Environmental Modelling & Software 25 (2010) 691–701

Contents lists available at ScienceDirect

Environmental Modelling & Software

journal homepage: www.elsevier.com/locate/envsoft

Exploring uncertainty and model predictive performance concepts via a modular snowmelt-runoff modeling framework

Tyler Jon Smith*, Lucy Amanda Marshall

Department of Land Resources and Environmental Sciences, Montana State University, Bozeman, MT 59717, United States

ARTICLE INFO

Article history:
Received 29 April 2009
Received in revised form
18 August 2009
Accepted 11 November 2009
Available online 6 January 2010

Keywords:
Modeling framework
Model comparison
Bayesian inference
Markov chain Monte Carlo simulation
Calibration
Uncertainty analysis
Snowmelt
Predictive performance

ABSTRACT

Model selection is an extremely important aspect of many hydrologic modeling studies because of the complexity, variability, and uncertainty that surrounds the current understanding of watershed-scale systems. However, development and implementation of a complete precipitation-runoff modeling framework, from model selection to calibration and uncertainty analysis, are rarely confronted. This paper introduces a modular precipitation-runoff modeling framework that has been developed and applied to a research site in Central Montana, USA. The case study focuses on an approach to hydrologic modeling that considers model development, selection, calibration, uncertainty analysis, and overall assessment. The results of this case study suggest that a modular framework is useful in identifying the interactions between and among different process representations and their resultant predictions of stream discharge. Such an approach can strengthen model building and address an oft ignored aspect of predictive uncertainty; namely, model structural uncertainty.

1. Introduction

Hydrologic model development covers an extensive range from black-box and empirical approaches to physics-based methods. However, neither black-box nor physics-based models are ideal due to the difficulties that surround each; black-box models lack the realism necessary to be made useful and reliable across settings (Donnelly-Makowecki and Moore, 1999), while physics-based models require theoretical understanding of a complex, changing system (Grayson et al., 1992). Because of these complications, a class of models that attempts to blend the simplicity of black-box models with some of the realism of physics-based models has become more prominent (for example, Atkinson et al., 2002; Farmer et al., 2003; Jothityangkoon et al., 2001; Littlewood et al., 2003). These conceptual models often follow a downward approach to model development (Sivapalan et al., 2003), starting with inclusion of the (perceived) most dominant drivers of the system, while lumping the less important processes into effective parameters requiring estimation via calibration (Klemeš, 1983).

Model selection is perhaps the single most important and often least scrutinized component to the precipitation-runoff modeling problem, especially in conceptual models where parameters are calibrated and exact process descriptions are not assumed. All too often a model structure is selected *a priori* with the objective of proving its appropriateness to the study location, with little or no consideration given to alternate structural formulations (Neuman, 2003). This type of approach often results in a model structure being deemed as either acceptable or unacceptable, but it does not address where the model lies on the spectrum (i.e., how acceptable or unacceptable it is).

Although commonly overlooked, some recent studies are beginning to incorporate this critical component; Fenicia et al. (2008), for example, introduced a modeling study that aimed to improve a conceptual model structure through iterative improvement of process representation (i.e., top-down model improvement) under a multi-objective scheme. While this work represents an important first step in model evaluation, it is limited by the use of a single model conceptualization under which alternatives are examined. Extending this further, Clark et al. (2008) considered 79 unique structures generated by mixing the individual flux equations of four commonly used rainfall-runoff models. The present study offers a compromise of these two prior studies by considering the interactions between alternate model components, but not individual

* Corresponding author. Department of Land Resources and Environmental Sciences, Montana State University, 334 Leon Johnson Hall, P.O. Box 173120, Bozeman, MT 59717, United States.

E-mail address: tyler.smith@msu.montana.edu (T.J. Smith).

1364-8152/$ – see front matter © 2009 Elsevier Ltd. All rights reserved.
doi:10.1016/j.envsoft.2009.11.010

flux representations. However, this study hopes to offer more insights into the relationship that exists between model components that extend beyond the soil moisture accounting context presented by Fenicia et al. (2008) and Clark et al. (2008).

