest that it can be confidently extrapolated to other catchments where subsurface shallow groundwater table dynamics dominate the hydrologic response.

Introduction

Hydrologic catchment modeling has evolved over the past several decades, moving from tools used exclusively to solve engineering problems [cf., *Nash and Sutcliffe*, 1970] to frameworks that address scientific questions or hypotheses [e.g., *Weiler and McDonnell*, 2004]. A plethora of models exist today that attempt to conceptualize catchment functioning and the underlying processes that drive streamflow generation. These models can be considered explicit (but constrained) hypotheses of

catchment behavior [*Andreassian et al.*, 2009] which may be tested through available catchment data. Model types span from simple soil moisture accounting models to detailed process-based models that attempt to characterize the physics of water movement through a catchment. The relative benefits and tradeoffs of different model types are well documented [e.g., *Beven*, 1989; *Klemeš*, 1983]. In developing process-based models, inferences may be made about hydrologic processes that are not observable at the temporal/spatial scale of interest. However, physically-based models suffer from issues of over-parameterization and lack of efficiency. To make predictions about future catchment states, there is an increasing desire to use simpler models that can take advantage of sophisticated optimization and uncertainty analysis algorithms. Despite this, it can be difficult (if not impossible) to relate conceptual model parameters to observable catchment characteristics thus affecting the reliability of models as they are extrapolated beyond calibration conditions.

Multiple recent studies have illustrated how the spatial distribution of channel networks and catchment characteristics may provide a template describing hydrologic and ecosystem functioning [e.g., *Jencso and McGlynn*, 2011; *Nippgen et al.*, 2011; *Riveros-Iregui et al.*, 2011]. Given the relative tradeoffs between complex physically-based models and efficient conceptual models there is an increasing desire to develop models that can characterize the catchment hydrologic template, providing efficient, process-based hypotheses about streamflow generation. These models would make explicit the relationship between catchment form (including landscape topography, vegetation patterns, and stream networks) and hydrologic functioning (including

streamflow patterns) that could be tested with data. Field data and hydrologic experiments are then critical to determine: (1) the distinguishing factors that affect hydrologic functioning that can be included in a model; (2) the relative importance of vegetation, topography, or climatic conditions in predicting hydrologic functioning so that model structures may be selected *a priori* based on expert knowledge; and (3) how this varies across different catchment types.

A lack of coordination between hydrologic experimentalists and modelers has been recognized as a hindrance to the continued advancement of hydrologic models as scientific tools [e.g., *Dunne*, 1983; *Seibert and McDonnell*, 2002]. Research in physical hydrology has continued to advance process understanding, however applied hydrology has struggled to synthesize these findings into widely applicable modeling frameworks [*McDonnell et al.*, 2007]. While numerous models are undoubtedly able to recreate observed hydrographs (following calibration) for many catchments, the simulation of internal variables is typically much less compelling [*Kirchner*, 2006]. It is precisely for these reasons that integration of experimental studies and modeling applications is so important [*Blöschl*, 2006]; the detailed information that many empirical studies contain can provide a new direction to the dated methodologies that saturate hydrologic models (both in their selection and their formulation).

Quantifying the relationship between catchment form and hydrologic functioning poses a significant challenge to the advancement of precipitation-runoff models. The issue becomes focused on how to develop models that are physically interpretable, conceptually simple, and consistent with observations of shallow water table

development and cessation. Here, we introduce a new hydrologic model that addresses this issue by drawing on collaboration between the experimentalist and the modeler. The Catchment Connectivity Model (CCM) provides a framework that emphasizes hydrologic interpretability and predictability, using landscape structure to generate simulations of catchment streamflow.

The Role of Landscape Hydrologic Connectivity in Total Catchment Response

Beginning with the benchmark studies of *Anderson and Burt* [1978] and *Beven* [1978], the role of topography as a driving mechanism of catchment discharge has been widely studied and widely supported. The topographic index [*Beven and Kirkby*, 1979] has since become a widespread means of describing the distribution of soil saturation via “partial areas” [*Betson*, 1964] or “variable source areas” [*Hewlett and Hibbert*, 1967].

Recent studies by *Jencso et al.* [2009; 2010] and *Jencso and McGlynn* [2011] explored the link between catchment structure and runoff characteristics in mountainous headwater basins. The study emphasized the predictability of hydrologic connectivity at the landscape scale and its control on total catchment response. Here, we define hydrologic connectivity as the hydrological linkages between landscape elements that can be uni-directional or bi-directional and temporally variable. Our model emphasizes the importance of hydrologic connections between hillslope-riparian-stream (HRS) zones. We define these as occurring when water table continuity develops across their interfaces and streamflow is present [*Jencso et al.*, 2009]. *Jencso et al.* [2009] concluded that it was the temporal frequency of these connections rather than the magnitude of flux at

individual HRS zones that is important for streamflow generation. We tested this connectivity frequency idea in this study. These principles underpin the CCM structure and its assumptions. A brief review of the study site, methods, and findings of *Jencso et al.* [2009] is discussed in the following sub-sections as a primer for the model description and test application.

Study Site Description

Tenderfoot Creek Experimental Forest (TCEF), located in Montana, USA (lat. 46°55'N, long. 110°52'W), served as the study location for the research of *Jencso et al.* [2009] (Figure 1). Seven gauged sub-catchments that form the headwaters of Tenderfoot Creek (22.8 km²) are located within TCEF, which drain into the Smith River (a tributary of the Missouri River). The historical mean annual temperature at TCEF is approximately 0 °C. TCEF receives the majority of its 850 mm of mean annual precipitation as snow (roughly 75%). The research site's vegetation is dependent upon the location in the landscape, with upland areas dominated by lodgepole pine and riparian zones dominated by sedges, rushes, and willows.

We focused on the Stringer Creek sub-catchment, one of TCEF's seven sub-catchments (Figure 1). The catchment area is 5.55 km²; its elevation ranges from approximately 1,990 to 2,430 meters. For additional information on available instrumentation within Stringer Creek and the greater TCEF, please refer to *Jencso et al.* [2009; 2010], *Jencso and McGlynn* [2011], and *Nippgen et al.* [2011].

Overview of Foundational Work

In their study, *Jencso et al.* [2009] sought to investigate the usefulness of HRS groundwater table connectivity in linking dominant catchment landscape components (hillslopes and riparian areas) to the channel network, considering both temporal and spatial dimensions. The relative contribution of landscape components to the stream response was quantified by the upslope accumulated area (UAA). Upslope accumulated area, defined as the amount of area draining to a particular point, is a useful topographic metric that can be derived from a digital elevation model (DEM). *Jencso et al.* [2009] divided the stream into 10m long “stream cells”, where the UAA is considered as the lateral area draining to a particular stream cell for both the left and right hand sides of the stream following the landscape analysis methods described in *Seibert and McGlynn* [2007] and *Grabs et al.* [2010].

Eighty-four recording groundwater wells across 24 hillslope-riparian-stream transects (14 on Stringer Creek, 10 on Tenderfoot Creek) were installed to assess the relationship between landscape structure and HRS groundwater connectivity. HRS connectivity was quantified based on the presence of saturation (groundwater levels above bedrock) across the hillslope to stream transition. Such HRS connections/disconnections were recorded for each transect on 30 minute intervals, focusing on the aspects of when and for how long such connections occurred rather than the specific mechanism driving the connection.

Jencso et al. [2009] found strong correlations between the duration of HRS groundwater table connectivity and the amount of upslope accumulated area draining to

each of the transects ($r^2 = 0.91$). Importantly, they also found that when this relationship was extrapolated to the entire catchment based on their topographic analysis, the shape of the “connectivity duration curve” (CDC) was highly correlated to the catchment’s flow duration curve ($r^2 = 0.95$). *Jencso et al.* [2009] observations suggest that topographically-driven lateral redistribution of water in the shallow subsurface (0.5–1.5 m soil zone above bedrock) occurs in response to inputs (rainfall and/or snowmelt) under high catchment storage states. Overland flow is limited in its overall contribution to runoff occurring in only approximately 2-5% of the Stinger Creek catchment (varying both spatially and temporally in response to input magnitude and durations) [*Jencso et al.*, 2010].

The Catchment Connectivity Model (CCM)

Based on the findings of *Jencso et al.* [2009], we developed a conceptual model to test the hypothesis that hydrologic connectivity frequency determines catchment runoff magnitude. The correlation between upslope accumulated area, hillslope-riparian-stream connectivity, and flow duration was exploited as the basis for a semi-distributed model of the stream’s response to precipitation. In our formulation, we principally aim to quantify the control that landscape structure exerts on hydrologic connectivity.

The CCM, highlighted in Figures 2 & 3, was designed to be a simple, conceptual precipitation-runoff model that maintains the ability to simulate the internal dynamics of the catchment. The model is based on a conceptualization of landscape hydrologic connectivity and its strong correlation to upslope accumulated area (hillslope size). The model functions in a semi-distributed fashion by dividing the catchment into discrete

units based on equal stream lengths. We divided the Stringer Creek catchment into a series of discrete lateral upslope units based on 10m stream cells (calculated for each side of the stream). The resulting area of each landscape unit is then equivalent to the lateral UAA calculated at the stream.

Following the empirical relationship between UAA and HRS connectivity documented by *Jencso et al.* [2009], we define a mathematical relationship to characterize differences in the duration of hydrologic connectivity among the defined landscape units. In this mathematical formulation, we assume that a hydrologic connection (shallow groundwater present across the hillslope, riparian, and stream zones) occurs when the volume of water stored in a landscape unit exceeds a threshold level that is dependent on the UAA of the landscape unit as calculated at the stream. This stream-centric perspective offers a significant departure from established landscape conceptualizations of catchment response. The storage threshold of each landscape unit is calculated as

$$S_j^* = \delta \cdot \left[1 - \left(\frac{UAA_j}{UAA_c} \right)^\kappa \right] \quad (1)$$

where S_j^* is the volumetric storage threshold for landscape unit j which must be exceeded for hydrologic connectivity to occur, UAA_j is the upslope accumulated area of landscape unit j , UAA_c is the value of upslope accumulated area at which continuous hydrologic connectivity occurs, δ is a scaling factor for catchment storage, and κ is the Pareto parameter describing the shape of the UAA-connectivity relationship. Equation 1 has the additional constraint that the landscape units must become “hydrologically active”

sequentially, with the landscape units representing larger UAAs becoming connected prior to any smaller landscape units. This simple mathematical expression is designed to capture the nonlinear relationship between storage state (antecedent wetness), UAA, and hydrologic connectivity that *Jencso et al.* [2009] emphasized in their findings.

Soil storage is conceptualized using a simple soil moisture accounting approach [e.g., *Manabe*, 1969], where the storage is increased by precipitation inputs at each time step and decreased by evapotranspiration losses. Each landscape unit has a unique storage to account for soil water content (i.e., there is one storage per landscape unit)

$$c_{i,j} = \max \begin{cases} 0 \\ c_{i-1,j} + P_i - ET_i \end{cases} \quad (2)$$

where i indexes time and j indexes landscape units, c is the water stored in the soil, P is precipitation (as rain plus snowmelt), and ET is the actual evapotranspiration. Discharge to the stream occurs once the storage reaches a threshold level. Such an approach to soil water storage accounting is common in many conceptual hydrologic models [see *Wagener et al.*, 2004]. Precipitation and evapotranspiration are assumed to be uniform across the catchment. Emphasizing the concept of connectivity further, all of the landscape units are defined as having the same outflow rate (when connected), regardless of hillslope size, with the outflow volume being a function of the duration of hydrologic connection across the hillslope-riparian-stream gradient such that

$$GW_{i,j} = \min \begin{cases} c_{i,j} \\ q^* \cdot \omega_{f_j} \end{cases} \quad (3a)$$

$$GW_{T_i} = \sum_{j=1}^m GW_{i,j} \quad (3b)$$

where $GW_{i,j}$ is the outflow from landscape unit j at time i , $c_{i,j}$ is the current water stored in the landscape unit, q^* is the parameter controlling the rate of discharge from the landscape unit, and ω_{f_j} is the frequency weighting of the landscape unit (i.e., the number of stream cells represented by the landscape unit divided by the total number of stream cells). The outflow from the landscape units are constrained such that more water cannot be released than what is stored. The total outflow from all of the landscape units for time i , GW_{T_i} , is then the sum of all the individual outflows, $GW_{i,j}$, for all the landscape units.

Based on the mathematical formulation of the perceptual functioning of Stringer Creek, landscape units with larger upslope accumulated areas have greater volumes of water stored, are more likely to become hydrologically connected to the stream, will have longer connectivity duration, and thus will contribute a larger portion of the volumetric discharge (even considering the assumption of equal discharge rates across the entire stream network). Conversely, units with smaller upslope accumulated areas will be much more temporally dynamic, storing less water, and achieving hydrologic connectivity intermittently. At peak discharge, the smaller landscape units are more likely to be connected, resulting in a larger proportion of the catchment contributing to streamflow generation. In this way, we hypothesize that it is the frequency of hydrologic connections to the stream that drive the aggregate catchment response, rather than the magnitude of flux at any one connection. We apply a simple exponential filter to represent catchment routing of these connections [Weiler *et al.*, 2003].

While the conceptual model is rooted in the perceived physical processes driving flow at Stringer Creek, the model structure itself maintains a relatively parsimonious sensibility in regard to its number of parameters (six) and modest data requirements. Table 1 describes each of the parameters that must be estimated by calibration. The only inputs needed to simulate discharge with the CCM are evapotranspiration, precipitation, and a DEM of the study catchment (from which UAA is computed). This simple, flexible model structure represents an alternative to the presently available conceptual precipitation-runoff models in that it has the potential to simulate internal catchment dynamics while maintaining the potential for efficient streamflow predictions. The following section introduces a test case for which the CCM was applied.

Test Case: Stringer Creek Catchment

The Stringer Creek catchment that was part of the *Jencso et al.* [2009] study on the influence of landscape structure on runoff dynamics serves as the initial application site for the CCM. The selection of Stringer Creek catchment was motivated by the availability of detailed hydrologic connectivity data that will allow the internal dynamics of the model's simulations to be directly used to assess the internal model performance. The following sub-sections describe the data used to drive the model, the calibration approach, and the results.

Data

Time series data of precipitation and evapotranspiration were used as inputs to the model. The study used data on a 6-hour time step for the period of study extending from

October 2005 to September 2008. Precipitation and snowmelt observations were derived from the Onion Park (2,258m) and Stringer Creek (1,996m) SNOTEL stations.

Evapotranspiration data was collected from collocated eddy-covariance towers within TCEF [Emanuel *et al.*, 2010]. Because Stringer Creek receives the majority of its precipitation as snowfall, an initial step of calculating the amount and timing of snowmelt events was performed. Snow water equivalent and cumulative precipitation data were used to infer actual snowmelt and rainfall amounts using a simple logic-based algorithm. While it would be equally possible to supplement the CCM with a snowmelt accounting model, this was not done here in order to remove any parameter interdependencies that could arise between the snowmelt model and the CCM. Measured stream discharge data from the flume at the outlet of Stringer Creek was available for model calibration and internal model validation exercises over the same time period. Additionally, all topographic data was obtained from a 10m grid digital elevation model (DEM) derived from 1m LiDAR bare earth topography data.

The model was calibrated for the spring runoff period (April through September) for water years 2007 and 2008 (WY2007, WY2008), following a model warm-up phase for WY2006. The warm-up phase allowed the initial unknown states of the soil storages to equilibrate. These years were selected for analysis due to the availability of hillslope-riparian-stream connectivity data, which allowed for a direct comparison of internal dynamics (connectivity) during the calibration period.

To discretize the catchment into separate landscape elements, the stream was divided into equal 10m reaches and the lateral upslope accumulated area for each side of

the reach was calculated. To improve model efficiency these units were then aggregated to 50 separate storages. The selection of 50 storages was made following a simple sensitivity analysis of model performance with respect to the number of storages, which were varied from 30 to 100. The individual 10m reaches were binned into storages with equal frequency, resulting in each storage having the same number of “member reaches” but with unequal “widths”. The binning method was selected for both its conceptual simplicity and conceptual resemblance to the underlying model structure.

Model Calibration

The CCM’s parameters were calibrated via Bayesian inference. The Bayesian approach is an attractive option because it combines parameter estimation and uncertainty analysis into a single process, where the parameter estimates are represented by probability distributions (called posterior distributions) rather than point estimates. Because of nonlinearities common in hydrologic models, analytic solutions of the posterior distributions are not possible. Markov chain Monte Carlo (MCMC) methods are a common way of numerically solving for the posterior distributions, and have gained attention for their increasing use in hydrology [e.g., *Kuczera and Parent, 1998; Marshall et al., 2004; Smith and Marshall, 2008*].

For this study, the Adaptive Metropolis (AM) algorithm [*Haario et al., 2001*] was selected as the MCMC method for implementing the Bayesian approach. This algorithm has shown to perform well in hydrologic problems [*Marshall et al., 2004*], and contains a logic that is very simple to implement. Bayesian methods require the definition of a formal likelihood function, and as such, make explicit the assumptions being made about

the form of the model residuals. Due to this, the selection of an appropriate likelihood function is of great importance.

In hydrology, this can be difficult because of the often positively skewed distribution of streamflows that can lead to heteroscedastic residuals [*Sorooshian and Dracup*, 1980]. To address this complication here, we incorporated a one parameter Box-Cox transformation [*Box and Cox*, 1964] on stream discharge into the likelihood function. This type of likelihood has been successfully implemented for use in a hydrologic applications [e.g., *Bates and Campbell*, 2001]. The Box-Cox transformation parameter was specified as an additional calibration parameter.

Model Results

Model Parameterization via Bayesian Inference: The posterior model parameter values are given in Table 2. To illustrate the degree of parameter uncertainty we report the median and upper/lower 10th percentile values. In addition, box plots of the parameter's posterior distributions are given in Figure 4. In general, each of the parameters is well defined with symmetric, single modes and has a clear region of high posterior density.

Model Simulations: Figure 5 shows the observed and simulated streamflow for the catchment. Note that the model calibration discards the first year of simulation as a model "warm-up" phase to account for unknown storage values at the start of the simulation, and only uses the spring melt-driven hydrograph (April-September) of the subsequent years for parameter estimation. Figure 5 additionally includes a binary time

series plot of simulated hydrologic connectivity for each landscape unit versus the UAA for each landscape unit. Because of the effect of the threshold function used to define hydrologic connectivity occurrence, larger landscape units are more likely to be hydrologically connected to the stream. The largest landscape units stay connected to the stream throughout the simulation period, reflecting their role in simulating baseflow conditions. Smaller landscape units are connected more intermittently and cumulatively contribute to peak spring runoff. Again, it is the frequency of connections to the stream that provides the variability in stream response, rather than the magnitude of any one connection. Under the main assumptions of the model, each hydrologic connection contributes to the stream discharge at equivalent flow rates.

Model Evaluation and Consistency with Field Observations: The model simulation allows assessment of the storage state of each landscape unit over time. Figure 6 shows the storage volume for each of the 50 landscape units, from which broad patterns of storage flux and variability can be inferred. From this figure, note that larger storages are generally associated with landscape units of larger UAA. The larger landscape units store more water (on a volumetric basis) so that it can be assumed that increasing storage volumes correspond to units with increasing UAA. Note that these larger landscape units also show greater variation and flux over time as they are more easily connected to the stream.

The model also allows assessment of the aggregate catchment connectivity duration over time. Figure 7 shows the connectivity duration curve for the catchment along with that estimated from the model simulation (derived from the binary

connectivities displayed in Figure 5). These connectivity curves can be used as an assessment of the internal model consistency with real catchment behavior beyond agreement with just streamflow.

Discussion of Results

We sought to determine the extent to which hydrologic connectivity and streamflow generation could be modeled via interpretation of landscape structure alone. Based on previous detailed field investigations [*Jencso et al.*, 2009] we hypothesized that streamflow production could be predicted by simulating hydrologic connectivity across hillslope, riparian, and stream zones (HRS) along a stream network. Patterns of hydrologic connectivity were assumed to be related to the distribution of a topographic index along the stream network, namely the distribution of lateral upslope accumulated area (UAA). We suggest that the Catchment Connectivity Model is as a first step toward developing a template of hydrologic behavior. We sought to examine the predictive power and internal consistency of the model when applied with real data. In this way, we aimed to test the perceptual model conceived by *Jencso et al.* [2009].

How Appropriate is the CCM from a Predictive Perspective?

While physically-based model parameters may be estimated (theoretically) without optimization the model structure described here represents a hybrid conceptual/empirical model, where model functioning is physically interpretable but specific parameters could not be estimated *a priori*. The model parameters were

estimated via Bayesian inference, and the predictive performance of the model was then inferred by examining the model simulations under uncertainty. The model simulation generally follows observed streamflow patterns, with the predictions capturing peak and recession flows indicating the general underlying model assumptions cannot be rejected. The model simulation had a coefficient of efficiency of 0.88 [*Nash and Sutcliffe*, 1970]. The general agreement between the observed and predicted stream discharge indicates the potential of this model to be more convincingly extrapolated to other hydrologic conditions and tested in additional catchments of varying topographic structure.

It must be noted that the model makes relatively strict assumptions about catchment functioning that limit the range of possible simulations. In light of this, the model performance is particularly promising. The model structure assumes there is no overland flow as it is observed only for a very small portion of the riparian zone ($\approx 2\text{-}5\%$) [*Jencso et al.*, 2010]. Because of this, an overland flow type model is not appropriate for the study catchment. The dominant hydrologic mechanism assumed in the model is shallow subsurface flow, as conceptualized via water table development between the stream and the uplands. Discharge variability is then limited according to the modeled hydrologic connectivity between the hillslope and the stream. The discharge rate to the stream is assumed constant along the stream network regardless of the contributing area of individual landscape units. This is a significant departure from typical storage models where storage size dictates the rate of stream contribution [e.g. *Boughton*, 2004].

How Appropriate is the CCM in Representing Internal Catchment Functioning?

The model structure provides an opportunity to make process-based interpretations of the model simulations and to relate different elements of the model to field observations. The patterns of hydrologic connectivity simulated by the model (Figure 5; Figure 7) agree with those observed by *Jencso et al.* [2009], as landscape units with the largest UAAs simulate baseflow conditions and landscape units with smaller UAAs represent spring runoff conditions as they progressively become hydrologically connected. This pattern arises out of the assumptions made by the storage threshold switching function (Eq. 1).

The simulated hydrologic connectivity curve was highly correlated to the observed field connectivity (Figure 7; coefficient of efficiency of 0.75), but showed overall less connectivity than was observed. It must be noted here that the observed hydrologic connectivity at Stringer Creek was not used to condition the model parameters, so this may be considered an unbiased check on the model assumptions and the appropriateness of the model structure. The disagreement between the observed and modeled hydrologic connectivity is hypothesized to be due to several potential factors. The DEM used for model discretization provides few landscape units with large UAA values and hence allows little flexibility in modeling baseflow conditions. Additionally, *Jencso and McGlynn* [2011] suggested that deeper groundwater streamflow sources become increasingly important at low flows. In turn, this affects the overall connectivity of the catchment over the entire simulation period. The disagreement could also be highlighting potential errors in the model structure and assumptions. The form of the

storage threshold function (Eq. 1) assumes that patterns of connectivity are dependent on storage alone and does not take into account other factors that may vary spatially or temporally including vegetation, slope, rates of precipitation, potential bedrock permeability, or evapotranspiration. The connectivity curve could also be illustrating issues with the model-driving data, as the observed precipitation (assumed to be uniform across the whole of the catchment) may not be fully representative of the true conditions.

Can Templates of Hydrologic Behavior be Developed Through Analysis of Landscape Structure Alone?

In our test case, we examined the extent to which catchment form represents hydrologic function. The CCM is based on the distribution of landscape elements along the stream as a template on which patterns of streamflow generation can be inferred. We expect that this is a first step upon which additional catchment characteristics that affect hydrologic processes can be added. Intensive and extensive field data and assessment of modeled streamflow and hydrologic connectivity can be used to diagnose model structural errors and make hypotheses about additional catchment factors that should be included.

We note that the model simulation generally underestimates peak streamflow values while providing good agreement with the wet up and recession periods of the hydrograph for each year of simulation (Figure 5). This could indicate that the assumption of equal flow rates from landscape units is not appropriate for the transient landscape units with small UAA. Alternatively, while only 2-5% of the catchment has

observed overland flow [Jencso *et al.*, 2010], these areas could provide a source of runoff important under peak flow conditions.

How Transferrable is the CCM Structure to Other Catchments?

The CCM makes strict assumptions about the importance of hydrologic connectivity in total catchment functioning. The results presented here provide initial testing of the model as they are applied to a single catchment over only a few years. However, the general agreement between streamflow and hydrologic connectivity simulations and observed data provide encouragement that the model represents overall hydrologic functioning and that landscape structure is an important control on streamflow generation at Stringer Creek.

We do not advocate, based on this limited case study, that "one size fits all". Rather, we suggest that by relating hydrologic functioning to catchment characteristics we can improve understanding of how catchment characteristics (including topography, vegetation, stream network geometry, soil structure, etc.) influence or reflect dominant hydrologic processes. Ideally, this can support *a priori* model selection to test hypotheses about streamflow patterns. The CCM described here has modest data requirements and unknown parameters that are generally equivalent to those required of simple soil storage accounting conceptual models and is based on a perceptual model that may be hypothesized for many other catchments with steep hillslopes and relatively uniform soils. We believe that this model formalizes new conceptualizations of the importance of hydrologic connectivity and should be tested across a range of catchments and climate

forcing to determine the limits of its applicability. Further work is focused on exploring the additional influence of vegetation patterns, the spatial variability of soil characteristics, and geology on runoff generation across flow states. We additionally aim to test the applicability of the model across catchments with varying landscape structure.

Conclusions

Our new model represents a synthesis of complex process understanding into a simple model structure that allows for verification of internal hydrologic consistency. This integration is critical to the development of mathematical model conceptualizations that are able to predict both internal and catchment outlet hydrologic dynamics. It is well recognized that models that are consistent with observed hydrologic processes are more flexible and have greater utility than those relying on the structural freedom of a model to curve-fit a single variable like streamflow [*Wagener and Gupta, 2005*].

The CCM was developed based on experimental observations by *Jencso et al.* [2009; 2010] who found strong correlations between UAA, hydrologic connectivity, and catchment flow duration curves. Incorporation of these findings into a runoff model was achieved by conceptualizing catchment runoff generation as a function of landscape hydrologic connectivity to the stream as approximated by the distribution of UAA across the stream network. The catchment was discretized into a series of landscape units according to the distribution of UAA along the stream network. Hydrologic connectivity was modeled for each landscape unit by a storage threshold value calculated by relating the UAA of the landscape unit to its water balance. Total discharge from each landscape

unit was assumed to be proportional to the duration that it was hydrologically connected to the stream; a significant departure from traditional conceptual hydrologic models. The detailed spatial and temporal observations at TCEF provided a unique opportunity to diagnose the model using hydrologic connectivity data (groundwater table continuity across the HRS continuum), while maintaining a conceptual simplicity that utilized only streamflow data for model conditioning. Overall, the model showed good agreement with observed data, including internal catchment dynamics. Future work will test the limits of this model's applicability across a range of catchment forms and climate variability.

Acknowledgements

This study was made possible by NSF grant EAR-0337650 and EAR-0943640. The authors thank the USDA Forest Service, Rocky Mountain Research Station, especially Robert E. Keane and Elaine Sutherland of the USDA Forest Service for site access and logistical support. Airborne Laser Mapping was provided by the National Science Foundation supported Center for Airborne Laser Mapping (NCALM) at the University of California, Berkeley.

Tables

Table 1. Names and descriptions of the CCM parameters and formal likelihood function parameters requiring calibration.

Parameter	Description
q^*	the rate of discharge from each landscape unit [volume/time]
τ_r	the residence time parameter of an exponential filter [time]
δ	a scaling parameter for catchment storage [depth]
κ	the Pareto parameter describing the shape of the UAA-connectivity relationship [-]
UAA_c	the value of upslope accumulated area above which continuous connectivity occurs [area]
γ	a data bias scaling factor [-]
λ	the Box-Cox parameter of the likelihood function used in model calibration [volume/time]
σ^2	the variance parameter of the likelihood function used in model calibration [volume/time]

Table 2. The 10th, median, and 90th percentile values for the calibrated parameters (CCM and likelihood function).

