THE APPLICATION OF APPROXIMATE BAYESIAN COMPUTATION IN THE CALIBRATION OF HYDROLOGICAL MODELS

by

Jason John Brown

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

of

Master of Science

In

Land Resources and Environmental Sciences

MONTANA STATE UNIVERSITY
Bozeman, Montana

April 2014

©COPYRIGHT

By

Jason John Brown

2014

All Rights Reserved

TABLE OF CONTENTS

1. INTRODUCTION TO THESIS....1

Uncertainty Analysis in Hydrologic Models....1
References....7

2. GENERALIZED LIKELIHOOD UNCERTAINTY ESTIMATION (GLUE) AND APPROXIMATE BAYESIAN COMPUTATION: WHAT'S THE CONNECTION?....11

Contribution of Authors and Co-Authors....11
Manuscript Information Page....12
Introduction....13
Background....14
Bayesian Inference....14
Generalized Likelihood Uncertainty Estimation....14
Approximate Bayesian Computation....15
The ABC-GLUE Connection....15
A Theoretical Analysis....15
A Case Study....16
Discussion....17
References....19

3. THE APPLICATION OF APPROXIMATE BAYESIAN COMPUTATION AND SEQUENTIAL MONTE CARLO IN THE CALIBRATION OF HYDROLOGICAL MODELS....20

Contribution of Authors and Co-Authors....20
Manuscript Information Page....21
Introduction....22
Methods....25
Conceptual Rainfall Runoff Model....25
Summary Statistics....28
ABC-SMC Algorithm....30
Posterior Post-Processing....33
Calibration and Model Assessment....34
Hydrological Data....35
Case Studies....36
Case Study 1: Synthetic Data....36
Case Study 2: Davey River, Tasmania....37
Case Study 3: Haughton River, Queensland....37

TABLE OF CONTENTS – CONTINUED

Results 38
Case Study 1 38
Case Study 2 40
Case Study 3 41
Discussion 42
ABC-SMC Algorithm Performance 43
Post-Processing with Local Linear Regression 44
Parameter and Predictive Uncertainty 45
Synthetic Data Set 45
Davey River and Haughton River Data Sets 46
Conclusion 47
References 71

4. CONCLUSION 75

References 78

REFERENCES CITED 79

LIST OF TABLES

Table Page

1: Uniform Prior Distributions for the three case studies 49

2: Parameter Estimations for the Synthetic data set using 49

3: Parameter Estimations for the Synthetic data set 49

4: Parameter Estimations for the Synthetic data set using 50

5: Predictive Uncertainty for the Synthetic Data Set for Case Study 1. 50

6: Parameter Estimations for the perennial flow data set using 50

7: Parameter Estimations for the perennial flow data set using 51

8: Parameter Estimations for the perennial flow data set using 51

9: Predictive Uncertainty for perennial flow for Case Study 2. 51

10: Parameter Estimations for the Haughton Stream data set using 52

11: Parameter Estimations for the Haughton Stream data set using 52

12: Parameter Estimations for the Haughton Stream data set using 52

13: Predictive Uncertainty for ephemeral flow for Case Study 3. 53

LIST OF FIGURES

Figure Page

1: The implemented PDM Model..... 54

2: Geographic Locations for Davy River and ephemeral flow Gauges, Australia 55

3: The hydrograph and hyetograph for the Davey River data set 56

4: The hydrograph and hyetograph for the Haughton River data 57

5: Dotty plots for the Synthetic data set calibration 58

6: Histograms and QQ-Norm plots for the Minimum Soil Storage
for the Synthetic data set 59

7: Histograms and QQ-Norm plots for the Groundwater Recharge Rate
for the Synthetic data set. 60

8: Parameter density plots for the Synthetic data set 61

9: Predictive uncertainty plots for the Synthetic data set.....62

10: Predictive uncertainty plots for the Synthetic data set
using the signature-based summary statistic 62

11: Predictive uncertainty plots for the Synthetic data set
using the composite summary statistic.....63

12: Parameter density plots for the Synthetic data set
using the Full (PDM1) and Reduced (PDM2) models 64

13: Parameter density plots for the Davey River data set.....65

14: Predictive uncertainty plots for the Davey River data set
using the hydrograph summary statistic 66

LIST OF FIGURES – CONTINUED

Figure Page

15: Predictive uncertainty plots for the Davey River data set using the signature-based summary statistic. 66

16: Predictive uncertainty plots for the Davey River data set using the composite summary statistic....67

17: Parameter desnity plots for the Haughton River dat set....68

18: Predictive uncertainty plots for the Haughton River data set using the hydrograph summary statistic....69

19: Predictive uncertainty plots for the Haughton River data set using the signature-based summary statistic....70

20: Predictive uncertainty plots for the Haughton River data set using the composite summary statistic. 70

ABSTRACT

There is an increasing need to obtain proper estimates for the uncertainty associated with Conceptual Rainfall-Runoff models and their predictions. Within hydrology, uncertainty analysis is commonly conducted using Bayesian inference or Generalized Likelihood Uncertainty Estimation (GLUE). Bayesian inference is a statistically rigorous method for estimating uncertainty, but it depends upon a formal likelihood function that may not be available. GLUE utilizes a generalized likelihood function that can operate as a proxy for a formal likelihood function. While this allows GLUE to effectively calibrate hydrological models with intractable likelihood functions, the lack of statistical rigor may negatively affect the uncertainty estimations.

Approximate Bayesian Computation (ABC) is a family of likelihood-free methods that have been recently introduced for calibrating hydrological models. While these methods are implemented using formal Bayesian inference for assessing uncertainty, they do not require any assumptions regarding the likelihood function. Thus they have the potential flexibility of GLUE with the statistical rigor inherent in Bayesian Inference.

The studies presented within this thesis demonstrate the theoretical links between GLUE and ABC. We then assess the efficacy of an implementation of ABC utilizing a Sequential Monte Carlo sampler (ABC-SMC) for calibrating Conceptual Rainfall-Runoff models. Two components of the ABC-SMC algorithm were evaluated. These included three classes of summary statistics used for evaluating model performance and post-processing techniques to adjust the final posterior distributions of the parameters.

ABC-SMC was computationally efficient in calibrating a six parameter hydrological model for one synthetic and two real world data sets. Post-processing using local linear regression generated marginal improvements to the posterior distributions. Summary statistics measuring the goodness-of-fit between the observed and predicted hydrographs performed well for the synthetic data where the total uncertainty was low. A composite summary statistic based upon matching both hydrograph and hydrological signatures of a basin were more effective for the real world data sets as total uncertainty increased. The results suggest a properly implemented ABC-SMC algorithm is an effective method for calibrating watershed models and for conducting uncertainty analysis.

CHAPTER 1

INTRODUCTION TO THESIS

Uncertainty Analysis in Hydrologic Models

The availability of water as a natural resource affects the genesis, sustainability, and degeneration of communities. As a limiting raw material, sufficient water availability is required to support domestic, agricultural, and industrial applications. In excess, water can threaten the infrastructure of the communities. Sustained water availability to communities is threatened by increased extraction and changes to precipitation regimes due to long-term climate variability (Matalas 1997, Milly et al. 2008, Taylor et al. 2013). Therefore, an improved understanding of the watershed processes that ultimately deliver precipitation water to communities is vital to managing future water budgets and assessing risks to public infrastructure. This necessitates the continued development of new approaches to help understand, describe, and forecast the behavior of watershed systems.

Watershed models have been a primary tool for describing watersheds since the Rational Method was introduced over 150 years ago (Todini 2007). Today, there are many different hydrological models that exist in hydrologic practice and research. These models range from parsimonious conceptual models that depict the watershed as a series of storages (e.g. the Probability Distribution Model (Moore and Clarke 1981) and the Sacramento Model (Kitanidis and Bras 1980)) to complex physical models that

simulate the physics of overland and subsurface flow (e.g. Système Hydrologique Européen (Abbott et al. 1986) and THALES (Moore et al. 1988)). The selection of an appropriate model structure will depend upon the purpose of the model, the data available to specify and constrain the model, and the physical properties of the modeled watershed (Beven 2011).

Nearly all hydrologic models require some type of model calibration to adjust the model behavior. This allows the model to be representative of the hydrological characteristics for a specific catchment. When possible, parameters may be directly measured, inferred from regionalized studies, or specified a priori based on the known watershed characteristics (Boyle et al. 2000). Otherwise, parameters may be estimated using optimization algorithms when a priori information is unavailable or inappropriate. The calibration process is used to define values for undefined, or tunable, parameters that control the behavior of the model. There are a plethora of ways that model parameters may be determined, depending largely on the model type and available field observations.

Regardless of methods used to estimate model parameters, it is of increasing importance to quantify the degree of uncertainty associated with hydrologic models and predictions (Beven 2013). Each component within the data, model, and calibration process will contribute to the overall predictive uncertainty in any modeling exercise (Pappenberger and Beven 2006). The uncertainty associated with each component can be characterized as stochastic errors (aleatory) or errors due to insufficient knowledge

(epistemic) (Beven 2013, Beven et al. 2011). In hydrologic studies, aleatory errors are generally associated with the natural variation within the data while epistemic uncertainty can be introduced throughout the entire calibration process. Sources of epistemic uncertainty may involve data collection errors, implementing insufficient model structures, and uncertainty resulting from the calibration process. By implementing uncertainty analysis throughout the calibration process, the results can be used to identify issues within the measured data, suggest changes in the model structure, and evaluate the resulting parameter estimates. This feedback loop can lead to an increased confidence in the model output and reduce the bounds representing the predictive uncertainty (Liu and Gupta 2007). The appropriate interpretation of uncertainty estimates is heavily dependent upon the methodologies used to calibrate watershed models and their ability to provide statistically relevant measures of uncertainty.

With the increase in available computational power over the last two decades, methods for simulation-based model calibration and uncertainty estimation have seen rapid advancements. The development of Monte Carlo simulation techniques allows researchers to visualize many potential model outcomes. By extracting the metrics of the simulation results, estimates on the model uncertainty can be directly assessed (Beven and Freer 2001, Kuczera and Parent 1998, Marshall et al. 2004). Within hydrological research, Markov Chain Monte Carlo methods (MCMC) (Beven et al. 2011, Marshall et al. 2004, Smith and Marshall 2008, Vrugt et al. 2013) and Generalized

Likelihood Uncertainty Estimation (GLUE) (Beven and Binley 1992, Freer et al. 1996, Zhang et al. 2011) have seen widespread use.

Markov Chain Monte Carlo methods were first introduced in hydrology as a way to realistically quantify parameter uncertainty and its potential impact on hydrologic model predictions (Kuczera and Parent 1998). MCMC utilizes kernel based random walks to explore the full parameter space. Formal likelihood functions are used for evaluating the quality of proposed parameter sets, thus allowing Bayesian inference to be applied in the estimation of the parameter and model uncertainty. MCMC algorithms can be sensitive to the model initiation and can be inefficient if the parameter space is complex (Bates and Campbell 2001). In recent years, MCMC methods in hydrology have been adapted to reduce the possibility of performance degradation (Smith and Marshall 2008, ter Braak and Vrugt 2008, Vrugt and Braak 2011, Vrugt et al. 2013), however, these methods also increase the complexity of the implementation (Smith and Marshall 2008).

Generalized Likelihood Uncertainty Estimation (GLUE) offers a different philosophy for hydrologic model parameter estimation and uncertainty analysis. GLUE was adapted from the Hornberger-Spear-Young method for sensitivity analysis (Freer et al. 1996) and processes randomly generated parameter sets drawn from the full parameter space through the proposed model (Beven and Binley 1992). A generalized likelihood function is used to measure the performance of each parameter set and a threshold is defined to remove those parameter sets that do not meet the threshold criteria. The distribution

of the remaining parameter sets can then be used to estimate the parameter value and to evaluate the uncertainty of the parameter estimations (Beven and Binley 1992).

The generalized likelihood function allows for flexibility in incorporating prior information that cannot be formally defined (Beven et al. 2008). However, the use of a non-sufficient function has been reported to produce unreliable estimations (Mantovan and Todini 2006, Montanari 2005, Stedinger et al. 2008). In these cases, issues of heteroscedasticity, non-normality, and serial correlations could affect the ability of GLUE to generate reliable uncertainty intervals (Stedinger et al. 2008), potentially leading to an underestimation of the uncertainty (Montanari 2005). GLUE has also been criticized for being incoherent and ignoring potential covariance by working strictly in the marginal parameter space (Jin et al. 2010). Despite these concerns, GLUE has remained a popular method for calibrating hydrological models due to the ease of implementation and the flexibility of the generalized likelihood function.

An emerging set of Bayesian methods, called Approximate Bayesian Computation, have been proposed as a way to estimate the posterior distribution in modeling scenarios where the likelihood function is either unknown or cannot be explicitly defined (Diggle and Gratton 1984, Rubin 1984). The unavailability of the likelihood function may be due to mathematical intractability or when the likelihood is not defined in a closed form (Marin et al. 2012). The first applied applications of ABC were in the field of genetics to make inferences from DNA sequence data (Beaumont et al. 2002, Marjoram et al. 2003, Pritchard et al. 1999, Tavare et al. 1997) and more recent

research has demonstrated the potential applicability of ABC for both ecological and epidemiology analysis (Sisson et al. 2007, Toni and Stumpf 2010, Toni et al. 2009).

Likelihood-free methods aim to eliminate the need for solvable likelihood functions. By utilizing Monte Carlo sampling techniques in conjunction with a summary statistic to estimate the performance of randomized parameter sets, the posterior distributions of the parameters can be estimated. If the summary statistic is sufficient, then the estimated posterior distribution of the parameters will reflect the true posterior distribution under a formal Bayesian analysis (Joyce and Marjoram 2008). Therefore, when properly implemented, ABC may offer the flexibility of GLUE while retaining the statistical rigor recognized in MCMC (Sadegh and Vrugt 2013).

This thesis is organized into four chapters. Chapter 1 is an introduction to the thesis, motivations, and contributions of the research. Chapter 2 is a published manuscript that explores the links between ABC and GLUE. Chapter 3 is a prepared manuscript that introduces innovative methodologies for improving the efficiency and performance of ABC methods useful for calibrating hydrological models. Chapter 4 is a summary of the research and findings reported in the two manuscripts.

References

Abbott, M.B., Bathurst, J.C., Cunge, J.A., O'Connell, P.E. and Rasmussen, J. (1986) An introduction to the European Hydrological System — Systeme Hydrologique Europeen, "SHE", 1: History and philosophy of a physically-based, distributed modelling system. Journal of Hydrology 87(1–2), 45-59.

Bates, B.C. and Campbell, E.P. (2001) A Markov Chain Monte Carlo Scheme for parameter estimation and inference in conceptual rainfall-runoff modeling. Water Resources Research 37(4), 937-947.

Beaumont, M.A., Zhang, W. and Balding, D.J. (2002) Approximate Bayesian Computation in Population Genetics. Genetics 162(4), 2025-2035.

Beven, K. (2013) So how much of your error is epistemic? Lessons from Japan and Italy. Hydrological Processes 27(11), 1677-1680.

Beven, K. and Binley, A. (1992) The future of distributed models: Model calibration and uncertainty prediction. Hydrological Processes 6(3), 279-298.

Beven, K. and Freer, J. (2001) Equifinality, data assimilation, and uncertainty estimation in mechanistic modelling of complex environmental systems using the GLUE methodology. Journal of Hydrology 249(1–4), 11-29.

Beven, K., Smith, P.J. and Wood, A. (2011) On the colour and spin of epistemic error (and what we might do about it). Hydrology & Earth System Sciences 15(10), 3123-3133.

Beven, K.J. (2011) Rainfall-Runoff Modelling: The Primer, Wiley.

Beven, K.J., Smith, P.J. and Freer, J.E. (2008) So just why would a modeller choose to be incoherent? Journal of Hydrology 354(1–4), 15-32.

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

Diggle, P.J. and Gratton, R.J. (1984) Monte Carlo Methods of Inference for Implicit Statistical Models. Journal of the Royal Statistical Society. Series B (Methodological) 46(2), 193-227.

Freer, J., Beven, K. and Ambroise, B. (1996) Bayesian Estimation of Uncertainty in Runoff Prediction and the Value of Data: An Application of the GLUE Approach. Water Resources Research 32(7), 2161-2173.

Jin, X., Xu, C.-Y., Zhang, Q. and Singh, V.P. (2010) Parameter and modeling uncertainty simulated by GLUE and a formal Bayesian method for a conceptual hydrological model. Journal of Hydrology 383(3–4), 147-155.

Joyce, P. and Marjoram, P. (2008) Approximately Sufficient Statistics and Bayesian Computation. Statistical Applications in Genetics & Molecular Biology 7(1), 1-18.

Kitanidis, P.K. and Bras, R.L. (1980) Real-time forecasting with a conceptual hydrologic model: 1. Analysis of uncertainty. Water Resources Research 16(6), 1025-1033.

Kuczera, G. and Parent, E. (1998) Monte Carlo assessment of parameter uncertainty in conceptual catchment models: the Metropolis algorithm. Journal of Hydrology 211(1–4), 69-85.

Liu, Y. and Gupta, H.V. (2007) Uncertainty in hydrologic modeling: Toward an integrated data assimilation framework. Water Resources Research 43(7), W07401.

Mantovan, P. and Todini, E. (2006) Hydrological forecasting uncertainty assessment: Incoherence of the GLUE methodology. Journal of Hydrology 330(1–2), 368-381.

Marin, J.-M., Pudlo, P., Robert, C. and Ryder, R. (2012) Approximate Bayesian computational methods. Statistics and Computing 22(6), 1167-1180.

Marjoram, P., Molitor, J., Plagnol, V. and Tavaré, S. (2003) Markov chain Monte Carlo without likelihoods. Proceedings of the National Academy of Sciences 100(26), 15324-15328.

Marshall, L., Nott, D. and Sharma, A. (2004) A comparative study of Markov chain Monte Carlo methods for conceptual rainfall-runoff modeling. Water Resources Research 40(2), W02501.

Matalas, N. (1997) Stochastic Hydrology in the Context of Climate Change. Climatic Change 37(1), 89-101.

Milly, P.C.D., Betancourt, J., Falkenmark, M., Hirsch, R.M., Kundzewicz, Z.W., Lettenmaier, D.P. and Stouffer, R.J. (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 Resources Research 41(8).

Moore, I.D., O'Loughlin, E.M. and Burch, G.J. (1988) A contour-based topographic model for hydrological and ecological applications. Earth Surface Processes and Landforms 13(4), 305-320.

Moore, R.J. and Clarke, R.T. (1981) A Distribution Function Approach to Rainfall Runoff Modeling. Water Resources Research Vol 17, No 5, p 1367-1382, October, 1981.

Pappenberger, F. and Beven, K.J. (2006) Ignorance is bliss: Or seven reasons not to use uncertainty analysis. Water Resources Research 42(5), W05302.

Pritchard, J.K., Seielstad, M.T., Perez-Lezaun, A. and Feldman, M.W. (1999) Population growth of human Y chromosomes: a study of Y chromosome microsatellites. Molecular Biology and Evolution 16(12), 1791-1798.

Rubin, D.B. (1984) Bayesianly Justifiable and Relevant Frequency Calculations for the Applies Statistician. The Annals of Statistics 12(4), 1151-1172.

Sadegh, M. and Vrugt, J.A. (2013) Approximate Bayesian Computation in hydrologic modeling: equifinality of formal and informal approaches. Hydrology & Earth System Sciences Discussions 10(4), 4739-4797.

Sisson, S.A., Fan, Y. and Tanaka, M.M. (2007) Sequential Monte Carlo without likelihoods. Proceedings of the National Academy of Sciences 104(6), 1760-1765.

Smith, T.J. and Marshall, L.A. (2008) Bayesian methods in hydrologic modeling: A study of recent advancements in Markov chain Monte Carlo techniques. Water Resources Research 44(12), W00B05.

Stedinger, J.R., Jery, R.S., Richard, M.V., Seung Uk, L. and Rebecca, B. (2008) Appraisal of the generalized likelihood uncertainty estimation (GLUE) method. Water Resources Research 44(12), n-a-n/a.

Tavare, S., Balding, D.J., Griffiths, R.C. and Donnelly, P. (1997) Inferring coalescence times from DNA sequence data. Genetics 145(2), 505-518.

Taylor, R.G., Scanlon, B., Doll, P., Rodell, M., van Beek, R., Wada, Y., Longuevergne, L., Leblanc, M., Famiglietti, J.S., Edmunds, M., Konikow, L., Green, T.R., Chen, J., Taniguchi, M., Bierkens, M.F.P., MacDonald, A., Fan, Y., Maxwell, R.M., Yechieli, Y., Gurdak, J.J., Allen, D.M., Shamsudduha, M., Hiscock, K., Yeh, P.J.F., Holman, I. and Treidel, H. (2013) Ground water and climate change. Nature Clim. Change 3(4), 322-329.

ter Braak, C.F. and Vrugt, J. (2008) Differential Evolution Markov Chain with snooker updater and fewer chains. Statistics and Computing 18(4), 435-446.

Todini, E. (2007) Hydrological catchment modelling: past, present and future. Hydrol. Earth Syst. Sci. 11(1), 468-482.

Toni, T. and Stumpf, M.P.H. (2010) Simulation-based model selection for dynamical systems in systems and population biology. Bioinformatics 26(1), 104-110.

Toni, T., Welch, D., Strelkowa, N., Ipsen, A. and Stumpf, M.P.H. (2009) Approximate Bayesian computation scheme for parameter inference and model selection in dynamical systems. Journal of the Royal Society Interface 6(31), 187-202.

Vrugt, J.A. and Braak, C.J.F.t. (2011) DREAM(D): an adaptive Markov Chain Monte Carlo simulation algorithm to solve discrete, noncontinuous, and combinatorial posterior parameter estimation problems. Hydrology and Earth System Sciences 15(12), 3701-3713.

Vrugt, J.A., ter Braak, C.J.F., Diks, C.G.H. and Schoups, G. (2013) Hydrologic data assimilation using particle Markov chain Monte Carlo simulation: Theory, concepts and applications. Advances in Water Resources 51(0), 457-478.

Zhang, Y., Liu, H. and Houseworth, J. (2011) Modified Generalized Likelihood Uncertainty Estimation (GLUE) Methodology for Considering the Subjectivity of Likelihood Measure Selection. Journal of Hydrologic Engineering 16(6), 558-561.