Beyond structural model improvement/assessment strategies, comprehensive precipitation-runoff modeling requires the modeler to consider the additional tasks of calibration, uncertainty analysis, and assessment. Many methodologies exist and have been applied to these essential components, none of which have become universally accepted as the best approach (e.g., Beven et al., 2007; Mantovan and Todini, 2006). Commensurate with these modeling duties, the objective of this study was to examine the relative performance of multiple, alternate model structures featured within a modular snowmelt-runoff modeling framework under a formal Bayesian inferential approach to parameter estimation and uncertainty analysis. The models conceptualize major watershed processes assumed to be important to the hydrologic system. We consider several elements important to the assessment of models within the modeling suite: their individual predictive performance, the uncertainty of individual models and the entire modeling suite, and the ability of the models to simulate variables not considered in the calibration process.

This paper is divided into the following sections: Section 2 presents an overview of the methods employed in this study, Section 3 describes the study location where the methods were applied, Section 4 introduces the modeling framework and each modular component, Section 5 details the assessment of the framework, Section 6 provides a discussion of the results, and Section 7 offers relevant conclusions.

2. Overview of methods

A multi-model approach to precipitation-runoff modeling problems is favorable, according to Neuman (2003), in that it reduces the potential of committing Type I errors (rejecting valid alternative model structures, by omission) and Type II errors (accepting an invalid model structure). The use of a multi-model selection framework not only reduces the potential Type I and II errors, it allows for the inclusion of multiple alternative representations and a means of comparison across the different formulations, in line with a top-down model development philosophy (e.g., Bai et al., 2009). In this study, we consider multiple conceptualizations of elements of the hydrologic system, including snowmelt rates, soil water storage, and runoff routing. Specific details of the modeling framework applied herein are described in Section 4.

The selection of simple conceptual modeling components within the framework necessitates a calibration technique to estimate the effective model parameters. More recently, it has become important in hydrologic studies to recognize that model parameters and simulations are inherently uncertain, as errors arise in using an imperfect model structure and in using inexact data. For the purposes of this study, all of the modular structures associated with the modeling framework were implemented under a Bayesian inferential framework. The use of Bayesian inference in complex modeling problems (like those encountered in hydrology) has gained attention because of the ease with which uncertainties can be considered via posterior parameter distributions (Gilks et al., 1996). However, because the posterior distribution is generally (and in this specific case) analytically intractable, a numerical/statistical estimation approach must be implemented.

Markov chain Monte Carlo (MCMC) simulation is a commonly used statistical method for approximating the posterior distribution, dating to the early 1950s and the development of the Metropolis algorithm (Metropolis et al., 1953). Such methods are now gaining particular popularity in hydrologic studies (e.g., Bates and Campbell, 2001; Kuczera and Parent, 1998; Marshall et al., 2004; Smith and Marshall, 2008). While a variety of different algorithms exist for constructing a random walk Markov chain that converges to the posterior distribution, the Delayed Rejection Adaptive Metropolis (DRAM) algorithm (Haario et al., 2006) was selected for use in this study. The selection of the DRAM algorithm was motivated by past findings of its superiority in hydrologic settings over other MCMC methods by Smith and Marshall (2008), building on the results of Marshall et al. (2004).

3. Study location

Snowpack accumulation and subsequent melt plays a critical role in water resource availability in many mountainous regions. Modeling these processes requires an understanding of snowmelt-runoff dynamics in such areas. For this study, models were implemented for the Stringer Creek watershed located within the Tenderfoot Creek Experimental Forest (TCEF, 46°55′ N, 110°52′ W) of Central Montana, USA (see Fig. 1). TCEF was established by the United States Department of Agriculture (USDA) in 1961 and is committed to the research of subalpine forests influenced by considerable expanses of lodgepole pine (*Pinus contorta*) typically found on the east slope of the northern Rocky Mountains. Average annual precipitation across TCEF is approximately 850 mm per year, largely in the form of snow. For further details on vegetation, climate, and geology of TCEF please refer to Ahl et al. (2008).