Parameter		10th	Median	90th
q^*	[m ³ /6-hr]	2.962 x 10 ⁴	2.963 x 10 ⁴	2.964 x 10 ⁴
τ_r	[6-hr]	15.11	15.39	16.10
κ	[-]	0.5126	0.5213	0.5275
δ	[m]	0.0980	0.0986	0.0991
UAA_c	[m ²]	5.060 x 10 ⁴	5.089 x 10 ⁴	5.097 x 10 ⁴
γ	[-]	0.8867	0.8876	0.8883
λ	[m ³ /6-hr]	0.71	4.64	13.75
σ^2	[m ³ /6-hr]	0.199	0.210	0.222

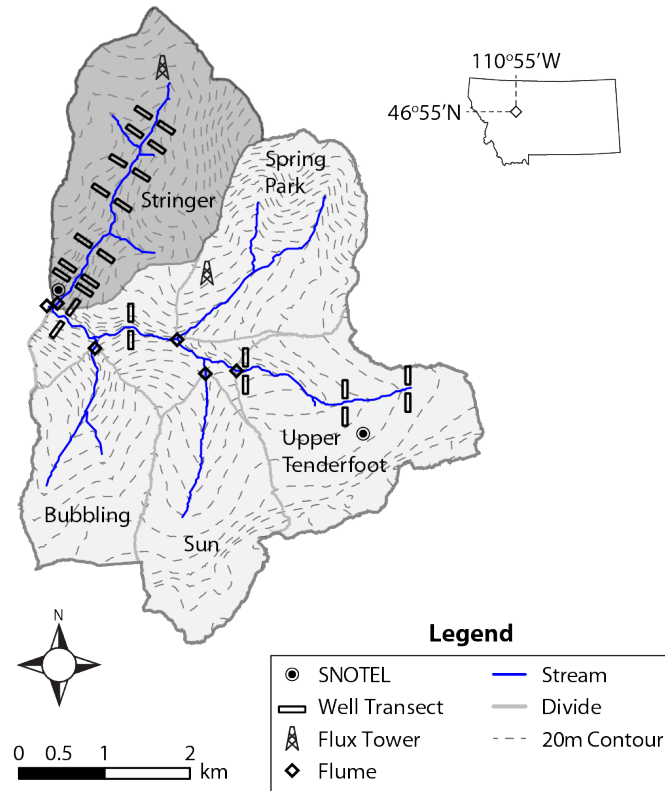
Figures

Figure 1. The Tenderfoot Creek Experimental Forest, with Stringer Creek catchment (used in modeling study) highlighted. Model input data was obtained from two sources: (1) the precipitation inputs (as rain plus snowmelt) were derived from the two SNOTEL stations located in TCEF (Onion Park and Stringer Creek) and (2) the evapotranspiration data was collected from the collocated eddy-covariance towers. The well transects measure shallow groundwater connectivity across the hillslope-riparian-stream continuum (on each side of the stream).

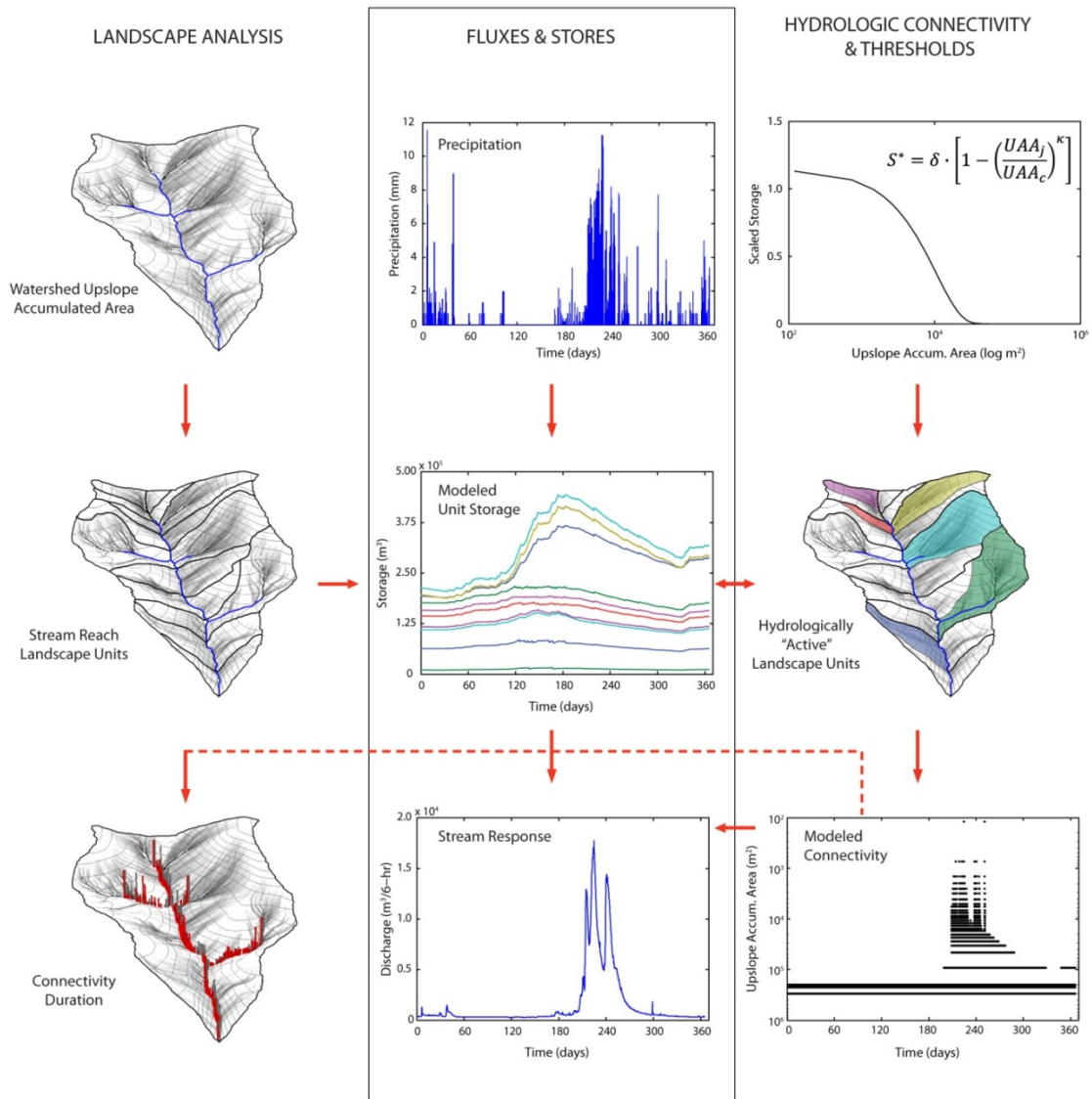


Figure 2. The conceptual form of the Catchment Connectivity Model as defined by the interactions of the drivers of the model structure: landscape analysis, fluxes and stores, and hydrologic connectivity and thresholds.

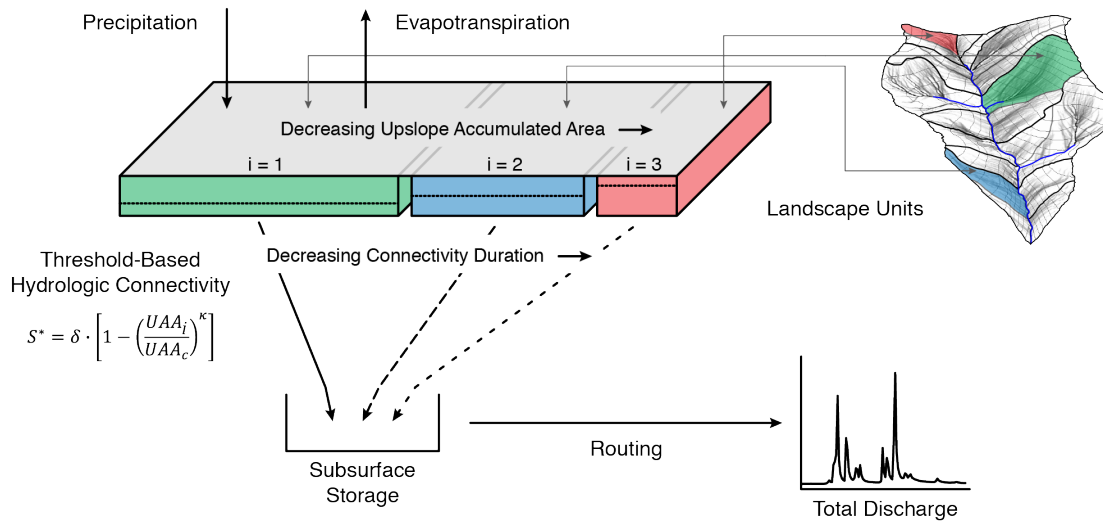


Figure 3. A graphic representation of the Catchment Connectivity Model structure that highlights the relationship between the size of landscape units (as upslope accumulated area) and threshold function mediated hydrologic connectivity.

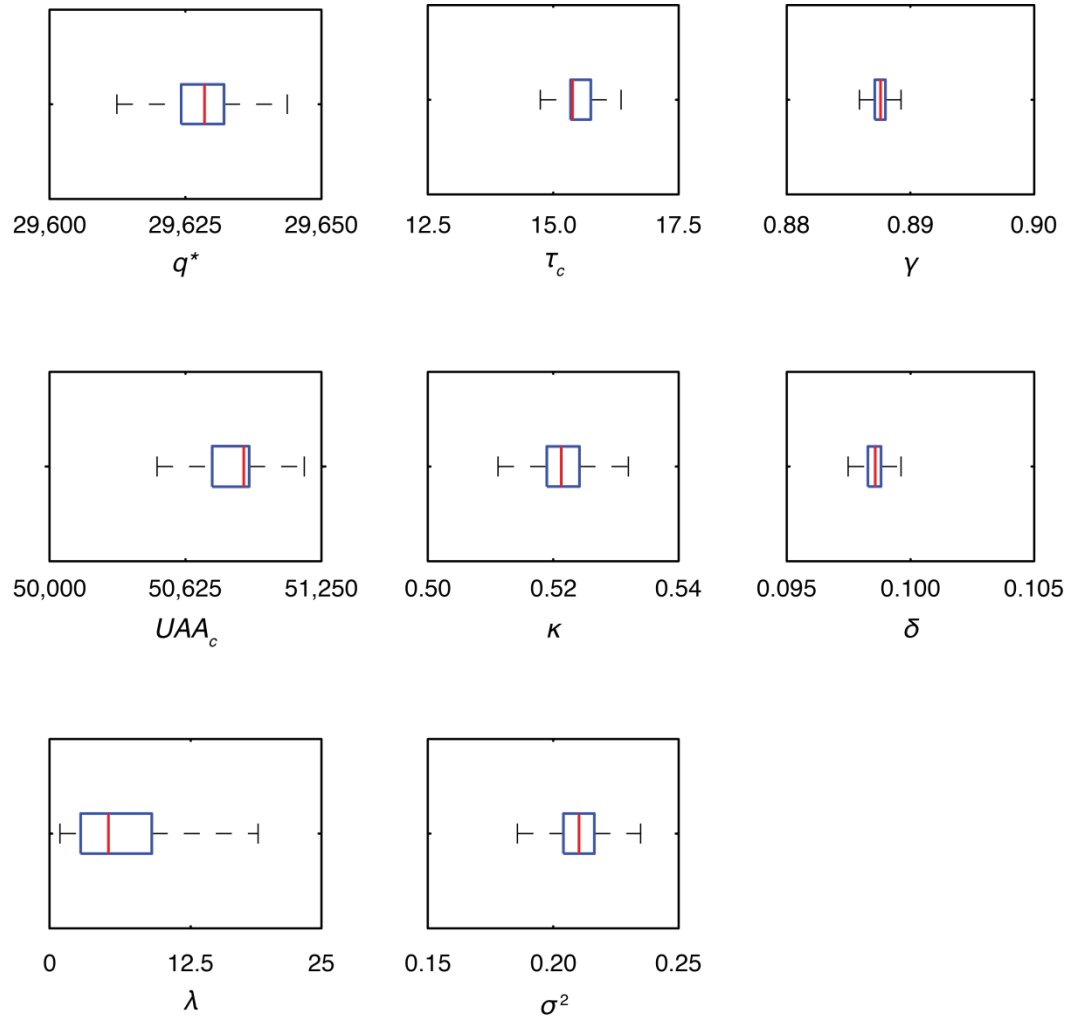


Figure 4. The marginal parameter posterior distributions (shown as boxplots) of the CCM and likelihood function. The posteriors for each parameter shown are generally well defined.

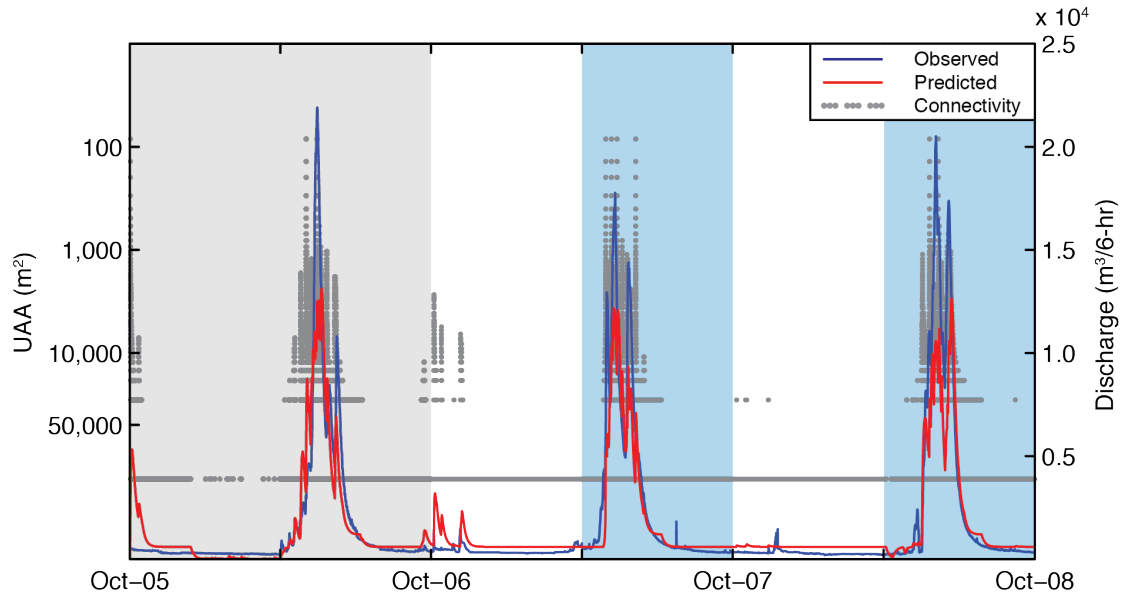


Figure 5. The observed and predicted streamflow hydrograph is shown, along with the simulated hydrologic connectivity for the warm-up year and the two subsequent analysis years. An active landscape unit (binary) connection to the stream is indicated by the presence of a gray circle marker. The blue shaded regions highlight the periods of data used for parameter estimation.

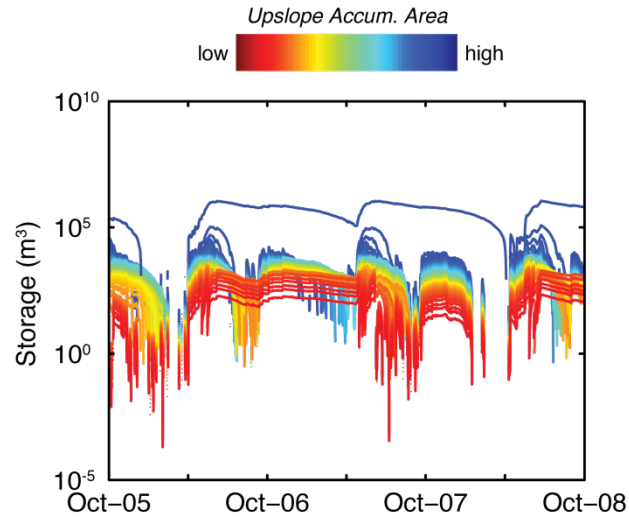


Figure 6. The storage content dynamics (simulated by CCM) over time for each of the 50 landscape unit storages. Note the log scale on the y-axis.

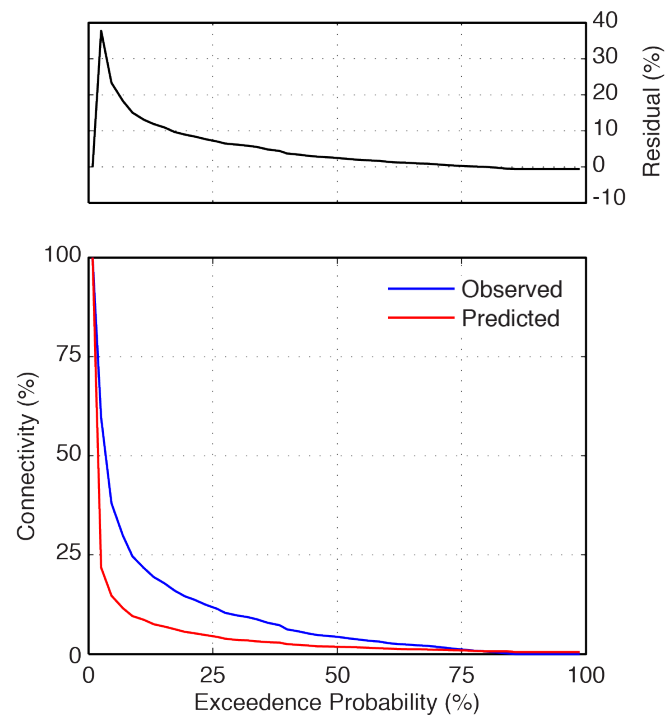


Figure 7. The predicted and observed connectivity duration curves (bottom) and the residual error (top) for the study analysis period. The plot indicates slight underestimation of connectivity for the model's predictions, most notable at intermediate discharge.

References

- Anderson, M. G., and T. P. Burt (1978), The role of topography in controlling throughflow generation, *Earth Surface Processes*, 3(4), 331-344.
- Andreassian, V., C. Perrin, L. Berthet, N. Le Moine, J. Lerat, C. Loumagne, L. Oudin, T. Mathevet, M. H. Ramos, and A. Valéry (2009), HESS Opinions "Crash tests for a standardized evaluation of hydrological models", *Hydrol. Earth Syst. Sci. Discuss.*, 6(3), 3669-3685.
- Bates, B. C., and E. P. Campbell (2001), A Markov chain Monte Carlo scheme for parameter estimation and inference in conceptual rainfall-runoff modeling, *Water Resour. Res.*, 37(4), 937-947.
- Betson, R. P. (1964), What Is Watershed Runoff?, *J. Geophys. Res.*, 69(8), 1541-1552.
- Beven, K. (1978), The hydrological response of headwater and sideslope areas, *Hydrol. Sci. Bull.*, 23(4), 419-437.
- Beven, K. (1989), Changing ideas in hydrology - The case of physically-based models, *J. Hydrol.*, 105(1-2), 157-172.
- Beven, K. J., and M. J. Kirkby (1979), A physically based, variable contributing area model of basin hydrology, *Hydrol. Sci. Bull.*, 24(1), 43-69.
- Blöschl, G. (2006), Hydrologic synthesis: Across processes, places, and scales, *Water Resour. Res.*, 42, W03S02, doi:10.1029/2005WR004319.
- Boughton, W. (2004), The Australian water balance model, *Environ. Modell. Software*, 19(10), 943-956.
- Box, G. E. P., and D. R. Cox (1964), An Analysis of Transformations, *J. Roy. Stat. Soc. B Meth.*, 26(2), 211-252.
- Dunne, T. (1983), Relation of field studies and modeling in the prediction of storm runoff, *J. Hydrol.*, 65(1-3), 25-48.
- Emanuel, R. E., H. E. Epstein, B. L. McGlynn, D. L. Welsch, D. J. Muth, and P. D'Odorico (2010), Spatial and temporal controls on watershed ecohydrology in the northern Rocky Mountains, *Water Resour. Res.*, 46, W11553, doi:10.1029/2009WR008890.
- Grabs, T. J., K. G. Jencso, B. L. McGlynn, and J. Seibert (2010), Calculating terrain indices along streams: A new method for separating stream sides, *Water Resour. Res.*, 46, W12536, doi:10.1029/2010WR009296.

- Haario, H., E. Saksman, and J. Tamminen (2001), An adaptive Metropolis algorithm, *Bernoulli*, 7(2), 223-242.
- Hewlett, J. D., and A. R. Hibbert (1967), Factors affecting the response of small watersheds to precipitation in humid areas, in *International Symposium on Forest Hydrology*, edited by W. E. Sopper and H. W. Lull, pp. 275-290, Pergamon, Oxford, UK.
- Jencso, K. G., B. L. McGlynn, M. N. Gooseff, S. M. Wondzell, K. E. Bencala, and L. A. Marshall (2009), Hydrologic connectivity between landscapes and streams: Transferring reach- and plot-scale understanding to the catchment scale, *Water Resour. Res.*, 45, W04428, doi:10.1029/2008WR007225.
- Jencso, K. G., B. L. McGlynn, M. N. Gooseff, K. E. Bencala, and S. M. Wondzell (2010), Hillslope hydrologic connectivity controls riparian groundwater turnover: Implications of catchment structure for riparian buffering and stream water sources, *Water Resour. Res.*, 46, W10524, doi:10.1029/2009WR008818.
- Jencso, K. G., and B. L. McGlynn (2011), Hierarchical controls on runoff generation: Topographically driven hydrologic connectivity, geology, and vegetation, *Water Resour. Res.*, doi:10.1029/2011WR010666, in press.
- Kirchner, J. W. (2006), Getting the right answers for the right reasons: Linking measurements, analyses, and models to advance the science of hydrology, *Water Resour. Res.*, 42, W03S04, doi:10.1029/2005WR004362.
- Klemeš, V. (1983), Conceptualization and scale in hydrology, *J. Hydrol.*, 65(1-3), 1-23.
- Kuczera, G., and E. Parent (1998), Monte Carlo assessment of parameter uncertainty in conceptual catchment models: the Metropolis algorithm, *J. Hydrol.*, 211(1-4), 69-85.
- Manabe, S. (1969), Climate and the Ocean Circulation: 1. The Atmospheric Circulation and the Hydrology of the Earth's Surface, *Mon. Weather Rev.*, 97(11), 739-774.
- Marshall, L., D. Nott, and A. Sharma (2004), A comparative study of Markov chain Monte Carlo methods for conceptual rainfall-runoff modeling, *Water Resour. Res.*, 40, W02501, doi:10.1029/2003WR002378.
- McDonnell, J. J., M. Sivapalan, K. Vaché, S. Dunn, G. Grant, R. Haggerty, C. Hinz, R. Hooper, J. Kirchner, M. L. Roderick, J. Selker, and M. Weiler (2007), Moving beyond heterogeneity and process complexity: A new vision for watershed hydrology, *Water Resour. Res.*, 43, W07301, doi:10.1029/2006WR005467.
- Nash, J. E., and J. V. Sutcliffe (1970), River flow forecasting through conceptual models part I - A discussion of principles, *J. Hydrol.*, 10(3), 282-290.

- Nippgen, F., B. L. McGlynn, L. A. Marshall, and R. E. Emanuel (2011), Landscape structure and climate influences on hydrologic response, *Water Resour. Res.*, doi:10.1029/2011WR011161, in press.
- Riveros-Iregui, D. A., B. L. McGlynn, L. A. Marshall, D. L. Welsch, R. E. Emanuel, and H. E. Epstein (2011), A watershed-scale assessment of a process soil CO₂ production and efflux model, *Water Resour. Res.*, 47, W00J04, doi:10.1029/2010wr009941.
- Seibert, J., and J. J. McDonnell (2002), On the dialog between experimentalist and modeler in catchment hydrology: Use of soft data for multicriteria model calibration, *Water Resour. Res.*, 38, 1241, doi:10.1029/2001WR000978.
- Seibert, J., and B. L. McGlynn (2007), A new triangular multiple flow direction algorithm for computing upslope areas from gridded digital elevation models, *Water Resour. Res.*, 43, W04501, doi:10.1029/2006WR005128.
- Smith, T. J., and L. A. Marshall (2008), Bayesian methods in hydrology: A study of recent advancements in Markov chain Monte Carlo techniques, *Water Resour. Res.*, 44, W00B05, doi:10.1029/2007WR006705.
- Sorooshian, S., and J. A. Dracup (1980), Stochastic parameter estimation procedures for hydrologic rainfall-runoff models: Correlated and heteroscedastic error cases, *Water Resour. Res.*, 16(2), 430-442.
- Wagener, T., H. S. Wheater, and H. V. Gupta (2004), *Rainfall-Runoff Modeling in Gauged and Ungauged Catchments*, Imperial College Press, London.
- Wagener, T., and H. V. Gupta (2005), Model identification for hydrological forecasting under uncertainty, *Stochast. Environ. Res. Risk Assess.*, 19(6), 378-387.
- Weiler, M., B. L. McGlynn, K. J. McGuire, and J. J. McDonnell (2003), How does rainfall become runoff? A combined tracer and runoff transfer function approach, *Water Resour. Res.*, 39, 1315, doi:10.1029/2003WR002331.
- Weiler, M., and J. McDonnell (2004), Virtual experiments: a new approach for improving process conceptualization in hillslope hydrology, *J. Hydrol.*, 285(1-4), 3-18.

CHAPTER SIX

A CROSS-CATCHMENT COMPARISON AND SENSITIVITY ANALYSIS
OF THE CATCHMENT CONNECTIVITY MODELContribution of Author and Co-Authors

Manuscript in Chapter 6

Author:	Tyler Smith
Contributions:	Responsible for the modeling and analysis performed in this study, the majority of the writing, and figure development.
Co-Author:	Lucy Marshall
Contributions:	Provided extensive discussions of the analysis used in the study and contributed to the writing and editing of the manuscript throughout its development.
Co-Author:	Brian McGlynn
Contributions:	Provided extensive discussions of the development of the methods used in the study and valuable critique and editing of the manuscript.
Co-Author:	Kelsey Jencso
Contributions:	Provided extensive discussions of the development of the methods used in the study and valuable critique and editing of the manuscript.

Manuscript Information

Smith, T., L. Marshall, B. McGlynn, and K. Jencso. A cross-catchment comparison and sensitivity analysis of the Catchment Connectivity Model. Manuscript for submittal to *Water Resources Research*.

- Journal: Water Resources Research
- Status of manuscript (check one)
 - ☒ Prepared for submission to a peer-reviewed journal
 - ☐ Officially submitted to a peer-reviewed journal
 - ☐ Accepted by a peer-reviewed journal
 - ☐ Published in a peer-reviewed journal
- Publisher: American Geophysical Union
- Anticipated date of submission: April 2012

A CROSS-CATCHMENT COMPARISON AND SENSITIVITY ANALYSIS OF THE CATCHMENT CONNECTIVITY MODEL

Abstract

The transferability of hydrologic models is of ever increasing importance due to the desire to make improved hydrologic predictions and test hypothesized hydrologic drivers. Here, we present an investigation into the transferability and input and structural sensitivities of the recently introduced Catchment Connectivity Model (CCM) [*Smith et al.*, in review]. The CCM was developed following extensive experimental observations identifying the key drivers of streamflow in the Tenderfoot Creek Experimental Forest (TCEF) [*Jencso et al.*, 2009; *Jencso et al.*, 2010], with the goal of creating a simple and internally consistent hydrologic model structure. The model was applied across eight catchments located within TCEF to investigate spatial sensitivity of model performance and parameterization, finding that model performance was very good across seven of the catchments in terms of both hydrograph fit and internal consistency (corroborated with experimental findings). The sensitivity to model inputs and structure was further examined at the most well studied of the catchments, finding that inputs primarily affected hydrograph fits while alternate structures had the ability to produce improved hydrograph fits but at the potential detriment to internal consistency (dependent on the credibility of the included process).

Introduction

Hydrologic models are often assessed in terms of their ability to reproduce observed streamflow hydrographs for a given catchment following the calibration of its unknown parameters [e.g., *Beven and Binley*, 1992; *Duan et al.*, 1992; *Gupta et al.*, 1998; *Nash and Sutcliffe*, 1970; *Wagner*, 2003]. However, a more important measure of a model's true utility is in its ability to be successfully transferred between hydrologically similar catchments [*Wagner and Gupta*, 2005]. This can reflect the appropriateness of the model structure (i.e., the model is actually representing the acting streamflow generation processes) and by extension can be considered a surrogate for the verification

of internal model consistency (i.e., processes internal to the calibration on streamflow are reliably simulated by the model).

Catchment classification (based upon the foundation of hydrologic similarity) has been an upward trending topic in hydrology over the last several years [e.g., *Blöschl*, 2006; *McDonnell and Woods*, 2004; *McDonnell et al.*, 2007; *Sawicz et al.*, 2011; *Sivapalan*, 2005; *Wagener et al.*, 2007]. At its simplest level, catchment classification is focused on understanding the organizing principles that make sense of the complexity seen across catchments. Through an understanding of the principles that guide classification of catchments, there exists the potential to construct hydrologic models upon the same principles to allow for both internally consistent and transferable model structures.