CHAPTER 2

GENERALIZED LIKELIHOOD UNCERTAINTY ESTIMATION (GLUE) AND APPROXIMATE BAYESIAN COMPUTATION: WHAT'S THE CONNECTION?

Contribution of Authors and Co-Authors

Manuscript in Chapter 2

Author: David J. Nott

Contributions: Primary author who conceptualized the paper and developed the theoretical components.

Co-Author: Lucy Marshall

Contributions: Conceptualized the hydrological case studies, wrote the case study, and edited the manuscript at all stages.

Co-Author: Jason J. Brown

Contributions: Developed the code for the ABC-SMC algorithm used in the case studies.

Manuscript Information Page

David J. Nott, Lucy A. Marshall, Jason J. Brown
Water Resources Research
Status of Manuscript:
____ Prepared for submission to a peer-reviewed journal
____ Officially submitted to a peer-review journal
____ Accepted by a peer-reviewed journal
__**X**_ Published in a peer-reviewed journal

Published by John Wiley and Sons
Volume 48, Issue 12

WATER RESOURCES RESEARCH, VOL. 48, W12602, doi:10.1029/2011WR011128, 2012

Generalized likelihood uncertainty estimation (GLUE) and approximate Bayesian computation: What's the connection?

David J. Nott,[1] Lucy Marshall,[2] and Jason Brown[2]

Received 5 July 2011; revised 11 September 2012; accepted 1 November 2012; published 20 December 2012.

[1] There has been a recent debate in the hydrological community about the relative merits of the informal generalized likelihood uncertainty estimation (GLUE) approach to uncertainty assessment in hydrological modeling versus formal probabilistic approaches. Some recent literature has suggested that the methods can give similar results in practice when properly applied. In this note, we show that the connection between formal Bayes and GLUE is not merely operational but goes deeper, with GLUE corresponding to a certain approximate Bayesian procedure even when the "generalized likelihood" is not a true likelihood. The connection we describe relates to recent approximate Bayesian computation (ABC) methods originating in genetics. ABC algorithms involve the use of a kernel function, and the generalized likelihood in GLUE can be thought of as relating to this kernel function rather than to the model likelihood. Two interpretations of GLUE emerge, one as a computational approximation to a Bayes procedure for a certain "error-free" model and the second as an exact Bayes procedure for a perturbation of that model in which the truncation of the generalized likelihood in GLUE plays a role. The intent of this study is to encourage cross-fertilization of ideas regarding GLUE and ABC in hydrologic applications. The connection we outline suggests the possibility of combining a formal likelihood with a kernel based on a generalized likelihood within the ABC framework and also allows advanced ABC computational methods to be used in GLUE applications. The model-based interpretation of GLUE may also be helpful in partially illuminating the implicit assumptions in different choices of generalized likelihood.

Citation: Nott, D. J., L. Marshall, and J. Brown (2012), Generalized likelihood uncertainty estimation (GLUE) and approximate Bayesian computation: What's the connection?, *Water Resour. Res.*, *48*, W12602, doi:10.1029/2011WR011128.

1. Introduction

[2] A common approach to uncertainty assessment in the hydrological literature is the generalized likelihood uncertainty estimation (GLUE) technique of *Beven and Binley* [1992]. There has been recent discussion in the hydrological literature concerning the relative merits of GLUE versus formal probabilistic approaches [see *Mantovan and Todini*, 2006; *Stedinger et al.*, 2008; *Beven et al.*, 2008; *Vrugt et al.*, 2009a; *Beven*, 2009, and references therein]. A recent comparison of the two methods concluded that "formal and informal Bayesian approaches have more common ground than the hydrologic literature and ongoing debate might suggest" [*Vrugt et al.*, 2009a, p. 1011]. In this note, we show that the connection between formal Bayes and GLUE goes beyond operational similarity by showing that it is possible to give a formal Bayesian interpretation of GLUE even when the corresponding generalized likelihood is not a true likelihood. In particular, recent developments in approximate Bayesian computation (ABC) methods (originating in genetics with related earlier precedents in the papers of *Diggle and Gratton* [1984] and *Gourieroux et al.* [1993]) provide an illuminating new way of looking at the GLUE procedure. In fact, while not originally motivated as an ABC algorithm, remarkably, GLUE predates the development of ABC methods: the first simple ABC algorithm was discussed in *Tavaré et al.* [1997].

[3] The ABC-GLUE connection is fruitful in a number of ways. First, even if the association is regarded as merely technical it allows advanced ABC computational methods and software to be brought to bear in GLUE applications. Second, in one of the ABC interpretations of GLUE, the generalized likelihood does not correspond to the likelihood but to a certain kernel function that appears in the algorithm; this naturally suggests a way to combine both a likelihood and generalized likelihood, with the probabilistic model allowing description of different sources of variability in a problem but the generalized likelihood reducing sensitivity to model misspecification in directions of interest to a modeler. Third, the model-based interpretation of GLUE may also be helpful in partially illuminating the implicit assumptions in different choices of generalized likelihood. The goal of this study is

[1]Department of Statistics and Applied Probability, National University of Singapore, Singapore.
[2]Department of Land Resources and Environmental Sciences, Montana State University, Bozeman, Montana, USA.

Corresponding author: L. Marshall, Department of Land Resources and Environmental Sciences, Montana State University, 334 Leon Johnson Hall, PO Box 173120, Bozeman, MT 59717-3120, USA. (lmarshall@montana.edu)

©2012. American Geophysical Union. All Rights Reserved.
0043-1397/12/2011WR011128

thus to promote cross-fertilization of ideas regarding GLUE and ABC, which we believe will motivate continuing interest in the underlying assumptions of GLUE and the numerical implementation of likelihood-free methods in hydrology.

[4] In the next section, we give a brief introduction to Bayesian inference, GLUE, and ABC methods. In section 3, we show how GLUE can be viewed as an ABC algorithm for a certain probabilistic model in which the corresponding generalized likelihood plays the role of the ABC kernel function. We also point out that the generalized likelihood (even when not a proper likelihood) can usually be considered to define a proper likelihood when the truncation in the GLUE algorithm is taken into account. We further include a case study that demonstrates how a representative ABC sampling algorithm and interpretation may be used for an established GLUE approach that attempts to address epistemic hydrologic model errors. Section 4 contains concluding discussion.

2. Background

2.1. Bayesian Inference

[5] Suppose we observe real data $y = (y_1, \ldots, y_n)^T$ and that we have a model that generates an output corresponding to this data $\hat{y} = (\hat{y}_1, \ldots, \hat{y}_n)^T$. The model output depends on some parameters θ that are unknown, so that $\hat{y} = \hat{y}(\theta)$. θ could in principle be high-dimensional, including uncertain model inputs as well as parameters that relate to understanding of dominant hydrological processes. Through the model outputs and a specification for the distribution of model errors $y - \hat{y}$, we can specify a probabilistic model for the observable data y, $p(y|\theta)$ say. We call $p(y|\theta)$ the likelihood function. In a Bayesian approach to inference, we also specify in probabilistic form what we know about the unknown θ before observing data through a so-called prior distribution $p(\theta)$. Together, the prior and likelihood specify a full probabilistic model for (y, θ), $p(y, \theta) = p(\theta)p(y|\theta)$, and then once data are observed we can condition on y to obtain

$$p(\theta|y) \propto p(\theta)p(y|\theta). \tag{1}$$

$p(\theta|y)$ is called the posterior distribution, and it is a probabilistic representation of what we know about the unknown θ after observing data. Inference and prediction from the model are then based on the posterior distribution, at least after appropriate model checking and diagnostics.

[6] The above is a quick summary of the formal Bayesian approach to inference as it applies to hydrological modeling (see, for example, *Gelman et al.* [2003] for a complete discussion of Bayesian statistical inference). Although the application of formal Bayesian methods in hydrology has been increasing, some modelers have cautioned that insufficient attention is paid to the difficulties of specification of a full probabilistic model, with the possibility that inferences of interest might be sensitive to some arbitrary aspects of such a specification. While appropriate diagnostic checks can certainly guard against this sensitivity to some extent it could be argued that at least in some situations an alternative to a full probabilistic model specification is needed. On the other hand, it has also been suggested that one strength of the full Bayesian approach lies in the ability to understand and model contributions of different kinds of uncertainty (such as parameter uncertainty, structural uncertainty, and input uncertainty). *Vrugt et al.* [2009a] and *Beven* [2009] provide a recent discussion of these concerns.

[7] We say something also about how to generate Monte Carlo samples from the posterior distribution (1), as this relates to the interpretation of GLUE given later. Markov chain Monte Carlo methods are most commonly used for this task (see *Gelman et al.* [2003] for an introduction), but here we will discuss only importance sampling approaches. In an importance sampling approach, we start with a so-called importance density $h(\theta)$ on the parameter space, which is hopefully a good approximation to $p(\theta|y)$ and is easy to simulate from directly. Then we generate $\theta_i \sim h(\theta)$, $i = 1, \ldots, n$, and calculate the weights

$$w_i \propto \frac{p(\theta_i)p(y|\theta_i)}{h(\theta_i)}, \tag{2}$$

where the weights are normalized to sum to one, $\sum_{i=1}^{n} w_i = 1$. The points θ_i with weights w_i are a weighted sample from the posterior, in the sense that if $r(\theta)$ is some measurable function of θ, then,

$$\sum_{i=1}^{n} r(\theta_i)w_i$$

is a consistent estimate of $E(r(\theta)|y)$ as $n \to \infty$.

2.2. Generalized Likelihood Uncertainty Estimation

[8] The philosophy behind GLUE is to avoid a difficult full probabilistic model specification but to instead seek out regions in the parameter space where the model predictions are consistent with the observations. It is often the case that there may be widely separated regions of the parameter space consistent with the data (the so-called equifinality problem). The GLUE algorithm is:

Algorithm 1:

[9] 1. Generate points $\eta_1, \ldots, \eta_n$ from the prior $p(\theta)$.

[10] 2. For each η_i, calculate the generalized likelihood values $G(\eta_i; y)$, $i = 1, \ldots, n$.

[11] 3. For a selected threshold c, discard all η_i with $G(\eta_i; y) < c$ and retain the rest, resulting in a remaining "behavioral" set of solutions, which we denote by $\theta_1, \ldots, \theta_m$.

[12] 4. Compute weights

$$w_i = \frac{G(\theta_i; y)}{\sum_{j=1}^{m} G(\theta_j; y)}, i = 1, \ldots, m.$$

[13] 5. For a predictive quantity Δ, estimate the distribution of Δ by the discrete distribution with weights w_i on Δ_i, where Δ_i is the prediction using parameter θ_i, $i = 1, \ldots, m$.

[14] The generalized likelihood $G(\theta; y)$ can be defined in a number of different ways. It can be an actual likelihood function $p(y|\theta)$, in which case the GLUE algorithm corresponds to an importance sampling method for generating samples from the posterior distribution if the truncation at step 3 is ignored. The importance density here is the prior $p(\theta)$. More commonly, however, the generalized likelihood need not be a true likelihood but is merely a function that measures the degree of misfit between the observations and

model predictions in ways that seem relevant to the modeler. We note later, however, that the combination of the generalized likelihood together with the truncation of step 3 usually does define a proper likelihood [*Smith et al.*, 2008]. One common choice for the generalized likelihood is

$$G(\theta; y) = (s^2_{y-\hat{y}})^{-T},$$

where $s_{y-\hat{y}}$ is the standard deviation of the model errors $y - \hat{y}$, and $T > 0$ is a selected constant (termed the shaping factor).

[15] The behavioral threshold, c, may be based on what is referred to as the *limits of acceptability* [*Beven*, 2006]. This refers to an estimate of the effective model error that attempts to take into account multiple sources of uncertainty in the modeling process (such as observational errors and model structure errors, e.g., *Liu et al.* [2009] and *Blazkova and Beven* [2009]).

2.3. Approximate Bayesian Computation

[16] ABC methods originated in genetics with *Tavaré et al.* [1997] and is commonly used in situations, where we have a model that is easy to simulate from, given the model parameters, but where the likelihood is difficult or impossible to calculate. ABC methods are able to simulate from the posterior distribution in Bayesian inference, producing summaries of the posterior useful for inference and prediction but without requiring that the likelihood be calculated. A recent survey of ABC methods is given by *Marin et al.* [2011].

[17] To describe the most basic form of ABC, suppose that the data y are discrete. Consider the following algorithm:

Algorithm 2:

[18] 1. Sample $(\theta, z) \sim p(\theta)p(z|\theta)$ (that is, first sample θ from the prior $p(\theta)$ and then for the sampled θ sample z from $p(z|\theta)$).

[19] 2. If $z = y$ (i.e. if the simulated data from $p(z|\theta)$ matches the observed data y) then output the value of θ simulated at step 1 and stop. Otherwise, return to step 1.

[20] It is not hard to see that the above algorithm is simply a rejection algorithm for simulating from $p(\theta|y)$. Furthermore, provided we can simulate from the model there is no need to evaluate the model likelihood in this algorithm.

[21] The rejection rate of the above algorithm might be very high if the number of possible values for y is large, and if y is continuous the algorithm in the form stated above is not applicable at all. This objection is usually addressed in a number of ways. First, often the data are summarized through summary statistics s that might be of much lower dimension than y (if the statistics s are sufficient statistics there is no loss of information in such a reduction). Second, in step 2, we in general do not require an exact match of z and y but instead require only that $d(y, z) < \epsilon$, where d is some distance function and $\epsilon > 0$ is a tolerance parameter. Third, in addition to the first two innovations there are many advanced computational methods that can be employed to make the methods work well in problems of moderate dimension (see the review of *Marin et al.* [2011], for an entry point into the literature). It is not our intention to discuss advanced ABC methods here. In what follows and with a slight abuse of notation, we continue to write y for the data that we condition on for inference, even if this is a summary statistic obtained from the full data set.

[22] Note that in Algorithm 2, if we accept the sampled θ in step 2 when $\mathrm{d}(y, z) < \epsilon$ we are in effect sampling from the distribution

$$p^*(z, \theta) = \frac{p(\theta)p(z|\theta)I(\mathrm{d}(y, z) < \epsilon)}{\int p(\theta)p(z|\theta)I(\mathrm{d}(y, z) < \epsilon)\mathrm{d}\theta\mathrm{d}z}, \tag{3}$$

i.e., we are sampling from the joint distribution of $p(\theta)p(z|\theta)$ conditioned by the indicator $I(\mathrm{d}(y, z) < \epsilon)$, such that we accept samples for θ which satisfy $\mathrm{d}(y, z) < \epsilon$. The resulting draws for θ then come from the corresponding marginal distribution, which is obtained by integrating equation (3)

$$p^*(\theta|y) \propto p(\theta) \int p(z|\theta)I(\mathrm{d}(y, z) < \epsilon)\mathrm{d}z, \tag{4}$$

which effectively corresponds to replacing the likelihood in equation (1) with an approximate likelihood obtained by averaging in a neighborhood of y. In equation (4), we have shown dependence on y explicitly in the notation to emphasize that this is an approximation to the posterior distribution $p(\theta|y)$ with the approximation increasingly good as $\epsilon \to 0$ (provided that y is a sufficient statistic).

[23] One can also replace the term $I(\mathrm{d}(y, z) < \epsilon)$ in equation (4) with a kernel function $K_\epsilon(y - z)$, where ϵ now denotes a bandwidth parameter, with $K_\epsilon(h)$ concentrating on zero as $\epsilon \to 0$. This results in approximations

$$p^*(z, \theta) \propto p(\theta)p(z|\theta)K_\epsilon(y - z) \tag{5}$$

and

$$p^*(\theta|y) \propto p(\theta) \int p(z|\theta)K_\epsilon(y - z)\mathrm{d}z \tag{6}$$

corresponding to equations (3) and (4), respectively. R. D. Wilkenson (Approximate Bayesian computation (ABC) gives exact results under the assumption of model error, 2008, available at http://arxiv.org/abs/0811.3355, hereinafter referred to as Wilkenson, 2008) has pointed out that if $K_\epsilon(h)$ is a probability density, then we can think of $p^*(\theta|y)$ as an exact posterior distribution in which the data are considered to be generated as $y = z + \delta$ with $z \sim p(z|\theta)$ and $\delta \sim K_\epsilon(\delta)$. Hence, equation (6) can be regarded as an approximation to $p(\theta|y)$, or as an exact posterior for a different model to the one embodied by the likelihood $p(y|\theta)$ where there is an additional additive random error. Wilkinson (2008) has discussed different interpretations that might be given to this error.

3. The ABC-GLUE Connection

3.1. A Theoretical Analysis

[24] Consider the model

$$y = \hat{y}(\theta) + \delta, \tag{7}$$

where $\delta = (\delta_1, \ldots, \delta_n)^T$ is specified as a zero mean random vector. We recognize that the assumption of additive

stochastic errors holds only for idealized cases in hydrologic modeling (see, for example, *Beven and Westerberg* [2011]), but note that this assumption is useful for illustrating the ABC-GLUE connection. We leave the exact distribution of δ unspecified, but consider the limiting case where simulating data from the model with parameter θ just results in the deterministic output $\hat{y}(\theta)$. For a generalized likelihood $G(\theta; y)$, suppose it is a function of $y - \hat{y}(\theta)$ only, $G(\theta; y) = g(y - \hat{y})$ say, with g having a global maximum at the origin. Commonly used generalized likelihoods that we know of satisfy these requirements. Now consider the ABC target equation (5) for this model where the kernel $K_\epsilon(h)$ is chosen to be

$$K_\epsilon(h) \propto \begin{cases} g(h), & \text{if } g(h) > c, \\ 0, & \text{otherwise,} \end{cases}$$

where $K_\epsilon(h)$ is normalized to integrate to one, so that $K_\epsilon(h)$ is a probability density function (because of the truncation this is usually possible, regardless of whether $G(\theta; y)$ is a real likelihood, and is certainly possible if g is bounded, say if g is unbounded at the origin and not integrable, then one may consider a truncation of g at some high level since usually $y - \hat{y} \approx 0$ is not attainable anyway in situations where GLUE is applied). Here the kernel parameter ϵ can be considered some decreasing function of c, $\epsilon = \epsilon(c)$ say. If we sample from equation (5) by importance sampling, using the importance density $p(\theta)p(z|\theta)$, we obtain exactly the GLUE algorithm, since if we generate $(\theta_i, z_i) \sim p(\theta)p(z|\theta)$, $i = 1, \ldots, n$, the importance weights equation (2) are of the form

$$w_i \propto \frac{p(\theta_i)p(z_i|\theta_i)K_\epsilon(y - z_i)}{p(\theta_i)p(z|\theta_i)} = K_\epsilon(y - z_i)$$

and since $K_\epsilon(y - z_i)$ is just $g(y - z_i)$ if this is bigger than c and zero otherwise, the importance sampling weights are zero for the nonbehavioral solutions of GLUE, and the behavioral solutions have the GLUE weights.

[25] So GLUE corresponds to an importance sampling ABC algorithm where the kernel function is defined from the generalized likelihood. Furthermore, as we pointed out in section 2.3 and following an observation by Wilkinson (2008), the above ABC algorithm can be considered as sampling from an exact posterior rather than an approximation, but where $\delta \sim K_\epsilon(\delta)$. Hence, the GLUE algorithm does correspond to a certain exact Bayesian analysis, where the error distribution depends on the generalized likelihood (even if not a proper likelihood) since the kernel defined by the truncation of the generalized likelihood generally does result in a well-defined model.

[26] If it is said that GLUE is incoherent, what does this mean if GLUE actually is a Bayesian analysis for a certain model? Actually, there are a number of serious points here. The model we have discussed is defined implicitly through the generalized likelihood $G(\theta; y)$ and c, and if we consider models for $y_1, \ldots, y_n$ and for $y_1, \ldots, y_m$ with $m < n$ given θ defined in this way, then the latter model is not obtained from the former in general by marginalization. So if one requires temporal consistency between models that might be considered at different times then GLUE does not have this kind of consistency. Often it is pointed out that with GLUE and certain choices of generalized likelihood there is no increase in information as the sample size increases (but it should be noted that if we were to allow a sample size-dependent choice of c, this would not be the case). Proponents of GLUE may counteract that the purpose of this is to use parameter uncertainty to guard against false confidence resulting from model misspecification and the formal Bayesian likelihood. To this in turn critics might respond that it is unlikely to be the case that parameter uncertainty can adequately represent predictive uncertainty or compensate for model misspecification, and that the uncertainty intervals produced by GLUE are often not well calibrated in the usual statistical sense. It has been argued that this is not always the purpose of such intervals anyway (see, for example, remarks in *Beven* [2006]). In addition, it is suggested that for real problems formal statistical measures are inappropriate (or incoherent in the qualitative sense, *Beven et al.* [2008]). We believe that consideration of the model-based interpretation of GLUE above may in any case be helpful in identifying the implicit assumptions in different choices of generalized likelihood function and might shed some light on how GLUE is best implemented.

3.2. A Case Study

[27] The practical connection between GLUE and ABC methods in hydrology can further be demonstrated by a simple case study. A recent GLUE-based study by *Westerberg et al.* [2011] considered hydrologic model calibration using a catchment's flow duration curve (FDC), arguing the method appeared less sensitive to epistemic model errors than applications using traditional model performance measures. The approach is particularly useful for modeling scenarios where observation time periods for streamflow and model input data do not overlap such that a likelihood cannot be formulated.

[28] We implement here a conceptual rainfall-runoff model used frequently in hydrologic model applications. The probability distributed model (PDM) [*Moore*, 2007] simulates catchment streamflow based on saturation excess runoff processes by conceptualizing the spatial distribution of soil moisture storage as a probability density function. Our version of the model is adapted from *Smith and Marshall* [2010] and contains six unknown parameters (Table 1): *cmax*, the maximum soil moisture storage capacity; *b*, the spatial variability of storage capacity; *kb*, the groundwater recharge constant; *tr*1, the rate of outflow of a surface storage; *tr*2, the rate of outflow of a subsurface storage; and *cf*, a storage threshold factor. The model is applied to the Stringer Creek subcatchment of Tenderfoot Creek Experimental Forest (TCEF) located in Montana. Inputs to the model are time-series data of observed precipitation and potential evapotranspiration on a 6-h time step for the period of study extending from October 2005 to September 2008. Snowmelt and precipitation data are calculated from meteorological observations at the Onion Park SNOTEL station located at TCEF [see *Nippgen et al.*, 2011]. TCEF is of particular interest to this application as it has been previously demonstrated that there is a strong connection between the FDC and shallow groundwater hydrologic connectivity [*Jencso et al.*, 2009; *Jencso and McGlynn*, 2011]. This suggests the FDC will be useful here for identifying hydrologic model parameters and for model diagnostics.