The Stringer Creek watershed encompasses 555 ha of the total 3700 ha of TCEF and covers a range in elevations from approximately 1990 to 2430 m. Beyond the natural factors that make Stinger Creek a useful study location, data useful for implementing conceptual-type models are available for this watershed as well. Time series data were selected for the period of study spanning from March 2002 to September 2007, using a 12-h time step. The data used for this study included streamflow observations from the flume at the outlet of Stringer Creek (managed by the Rocky Mountain Research Station, USFS) and meteorological observations (including snow water equivalents) from the Onion Park SNOTEL (2258 m, managed by the NRCS). Additionally, all topographic data was obtained from a 10 m grid digital elevation model (DEM). For additional information on available instrumentation within Stringer Creek and the greater TCEF, please refer to Jencso et al. (2009).

4. Modular hydrologic modeling framework: development

In acknowledgement of the potential shortcomings of a single model approach to precipitation-runoff modeling, a modular hydrologic modeling framework was developed and applied to the Stringer Creek watershed. In order to account for the dominant processes three primary modules (each consisting of alternate conceptualizations) were built into the framework: soil moisture accounting and runoff routing, snowmelt accounting, and the method of semi-distributing precipitation inputs across the watershed area. The framework modules are then combined in all possible manners to create a total of thirty possible "modular structures". Given that each structure is a conceptualization of reality developed under a top-down modeling approach, the modules attempt to characterize the perceived dominant processes while lumping all others into several effective parameters.

4.1. Soil moisture accounting and runoff routing

Representing the water accounting processes of the physical system, three competing model base structures for soil moisture accounting and runoff routing were selected (refer to Fig. 2). The three approaches chosen for this study were a simple bucket model