Model assessment is traditionally performed following an *a priori* selection of a model structure (i.e., model X, Y, or Z is chosen by the modeler with the underlying assumption regarding its appropriateness relative to the study site) and on the basis of its ability to recreate a past streamflow hydrograph following calibration of the model parameters [*Klemeš*, 1983]. Recently, there has been a push to utilize flexible model structures that consist of generic components which can be mixed and matched to create an “appropriate” model structure for the site of interest [e.g., *Clark et al.*, 2008; *Clark et al.*, 2011; *Fenicia et al.*, 2011; *Kavetski and Fenicia*, 2011; *Staudinger et al.*, 2011]. This of course requires the testing of multiple model structures for the site of interest. However, as many model comparison studies have shown, alternate model structures often perform quite similarly to one another (in terms of hydrograph fit) despite

differences in complexity and process representation [e.g., *Duan et al.*, 2006; *Reed et al.*, 2004]. Given this, there is a need to better constrain the assessment of hydrologic models beyond hydrograph fit, to other indicators such as internal process consistency [e.g., *McMillan et al.*, 2011; *Son and Sivapalan*, 2007]. The utility of internally consistent and transferable hydrologic models is immense. Such models are not only critical to improving water resource management (particularly for periods outside the calibration period), but also provide insights into model development and assessment [*Klemeš*, 1983; *Klemeš*, 1986; *Sivapalan et al.*, 2003]. As a consequence of transferable and internally consistent model structures, the ability to make predictions at ungauged catchments is undoubtedly enhanced.

In this study, we assess the recently introduced Catchment Connectivity Model (CCM) [*Smith et al.*, in review] across multiple neighboring catchments in terms of its transferability and internal consistency. We achieve this through cross-catchment diagnostics (variability in model parameters, performance, etc.), a consideration of sensitivities brought about by potential bias in model input data, and the merit of alternative representations of the overland flow process within the original model structure. By considering adjacent catchments, many of the difficulties in assessing model transferability are reduced; forcing climate, dominant processes, and topography are all roughly equivalent. Despite these similarities, there are differences across these sub-catchments that lead to variability in hydrologic response [*Jencso et al.*, 2010; *Nippgen et al.*, 2011]. This paper addresses three main questions: (1) How much variation is there in model performance across catchments for a model developed based on

empirical evidence of the dominant processes acting in the area?; (2) How sensitive is model output to model input?; and (3) Can the model structure be improved to address such differences?

Methods

Study Site Description

The Tenderfoot Creek Experimental Forest (TCEF), located in the Little Belt Mountains of Montana, USA, serves as the study location for this investigation. TCEF is a highly active research site [e.g., *Jencso et al.*, 2009; *Jencso et al.*, 2010; *Jencso and McGlynn*, 2011; *Nippgen et al.*, 2011; *Payn et al.*, 2009; *Smith and Marshall*, 2008; 2010] managed by the U.S. Forest Service, Rocky Mountain Research Station. TCEF is comprised of eight sub-catchments that form the Tenderfoot Creek catchment which drains to the Smith River, a tributary of the Missouri River (Figure 1).

In total, TCEF is approximately 3,700 hectares in area with the uplands being largely occupied by stands of lodgepole pine while shrubs and grasses characterize the riparian areas. Elevations at Tenderfoot Creek Experimental Forest range from approximately 1,800 meters to 2,450 meters. The climate at TCEF is primarily continental, resulting in an annual average precipitation total of approximately 900 mm (strongly dependent on elevation), with about 75% of precipitation falling as snow.

Data

Precipitation, evapotranspiration, a digital elevation model (DEM), and streamflow data were obtained for each of the eight TCEF catchments included in this

study. The analysis period included October 1, 2005 through September 30, 2008, with data procured on a 6-hourly time step. The evapotranspiration data was obtained from collocated eddy-covariance towers located within TCEF [*Emanuel et al.*, 2010] and is assumed to be representative of each TCEF sub-catchment. The precipitation data is based on data collected from the Stringer Creek (1,996 m) and Onion Park (2,259 m) SNOTEL stations located within TCEF. The model precipitation inputs were rain plus snowmelt, with snowmelt computed using a simple logic-based algorithm [refer to *Nippgen et al.*, 2011] that is based on snowpack dynamics. Three different precipitation input series were considered in this study: (1) input from the Stringer Creek SNOTEL station; (2) input from the Onion Park SNOTEL; and (3) an elevation-weighted input that utilized the data from both SNOTEL stations. Figure 2 highlights the differences between the three precipitation input time series for WY2008 (with patterns typical for all years used in this study). For each catchment 10 meter DEMs (derived from 1 meter LiDAR bare earth topography data) were used for landscape analysis. Stream discharge was measured (as stage) on 30 minute intervals at the flumes located at the outlet of each catchment (see Figure 1) using capacitance rods with ± 1 mm resolution that was later aggregated to the 6-hourly time step [refer to *Nippgen et al.*, 2011].

Hydrologic Model

We applied the Catchment Connectivity Model (CCM) [*Smith et al.*, in review] to each TCEF sub-catchment. The CCM, conceptualized in Figure 3, was developed based on extensive empirical observations [*Jencso et al.*, 2009; *Jencso et al.*, 2010] made within TCEF that showed strong correlations between upslope accumulated area (UAA; the

lateral area draining to a particular point along the stream) and shallow groundwater hydrologic connectivity across the hillslope-riparian-stream continuum. The foundational concept of the model structure is that streamflow processes are driven by the frequency of hydrologic connections of the hillslopes to the stream rather than the magnitude of any single hydrologic connection and that the frequency of connections is temporally dynamic. The upslope accumulated area for a given hillslope is used to simulate the threshold driven connectivity between the hillslopes and the stream (equation 3 of Figure 3).

Modeled hillslope (landscape unit) storage (equation 1 of Figure 3) is computed as the difference between inputs (previous hillslope storage and current precipitation) and outputs (evapotranspiration). Hydrologic connectivity of the hillslope to the stream is determined to be present when hillslope storage exceeds the hillslope storage threshold (equation 2 of Figure 3), which is based on the hillslope's size (as UAA). During an active hydrologic connections ($h_c = 1$, equation 3 of Figure 3) water is released from the hillslope as shallow groundwater at a rate equal to q^* (a calibrated parameter) times the frequency weight of hillslope, with an upper limit equal to the current storage (equation 4 of Figure 3). Each connected hillslope contributes water to the conceptual subsurface storage that is then routed to catchment streamflow via an exponential filter model (equation 5 of Figure 3). Modeled hillslope storage is updated at the end of the time step to account for the output of water as streamflow (equation 6 of Figure 3). Despite its distributed nature, the CCM is a parsimonious, six parameter model (parameter names and descriptions given in Table 2) due to the empirically-based conceptualization of

streamflow being driven by the frequency of hydrologic connections and the upslope accumulated area-dependent threshold function.

The model was implemented using a semi-distributed approach in this application. The catchment was discretized into hillslopes based on discrete 10m units of stream length (referred to as stream cells; for both the left- and right-hand sides of the stream). For computational convenience, these hillslopes can then be aggregated into non-contiguous landscape units based on their size. Details regarding the study catchments, including area, channel length, and the number of landscape units for each catchment considered in this study are provided in Table 1. In each of the catchments, the hillslopes were aggregated such that the ratio of hillslopes to landscape units was approximately equal, noting that the landscape units are proportional to catchment stream length and not catchment area.

Model Calibration & Uncertainty Analysis Framework

A Bayesian approach was used to perform model calibration and uncertainty analysis in this study, a common and useful method for hydrologic modeling [e.g., *Bates and Campbell, 2001; Huard and Mailhot, 2008; Kuczera and Parent, 1998; Schaefli et al., 2007*]. The Adaptive Metropolis (AM) algorithm [*Haario et al., 2001*] has been successfully implemented in a number of hydrologic modeling studies [e.g., *Marshall et al., 2004*] including previous work within TCEF [*Smith and Marshall, 2008; 2010*] and was chosen to carry out the Bayesian approach here. A Box-Cox transformed Gaussian likelihood function [refer to *Bates and Campbell, 2001*] was used to address potential

heteroscedasticity in the model residuals. The likelihood function is represented by two additional parameters (a Box-Cox transformation parameter and a variance parameter) that were calibrated along with the CCM parameters.

Applications

The sensitivities of the Catchment Connectivity Model are described in the following subsections, with a focus on three aspects: (1) spatial variability and transferability – the variability of model performance and parameterizations across multiple catchments; (2) input sensitivity – the sensitivity of model performance to varying precipitation input sources; and (3) model structural variability – the variability in results when including additional process components to the structure. In each of the last two analyses, only the Lower Stringer Creek (LSC) catchment was considered due to it being the original catchment at which the CCM was developed and also to simplify the presentation of results.

Spatial Variability and Transferability: A Cross-Catchment Analysis

The Catchment Connectivity Model was calibrated for each of the eight catchments shown in Figure 1. The first year of data was used as a model warm-up period to allow the model storages to be initialized for the start of the second year. The model was specifically calibrated for the April-September runoff period for years two and three, since this period represents the primary period of interest; the remainder of the year is marked primarily by baseflow due to a lack of input (i.e., rain or snowmelt) precipitation.

The same evapotranspiration time series was used for each catchment and a catchment-specific, elevation-weighted precipitation time series was used (based on the Onion Park and Stringer Creek SNOTEL stations) [Nippgen *et al.*, 2011].

The ability of the model to simulate the observed streamflow hydrograph at each of the catchments, in addition to the model predicted HRS connectivity (areas shown in gray shading represent those *not* used for parameter calibration) is shown in Figure 4. In general, model fits are fairly consistent (in terms of the Nash-Sutcliffe Efficiency; shown in Figure 4) across each of the eight catchments with the exception of the BUB catchment. All of the catchments tend to under-predict the hydrograph peaks. Note that the results presented in Figure 4 are shown in untransformed space; however, the model parameters were calibrated in Box-Cox transformed (log) space. This transformation effectively reduced the weight that high streamflow observations were given in untransformed space; as a result, in transformed space the under predictions are less obvious and are balanced by slight over predictions in baseflow.

Examining the results further, we plotted the flow duration curves (FDC) for each of the catchments for the entirety of the last two years of the time series in Figure 5. FDCs allow the model predictions to be considered cumulatively without respect to timing and also afford an opportunity to visually assess some of the differences between the observed and predicted that can be obscured in hydrographs. The general under prediction of high streamflows is highlighted, but at the same time the ability of the model to accurately simulate streamflows of approximately 20% exceedence and greater in all catchments is also accentuated. Further, catchments SPC and UTEN showed good

fits across all streamflow magnitudes and to a slightly lesser extent the same is true for LONE, LSC, and SUN.

The HRS connectivity results can be presented in a similar manner to the streamflow FDC by means of the connectivity duration curve (CDC). The CDCs (Figure 6) represent the binary connectivity relationship shown in Figure 4 in terms of percent connectivity versus exceedence probability and allow for the internal consistency of the model to be assessed. Like the FDCs shown in Figure 5, the simulated CDCs tend to under-predict the observed CDCs (i.e. the model simulates too little connectivity across the stream network). The results show significant variation across the best and worst performing catchments, with PACK being very similar to the observed and BUB being very dissimilar to the observed. It should be noted that the observed line was created by extrapolating the empirical relationship between upslope accumulated area and connectivity [refer to *Jencso et al.*, 2009] to each individual catchment.

The sensitivity/variability across catchments, in terms of model performance, can be linked to the variability in model parameters across catchments. This variability is shown in terms of boxplots of the posterior parameter distributions (Figure 7) for each model parameter at each catchment. The parameter distributions across each of the catchments show both variation (parameters q^* , δ , uaa_c) and consistency (parameters τ , γ , κ with exception to SUN); refer to “Hydrologic Model” section, Table 2, and Figure 3 for further details regarding the model parameters. The variability in parameters found across catchments is largely expected as the model structure is based on catchment-specific upslope accumulated areas to drive the threshold function, which controls hydrologic

connectivity and thus streamflow generation. However, of more interest are the parameters that show systematic consistency across most but not all of catchments. These parameters will be discussed in further detail in the discussion.

Input Sensitivity

We investigated the systematic under prediction of streamflows across catchments shown in Figure 4. The sensitivity of the CCM to variations in the precipitation inputs, in terms of performance and parameterization, for the Lower Stringer Creek (LSC) catchment was analyzed. The original model simulation driven by the elevation-weighted precipitation inputs was compared with new model simulations for precipitation inputs arising from each of the individual SNOTEL stations (Onion Park and Stringer Creek) located within Tenderfoot Creek Experimental Forest. By considering these three precipitation data sets, we can begin to investigate the potential observational variability and/or biases that may exist in the data sets. As a result of these three unique precipitation time series, three unique CCM parameterizations and streamflow simulations were computed.

The impact of input variability of modeled hydrographs in the Lower Stringer Creek catchment was evident (Figure 8). In general, the Stringer Creek SNOTEL station precipitation results in the largest under predictions of the observed hydrograph followed by the elevation-weighted precipitation with the Onion Park SNOTEL station precipitation time series resulting in the highest predictions of streamflow.

Simulated flow duration curves, shown in Figure 9, highlight these differences that are most notable at higher streamflows. The 0% exceeded streamflow value for the

Stringer Creek SNOTEL prediction is significantly less than that of the observed hydrograph (53% less). Likewise, the Stinger Creek SNOTEL data set resulted in peak streamflows much lower than the other two data sets also, 31% less than the Onion Park SNOTEL peak predictions and 25% less than the elevation-weighted peak predictions. Despite these differences in peak streamflow, the NSE values remain fairly consistent (refer to Figure 8) due to model calibration being conducted in Box-Cox transformed (log) space. The sensitivity of the modeled hydrologic connectivity was investigated in response to the varying precipitation input time series. The connectivity duration curves for each time series (Figure 10) indicated that there is little variation in modeled connectivity in response to varying precipitation input data, despite there being observed differences in calibrated model parameters presented in Figure 11.

Model Structural Variability

In the final element of this study, the sensitivity of the model predictions to model structure was assessed. We considered three alternative formulations of the overland flow process within the CCM structure: (1) no overland flow (the original version of the CCM); (2) a riparian area limited overland flow component; and (3) an unrestricted overland flow component. The decision to test the addition of an overland flow component was selected for two reasons. First, overland flow is a process that is known to act in these catchments but was not included in the original structure (due to it being perceived as a minor process). Second, based on the general under prediction of hydrograph peaks (refer to Figure 4), it was believed that this process might be more important to peak streamflow prediction than originally thought. This assessment was

again focused on the Lower Stringer Creek catchment, and uses the elevation-weighted precipitation time series data.

The riparian area limited overland flow component was included such that the overland flow at any time was equal only to the proportion of precipitation occurring over the saturated riparian area. *Jencso et al.* [2010] delineated riparian area extents for the Lower Stringer Creek catchment, classifying 3% of the catchment as riparian area. The unrestricted overland flow component involved placing an upper threshold (an additional calibrated parameter, see Table 2) on the model storages such that any precipitation in excess of the upper limit would be routed to the stream as overland flow, provided the modeled storage was hydrologically connected to the stream.

The streamflow hydrograph for each of the three model conceptualizations is compared to the observed hydrograph in Figure 12. From this figure, it is immediately clear that the structure that includes the unrestricted overland flow component results in much higher predictions of streamflow during the peaks of the hydrograph, as compared to the original structure and the riparian area-limited overland flow supplemented structure. This is further stressed by the FDC shown in Figure 13. While the original and riparian overland flow versions of the model result in under predictions of the peak streamflow (0% exceedence value), the unrestricted overland flow version actually over estimates peak flows.

However, when considering internal model performance (as the ability of the structures to reproduce observed hydrologic connectivity), Figure 14 emphasizes the lack of appropriateness of the unrestricted overland flow version of the CCM. The modeled

connectivity duration curve for the unrestricted overland flow version of the CCM fails to adequately reproduce the observed curve, whereas the unmodified and riparian area-limited overland flow versions of the CCM structure perform almost identically and represent a significantly better fit to the observed CDC.

Discussion

The transferability of hydrologic models is of increasing relevance as interest in better managing water resources becomes more critical to societal needs and pressures. In response to this, we performed a multifaceted sensitivity analysis and model assessment using the recently introduced Catchment Connectivity Model [*Smith et al.*, in review] that was developed based on extensive empirical observations at the Tenderfoot Creek Experimental Forest [*Jencso et al.*, 2009; *Jencso et al.*, 2010]. We considered the transferability of the model across eight catchments located within TCEF, as well as the sensitivity of the model performance at the Lower Stringer Creek catchment to precipitation inputs based on different local gauging stations (Stringer Creek SNOTEL, Onion Park SNOTEL, elevation-weighting of both SNOTEL stations) and to the inclusion of alternate representations of overland flow (none, riparian-area limited, unrestricted).

Our analysis investigated the question of how transferrable the CCM structure is across catchments. The study revealed generally good fitting hydrographs (Figure 4), flow duration curves (Figure 5), and internal model consistency (in terms of connectivity duration curves, Figure 6) across all catchments. The CCM performed quite well with an

average Nash-Sutcliffe efficiency (computed in transformed space) of 0.82 despite the Bubbling Creek (BUB) catchment having an NSE equal to 0.66.

The Bubbling Creek catchment had by far the poorest performance not only in terms of hydrograph fit (i.e., NSE) but also in terms of predicted hydrologic connectivity. Linking the poor hydrograph fit and poor internal consistency to the catchment information provided in Table 1, we see that the Bubbling Creek catchment had only ten landscape units due to shorter channel length than the other catchments. These results together indicate that there potentially exists a lower limit to the number of landscape units required for the CCM to be successful. Previous investigations into the sensitivity of model performance to the number of landscape units at the Lower Stringer Creek (LSC) catchment found little difference in performance across a range of 30-100 landscape units [Smith *et al.*, in review]; however, further investigations are planned to focus on the role of catchment discretization (at the lower limit) on model simulations, especially for simulating baseflow.

In terms of model parameters, the calibrated parameter distributions (as boxplots; Figure 7) for each catchment highlight both consistency and variation across parameters. In two of the CCM parameters (τ and γ), the BUB catchment parameters were substantially different from the other seven catchments. The γ parameter represents a data bias scaling factor applied to the evapotranspiration time series. Functionally when this parameter is greater than 1, streamflow is modulated by allowing less water to be released from the catchment. At the same time, the τ parameter represents the response time for the routing (exponential filter) and larger values result in water being released

from the catchment more slowly. Together these values combine to limit the amount of water released from the catchment. This result is logical as experimental observation suggests that the Bubbling Creek catchment is “leaky” (i.e., the geologic layers are positioned such that groundwater is lost to other catchments). The Sun Creek catchment likewise represents an “outlier” for the κ parameter (an exponent in the model storage threshold function). The exact role of this parameter is more difficult to elucidate as it is related to the other threshold function parameters (δ and uaa_c).

Other parameters show some degree of variability but with the values of each of the catchments falling within similar magnitudes. The question is, should the parameters vary if the streamflow generating processes and the climate are the same? The answer to this question is obviously complicated, but a simple understanding of how the CCM functions structurally can provide some insights. For instance, the q^* parameter represents the rate of discharge from any connected landscape unit. It should be expected, coupled with the fact that the largest landscape unit is structurally enforced to be always connected, that q^* is representative of low flow (baseflow) conditions. This understanding results in the expectation that q^* will vary in response to baseflow magnitudes at the catchments under consideration. Likewise, the uaa_c parameter (value of upslope accumulated area for which continuous connectivity occurs) will be a function of the catchment-specific upslope accumulated area.

All model predictions of streamflow tend to under-predict peak streamflows, clearly evidenced in the FDCs (Figure 5). In turn, the predictions of hydrologic connectivity (internal to the model calibration; Figure 6) also showed under predictions

of connectivity relative to the observed connectivity based on the empirical relationship of *Jencso et al.* [2009]. Is this indicative of a bias in the model structure or the model input data? To address this question, we further investigated these issues at the Lower Stringer Creek (LSC) catchment.

Uncertainty in the model inputs is a major concern in many hydrologic models [e.g., *Kavetski et al.*, 2006]. Here we simply investigated the response of the model simulations to three alternate precipitation input series arising from two different SNOTEL stations located within the study area. Figure 8 and Figure 9 revealed that the precipitation series with the smallest magnitude (Stinger Creek SNOTEL) resulted in greater under predictions of streamflow (particularly on the hydrograph peaks) than the other two input series (which are much more similar in terms of cumulative magnitudes). This result was somewhat interesting considering the model structure included a data bias scaling factor (γ parameter). Although this parameter is applied to the evapotranspiration input data, it is indirectly linked to the precipitation input through the model equations (water balance).

However, while this parameter affords the model the flexibility to increase the proportion of water entering the system (by adjusting the amount leaving the system), the results shown in Figure 11 reveal the complex interactions that this involves as it moves through the model structure in response to the end objective of optimizing the model parameters. As the γ parameter is decreased (to increase the precipitation to evapotranspiration ratio), the value of q^* (a baseflow surrogate) is likewise decreased and is accompanied by an increase in τ (response time). Similar shifts were also brought about

by the shift in the data bias scaling parameter in the model threshold function (that drives connectivity prediction) parameters (κ , δ , uaa_c). These results suggest that although the γ parameter can modulate the precipitation to evapotranspiration ratio, doing so leads to shifts in other parameter values in response to the model's structural composition and assumptions. Despite the shifts in model simulations of the observed hydrograph (in response to shifts in parameter values), little difference in terms of internal predictions of hydrologic connectivity across the three precipitation input time series were observed (Figure 10). This analysis highlights the fact that the model parameters are indeed sensitive to the input data.

The final component of the study then examined whether the under prediction of peak streamflow likely originates with inputs or with the assumed model structure. By comparing three alternate conceptualizations of the overland flow component of the CCM structure (none, riparian-limited, unrestricted) for the elevation-weighted precipitation input series at the LSC catchment, we were able to refine our assessment of the CCM's performance in conjunction with the cross-catchment assessment and input data sensitivity analysis.

Based on the hydrographs and FDCs (Figure 12 and Figure 13, respectively), we found that similar performance was achieved for the version of the CCM that had no overland flow component and the version that included the riparian area-limited overland flow component that was based on extensive empirical observation at the catchment [Jencso *et al.*, 2009; Jencso *et al.*, 2010]. In contrast, the unrestricted overland flow conceptualization resulted in a better fit to peak streamflow events (though at some

expense to the fits at lower streamflows). Moving to the predictions of internal model consistency (Figure 14), again we saw similar performance for the two more realistic (i.e., empirically justified) conceptualizations of overland flow. However, the unrealistic representation of overland flow resulted in a significant degradation in internal model consistency.

Flexible model structures have become increasingly popular recently [e.g., *Clark et al.*, 2011; *Fenicia et al.*, 2011; *Kavetski and Fenicia*, 2011; *Staudinger et al.*, 2011]. These structural identification frameworks have the ability to incorporate new/alternate processes to address perceived model structural shortcomings [*Clark et al.*, 2008]. However, the findings presented here, when taken as a whole, suggest that while alterations to the model structure can be effective at improving hydrograph fits, they can also be unjustifiable based on experimental evidence and result in simulations that get the right answers for the wrong reasons [*Kirchner*, 2006]. A more promising avenue toward improving model performance while maintaining internal consistency appears to be related to the quality of input data. Though the ultimate differences for the three precipitation time series were not great, they suggest that all of the model parameters are affected by such uncertainty. The sensitivity analysis to inputs highlighted the potential for improvement of the threshold function that determines predicted hydrologic connectivity, which was found to be less sensitive to the input data. Perhaps improving this function's responsiveness to inputs is what is needed to improve the internal predictions of hydrologic connectivity rather than the addition of more streamflow generating processes.

Conclusions

This study represents a detailed examination of the transferability and sensitivity of the Catchment Connectivity Model, a conceptual hydrologic model developed following detailed experimental observations made within the Tenderfoot Creek Experimental Forest. The intent of the model development was to create a simple, yet internally consistent (and verifiable) structure with the hope of improving model transferability for catchments of similar physical controls.

An examination of the model, individually calibrated across eight adjacent catchments, highlighted the consistency of the model simulations in terms of both external (i.e., streamflow) and internal (i.e., hydrologic connectivity) processes. Despite this, it was shown that model parameters did (and in some cases should) vary across catchments due to their relation to physical processes that vary with catchment scale. This stresses the fact that direct transfer of parameters from one catchment to another is fraught with complications. The sensitivity analysis to model structure underscored the fact that internal consistency is highly dependent upon realistic mechanisms being represented in the model structure. The sensitivity to input indicated that biases in model input are a key area for further research to improve overall model performance, while maintaining acceptable levels of internal consistency.

Future research is planned that will continue to examine these issues, including the sensitivity of the model simulations to alternate representations of the storage

threshold function (determining hydrologic connectivity) and further investigation into the utility of the CCM for making predictions in ungauged basins.

Acknowledgements

This study was funded by NSF grant EAR-0943640. The first author received additional support as a Graduate Fellow of the Montana Institute on Ecosystems. The authors thank the U.S. Forest Service, Rocky Mountain Research Station for site access and logistical support. Airborne Laser Mapping was provided by the National Science Foundation supported Center for Airborne Laser Mapping (NCALM) at the University of California, Berkeley.

Tables

Table 1. Catchment Characteristics for Each Catchment Modeled in this Study.

<i>Catchment Name</i>	<i>Abbreviation</i>	<i>Area (ha)</i>	<i>Stream Length (m)</i>	<i>Number of Landscape Units</i>
Bubbling Creek	BUB	318	740	10
Lonesome Creek	LONE	311	2,090	30
Lower Stringer Creek	LSC	555	3,780	50
Pack Creek	PACK	333	1,540	20
Passionate Creek	PASS	201	2,380	30
Spring Park Creek	SPC	418	3,000	40
Sun Creek	SUN	347	1,710	25
Upper Tenderfoot Creek	UTEN	444	2,200	30

Table 2. Names and Descriptions of the CCM Parameters Requiring Calibration.

<i>Parameter</i>	<i>Description</i>
q^*	the rate of discharge from each landscape unit [volume/time]
τ	the response time parameter of an exponential filter [time]
δ	a scaling parameter for catchment storage [depth]
κ	the parameter describing the shape of the UAA-connectivity relationship [-]
uaa_c	the value of upslope accumulated area above which continuous connectivity occurs [area]
γ	a data bias scaling factor [-]
$cmax$	the maximum soil storage capacity (only used in unrestricted overland flow version of the model) [volume]

Figures

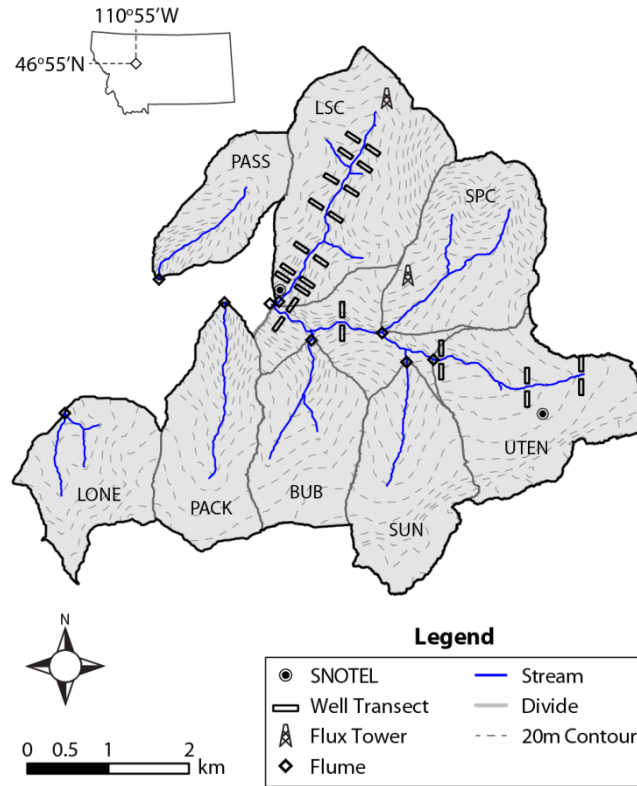


Figure 1. The Tenderfoot Creek Experimental Forest, with each of the eight catchments used in this study shown. Model forcing data was retrieved from two sources: (1) the precipitation inputs (as rain plus snowmelt) were derived from SNOTEL stations located within TCEF (the Onion Park station in the UTEN catchment and the Stringer Creek station near the outlet of the LSC catchment) and (2) the evapotranspiration data was gathered from the collocated eddy-covariance towers. The well transects measure shallow groundwater connectivity across the hillslope-riparian-stream continuum (on each side of the stream).

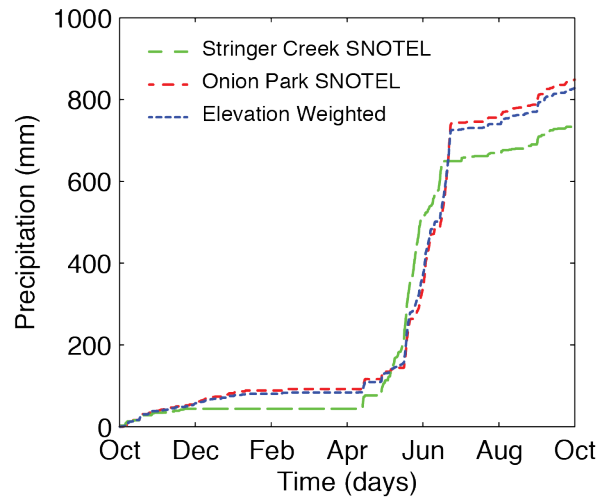


Figure 2. Cumulative precipitation (as rain plus snowmelt) over the course of water year 2008 using three different data sources for the Stringer Creek catchment: (1) the Onion Park SNOTEL station; (2) the Stringer Creek SNOTEL station; and (3) and elevation-weighting of the two SNOTEL stations.