Table 1. PDM Parameters and Sampling Ranges

Parameter	Description (Units)	Sampling Range
cmax	Maximum storage capacity (mm)	[0–400]
b	Exponent of Pareto distribution controlling spatial distribution of storage (none)	[0–1.5]
kb	Groundwater recharge constant (none)	[0–0.05]
tr1	Rate of outflow from surface storage (none)	[0–0.2]
tr2	Rate of outflow from groundwater storage (none)	[0.2–1.0]
cf	Minimum storage threshold (mm)	[0–30]

[29] We calibrate the model using two methods: a GLUE approach using simple Monte Carlo sampling and specifying a generalized likelihood based on the catchment's FDC; and an ABC approach that uses sequential Monte Carlo (SMC) sampling and a summary statistic of the same form as the GLUE generalized likelihood. We set this summary statistic (or generalized likelihood) as the mean absolute error (MAE) of the observed and simulated FDC. Adapting the method used by *Westerberg et al.* [2011], we specify 30 evaluation points (EPs) along the FDC based on equal intervals of observed discharge and use these to calculate MAE. In both GLUE and ABC approach, we specify a model error tolerance, ϵ, (or limit of acceptability) equal to 5% of the FDC exceedance percentage. It should be noted that this tolerance may be based on discharge measurement error as in *Westerberg et al.* [2011], but in the absence of stage-duration data, we select a value such that we can readily demonstrate the connection between GLUE and ABC.

[30] For our GLUE algorithm, we randomly sample candidate parameter sets from specified parameter ranges (Table 1) and simulate the catchment's FDC. Using GLUE terminology, parameter sets that provide an FDC within the identified limit of acceptability are deemed "behavioral" and retained to indicate the parameter uncertainty. For this application, we continue sampling until 500 behavioral parameters are obtained.

[31] Our ABC-SMC approach uses the algorithm developed by *Sisson et al.* [2007]. In SMC-based simulation, a population of parameters (referred to as particles) is evolved from an initial prior distribution, through a sequence of intermediary distributions, until it ultimately represents a sample from the target parameter distribution. In the ABC framework, particles are sampled from preceding populations, and each sampled particle is moved according to a Markov kernel to improve particle dispersion. A decreasing sequence of tolerances $\epsilon(t)$ is defined such that as $\epsilon(t) \rightarrow 0$ we obtain the target distribution of interest. Sampled particles that fall outside the prespecified tolerance are rejected. Repeated resampling (and moving) of particles is undertaken until a fixed number of particles are obtained for each population. For our case study, we define the sequence of error tolerances as an exponential decay function for the MAE of the FDC. Particles are initially sampled from the specified parameter ranges (Table 1). We sample 10 sequential populations with the final error tolerance set at 5% of the FDC. The population size is specified to be 500 (the same as our GLUE-based study).

[32] Weighted histograms of the resulting GLUE and ABC sampled parameters for the PDM are given in Figure 1. It should be noted that the GLUE-derived parameter distributions adhere reasonably well to the ABC-SMC parameters. Differences in the distributions may be attributed to sampling density, parameter insensitivity, or differences in the robustness of the simple Monte Carlo GLUE approach and the SMC algorithm. It is obvious that the GLUE algorithm applied here is simply an ABC rejection algorithm as described in Algorithm 2. However, these results also demonstrate the practical similarity between the GLUE and ABC algorithms for hydrologic applications. In the case presented here, the summary statistic is selected based on the lack of streamflow observations during the period of model simulation. In related cases (especially those regarding predictions in ungauged basins) summary statistics may be derived based on certain hydrologic signatures [*Wagener and Montanari*, 2011]. Following the connection outlined in the preceding connection, the use of generalized likelihoods based on such signatures combined with probabilistic error models may provide a way to explicitly model different sources of error but also address epistemic model errors. Recent discussions regarding the sufficiency of certain types of ABC summary statistics [*Blum et al.*, 2012] may thus prove useful to the hydrologic community for understanding the implications of certain types of objective functions in model inference.

4. Discussion

[33] By making a link between the GLUE algorithm and ABC, we have further clarified the relationship between GLUE and formal Bayesian methods. It is remarkable that GLUE is an example of an ABC method that predates the development of ABC [*Tavaré et al.*, 1997]. It might be argued that the link we have outlined here is merely technical. However, even if this is the case, there are some immediate and concrete benefits. The frenetic recent activity on developing more efficient ABC approximations and software [*Marjoram et al.*, 2003; *Sisson et al.*, 2007, 2009; *Beaumont et al.*, 2009] can be brought to bear for GLUE-based applications. A common objection to the use of GLUE revolves essentially around the computational inefficiency of importance sampling. For example, *Vrugt et al.* [2009b], in discussing the Differential evolution adaptive metropolis (DREAM) Markov chain Monte Carlo (MCMC) scheme of *Vrugt et al.* [2008, 2009a], state that "However, DREAM will have a much better efficiency in finding 'acceptable' models as it uses adaptive proposal updating to search for high quality solutions. Use of a simple uniform sampling distribution... as typically done in GLUE, can result in an algorithm that, even after billions of model evaluations, may only have generated a handful of good solutions...". Recent work has then shown how likelihood-based MCMC algorithms may be modified to address GLUE problems [e.g., *Blasone et al.*, 2008; *McMillan and Clark*, 2009]. We believe that the connection outlined here could further encourage adoption of new numerical methods for GLUE, and the availability of ABC software will also help in this regard. Once the distribution targeted by GLUE is understood, it becomes possible to advance the use of sophisticated SMC algorithms, Markov chain Monte Carlo algorithms, and so on with GLUE.

[34] The continuing interest in characterizing input uncertainty in hydrologic modeling (such as the Bayesian total error analysis [BATEA] method of *Kavetski et al.* [2006a, 2006b]) may also benefit from ABC methods. The quantification of hydrologic input error often requires parameter

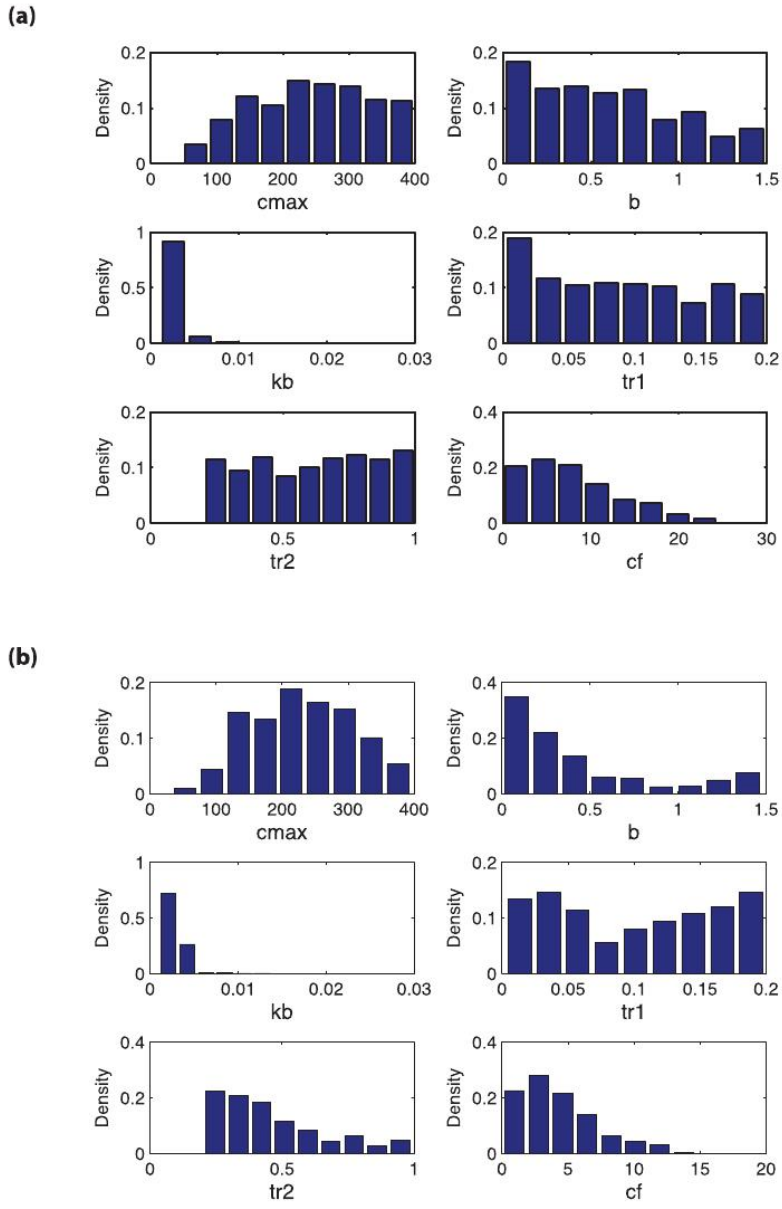


Figure 1. Weighted histograms of (a) sampled GLUE parameters and (b)ABC-SMC particles. A description of the model parameters may be found in Table 1.

sampling from very high-dimensional models due to the introduction of probabilistic models representing rainfall uncertainty. The ABC approach can provide a way of integrating out input error parameters so that model predictive uncertainty may be more easily addressed (see *Nott et al.* [2012] for initial work in this regard).

[35] The connection with ABC also hints at ways to extend GLUE where we want to specify both a probabilistic model as well as a generalized likelihood (since the generalized likelihood is identified with the ABC kernel and not the likelihood we are free to specify a likelihood as well). This allows us to model different sources of uncertainty but to also mitigate the effects of misspecification in directions of interest through choosing the kernel function based on a generalized likelihood. We are currently pursuing work based on these extensions.

[36] **Acknowledgments.** The authors thank Keith Beven, Jasper Vrugt, and a third anonymous reviewer for their thoughtful comments provided on this article. David Nott gratefully acknowledges the support and contributions of the Singapore-Delft Water Alliance (SDWA). The research presented in this work was carried out as part of the SDWAs multireservoir research program (R-264-001-001-272).

References

Beaumont, M., J.-M. Cornuet, J.-M. Marin, and C. Robert (2009), Adaptive approximate Bayesian computation, *Biometrika*, *96*, 983–990.

Beven, K. J. (2006), A manifesto for the equifinality thesis, *J. Hydrol.*, *320*, 18–36.

Beven, K. J. (2009), Comment on "Equifinality of formal (DREAM) and informal (GLUE) Bayesian approaches in hydrologic modeling?" by Jasper A. Vrugt, Cajo J. F. ter Braak, Hoshin V. Gupta and Bruce A. Robinson, *Stoch. Environ. Res. Risk Assess.*, *23*, 1059–1060.

Beven, K. J., and A. M. Binley (1992), The future of distributed models: Model calibration and uncertainty prediction, *Hydrol. Process.*, *6*, 279–298.

Beven, K. J., and I. Westerberg (2011), On red herrings and real herrings: Disinformation and information in hydrological inference, *Hydrol. Process.*, *25*(10), 1676–1680.

Beven, K. J., P. J. Smith, and J. E. Freer (2008), So just why would a modeller choose to be incoherent?, *J. Hydrol.*, *354*, 15–32.

Blasone, R.-S., J. A. Vrugt, H. Madsen, D. Rosbjerg, M. A. Robinson, and G. A. Zyvoloski (2008), Generalized likelihood uncertainty estimation (GLUE) using adaptive Markov Chain Monte Carlo sampling, *Adv. Water Res.*, *31*, 630–648.

Blazkova, S., and K. Beven (2009), A limits of acceptability approach to model evaluation and uncertainty estimation in flood frequency estimation by continuous simulation: Skalka catchment, Czech Republic, *Water Resour. Res.*, *45*, W00B16, doi:10.1029/2007WR006726.

Blum, M. G., M. A. Nunes, D. Prangle, and S. A. Sisson (2012), A comparative review of dimension reduction methods in approximate Bayesian computation, *Stat. Sci.*, in press.

Diggle, P. J., and R. J. Gratton (1984), Monte Carlo methods of inference for implicit statistical models, *J. R. Stat. Soc. Ser. B*, *46*, 193–227.

Gelman, A., J. B. Carlin, H. S. Stern, and D. B. Rubin (2003), *Bayesian Data Analysis*, 2nd ed., CRC Press, Boca Raton, Fla.

Gourieroux, C., A. Monfort, and E. Renault (1993), Indirect inference, *J. Appl. Econ.*, *8*, 85–118.

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., and B. L. McGlynn (2011), Hierarchical controls on runoff generation: Topographically driven hydrologic connectivity, geology, and vegetation, *Water Resour. Res.*, *47*, W11527, doi:10.1029/2011WR010666.

Kavetski, D., G. Kuczera, and S. W. Franks (2006a), Bayesian analysis of input uncertainty in hydrological modeling: 1. Theory, *Water Resour. Res.*, *42*, W03407, doi:10.1029/2005WR004368.

Kavetski, D., G. Kuczera, and S. W. Franks (2006b), Bayesian analysis of input uncertainty in hydrological modeling: 2. Application, *Water Resour. Res.*, *42*, W03408, doi:10.1029/2005WR004376.

Liu, Y., J. Freer, K. Beven, and P. Matgen (2009), Towards a limits of acceptability approach to the calibration of hydrological models: Extending observation error, *J. Hydrol.*, *367*, 93–103.

Mantovan, P., and E. Todini (2006), Hydrological forecasting uncertainty assessment: Incoherence of the GLUE methodology, *J. Hydrol.*, *330*, 368–381.

Marin, J.-M., P. Pudlo, C. P. Robert, and R. Ryder (2011), Approximate Bayesian computational methods, *Stat. Comput.*, *22*(6), 1167–1180, doi:10.1007/s11222-011-9288-2.

Marjoram, P., J. Molitor, V. Plagnol, and S. Tavaré (2003), Markov chain Monte Carlo without likelihoods, *Proc. Natl. Acad. Sci. U. S. A.*, *100*, 15,324–15,328.

McMillan, H., and M. Clark (2009), Rainfall-runoff model calibration using informal likelihood measures within a Markov chain Monte Carlo sampling scheme, *Water Resour. Res.*, *45*, W04418, doi:10.1029/2008WR007288.

Moore, R. J. (2007), The PDM rainfall runoff model, *Hydrol. Earth Syst. Sci.*, *11*, 483–499, doi:10.5194/hess-11-483-2007.

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*, W12528, doi:10.1029/2011WR011161.

Nott, D. J., Y. Fan, L. Marshall, and S. A. Sisson (2012), Approximate Bayesian computation and Bayes linear analysis: Towards high-dimensional ABC, *J. Comp. Graph. Stat.*, in press.

Sisson, S. A., Y. Fan, and M. Tanaka (2007), Sequential Monte Carlo without likelihoods, *Proc. Natl. Acad. Sci. U. S. A.*, *104*, 1760–1765.

Sisson, S. A., Y. Fan, and M. Tanaka (2009), Sequential Monte Carlo without likelihoods: Errata, *Proc. Natl. Acad. Sci. U. S. A.*, *106*, p. 16889.

Smith, T. J., and L. A. Marshall (2010), Exploring uncertainty and model predictive performance concepts via a modular snowmelt-runoff modelling framework, *Environ. Modell. Software*, *25*(6), 691–701.

Smith, P. J., K. J. Beven, and J. Tawn (2008), Informal likelihood measures in model assessment: Theoretic development and investigation, *Adv. Water Res.*, *31*, 1087–1100.

Stedinger, J. R., R. M. Vogel, R. Batchelder, and S. U. Lee (2008), Appraisal of the generalized likelihood uncertainty estimation (GLUE) method, *Water Resour. Res.*, *44*, W00B06, doi:10.1029/2008WR006822.

Tavaré, S., D. J. Balding, R. C. Griffiths, and P. Donnelly (1997), Inferring coalescence times from DNA sequence data, *Genetics*, *145*, 505–518.

Vrugt, J. A., C. J. F. ter Braak, M. P. J. M. Clark Hyman, and B. A. Robinson (2008), Treatment of input uncertainty in hydrologic modeling: Doing hydrology backward with Markov chain Monte Carlo simulation, *Water Resour. Res.*, *44*, W00B09, doi:10.1029/2007WR006720.

Vrugt, J. A., C. J. F. ter Braak, H. V. Gupta, and B. A. Robinson (2009a), Equifinality of formal (DREAM) and informal (GLUE) Bayesian approaches in hydrologic modeling?, *Stoch. Environ. Res. Risk Assess.*, *23*, 1011–1026.

Vrugt, J. A., C. J. F. ter Braak, C. G. H. Diks, B. A. Robinson, J. M. Hyman, and D. Higdon (2009b), Accelerating Markov chain Monte Carlo simulation by differential evolution with self-adaptive randomized subspace sampling, *Int. J. Nonlin. Sci. Numer. Simul.*, *10*, 273–290.

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.

Westerberg, I. K., J.-L. Guerrero, P. M. Younger, K. J. Beven, J. Seibert, S. Halldin, J. E. Freer, and C.-Y. Xu (2011), Calibration of hydrological models using flow-duration curves, *Hydrol. Earth Syst. Sci.*, *15*, 2205–2227, doi:10.5194/hess-15-2205-2011.

CHAPTER 3

THE APPLICATION OF APPROXIMATE BAYESIAN COMPUTATION AND SEQUENTIAL MONTE CARLO IN THE CALIBRATION OF HYDROLOGICAL MODELS

Contribution of Authors and Co-Authors

Manuscript in Chapter 3

Author: Jason J. Brown

Contributions: Researched the theoretical background of ABC-SMC, developed the code base for processing the ABC-SMC algorithm, completed the ABC-SMC case studies, analyzed the results, and wrote the initial draft.

Co-Author: Lucy A. Marshall

Contributions: Assisted in developing the ABC-SMC implementation and hydrological model, interpretation of the results, and edited the manuscript at all stages.

Co-Author: Robert A. Payn

Contributions: Assisted in conceptual development and final editing of the manuscript.

Co-Author: Mark C. Greenwood

Contributions: Assisted in conceptual development and final editing of the manuscript.

Manuscript Information Page

Jason J. Brown, Lucy A. Marshall, Rob A. Payn, Mark C. Greenwood
Hydrology and Earth System Sciences
Status of Manuscript:
__X_ Prepared for submission to a peer-reviewed journal
____ Officially submitted to a peer-review journal
____ Accepted by a peer-reviewed journal
____ Published in a peer-reviewed journal

Copernicus Publications

THE APPLICATION OF APPROXIMATE BAYESIAN COMPUTATION AND SEQUENTIAL MONTE CARLO IN THE CALIBRATION OF HYDROLOGICAL MODELS

The following work is currently in progress to be submitted for publication

Jason J. Brown[1], Lucy A. Marshall[1], Robert A. Payn[1], and Mark C. Greenwood[2]

1. Department of Land Resources and Environmental Sciences, Montana State University, Bozeman, MT 59717
2. Department of Mathematical Sciences Montana State University, Bozeman, MT 59717

Approximate Bayesian Computation (ABC) is a family of likelihood-free methods that have been recently introduced for calibrating hydrological models. This study assesses the application of an ABC Rejection scheme using a Sequential Monte Carlo sampling algorithm (ABC-SMC), and introduces the concepts of post-processing adjustments on the posterior distributions during hydrological model calibration. Model performance was evaluated using summary statistics from simulations comparing the hydrograph, flow duration curve, and a composite of the two. ABC-SMC was computationally efficient in calibrating a six parameter Probability Distribution Model for one synthetic and two real world data sets. Post-processing methods were implemented using local linear regression to adjust the final parameter estimations, generating marginal improvements to the posterior distributions. Statistical-based summary statistics performed well when the total uncertainty was low. As total uncertainty increased with the real world data sets, the composite summary statistic had the highest percentage of observed streamflow inside the 95% credible intervals.

1. Introduction

Since their introduction, hydrologic models have been an important tool in the study and management of water resources. Concerns for these resources have grown with the increasing levels of utilization and potential changes in the global climate patterns affecting resource availability (Matalas 1997, Milly et al. 2008, Taylor et al. 2013). To improve the efficacy of hydrologic model applications, quantifying the degree of uncertainty associated with these models is important to the appropriate

interpretations of their output (Beven 2013). Yet it has been argued that uncertainty analysis is under-utilized within the hydrological community (Pappenberger and Beven 2006). As the methods for quantifying uncertainty improve, incorporating uncertainty analysis into hydrological planning and projects is becoming more practical and meaningful for those who use model outputs to make management decisions (Pappenberger and Beven 2006).

With the increase in available computational power over the last two decades, methods of simulation-based model calibration and uncertainty estimation have seen rapid advancements. The development of Monte Carlo simulation techniques allows researchers to visualize many potential model outcomes. By extracting the metrics of the simulation results, estimates on the model uncertainty can be made (Beven and Freer 2001, Kuczera and Parent 1998, Marshall et al. 2004). Within hydrological research, uncertainty assessment is most commonly conducted using Bayesian inference with a Markov Chain Monte Carlo (MCMC) sampler (Marshall et al. 2004, Smith and Marshall 2008, Vrugt and Braak 2011, Vrugt et al. 2013) or Generalized Likelihood Uncertainty Estimation (GLUE) (Beven and Binley 1992, Freer et al. 1996, Zhang et al. 2011). More recently, another set of Bayesian methods, called Approximate Bayesian Computation (ABC), has emerged for calibrating ecological and environmental models.

The basic concepts of ABC were proposed as a way to estimate the posterior distribution in modeling scenarios where the likelihood function is either unknown or

cannot be explicitly defined (Diggle and Gratton 1984, Rubin 1984). The unavailability of the likelihood function may be due to mathematical intractability or when the likelihood is not defined in a closed form (Marin et al. 2012). The first applied uses of ABC were in the field of genetics to make inferences from DNA sequence data (Beaumont et al. 2002, Marjoram et al. 2003, Pritchard et al. 1999, Tavare et al. 1997) and more recent research has demonstrated the potential applicability of ABC for both ecological and epidemiology analysis (Beaumont 2010, Sisson et al. 2007, Toni and Stumpf 2010, Toni et al. 2009).