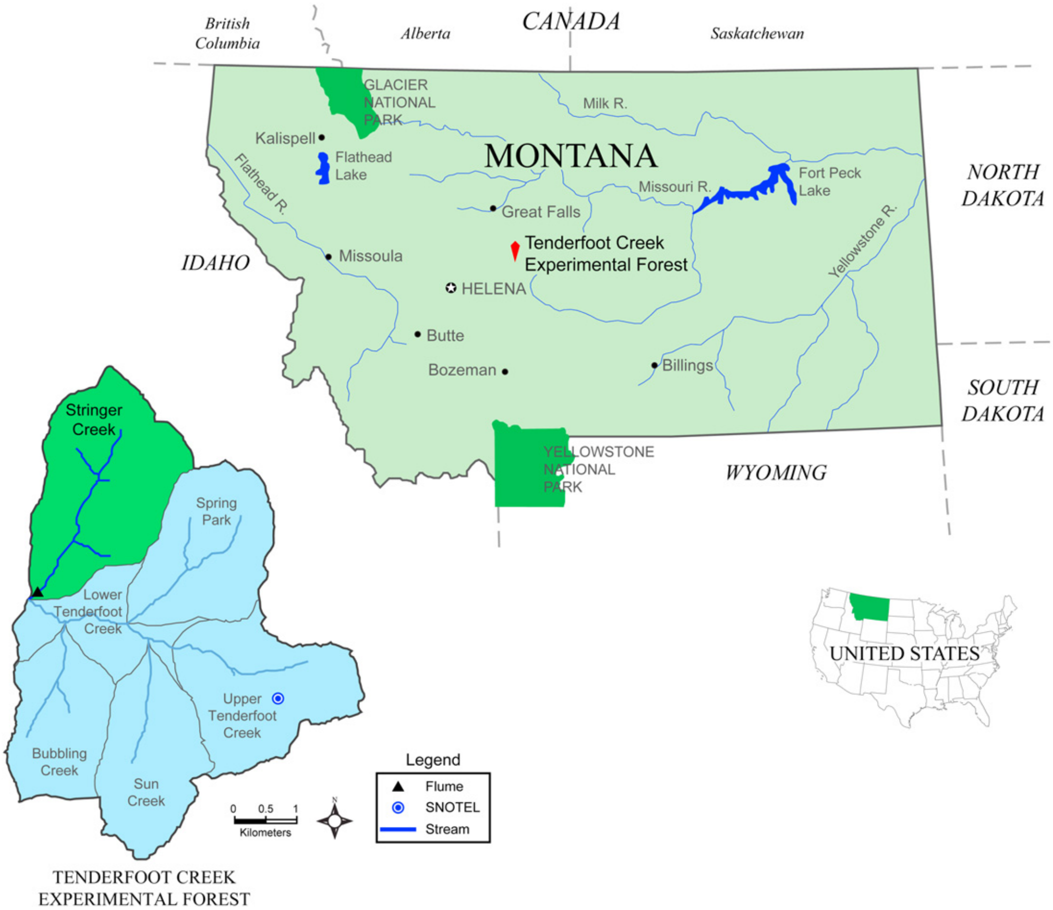

Fig. 1. The greater Tenderfoot Creek Experimental Forest, with Stringer Creek watershed delineated.

(BUC; e.g., Manabe, 1969), the Probability Distributed Model (PDM; Moore, 1985, 2007) and TOPMODEL (TOP; Beven, 1997; Beven and Kirkby, 1979).

Each of the approaches selected for inclusion in the soil moisture accounting and runoff routing module seek to route water to the watershed outlet via conceptual storages. The primary difference among the three conceptualizations is in the manner that the soil moisture is distributed. The bucket model integrates variability across the watershed by applying a fixed storage capacity to soil moisture via a uniform distribution. The Probability Distributed Model considers storage in a similar manner, with the exception that the soil moisture is distributed by a Pareto distribution, with a parameter defining the exact shape. TOPMODEL approaches the soil moisture accounting in a more structured manner based on the calculation of the topographic index (Kirkby, 1975), derived from a digital elevation model of the watershed.

The value of the three components selected for soil moisture accounting and runoff routing in this study is in the unique similarities and differences among each in their first-order representations of the Stringer Creek watershed. The similarity of the bucket and Probability Distributed models allows for a better understanding of the worth associated with the soil moisture accounting procedure

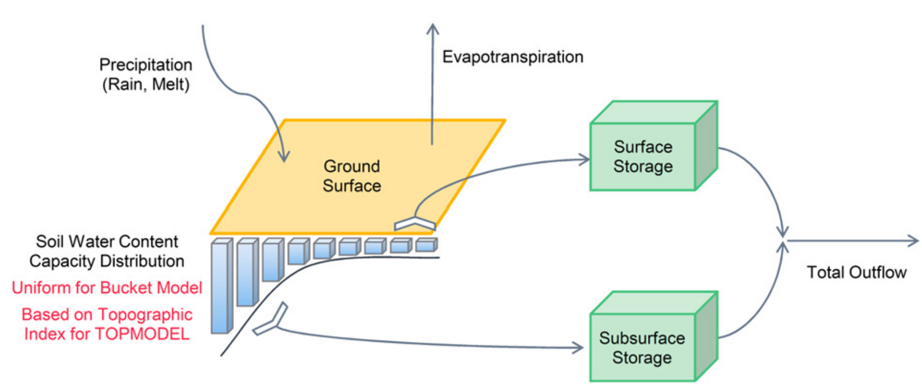

Fig. 2. Schematic diagram of the Probability Distributed Model soil moisture accounting and runoff routing component. Each of the three competing formulations follow the same basic design, but have alternate descriptions of the soil water content capacity distribution. Adapted after Smith and Marshall (2008).

used in each (i.e., whether the variable soil water content storage of the PDM provides a benefit). The dissimilarity of TOPMODEL to the bucket model or the PDM allows for the opportunity to better understand the merit of the variable contributing area approach implemented in TOPMODEL.

The data requirements for each of the three soil moisture accounting and runoff routing modules are quite similar. Each base structure requires the input of precipitation (as rainfall or snowmelt) and evapotranspiration data to produce simulated stream discharge. TOPMODEL has the additional data requirement of a digital elevation model of the study site, necessary for the calculation of the topographic index.

