18th World IMACS / MODSIM Congress, Cairns, Australia 13-17 July 2009
http://mssanz.org.au/modsim09

A conceptual precipitation-runoff modeling suite: Model selection, calibration and predictive uncertainty assessment

Smith, T.J. [1] **and L.A. Marshall** [1]

[1] *Department of Land Resources and Environmental Sciences, Montana State University, Bozeman, Montana, USA*
Email: tylersmith@montana.edu

Abstract: Developing an understanding of the dominant controls on runoff processes at the catchment scale is vital to the effectiveness of model predictions. This is especially true for watersheds that receive precipitation predominantly as snowfall. In these regions, the water responsible for generating a majority of the streamflow is stored in the snowpack and released during a rapid melt period. Recognizing the uncertainty that exists in the process of selecting one model structure over another, a conceptual precipitation-runoff modeling framework featuring a modular construction has been developed. The Simulation and Prediction Lab for Analysis of Snowmelt Hydrology (SPLASH), a companion graphical user interface, was developed to encompass all of the components of the modeling framework. SPLASH incorporates the tools necessary for implementing the framework, including model calibration, simulation, and uncertainty analysis, while also providing the tools necessary to perform model analysis including generating data files of model output and plotting tools for uncertainty analysis.

The modular construction of the modeling approach allowed for the inclusion of a suite of competing component structures; namely, a soil moisture accounting and runoff routing component, a snowmelt accounting component, and a semi-distribution component (for the precipitation inputs). A total of thirty modular structures were investigated, representing the total number of possible combinations of the included components. The framework was applied to a test location, the Stringer Creek watershed, located in the Tenderfoot Creek Experimental Forest (TCEF) of Central Montana, USA. Stringer Creek watershed is heavily influenced by snowfall, accounting for more than 75% of the total yearly precipitation. The use of minimally parameterized models is motivated by the desire to capture first-order processes, in line with a top-down modeling philosophy. Such models have the capability to be more efficient in modeling the system by having less uncertainty with similar predictive power when compared to more complex model structures. Because of the conceptual nature of the framework, some form of model calibration was necessary to obtain appropriate parameter approximations. In this case, a Bayesian inferential approach was adopted and implemented via the Delayed Rejection Adaptive Metropolis (DRAM) algorithm, a Markov chain Monte Carlo (MCMC) approach. The application of a Bayesian approach provides a statistical means of estimating uncertainty associated with the parameters, while also performing parameter estimation.

This research focused on developing a greater understanding of the effect of model composition through identifying the benefit of each component on the model prediction. The modular structures were calibrated over a four year period and tested during a separate two year period. The benefits and limitations of this conceptual catchment modeling framework have been addressed through a focus on the three elements of model performance (fit to streamflow), uncertainty, and 'realism' (fit to snow water equivalents). Further consideration was also focused on the trade-off between model structural complexity and predictive performance. The application of the modeling framework indicated model performance improved with the incorporation of a probability distributed soil moisture accounting approach. Additionally, the importance of incorporating net radiation into the snowmelt accounting method greatly enhanced overall model performance (simulation of stream discharge), as well as the simulation of the internal dynamics of snow water equivalents. Beyond enabling the identification of model components that reproduce stream discharge well, the modeling framework also assisted in identifying specific traits of model components that led to the favorable simulations they generated.

Keywords: *model calibration, Bayesian statistics, uncertainty assessment, model selection, Markov chain Monte Carlo, snowmelt*

1. INTRODUCTION

Hydrologic modeling has taken on many forms in past studies, covering the entire spectrum from physically-based to conceptual to black-box approaches. The bulk of studies focusing on the ability of a model to represent a physical system fail to consider a range of competing, alternate hypotheses aimed at addressing the modeling question (Neuman, 2003). While the task of estimating uncertainty in the model and its parameter estimates has become increasingly studied (e.g., Bates and Campbell, 2001; Beven and Binley, 1992), Neuman (2003, p. 292) points out that "[t]he bias and uncertainty that result from reliance on an inadequate conceptual-mathematical model are typically much larger than those introduced through an inadequate choice of model parameter values."