Likelihood-free methods aim to eliminate the need for solvable likelihood functions. By utilizing Monte Carlo sampling techniques in conjunction with an appropriate summary statistic, the posterior distribution of the parameters can be estimated. ABC implements a performance threshold that is used to remove areas of the parameter space whose parameter sets perform poorly. While the threshold is a binary operator, post-processing techniques can be used to adjust the posterior distributions of the parameters based on their performance as measured by the summary statistic. If the summary statistic is sufficient, then the estimated posterior distribution of the parameters will reflect the true posterior distribution under a formal Bayesian analysis (Joyce and Marjoram 2008). Therefore, when properly implemented, ABC may offer the flexibility of the generalized likelihood function utilized by GLUE while retaining the statistical rigor recognized in Bayesian Inference (Sadegh and Vrugt 2013).

In light of the recent interest in ABC computation in hydrology and other disciplines, the goal of this paper is to critically assess and review the use of ABC in a hydrologic application. Via a commonly used hydrologic model, we implement and assess an ABC algorithm using a Sequential Monte Carlo Sampler (ABC-SMC) for model calibration and uncertainty assessment. We investigate the usefulness of the approach across both synthetic and real case studies. We also demonstrate the effect of several summary statistics for model calibration, and look at the utility of post-processing techniques to adjust the posterior based upon the summary statistics. The goal of this paper is to provide benchmark guidance on the potential paths for ABC methods in hydrology.

2. Methods

2.1 Conceptual Rainfall Runoff Model

A widely used and broadly applicable conceptual rainfall-runoff model was selected to evaluate the effectiveness of the ABC-SMC algorithm for calibrating conceptual rainfall-runoff models. The conceptual rainfall-runoff model was based on the Probability Distribution Model (PDM), a lumped model that utilizes a probability function to represent the spatial variability of the soil moisture storage capacity across a catchment (Moore and Clarke 1981). Since being introduced in 1981, the PDM has developed into a framework that can account for multiple storage and routing components, allowing the model to be adapted to differing watershed systems (Moore 2007). The PDM has been implemented in hydrologic case studies investigating model

optimization (Jeremiah et al. 2012), predictions in ungauged basins (McIntyre et al. 2005), and assessing regionalization impacts on flood risks (Bell et al. 2007).

This study implements a version of the PDM that incorporates overland flow and baseflow (Figure 1). The probability distributed soil moisture store (*S*) receives inputs through precipitation (*P*). Water losses include actual evapotranspiration (*AET*) and percolation into an unconstrained groundwater storage (*GW)*. Routing components account for contributions to stream flow via direct runoff and ground water discharge. The water balance equation for the soil moisture storage at time step *t* can be expressed as:

$$S_t = \max(0,\ S_{t-1} + \quad P_t - AET_t - \ GW_t) \qquad (4)$$

The spatial variability in the soil moisture store is represented using a Pareto distribution that defines the available soil moisture capacity. This distribution is defined by two parameters: the maximum storage capacity of the basin (C_{max}) and the degree of spatial variability (c_b) in the Pareto distribution. The critical capacity of the storage at any given time is dependent upon the proportion of the basin generating runoff. The critical maximum storage (S_{max}) is derived from C_{max} and c_b (Equation 5). The relationships between storage capacity (C_t) and adjusted critical storage (S_t) at time *t* are defined as:

$$s_{max} = C_{max}/(c_b + 1) \qquad (5)$$

$$s_t = s_{max} * \left(1 - \left(1 - \frac{C_{t-1}}{C_{\max}}\right)\right)^{c_b+1} \quad (6)$$

At each time step, the catchment infiltration (I_t) can be estimated based on the observed precipitation (P_t), the critical storage of the previous time step, and the maximum critical storage capacity. Any precipitation that is in excess of the infiltration capacity contributes to stream flow as surface runoff (P_r).

The soil moisture storage has two paths for water outputs; *AET* and percolation to the groundwater storage. The *AET* loss is based upon the potential *ET* (*PET*) adjusted by the percentage of the critical capacity of the soil moisture storage and an exponent (b_e) to define the linearity of the relationship between *ET* and *AET* (Equation 7). The model incorporates groundwater recharge (GWR) from the soil moisture store. This route is implemented as a rate (GW_r) and is activated when the critical storage is greater than a minimum basin storage threshold (C_{min}) (Equation 8).

$$AET_t = PET_t \left(\frac{s_{t-1}}{s_{max}}\right)^{b_e} \quad (7)$$

$$if\ s_t > C_{min};\ GWR_t = GW_r * S_{t-1} \quad (8)$$

The groundwater discharge (*GWD*) is implemented as a rate (GW_q) of available GW from the previous time step (Equation 9).

$$GWD_t = GW_q * GWD_{t-1} \quad (9)$$

The total stream flow from each step on the hydrograph was the sum the surface runoff and GWD (Equation 10).

$$Q_s = P_r + GWD_t \quad (10)$$

2.2 Summary statistics

We selected three different summary statistics to evaluate the model goodness-of-fit to the observed data. These summary statistics represent three categories of metrics that can be used in ABC algorithms. These categories include statistical measures of the goodness-of-fit between the observed and predicted hydrograph, signature-based measures that incorporate contextual information instead of a pure statistical measure of fit (Vrugt and Sadegh 2013), and composite summary statistics which consist of measures that represent different pieces of information applicable to the question of interest and then reduced into a one dimensional measure (Beaumont 2010, Blum et al. 2013).

The Nash-Sutcliffe Efficiency (NSE) represents a statistical measurement that is well established within the hydrological literature (Gupta and Kling 2011, Schaefli and Gupta 2007). The NSE is based upon the normalization of the mean square error by the variance of the observed data (Nash and Sutcliffe 1970). This represents a signal-to-noise measure of performance (Schaefli and Gupta 2007) and implicitly measures the proposed model compared to the simplest model possible, allowing relative comparisons between models of the same context (Schaefli and Gupta 2007). While the standard NSE score ranges from negative infinity to one, this study adjusted the NSE score to a scale of zero to one, with zero being the optimum (Equation 11). This

allowed all three summary statistics to have a common optimum value and to help facilitate the use of the composite summary statistic.

$$\text{Minimized NSE } = 1 - \max\left(0,\ 1 - \frac{\sum_{t=1}^{T}(Q_o^t - Q_p^t)^2}{\sum_{t=1}^{T}(Q_o^t - \overline{Q_o})^2}\right) \quad (11)$$

where *t* is the daily time step, Q_p is the observed daily streamflow and Q_o is the predicted daily streamflow.

In contrast, the use of the Sum of the Absolute Difference of the Flow Duration Curve as an summary statistic focuses on reproducing a hydrologic signature that might represent broader hydrologic patterns (Westerberg et al. 2011) or physical properties (Gupta et al. 2008) of a watershed. This approach may be useful when implementing highly complex models, dealing with non-overlapping data sets, or if there is insufficient data for a statistical measure to be effective (Gupta et al. 2008, Westerberg et al. 2011) . Using the flow duration curve as an summary statistic may be beneficial as it has been reported to be less sensitive to flow magnitudes and epistemic errors (Westerberg et al. 2011). In this analysis, the flow duration curves were compared using twenty-five equally sized bins for the predicted and observed discharge (Equation 12).

$$\text{S.A.D of the Flow Duration Curve} = \sum_{i=1}^{25}\left|P\left(B_p^i\right) - P\left(B_o^i\right)\right| \quad (12)$$

where *b* is the bin number, $P\left(B_p^i\right)$ is the probability associated with the signature statistic for the predicted streamflow and $P\left(B_o^i\right)$ is the probability associated with the signature statistic of the observed streamflow.

The third summary statistic, a composite statistic, was an equally weighted linear combination of the hydrograph statistic and the signature statistic. The composite statistic was implemented to explore the effect of combining statistical and signature-based summary statistics. As the hydrograph statistic and signature statistic are on different scales, the maximum signature statistic value recorded during a tuning run was used to normalize the signature statistic values. The composite statistic was then derived from the mean of the hydrograph statistic and the normalized signature statistic (Equation 13).

$$\text{composite statistic} = \left(Minimized\ NSE + \frac{S.A.D\ FDC}{\max(S.A.D\ FDC)}\right)/2 \qquad (13)$$

2.3 ABC-SMC Algorithm

Multiple implementations of ABC have been proposed using different sampling algorithms. The basic ABC Rejection sampler (Beaumont et al. 2002) is considered an inefficient method for the exploration of the parameter space (Sisson et al. 2007). Later adaptations have been proposed to improve the sampling efficiency by implementing more complex sampling algorithms, such as Markov Chain Monte Carlo (Marjoram et al. 2003) and Sequential Monte Carlo (SMC) samplers (Beaumont et al. 2009, Sisson et al. 2007, Toni et al. 2009). Reported benchmarks indicated that ABC-SMC provided similar estimations as the ABC Rejection sampler with up to fifty times fewer particles (Toni et al. 2009). Several recent ABC-SMC implementations have been proposed that serve as

the benchmark algorithms currently used (Beaumont et al. 2009, Sisson 2009, Toni et al. 2009).

ABC-SMC uses an optimized sampling scheme to allow for a more efficient exploration of the parameter space (Toni et al. 2009). Within ABC-SMC, the parameter sets, called particles, are processed through a series of intermediate distributions. Each intermediate distribution is based upon a collection of particles that meets the threshold criteria and forms a population of parameter sets. An Importance Sampling weight is used to weight each particle in the population. Successive populations are comprised of particles sampled from the previous population based on weighted random sampling. The selected particles are then shifted with a perturbation kernel to randomly generate a new parameter set. By reducing the threshold for each population, poorly performing areas of the parameter space are eliminated from further exploration. The final populations will occupy areas of the parameter space that contain high densities of behavioral particles. The ABC-SMC algorithm proposed by Toni (Toni et al. 2009) is as follows:

1. $At\ iteration\ i = 1,$ (*Algorithm 1)*

   $For\ j = 1, \ldots\ N$

   1.1 $Sample\ \theta_j^{**}\ independently\ from\ \pi(\theta)$

   $if\ d(x^*, x_0) \geq\ \varepsilon_i\ return\ to$ 1.1

   $Set\ \omega_i^j = \frac{1}{N}$

2. $At\ iteration\ 2\ \leq i\ \leq I,$

   $For\ j = 1, \ldots\ , N, repeat$

   2.1 $Sample\ \theta_j^*\ from\ the\ \theta_{(i-1)}^{(j)} with\ weights\ \omega_{i-1}$

   $Obtain\ \theta_j^{**} \sim K_i(\theta^*|\theta)$

   $Simulate\ a\ candidate\ data\ set\ x^* \sim$ f$(x|\theta^{**})$

   $if\ d(x^*, x_0) \geq\ \varepsilon_i\ return\ to$ 2.1

   2.2 $Set\ \theta_i^{(j)} =\ \theta^{**}\ and\ calculate\ the\ weight\ for\ particle\ \theta_i^{(j)}$

$$\omega_i^{(j)} = \frac{\pi\left(\theta_i^{(j)}\right)}{\sum_{i=1}^{N} \omega_{i-1}^{(j)} * K_i\left(\theta_i^{(j)} \middle| \theta_{i-1}^{(j)}\right)}$$

3. $Normalize\ the\ Weights$

$$\omega_i^{(j)} = \frac{\omega_i^{(j)}}{\sum \omega_i^*}$$

4. $If\ \ i = I, apply\ local\ linear\ regression\ adjustment\ to\ \theta_i$

where *i* is the current population, *T* is the number of populations, *j* represents particles, and *N* is the number of particles present in each population. Parameter sets associated with a particle are represented by θ and are sampled from the joint prior distribution, $\pi(\theta)$. A kernel function, K_i, is used to perturb samples particles beginning with the second population. Each particle has a summary statistic value, $d(x^*, x_0)$ and an Importance Sampling weight, ω_j.

For this project, each population consisted of 1000 particles. The parameters were initially defined with uniform distributions (Table 1) based on a tuning run, which consisted of a single population of 5000 particles. A multivariate Gaussian kernel

utilizing an adaptive covariance matrix was used for the particle perturbations, K_i, and Importance Sampling weights. The tuning runs were also used to define the threshold decay, ε_i. The 0.5 ($X_{0.5}$), 1.0 ($X_{1.0}$), and 25th (X_{25}) percentiles from the tuning run were recorded and used in conjunction with a geometric decay for the threshold. The resulting threshold value at the i^{th} iteration was:

$$\varepsilon_i = X_{0.5} + X_{25}(0.5)^{i-1} \quad (14)$$

This function was selected to provide a smooth but rapid decay rate as the ABC-SMC moved through the sequence of populations. Subsequent populations were processed until all particles had a summary statistic value below $X_{1.0}$.

2.4 Posterior Post-Processing

The final population for each calibration process was adjusted using local linear regression. Regression methods can be used as a post-processing technique to model the parameter values as a function of the summary statistic. The resulting model is then used to adjust the parameter values for the particles that define the posterior distribution. The use of a local linear regression is used to help satisfy the linearity and additive assumptions that may not be met globally (Beaumont et al. 2002). The general form of the local linear regression is:

$$\theta_i^* = \theta_i - (\boldsymbol{s}_i - \boldsymbol{s})^T \hat{\beta} \quad (15)$$

where $(\boldsymbol{s}_i - \boldsymbol{s})$ is the scalar or vector distance between the summary statistic of a parameter set and theoretical optimal value. The weights associated with $\hat{\beta}$ are based on a non-parametric kernel and the distance of $(\boldsymbol{s}_i - \boldsymbol{s})$ (Equation 2).

$$K_\delta(\llbracket \boldsymbol{s}_i - \boldsymbol{s} \rrbracket) \quad (16)$$

where δ is a tolerance value that determines the sphere of influence. The posterior density of the adjusted particles ($\hat{\pi}(\theta_0|\boldsymbol{s})$) is then approximated using a density-estimation bandwidth (Δ) and the weighting kernel (Equation 3).

$$\hat{\pi}(\theta_0|\boldsymbol{s}) = \frac{\sum_i K_\Delta(\theta_i^* - \theta_0) K_\delta(\llbracket \boldsymbol{s}_i - \boldsymbol{s} \rrbracket)}{\sum_i K_\delta(\llbracket \boldsymbol{s}_i - \boldsymbol{s} \rrbracket)} \quad (17)$$

In this study, the particles in the final populations were adjusted using the local linear regression functions within the ABC library for R (Csilléry et al. 2012). These methods account for potential issues of collinearity and heteroscedasticity between the parameters in the model. The adjusted particles were then processed through the PDM model to generate a summary statistic for the LLR adjusted parameter sets.

2.5 Calibration and Model Assessment

For each calibration run, diagnostic plots were used to evaluate the ABC algorithm and the uncertainty analysis. Dotty plots were generated to visualize the evolution of the parameter space through the populations by plotting the parameter value versus the summary statistic value. Parameter uncertainty was assessed with histograms, QQ-Norm plots, and kernel density plots of the marginal distributions for the estimated parameters. Predictive uncertainty was assessed with the use of predicted versus observed plots, predicted hydrographs, the mean width of the 95% quartile based credible interval, and the percentage of the observed streamflow that were outside the 95% quartile based credible interval for the predicted streamflow.

2.6 Hydrologic Data

To demonstrate the effectiveness of the ABC-SMC algorithm over a variety of hydrologic conditions, three separate data sets were used during this study. Two of these data sets consisted of real world hydrological data with daily precipitation, stream flow and estimated PET spanning the time period from June 12, 1972 to December 09, 1982. To test the algorithm under known conditions, a third data set with synthetic streamflow was derived using measured precipitation and ET.

The real world data sets were chosen to examine two diverse catchments with differing climatic conditions, hydrological processes, and hydrological signatures. The first was data set was recorded at Station 307201 on the Davey River in Tasmania (Figure 2). The catchment covers an area of 686 km^2 with a mean annual precipitation of 2322 mm and a mean annual runoff of 2068 mm. Therefore, this catchment has a high runoff ratio and shows predominately perennial flows with a mean daily streamflow of 2.7 mm. During the sample period used, 3.6% of the days had zero flow recorded.

The second site was recorded at Station 119003 on the Haughton River at Powerline in Queensland, Australia (Figure 3). The watershed covers an area of 1773 km^2, with a mean annual precipitation 898 mm. The mean annual runoff is 220 mm with a mean stream flow of 0.008 mm. Within the sample period, 29.4% of the days had an observed streamflow of zero.

The third data set was synthesized using the implemented PDM model. The precipitation and ET observations from Davey River were used to generate synthetic runoff by fixing the values of the six unknown parameters as follows: c_b = 2.0, C_{min} = 20mm, C_{max} = 100mm, b_e = 2.0, GW_r = 0.1, GW_d = 0.5. White noise (*N(0,0.05)*) was added to the synthetic runoff to introduce a small amount of aleatory error in the streamflow.

2.7 Case Studies

The ABC-SMC algorithm was implemented for three separate case studies corresponding to the data sets. Through these case studies, the utility, flexibility and robustness of the ABC-SMC approach was investigated.

2.7.1 Case Study 1: Synthetic Data

The first case study was designed to examine the effectiveness and robustness of the ABC-SMC algorithm under known conditions. The ABC-SMC algorithm was implemented for three scenarios to demonstrate the ability of the approach to derive parameter distributions comparable to the parameters used to generate the synthetic data. The first scenario compared the results from the three calibration processes using the synthetic data, the full model, and the three summary statistics. In this scenario, the parameter and predictive uncertainties from the three calibration processes were compared through the diagnostic plots and 95% credible intervals for the predicted streamflow.

The second scenario was conducted using a reduced version of the PDM model. In this scenario, the C_{min} parameter was removed from the model. This introduced model uncertainty into the calibration process and was used to address whether ABC-SMC would be sensitive to introduced uncertainty due to model structure.

2.7.2 Case Study 2: Davey River, Tasmania

The second case study used the Davey River data. This data set was used to examine the utility of ABC-SMC on a hydrograph demonstrating regular precipitation events and perennial stream flows (Figure 2). For each summary statistic, the model was calibrated with the observed data. Summary plots and the 95% credible interval were then used to assess the parameter and predictive uncertainty for each summary statistic.

2.7.3 Case Study 3: Haughton River, Queensland

The third case study was conducted using the Haughton River data set. This data set contains a high percentage of zero data values nestled between precipitation events (Figure 3). As with the second case study, the model was calibrated for each summary statistic using the observed data. Summary plots and the 95% credible interval were then used to assess the parameter and predictive uncertainty for each summary statistic.

3. Results

3.1 Case Study 1

The ABC-SMC algorithm (Section 2.4) was implemented to calibrate the PDM model using the synthetic data set. This case was used to assess the effectiveness of the calibration process for estimating streamflow from known parameters values using the implemented model.

During the ABC-SMC calibrations with the synthetic data, the groundwater recharge rate (GW_r) (Figure 5e) strongly converges to the true parameter value (0.1). The minimum soil storage (C_{min}) (Figure 5d) and groundwater discharge rate (Figure 5e) shows the least convergence with a relatively flat final marginal distribution.

Model parameters were generally unimodal and symmetric (e.g. Figures 6a-b, 7a-b). Comparisons were made between the unadjusted marginal distributions and those adjusted using local linear regression and posterior weighting. Using the hydrograph statistic results as an example, the LLR adjustment for the Minimum Soil Moisture Storage (C_{min}) (Figure 6c-d) shows a reduction to the tails of the distribution as the particles are adjusted towards the mode. The distribution for the Groundwater Recharge Rate (GW_r) (Figure 7c-d) shows a reduction in the skewness and a narrowing of the overall distribution.

When calibrating to the synthetic data set, the lowest parameter and predictive uncertainty occurred when the hydrograph statistic was used as the summary statistic. All parameters had modes close to the true value (Table 2, Figures 5 and 8). The Ground

Water Recharge Rate (GW_r) showed the least uncertainty and a well-defined mode, while the other five parameters had more uniform distributions. When using the signature-based metric, the parameter uncertainty increased and the weighted medians and modes for two parameters, Minimum Soils Moisture Storage (C_{min}) and Maximum Soil Moisture Storage (C_{max}), were underestimated (Table 3, Figure 8). The results using the composite summary statistic had more uncertainty in the parameter distributions (Figure 8) than those using the hydrograph statistic, but the weighted medians were close to the true parameter values (Table 4).

None of the three calibration process showed significant bias in the predicted hydrographs (Table 5). When using the hydrograph statistic, the predicted versus observed streamflow (Figure 9a) showed more deviations at low flows, but had a strong association. The predicted hydrograph (Figure 9b) had the narrowest mean width for the credible intervals, but also had highest percentage of observed streamflow outside the 95% credible interval (Table 5).

The results from using the signature-based metric showed a strong association between the predicted and observed streamflow (Figure 10a), though there are more deviations at the lower flows. These results had the widest mean width for the 95% credible interval of the predicted hydrograph (Table 5, Figure 10b) and had fewer observed streamflow outside the 95% credible interval. The composite summary statistic results had a stronger association between the predicted and observed streamflow (Figure 11a) than those using the signature-based metric, but showed more

deviations at low streamflow than the results using the hydrograph statistic. The results using the composite summary statistic had the fewest observed streamflow outside the 95% credible interval while having a mean credible interval width between the other calibration results (Table 5).

When calibrating the synthetic data with the hydrograph statistic and the reduced model, the parameter uncertainty increased. When the Minimum Soil Moisture Storage (C_{min}) was removed, both the Maximum Soil Moisture Store (C_{max}) and the Groundwater Discharge Rate (GW_d) were underestimated (Figures 12).

3.2 Case Study 2

The weighted medians (Tables 6 and 8) and distributions (Figure 13) for the results using the hydrograph and composite summary statistics exhibited similar parameter estimates and uncertainty for all six parameters. The results using the signature-based metric (Table 7) for the Spatial Variability Coefficient (c_b), Minimum Soil Moisture Storage (C_{min}), and Groundwater Recharge Rate (GW_d) had different modes than the results using the other two summary statistics. The parameter uncertainty was also higher with the results using the signature-based metric, especially with the Groundwater Recharge Rate (GW_r) and the Groundwater Discharge Rate (GW_d).