4.2. Snowmelt accounting

Given the snowfall-dominated precipitation pattern of the Stringer Creek watershed, the soil moisture accounting and runoff routing modules must be augmented with a snowmelt accounting module to appropriately characterize the inputs of the physical system. Continuing the use of downward model development strategies, three related, conceptual snowmelt accounting techniques were selected for use within the modeling suite: the temperature index (e.g., Boggild et al., 1999; Hock, 2003; Lindström et al., 1997; Martinec and Rango, 1986), the radiation index, and the temperature and radiation index approaches (e.g., Brubaker et al., 1996; Cazorzi and Fontana, 1996; Kustas et al., 1994). The temperature and radiation index approach is a simple summation of the temperature index approach and the radiation index approach and is given by

$$Melt_{\mathrm{i}} = DDF \cdot \max\left\{ \begin{matrix} 0 \\ Tbar_{\mathrm{i}} - TF \end{matrix} \right\} + NRF \cdot \max\left\{ \begin{matrix} 0 \\ Rn_{\mathrm{i}} \end{matrix} \right\} \qquad (1)$$

where i indexes time, *Melt* is the predicted snowmelt, *Tbar* is the average air temperature over 36 h (to account for snowpack cold content), *Rn* is the net radiation, *TF* is the "threshold temperature for snowmelt", *DDF* is the "degree-day factor", and *NRF* is the "net radiation factor".

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; Barry et al., 1990; Blöschl et al., 1991; Marks and Dozier, 1992), the constraints posed by such methods are extensive (e.g., Rango and Martinec, 1995). With limited point data from the Onion Park SNOTEL station the application of an energy balance approach was not feasible, and instead the simplified methods outlined previously were selected. The use of conceptual, index-based approaches to snowmelt accounting is well established in the literature (e.g., Blöschl and Kirnbauer, 1991; Braithwaite, 1995; Daly et al., 2000; Hock, 1999), with Ohmura (2001) providing a link to the physical basis of such approaches. Conceptual, temperature index approaches have been found to often perform comparably to full energy balance methods on a catchment scale (WMO, 1986).

4.3. Distribution of snowmelt accounting

The implementation of a spatial distribution scheme associated with snowmelt modeling 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 while not significantly increasing the data required to implement the model. In an effort to account for the spatial variability (potentially) present in the precipitation inputs to the system, four different conceptual semi-distribution modules have been assimilated into the modeling framework. Each semi-distribution method requires the availability of a digital elevation model for 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; this provides a weighted description of snowmelt and rainfall that are to be used as the "precipitation" inputs to the soil moisture accounting and runoff routing module. The four semi-distribution modules considered in the framework are: no semi-distribution, semi-distribution by aspect, semi-distribution by elevation, and semi-distribution by aspect and elevation; the procedure for each semi-distribution method is outlined in the following subsections.

4.3.1. Semi-distribution by aspect

Watershed aspect is an important control on snowmelt rates due to its relationship with to the energy balance of a snowpack, manifested as solar radiation (Cazorzi and Fontana, 1996). Given the desire to better characterize the physical processes driving the hydrologic system and the correlation between aspect and radiation, disaggregation of a watershed into hydrologic response units (HRUs) based on aspect is not unfounded (e.g., Braun et al., 1994; Charbonneau et al., 1981).

Semi-distribution by aspect considers only the calculation of snowmelt input to the system by breaking up the watershed into three aspect classes; North-facing (315°–45°), South-facing (135°–225°) and East-West-facing (45°–135° and 225°–315°). Each aspect class has been assigned an *aspect factor* (−1 for North-facing, 1 for South-facing, and 0 for East-West-facing). The formulation of aspect discretization used for this study is based on the assumption that South-facing aspects produce snowmelt first, followed by East-West-facing and North-facing, an assumption supported by Cazorzi and Fontana (1996). The aspect factor is then used as a multiplier in the calculation of snowmelt that assesses the aspect-specific snowmelt (for the North and South aspect classes) produced above the base snowmelt level (assumed to be that of the East-West aspect class). Eq. (2) demonstrates the implementation of the aspect factor with the temperature and radiation index approach to snowmelt accounting

$$\begin{aligned} Melt_{\mathrm{i}} &= DDF \cdot \max\left\{ \begin{matrix} 0 \\ Tbar_{\mathrm{i}} - TF \end{matrix} \right\} \\ &\quad + (NRF_0 + NRF_1 \cdot ASPF_{\mathrm{j}}) \cdot \max\left\{ \begin{matrix} 0 \\ Rn_{\mathrm{i}} \end{matrix} \right\} \end{aligned} \qquad (2)$$

where i is the index on time, j is the index on aspect class, *ASPF* is the aspect factor, NRF_0 (model parameter) is the base net radiation factor (equivalent to *NRF* for non-aspect semi-distribution), and NRF_1 (model parameter) is the aspect-dependent net radiation factor. The weight of each aspect class is determined by the ratio of the number of DEM cells in a given class to the total number of cells in the watershed.