Despite the abundance of research that has been focused on the development of models to describe the physical system, much is still unknown about the primary drivers of watershed processes. Because of the large uncertainties surrounding current watershed models and their appropriate application, this paper introduces a modeling framework consisting of a collection of flexible, modular structures with variable complexities (number of model dimensions) derived from a suite of previously established conceptual precipitation-runoff models. The motivation of this design is threefold: (1) to analyze how a variety of different structures specifically relate to the prediction of a streamflow hydrograph, (2) to perform a comparative analysis concentrating on characterizing the relationship between structural complexity and predictive power, and (3) to synthesize the framework into a freely distributable, standalone software package. The following sections introduce the study location, the models (or model components) selected for inclusion in the study, the methods used to carry out the analysis, and the pertinent conclusions drawn from the research.

2. SITE DESCRIPTION

Recognizing the need to select a site that offers broad applicability to the modeling approach laid out in the research and that is also representative of the conditions typical to mountain watersheds in the Rocky Mountain West, a location in Central Montana, USA has been selected. Positioned at the headwaters of Tenderfoot Creek in the Little Belt Mountains of the Lewis and Clark National Forest, Tenderfoot Creek Experimental Forest (TCEF, 46^{o}55' N, 110^{o}52' W) is approximately 115 kilometers southeast of Great Falls, Montana (Figure 1). The selection of the Stinger Creek watershed (555 hectares) for this study was motivated by a climatic pattern typical of regional mountain locations and the availability of historical long-term records. Specific data for the Stringer Creek watershed that were assembled for this study include meteorological data from the Onion Park SNOTEL (Figure 1), streamflow data from the gauging station at the outlet of Stringer Creek (Figure 1), and topographic data in the form of a 10 meter grid digital elevation model. Time series data were selected for the period from March 2002 to September 2007.

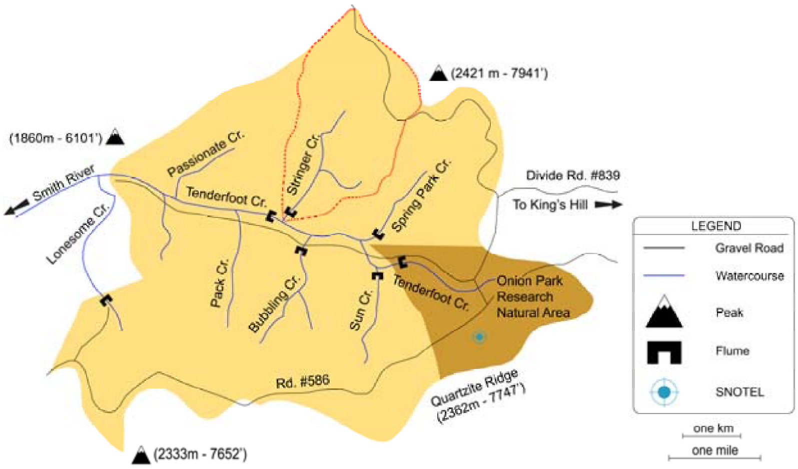

Figure 1. Tenderfoot Creek Experimental Forest, MT, with important features shown.

3. MODELING FRAMEWORK

A suite of conceptual, modular structures was developed and evaluated on a 12 hour time step for this study. The structures were selected ranging in structural representation of the physical system and in structural complexity (i.e., the number of model dimensions), acknowledging the complexity of hydrologic systems and the ambiguity associated with the understanding of the driving processes. Compounding this uncertainty is the need to characterize snowmelt inputs to the watershed system for areas dominated by snowfall. In consideration of the Stringer Creek watershed, several main processes could clearly be seen as dominant

(Jencso et al., 2009), necessitating their inclusion in the overall conceptual model as model components (or modules).