None of the three calibration processes showed significant bias in the predicted hydrographs (Table 9). The results using the hydrograph statistic tended to underestimate the observed streamflow for observed streamflow between 15 and 25mm, but had good representation of the streamflow events (Figure 14b). The results

using the signature-based metric were in the middle of the three summary statistics in terms of the mean credible interval width and the percentage of observed streamflow outside the 95% credible interval (Table 9).

The results using the signature-based metric had a weaker association in the predicted versus observed streamflow, though it appears better centered than the results using the hydrograph statistic. The predicted hydrograph had good representation of the streamflow events. These results had the narrowest mean credible interval width, but also had the largest percentage of observed streamflow outside the 95% credible interval (Table 9). The results using the composite summary statistic were similar to those using the signature-based metric (Figure 15). The mean 95% credible interval width was the largest of the three summary statistics and the percentage of observed streamflow outside the credible interval was the smallest (Table 9).

3.3 Case Study 3

The three summary statistics showed similar results (Figure 17, Tables 10-13) for the Evapotranspiration exponent (b_e) and the Groundwater Recharge Rate (GW_r). The mode and uncertainty for the Maximum Soil Storage (C_{max}) and the Spatial Variability Coefficient (c_b) were most similar between the results from the hydrograph composite summary statistics. The mode and uncertainty for the Minimum Soil Moisture Storage (C_{max}) and the Groundwater Discharge Rate (GW_d) were most similar between the signature-based metric and the composite summary statistic results.

The results using all three summary statistics had similar predictive uncertainty when comparing predicted and observed streamflow (Figures 18a, 19a, 20a). The streamflow events were equally replicated (Figures 18b, 19b, 20b), but the mean width of the 95% credible intervals for the results using the hydrograph statistic were narrower than the other results (Table 13). The results using the composite summary statistic had a higher mean 95% credible interval width than those using the signature-based statistic, but had the smallest percentage of observed streamflow outside the credible interval (Table 13).

4. Discussion

This study demonstrates the potential usefulness and limitations using ABC-SMC algorithms for calibration of lumped conceptual watershed models with minimal parameter requirements. This included an assessment of the convergence efficiency of the algorithm, the effects of post-processing adjustments, and the effects of the three types of summary statistics on the parameter and predictive uncertainty on the calibrated models.

4.1 ABC-SMC Algorithm Implementation

When implementing ABC-SMC, there are multiple factors that affect the efficiency of the algorithm. These include the number of unknown parameters in the model, the initial bounds on the parameters, the threshold decay function, and the target population size. The first two factors affect the size of the joint parameter space that the ABC-SMC algorithm is required to search. Increasing the number of parameters or

their initial bounds will expand the parameter space, requiring additional particles and intermediate populations during the calibration process. While the number of parameters is dependent on the model, using tuning runs as an exploratory tool to help define the joint parameter spaces can be useful in narrowing the initial bounds and increasing the efficiency of the algorithm.

The threshold decay function has a substantial impact on the efficiency of the ABC-SMC algorithm. When possible, statically defined thresholds can be implemented, though this requires prior knowledge and assumptions regarding the potential values for the summary statistics. An ideal decay rate would have a rapid decay for the initial intermediate populations to quickly eliminate the poorest performing areas of the parameter space. The decay rate should slow as the process advances through later intermediate populations, allowing a more thorough exploration as the joint parameter space decreases in size. A geometric decay rate was found to be an effective function for threshold decay rate.

The target number of particles for each population is the final factor that was addressed. The number of particles will depend upon the size of the parameter space, the computational cost to process each particle through the model, and total uncertainty. For the data sets calibrated for this project, a population size of 1000 particles was considered adequate for properly representing the marginal distributions of the parameters. With this targeted population size, the average number of particles processed for each calibration run was 78,530 particles. By comparison, research

comparing several MCMC algorithms to calibrate a similar PDM model using different data required approximately 2.5 to 4 times more model runs (Smith and Marshall 2008).

4.2 Post-Processing with Local Linear Regression

One of the potential strengths of ABC methodologies is the ability to conduct post-processing adjustments on the joint distribution of the posterior. These methods may be implemented for all the populations in the calibration process, or as an adjustment to the final population. For this study, a local linear regression method was implemented to adjust the posterior distribution of the final population.

The use of the local linear regression had marginal benefits for the posterior distributions. The QQ-Norm plots (eg Figures 6b,d and 7b,d) did not indicate the presence of normality violations resulting from the LLR adjustments. In general, the LLR adjustments did adjust the parameter values closer to the mode and reduced the overall variability (eg Figure 6a,c and Figure 7a,c). However, using the multivariate local linear regression did lead to particle values being adjusted beyond the boundary conditions as defined by the initial prior distributions. This indicates that there may have been violations in the boundary assumptions when using local linear regression as a post-processing method in these case studies. Alternative methods have been proposed that implement non-linear or conditional adjustments to the final parameters (Blum et al. 2013, Blum and François 2010). These methods, such as non-linear regression, ridge regression, and neural networks, may be more appropriate for making adjustments to the posterior distributions when calibrating hydrologic models with ABC methods.

4.3 Parameter and Predictive Uncertainty

4.3.1 Synthetic Data Set

For the synthetic case study, the modes of the parameter density plots using the hydrograph statistic (Figure 8) were more consistently associated with the known values than those using the signature-based and composite summary statistics. The parameter estimates using the signature-based statistic were the least consistent. There were no real trends visible concerning the variances of the parameter distributions in relation to the summary statistic. The final parameter estimates yielded similar predictive results (Figures 9a, 10a, and 11a), though the 95% credible intervals (Figures 9b, 10b, 11b) differ. Based on these plots and the percentage of observed streamflow outside the 95% credible interval (Table 5), using the hydrograph summary statistic predicted the value of the parameters better than the signature-based and composite summary statistics, but underestimated the predictive uncertainty. The results from the synthetic case study suggest that statistical summary statistics may perform the best when the total uncertainty is small as they may capture the most information when comparing the goodness-of-fit between the observed and predicted streamflow.

4.3.2 Davey River and Haughton River Data Sets

For the two real world data sets, the increasing uncertainty within the data sets affected the performance of the summary statistics. For the perennial flow data in Case 2, the results using the signature-based metric had the narrowest 95% credible interval for the predicted hydrograph. Consequently, it also had the highest percent of observed

streamflow that fell outside of the 95% credible interval. The results using the composite summary statistic had the widest mean width and the lowest number of observed streamflow outside the 95% credible interval. For Case 2, using the signature-based statistic resulted in an underestimated credible interval and poor predictive uncertainty when compared to the hydrograph statistic and composite statistic.

For the ephemeral flow data used in Case 3, the results were more variable. Approximately 29.4% of the daily recorded streamflow were zero, thus a signature-based summary statistic may be considered more appropriate (Westerberg et al. 2011). The calibration results using the hydrograph statistic had the narrowest 95% credible interval and the largest percent of observed streamflow outside the credible interval. The results from the signature-based and composite summary statistics had similar widths for their 95% credible intervals, but the composite summary statistic had a substantially lower percentage of observed streamflow outside the credible interval.

The results from the two real world data sets show potential benefits from implementing signature-based and composite summary statistics. As the total uncertainty within the data and model structure increases, these two categories can harness additional contextual information or reduce the uncertainty through smoothing processes that improve the predictive uncertainty present in the final model. Using the hydrograph statistic as a summary statistic tended to generate over-fitted models with 95% credible intervals that were too narrow, leading to higher predictive uncertainty. Using the composite statistic as the summary statistic yielded the lowest percentage of

predictions that were outside the 95% credible interval. This suggests that incorporating both the hydrograph statistic and signature statistic into a composite summary statistic reduced the predictive uncertainty in the final predicted model.

5. Conclusion

For the watershed hydrologist, ABC methodologies are a promising family of techniques for calibrating watershed models. These methods have multiple benefits, including the ability to calibrate models with intractable likelihood functions, implementing uncertainty analyses based on formal Bayesian inference, post-processing techniques to adjust the posterior distributions, and the potential of utilizing innovative summary statistics that can be tailored to the hydrological characteristics of individual basins. When coupled with efficient sampling algorithms, such as SMC, calibrating hydrologic models using ABC methods requires less computational effort.

While the general algorithms for ABC-SMC have been well documented, two components of the ABC process should be of primary interest to the hydrological modeler. The first is the ability to use multivariate post-processing methods to adjust the posterior distributions. The post-processing can be conducted with existing methodologies, such as LLR and neural networks, but the application of specific methods should be evaluated within the context of hydrological modeling.

The second benefit of ABC methods for hydrologists is the ease in which composite summary statistics that can be implemented in order to incorporate signature-based information into a calibration scheme. This facilitates two important capabilities. The

first is to use performance measures that are tailored to the specific characteristics of a watershed. The second benefit is the potential for calibrating models to watersheds where the observed data is sparse or non-contiguous. The use of signature-based or composite summary statistics with ABC methods is a promising development in calibrating hydrological that needs further exploration and evaluation.

©2013 The MathWorks, Inc. MATLAB and Simulink are registered trademarks of The MathWorks, Inc. See www.mathworks.com/trademarks for a list of additional trademarks. Other product or brand names may be trademarks or registered trademarks of their respective holders.

6. Tables

	Parameter					
Case	c_b	C_{max}	b_e	C_{min}	GW_r	GW_d
Synthetic	0-3.0	5-200	0-3.0	0-40.0	0-1.0	0-1.0
perennial flow	0-3.0	50-750	0-3.0	0-200.0	0-1.0	0-1.0
Haughton Stream	0-3.0	50-750	0-3.0	0-200.0	0-1.0	0-1.0

Table 1: Uniform Prior Distributions for the three case studies

Parameter	Weighted Median	Weighted 2.5 Percentile	Weighted 97.5 Percentile
c_b	2.126	1.771	2.870
C_{max}	106.608	45.786	183.379
b_e	2.232	1.738	2.864
C_{min}	20.203	2.719	36.395
GW_r	0.090	0.049	0.133
GW_q	0.549	0.073	0.980

Table 2: Parameter Estimations for the Synthetic data set using the hydrograph summary statistic for Case Study 1.

Parameter	Weighted Median	Weighted 2.5 Percentile	Weighted 97.5 Percentile
c_b	2.015	0.927	2.858
C_{max}	67.802	23.917	150.969
b_e	2.257	1.451	2.903
C_{min}	13.970	0.000	37.971
GW_r	0.199	0.012	0.738
GW_q	0.555	0.242	0.875

Table 3: Parameter Estimations for the Synthetic data set using the signature-based summary statistic for Case Study 1.

Parameter	Weighted Median	Weighted 2.5 Percentile	Weighted 97.5 Percentile
c_b	2.134	0.965	3.065
C_{max}	96.051	47.972	174.716
b_e	2.186	1.456	3.016
C_{min}	17.736	4.570	37.817
GW_r	0.078	0.006	0.217
GW_q	0.549	0.163	0.985

Table 4: Parameter Estimations for the Synthetic data set using the composite summary statistic for Case Study 1.

	hydrograph statistic	signature statistic	composite statistic
Mean Width of 95% Credible Interval	0.45	1.59	1.36
Percent Outside Credible Interval	11.30	3.05	2.53
Percent Below Observed	46.27	47.30	49.62
Percent Above Observed	53.73	52.70	50.38

Table 5: Predictive Uncertainty for the Synthetic Data Set for Case Study 1.

Parameter	Weighted Median	Weighted 2.5 Percentile	Weighted 97.5 Percentile
c_b	1.522	0.557	2.669
C_{max}	373.028	146.98	694.364
b_e	0.948	0.583	1.716
C_{min}	136.307	55.004	195.892
GW_r	0.835	0.652	1.022
GW_q	0.490	0.387	0.621

Table 6: Parameter Estimations for the perennial flow data set using the hydrograph summary statistic for Case Study 2.

Parameter	Weighted Median	Weighted 2.5 Percentile	Weighted 97.5 Percentile
c_b	0.987	0.124	2.499
C_{max}	503.831	192.824	754.857
b_e	0.762	0.291	1.575
C_{min}	117.360	6.491	200.656
GW_r	0.796	0.578	1.017
GW_q	0.771	0.598	0.969

Table 7: Parameter Estimations for the perennial flow data set using the signature-based summary statistic for Case Study 2.

Parameter	Weighted Median	Weighted 2.5 Percentile	Weighted 97.5 Percentile
c_b	1.650	0.184	2.987
C_{max}	316.578	118.677	584.466
b_e	0.993	0.481	2.413
C_{min}	105.084	7.883	207.184
GW_r	0.804	0.644	0.986
GW_q	0.689	0.504	0.882

Table 8: Parameter Estimations for the perennial flow data set using the composite summary statistic for Case Study 2.

	hydrograph statistic	signature statistic	composite statistic
Mean Width of 90% Credible Interval	4.47	1.76	5.55
Percent Outside Credible Interval	35.66	65.67	31.10
Percent Below Observed	53.96	54.70	55.46
Percent Above Observed	46.04	45.30	44.54

Table 9: Predictive Uncertainty for perennial flow for Case Study 2.

Parameter	Weighted Median	Weighted 2.5 Percentile	Weighted 97.5 Percentile
c_b	1.970	0.382	3.129
C_{max}	478.205	234.012	704.006
b_e	0.103	0.000	0.440
C_{min}	126.325	83.583	158.456
GW_r	0.738	0.449	1.062
GW_q	0.050	0.00	0.228

Table 10: Parameter Estimations for the Haughton Stream data set using the hydrograph summary statistic for Case Study 3.

Parameter	Weighted Median	Weighted 2.5 Percentile	Weighted 97.5 Percentile
c_b	1.187	0.206	2.492
C_{max}	495.205	210.832	759.226
b_e	0.067	0.000	0.673
C_{min}	165.440	59.013	208.984
GW_r	0.339	0.000	1.162
GW_q	0.265	0.000	0.844

Table 11: Parameter Estimations for the Haughton Stream data set using the signature-based summary statistic for Case Study 3.

Parameter	Weighted Median	Weighted 2.5 Percentile	Weighted 97.5 Percentile
c_b	1.775	0.000	3.002
C_{max}	474.629	104.958	701.161
b_e	0.130	0.000	0.451
C_{min}	133.772	91.185	166.192
GW_r	0.779	0.503	1.013
GW_q	0.053	0.000	0.277

Table 12: Parameter Estimations for the Haughton Stream data set using the composite summary statistic for Case Study 3.

	hydrograph statistic	signature statistic	composite statistic
Mean Width of 90% Credible Interval	1.91	2.33	2.30
Percent Outside Credible Interval	32.43	21.86	1.59
Percent Below Observed	34.85	42.91	40.84
Percent Above Observed	62.88	53.61	55.13

Table 13: Predictive Uncertainty for ephemeral flow for Case Study 3.

7. Figures

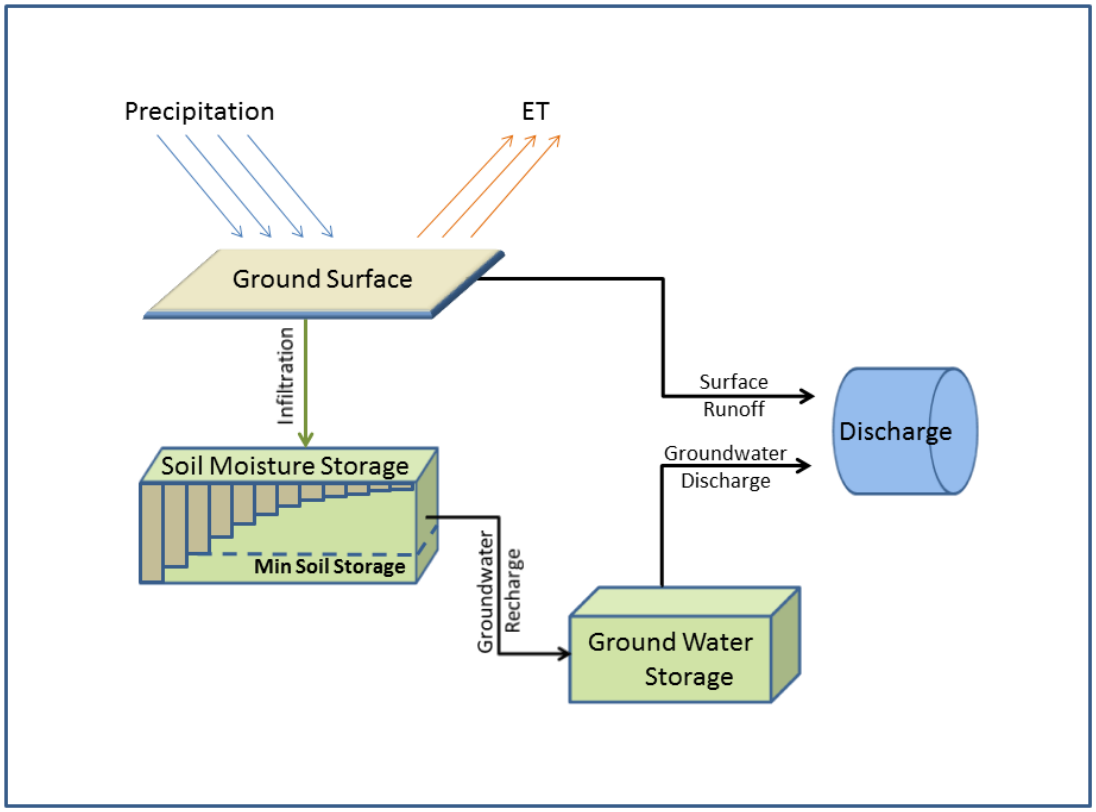


Figure 1: The implemented PDM Model. This model consists of precipitation inputs, evapotranspiration losses, a soil moisture store, groundwater store, and outputs through streamflow.



Figure 2: Geographic Locations for Davy River and Haughton River Gauges, Australia

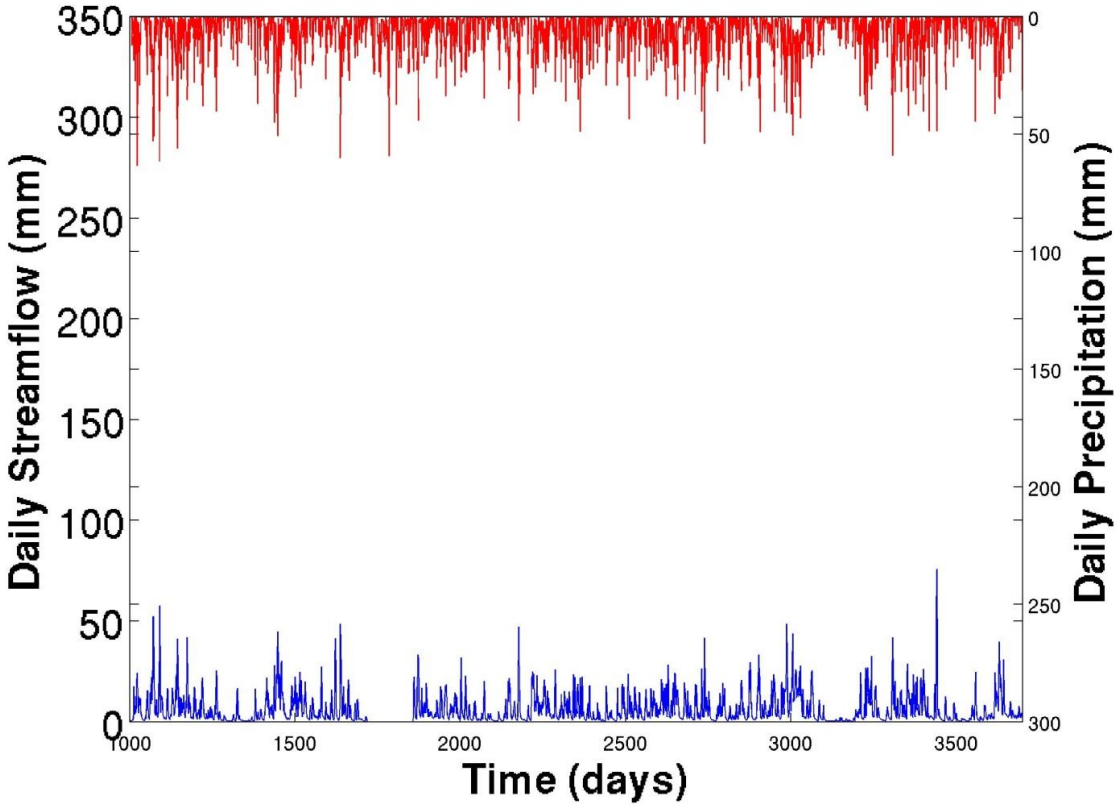


Figure 3: The hydrograph and hyetograph for the Davey River data set from June 12, 1972 to December 09, 1982.

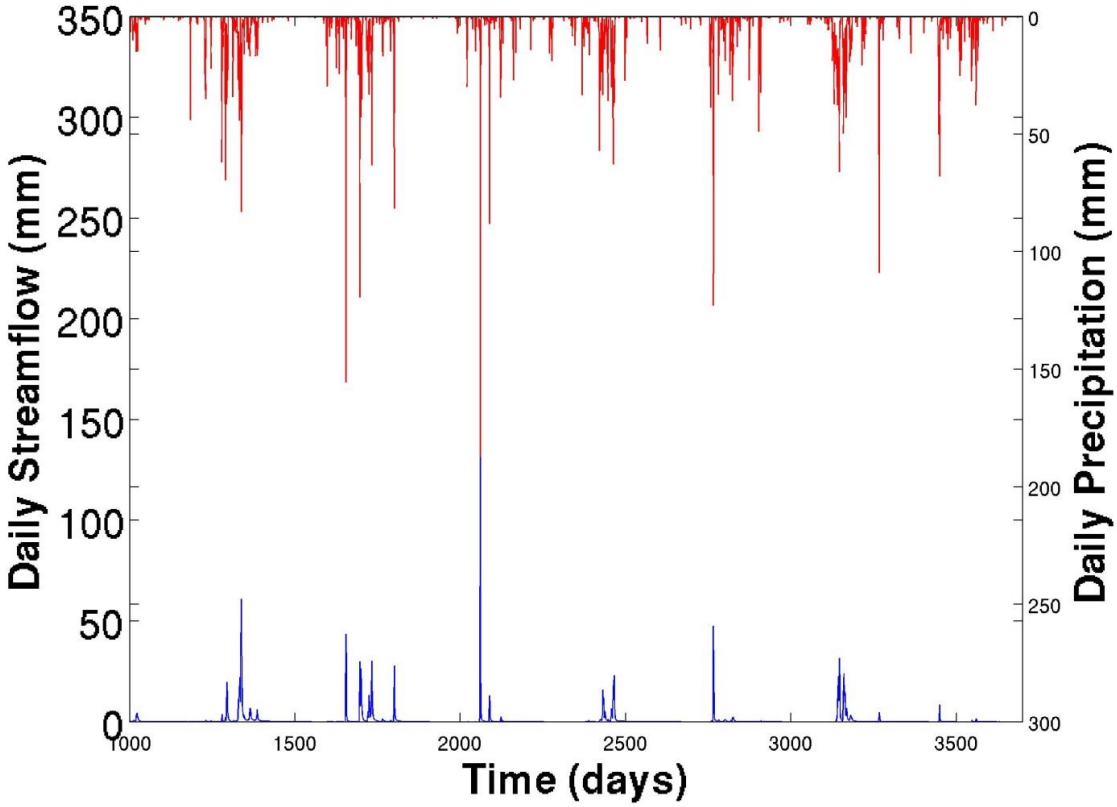


Figure 4: The hydrograph and hyetograph for the Haughton River data set from June 12, 1972 to December 09, 1982.