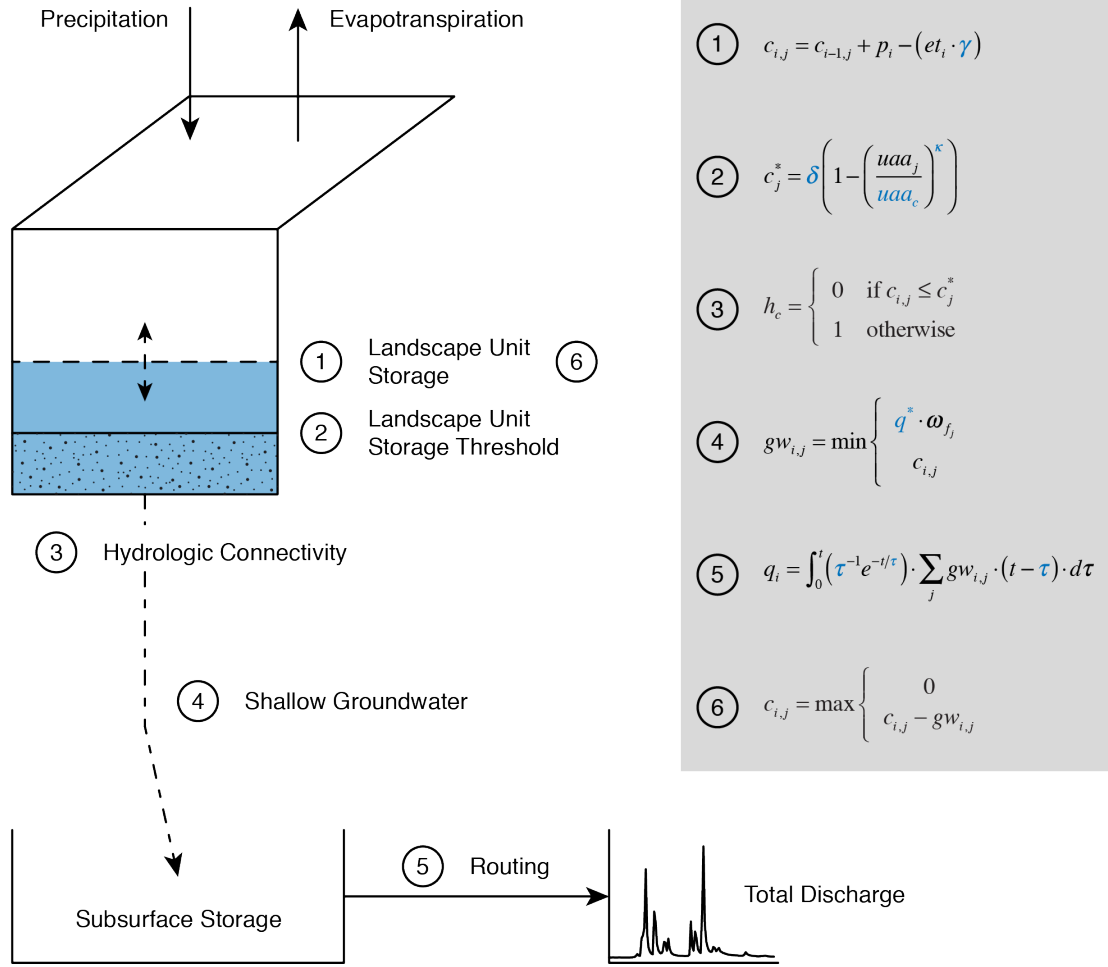


Figure 3. A schematic of the Catchment Connectivity Model, highlighting the computational implementation for a single landscape unit.

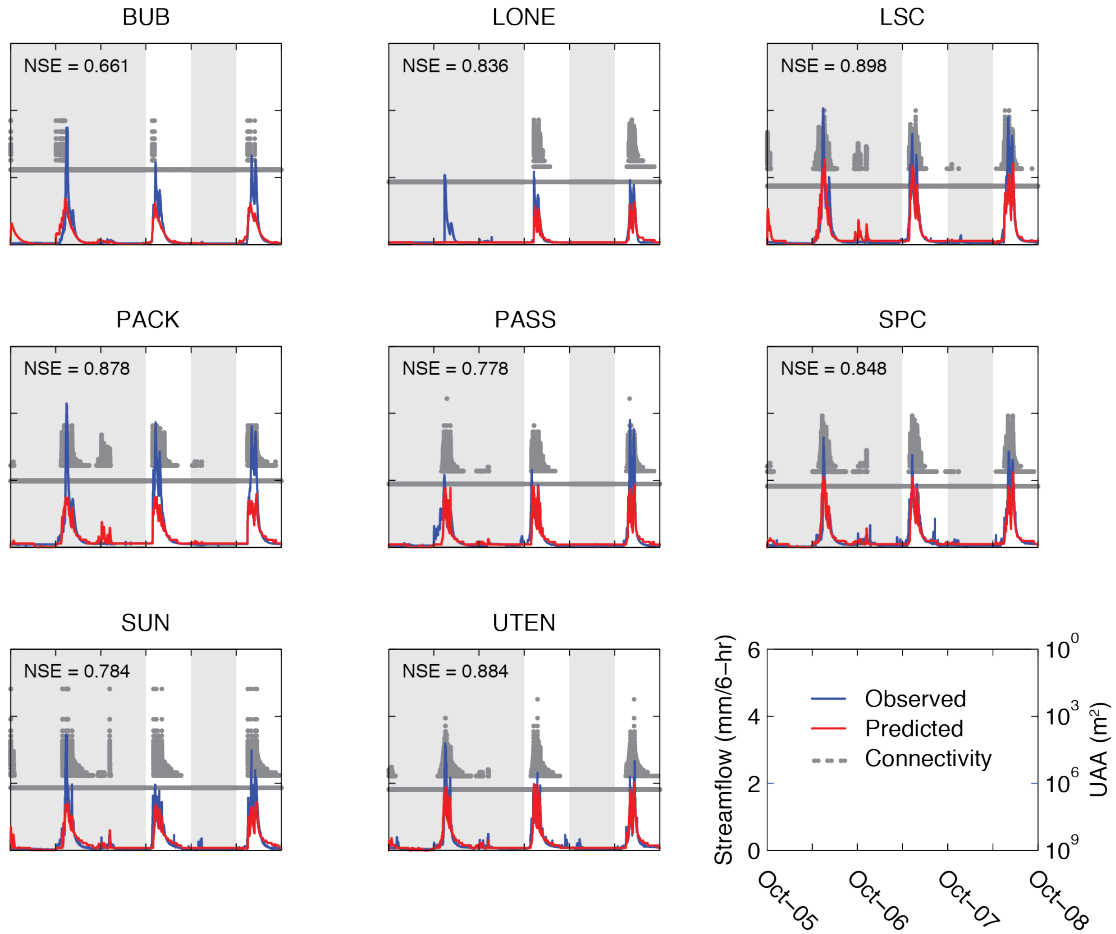


Figure 4. A comparison of the streamflow hydrographs and model predicted binary connections for each catchment. Note that the first year of data was used for model warm-up (to allow for initialization of the model storages) and that the model was calibrated over the peak streamflow season (April-September) for the remaining two years of data (periods shown without shading). The NSE value shown is for the model calibration period only and is based on the Box-Cox transformation used in the model calibration. Refer to Figure 1 and Table 1 for catchment details.

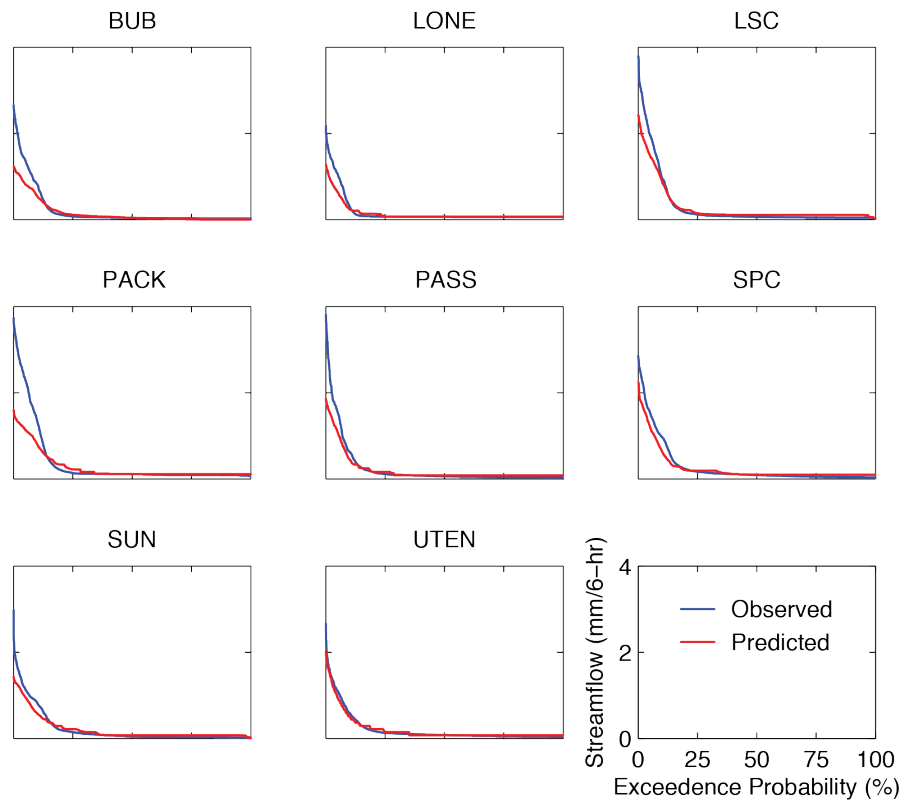


Figure 5. Flow duration curves (both observed and predicted) for each of the catchments used in this study for the entirety of the second and third years of data.

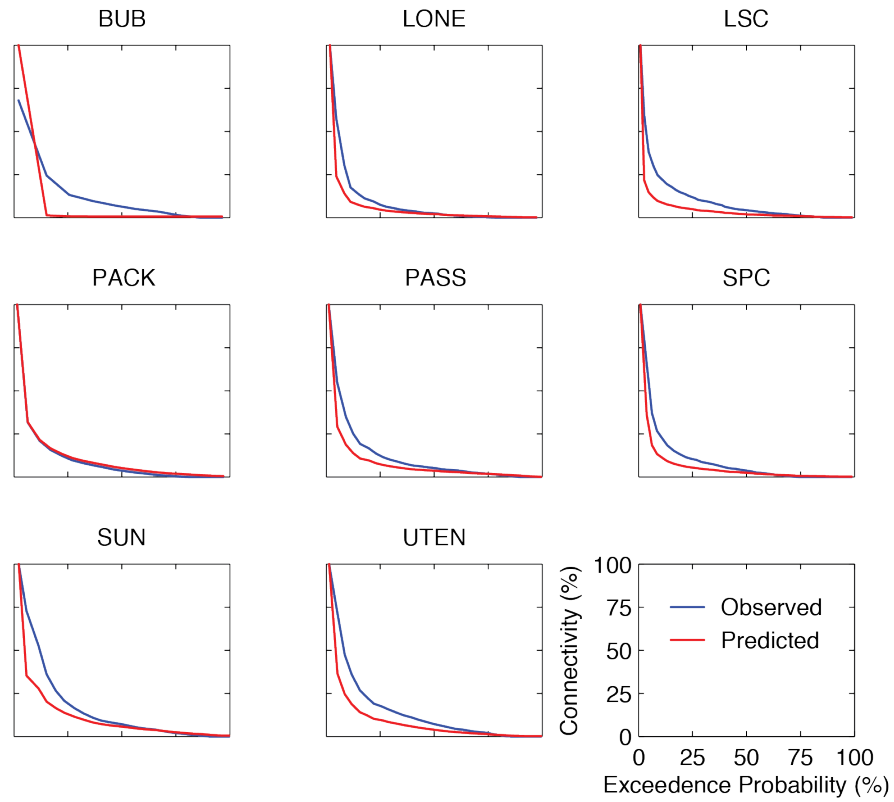


Figure 6. Cross-catchment comparison of the connectivity duration curve for each of the eight catchments included in the study for the entirety of the second and third years of data. Note that the “observed” curve is based on an experimentally-derived relationship that has been extrapolated to each of the individual catchments.

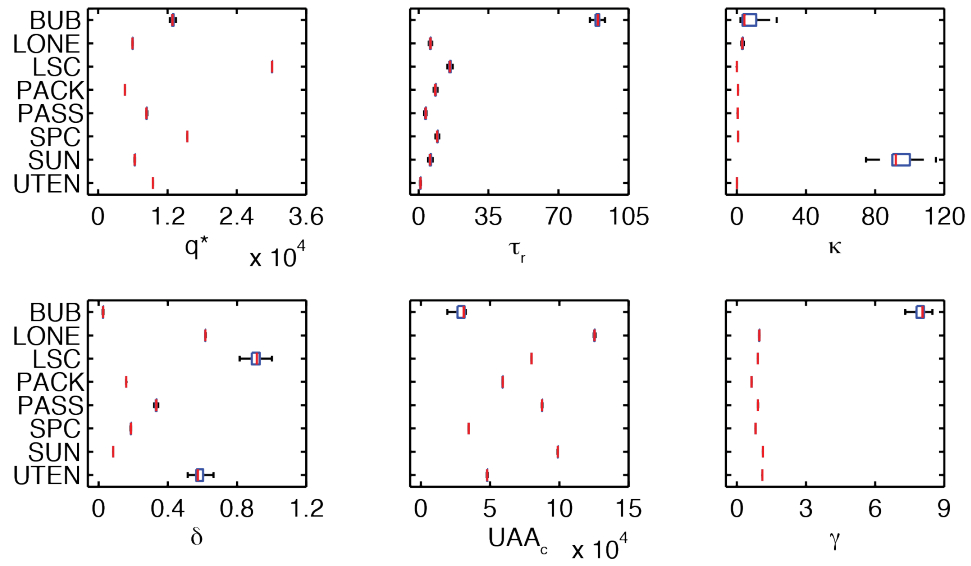


Figure 7. Boxplots of the posterior distributions for each of the calibrated CCM parameters and each of the eight catchments investigated.

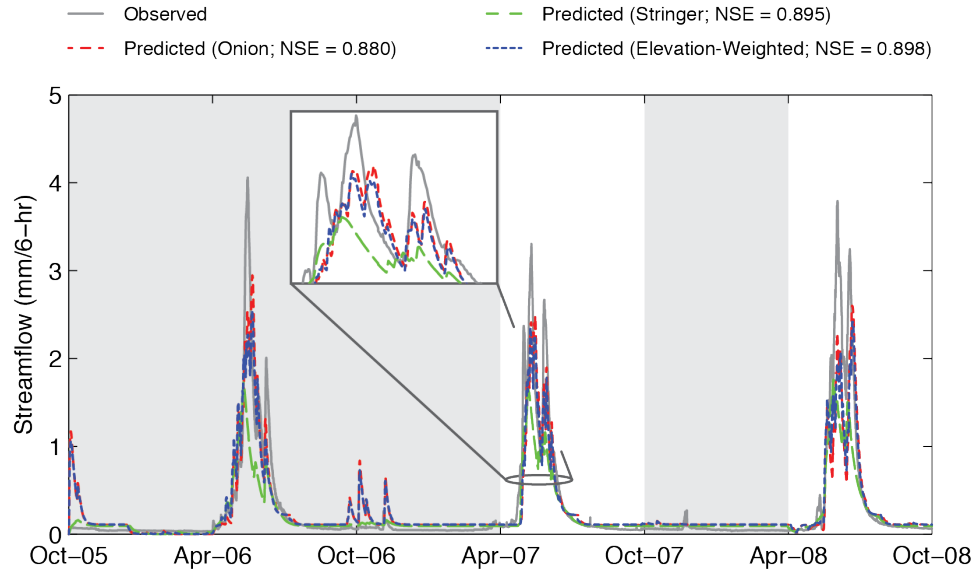


Figure 8. Comparison of streamflow hydrographs for the three different precipitation input time series. In each case, the first year was used for warm-up and the parameters were calibrated over the April-September period for each of the subsequent years (no shading).

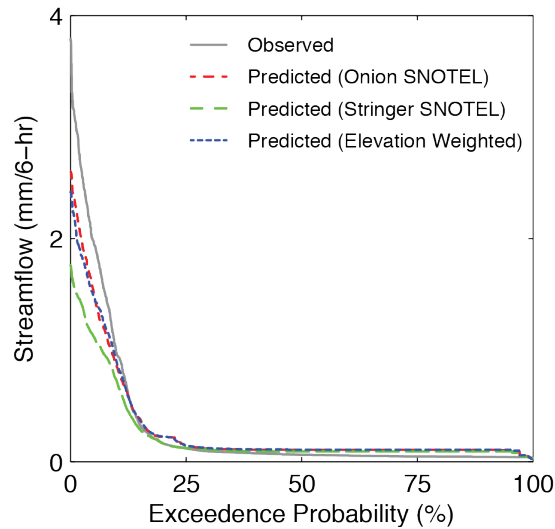


Figure 9. A flow duration curve (computed for years two and three of data) highlighting the differences in performance of the CCM in response to varying precipitation inputs at the Stringer Creek catchment.

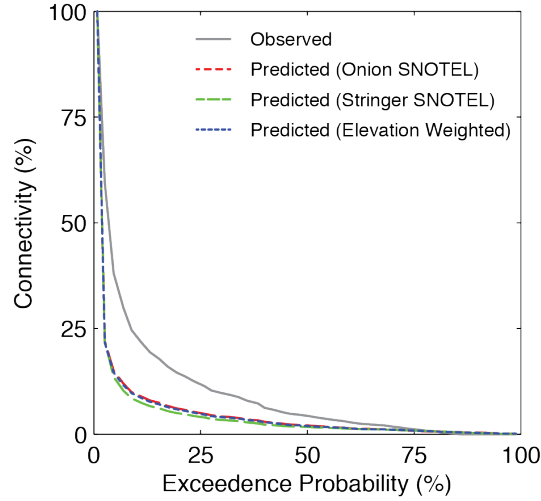


Figure 10. A connectivity duration curve (computed for years two and three of data) highlighting the differences in the internal consistency of the CCM in response to varying precipitation inputs at the Stringer Creek catchment.

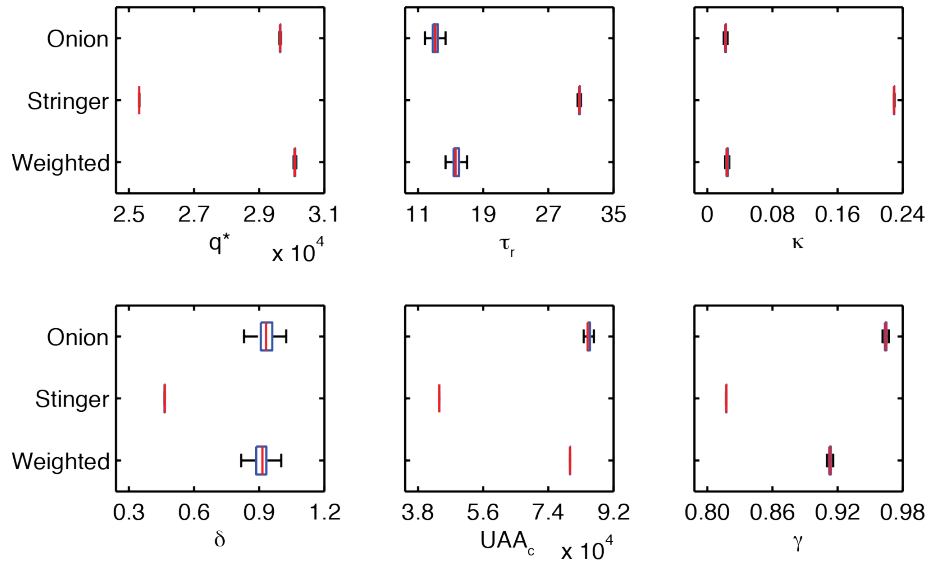


Figure 11. Boxplots of the posterior distributions for each of the calibrated CCM parameters at the Lower Stringer Creek (LSC) catchment for each of the three precipitation input data sets. Recall that the Onion Park SNOTEL has the highest measured precipitation followed by the elevation-weighted series and Stringer Creek SNOTEL.

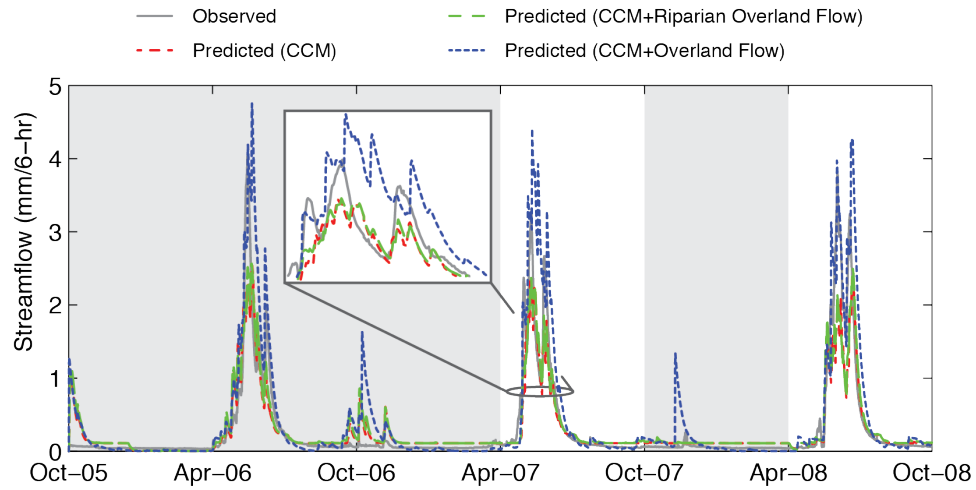


Figure 12. Comparison of streamflow hydrographs for the basic CCM structure, the CCM with a riparian area-limited overland flow component, and the CCM with an unrestricted overland flow component. In each case, the first year was used for warm-up and the parameters were calibrated over the April-September period for each of the subsequent years (no shading).

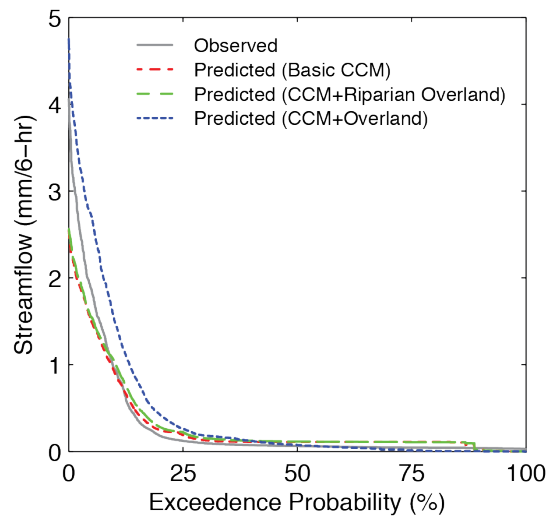


Figure 13. A comparison of the flow duration curves for the CCM, the CCM with riparian area-limited overland flow, and CCM with unrestricted overland flow. The first year of data was used for model warm-up and was excluded from the computation of flow durations.

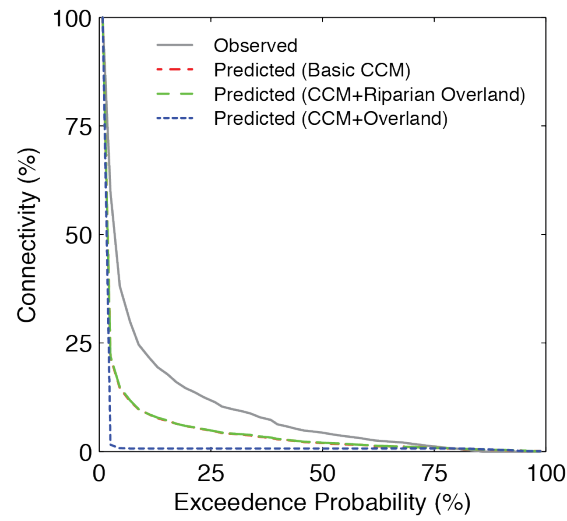


Figure 14. A comparison of the connectivity duration curves for the CCM, the CCM with riparian area-limited overland flow, and CCM with unrestricted overland flow. The first year of data was used for model warm-up and was excluded from the computation of connectivity durations.

References

- Bates, B. C., and E. P. Campbell (2001), A Markov chain Monte Carlo scheme for parameter estimation and inference in conceptual rainfall-runoff modeling, *Water Resour. Res.*, 37(4), 937-947.
- Beven, K., and A. Binley (1992), The future of distributed models: Model calibration and uncertainty prediction, *Hydrol. Processes*, 6(3), 279-298.
- Blöschl, G. (2006), Hydrologic synthesis: Across processes, places, and scales, *Water Resour. Res.*, 42, W03S02, doi:10.1029/2005WR004319.
- Clark, M. P., A. G. Slater, D. E. Rupp, R. A. Woods, J. A. Vrugt, H. V. Gupta, T. Wagener, and L. E. Hay (2008), Framework for Understanding Structural Errors (FUSE): A modular framework to diagnose differences between hydrological models, *Water Resour. Res.*, 44, W00B02, doi:10.1029/2007WR006735.
- Clark, M. P., D. Kavetski, and F. Fenicia (2011), Pursuing the method of multiple working hypotheses for hydrological modeling, *Water Resour. Res.*, 47, W09301, doi:10.1029/2010wr009827.
- Duan, Q., S. Sorooshian, and V. Gupta (1992), Effective and efficient global optimization for conceptual rainfall-runoff models, *Water Resour. Res.*, 28(4), 1015-1031.
- Duan, Q., J. Schaake, V. Andréassian, S. Franks, G. Goteti, H. V. Gupta, Y. M. Gusev, F. Habets, A. Hall, L. Hay, T. Hogue, M. Huang, G. Leavesley, X. Liang, O. N. Nasonova, J. Noilhan, L. Oudin, S. Sorooshian, T. Wagener, and E. F. Wood (2006), Model Parameter Estimation Experiment (MOPEX): An overview of science strategy and major results from the second and third workshops, *J. Hydrol.*, 320(1-2), 3-17.
- Emanuel, R. E., H. E. Epstein, B. L. McGlynn, D. L. Welsch, D. J. Muth, and P. D'Odorico (2010), Spatial and temporal controls on watershed ecohydrology in the northern Rocky Mountains, *Water Resour. Res.*, 46, W11553, doi:10.1029/2009WR008890.
- Fenicia, F., D. Kavetski, and H. H. G. Savenije (2011), Elements of a flexible approach for conceptual hydrological modeling: 1. Motivation and theoretical development, *Water Resour. Res.*, 47(11), W11510.
- Gupta, H. V., S. Sorooshian, and P. O. Yapo (1998), Toward improved calibration of hydrologic models: Multiple and noncommensurable measures of information, *Water Resour. Res.*, 34(4), 751-763.

- Haario, H., E. Saksman, and J. Tamminen (2001), An adaptive Metropolis algorithm, *Bernoulli*, 7(2), 223-242.
- Huard, D., and A. Mailhot (2008), Calibration of hydrological model GR2M using Bayesian uncertainty analysis, *Water Resour. Res.*, 44, W02424, doi:10.1029/2007WR005949.
- Jencso, K. G., B. L. McGlynn, M. N. Gooseff, S. M. Wondzell, K. E. Bencala, and L. A. Marshall (2009), Hydrologic connectivity between landscapes and streams: Transferring reach- and plot-scale understanding to the catchment scale, *Water Resour. Res.*, 45, W04428, doi:10.1029/2008WR007225.
- Jencso, K. G., B. L. McGlynn, M. N. Gooseff, K. E. Bencala, and S. M. Wondzell (2010), Hillslope hydrologic connectivity controls riparian groundwater turnover: Implications of catchment structure for riparian buffering and stream water sources, *Water Resour. Res.*, 46, W10524, doi:10.1029/2009WR008818.
- Jencso, K. G., and B. L. McGlynn (2011), Hierarchical controls on runoff generation: Topographically driven hydrologic connectivity, geology, and vegetation, *Water Resour. Res.*, 47(11), W11527.
- Kavetski, D., G. Kuczera, and S. W. Franks (2006), Bayesian analysis of input uncertainty in hydrological modeling: Theory, *Water Resour. Res.*, 42, W03407, doi:10.1029/2005WR004368.
- Kavetski, D., and F. Fenicia (2011), Elements of a flexible approach for conceptual hydrological modeling: 2. Application and experimental insights, *Water Resour. Res.*, 47(11), W11511.
- Kirchner, J. W. (2006), Getting the right answers for the right reasons: Linking measurements, analyses, and models to advance the science of hydrology, *Water Resour. Res.*, 42, W03S04, doi:10.1029/2005WR004362.
- Klemeš, V. (1983), Conceptualization and scale in hydrology, *J. Hydrol.*, 65(1-3), 1-23.
- Klemeš, V. (1986), Operational testing of hydrological simulation models, *Hydrol. Sci. J.*, 31(1), 13-24.
- Kuczera, G., and E. Parent (1998), Monte Carlo assessment of parameter uncertainty in conceptual catchment models: the Metropolis algorithm, *J. Hydrol.*, 211(1-4), 69-85.
- Marshall, L., D. Nott, and A. Sharma (2004), A comparative study of Markov chain Monte Carlo methods for conceptual rainfall-runoff modeling, *Water Resour. Res.*, 40, W02501, doi:10.1029/2003WR002378.