4.3.2. Semi-distribution by elevation

A known relationship exists between precipitation (and temperature) and elevation (see Daly et al., 1994 for a comprehensive review). Invoking this (in a similar fashion to the previously described approach to dividing a watershed according to aspect), a watershed can be simply divided into HRUs based on evenly spaced elevation bands to help improve the link between process understanding and model conceptualization. Such a practice has gained growing support in snow-dominated hydrologic systems (e.g., Fontaine et al., 2002; Hartman et al., 1999).

In this application, semi-distribution by elevation considers the calculation of all precipitation (and temperature) input to the system by breaking up the watershed into five uniform width elevation classes. The precipitation input is calculated based on the use of lapse rates (calibrated model parameters) applied to the data (e.g., Liston and Elder, 2006), utilizing a reference point (Onion Park SNOTEL) located within the elevation range of the Stringer Creek watershed itself. Because snowmelt accounting relies on air temperature, both precipitation (and snow water equivalents) and temperature are computed for each elevation class. The total precipitation input to the soil moisture accounting and runoff routing module is then a mean weighted input of the five individual classes. The weight of each elevation class is determined by the ratio of the number of DEM cells in a given class to the total number of cells in the watershed.

4.3.3. Semi-distribution by aspect and elevation

The final approach to watershed semi-distribution included in the modeling framework is based on a combination of the aspect semi-distribution and the elevation semi-distribution approaches. Such an approach has been used previously to enhance the physical reality of the snowmelt accounting model component (e.g., Braun et al., 1994). Calculation of semi-distributed input data proceeds with the discretization of the watershed into three aspect classes for each of the five elevation classes (a total of 15 classes), as described in the previous subsections. The semi-distributed input data is then passed to the snowmelt accounting routine.

As with the aspect semi-distribution approach, the temperature index method for snowmelt accounting is precluded by the use of this semi-distribution scheme. This exclusion is a result of the assumption in the aspect semi-distribution conceptualization that snowmelt varies with aspect as a result of differing net radiation influence on different aspects; because the temperature index snowmelt accounting approach is independent of net radiation it must also be independent of aspect, based on this underlying premise.

4.4. Summary of modular modeling framework

The modular modeling framework specified above represents a mix of alternate conceptualizations of soil moisture accounting, snowmelt accounting, and catchment semi-distribution. The modular components cover a range of commonly used structures. It should be noted that some of the components represent special cases of others. For the soil moisture accounting module, the bucket model is a special case of the PDM (where the shape of the soil water content capacity distribution is uniform). In this case, comparison of these structures is very much an analysis of the value of the shape parameter applied to the soil moisture storage. In the case of the snowmelt accounting modules, the temperature index and radiation index approaches are special cases of the temperature and radiation index approach (as it is a simple sum of these methods). However, in this case it is not simply a question of the value of adding a parameter; in these, alternate snowmelt accounting procedures require and utilize alternate sources of data (radiation index uses only radiation data, temperature index uses only air temperature data, etc.). These cases represent a study of the value of different data to predict snowmelt while utilizing alternate parameterizations of the temperature and radiation index approach. To maintain a continuity of presentation associated with these "special cases", all comparisons will be analyzed under a model comparison context.

5. Modular hydrologic modeling framework: MCMC algorithm implementation and model assessment

The assessment of model suitability can take on many forms. While studies often focus primarily on how well a model's prediction matches the observed data, a more thorough approach considering at least performance, uncertainty, and realism is advocated by Wagener (2003). By implementing this type of multi-faceted approach to evaluate the entire modeling framework of this study, an even more comprehensive understanding of reasons behind specific successes or failures of a given modular structure can be extracted.