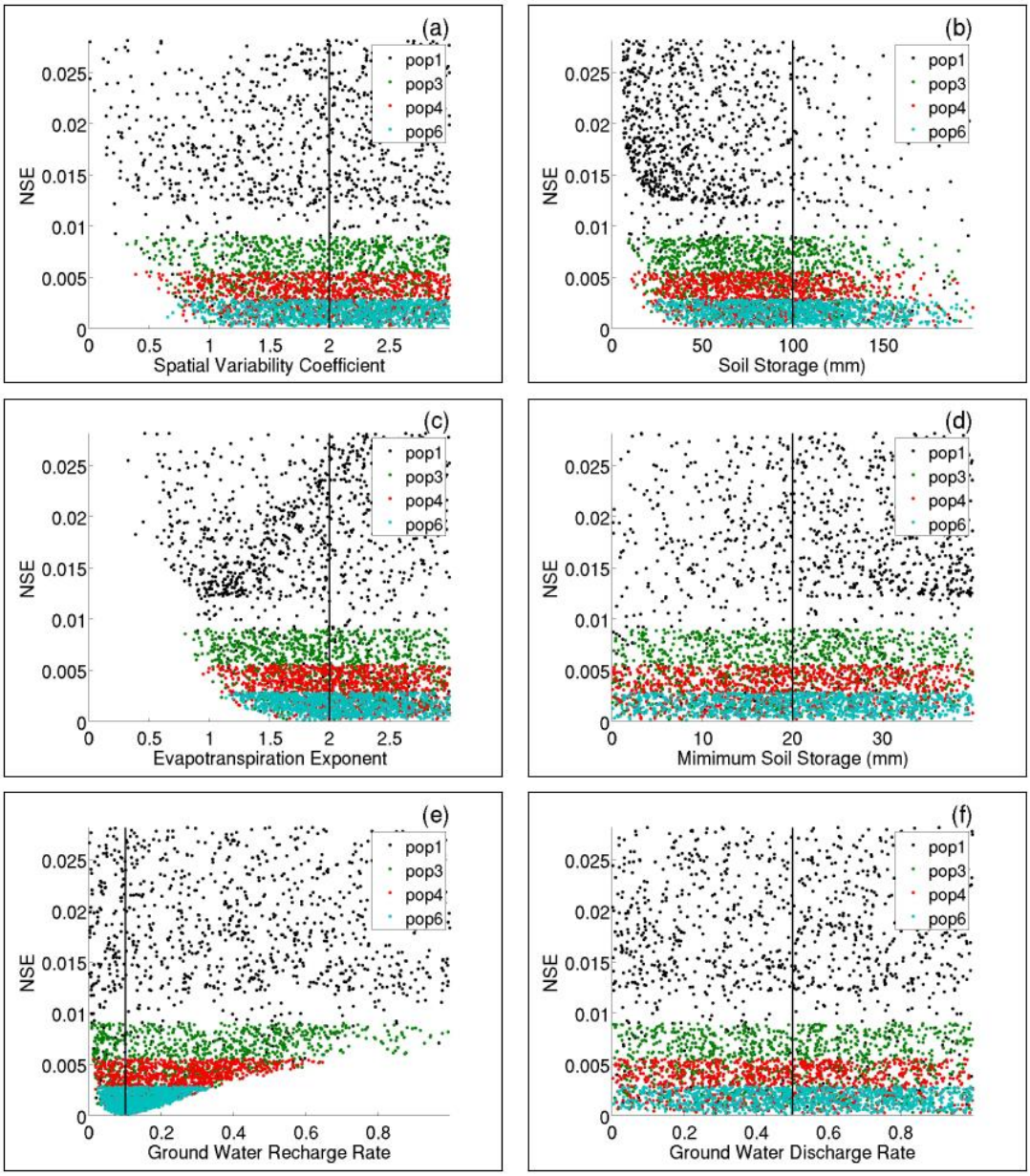


Figure 5: Dotty plots representing the distributions and convergence for each parameter of the PDM as calibrated for the Synthetic data set by the ABC-SMC algorithm using the hydrograph summary statistic. Sensitive parameters, such as GWU, demonstrate a strong convergence to the mode as the populations progress through the intermediate distributions. Less sensitive parameters exhibit a higher variability and more uniform distributions.

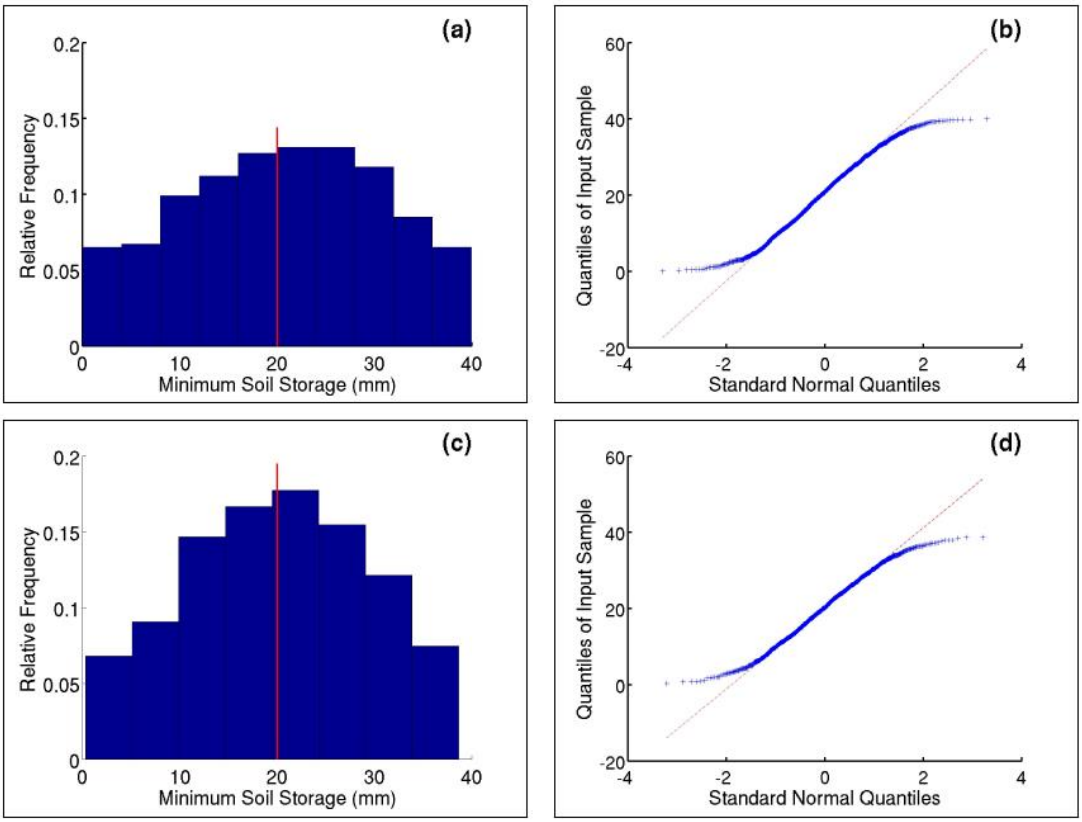


Figure 6: Histograms and QQ-Norm plots for the Minimum Soil Storage for the Synthetic data set calibrated using the hydrograph summary statistic. The Local Linear Regression adjustment (c-d) adjusts the marginal distribution of the raw parameters ((a-b) towards the true value as represented by the red line.

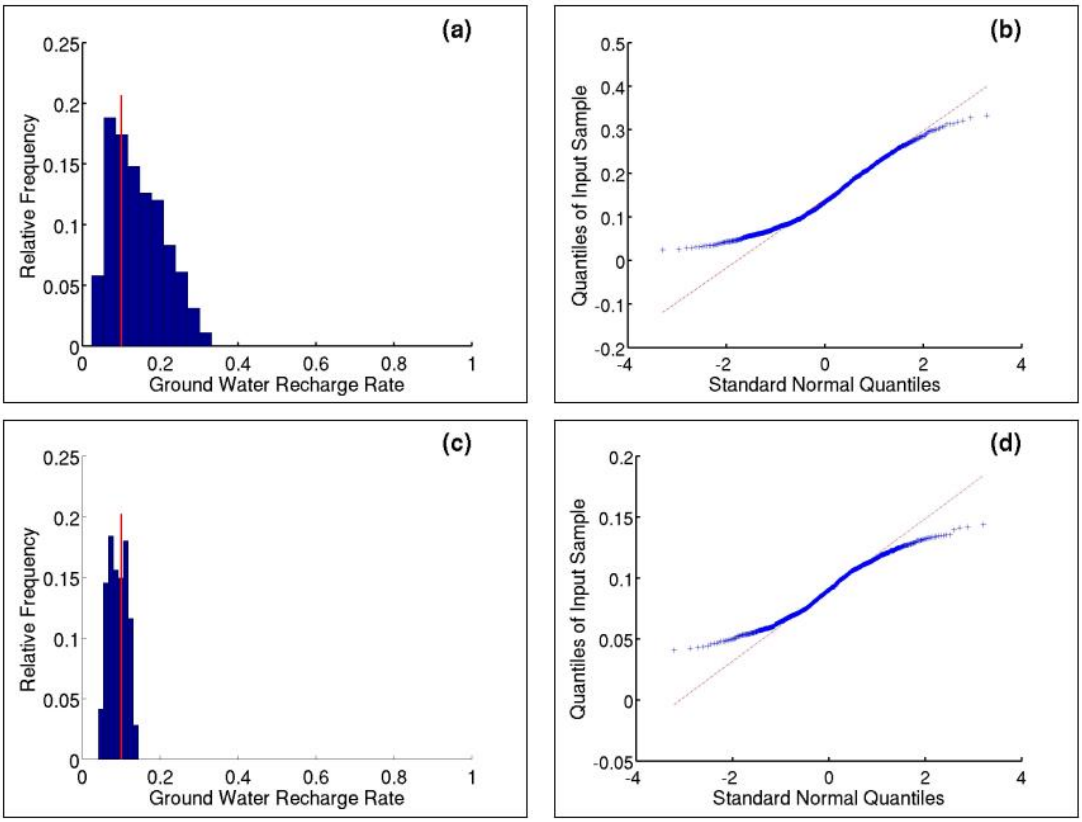


Figure 7: Histograms and QQ-Norm plots for the Groundwater Recharge Rate for the Synthetic data set calibrated using the hydrograph summary statistic. This is an example of the effect of the Local Linear Regression adjustment on a sensitive parameter. The bounds and the variance of the adjusted distribution (c-d) are narrower than the raw parameters (a-b).

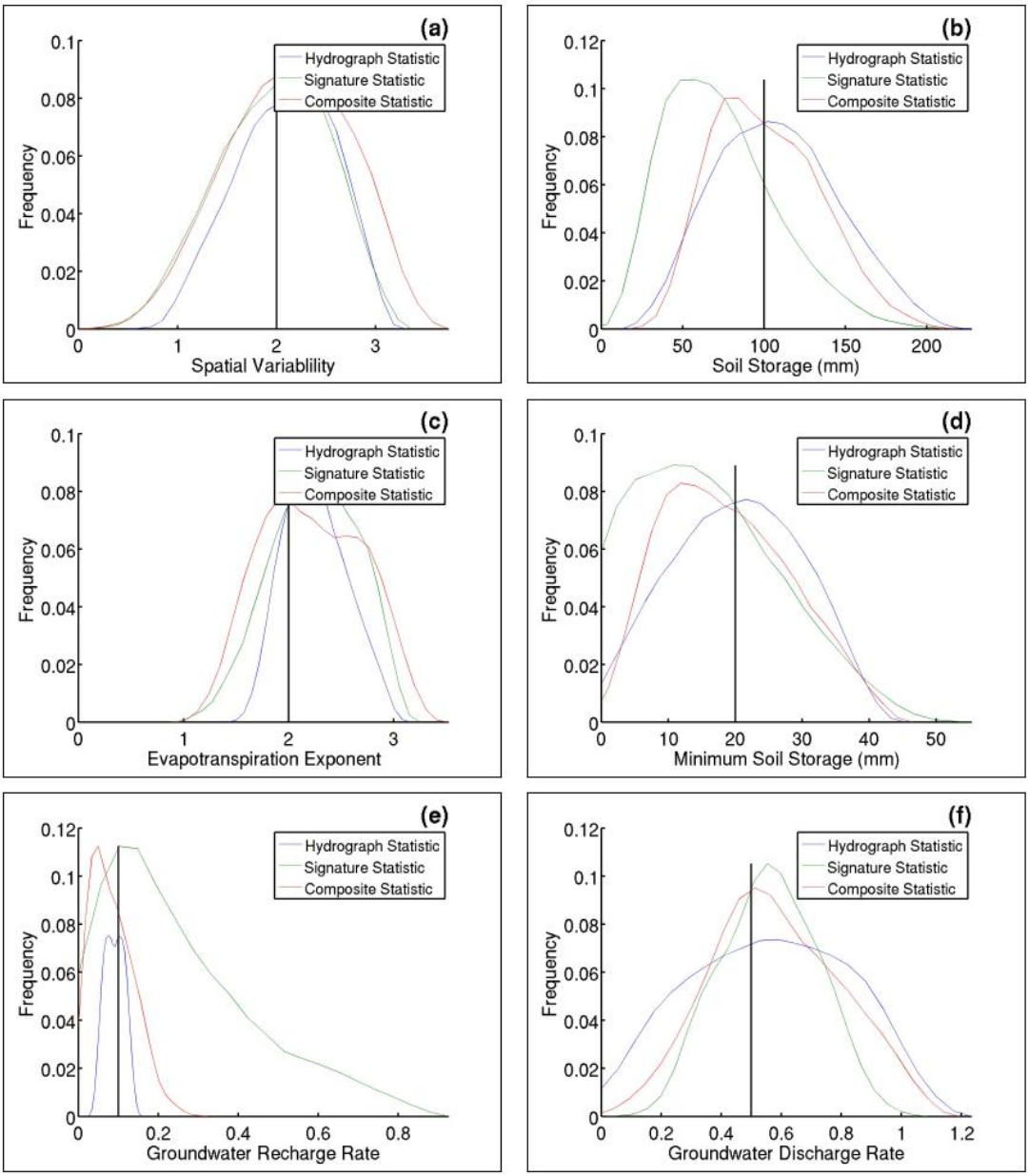


Figure 8: Parameter density plots for the Synthetic data set using the hydrograph statistic, signature statistic, and composite summary statistics. The plots suggest that the hydrograph summary statistic was able to have more reliable estimations of the true parameter value as indicated by the black line. Notable differences in the marginal distribution were the Maximum Soil Moisture Storage (b) and the Minimum Soil Moisture Storage (d). The parameter uncertainty for the signature statistic and composite statistic is higher than the hydrograph statistic.

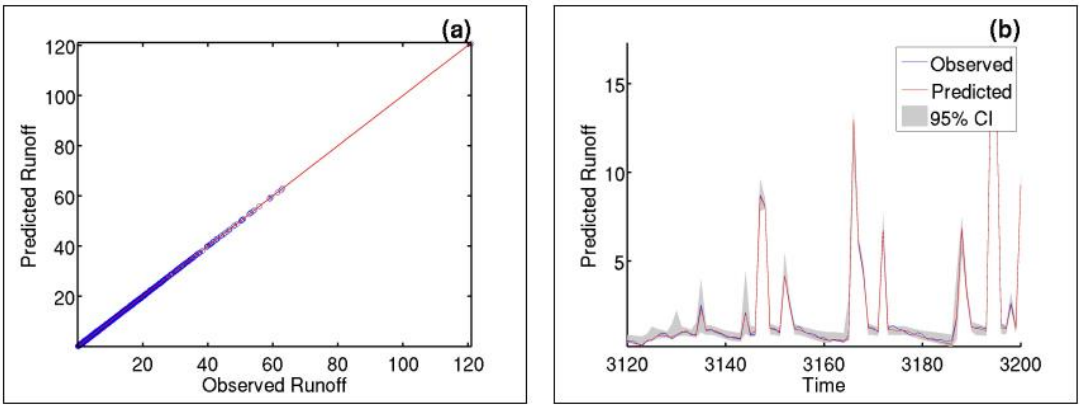


Figure 9: Predictive uncertainty plots for the Synthetic data set using the hydrograph summary statistic. The predicted versus observed plot (a) suggests a strong relationship. The hydrographs (b) suggest little predictive uncertainty with both the observed time series closely follows the predicted time series.

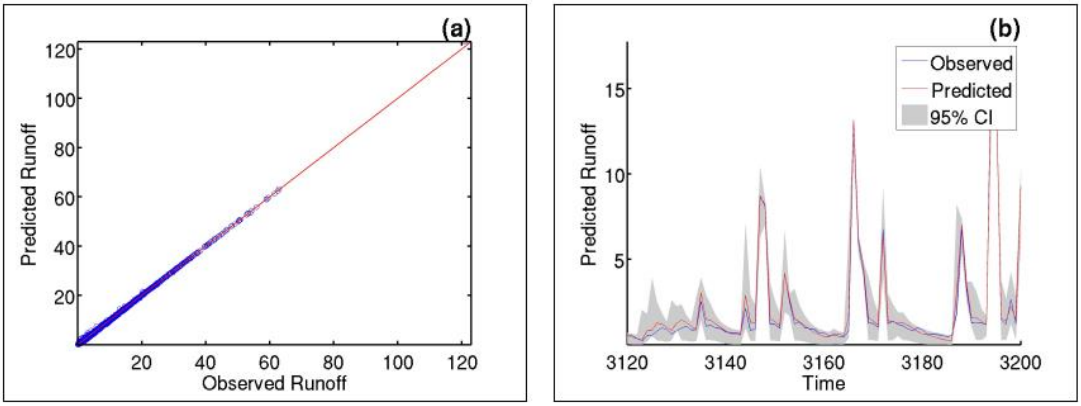


Figure 10: Predictive uncertainty plots for the Synthetic data set using the signature-based summary statistic. The predicted versus observed plot (a) suggests a strong relationship, though the lower streamflow values show more variability than the hydrograph statistic. The hydrographs (b) shows the observed time series follows the same predicted with noticeable deviations. The overall predictive uncertainty is higher than the hydrograph statistic estimates as shown by the deviations and larger 90% credible intervals.

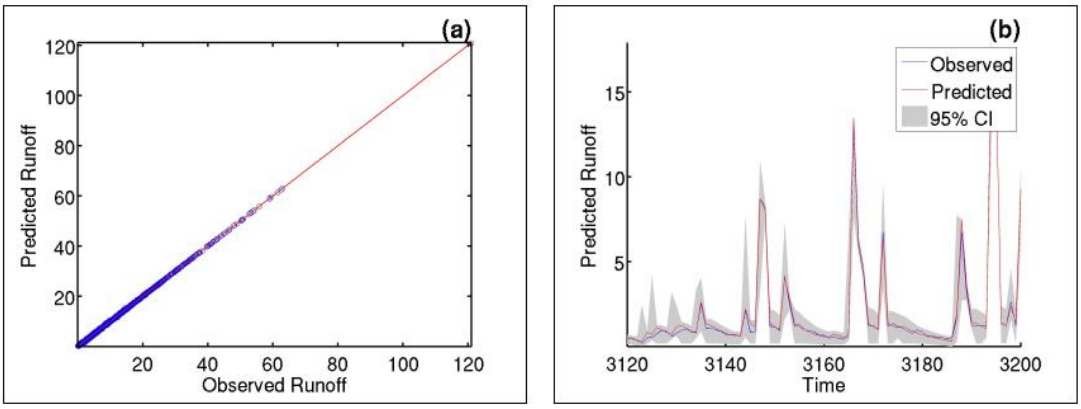


Figure 11: Predictive uncertainty plots for the Synthetic data set using the composite summary statistic. The predicted versus observed plot (a) suggests a strong relationship between the predicted and observed streamflow, but as with the signature statistic results, the hydrographs (b) indicate deviations from the predicted and observed time series. These plots suggest deviations in the results using the composite summary statistic lie between the predictive uncertainty of the hydrograph statistic and signature-based statistics.