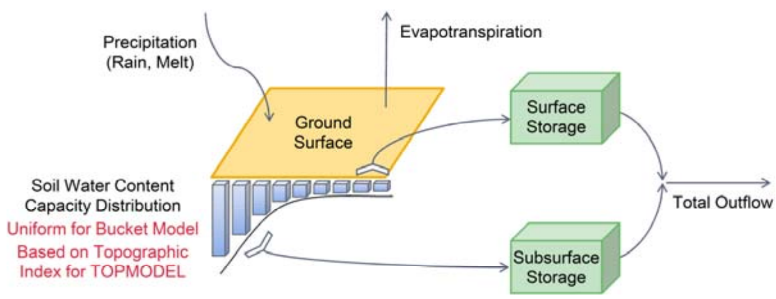

Figure 2. Basic model configuration, where the snowmelt accounting procedure (not shown) determines the amount of melt water input to the soil moisture accounting module.

From the identified dominant processes, model structures were constructed to include a soil moisture accounting and runoff routing module, a snowmelt accounting module, and a semi-distribution module. Under this framework thirty unique modular structures were developed comprising all possible combinations of modules. The details of each module will be discussed in the following subsections.

3.1. Soil Moisture Accounting and Runoff Routing

Representing the water accounting processes of the physical system, three competing soil moisture accounting and runoff routing modules were selected. The three approaches chosen for this study were: (1) a simple bucket model (BUC, e.g., Manabe, 1969), (2) the Probability Distributed Model (PDM, Moore, 2007), and (3) the TOPMODEL (TOP, Beven, 1997). Given that each is a conceptualization of reality, developed under a top-down modeling approach, the modules attempt to characterize the dominant process (specifically, the partitioning of the soil water content capacity distribution) while lumping all others into effective parameters. Figure 2 illustrates the basic functioning of the three alternate soil moisture accounting modules, where water is routed from the conceptual storages to the stream via time-delay release parameters.

3.2. Snowmelt Accounting

Given the snowfall-dominated precipitation of the Stringer Creek watershed, a snowmelt/rainfall accounting module was implemented to appropriately characterize the inputs to the system. Continuing the use of downward model development strategies, three conceptual snowmelt accounting techniques were selected for use within the modeling suite: (1) the temperature index approach, (2) the radiation index approach, and (3) the combined temperature and radiation index approach.

Although significant research has gone into the physical understanding of snowpack metamorphosis and the development of full energy balance methods to simulate snowmelt (e.g., Anderson, 1968), the constraints posed by such methods are extensive (e.g., Rango and Martinec, 1995). The use of conceptual, index-based approaches to snowmelt accounting is well established in the literature (e.g., Blöschl and Kirnbauer, 1991), with Ohmura (2001) providing a physical basis for such approaches.

3.3. Semi-Distribution of Inputs

The implementation of a spatial distribution scheme associated with inputs to the watershed has the ability to improve the physical representation of system processes (Garen and Marks, 2005), as well as the predictive result of the model simulation. In an effort to account for the (potential) spatial variability present in the precipitation inputs to the system, four different conceptual methods were considered: (1) no semi-distribution, (2) semi-distribution by aspect, (3) semi-distribution by elevation, and (4) semi-distribution by aspect and elevation. Each semi-distribution method uses a 10 m digital elevation model of the study watershed. For each method, the semi-distributed results are re-aggregated using a weighted average approach after the snowmelt accounting procedures are called and the weighted average value is then propagated through the soil moisture accounting and runoff routing module.

4. RESULTS

4.1. Calibration and Uncertainty Analysis Approach

The importance of model calibration, uncertainty analysis, and assessment in hydrologic modeling necessitates a comprehensive approach to the precipitation-runoff problem. Many different methodologies exist and have been applied to the calibration of hydrologic models, such as manual approaches including visual goodness-of-fit tests (e.g., Madsen et al., 2002) or fully automatic techniques including global optimization (e.g., Duan et al., 1992), Monte Carlo simulation (e.g., Beven and Binley, 1992), or Markov chain Monte Carlo (MCMC) simulation (e.g., Bates and Campbell, 2001).

The implementation of a Bayesian statistical approach offers the ability to perform both parameter estimation and uncertainty analysis, while also serving as an attractive framework within which multiple model structures can be compared (Marshall et al., 2005). Despite the obvious benefits of such an approach, Bayesian inference has seen limited use in hydrology until recently due to the complications associated with calculating the posterior distribution for highly complex, non-linear models. In hydrology, Bayesian methods usually rely on Markov chain Monte Carlo simulation for the characterization of a probability distribution for the model parameters and outputs, known as the posterior distribution. While many MCMC algorithms have been developed in an effort to describe the posterior distribution, the Delayed Rejection Adaptive Metropolis (DRAM, Haario et al., 2006) algorithm has been selected for calibration and uncertainty estimation for the conceptual modeling suite presented in Section 3. Previous research has shown the DRAM algorithm to be an efficient means of characterizing complex parameter spaces relative to other recently developed approaches (Smith and Marshall, 2008).