Model calibration was performed on each of the thirty modular structures of the modeling framework using the calibration data set collected from March 2002 to September 2005. An *a priori* decision must be made on the appropriate likelihood function to use in the DRAM algorithm. For this study, all models were assessed using a likelihood that assumes the errors are normally, independently, and identically distributed

$$P(Q|\theta) = \left(2\pi\sigma^2\right)^{-n/2}\prod\exp\left\{-\frac{(Q_{\text{obs}}-Q_p)^2}{2\sigma^2}\right\} \tag{3}$$

where n is the number of data points, σ^2 is the variance (whose posterior distribution was sampled by the DRAM algorithm), Q_{obs} is the observed streamflow, and Q_p is the predicted streamflow for a given parameter set. The selection of this likelihood function (Eq. (3)) has been commonly made in the literature (for example, Bates and Campbell, 2001; Samanta et al., 2007; Thiemann et al., 2001; Vrugt et al., 2003). The appropriateness of Eq. (3) will be addressed in further detail in Section 5.2.

Convergence of a model's parameters to a stationary posterior probability distribution is of utmost importance in proper calibration of a model when using an MCMC approach such as the DRAM algorithm implemented in this study. Criteria establishing convergence of MCMC chains are common and varied in their approach (e.g., Brooks and Gelman, 1998; Brooks and Roberts, 1998). For the purposes of this study, the approach proposed by Gelman and Rubin (1992) was selected, where an R-statistic is calculated based on a ratio of the variance within a single chain and the variance between multiple parallel chains (for each parameter). However, due to the sizeable number of models and substantial run-times associated with each, convergence was only directly tested for a representative subset of modular structures within the framework. The necessary number of parameter samples required for convergence of these modular structures was used to guide inference in the others. Those not directly checked with the R-statistic underwent visual assessment of convergence using the trace of each parameter's sampled values from the posterior distribution. Each of the modular structures contained in the modeling framework will be analyzed in the following subsections with respect to each aspect of the outlined assessment strategy (i.e., performance, uncertainty, and realism). Note that for the purposes of this paper, we focus on the utility of the methodology for comparing alternate model structures.

5.1. Assessment metric: performance

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 set being used to calibrate model parameters and the other being used for an "independent" assessment of model performance (often referred to as validation). The calibration phase covered approximately four years, from March 2002 to September 2005, whereas the "validation" period covered two years, from October 2005 to September 2007. The validation period experiences an increase of approximately 15% in the mean total precipitation (and similarly snow water equivalents), but the timing of the events were consistent with that observed for the calibration period (see Fig. 3). This extends the meteorological conditions under which the model parameters were estimated, allowing for a truer test of each model's merits during the assessment via the validation data set.

Owing to the fact that more complex models often provide an improvement in the calibration period due to a greater flexibility in structure afforded by additional parameters (Tripp and Niemann, 2008), the application of performance assessment that goes beyond the traditional split-sample approach (i.e., calibration-validation testing) is needed. As a way of understanding the effect of dimensionality on the performance across the suite of structures, two related measures that penalize models based on complexity, Akaike's information criterion (AIC, Akaike, 1974) and Bayes' information criterion (BIC, Schwarz, 1978), were computed. Eqs. (4) and (5) represent the AIC and BIC statistics, respectively.

$$AIC = -2 \cdot \ln[P(Q|\theta)] + 2k \tag{4}$$

$$BIC = -2 \cdot \ln[P(Q|\theta)] + k \cdot \ln(n) \tag{5}$$

where $\ln[P(Q|\theta)]$ is the maximum log-likelihood, k is the number of model parameters, and n is the number of observations in the data set.

Taking each modular structure's complexity into account resulted in performance rankings slightly different to those obtained based on the optimum log-likelihood. Although the AIC and BIC measures also differed slightly to one another (as one would expect), the rankings on a whole were very similar. Ye et al. (2008) note that the AIC is preferred to the BIC when the ratio of the number of observations to the number of calibrated parameters is less than 40 and that when this ratio is large the BIC is favored. For the case study examined here, this ratio is on the order of 200 or greater for all models. Because of this, the BIC was used as the criteria for all future rankings of stream discharge.