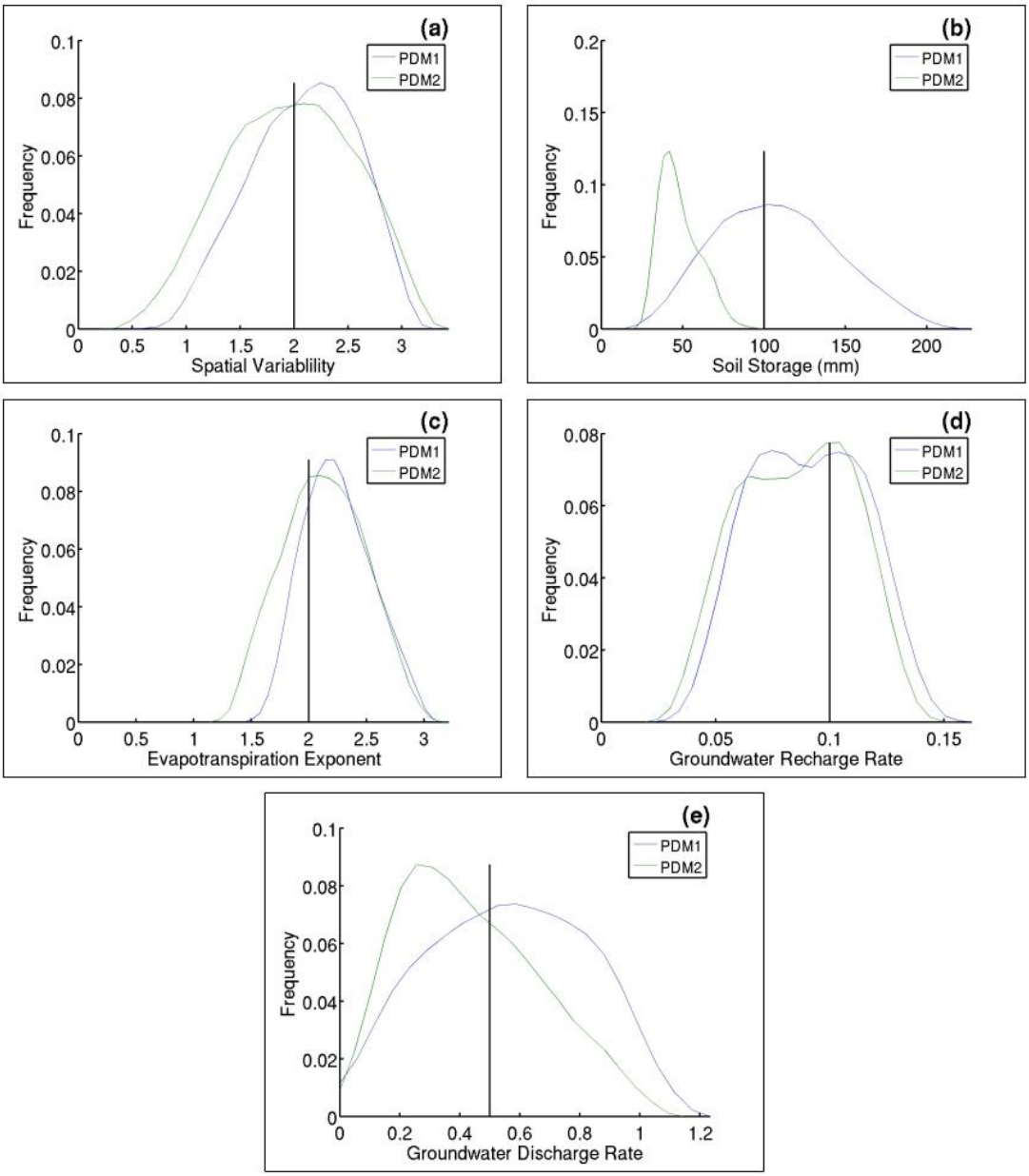


Figure 12: Parameter density plots for the Synthetic data set using the Full (PDM1) and Reduced (PDM2) models and the hydrograph summary statistic. Altering the model structure by removing the Minimum Soil Moisture Storage Parameter, increased parameter uncertainty is demonstrated for the Maximum Soils Moisture Storage and the Groundwater Discharge Rate.

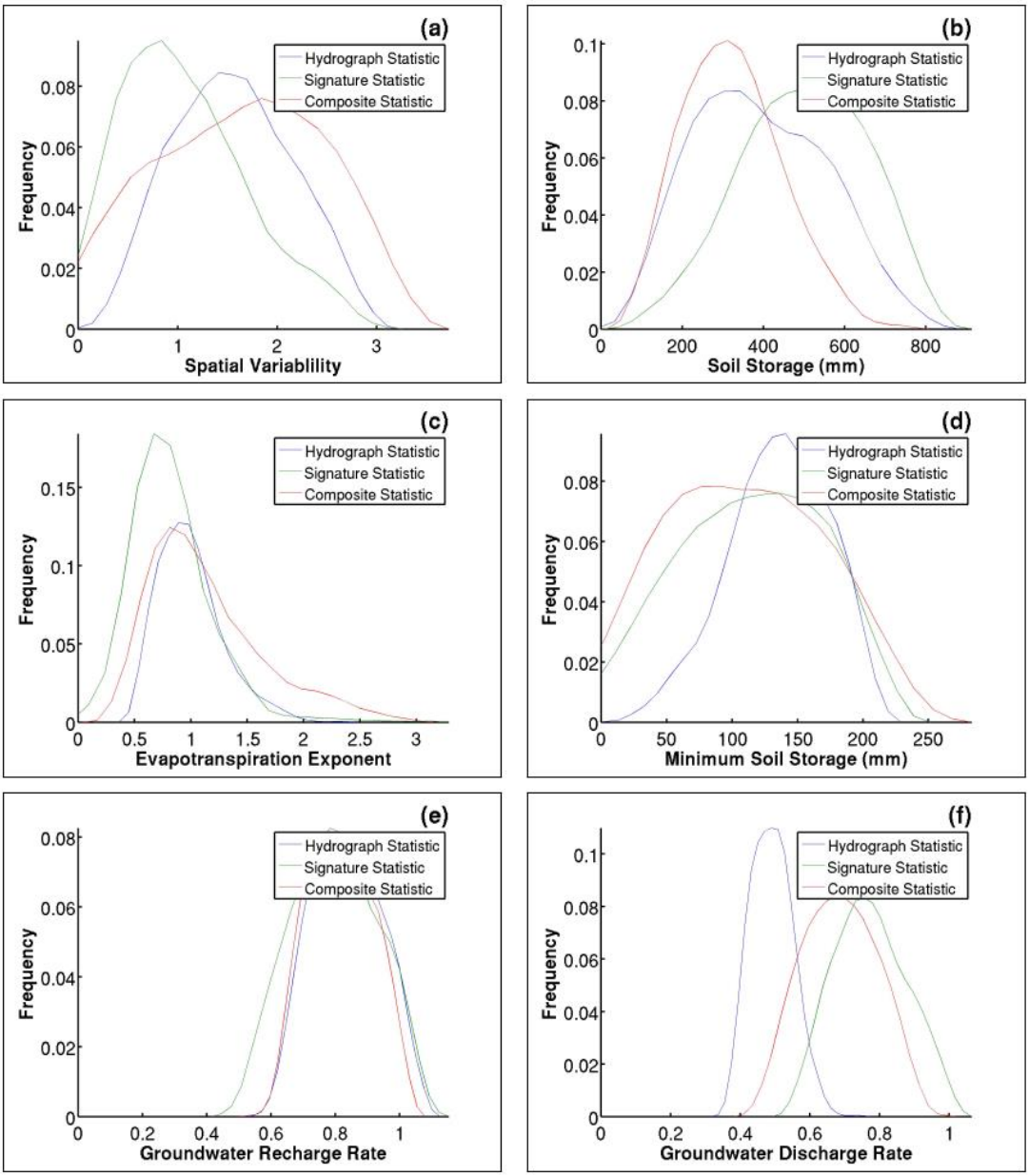


Figure 13: Parameter density plots for the Davey River data set. For all six parameters, the density plots for the results using the hydrograph statistic and composite statistics had similar modes and variances. The results using the signature-based summary statistic were similar for the evapotranspiration coefficient and the maximum soil moisture storage and differed for the other four parameters.

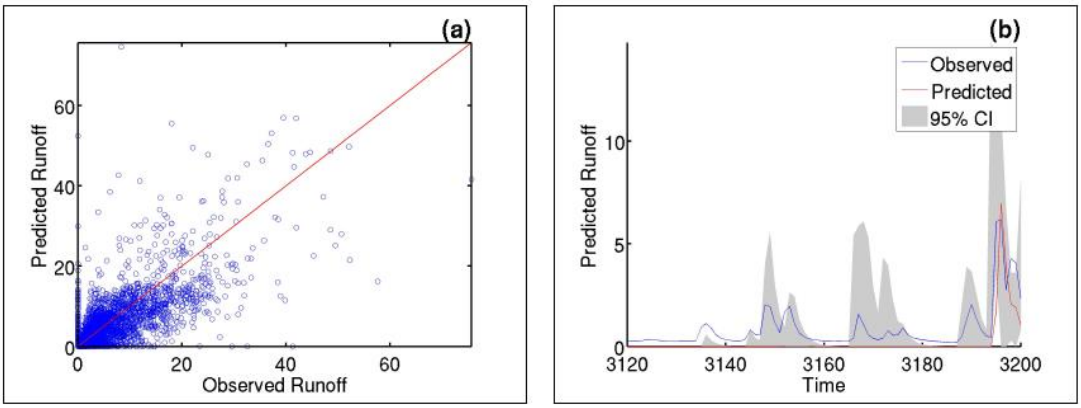


Figure 14: Predictive uncertainty plots for the Davey River data set using the hydrograph summary statistic. The observed versus predicted plot (a) indicates the implemented model tends to underestimate the observed values. This is also visible for the 180 days shown in the hydrographs (b).

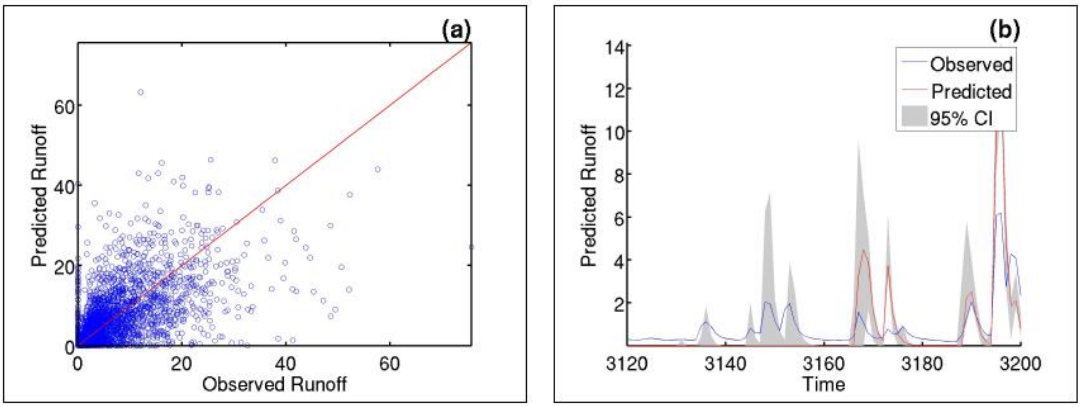


Figure 15: Predictive uncertainty plots for the Davey River data set using the signature-based summary statistic. The observed versus predicted plot (a) indicates the implemented model may underestimate the observed streamflow, but not as much as the results using the hydrograph statistic. This is also visible for the 180 days shown, where the events from day 3165 to 3200 are better represented.

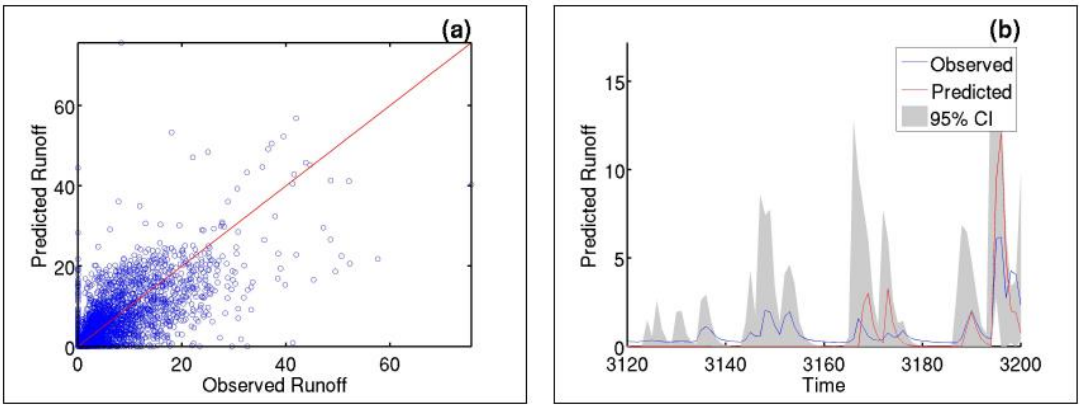


Figure 16: Predictive uncertainty plots for the Davey River data set using the composite summary statistic. The observed versus predicted plot (a) indicates the implemented model is comparable to the signature statistic results. The observed streamflow still shows underestimation, but the peaks are matched as well as the signature statistic results, but the magnitude of these peaks are closer to the observed streamflow.

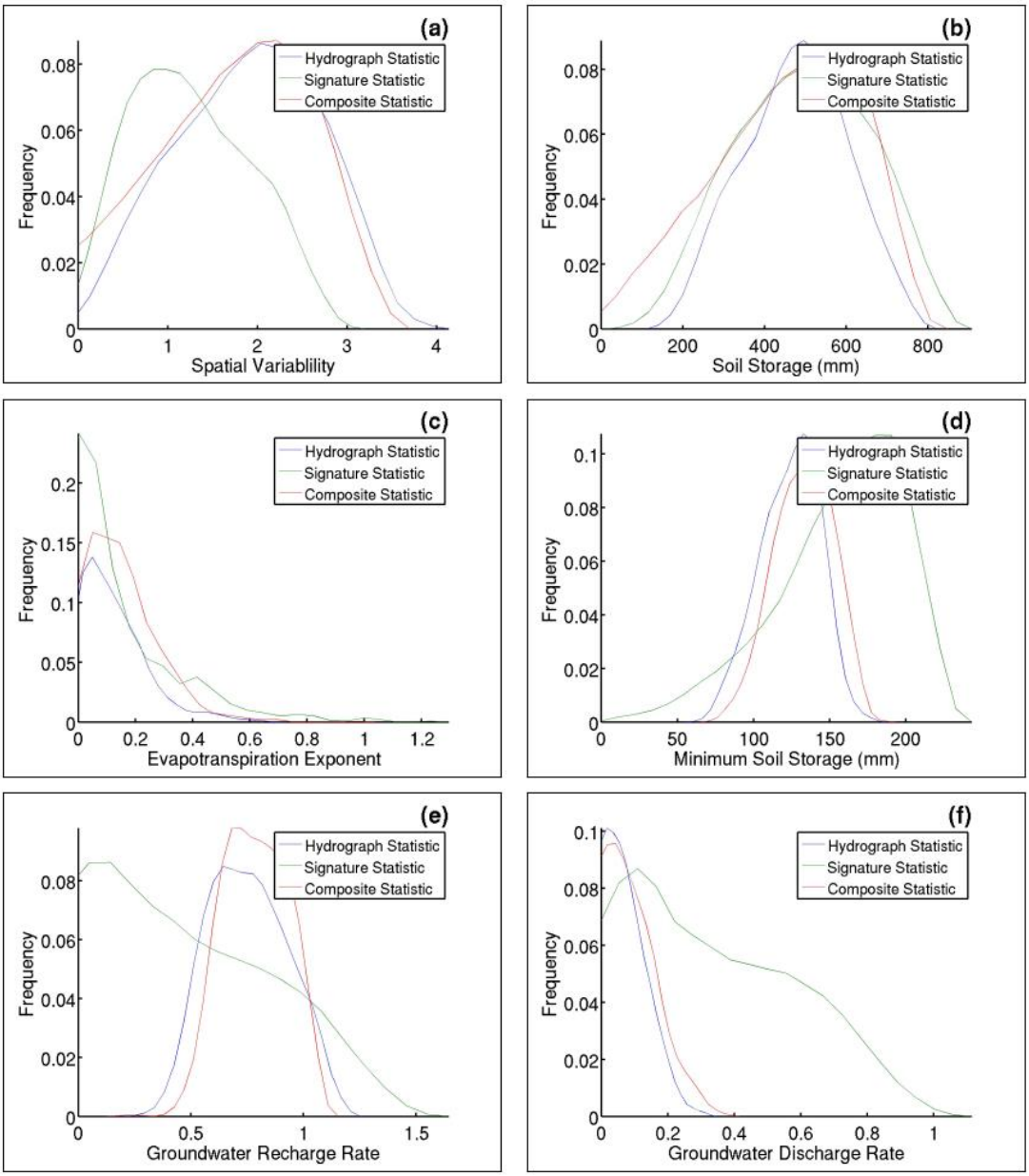


Figure 17: Kernel density plots for the estimated parameters from the Haughton River calibration processes. All six parameters show strong unimodal shapes for all three summary statistics. The value for the modes varied between the summary statistics non-systematically.

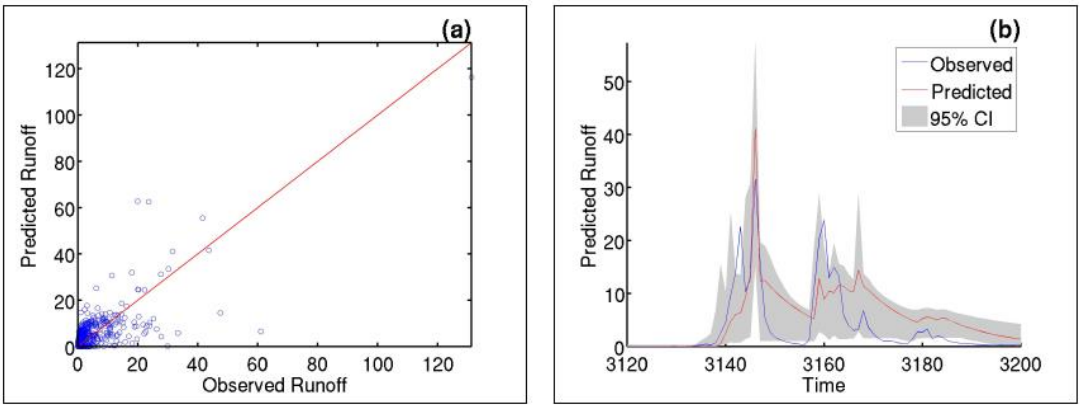


Figure 18: Predictive uncertainty plots for the Haughton River data set using the hydrograph summary statistic. The observed versus predicted plot (a) indicates the implemented model tends to overestimate the observed values. The predicted streamflow (b) suggests that the implemented model matches most of the event peaks, but overestimates the streamflow.

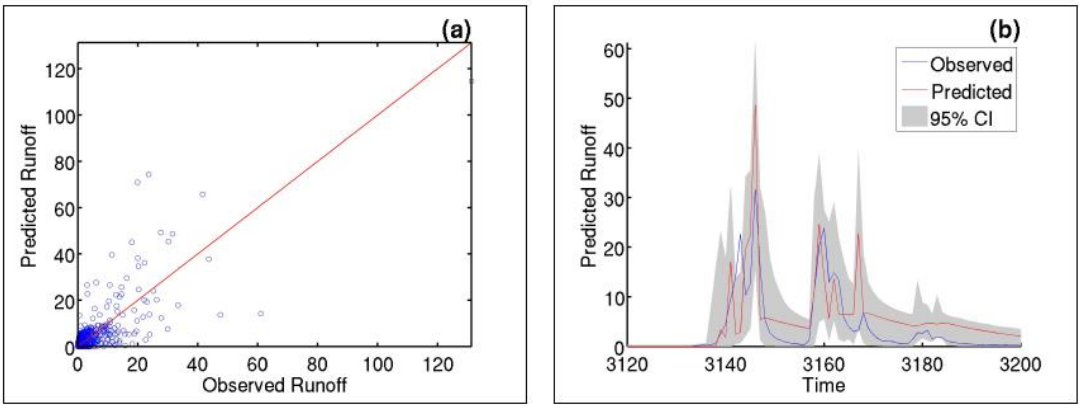


Figure 19: Predictive uncertainty plots for the Haughton RIver data set using the signature-based summary statistic. The observed versus predicted plot (a) does not have a clear indication that there is a systematic error in the observed values. The predicted streamflow (b) suggests that there is a pattern of overestimations, but the streamflow events are well represented. The magnitudes of the errors do not appear as high as the predictions when using the hydrograph statistic.

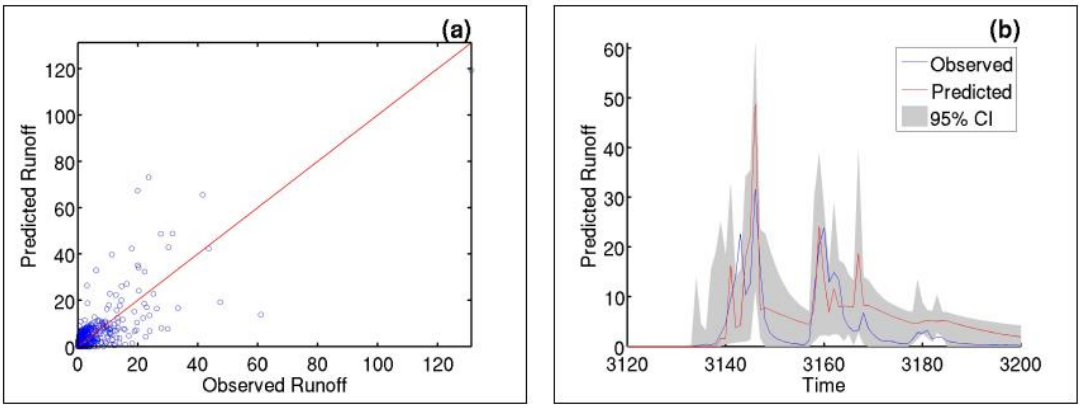


Figure 20: Predictive uncertainty plots for the Haughton River data set using the composite summary statistic. The observed versus predicted plot (a) suggests a tendency for underestimating the streamflow. The predicted streamflow (b) shows similar performance as with the signature statistic results, however, the uncertainty bounds are wider.

References

Beaumont, M.A. (2010) Approximate Bayesian Computation in Evolution and Ecology. Annual review of ecology, evolution, and systematics 41(1), 379-406.

Beaumont, M.A., Cornuet, J.-M., Marin, J.-M. and Robert, C.P. (2009) Adaptive approximate Bayesian computation. Biometrika 96(4), 983-990.

Beaumont, M.A., Zhang, W. and Balding, D.J. (2002) Approximate Bayesian Computation in Population Genetics. Genetics 162(4), 2025-2035.