4.2. Assessment Strategy

The assessment of model suitability can take many identities; however, studies often focus solely on direct simulation performance. A more thorough approach of considering at least model performance, uncertainty, and realism (i.e., the ability of the model to simulate quantities internal to calibration) is advocated by Wagener (2003). By implementing this type of multi-faceted approach to evaluate models, a more comprehensive understanding of reasons behind model successes or failures can be extracted. Each of the modular structures contained in the modeling suite is analyzed in the following subsections with respect to each of these aspects directly, as well as with respect to structural complexity.

4.3. Performance Analysis

Evaluation of model performance is traditionally based on the results of model predictions relative to the observed data (see Klemeš, 1986). In this study, the data has been split into two sets, with one being used to calibrate model parameters and the other being used for an 'independent' assessment of model performance. The calibration phase covered approximately four years, from March 2002 through September 2005 and the 'validation' period covered two years, from October 2005 through September 2007; this sample split ensured the data in each period were representative of past typical conditions in Stringer Creek.

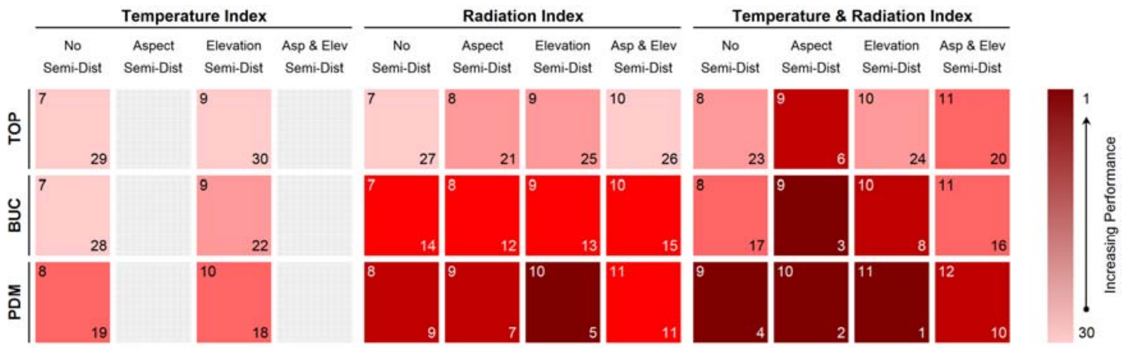

Figure 3. Summary of calibration results for stream discharge, where models were ranked based on the maximum log-likelihood obtained for each modular structure during calibration. Model dimension is given in the upper left corner and model rank is given in the lower right corner of each cell.

A summary of the calibration results for each of the thirty modular structures contained within the framework is given in Figure 3, where the calibration performances have been binned into six equal frequency classes as summarized by the maximum calibrated log-likelihood value. The number of parameters associated with each structure and the model rank are also given in Figure 3. The patterns present in Figure 3 begin to reveal the interactions taking place among the three different modules and those occurring within a single module. In general, improvement occurred when moving down (toward the PDM module) and to the right (toward radiation-based snowmelt modules) in Figure 3. The deficiencies in TOPMODEL and the temperature index modules are clear. The semi-distribution module appears less critical to simulation success, as 'good' predictions were observed across all semi-distribution methods. Assessment of performance that considered complexity, quantified by Akaike's information criterion (Akaike, 1974) resulted in the same ranking of models as was found with 'raw' performance measures.