- McDonnell, J. J., and R. Woods (2004), On the need for catchment classification, *J. Hydrol.*, 299(1-2), 2-3.
- McDonnell, J. J., M. Sivapalan, K. Vaché, S. Dunn, G. Grant, R. Haggerty, C. Hinz, R. Hooper, J. Kirchner, M. L. Roderick, J. Selker, and M. Weiler (2007), Moving beyond heterogeneity and process complexity: A new vision for watershed hydrology, *Water Resour. Res.*, 43, W07301, doi:10.1029/2006WR005467.
- McMillan, H. K., M. P. Clark, W. B. Bowden, M. Duncan, and R. A. Woods (2011), Hydrological field data from a modeller's perspective: Part 1. Diagnostic tests for model structure, *Hydrol. Processes*, 25(4), 511-522.
- Nash, J. E., and J. V. Sutcliffe (1970), River flow forecasting through conceptual models part I - A discussion of principles, *J. Hydrol.*, 10(3), 282-290.
- Nippgen, F., B. L. McGlynn, L. A. Marshall, and R. E. Emanuel (2011), Landscape structure and climate influences on hydrologic response, *Water Resour. Res.*, 47(12), W12528.
- Payn, R. A., M. N. Gooseff, B. L. McGlynn, K. E. Bencala, and S. M. Wondzell (2009), Channel water balance and exchange with subsurface flow along a mountain headwater stream in Montana, United States, *Water Resour. Res.*, 45, W11427, doi:10.1029/2008WR007644.
- Reed, S., V. Koren, M. Smith, Z. Zhang, F. Moreda, and D.-J. Seo (2004), Overall distributed model intercomparison project results, *J. Hydrol.*, 298(1-4), 27-60.
- Sawicz, K., T. Wagener, M. Sivapalan, P. A. Troch, and G. Carrillo (2011), Catchment classification: empirical analysis of hydrologic similarity based on catchment function in the eastern USA, *Hydrol. Earth Syst. Sci.*, 15(9), 2895-2911.
- Schaefli, B., D. B. Talamba, and A. Musy (2007), Quantifying hydrological modeling errors through a mixture of normal distributions, *J. Hydrol.*, 332(3-4), 303-315.
- Sivapalan, M., G. Blöschl, L. Zhang, and R. Vertessy (2003), Downward approach to hydrological prediction, *Hydrol. Processes*, 17(11), 2101-2111.
- Sivapalan, M. (2005), Pattern, Process and Function: Elements of a Unified Theory of Hydrology at the Catchment Scale, in *Encyclopedia of Hydrological Sciences*, edited by M. G. Anderson, John Wiley, Chichester, U.K.
- Smith, T., L. Marshall, B. McGlynn, and K. Jencso (in review). Using field data to inform and evaluate a new model of catchment hydrologic connectivity. Submitted to *Water Resour. Res.*

- Smith, T. J., and L. A. Marshall (2008), Bayesian methods in hydrology: A study of recent advancements in Markov chain Monte Carlo techniques, *Water Resour. Res.*, 44, W00B05, doi:10.1029/2007WR006705.
- Smith, T. J., and L. A. Marshall (2010), Exploring uncertainty and model predictive performance concepts via a modular snowmelt-runoff modeling framework, *Environ. Modell. Software*, 25(6), 691-701.
- Son, K., and M. Sivapalan (2007), Improving model structure and reducing parameter uncertainty in conceptual water balance models through the use of auxiliary data, *Water Resour. Res.*, 43, W01415, doi:10.1029/2006WR005032.
- Staudinger, M., K. Stahl, J. Seibert, M. P. Clark, and L. M. Tallaksen (2011), Comparison of hydrological model structures based on recession and low flow simulations, *Hydrol. Earth Syst. Sci.*, 15(11), 3447-3459.
- Wagener, T. (2003), Evaluation of catchment models, *Hydrol. Processes*, 17(16), 3375-3378.
- Wagener, T., and H. V. Gupta (2005), Model identification for hydrological forecasting under uncertainty, *Stochast. Environ. Res. Risk Assess.*, 19(6), 378-387.
- Wagener, T., M. Sivapalan, P. Troch, and R. Woods (2007), Catchment Classification and Hydrologic Similarity, *Geography Compass*, 1(4), 901-931.

CHAPTER SEVEN

PREDICTING HYDROLOGIC RESPONSE THROUGH A POOLED
CATCHMENT KNOWLEDGEBASEContribution of Author and Co-Authors

Manuscript in Chapter 7

- Author: Tyler Smith
- Contributions: Responsible for the development and analysis of the technique introduced in this study, the majority of the writing, and figure development.
- Co-Author: Lucy Marshall
- Contributions: Provided extensive discussions on the development of the methods used in the study and contributed to the writing and editing of the manuscript throughout its development.
- Co-Author: Ashish Sharma
- Contributions: Provided extensive discussions on the development of the methods used in the study and contributed to the writing and editing of the manuscript throughout its development.

Manuscript Information

Smith, T., L. Marshall, and A. Sharma. Predicting Hydrologic Response Through a Pooled Catchment Knowledgebase, *Water Resources Research*, submitted.

- Journal: Water Resources Research
- Status of manuscript (check one)
 - ☐ Prepared for submission to a peer-reviewed journal
 - ☒ Officially submitted to a peer-reviewed journal
 - ☐ Accepted by a peer-reviewed journal
 - ☐ Published in a peer-reviewed journal
- Publisher: American Geophysical Union
- Date of submission: February 14, 2012

PREDICTING HYDROLOGIC RESPONSE THROUGH A POOLED CATCHMENT KNOWLEDGEBASE

Abstract

Making useful predictions in ungauged basins is an incredibly difficult task given the limitations of hydrologic models to represent physical processes appropriately across a vast array of heterogeneity within and between different catchments. Here, we introduce a new method for this challenge, Bayes Empirical Bayes, that allows for the statistical pooling of information from multiple donor catchments and provides the ability to transfer full parameter distributions rather than single parameter sets to the ungauged catchment. Further, the methodology provides an efficient framework with which to formally assess predictive uncertainty at the ungauged catchment. We investigated the utility of the methodology under both synthetic and real data conditions (utilizing catchments belonging to a set of 240 catchments across Australia), and with respect to its sensitivity to the number and quality of the donor catchments used. This study highlighted the ability of the hierarchical Bayes Empirical Bayes approach to produce expected outcomes in both the synthetic and real data applications. The method was found to be sensitive to the quality (hydrologic similarity) of the donor catchments used. Results were less sensitive to the number of donor catchments, but indicated that predictive uncertainty was best constrained with larger numbers of donor catchments (but still adequate with fewer donors).

Introduction

Hydrologic modeling faces a major challenge when it comes to simulating streamflow in ungauged catchments. Simple conceptual hydrologic models require observed streamflow data to calibrate the parameters to produce reasonable predictions, while complex physics-based hydrologic models can theoretically avoid the need for calibration but require extensive data to inform parameterization. In each case, catchment data is necessary and in the case of the ungauged catchment, data is absent. This problem has spawned an expansive assemblage of research on the topic [e.g., Bárdossy, 2007; Blöschl, 2005; Oudin *et al.*, 2008; Zhang and Chiew, 2009].

In 2003, the International Association of Hydrological Sciences (IAHS) introduced the Predictions in Ungauged Basins (PUB) decadal initiative [*Sivapalan et al.*, 2003], in response to the fact that a majority of the world's catchments are ungauged and a growing realization that the ability of modeling such catchments is important from both operational (forecasting, management, design, etc.) and scientific (catchment functioning, process understanding, etc.) perspectives [*Blöschl*, 2005]. At its core, the PUB initiative was formed around a central focus on the quantification of predictive uncertainty and its reduction for predictions made at ungauged catchments; however, to date a significant proportion of research under the PUB umbrella has focused more on methods aimed at reducing uncertainty [e.g., catchment classification/ regionalization studies; *Archfield and Vogel*, 2010; *Bulygina et al.*, 2011; *Kling and Gupta*, 2009; *Oudin et al.*, 2008; *Reichl et al.*, 2009; *Zhang et al.*, 2008] than those targeted towards quantifying it. Many of these types of studies tend to focus on calibrating a model at a gauged location and transferring the optimal parameter estimates to an adjacent ungauged location deemed to be “similar” by some type of regionalization scheme (geographic proximity, climate, topography, etc.).

Undoubtedly, catchment classification/regionalization studies play an important and vital role in PUB studies. At the same time there is a need for developing a better understanding of how to formally and appropriately quantify uncertainty in predictions at ungauged catchments, as well as the extent to which the uncertainty may be reduced, if at all. Of course there have been some studies focused on addressing this aspect of the PUB initiative [e.g., *Bastola et al.*, 2008; *Bulygina et al.*, 2011; *Zhang et al.*, 2008]. The study

by *McIntyre et al.* [2005], in particular, sought to use formal Bayesian techniques through the implementation of the Bayesian Model Averaging framework to create ensembles resulting from multiple parameter sets to bracket predictions at ungauged catchments.

In this study, the predictions in ungauged basins initiative is explored with respect to developing a useful technique for quantifying predictive uncertainty at ungauged sites rather than focusing on the catchment classification (donor selection) aspect common to many PUB studies. Here, we introduce a Bayesian statistical approach, whereby a joint probability distribution will be computed for the ungauged location based on information from “similar” gauged sites using the so-called Bayes Empirical Bayes (BEB) methodology [*Deely and Lindley*, 1981].

The remainder of the paper will be divided into the following sections: section 2 will formally introduce the Bayes Empirical Bayes approach and emphasize its conceptual underpinnings, section 3 will present the materials (data, model) used throughout the paper, section 4 will detail two case studies (one based on synthetic data, one based on real data), section 5 will provide a discussion of the results, and section 6 will highlight the important conclusions to be drawn from this study.

Bayes Empirical Bayes

The Bayes Empirical Bayes approach was developed as a method for empirically deriving appropriate, non-vague prior distributions for use within Bayesian inference [*Goodman*, 2004]. In order to obtain such distributions, the traditional Bayesian problem is extended to a hierarchical framework that takes advantage of auxiliary data resources.

Here, within the context of the BEB hierarchy, we define a *case* as a unique sample that belongs to a larger, aggregate *population*. With respect to hydrology, a single gauged watershed can be thought of as a *case* that belongs to a *population* of hydrologically similar watersheds. Figure 1 outlines the hierarchical structure used within the BEB technique. Working from the bottom of the figure upward, each case (watershed) has its own forcing data (x_i) that is used to inform its case-specific model parameters (θ_i). Instead of each parameter having its own prior distribution a *global prior* is defined that is common to the parameters across each of the cases. Each case can be thought of as a random sample from this global prior distribution. As an example, for a three-parameter hydrologic model, it would be necessary to define a global prior for each of the three hydrologic parameters which is done only in terms of the distributional form (e.g., parameter θ_1 is defined by a normal distribution with unknown mean and variance, etc.) rather than explicitly specifying both the prior distributional form and prior distribution parameter values as is required under a standard Bayesian approach (e.g., hydrologic parameter θ_1 is defined by a normal prior distribution with mean equal to zero and variance equal to 1.0). The parameters that define the Bayes Empirical Bayes global prior distribution are unknown parameters (referred to as *hyperparameters*, γ) that must also be estimated via the cases. Because the global priors have *hyperparameters* that are to be estimated, these hyperparameters must also be assigned prior distributions (to maintain formal Bayesian principles). The priors on the hyperparameters are known as *hyperpriors* and are commonly assumed to be vague.

Ultimately, the hierarchical BEB architecture seeks to exploit the case-specific data to improve our characterization of the true prior distribution (the global prior) on the hydrologic model parameters. Given this, the following notation has been adopted: let i index the cases (watersheds), x_i the hydrologic forcing data for case i from the ensemble of hydrologic forcing data $\underline{x}$, and the residual model errors for case i as ε_i from the ensemble of errors $\underline{\varepsilon}$. The likelihood for case i is then selected such that its inherent assumptions are adequately representative of the distribution of the errors. The choice of the likelihood function is flexible within the BEB methodology, allowing for different likelihood functions to be used for each of the different cases. The distribution relating the cases to the global prior is defined such that

$$\theta \sim g(\theta|\gamma) \quad (1)$$

where the distribution $g(\theta|\gamma)$ serves as the prior for θ in each case and all variables are as previously defined. The hyperparameters are assigned a prior (the *hyperprior*)

$$\gamma \sim \pi(\gamma|\psi) \quad (2)$$

which is conditional on the fixed parameters of the hyperprior (ψ). The values of ψ are generally chosen such that the hyperprior be vague.

Given these components, the global prior distribution can be estimated under a joint inference on all parameters and hyperparameters. In doing so, the posterior becomes

$$p(\underline{\theta}, \gamma | \underline{x}) \propto \pi(\gamma) \prod_i \{\mathcal{L}_i(x_i | \theta_i) \cdot g(\theta_i | \gamma)\} \quad (3)$$

where the contribution of the posterior for any case depends on the likelihood for the case-specific parameters (both hydrologic model parameters and parameters associated

with the likelihood function itself) and the probability of the case-specific parameters which is conditional on the common hyperparameters.

Ultimately, the objective of this procedure is to obtain an estimate of the global prior distributions for each hydrologic model parameter. The global prior distributions can be thought of as the collective knowledge about the individual model parameters based on the information provided by the gauged, donor watersheds. Using this knowledgebase, this information may be transferred to any ungauged, target watershed in terms of full parameter distributions. These formal probability distributions then allow for the formal consideration/quantification of uncertainty at the ungauged watershed of interest.

Kuczera [1982] employed a similar technique, known as Empirical Bayes [Robbins, 1964], in a regionalization/flood frequency analysis context and found it to be of great potential in making predictions at sites with limited data by pooling the data from similar sites. Although conceptually similar to the BEB method, the Empirical Bayes method is not fully Bayesian in its implementation (giving rise to the name *Bayes* Empirical Bayes).

Methods and Materials

We aimed to explore the utility of the Bayes Empirical Bayes methodology via multiple case studies using both real and synthetic data applications. In each case, we utilize a conceptual rainfall-runoff model to gauged watersheds and investigate the

transferability of model parameters from gauged to “ungauged” sites, when the model parameters are estimated via the BEB architecture.

Model Description

Each case study implements a model based on the probability distributed model (PDM, Figure 2), first developed by *Moore* [1985]. The PDM is a flexible and parsimonious conceptual rainfall-runoff model. As a lumped conceptual model, the PDM is primarily concerned in simulating the frequency of occurrences of hydrologic variables rather than the particular location of such occurrences within the watershed [*Moore*, 1985].

Runoff production simulated by the model is controlled by soil absorption capacity on the basis of the spatial variability of soil capacities across the watershed. The model features two storages: (1) a surface store that receives water in excess of the soil capacity and (2) a subsurface store that receives infiltrated water. The total discharge simulated by the model at the watershed outlet is simply the sum of the outflows from each of the storages.

The hydrologic model then is comprised of six effective parameters: maximum soil storage capacity ($CMAX$), spatial variability within the watershed (B), rate of drainage into subsurface storage (KB), fraction of subsurface storage released to outflow (TRI), fraction of surface storage released to outflow ($TR2$), and soil storage threshold for subsurface outflow (CF).

Parameter Estimation

A two-phase parameter estimation scheme is implemented in the following case studies presented in this paper. The first phase is aimed at locating regions of high posterior density for each of the gauged watersheds serving as *cases* in the BEB architecture and is achieved via the Dynamically Dimensioned Search (DDS) algorithm introduced by *Tolson and Shoemaker* [2007]. The DDS algorithm was selected for the initial phase due to its demonstrated usefulness in finding global optimums of hydrologic model parameters with a minimal number of function evaluations. The second phase uses the optimum parameter values found by the DDS algorithm as initial starting points for the Adaptive Metropolis (AM) algorithm [*Haario et al.*, 2001]. The AM algorithm, a Markov chain Monte Carlo algorithm, is a modification to the standard random walk Metropolis algorithm [*Metropolis et al.*, 1953] and is used in this study to carry out the estimation of the posterior distributions and global prior distributions. Although the DDS step is certainly not required, it was used here to improve the robustness of the sampling algorithm and reduce the potential of convergence to local optima. The steps involved in implementing the AM algorithm for hydrologic problems are discussed by *Marshall et al.* [2004].

The sampling of parameters within the AM algorithm was performed in multiple blocks, where each block corresponded to each of the cases (gauged catchments) with a final block corresponding to the hyperparameters. Under this multi-block sampling scheme, parameters were sampled within one block, accepted or rejected, and then held constant while the next block was sampled; this procedure repeats until each block is

sampled within the current iteration before moving to the next iteration. Such a sampling approach was adopted to reduce the complications associated with estimating large numbers of parameters at once, particularly relating to the definition of the covariance matrix required to draw samples from the multivariate jump distribution and the maintenance of an appropriate acceptance rate.

Test Case Applications

The following section introduces two test cases. The first is a synthetic application aimed at highlighting the utility of the Bayes Empirical Bayes methodology for hydrologic problems (under controlled conditions), while the second extends the approach to an application that features real data. In both cases we apply the logic in terms of addressing the Predictions in Ungauged Basins research initiative's science focus of improved quantification of predictive uncertainty [Sivapalan *et al.*, 2003]. It is not our intent to address the more commonly studied aspect of the PUB initiative, catchment classification/donor watershed selection [e.g., Archfield and Vogel, 2010; Carrillo *et al.*, 2011; Wagener and Wheeler, 2006]. Rather, we begin each of these case studies at the point where suitable donors have been selected based on some metric of similarity; such an approach is justified when considering the fact that no method can be fully vetted if it depends on unreliable components. Here our primary objective is to examine the utility of the BEB approach to appropriately quantify predictive uncertainty at the ungauged location given hydrologically similar donor catchments.

Test Case 1: An Application to Synthetic Data

In this “proof of concept” application, we selected six data sets (each covering five years) from a collection of 240 gauged catchments located across Southeastern Australia (Figure 3) [Vaze *et al.*, 2010]. From the six catchments, we randomly selected one to serve as the ungauged catchment (with the other five to be used as donors) and assigned values for each of the PDM’s parameters. To ensure hydrologic similarity of the catchments (donors and target), the values of the parameters for each of the donors was randomly sampled as $\pm 10\%$ of the ungauged, target catchment’s parameter values. The PDM was run with the true parameter values and the resultant streamflow was corrupted with Gaussian noise ($\varepsilon \sim N(0, 0.1)$) for each of the six catchments. The distributional form of each of the global priors was then defined based on the known constraints of the parameters (refer to Table 1).

Given the synthetic data set (5 years of daily observations, 5 donor catchments, 1 “ungauged” catchment), the methodology outlined in sections 2 and 3 was implemented. First, the gauged, donor catchments were optimized using the DDS algorithm to locate regions of high posterior density (while also confirming the ability of the DDS to quickly locate parameter sets extremely close to the known, true values). Next, the AM algorithm was run using the DDS optimums as starting points. The AM algorithm was used to sample from the posterior distributions for each of the unknown parameters (including the hydrologic model parameters, likelihood function parameters, and global prior distribution parameters). Algorithm convergence was assessed visually based on parameter traces and the known optimum parameter values. Recall that for each of the

five cases (donor catchments), the PDM model was applied to produce estimates of streamflow and thus required six parameters to be estimated (per case). Additionally, the likelihood function features an unknown variance parameter that was also estimated on a per case basis. Finally, the hyperparameters associated with the global prior were also estimated; this represented an additional 13 parameters (Table 1). A total of 48 parameters ($6 \times 5 + 1 \times 5 + 13$) were estimated by the algorithm under the hierarchical Bayesian framework described previously.

Figure 4 shows the global prior distributions for each model parameter constructed based on the optimum values (joint posterior maximum) of the sampled hyperparameters. Included for each parameter distribution is the location of the posterior distribution mode for each case (donor catchment). From this figure, the shape of the global priors for each of the local parameters (hydrologic model parameters & likelihood function parameters) reveals the way in which the different cases are linked to one another via the global prior. Figure 5 highlights the ability of the BEB method to appropriately identify the true model parameters for the ungauged catchment, given hydrologically similar donor catchments. For all model parameters the true value for the ungauged catchment fell within the whiskers of the box plots, with the majority of the true values being within the 25th and 75th percentiles.

Finally, Figure 6 provides information regarding the prediction interval based on the BEB analysis for the ungauged catchment rather than a single model prediction. This figure shows good correspondence between the 90% confidence interval and the observed (true) streamflow for the “ungauged” catchment. The 90% interval captured

96.6% of the observed streamflow, while maintaining a modest mean interval width of 1.13 mm/day.

Test Case 2: An Application to Real Data

In this application, we endeavored to investigate the sensitivity of the BEB method in two ways: (1) with respect to the number of donor catchments and (2) with respect to the hydrologic similarity of the donor catchments. In addressing the first objective, we will compare the sensitivity of the BEB method using five hydrologically similar donors against ten hydrologically similar donors. In the second objective, we will compare BEB performance using (known) hydrologically similar donors with that for randomly selected donors. In each case we will again make use of the collection of 240 gauged catchments located across Southeast Australia (Figure 3) [Vaze *et al.*, 2010] introduced in the previous example.

In order to determine hydrological similarity across the catchments, we first calibrated the PDM for all 240 catchments (most recent 10 years without data gaps) using the DDS algorithm. Following this step, we performed a hierarchical cluster analysis [Hastie *et al.*, 2009] on the DDS optimum parameter sets, where equal weight was assigned to each model parameter in the cluster analysis. This procedure resulted in 30 clusters of hydrologically similar catchments (in terms of PDM parameterization), where the number of nodes was selected to balance the need to have clusters with a suitable number members and the need for the members to be reasonably similar. Figure 7 shows the dendrogram resulting from the cluster analysis, where the y-axis represents the distance between clusters (signified by the height of the linkages), the x-axis represents

cluster nodes (1-30, containing the totality of the 240 catchments), and the clusters are grouped by color based on a distance threshold (< 0.8). The cluster utilized in this test application is further highlighted in the inset of Figure 3, with the sites from the cluster to be used as donors shown by circles and the ungauged site shown by a square.

Following along with the first application, in all cases we will again define the distributional form of each of the global priors based on the known constraints of the parameters of the PDM structure (refer to Table 1), and employ the AM algorithm to carry out parameter estimation (hydrologic, likelihood, hyperparameters) under the BEB framework for each of the separate settings (cluster analysis selected donors; randomly selected donors) until all of the sampled parameters had converged to their stationary, posterior distributions.

In seeking to examine the sensitivity of the BEB methodology to the number of hydrologically similar donor catchments, we will compare the performance of the method using both five and ten donor catchments. The donor catchments and the “ungauged” catchment arise from node 6 of the cluster analysis (Figure 7). This node contained 12 catchments of which one catchment was selected at random to be excluded from the study, another was selected at random to serve as the “ungauged” catchment, and the remaining ten were selected to serve as the donors. The set of ten donors was then randomly sampled to identify the subset of five donors to be used. The DDS optimal parameter sets for the ungauged, the ten donors, and the subset of five donors are given in Table 2.

Recall that the global prior distributions are a summary of the information contributed by the donor catchments in the BEB analysis, and are to be used for transfer to the ungauged catchment. Figure 8 shows box plots of the global prior distributions (based on the optimal set of hyperparameters – the parameters that define the global prior distributions) for the comparison between the five (green) and ten (orange) hydrologically similar donor catchments, with the DDS optimal parameter values for the ungauged site (dashed vertical line) to serve as a reference.

From Figure 8, it is clear that adding more donors does not necessarily lead to more constrained parameter estimates. This result is logical considering that a single donor alone will give an exact parameter estimate. Adding more donors has the advantage of accounting for more extreme behavior (simply by having a larger sample size) and at characterizing the overall predictive uncertainty as a result. At the same time, the method appears to be fairly consistent in terms of the median (central line of box plots) between the five and ten donor levels.

In exploring the sensitivity of the BEB methodology to the hydrological similarity of the donor catchments, we made use of the five donor catchments from the previous part of the study and the same “ungauged” catchment. To serve as a benchmark reference, a separate set of five donor catchments were randomly selected from the reduced catchment population ($n = 118$; see Figure 3) to be used within the BEB framework. The DDS optimum parameter sets for each of the catchments that were selected (1 ungauged target, 5 clustered donors, and 5 random donors) are again given in Table 2.

Again referring to Figure 8, we can see the sensitivity of the method when the donor catchments are not constrained by hydrological similarity (randomly selected donors; blue boxes). Importantly, when the donors are chosen randomly the uncertainty is greatly inflated (represented by the width of the box plots) leading to flat distributions that are largely uninformative. Furthermore, the medians for the randomly selected donors tend to fall much farther from the ungauged optimal values than those for the clustered catchments, as would logically be expected.

In exploring the uncertainty issue further across each of the three scenarios of interest, Figure 9 reveals the 90% confidence intervals (5 random donors, 5 clustered donors, 10 clustered donors) along with the observed streamflow for the “ungauged” catchment for a three-month period during the sixth of ten years used in this application (for enhanced readability). It is visually clear that the interval is much wider for the random donors scenario. Quantitatively, the mean interval width (for the entire ten years) for the 5 random donors, 5 clustered donors, and 10 clustered donors were 1.14, 0.51, and 0.66 mm/day, respectively. At the same time, the percent of observations captured by the 90% intervals (for the entire ten years) were 47.1% for the 5 random donors, 79.6% for the 5 clustered donors, and 86.7% for the 10 clustered donors. These results highlight the over estimation of uncertainty when using the random donors (signified by the mean interval width). Further, the issue of number of donors used is highlighted by the increased interval width with increased number of donors (5 clustered to 10 clustered), which corresponds to an improvement in capture of the observed discharge at the ungauged catchment for the 10 clustered donors relative to the 5 clustered donors.

Discussion

The Predictions in Ungauged Basins research initiative has resulted in a multitude of studies and attention focused on improving our ability to predict streamflow at locations where the hydrologic model cannot be calibrated. A vast majority of these studies have focused on developing ways of identifying catchments that are “hydrologically similar”, while others have attempted to develop strategies for transferring knowledge of the model at gauged locations (deemed hydrologically similar) to the ungauged location. These studies have resulted in an invaluable foundation for PUB, but the issue of uncertainty relies on two components working in unison: (1) methods that quantify the uncertainty and (2) methods that reduce the uncertainty. In order to gain a full understanding of the value of the methods aimed at reducing uncertainty, we first must have reliable methods of formally quantify it. The study presented here aimed to address this via the novel application of a hierarchical Bayesian technique. We utilized both synthetic data and real data from a collection of 240 catchments across Southeast Australia to test the hierarchical Bayesian method’s ability to transfer entire probability distributions for the hydrologic model parameters (rather than single parameter sets), while appropriately quantifying the predictive uncertainty for the predictions made at the ungauged location.

In order to fully assess the utility of the Bayes Empirical Bayes methodology, we began each test application at the point where hydrologically similar donor catchments had been identified. In the synthetic application this was done by simply perturbing the values of the true parameter values for the “ungauged” catchment by $\pm 10\%$. For the

application making use of the real data from the collection of Australian catchments, this was done by calibrating the PDM structure for each catchment using the Dynamically Dimensioned Search algorithm and then performing a hierarchical cluster analysis (Table 2; Figure 7). The result of this procedure was clusters of catchments that are similar in terms of their parameter sets (where each parameter had equal weight in the clustering analysis; Figure 3). At the same time, this provided initial parameter sets of high posterior density to be used as the starting parameter values within the application of the Bayesian inference (reducing possible complications due to poor parameter initialization and time to convergence). In beginning from the point of having hydrologically similar donors we assumed two things: (1) the method studied here is independent of the problem of finding suitable donors, but is rather interested in whether predictive uncertainty is appropriately quantified given good donors and (2) the determination of hydrologically similar donors is an extensive and active research topic in its own right.

The BEB method is founded on the concept of using data to empirically estimate the prior distribution, done via a hierarchical architecture (Figure 1). The method can be summarized as consisting of cases with their own specific data and unknown hydrologic model parameters (the donor catchments in this application) that are related to one another via global prior distributions (that also have unknown “hyperparameters”). In order to carry out this methodology, the global prior distributions must be defined a priori in terms of their form (e.g., Normal, Gamma, Beta, etc.) only as opposed to priors in traditional Bayesian inference that must be fully defined (e.g., Normal with mean and standard deviation). For both the synthetic and real data cases the global prior

distributional forms were defined based on the physical constraints of the parameters themselves (e.g., bounded by zero, bounded by zero and one, etc.) and were given in Table 1.