Bell, V.A., Kay, A.L., Jones, R.G. and Moore, R.J. (2007) Use of a grid-based hydrological model and regional climate model outputs to assess changing flood risk. International Journal of Climatology 27(12), 1657-1671.

Beven, K. (2013) So how much of your error is epistemic? Lessons from Japan and Italy. Hydrological Processes 27(11), 1677-1680.

Beven, K. and Binley, A. (1992) The future of distributed models: Model calibration and uncertainty prediction. Hydrological Processes 6(3), 279-298.

Beven, K. and Freer, J. (2001) Equifinality, data assimilation, and uncertainty estimation in mechanistic modelling of complex environmental systems using the GLUE methodology. Journal of Hydrology 249(1–4), 11-29.

Blum, M., Nunes, M., Prangle, D. and Sisson, S. (2013) A comparative review of dimension reduction methods in approximate Bayesian computation. Statistical Science 28(2), 189-208.

Blum, M.B. and François, O. (2010) Non-linear regression models for Approximate Bayesian Computation. Statistics and Computing 20(1), 63-73.

Csilléry, K., François, O. and Blum, M.G.B. (2012) abc: an R package for approximate Bayesian computation (ABC). Methods in Ecology and Evolution 3(3), 475-479.

Diggle, P.J. and Gratton, R.J. (1984) Monte Carlo Methods of Inference for Implicit Statistical Models. Journal of the Royal Statistical Society. Series B (Methodological) 46(2), 193-227.

Freer, J., Beven, K. and Ambroise, B. (1996) Bayesian Estimation of Uncertainty in Runoff Prediction and the Value of Data: An Application of the GLUE Approach. Water Resources Research 32(7), 2161-2173.

Gupta, H.V. and Kling, H. (2011) On typical range, sensitivity, and normalization of Mean Squared Error and Nash-Sutcliffe Efficiency type metrics. Water Resources Research 47(10), W10601.

Gupta, H.V., Wagener, T. and Liu, Y. (2008) Reconciling theory with observations: elements of a diagnostic approach to model evaluation. Hydrological Processes 22(18), 3802-3813.

Jeremiah, E., Sisson, S.A., Sharma, A. and Marshall, L. (2012) Efficient hydrological model parameter optimization with Sequential Monte Carlo sampling. Environmental modelling & software : with environment data news 38, 283-295.

Joyce, P. and Marjoram, P. (2008) Approximately Sufficient Statistics and Bayesian Computation. Statistical Applications in Genetics & Molecular Biology 7(1), 1-18.

Kuczera, G. and Parent, E. (1998) Monte Carlo assessment of parameter uncertainty in conceptual catchment models: the Metropolis algorithm. Journal of Hydrology 211(1–4), 69-85.

Marin, J.-M., Pudlo, P., Robert, C. and Ryder, R. (2012) Approximate Bayesian computational methods. Statistics and Computing 22(6), 1167-1180.

Marjoram, P., Molitor, J., Plagnol, V. and Tavaré, S. (2003) Markov chain Monte Carlo without likelihoods. Proceedings of the National Academy of Sciences 100(26), 15324-15328.

Marshall, L., Nott, D. and Sharma, A. (2004) A comparative study of Markov chain Monte Carlo methods for conceptual rainfall-runoff modeling. Water Resources Research 40(2), W02501.

Matalas, N. (1997) Stochastic Hydrology in the Context of Climate Change. Climatic Change 37(1), 89-101.

McIntyre, N., Lee, H., Wheater, H., Young, A. and Wagener, T. (2005) Ensemble predictions of runoff in ungauged catchments. Water Resources Research 41(12), W12434.

Milly, P.C.D., Betancourt, J., Falkenmark, M., Hirsch, R.M., Kundzewicz, Z.W., Lettenmaier, D.P. and Stouffer, R.J. (2008) Stationarity Is Dead: Whither Water Management? Science 319(5863), 573-574.

Moore, R.J. (2007) The PDM rainfall-runoff model. Hydrology and Earth System Sciences 11(1), 483-499.

Moore, R.J. and Clarke, R.T. (1981) A Distribution Function Approach to Rainfall Runoff Modeling. Water Resources Research Vol 17, No 5, p 1367-1382, October, 1981. 11 Fig, 5 Tab, 25 Ref.

Nash, J.E. and Sutcliffe, J.V. (1970) River flow forecasting through conceptual models part I — A discussion of principles. Journal of Hydrology 10(3), 282-290.

Pappenberger, F. and Beven, K.J. (2006) Ignorance is bliss: Or seven reasons not to use uncertainty analysis. Water Resources Research 42(5), W05302.

Pritchard, J.K., Seielstad, M.T., Perez-Lezaun, A. and Feldman, M.W. (1999) Population growth of human Y chromosomes: a study of Y chromosome microsatellites. Molecular Biology and Evolution 16(12), 1791-1798.

Rubin, D.B. (1984) Bayesianly Justifiable and Relevant Frequency Calculations for the Applies Statistician. The Annals of Statistics 12(4), 1151-1172.

Sadegh, M. and Vrugt, J.A. (2013) Approximate Bayesian Computation in hydrologic modeling: equifinality of formal and informal approaches. Hydrology & Earth System Sciences Discussions 10(4), 4739-4797.

Schaefli, B. and Gupta, H.V. (2007) Do Nash values have value? Hydrological Processes 21(15), 2075-2080.

Sisson, S.A. (2009) Correction for Sisson et al., Sequential Monte Carlo without likelihoods. Proceedings of the National Academy of Sciences 106(39), 16889.

Sisson, S.A., Fan, Y. and Tanaka, M.M. (2007) Sequential Monte Carlo without likelihoods. Proceedings of the National Academy of Sciences 104(6), 1760-1765.

Smith, T.J. and Marshall, L.A. (2008) Bayesian methods in hydrologic modeling: A study of recent advancements in Markov chain Monte Carlo techniques. Water Resources Research 44(12), W00B05.

Tavare, S., Balding, D.J., Griffiths, R.C. and Donnelly, P. (1997) Inferring coalescence times from DNA sequence data. Genetics 145(2), 505-518.

Taylor, R.G., Scanlon, B., Doll, P., Rodell, M., van Beek, R., Wada, Y., Longuevergne, L., Leblanc, M., Famiglietti, J.S., Edmunds, M., Konikow, L., Green, T.R., Chen, J., Taniguchi, M., Bierkens, M.F.P., MacDonald, A., Fan, Y., Maxwell, R.M., Yechieli, Y., Gurdak, J.J., Allen, D.M., Shamsudduha, M., Hiscock, K., Yeh, P.J.F., Holman, I. and Treidel, H. (2013) Ground water and climate change. Nature Clim. Change 3(4), 322-329.

Toni, T. and Stumpf, M.P.H. (2010) Simulation-based model selection for dynamical systems in systems and population biology. Bioinformatics 26(1), 104-110.

Toni, T., Welch, D., Strelkowa, N., Ipsen, A. and Stumpf, M.P.H. (2009) Approximate Bayesian computation scheme for parameter inference and model selection in dynamical systems. Journal of the Royal Society Interface 6(31), 187-202.

Vrugt, J.A. and Braak, C.J.F.t. (2011) DREAM(D): an adaptive Markov Chain Monte Carlo simulation algorithm to solve discrete, noncontinuous, and combinatorial posterior parameter estimation problems. Hydrology and Earth System Sciences 15(12), 3701-3713.

Vrugt, J.A. and Sadegh, M. (2013) Toward diagnostic model calibration and evaluation: Approximate Bayesian computation. Water Resources Research 49(7), 4335-4345.

Vrugt, J.A., ter Braak, C.J.F., Diks, C.G.H. and Schoups, G. (2013) Hydrologic data assimilation using particle Markov chain Monte Carlo simulation: Theory, concepts and applications. Advances in Water Resources 51(0), 457-478.

Westerberg, I.K., Guerrero, J., Younger, P.M. and Beven, K.J. (2011) Calibration of hydrological models using flow-duration curves. Hydrology and Earth System Sciences.

Zhang, Y., Liu, H. and Houseworth, J. (2011) Modified Generalized Likelihood Uncertainty Estimation (GLUE) Methodology for Considering the Subjectivity of Likelihood Measure Selection. Journal of Hydrologic Engineering 16(6), 558-561.

CHAPTER FOUR

CONCLUSION

Hydrological models are one of the most useful tools for understanding and managing watersheds. In the last two decades, the rapid increase in available computing power has led to the adoption and dependence on hydrologic models for managing water resources. However, the methodology allowing for the effective estimation of the uncertainty present within the models and the resulting predictions is still evolving.

Within hydrology, the two most commonly used methods for calibrating models are GLUE and MCMC. While MCMC implements Bayesian inference and is considered to have more statistically rigor than GLUE (Montanari 2005, Stedinger et al. 2008), GLUE is still popular due to the ease of usage when compared to MCMC methods.

Chapter 1 introduced ABC methods to the hydrological community. The purpose of this chapter was to present the theoretical links between ABC and GLUE. In fact, the sampling process of the basic ABC Rejection sampler is similar to the GLUE algorithm. The benefits to adopting the ABC framework are the increased statistical rigor, improved weighting, flexible summary statistics, and the ability to work within a multivariate environment. This improves the estimates of the predictive uncertainty associated with the final proposed model allowing proper uncertainty assessments to be easier to conduct and more useful to the consumers of the model predictions.

Chapter 3 expanded upon the basic ABC algorithm and introduced three features of the ABC methods that are designed to improve the performance and uncertainty estimations for the hydrological modeler. When the ABC algorithm is implemented using the SMC sampler, the efficiency of the calibration process improves. If individual calibration processes are more efficient, a modeler has the flexibility to run additional and/or more complex models using the same computational resources.

The second feature that may improve the assessing the parameter uncertainty is the use of post-processing techniques to improve the posterior distribution for the parameters. Unlike GLUE, these adjustments can be performed on the joint distribution which accounts for potential issues of covariance and heteroscedasticity. While the local linear regression results showed marginal benefits, the marginal distributions of the parameters were improved.

The third feature that was presented in Chapter 3 was the application of customized summary statistics. Three categories of summary statistics were used, statistical based, signature-based, and composite summary statistics. The results indicated that with small amounts of uncertainty in the data and model structure, the statistical measure performed well. However, the inclusion of signature information improved the predictive uncertainty for the real world data sets used in Case 2 and Case 3. In these cases, the composite statistic composite summary statistic had the fewest predictions that fell outside the 95% credible interval, demonstrating the potential benefits for innovative summary statistics tailored to the characteristics of a specific watershed.

As presented in Chapters 2 and 3, ABC-SMC provides an efficient method to calibrate hydrological models while maintaining the statistical rigor attributed with Bayesian inference. ABC methods can implemented as easily as GLUE and provide realistic estimates of the predictive uncertainty that are associated with MCMC. With the addition of post-processing techniques and innovative summary statistic, ABC-SMC methods provide a viable set of tools to improve the hydrological modeling process and uncertainty assessment.

References

Montanari, A. (2005) Large sample behaviors of the generalized likelihood uncertainty estimation (GLUE) in assessing the uncertainty of rainfall-runoff simulations. Water Resources Research 41(8), n-a-n/a.

Stedinger, J.R., Jery, R.S., Richard, M.V., Seung Uk, L. and Rebecca, B. (2008) Appraisal of the generalized likelihood uncertainty estimation (GLUE) method. Water Resources Research 44(12).

REFERENCES CITED

Abbott, M.B., Bathurst, J.C., Cunge, J.A., O'Connell, P.E. and Rasmussen, J. (1986) An introduction to the European Hydrological System — Systeme Hydrologique Europeen, “SHE”, 1: History and philosophy of a physically-based, distributed modelling system. Journal of Hydrology 87(1–2), 45-59.

Bates, B.C. and Campbell, E.P. (2001) A Markov Chain Monte Carlo Scheme for parameter estimation and inference in conceptual rainfall-runoff modeling. Water Resources Research 37(4), 937-947.

Beaumont, M.A. (2010) Approximate Bayesian Computation in Evolution and Ecology. Annual review of ecology, evolution, and systematics 41(1), 379-406.

Beaumont, M.A., Cornuet, J.-M., Marin, J.-M. and Robert, C.P. (2009) Adaptive approximate Bayesian computation. Biometrika 96(4), 983-990.

Beaumont, M.A., Zhang, W. and Balding, D.J. (2002) Approximate Bayesian Computation in Population Genetics. Genetics 162(4), 2025-2035.

Bell, V.A., Kay, A.L., Jones, R.G. and Moore, R.J. (2007) Use of a grid-based hydrological model and regional climate model outputs to assess changing flood risk. International Journal of Climatology 27(12), 1657-1671.

Beven, K. (2013) So how much of your error is epistemic? Lessons from Japan and Italy. Hydrological Processes 27(11), 1677-1680.

Beven, K. and Binley, A. (1992) The future of distributed models: Model calibration and uncertainty prediction. Hydrological Processes 6(3), 279-298.

Beven, K. and Freer, J. (2001) Equifinality, data assimilation, and uncertainty estimation in mechanistic modelling of complex environmental systems using the GLUE methodology. Journal of Hydrology 249(1–4), 11-29.

Beven, K., Smith, P.J. and Wood, A. (2011) On the colour and spin of epistemic error (and what we might do about it). Hydrology & Earth System Sciences 15(10), 3123-3133.

Beven, K.J. (2011) Rainfall-Runoff Modelling: The Primer, Wiley.

Beven, K.J., Smith, P.J. and Freer, J.E. (2008) So just why would a modeller choose to be incoherent? Journal of Hydrology 354(1–4), 15-32.

Blum, M., Nunes, M., Prangle, D. and Sisson, S. (2013) A comparative review of dimension reduction methods in approximate Bayesian computation. Statistical Science 28(2), 189-208.

Blum, M.B. and François, O. (2010) Non-linear regression models for Approximate Bayesian Computation. Statistics and Computing 20(1), 63-73.

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

Csilléry, K., François, O. and Blum, M.G.B. (2012) abc: an R package for approximate Bayesian computation (ABC). Methods in Ecology and Evolution 3(3), 475-479.

Diggle, P.J. and Gratton, R.J. (1984) Monte Carlo Methods of Inference for Implicit Statistical Models. Journal of the Royal Statistical Society. Series B (Methodological) 46(2), 193-227.

Freer, J., Beven, K. and Ambroise, B. (1996) Bayesian Estimation of Uncertainty in Runoff Prediction and the Value of Data: An Application of the GLUE Approach. Water Resources Research 32(7), 2161-2173.

Gupta, H.V. and Kling, H. (2011) On typical range, sensitivity, and normalization of Mean Squared Error and Nash-Sutcliffe Efficiency type metrics. Water Resources Research 47(10), W10601.

Gupta, H.V., Wagener, T. and Liu, Y. (2008) Reconciling theory with observations: elements of a diagnostic approach to model evaluation. Hydrological Processes 22(18), 3802-3813.

Jeremiah, E., Sisson, S.A., Sharma, A. and Marshall, L. (2012) Efficient hydrological model parameter optimization with Sequential Monte Carlo sampling. Environmental modelling & software : with environment data news 38, 283-295.

Jin, X., Xu, C.-Y., Zhang, Q. and Singh, V.P. (2010) Parameter and modeling uncertainty simulated by GLUE and a formal Bayesian method for a conceptual hydrological model. Journal of Hydrology 383(3–4), 147-155.

Joyce, P. and Marjoram, P. (2008) Approximately Sufficient Statistics and Bayesian Computation. Statistical Applications in Genetics & Molecular Biology 7(1), 1-18.

Kitanidis, P.K. and Bras, R.L. (1980) Real-time forecasting with a conceptual hydrologic model: 1. Analysis of uncertainty. Water Resources Research 16(6), 1025-1033.

Kuczera, G. and Parent, E. (1998) Monte Carlo assessment of parameter uncertainty in conceptual catchment models: the Metropolis algorithm. Journal of Hydrology 211(1–4), 69-85.

Liu, Y. and Gupta, H.V. (2007) Uncertainty in hydrologic modeling: Toward an integrated data assimilation framework. Water Resources Research 43(7), W07401.

Mantovan, P. and Todini, E. (2006) Hydrological forecasting uncertainty assessment: Incoherence of the GLUE methodology. Journal of Hydrology 330(1–2), 368-381.

Marin, J.-M., Pudlo, P., Robert, C. and Ryder, R. (2012) Approximate Bayesian computational methods. Statistics and Computing 22(6), 1167-1180.

Marjoram, P., Molitor, J., Plagnol, V. and Tavaré, S. (2003) Markov chain Monte Carlo without likelihoods. Proceedings of the National Academy of Sciences 100(26), 15324-15328.

Marshall, L., Nott, D. and Sharma, A. (2004) A comparative study of Markov chain Monte Carlo methods for conceptual rainfall-runoff modeling. Water Resources Research 40(2), W02501.

Matalas, N. (1997) Stochastic Hydrology in the Context of Climate Change. Climatic Change 37(1), 89-101.

McIntyre, N., Lee, H., Wheater, H., Young, A. and Wagener, T. (2005) Ensemble predictions of runoff in ungauged catchments. Water Resources Research 41(12), W12434.

Milly, P.C.D., Betancourt, J., Falkenmark, M., Hirsch, R.M., Kundzewicz, Z.W., Lettenmaier, D.P. and Stouffer, R.J. (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 Resources Research 41(8), n-a-n/a.

Moore, I.D., O'Loughlin, E.M. and Burch, G.J. (1988) A contour-based topographic model for hydrological and ecological applications. Earth Surface Processes and Landforms 13(4), 305-320.

Moore, R.J. (2007) The PDM rainfall-runoff model. Hydrology and Earth System Sciences 11(1), 483-499.

Moore, R.J. and Clarke, R.T. (1981) A Distribution Function Approach to Rainfall Runoff Modeling. Water Resources Research Vol 17, No 5, p 1367-1382, October, 1981. 11 Fig, 5 Tab, 25 Ref.

Nash, J.E. and Sutcliffe, J.V. (1970) River flow forecasting through conceptual models part I — A discussion of principles. Journal of Hydrology 10(3), 282-290.

Pappenberger, F. and Beven, K.J. (2006) Ignorance is bliss: Or seven reasons not to use uncertainty analysis. Water Resources Research 42(5), W05302.

Pritchard, J.K., Seielstad, M.T., Perez-Lezaun, A. and Feldman, M.W. (1999) Population growth of human Y chromosomes: a study of Y chromosome microsatellites. Molecular Biology and Evolution 16(12), 1791-1798.

Rubin, D.B. (1984) Bayesianly Justifiable and Relevant Frequency Calculations for the Applies Statistician. The Annals of Statistics 12(4), 1151-1172.

Sadegh, M. and Vrugt, J.A. (2013) Approximate Bayesian Computation in hydrologic modeling: equifinality of formal and informal approaches. Hydrology & Earth System Sciences Discussions 10(4), 4739-4797.

Schaefli, B. and Gupta, H.V. (2007) Do Nash values have value? Hydrological Processes 21(15), 2075-2080.

Sisson, S.A. (2009) Correction for Sisson et al., Sequential Monte Carlo without likelihoods. Proceedings of the National Academy of Sciences 106(39), 16889.

Sisson, S.A., Fan, Y. and Tanaka, M.M. (2007) Sequential Monte Carlo without likelihoods. Proceedings of the National Academy of Sciences 104(6), 1760-1765.

Smith, T.J. and Marshall, L.A. (2008) Bayesian methods in hydrologic modeling: A study of recent advancements in Markov chain Monte Carlo techniques. Water Resources Research 44(12), W00B05.

Stedinger, J.R., Jery, R.S., Richard, M.V., Seung Uk, L. and Rebecca, B. (2008) Appraisal of the generalized likelihood uncertainty estimation (GLUE) method. Water Resources Research 44(12).

Tavare, S., Balding, D.J., Griffiths, R.C. and Donnelly, P. (1997) Inferring coalescence times from DNA sequence data. Genetics 145(2), 505-518.

Taylor, R.G., Scanlon, B., Doll, P., Rodell, M., van Beek, R., Wada, Y., Longuevergne, L., Leblanc, M., Famiglietti, J.S., Edmunds, M., Konikow, L., Green, T.R., Chen, J., Taniguchi, M., Bierkens, M.F.P., MacDonald, A., Fan, Y., Maxwell, R.M., Yechieli, Y., Gurdak, J.J., Allen, D.M., Shamsudduha, M., Hiscock, K., Yeh, P.J.F., Holman, I. and Treidel, H. (2013) Ground water and climate change. Nature Clim. Change 3(4), 322-329.

ter Braak, C.F. and Vrugt, J. (2008) Differential Evolution Markov Chain with snooker updater and fewer chains. Statistics and Computing 18(4), 435-446.

Todini, E. (2007) Hydrological catchment modelling: past, present and future. Hydrol. Earth Syst. Sci. 11(1), 468-482.

Toni, T. and Stumpf, M.P.H. (2010) Simulation-based model selection for dynamical systems in systems and population biology. Bioinformatics 26(1), 104-110.

Toni, T., Welch, D., Strelkowa, N., Ipsen, A. and Stumpf, M.P.H. (2009) Approximate Bayesian computation scheme for parameter inference and model selection in dynamical systems. Journal of the Royal Society Interface 6(31), 187-202.

Vrugt, J.A. and Braak, C.J.F.t. (2011) DREAM(D): an adaptive Markov Chain Monte Carlo simulation algorithm to solve discrete, noncontinuous, and combinatorial posterior parameter estimation problems. Hydrology and Earth System Sciences 15(12), 3701-3713.

Vrugt, J.A. and Sadegh, M. (2013) Toward diagnostic model calibration and evaluation: Approximate Bayesian computation. Water Resources Research 49(7), 4335-4345.

Vrugt, J.A., ter Braak, C.J.F., Diks, C.G.H. and Schoups, G. (2013) Hydrologic data assimilation using particle Markov chain Monte Carlo simulation: Theory, concepts and applications. Advances in Water Resources 51(0), 457-478.

Westerberg, I.K., Guerrero, J., Younger, P.M. and Beven, K.J. (2011) Calibration of hydrological models using flow-duration curves. Hydrology and Earth System Sciences.

Zhang, Y., Liu, H. and Houseworth, J. (2011) Modified Generalized Likelihood Uncertainty Estimation (GLUE) Methodology for Considering the Subjectivity of Likelihood Measure Selection. Journal of Hydrologic Engineering 16(6), 558-561.