The modeling framework was further broken into 31 comparison units, covering each possible module combination produced by holding two modules constant and varying the third module. This allowed for the isolation of the effect of the varying component to the overall performance of a structure. As was indicated in Figure 3, the temperature index approach to snowmelt accounting was consistently the least effective of the three alternatives. Focusing on linking this module's inefficiency to the hydrograph, Figure 4 (Inset) reveals that the primary area of difference among the three snowmelt accounting methods was in the rising limb of the hydrograph. The temperature index method routinely resulted in a much 'flashier' response during this segment of the hydrograph (where streamflow responds quickly to fast melting accumulated snow), especially in combination with the no semi-distribution component. When semi-distribution by elevation was considered, this 'flashy' signal was somewhat muted but still evident.

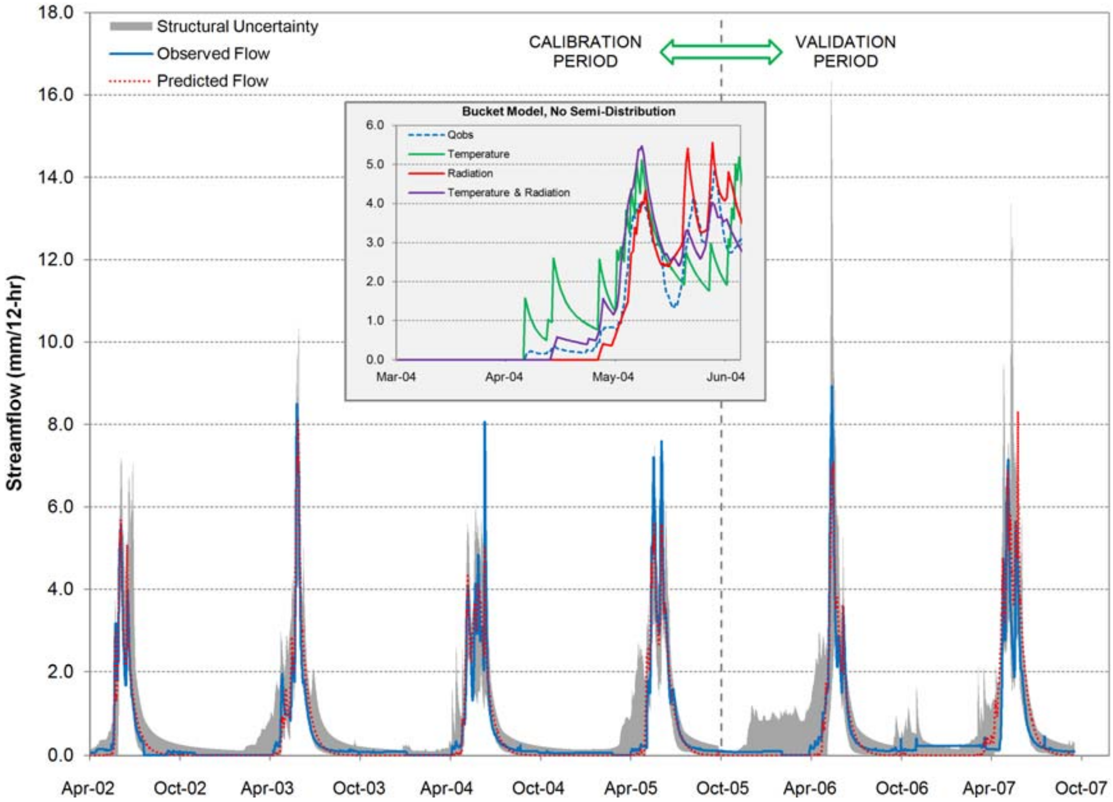

Figure 4. Structural uncertainty and streamflow simulation for 'optimal' structure. Simulations shown for both the calibration and validation periods. **Inset** shows the 'flashy' behavior of the temperature index approach, when other components are constant.

Focusing on the validation period, trends similar to those found in the calibration period exist with regard to favored components (e.g., temperature and radiation index approach to snowmelt accounting). Although the patterns present across the modeling framework can be seen in both the calibration and validation periods, the overall performances during the validation period deteriorate and the interactions are generally 'looser' among the different components.