The synthetic data application was utilized to highlight how the method works and is implemented under controlled conditions (i.e., where input and structural uncertainty are not factors). The posterior mode for each of the five donors was shown superimposed on the global prior distribution (based on hyperparameters corresponding to the maximum posterior density) in Figure 4. This figure qualitatively reveals the similarity of the donor catchments based on the spread of the distribution, where narrow distributions represent congruous information among the donors and wide/flat distributions represent incongruous information. Similarly, this figure shows the differences between the individual donors based on the locations of the individual posterior modes (represented by the symbols in the figure). Figure 5 presented complimentary information in a different way by using box plots. The box plots again highlight the spread in the global prior distribution (a summary of the information provided by the donors that is being transferred to the ungauged catchment), but this time reflects how well this probability distribution transfers to the ungauged site. For this instance, the true parameter values at the ungauged site were captured by the probability distributions computed based on the donor catchments for all parameters. Finally, Figure 6 provided the results in terms of a 90% prediction interval about the observed streamflow at the ungauged catchment with a mean interval width of only 1.13 mm/day that captured 96.6% of the streamflow observations at the ungauged site.

The application to real catchment data was used to investigate not only the usefulness of the BEB method under realistic data conditions and uncertainties, but also to examine the sensitivity of the method to the number and degree of similarity of donor catchments. To that end, Figure 8 and Figure 9 provide an intercomparison of results across three donor scenarios: (1) five randomly selected donors from the full set of catchments; (2) five cluster selected donors; and (3) ten cluster selected donors. In all cases, the same ungauged catchment was used which came from the same cluster as the sets of five and ten clustered donor catchments. Figure 8 highlights the sensitivity of the method to the use of donors that are hydrologically similar to the ungauged catchment, signified by the width of the box plots. These are much wider for the set of randomly selected donors than either of the clustered donor conditions. At the same time, we can see that having too few donors may lead to an underestimation of uncertainty due to a lack of sampling of more “extreme” catchment behavior which is better accounted for with a large sample size. This statement is supported by the visual evidence provided in Figure 9, which shows the 90% confidence interval for each of the donor scenarios. Here, the ten clustered donors resulted in a confidence interval that was best able to capture the observed streamflows at the ungauged site (86.7% versus 79.6% for five clustered and 47.1% for five random) at only slightly greater mean interval width than the five clustered donors (0.66 mm/day versus 0.51 for five clustered and 1.14 for five random). Of course, unlike with the synthetic data application, the real data application necessarily involves additional complications. Among them are the definition of an appropriate hydrologic model structure and a metric for determining hydrologically similar

catchments for the analysis. These two issues are common in hydrologic modeling and lack simple answers. Here, we sought to use a hydrologic model that was flexible enough to adapt to a variety of hydrologic drivers of streamflow across a wide variety of catchments. A simple first step at addressing the appropriateness of a model structure is in terms of its performance. For the cluster used in the real data application the mean cluster Nash-Sutcliffe Efficiency was 0.791. Although by no means definitive of model appropriateness this result supports the premise that the model is at least acceptable for these catchments. A more thorough method for addressing model suitability such as using Bayes factors to assess model appropriateness [see *Marshall et al.*, 2005] could be considered to explore this matter further. In terms of similarity metrics for selecting catchments in the cluster analysis, there are obviously many different approaches that can be adopted. The important thing to recognize is that the clustering results are sensitive to these decisions (i.e., clustering based on normalized parameter values will produce different results to clustering based on raw parameter values). Here we adopted a metric based on normalized parameter values to ensure each parameter carried equal weight in the analysis, otherwise parameters defining storage capacity would undoubtedly dominate the clustering as they have an unbounded upper limit while many of the other parameters are bounded on 0-1.

Conclusions

The results of the hierarchical Bayesian application to the Predictions in Ungauged Basins initiative presented in this study, featuring both a synthetic data application and a real data application drawn from a set of 240 catchments located in Southeast Australia, highlights the potential utility of such an approach under conditions where there is limited to no data available at the catchment of interest but there is data available at other “hydrologically similar” catchments. The hierarchical Bayes Empirical Bayes methodology introduced here provides not only a means of transferring knowledge (parameter sets) from gauged catchments to ungauged catchments but also provides a way to convert this knowledge into statistically formal probability distributions that directly allow for the quantification of uncertainty. This is an important result given the primary science focus of PUB is geared toward improved quantification of predictive uncertainty for simulations at ungauged locations.

In order to best frame the contributions of the BEB methodology for PUB, we focused on starting with “hydrologically similar” donors. This was done in the real data application via an initial calibration of the PDM structure for each of the 240 catchments using the Dynamically Dimensioned Search algorithm, followed by a hierarchical clustering analysis to group catchments based on the similarity of the optimized model parameters. The Adaptive Metropolis algorithm was used to carry out all Bayesian inference.

The results of the test applications highlight the utility and feasibility of using hierarchical Bayesian approaches to combine and summarize the information from

multiple donor catchments into a single joint probability distribution that can be easily and effectively transferred to the ungauged catchment, while providing statistically formal estimates of predictive uncertainty. Such an approach represents a significant departure from methods that rely on regression techniques to transfer parameter sets to the ungauged catchment and overcomes the limitations of these methods at quantifying uncertainty.

Future work is planned to further investigate the use of this technique on additional sets of catchments and using more physically realistic model representations. Additionally, work is planned to integrate the selection of hydrologically similar catchments into future studies rather than using donors that have been predetermined to be hydrologically similar.

Acknowledgements

This research was funded by the Australian Research Council (DP0879218) and the National Science Foundation (EAR-0943640). The first author received additional support as a Graduate Fellow of the Montana Institute on Ecosystems. We are thankful to Dan Goodman and George Kuczera for their insight regarding the Bayes Empirical Bayes methodology during the initial stages of this research. Finally, we thank Francis Chiew and Jai Vaze for sharing the streamflow data used in this study.

Tables

Table 1. A Description of Global Priors Used Under the Hierarchical Bayesian Modeling Framework

Parameter	Value Range	Prior	Hyperparameters
C_{MAX}	> 0 ($> CF$)	$C_{MAX} \sim \text{Gamma}$ (A_1, A_2)	A_1 (shape); A_2 (scale)
B	> 0	$B \sim \text{Gamma}$ (B_1, B_2)	B_1 (shape); B_2 (scale)
KB	$0 < KB < 1$	$KB \sim \text{Beta}$ (C_1, C_2)	C_1 (shape); C_2 (shape)
$TR1$	$0 < TR1 < 1$ ($< TR2$)	$TR1 \sim \text{Beta}$ (D_1, D_2)	D_1 (shape); D_2 (shape)
$TR2$	$0 < TR2 < 1$ ($> TR1$)	$TR2 \sim \text{Beta}$ (E_1, E_2)	E_1 (shape); E_2 (shape)
CF	> 0 ($< C_{MAX}$)	$CF \sim \text{Gamma}$ (F_1, F_2)	F_1 (shape); F_2 (scale)
VAR (likelihood variance)	> 0	$VAR \sim \text{Exp}$ (G_1)	G_1 (rate)

Table 2. A Description of the Catchments Selected for Analysis in Test Case 2^a

	Catchment ID ^b	DDS Optimal Parameter Sets					
		<i>C_{MAX}</i>	<i>B</i>	<i>KB</i>	<i>TR1</i>	<i>TR2</i>	<i>CF</i>
<i>Clustered Catchments (5^c & 10 donors)</i>	<i>C1^c</i>	356.4	0.024	0.014	0.018	0.538	341.0
	<i>C2^c</i>	392.4	0.016	0.016	0.011	0.502	382.1
	<i>C3</i>	351.7	0.009	0.016	0.017	0.641	345.8
	<i>C4</i>	357.5	0.019	0.013	0.007	0.709	342.2
	<i>C5</i>	399.6	0.026	0.004	4.1×10^{-4}	0.492	385.3
	<i>C6</i>	399.6	0.018	0.009	0.017	0.455	390.4
	<i>C7^c</i>	398.7	0.024	0.017	0.027	0.532	392.1
	<i>C8</i>	381.3	0.040	0.003	0.029	0.409	354.5
	<i>C9^c</i>	371.1	0.016	0.004	0.024	0.599	360.5
	<i>C10^c</i>	399.4	0.044	0.024	2.0×10^{-4}	0.526	396.8
<i>Random Catchments (5 donors)</i>	<i>R1</i>	328.7	0.033	0.026	0.010	0.640	328.1
	<i>R2</i>	392.4	0.016	0.016	0.011	0.502	382.1
	<i>R3</i>	268.7	0.102	0.042	0.007	0.261	257.7
	<i>R4</i>	317.5	0.058	0.060	0.003	0.107	309.1
	<i>R5</i>	373.2	0.257	0.111	0.021	0.305	226.2
<i>Target Catchment</i>	<i>UG</i>	396.0	0.021	0.007	0.029	0.536	385.4

^aHierarchical clustering was performed on DDS optimal parameter sets with normalized parameter values, giving all parameters equal weight in the analysis.

^bWhere IDs *C1-C10* represent the gauged, donor catchments from the cluster analysis, IDs *R1-R5* represent the gauged, donor catchments selected randomly, and ID *UG* represents the ungauged, target catchment.

^cFive randomly selected catchments from the set of ten clustered donors to be used for the five clustered donors scenario.

Figures

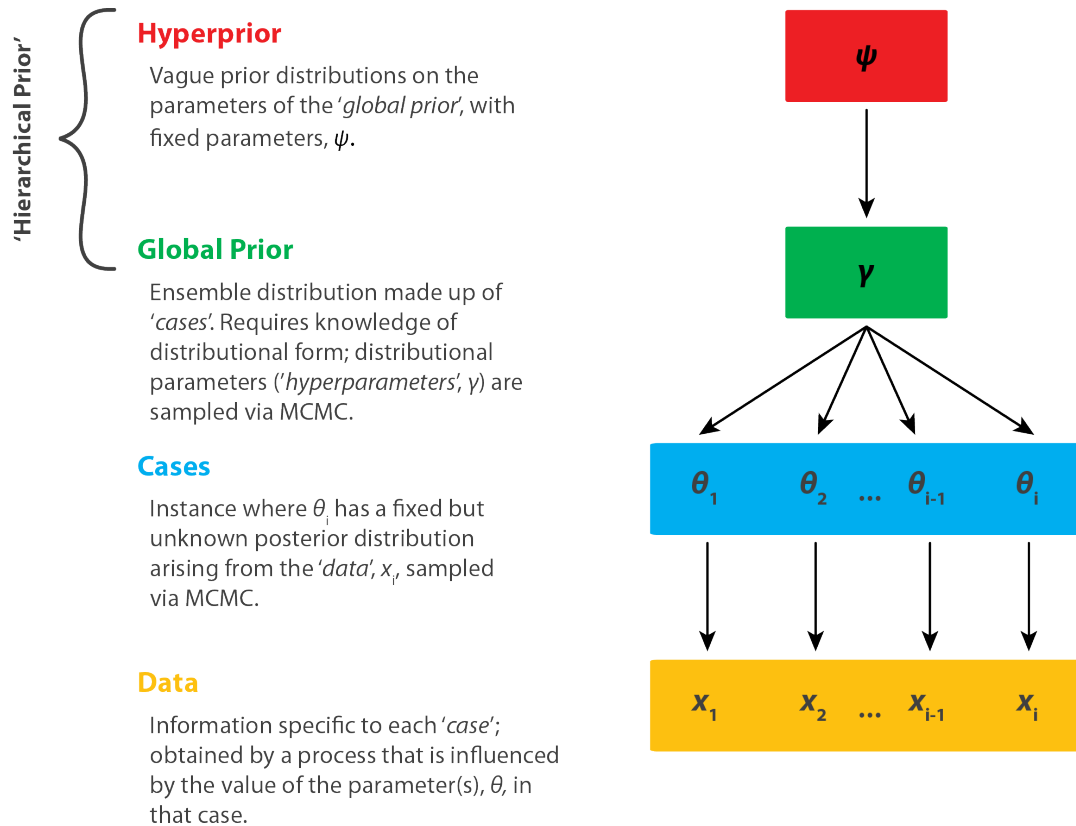


Figure 1. Diagram outlining the hierarchical architecture of the Bayes Empirical Bayes methodology.

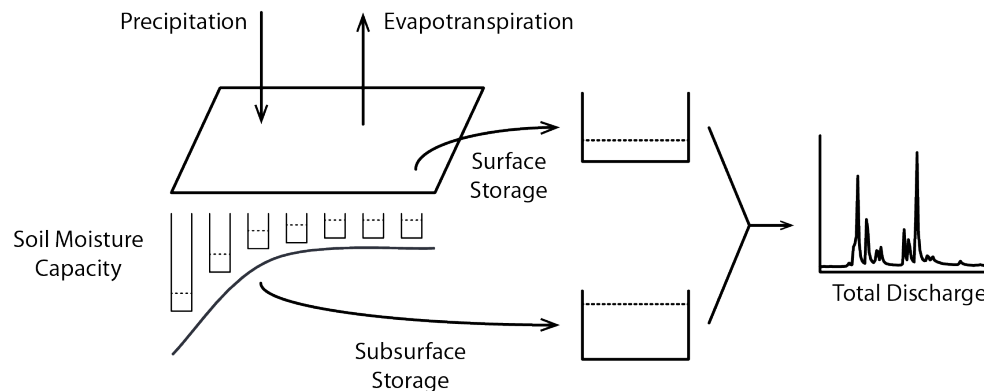


Figure 2. Schematic diagram of the probability distributed model (PDM).

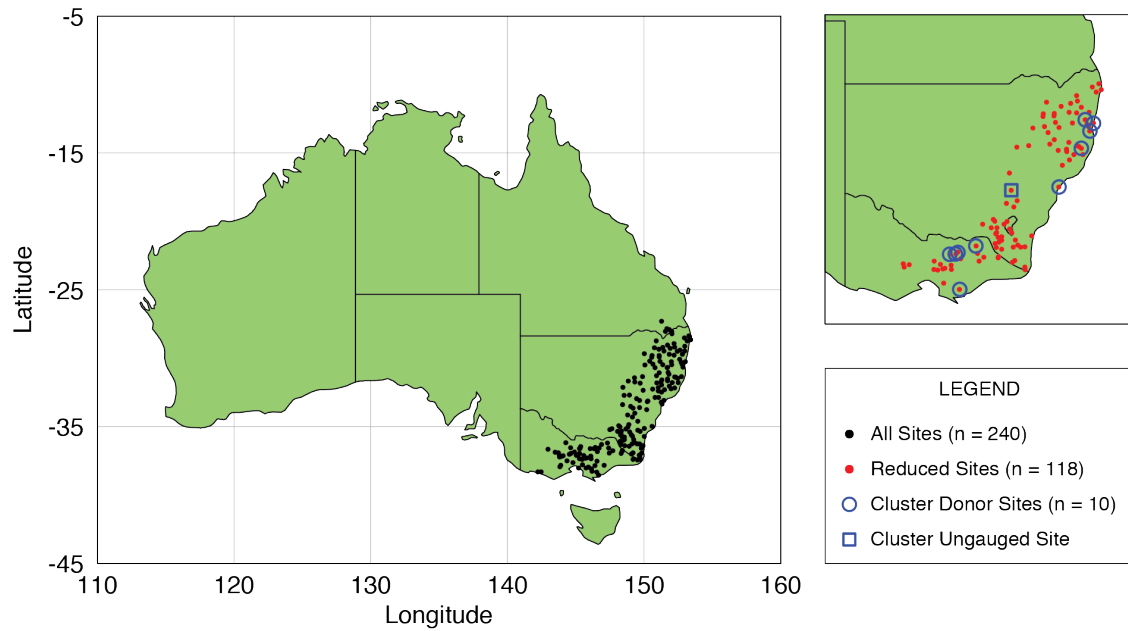


Figure 3. Location of the catchments used in this study. (Left) Each of the 240 catchments in the dataset is shown. (Right) The reduced set of 118 catchments are shown (non-ephemeral with at least ten years of continuous data), along with the sites used as donors for the ungauged target catchment based on hierarchical cluster analysis.

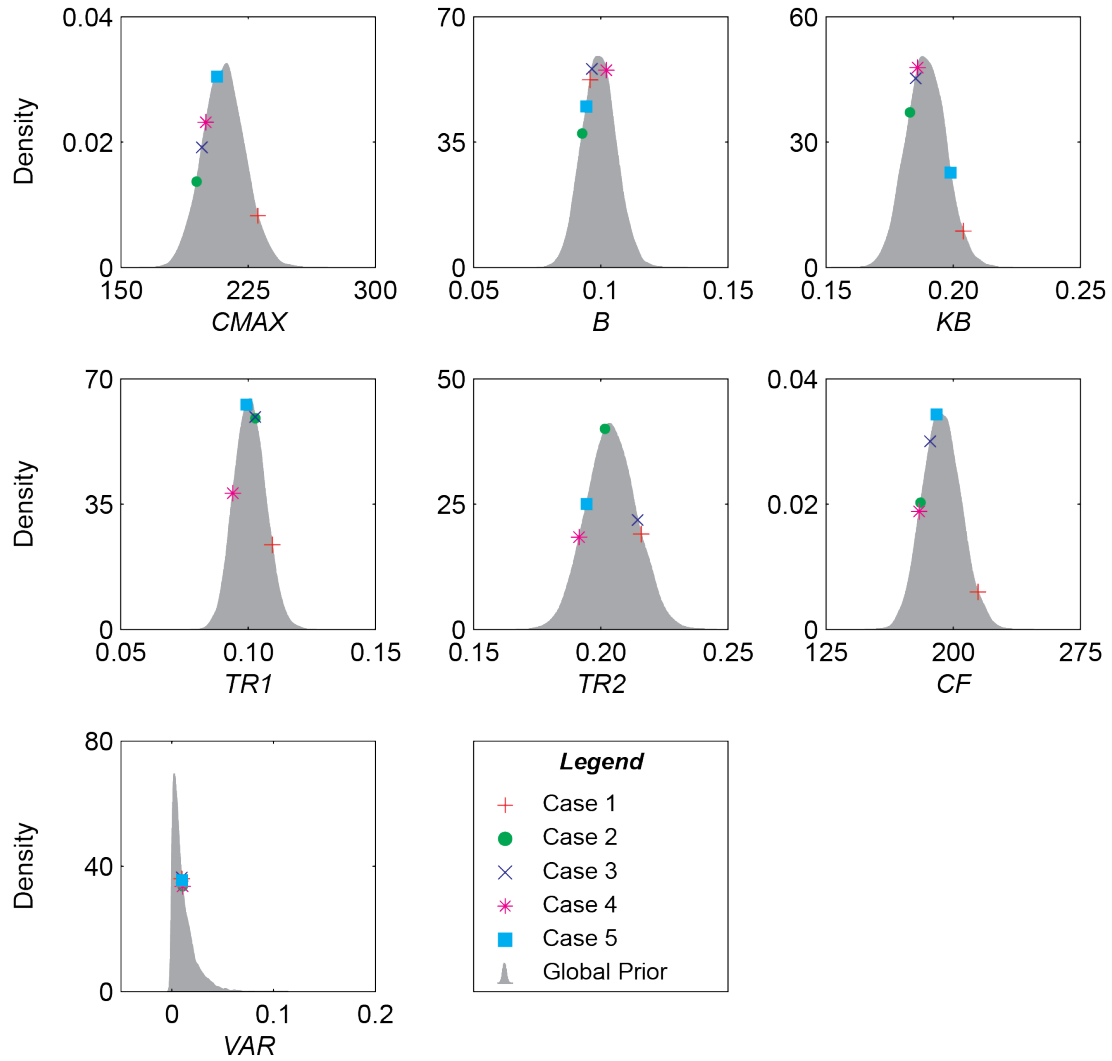


Figure 4. Plot displaying the posterior mode for each parameter and each case, superimposed on the global prior distribution estimated based on the composite of the cases for the synthetic test application.

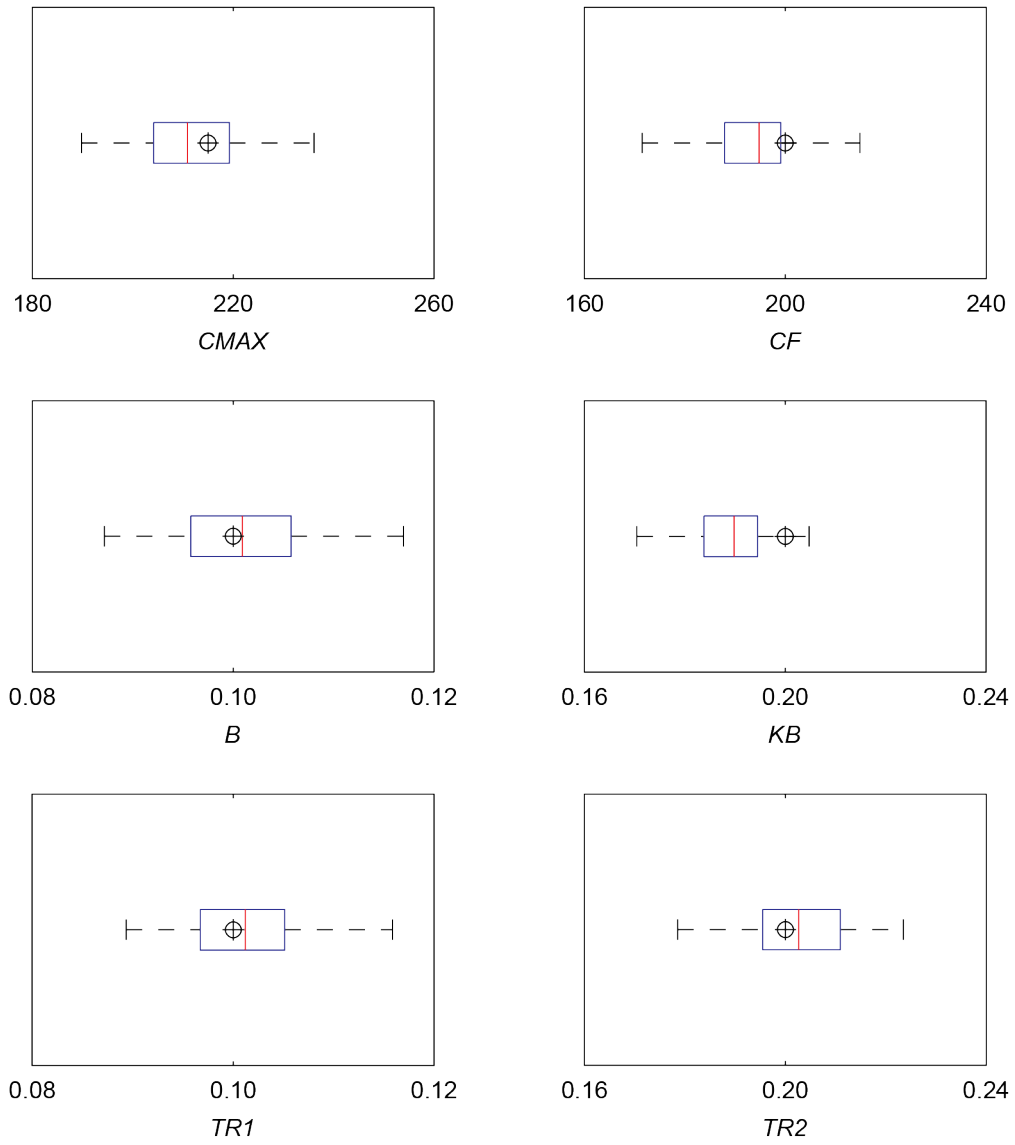


Figure 5. Box plots of the global prior distributions (based on “optimal” hyperparameters) with the true parameter values for the “ ungauged ” catchment shown with the target symbol. Note that the true parameter value is captured within the whiskers ($\pm 2.7\sigma$; 99.3% coverage) in all cases and within the 25th and 75th percentiles for parameters *CMAX*, *B*, *TR1*, and *TR2*.

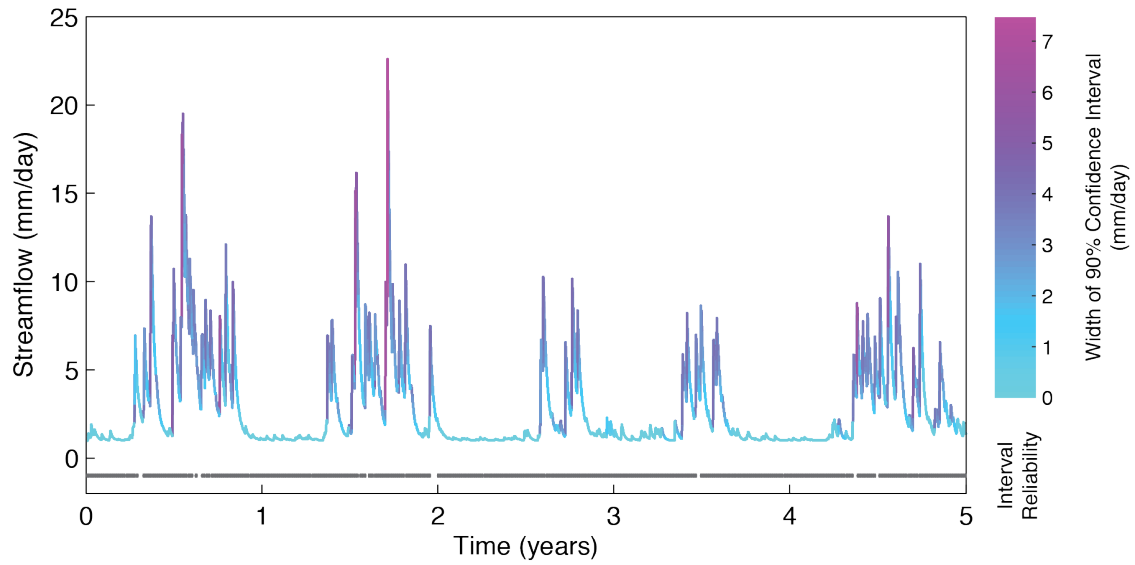


Figure 6. Results of the BEB analysis for the parameter transfer to the “ ungauged ” catchment for the synthetic data application. The plot shows the observed hydrograph of the “ ungauged ” catchment, the 90% confidence interval width shown using the corresponding color gradient, and the interval reliability (96.6% observations captured) represented as a binary plot (present = interval captured observation; absent = interval failed to capture observation).

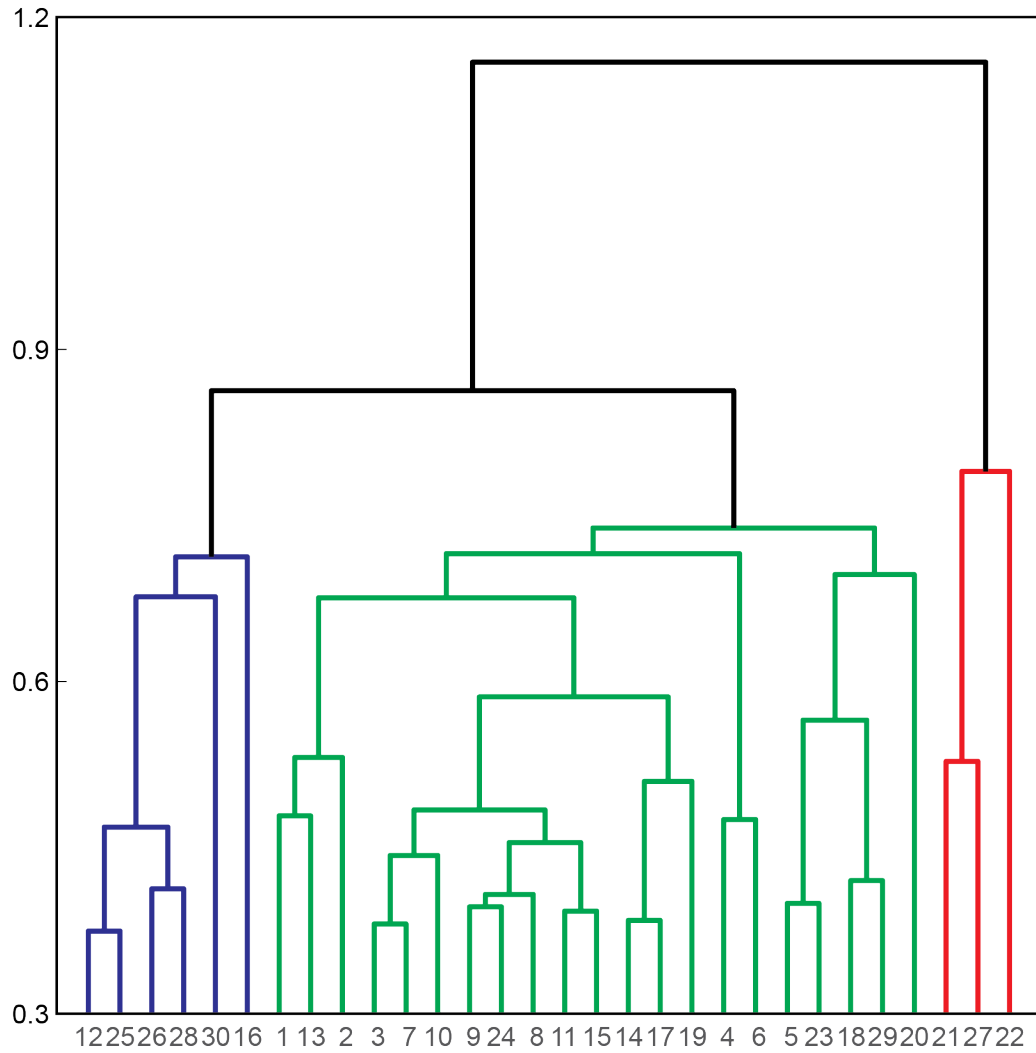


Figure 7. A dendrogram illustrating the results of the hierarchical clustering analysis based on the optimum parameter sets determined with the DDS algorithm. Note the x-axis indicates nodes (1-30) that contain all of the 240 catchments in the data set and the y-axis represents the relative distance between nodes (based on linkage height). For Test Case 2, the ungauged catchment and the catchments corresponding to the five and ten cluster selected scenarios were contained within node 6.

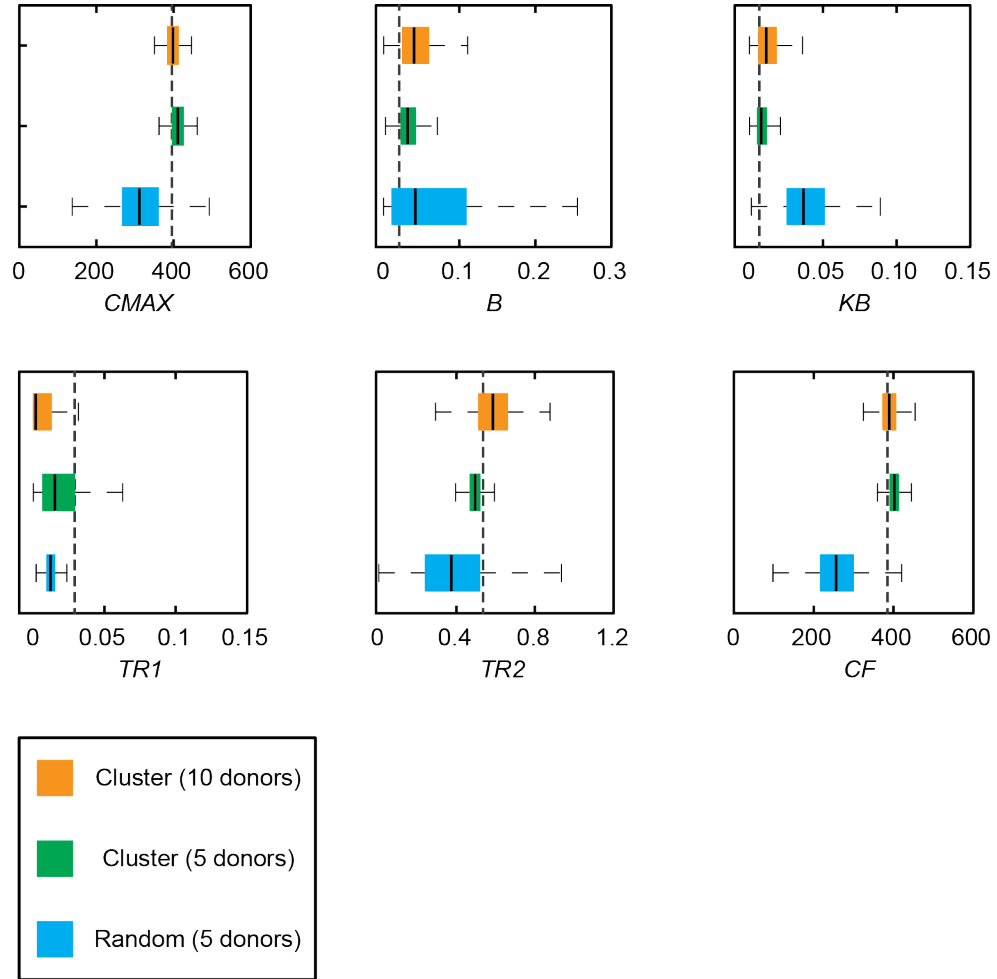


Figure 8. Box plots of the global prior distributions (based on “optimal” hyperparameters) for each of the scenarios (5 randomly selected donors; 5 cluster selected donors; 10 cluster selected donors) examined in the real data test application, with the optimal parameter values for the “ungauged” catchment (obtained with DDS algorithm) shown by the vertical dashed line in each subplot.

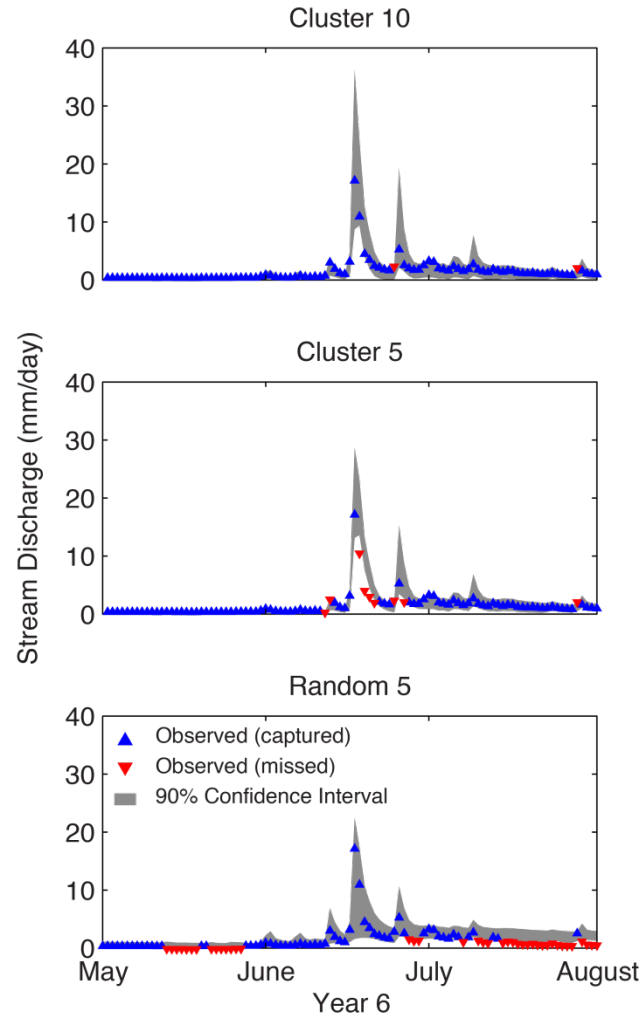


Figure 9. Plots of the 90% confidence intervals and observed streamflow for the “ungauged” catchment for a subset of the real data used in the study. The blue triangles represent the observations captured by the 90% interval and the inverted red triangles represent the observations missed by the 90% interval.

References

- Archfield, S. A., and R. M. Vogel (2010), Map correlation method: Selection of a reference streamgage to estimate daily streamflow at ungauged catchments, *Water Resour. Res.*, 46, W10513, doi:10.1029/2009WR008481.
- Bárdossy, A. (2007), Calibration of hydrological model parameters for ungauged catchments, *Hydrol. Earth Syst. Sci.*, 11(2), 703-712.
- Bastola, S., H. Ishidaira, and K. Takeuchi (2008), Regionalisation of hydrological model parameters under parameter uncertainty: A case study involving TOPMODEL and basins across the globe, *J. Hydrol.*, 357(3-4), 188-206.
- Blöschl, G. (2005), Rainfall-Runoff Modeling of Ungauged Catchments, in *Encyclopedia of Hydrological Sciences*, edited by M. G. Anderson, John Wiley, Chichester, U.K.
- Bulygina, N., N. McIntyre, and H. Wheeler (2011), Bayesian conditioning of a rainfall-runoff model for predicting flows in ungauged catchments and under land use changes, *Water Resour. Res.*, 47, W02503, doi:10.1029/2010WR009240.
- Carrillo, G., P. A. Troch, M. Sivapalan, T. Wagener, C. Harman, and K. Sawicz (2011), Catchment classification: hydrological analysis of catchment behavior through process-based modeling along a climate gradient, *Hydrol. Earth Syst. Sci.*, 15(11), 3411-3430.
- Deely, J. J., and D. V. Lindley (1981), Bayes Empirical Bayes, *J. Am. Stat. Assoc.*, 76(376), 833-841.
- Goodman, D. (2004), Taking the Prior Seriously: Bayesian Analysis without Subjective Probability, in *The Nature of Scientific Evidence: Statistical, Philosophical, and Empirical Considerations*, edited by M. L. Taper and S. R. Lele, pp. 379-409, The University of Chicago Press, Chicago.
- Haario, H., E. Saksman, and J. Tamminen (2001), An adaptive Metropolis algorithm, *Bernoulli*, 7(2), 223-242.
- Hastie, T., R. Tibshirani, and J. Friedman (2009), *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, 2nd ed., 745 pp., Springer-Verlag.
- Kling, H., and H. Gupta (2009), On the development of regionalization relationships for lumped watershed models: The impact of ignoring sub-basin scale variability, *J. Hydrol.*, 373(3-4), 337-351.

- Kuczera, G. (1982), Combining Site-Specific and Regional Information: An Empirical Bayes Approach, *Water Resour. Res.*, 18(2), 306-314.
- Marshall, L., D. Nott, and A. Sharma (2004), A comparative study of Markov chain Monte Carlo methods for conceptual rainfall-runoff modeling, *Water Resour. Res.*, 40, W02501, doi:10.1029/2003WR002378.
- Marshall, L., D. Nott, and A. Sharma (2005), Hydrological model selection: A Bayesian alternative, *Water Resour. Res.*, 41, W10422, doi:10.1029/2004WR003719.
- McIntyre, N., H. Lee, H. Wheeler, A. Young, and T. Wagener (2005), Ensemble predictions of runoff in ungauged catchments, *Water Resour. Res.*, 41, W12434, doi:10.1029/2005WR004289.
- Metropolis, N., A. W. Rosenbluth, N. Rosenbluth, A. H. Teller, and E. Teller (1953), Equations of state calculations by fast computing machines, *J. Chem. Phys.*, 21(6), 1087-1092.
- Moore, R. J. (1985), The probability-distributed principle and runoff production at point and basin scales, *Hydrol. Sci. J.*, 30(2), 273-297.
- Oudin, L., V. Andréassian, C. Perrin, C. Michel, and N. Le Moine (2008), Spatial proximity, physical similarity, regression and ungauged catchments: A comparison of regionalization approaches based on 913 French catchments, *Water Resour. Res.*, 44, W03413, doi:10.1029/2007WR006240.
- Reichl, J. P. C., A. W. Western, N. R. McIntyre, and F. H. S. Chiew (2009), Optimization of a similarity measure for estimating ungauged streamflow, *Water Resour. Res.*, 45, W10423, doi:10.1029/2008WR007248.
- Robbins, H. (1964), The empirical Bayes approach to statistical decision problems, *Annals of Mathematical Statistics*, 35(1), 1-20.
- Sivapalan, M., K. Takeuchi, S. W. Franks, V. K. Gupta, H. Karambiri, V. Lakshmi, X. Liang, J. J. McDonnell, E. M. Mendenso, P. E. O'Connell, T. Oki, J. W. Pomeroy, D. Schertzer, S. Uhlenbrook, and E. Zehe (2003), IAHS Decade on Predictions in Ungauged Basins (PUB), 2003-2012: Shaping an exciting future for the hydrological sciences, *Hydrol. Sci. J.*, 48(6), 857-880.
- Tolson, B. A., and C. A. Shoemaker (2007), Dynamically dimensioned search algorithm for computationally efficient watershed model calibration, *Water Resour. Res.*, 43, W01413, doi:10.1029/2005WR004723.
- Vaze, J., D. A. Post, F. H. S. Chiew, J. M. Perraud, N. R. Viney, and J. Teng (2010), Climate non-stationarity – Validity of calibrated rainfall-runoff models for use in climate change studies, *J. Hydrol.*, 394(3-4), 447-457.

- Wagener, T., and H. S. Wheater (2006), Parameter estimation and regionalization for continuous rainfall-runoff models including uncertainty, *J. Hydrol.*, 320(1-2), 132-154.
- Zhang, Y., and F. H. S. Chiew (2009), Relative merits of different methods for runoff predictions in ungauged catchments, *Water Resour. Res.*, 45, W07412, doi:10.1029/2008WR007504.
- Zhang, Z., T. Wagener, P. Reed, and R. Bhushan (2008), Reducing uncertainty in predictions in ungauged basins by combining hydrologic indices, regionalization and multiobjective optimization, *Water Resour. Res.*, 44, W00B04, doi:10.1029/2008WR006833.

CONCLUSION

Summary

The challenge of making accurate predictions of streamflow using hydrologic models is naturally complicated by uncertainties associated with the model input data, the model structure, and the model parameters. Furthermore, any attempts to model streamflow are generally limited by the availability of streamflow observations that are needed to estimate the model parameters to begin with. Bayesian inference offers an attractive framework within which hydrologic modeling studies can be conducted due to its ability to account for all the associated uncertainties (in a statistically formal way).

The main focus of this dissertation has been three-fold: (1) to investigate how to more appropriately apply the Bayesian framework through an investigation into its inherent assumptions; (2) to explore hydrologic model development by allowing experimental observations drive model formulation; and (3) to improve our ability to make reasonable hydrologic predictions of streamflow where observations are not available to use for parameter estimation. The following paragraphs summarize the major findings of each chapter.

In Chapter 2, we investigated the use of Bayesian analysis in hydrologic modeling studies of ephemeral catchments. In particular, we were interested in better understanding how the statistical assumptions of the Bayesian approach manifested themselves and impacted model performance and uncertainty estimates for catchments that exhibit distinct flow/no-flow states. It was found that failing to satisfy the statistical assumptions

of the likelihood function resulted in a failure to identify true values of known parameters (for synthetic data applications) and improper quantification of uncertainty estimates. Effects on model performance were variable, with degradations in performance in some cases and little change in others. This finding suggested the necessity for a multifaceted assessment strategy.

Building on the findings of Chapter 2, in Chapter 3 we extended the informal procedures for testing and identifying a “good” likelihood function into a formal procedure applicable across catchment conditions (ephemeral, perennial). The ultimate objective of this study was to build a framework that reliably guides the modeler to the simplest likelihood function whose statistical assumptions are reasonably satisfied for the given hydrologic modeling study. The study included components related to commonly violated assumptions (in hydrologic modeling) and showed good performance across both synthetic and real data applications. At the same time, the method was built in a generic way, allowing for easy adaptation and extension to include additional components as needed.

In Chapter 4, we explored the usage of transformations in hydrologic modeling as a means of satisfying statistical assumptions (which are violated in regular space but satisfied in transformed space). We compared and contrasted the commonly used Box-Cox transformation with our flow-corrected time transformation which was previously unused for hydrologic model calibration purposes. We found that the transformation was useful in de-skewing (often) positively skewed streamflow data better than the Box-Cox transformation. This resulted in statistical assumptions being more adequately satisfied,

leading to superior model inference. However, transformation does favor high flows to low flows resulting in limitations when low flows are of primary interest.

In Chapter 5, the focus of interest is shifted from the Bayesian implementation of hydrologic models to the development of the model itself. We began with the study of extensive experimental field observations at the Tenderfoot Creek Experimental Forest that provided the template upon which a new model structure was founded. We sought to emphasize the concept of doing hydrology (modeling) forward; that is, starting with observations and constructing a model based on them rather than the commonly employed method of selecting a model (often based on personal preference rather than appropriateness) in a one-size fits all manner (without regard for empirical understanding of catchment functioning). We created a simple and efficient hydrologic model, the Catchment Connectivity Model, that was able to reproduce both streamflow observations and hydrologic connectivity.

Chapter 6 continues with the Catchment Connectivity Model in an application that investigated its utility in being transferred to catchments beyond that for which it was developed. Additional case studies were considered to better understand the model's sensitivity to input data arising from various stations located within the larger study area and to alternate conceptualizations of the overland flow mechanism. It was found that the model showed consistent performance (both to streamflow and internal variables – hydrologic connectivity) across eight different catchments. The model was found to be somewhat sensitive to the inputs, especially in terms of streamflow predictions and less so in terms of hydrologic connectivity predictions. On the other hand, the model showed

strong sensitivities to alternate conceptualizations of overland flow; in particular it was found that streamflow performance could be improved but often at the expense of consistency to internal variables. This was particularly true when known unrealistic conceptualizations were tested.

Finally, in Chapter 7, we sought to further the study of model transferability as it relates to predictions in ungauged catchments. Specifically, we were interested in investigating whether a hierarchical Bayesian approach would be useful in quantifying and constraining predictive uncertainty at the ungauged catchment. In doing so, we made use of an extensive collection of more than 200 catchments located throughout Southeast Australia. We tested the method for both synthetic and real data cases and with respect to its sensitivities to the quality (hydrologic similarity) and quantity (number) of gauged donor catchments. We found that the hierarchical Bayesian approach was able to accurately quantify predictive uncertainty that captured a high percentage of the ungauged catchments streamflow. The method was found to be sensitive to donor quality (as expected) and to a lesser extent to the number of donors used.

Hydrologic modeling is continually confronted with the challenge of synthesizing heterogeneous findings arising from detailed studies into the physical controls on streamflow generation across a vast range of study catchments. This is an especially daunting task given the need for hydrologic models to be flexible and transferable across multiple catchments. The chapters of this dissertation sought to address many of the most pressing issues common across hydrological modeling studies through insights into the usage of statistical techniques for parameter estimation and uncertainty analysis, the

development of transferable and reliable model structures, and for applications where typical methodologies are limited due to the unavailability of previously measured streamflow data. The chapters also represent a unique examination of two disparate data sets, one made up of basic measurements over 200+ Australian catchments and the other made up of highly detailed measurements over just a few catchments in the Tenderfoot Creek Experimental Forest of Montana, USA.

Limitations and Recommendations for Future Research

The research presented in this dissertation explored hydrologic modeling through insights into many of the components upon which modeling is founded. As a result of this, limitations of the methods and potential for future studies have been uncovered. The limitations of the research presented here are primarily related to computational difficulties and data availability, while future research is aimed at exploiting the computational power and data that do exist to further improve our models and predictions.

The value of the Bayesian inferential framework detailed in Chapters 2 and 3 (and to a lesser extent in Chapter 4) was shown to be dependent upon the proper description of the model residuals via the likelihood function. For hydrologic modeling applications this can pose particular difficulties. However, the framework for selecting appropriate likelihood functions was shown to be extendable and adaptable. *Schoops and Vrugt* [2010] introduced a likelihood function that avoids the underlying assumption of residual normality (commonly attempted to achieve via transformations). Such approaches, which

can be built into the framework introduced in Chapter 3, show promise for future research into improving the ability to accurately satisfy the statistical assumptions made in Bayesian inference. Studies into the development of a formal likelihood function that makes use of the flow-corrected time transformation presented hydrologic model calibration in Chapter 4 would also be of benefit to the expanding suite of likelihood functions available to users of Bayesian analysis.

In terms of model development, we focused in Chapters 5 and 6 on model development by using extensive empirical observations to form the basis of the Catchment Connectivity Model [*Smith et al.*, in review]. These studies showed consistent and reliable model simulations for both external variables (streamflow) and internal variables (hydrologic connectivity). An obvious limitation to this approach to model development is that it is only possible when extensive observations are available to begin with. This fact highlights the importance of conducting targeted studies that are useful to a range of scientists across disciplines to maximize the value of the studies [*McDonnell et al.*, 2007]. At the same time, these chapters highlighted the potential for future work with the CCM structure itself. Elements of particular interest include: (1) an extensive testing of alternate forms of the threshold-based storage function that determines the hydrologic connectivity of the modeled landscape units and (2) modeling using the CCM at catchments progressively further from the catchment conditions for which it was developed (e.g., a flat-land catchment). In each case, these future studies would help to set the limits of applicability of the model structure and develop a better understanding of

the CCM structure's strengths and weaknesses, while potentially providing insights into that applicability of a wider range of hydrologic models.

Finally, in Chapter 7 we concentrated on improving upon the current ability to make meaningful predictions in ungauged catchments via an approach that attempts to better quantify predictive uncertainty. Being able to make predictions in ungauged locations is obviously of societal importance, particularly in areas where the cost of maintaining an infrastructure of an observational network is unrealistic [*Sivapalan, 2003*]. The Bayes Empirical Bayes (BEB) [*Deely and Lindley, 1981*] approach, a hierarchical Bayesian method, was introduced as a means of improving upon the ability to properly quantify predictive uncertainty in predictions made at the ungauged site. In the application presented in this chapter, we relied upon an extensive network of gauged catchments and a generic hydrologic model to find suitable donor catchments. This is a clear limitation to the method. In future work, we hope to integrate the studies at the Tenderfoot Creek Experimental Forest (Chapters 5 & 6) with this work to investigate the utility of the BEB approach with a more limited number of catchments but using a much more tailored hydrologic model (the CCM). This research would have the potential of both improving our understanding of the spatiotemporal sensitivities of the BEB approach and also further our understanding of the utility of models developed based on observation to those constructed in a more flexible and generic way. Additionally, the clustering method used as a preliminary step for finding hydrologically similar catchments in the study presented in Chapter 7 offers potential for looking at the problem of catchment classification from an alternative viewpoint. A common approach in

catchment classification is to use similar catchment characteristics (area, slope, elevation, climate, etc.) to infer similar hydrologic model parameters. This of course assumes that the hydrologic model parameters are somehow related to easily attainable catchment characteristics. There exists strong potential for studies that target this problem from the opposite perspective, starting with a calibrated model across a large number of catchments and using clustering to group them based on model parameters. This approach would then make it possible to assess the similarity of the physical catchment characteristics within each cluster in an attempt to determine the characteristics that truly do result in similar hydrologic functioning (via the model).

In this dissertation, the three primary topics of (1) statistical implementation of models under uncertainty, (2) hydrologic model development, and (3) model applicability/transferability under data scarcity were considered. As a result, the research presented here has addressed numerous aspects involved in improving hydrologic model performance, whose cumulative contribution results in a better understanding of how to construct, implement, and apply hydrologic models to lead to improved water resource management and planning problems of societal importance.

References

- Deely, J. J., and D. V. Lindley (1981), Bayes Empirical Bayes, *J. Am. Stat. Assoc.*, 76(376), 833-841.
- McDonnell, J. J., M. Sivapalan, K. Vaché, S. Dunn, G. Grant, R. Haggerty, C. Hinz, R. Hooper, J. Kirchner, M. L. Roderick, J. Selker, and M. Weiler (2007), Moving beyond heterogeneity and process complexity: A new vision for watershed hydrology, *Water Resour. Res.*, 43, W07301, doi:10.1029/2006WR005467.
- Schoups, G., and J. A. Vrugt (2010), A formal likelihood function for parameter and predictive inference of hydrologic models with correlated, heteroscedastic, and non-Gaussian errors, *Water Resour. Res.*, 46, W10531, doi:10.1029/2009WR008933.
- Sivapalan, M. (2003), Prediction in ungauged basins: a grand challenge for theoretical hydrology, *Hydrol. Processes*, 17(15), 3163-3170.
- Smith, T., L. Marshall, B. McGlynn, and K. Jencso (in review). Using field data to inform and evaluate a new model of catchment hydrologic connectivity. Submitted to *Water Resour. Res.*

REFERENCES

- Aitchison, J. (1955), On the Distribution of a Positive Random Variable Having a Discrete Probability Mass at the Origin, *J. Am. Stat. Assoc.*, 50(271), 901-908.
- Ajami, N. K., Q. Duan, and S. Sorooshian (2007), An integrated hydrologic Bayesian multimodel combination framework: Confronting input, parameter, and model structural uncertainty in hydrologic prediction, *Water Resour. Res.*, 43, W01403, doi:10.1029/2005WR004745.
- Anderson, M. G., and T. P. Burt (1978), The role of topography in controlling throughflow generation, *Earth Surface Processes*, 3(4), 331-344.
- Archfield, S. A., and R. M. Vogel (2010), Map correlation method: Selection of a reference streamgage to estimate daily streamflow at ungauged catchments, *Water Resour. Res.*, 46, W10513, doi:10.1029/2009WR008481.
- Bárdossy, A. (2007), Calibration of hydrological model parameters for ungauged catchments, *Hydrol. Earth Syst. Sci.*, 11(2), 703-712.
- Bastola, S., H. Ishidaira, and K. Takeuchi (2008), Regionalisation of hydrological model parameters under parameter uncertainty: A case study involving TOPMODEL and basins across the globe, *J. Hydrol.*, 357(3-4), 188-206.
- Bates, B. C., and E. P. Campbell (2001), A Markov chain Monte Carlo scheme for parameter estimation and inference in conceptual rainfall-runoff modeling, *Water Resour. Res.*, 37(4), 937-947.
- Beasley, R. S. (1976), Contribution of Subsurface Flow from the Upper Slopes of Forested Watersheds to Channel Flow, *Soil Sci. Soc. Am. J.*, 40(6), 955-957.
- Betson, R. P. (1964), What Is Watershed Runoff?, *J. Geophys. Res.*, 69(8), 1541-1552.
- Beven, K. (1978), The hydrological response of headwater and sideslope areas, *Hydrol. Sci. Bull.*, 23(4), 419-437.
- Beven, K. (1989), Changing ideas in hydrology - The case of physically-based models, *J. Hydrol.*, 105(1-2), 157-172.
- Beven, K., and A. Binley (1992), The future of distributed models: Model calibration and uncertainty prediction, *Hydrol. Processes*, 6(3), 279-298.
- Beven, K. (2006), A manifesto for the equifinality thesis, *J. Hydrol.*, 320(1-2), 18-36.
- Beven, K. (2006), On undermining the science?, *Hydrol. Processes*, 20(14), 3141-3146.

- Beven, K., P. Smith, and J. Freer (2007), Comment on "Hydrological forecasting uncertainty assessment: Incoherence of the GLUE methodology" by Pietro Mantovan and Ezio Todini, *J. Hydrol.*, 338(3-4), 315-318.
- Beven, K. J., and M. J. Kirkby (1979), A physically based, variable contributing area model of basin hydrology, *Hydrol. Sci. Bull.*, 24(1), 43-69.
- Beven, K. J., P. J. Smith, and J. E. Freer (2008), So just why would a modeller choose to be incoherent?, *J. Hydrol.*, 354(1-4), 15-32.
- Blöschl, G., and M. Sivapalan (1995), Scale issues in hydrological modeling: A review, *Hydrol. Processes*, 9(3-4), 251-290.
- Blöschl, G. (2005), Rainfall-Runoff Modeling of Ungauged Catchments, in *Encyclopedia of Hydrological Sciences*, edited by M. G. Anderson, John Wiley, Chichester, U.K.
- Blöschl, G. (2006), Hydrologic synthesis: Across processes, places, and scales, *Water Resour. Res.*, 42, W03S02, doi:10.1029/2005WR004319.
- Blöschl, G., and A. Montanari (2010), Climate change impacts - throwing the dice?, *Hydrol. Processes*, 24(3), 374-381.
- Bonell, M., J. J. McDonnell, F. N. Scatena, J. Seibert, S. Uhlenbrook, and H. A. J. van Lanen (2006), HELPing FRIENDs in PUBs: charting a course for synergies within international water research programmes in gauged and ungauged basins, *Hydrol. Processes*, 20(8), 1867-1874.
- Boughton, W. (2004), The Australian water balance model, *Environ. Modell. Software*, 19(10), 943-956.
- Box, G. E. P., and D. R. Cox (1964), An Analysis of Transformations, *J. Roy. Stat. Soc. B Meth.*, 26(2), 211-252.
- Box, G. E. P., and G. C. Tiao (1973), *Bayesian Inference in Statistical Analysis*, Addison-Wesley, Reading, MA.
- Bulygina, N., N. McIntyre, and H. Wheeler (2011), Bayesian conditioning of a rainfall-runoff model for predicting flows in ungauged catchments and under land use changes, *Water Resour. Res.*, 47, W02503, doi:10.1029/2010WR009240.
- Butts, M. B., J. T. Payne, M. Kristensen, and H. Madsen (2004), An evaluation of the impact of model structure on hydrological modelling uncertainty for streamflow simulation, *J. Hydrol.*, 298(1-4), 242-266.