4.4. Uncertainty Analysis

The benefit of implementing a Bayesian approach is in its ability to characterize the parameter uncertainty via parameter posterior distributions. These distributions have the potential to offer insights into the behaviors of each parameter. For example, it may be found that a certain parameter is in essence uniformly distributed and may be unnecessary to the structure's success. Identification of characteristic distributions (e.g., uniform, flat, peaked, bimodal, thresholds, etc.) is an important tool in model checking and model improvement. The posterior distributions also aid in the calculation of statistically valid confidence and prediction intervals. Such intervals can then be considered to represent the uncertainty in the parameters, input data, and individual model structure.

The structural uncertainty present in the entire modeling framework is brought about by alternative representations of the key watershed processes of soil moisture accounting and runoff routing, snowmelt accounting, and semi-distribution. By pooling the model simulations of each modular structure (at every time step), the overall structural uncertainty of the modeling framework can be investigated. Figure 4 illustrates the structural uncertainty across both the calibration and validation periods, as well as the prediction of the modular structure that had the best performance during calibration (*PDMsdE_TRn*). The uncertainty in the validation period tends to be greater than during the calibration period, which is consistent with expectations. However, model simulations were reasonably consistent in pattern across the two analysis periods.

4.5. Realism Analysis

The assessment of precipitation-runoff models based on one or more internal system processes not used in model calibration is valuable in relation to a model's ultimate usefulness in a forecasting mode. Previous studies have found that precipitation and streamflow data lack the information content required to prove a structure to be a realistic description of the watershed (e.g., Jakeman and Hornberger, 1993). Accepting the limitations of the forcing data (precipitation, streamflow), snow water equivalents data were used as a first step toward checking the framework's realism. This data represents an objective outlook on the ability of each modular structure to simulate a vital watershed process, specifically the development and melting of a snowpack. Although useful in a broad sense, this data on snowpack is invariably biased due to the extrapolation of single site measurements to the greater system. An assessment of model realism indicated that the same models favored during the calibration period for discharge are apt to be favored in their representation of the internal snowmelt dynamics of the Stringer Creek watershed.

4.6. Simulation and Prediction Lab for Analysis of Snowmelt Hydrology (SPLASH)

The enduring characteristic of many successful applied research studies is a quality of usability beyond its research scope. In the modeling of precipitation-runoff processes in particular, there has been a trend toward computer-intensive methods that often rely strongly on the application of complex statistical algorithms (e.g., Bates and Campbell, 2001). While numerous studies praise the merits of such methods (e.g., Marshall et al., 2004), the audience is left with the task of researching, programming, and implementing all of the interrelated components associated with such advantageous methodologies.

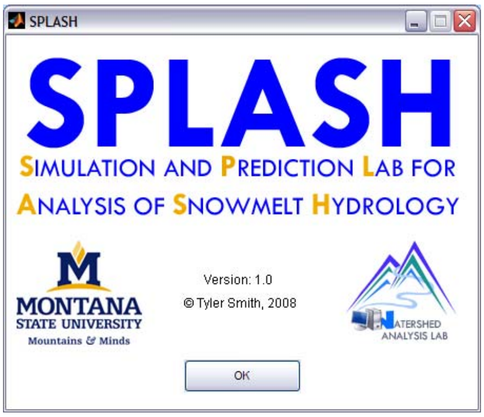

Figure 5. SPLASH modeling software.

The true utility of many emergent approaches to hydrologic modeling is in their application. However, traditional, well-understood, and easily available approaches will logically be the first choice in most operational settings due to the difficulty associated with implementing the cutting-edge (Mantovan and Todini, 2006). In response to this challenge, the Simulation and Prediction Lab for Uncertainty Analysis of Snowmelt Hydrology (SPLASH, Figure 5) was developed to allow the user to easily carry out the tasks of model selection, calibration, evaluation, and uncertainty analysis under a Bayesian framework. SPLASH was built as a complete modeling tool, with the capability to perform the varied tasks required by conceptual modeling under uncertainty.

5. CONCLUSIONS

The overall assessment of hydrologic models is of crucial importance, and the consideration of multiple alternate representations of the physical system should be a fundamental component of any such assessment. Because watershed-scale systems are extremely complex, variable, and uncertain failure to consider multiple model structures will undoubtedly lead to errors. The present study illustrated the benefits of developing a modeling framework based on a top-down assessment of the dominant processes of the snow-dominated Stringer Creek watershed. Assessment focused primarily on the performance of individual modules and their direct impact on hydrograph prediction, the accounting of structural uncertainties of the framework itself, and the consistency of models to represent the internal dynamic of snow accumulation and melt.