- Campbell, E. P., D. R. Fox, and B. C. Bates (1999), A Bayesian approach to parameter estimation and pooling in nonlinear flood event models, *Water Resour. Res.*, 35(1), 211-220.
- Carrillo, G., P. A. Troch, M. Sivapalan, T. Wagener, C. Harman, and K. Sawicz (2011), Catchment classification: hydrological analysis of catchment behavior through process-based modeling along a climate gradient, *Hydrol. Earth Syst. Sci.*, 15(11), 3411-3430.
- Clark, M. P., A. G. Slater, D. E. Rupp, R. A. Woods, J. A. Vrugt, H. V. Gupta, T. Wagener, and L. E. Hay (2008), Framework for Understanding Structural Errors (FUSE): A modular framework to diagnose differences between hydrological models, *Water Resour. Res.*, 44, W00B02, doi:10.1029/2007WR006735.
- Clark, M. P., and D. Kavetski (2010), Ancient numerical demons of conceptual hydrological modeling: 1. Fidelity and efficiency of time stepping schemes, *Water Resour. Res.*, 46, W10510, doi:10.1029/2009WR008894.
- Clark, M. P., D. Kavetski, and F. Fenicia (2011), Pursuing the method of multiple working hypotheses for hydrological modeling, *Water Resour. Res.*, 47, W09301, doi:10.1029/2010wr009827.
- Clark, M. P., H. K. McMillan, D. B. G. Collins, D. Kavetski, and R. A. Woods (2011), Hydrological field data from a modeller's perspective: Part 2: process-based evaluation of model hypotheses, *Hydrol. Processes*, 25(4), 523-543.
- Deely, J. J., and D. V. Lindley (1981), Bayes Empirical Bayes, *J. Am. Stat. Assoc.*, 76(376), 833-841.
- Duan, Q., S. Sorooshian, and V. Gupta (1992), Effective and efficient global optimization for conceptual rainfall-runoff models, *Water Resour. Res.*, 28(4), 1015-1031.
- Duan, Q., J. Schaake, V. Andréassian, S. Franks, G. Goteti, H. V. Gupta, Y. M. Gusev, F. Habets, A. Hall, L. Hay, T. Hogue, M. Huang, G. Leavesley, X. Liang, O. N. Nasonova, J. Noilhan, L. Oudin, S. Sorooshian, T. Wagener, and E. F. Wood (2006), Model Parameter Estimation Experiment (MOPEX): An overview of science strategy and major results from the second and third workshops, *J. Hydrol.*, 320(1-2), 3-17.
- Duan, Q., N. K. Ajami, X. Gao, and S. Sorooshian (2007), Multi-model ensemble hydrologic prediction using Bayesian model averaging, *Adv. Water Resour.*, 30(5), 1371-1386.
- Dunn, S. M., J. Freer, M. Weiler, M. J. Kirkby, J. Seibert, P. F. Quinn, G. Lischeid, D. Tetzlaff, and C. Soulsby (2008), Conceptualization in catchment modelling: simply learning?, *Hydrol. Processes*, 22(13), 2389-2393.

- Dunne, T., and R. D. Black (1970), Partial Area Contributions to Storm Runoff in a Small New England Watershed, *Water Resour. Res.*, 6(5), 1296-1311.
- Dunne, T. (1983), Relation of field studies and modeling in the prediction of storm runoff, *J. Hydrol.*, 65(1-3), 25-48.
- Emanuel, R. E., H. E. Epstein, B. L. McGlynn, D. L. Welsch, D. J. Muth, and P. D'Odorico (2010), Spatial and temporal controls on watershed ecohydrology in the northern Rocky Mountains, *Water Resour. Res.*, 46, W11553, doi:10.1029/2009WR008890.
- Engman, E. T., and A. S. Rogowski (1974), A Partial Area Model for Storm Flow Synthesis, *Water Resour. Res.*, 10(3), 464-472.
- Fenicia, F., J. J. McDonnell, and H. H. G. Savenije (2008), Learning from model improvement: On the contribution of complementary data to process understanding, *Water Resour. Res.*, 44, W06419, doi:10.1029/2007WR006386.
- Fenicia, F., D. Kavetski, and H. H. G. Savenije (2011), Elements of a flexible approach for conceptual hydrological modeling: 1. Motivation and theoretical development, *Water Resour. Res.*, 47(11), W11510.
- Filliben, J. J. (1975), The Probability Plot Correlation Coefficient Test for Normality, *Technometrics*, 17(1), 111-117.
- Freeze, R. A. (1972), Role of Subsurface Flow in Generating Surface Runoff 2. Upstream Source Areas, *Water Resour. Res.*, 8(5), 1272-1283.
- Freeze, R. A. (1972), Role of subsurface flow in generating surface runoff: 1. Base flow contributions to channel flow, *Water Resour. Res.*, 8(3), 609-623.
- Gelman, A., and D. B. Rubin (1992), Inference from iterative simulation using multiple sequences, *Stat. Sci.*, 7(4), 457-472.
- Gelman, A., J. B. Carlin, H. S. Stern, and D. B. Rubin (2004), *Bayesian Data Analysis*, 2nd ed., Chapman & Hall/CRC, New York.
- Goodman, D. (2004), Taking the Prior Seriously: Bayesian Analysis without Subjective Probability, in *The Nature of Scientific Evidence: Statistical, Philosophical, and Empirical Considerations*, edited by M. L. Taper and S. R. Lele, pp. 379-409, The University of Chicago Press, Chicago.
- Grabs, T. J., K. G. Jencso, B. L. McGlynn, and J. Seibert (2010), Calculating terrain indices along streams: A new method for separating stream sides, *Water Resour. Res.*, 46, W12536, doi:10.1029/2010WR009296.

- Gupta, H. V., S. Sorooshian, and P. O. Yapo (1998), Toward improved calibration of hydrologic models: Multiple and noncommensurable measures of information, *Water Resour. Res.*, 34(4), 751-763.
- Gupta, H. V., T. Wagener, and Y. Liu (2008), Reconciling theory with observations: elements of a diagnostic approach to model evaluation, *Hydrol. Processes*, 22(18), 3802-3813.
- Gupta, H. V., and H. Kling (2011), On typical range, sensitivity, and normalization of Mean Squared Error and Nash-Sutcliffe Efficiency type metrics, *Water Resour. Res.*, 47, W10601, doi:10.1029/2011wr010962.
- Haario, H., E. Saksman, and J. Tamminen (2001), An adaptive Metropolis algorithm, *Bernoulli*, 7(2), 223-242.
- Hastie, T., R. Tibshirani, and J. Friedman (2009), *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, 2nd ed., 745 pp., Springer-Verlag.
- Hewlett, J. D., and A. R. Hibbert (1967), Factors affecting the response of small watersheds to precipitation in humid areas, in *International Symposium on Forest Hydrology*, edited by W. E. Sopper and H. W. Lull, pp. 275-290, Pergamon, Oxford, UK.
- Horton, R. E. (1933), The role of infiltration in the hydrologic cycle, *Trans. Am. Geophys. Un.*, 14, 446-460.
- Hsu, K.-l., H. Moradkhani, and S. Sorooshian (2009), A sequential Bayesian approach for hydrologic model selection and prediction, *Water Resour. Res.*, 45, W00B12, doi:10.1029/2008WR006824.
- Huard, D., and A. Mailhot (2008), Calibration of hydrological model GR2M using Bayesian uncertainty analysis, *Water Resour. Res.*, 44, W02424, doi:10.1029/2007WR005949.
- Jencso, K. G., B. L. McGlynn, M. N. Gooseff, S. M. Wondzell, K. E. Bencala, and L. A. Marshall (2009), Hydrologic connectivity between landscapes and streams: Transferring reach- and plot-scale understanding to the catchment scale, *Water Resour. Res.*, 45, W04428, doi:10.1029/2008WR007225.
- Jencso, K. G., B. L. McGlynn, M. N. Gooseff, K. E. Bencala, and S. M. Wondzell (2010), Hillslope hydrologic connectivity controls riparian groundwater turnover: Implications of catchment structure for riparian buffering and stream water sources, *Water Resour. Res.*, 46, W10524, doi:10.1029/2009WR008818.

- Jencso, K. G., and B. L. McGlynn (2011), Hierarchical controls on runoff generation: Topographically driven hydrologic connectivity, geology, and vegetation, *Water Resour. Res.*, 47(11), W11527.
- Kavetski, D., S. W. Franks, and G. Kuczera (2003), Confronting input uncertainty in environmental modelling, in *Calibration of Watershed Models*, edited by Q. Duan, et al., pp. 49-68, American Geophysical Union.
- Kavetski, D., G. Kuczera, and S. W. Franks (2006), Bayesian analysis of input uncertainty in hydrological modeling: Theory, *Water Resour. Res.*, 42, W03407, doi:10.1029/2005WR004368.
- Kavetski, D., and M. P. Clark (2010), Ancient numerical demons of conceptual hydrological modeling: 2. Impact of time stepping schemes on model analysis and prediction, *Water Resour. Res.*, 46, W10511, doi:10.1029/2009WR008896.
- Kavetski, D., and F. Fenicia (2011), Elements of a flexible approach for conceptual hydrological modeling: 2. Application and experimental insights, *Water Resour. Res.*, 47(11), W11511.
- Kirchner, J. W. (2006), Getting the right answers for the right reasons: Linking measurements, analyses, and models to advance the science of hydrology, *Water Resour. Res.*, 42, W03S04, doi:10.1029/2005WR004362.
- Klemeš, V. (1983), Conceptualization and scale in hydrology, *J. Hydrol.*, 65(1-3), 1-23.
- Klemeš, V. (1986), Operational testing of hydrological simulation models, *Hydrol. Sci. J.*, 31(1), 13-24.
- Kling, H., and H. Gupta (2009), On the development of regionalization relationships for lumped watershed models: The impact of ignoring sub-basin scale variability, *J. Hydrol.*, 373(3-4), 337-351.
- Kuczera, G. (1982), Combining Site-Specific and Regional Information: An Empirical Bayes Approach, *Water Resour. Res.*, 18(2), 306-314.
- Kuczera, G. (1983), Improved Parameter Inference in Catchment Models 1. Evaluating Parameter Uncertainty, *Water Resour. Res.*, 19(5), 1151-1162.
- Kuczera, G., and E. Parent (1998), Monte Carlo assessment of parameter uncertainty in conceptual catchment models: the Metropolis algorithm, *J. Hydrol.*, 211(1-4), 69-85.
- Kuczera, G., D. Kavetski, S. Franks, and M. Thyer (2006), Towards a Bayesian total error analysis of conceptual rainfall-runoff models: Characterising model error using storm-dependent parameters, *J. Hydrol.*, 331(1-2), 161-177.

- Ljung, G. M., and G. E. P. Box (1978), On a Measure of Lack of Fit in Time Series Models, *Biometrika*, 65(2), 297-303.
- Madsen, H., G. Wilson, and H. C. Ammentorp (2002), Comparison of different automated strategies for calibration of rainfall-runoff models, *J. Hydrol.*, 261(1-4), 48-59.
- Manabe, S. (1969), Climate and the Ocean Circulation: 1. The Atmospheric Circulation and the Hydrology of the Earth's Surface, *Mon. Weather Rev.*, 97(11), 739-774.
- Mantovan, P., and E. Todini (2006), Hydrological forecasting uncertainty assessment: Incoherence of the GLUE methodology, *J. Hydrol.*, 330(1-2), 368-381.
- Mantovan, P., E. Todini, and M. L. V. Martina (2007), Reply to comment by Keith Beven, Paul Smith and Jim Freer on "Hydrological forecasting uncertainty assessment: Incoherence of the GLUE methodology", *J. Hydrol.*, 338(3-4), 319-324.
- Marshall, L., D. Nott, and A. Sharma (2004), A comparative study of Markov chain Monte Carlo methods for conceptual rainfall-runoff modeling, *Water Resour. Res.*, 40, W02501, doi:10.1029/2003WR002378.
- Marshall, L., D. Nott, and A. Sharma (2005), Hydrological model selection: A Bayesian alternative, *Water Resour. Res.*, 41, W10422, doi:10.1029/2004WR003719.
- Marshall, L., A. Sharma, and D. Nott (2006), Modeling the catchment via mixtures: Issues of model specification and validation, *Water Resour. Res.*, 42, W11409, doi:10.1029/2005WR004613.
- McDonnell, J. J., and R. Woods (2004), On the need for catchment classification, *J. Hydrol.*, 299(1-2), 2-3.
- McDonnell, J. J., M. Sivapalan, K. Vaché, S. Dunn, G. Grant, R. Haggerty, C. Hinz, R. Hooper, J. Kirchner, M. L. Roderick, J. Selker, and M. Weiler (2007), Moving beyond heterogeneity and process complexity: A new vision for watershed hydrology, *Water Resour. Res.*, 43, W07301, doi:10.1029/2006WR005467.
- McGuire, K. J., and J. J. McDonnell (2006), A review and evaluation of catchment transit time modeling, *J. Hydrol.*, 330(3-4), 543-563.
- McIntyre, N., H. Lee, H. Wheeler, A. Young, and T. Wagener (2005), Ensemble predictions of runoff in ungauged catchments, *Water Resour. Res.*, 41, W12434, doi:10.1029/2005WR004289.

- McMillan, H. K., M. P. Clark, W. B. Bowden, M. Duncan, and R. A. Woods (2011), Hydrological field data from a modeller's perspective: Part 1. Diagnostic tests for model structure, *Hydrol. Processes*, 25(4), 511-522.
- Metropolis, N., A. W. Rosenbluth, N. Rosenbluth, A. H. Teller, and E. Teller (1953), Equations of state calculations by fast computing machines, *J. Chem. Phys.*, 21(6), 1087-1092.
- Milly, P. C. D., J. Betancourt, M. Falkenmark, R. M. Hirsch, Z. W. Kundzewicz, D. P. Lettenmaier, and R. J. Stouffer (2008), Stationarity Is Dead: Whither Water Management?, *Science*, 319(5863), 573-574.
- Montanari, A. (2005), Large sample behaviors of the generalized likelihood uncertainty estimation (GLUE) in assessing the uncertainty of rainfall-runoff simulations, *Water Resour. Res.*, 41, W08406, doi:10.1029/2004WR003826.
- Montanari, A., C. A. Shoemaker, and N. van de Giesen (2009), Introduction to special section on Uncertainty Assessment in Surface and Subsurface Hydrology: An overview of issues and challenges, *Water Resour. Res.*, 45, W00B00, doi:10.1029/2009WR008471.
- Moore, R. J. (1985), The probability-distributed principle and runoff production at point and basin scales, *Hydrol. Sci. J.*, 30(2), 273-297.
- Moore, R. J. (2007), The PDM rainfall-runoff model, *Hydrol. Earth Syst. Sci.*, 11(1), 483-499.
- Nash, J. E., and J. V. Sutcliffe (1970), River flow forecasting through conceptual models part I - A discussion of principles, *J. Hydrol.*, 10(3), 282-290.
- Nippgen, F., B. L. McGlynn, L. A. Marshall, and R. E. Emanuel (2011), Landscape structure and climate influences on hydrologic response, *Water Resour. Res.*, 47(12), W12528.
- Oudin, L., V. Andréassian, C. Perrin, C. Michel, and N. Le Moine (2008), Spatial proximity, physical similarity, regression and ungaged catchments: A comparison of regionalization approaches based on 913 French catchments, *Water Resour. Res.*, 44, W03413, doi:10.1029/2007WR006240.
- Payn, R. A., M. N. Gooseff, B. L. McGlynn, K. E. Bencala, and S. M. Wondzell (2009), Channel water balance and exchange with subsurface flow along a mountain headwater stream in Montana, United States, *Water Resour. Res.*, 45, W11427, doi:10.1029/2008WR007644.
- Peel, M. C., F. H. S. Chiew, A. W. Western, and T. A. McMahon (2000), Extension of unimpaired monthly streamflow data and regionalisation of parameter values to

- estimate streamflow in ungauged catchments, in *National Land and Water Resources Audit*, edited, CRC for Catchment Hydrology, Melbourne.
- Ragan, R. M. (1968), An experimental investigation of partial area contributions, in *Hydrological Aspects of the Utilization of Water, Reports and Discussions (Proc. IAHS Assembly at Bern)*, edited, pp. 241–251.
- Reed, S., V. Koren, M. Smith, Z. Zhang, F. Moreda, and D.-J. Seo (2004), Overall distributed model intercomparison project results, *J. Hydrol.*, 298(1-4), 27-60.
- Reichert, P., and J. Mieleitner (2009), Analyzing input and structural uncertainty of nonlinear dynamic models with stochastic, time-dependent parameters, *Water Resour. Res.*, 45, W10402, doi:10.1029/2009wr007814.
- Reichl, J. P. C., A. W. Western, N. R. McIntyre, and F. H. S. Chiew (2009), Optimization of a similarity measure for estimating ungauged streamflow, *Water Resour. Res.*, 45, W10423, doi:10.1029/2008WR007248.
- Renard, B., D. Kavetski, G. Kuczera, M. Thyer, and S. W. Franks (2010), Understanding predictive uncertainty in hydrologic modeling: The challenge of identifying input and structural errors, *Water Resour. Res.*, 46, W05521, doi:10.1029/2009WR008328.
- Reusser, D. E., and E. Zehe (2011), Inferring model structural deficits by analyzing temporal dynamics of model performance and parameter sensitivity, *Water Resour. Res.*, 47, W07550, doi:10.1029/2010wr009946.
- Riveros-Iregui, D. A., B. L. McGlynn, L. A. Marshall, D. L. Welsch, R. E. Emanuel, and H. E. Epstein (2011), A watershed-scale assessment of a process soil CO₂ production and efflux model, *Water Resour. Res.*, 47, W00J04, doi:10.1029/2010wr009941.
- Robbins, H. (1964), The empirical Bayes approach to statistical decision problems, *Annals of Mathematical Statistics*, 35(1), 1-20.
- Robert, C. P. (1996), Mixtures of distributions: inference and estimation, in *Markov Chain Monte Carlo in Practice*, edited by W. R. Gilks, et al., pp. 441-464, Chapman & Hall, London.
- Rodhe, A., L. Nyberg, and K. Bishop (1996), Transit Times for Water in a Small Till Catchment from a Step Shift in the Oxygen 18 Content of the Water Input, *Water Resour. Res.*, 32(12), 3497-3511.
- Samanta, S., D. S. Mackay, M. K. Clayton, E. L. Kruger, and B. E. Ewers (2007), Bayesian analysis for uncertainty estimation of a canopy transpiration model, *Water Resour. Res.*, 43, W04424, doi:10.1029/2006WR005028.

- Samanta, S., M. K. Clayton, D. S. Mackay, E. L. Kruger, and B. E. Ewers (2008), Quantitative comparison of canopy conductance models using a Bayesian approach, *Water Resour. Res.*, *44*, W09431, doi:10.1029/2007WR006761.
- Sawicz, K., T. Wagener, M. Sivapalan, P. A. Troch, and G. Carrillo (2011), Catchment classification: empirical analysis of hydrologic similarity based on catchment function in the eastern USA, *Hydrol. Earth Syst. Sci.*, *15*(9), 2895-2911.
- Schaefli, B., D. B. Talamba, and A. Musy (2007), Quantifying hydrological modeling errors through a mixture of normal distributions, *J. Hydrol.*, *332*(3-4), 303-315.
- Schoups, G., and J. A. Vrugt (2010), A formal likelihood function for parameter and predictive inference of hydrologic models with correlated, heteroscedastic, and non-Gaussian errors, *Water Resour. Res.*, *46*, W10531, doi:10.1029/2009WR008933.
- Seibert, J., and J. J. McDonnell (2002), On the dialog between experimentalist and modeler in catchment hydrology: Use of soft data for multicriteria model calibration, *Water Resour. Res.*, *38*, 1241, doi:10.1029/2001WR000978.
- Seibert, J., and B. L. McGlynn (2007), A new triangular multiple flow direction algorithm for computing upslope areas from gridded digital elevation models, *Water Resour. Res.*, *43*, W04501, doi:10.1029/2006WR005128.
- Sivapalan, M. (2003), Prediction in ungauged basins: a grand challenge for theoretical hydrology, *Hydrol. Processes*, *17*(15), 3163-3170.
- Sivapalan, M., G. Blöschl, L. Zhang, and R. Vertessy (2003), Downward approach to hydrological prediction, *Hydrol. Processes*, *17*(11), 2101-2111.
- Sivapalan, M., K. Takeuchi, S. W. Franks, V. K. Gupta, H. Karambiri, V. Lakshmi, X. Liang, J. J. McDonnell, E. M. Mendiondo, P. E. O'Connell, T. Oki, J. W. Pomeroy, D. Schertzer, S. Uhlenbrook, and E. Zehe (2003), IAHS Decade on Predictions in Ungauged Basins (PUB), 2003-2012: Shaping an exciting future for the hydrological sciences, *Hydrol. Sci. J.*, *48*(6), 857-880.
- Sivapalan, M. (2005), Pattern, Process and Function: Elements of a Unified Theory of Hydrology at the Catchment Scale, in *Encyclopedia of Hydrological Sciences*, edited by M. G. Anderson, John Wiley, Chichester, U.K.
- Sklash, M. G., and R. N. Farvolden (1979), The role of groundwater in storm runoff, *J. Hydrol.*, *43*(1-4), 45-65.
- Smith, T., A. Sharma, L. Marshall, R. Mehrotra, and S. Sisson (2010), Development of a formal likelihood function for improved Bayesian inference of ephemeral catchments, *Water Resour. Res.*, *46*, W12551, doi:10.1029/2010WR009514.

- Smith, T., L. Marshall, B. McGlynn, and K. Jencso (in review). Using field data to inform and evaluate a new model of catchment hydrologic connectivity. Submitted to *Water Resour. Res.*
- Smith, T. J., and L. A. Marshall (2008), Bayesian methods in hydrology: A study of recent advancements in Markov chain Monte Carlo techniques, *Water Resour. Res.*, 44, W00B05, doi:10.1029/2007WR006705.
- Smith, T. J., and L. A. Marshall (2010), Exploring uncertainty and model predictive performance concepts via a modular snowmelt-runoff modeling framework, *Environ. Modell. Software*, 25(6), 691-701.
- Son, K., and M. Sivapalan (2007), Improving model structure and reducing parameter uncertainty in conceptual water balance models through the use of auxiliary data, *Water Resour. Res.*, 43, W01415, doi:10.1029/2006WR005032.
- Sorooshian, S., and J. A. Dracup (1980), Stochastic parameter estimation procedures for hydrologic rainfall-runoff models: Correlated and heteroscedastic error cases, *Water Resour. Res.*, 16(2), 430-442.
- Staudinger, M., K. Stahl, J. Seibert, M. P. Clark, and L. M. Tallaksen (2011), Comparison of hydrological model structures based on recession and low flow simulations, *Hydrol. Earth Syst. Sci.*, 15(11), 3447-3459.
- Stedinger, J. R., R. M. Vogel, S. U. Lee, and R. Batchelder (2008), Appraisal of the generalized likelihood uncertainty estimation (GLUE) method, *Water Resour. Res.*, 44, W00B06, doi:10.1029/2008WR006822.
- Thiemann, M., M. Trosset, H. V. Gupta, and S. Sorooshian (2001), Bayesian recursive parameter estimation for hydrologic models, *Water Resour. Res.*, 37(10), 2521-2535.
- Thyer, M., G. Kuczera, and Q. J. Wang (2002), Quantifying parameter uncertainty in stochastic models using the Box-Cox transformation, *J. Hydrol.*, 265(1-4), 246-257.
- Thyer, M., B. Renard, D. Kavetski, G. Kuczera, S. W. Franks, and S. Srikanthan (2009), Critical evaluation of parameter consistency and predictive uncertainty in hydrological modeling: A case study using Bayesian total error analysis, *Water Resour. Res.*, 45, W00B14, doi:10.1029/2008WR006825.
- Tolson, B. A., and C. A. Shoemaker (2007), Dynamically dimensioned search algorithm for computationally efficient watershed model calibration, *Water Resour. Res.*, 43, W01413, doi:10.1029/2005WR004723.

- Tu, W., and X.-H. Zhou (1999), A Wald test comparing medical costs based on log-normal distributions with zero valued costs, *Stat. Med.*, 18(20), 2749-2761.
- Tu, W. (2002), Zero-Inflated Data, in *Encyclopedia of Environmetrics*, edited by A. H. El-Shaarawi and W. W. Piegorsch, pp. 2387-2391, John Wiley & Sons, Ltd., Chichester, U.K.
- Vaché, K. B., J. J. McDonnell, and J. Bolte (2004), On the use of multiple criteria for a posteriori model rejection: Soft data to characterize model performance, *Geophys. Res. Lett.*, 31, L21504, doi:10.1029/2004GL021577.
- Vaze, J., D. A. Post, F. H. S. Chiew, J. M. Perraud, N. R. Viney, and J. Teng (2010), Climate non-stationarity – Validity of calibrated rainfall–runoff models for use in climate change studies, *J. Hydrol.*, 394(3-4), 447-457.
- Vrugt, J., C. ter Braak, H. Gupta, and B. Robinson (2009), Equifinality of formal (DREAM) and informal (GLUE) Bayesian approaches in hydrologic modeling?, *Stochast. Environ. Res. Risk Assess.*, 23(7), 1011-1026.
- Vrugt, J. A., H. V. Gupta, W. Bouten, and S. Sorooshian (2003), A Shuffled Complex Evolution Metropolis algorithm for optimization and uncertainty assessment of hydrologic model parameters, *Water Resour. Res.*, 39, 1201, doi:10.1029/2002WR001642.
- Vrugt, J. A., H. V. Gupta, S. C. Dekker, S. Sorooshian, T. Wagener, and W. Bouten (2006), Application of stochastic parameter optimization to the Sacramento Soil Moisture Accounting model, *J. Hydrol.*, 325(1-4), 288-307.
- Vrugt, J. A., C. J. F. ter Braak, M. P. Clark, J. M. Hyman, and B. A. Robinson (2008), Treatment of input uncertainty in hydrologic modeling: Doing hydrology backwards with Markov chain Monte Carlo simulation, *Water Resour. Res.*, 44, W00B09, doi:10.1029/2007WR006720.
- Wagener, T. (2003), Evaluation of catchment models, *Hydrol. Processes*, 17(16), 3375-3378.
- Wagener, T., H. S. Wheater, and H. V. Gupta (2004), *Rainfall-Runoff Modeling in Gauged and Ungauged Catchments*, Imperial College Press, London.
- Wagener, T., and H. V. Gupta (2005), Model identification for hydrological forecasting under uncertainty, *Stochast. Environ. Res. Risk Assess.*, 19(6), 378-387.
- Wagener, T., and N. McIntyre (2005), Identification of rainfall-runoff models for operational applications, *Hydrol. Sci. J.*, 50(5), 735-751.

- Wagener, T., and H. S. Wheater (2006), Parameter estimation and regionalization for continuous rainfall-runoff models including uncertainty, *J. Hydrol.*, 320(1-2), 132-154.
- Wagener, T., M. Sivapalan, P. Troch, and R. Woods (2007), Catchment Classification and Hydrologic Similarity, *Geography Compass*, 1(4), 901-931.
- Wagener, T., K. van Werkhoven, P. Reed, and Y. Tang (2009), Multiobjective sensitivity analysis to understand the information content in streamflow observations for distributed watershed modeling, *Water Resour. Res.*, 45, W02501, doi:10.1029/2008WR007347.
- Wagener, T., and A. Montanari (2011), Convergence of approaches toward reducing uncertainty in predictions in ungauged basins, *Water Resour. Res.*, 47, W06301, doi:10.1029/2010WR009469.
- Weiler, M., B. L. McGlynn, K. J. McGuire, and J. J. McDonnell (2003), How does rainfall become runoff? A combined tracer and runoff transfer function approach, *Water Resour. Res.*, 39, 1315, doi:10.1029/2003WR002331.
- Weiler, M., and J. McDonnell (2004), Virtual experiments: a new approach for improving process conceptualization in hillslope hydrology, *J. Hydrol.*, 285(1-4), 3-18.
- Weyman, D. R. (1970), Throughflow on hillslopes and its relation to the stream hydrograph, *Bull. Int. Assoc. Sci. Hydrol.*, 15(2), 25-33.
- Xu, C.-Y. (2001), Statistical Analysis of Parameters and Residuals of a Conceptual Water Balance Model – Methodology and Case Study, *Water Resour. Manag.*, 15(2), 75-92.
- Yadav, M., T. Wagener, and H. Gupta (2007), Regionalization of constraints on expected watershed response behavior for improved predictions in ungauged basins, *Adv. Water Resour.*, 30(8), 1756-1774.
- Zhang, Y., and F. H. S. Chiew (2009), Relative merits of different methods for runoff predictions in ungauged catchments, *Water Resour. Res.*, 45, W07412, doi:10.1029/2008WR007504.
- Zhang, Z., T. Wagener, P. Reed, and R. Bhushan (2008), Reducing uncertainty in predictions in ungauged basins by combining hydrologic indices, regionalization and multiobjective optimization, *Water Resour. Res.*, 44, W00B04, doi:10.1029/2008WR006833.

Zhou, X.-H., and W. Tu (1999), Comparison of Several Independent Population Means When Their Samples Contain Log-Normal and Possibly Zero Observations, *Biometrics*, 55(2), 645-651.

Zhou, X.-H., and W. Tu (2000), Confidence Intervals for the Mean of Diagnostic Test Charge Data Containing Zeros, *Biometrics*, 56(4), 1118-1125.