While the results of this study do not conclusively establish the best single model for mountainous, snow-dominated watersheds, they do offer insight into the interplay among the modular components and predictive performance for our case study. The overriding conclusions found were: (1) precipitation semi-distribution methods did little more than provide another degree of freedom to the model structures, (2) net radiation is an important component in conceptualization of snowmelt accounting methods, and (3) the TOPMODEL component's rigidity brought on by the inclusion of topographic index classes may inhibit the overall calibration performance but may also provide a beneficial amount of physical realism in a forecasting mode.

ACKNOWLEDGEMENTS

This work was made possible by partial funding from the Montana Water Center. We wish to thank the U.S. Forest Service Rocky Mountain Research Station and NRCS National Weather and Climate Center for providing the data necessary to perform this study.

REFERENCES

Akaike, H. (1974), A new look at statistical model identification. *IEEE Transactions on Automatic Control*, 19(6), 716-723.

Anderson, E.A. (1968), Development and Testing of Snow Pack Energy Balance Equations. *Water Resources Research*, 4(1), 19-37.

Bates, B.C., 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.

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

Beven, K. (1997), TOPMODEL: A critique. *Hydrological Processes*, 11(9), 1069-1085.

Blöschl, G., Kirnbauer, R. (1991), Point snowmelt models with different degrees of complexity - Internal processes. *Journal of Hydrology*, 129(1-4), 127-147.

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

Garen, D.C., Marks, D. (2005), Spatially distributed energy balance snowmelt modelling in a mountainous river basin: estimation of meteorological inputs and verification of model results. *Journal of Hydrology*, 315(1-4), 126-153.

Haario, H., Laine, M., Mira, A., Saksman, E. (2006), DRAM: Efficient adaptive MCMC. *Statistics and Computing*, 16(4), 339-354.

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

Jencso, K.G., McGlynn, B.L., Gooseff, M.N., Wondzell, S.M., Bencala, K.E., Marshall, L.A. (2009), Hydrologic connectivity between landscapes and streams: Transferring reach- and plot-scale understanding to the catchment scale. *Water Resources Research*, 45, W04428, doi:10.1029/2008WR007225.

Klemeš, V. (1986), Operational testing of hydrological simulation models. *Hydrological Sciences Journal*, 31(1), 13-24.

Madsen, H., Wilson, G., Ammentorp, H.C. (2002), Comparison of different automated strategies for calibration of rainfall-runoff models. *Journal of Hydrology*, 261(1-4), 48-59.

Manabe, S. (1969), Climate and the Ocean Circulation: 1. The Atmospheric Circulation and the Hydrology of the Earth's Surface. *Monthly Weather Review*, 97(11), 739-774.

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

Marshall, L., Nott, D., Sharma, A. (2004), A comparative study of Markov chain Monte Carlo methods for conceptual rainfall-runoff modeling. *Water Resources Research*, 40, W02501, doi:10.1029/2003WR002378.

Marshall, L., Nott, D., Sharma, A. (2005), Hydrological model selection: A Bayesian alternative. *Water Resources Research*, 41, W10422, doi:10.1029/2004WR003719.

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

Neuman, S.P. (2003), Maximum likelihood Bayesian averaging of uncertain model predictions. *Stochastic Environmental Research and Risk Assessment*, 17(5), 291-305.

Ohmura, A. (2001), Physical basis for the temperature-based melt-index method. *Journal of Applied Meteorology*, 40(4), 753-763.

Rango, A., Martinec, J. (1995), Revisiting the degree-day method for snowmelt computations. *Journal of the American Water Resources Association*, 31(4), 657-669.

Smith, T.J., Marshall, L.A. (2008), Bayesian methods in hydrology: A study of recent advancements in Markov chain Monte Carlo techniques. *Water Resources Research*, 44, W00B05, doi:10.1029/2007WR006705.

Wagener, T. (2003), Evaluation of catchment models. *Hydrological Processes*, 17(16), 3375-3378.