A summary of the stream discharge calibration results for each of the thirty modular structures contained within the modeling framework are given in Fig. 4(a), where the calibration performances (based on the BIC, refer to Eq. (5)) have been binned into six evenly populated classes. The patterns present in Fig. 4(a) 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 and to the right in Fig. 4(a) (i.e., toward the PDM soil moisture accounting module and the temperature and radiation snowmelt accounting module), indicating better general performance by more complex structures (where complexity is based on number of parameters; given in upper left corner of cells in Fig. 4).

The relative value of module components is also clearly illustrated in Fig. 4(a). This is especially apparent in the snowmelt accounting routines, illustrated by the generally improved performances of models along the right portion of the figure with net radiation performing better than ambient air temperature as a predictor of snowmelt. The move downward in this figure exposes the prevalence of better performing models in structures that included the PDM module. Less clear is the value of the semi-distribution module to the overall structure. There was often little change in overall performance generated by a change in semi-distribution method.

The modeling framework was further broken into 31 comparison units covering each possible combination produced by holding two modules constant and varying the third module. This allowed for the isolation of the effect of the variable component in the modular structures, relative to the constant components. As was indicated in Fig. 4(a), the temperature index approach to snowmelt accounting was consistently the least effective of the three different options. Focusing on linking this inefficiency to the hydrograph, Fig. 5(Inset) reveals that the primary area of difference between the three snowmelt accounting options lies in the rising limb of the hydrograph peaks (where each simulation is the result of the maximum log-likelihood parameter set associated with the given structure). The temperature index method routinely resulted in a much "flashier" response during this segment of the hydrograph (where streamflow responds quickly to faster melting accumulated snow), especially in combination with the no semi-distribution component. This can be quantified by the flashiness index presented by Baker et al. (2004), where discharge predictions were approximately 45% flashier for the structure utilizing the temperature index approach as compared to structures employing either the radiation index or temperature and radiation index approaches. When the semi-distribution by elevation was considered, this "flashy" signal was largely muted. This example provides a clear and obvious interaction that can exist between model structural components.

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.

Separation of the streamflow record into categories of low, moderate, and high flows allowed for further analysis of model fits (based on MSE) to the observed hydrograph. The flow classes were determined by setting a threshold of 0.1 (mm/12-h) or less for low flows and a threshold of 1.85 (mm/12-h) or greater for high flows, with moderate flows falling in between. Fig. 6 shows the portion of the total mean squared error corresponding to the prescribed flow categories. From this illustration the importance of modeling the high flows becomes apparent. Although not surprising since the mean squared error tends to accentuate errors in high flows (Clark et al., 2008), the use of such an objective function for understanding model simulations across flow categories is not uncommon (Wagener and McIntyre, 2005). Fig. 6 is additionally valuable in highlighting the nature with which the relative importance of the errors varies across the different flow categories and structures. For instance, the TOPMODEL-based predictions perform poorly on the low flows relative to either bucket- or PDM-based structures.

Table 1 follows closely with the information presented graphically in Fig. 6, showing the rank of each modular structure (1–30) for each of the flow categories considered (low, moderate, high), as well as an overall performance rank. Note that these ranks are based on MSE, which produce the same ranks as the optimum log-likelihood value. However, they do not take into account the complexity of the structure as opposed to the overall ranks (based on BIC) presented visually in Fig. 4(a). Interestingly, several of the modular structures perform very well under one flow condition, while performing quite poorly under others.

For example, the *PDMsdA_TRn* structure models high flows much more effectively than low and moderate flows, resulting in a high overall performance ranking (because the MSE accentuates error in high flows). The *PDMsdAE_TRn* structure models low and moderate flows much better than the *PDMsdA_TRn* structure, while maintaining good performance in the high flows. In this example, the only difference between the two structures is the distribution of snowmelt accounting (aspect semi-distribution versus aspect and elevation semi-distribution). This perhaps points to an increased value for the aspect and elevation semi-distribution method for simulation of low and moderate flows. Although many such

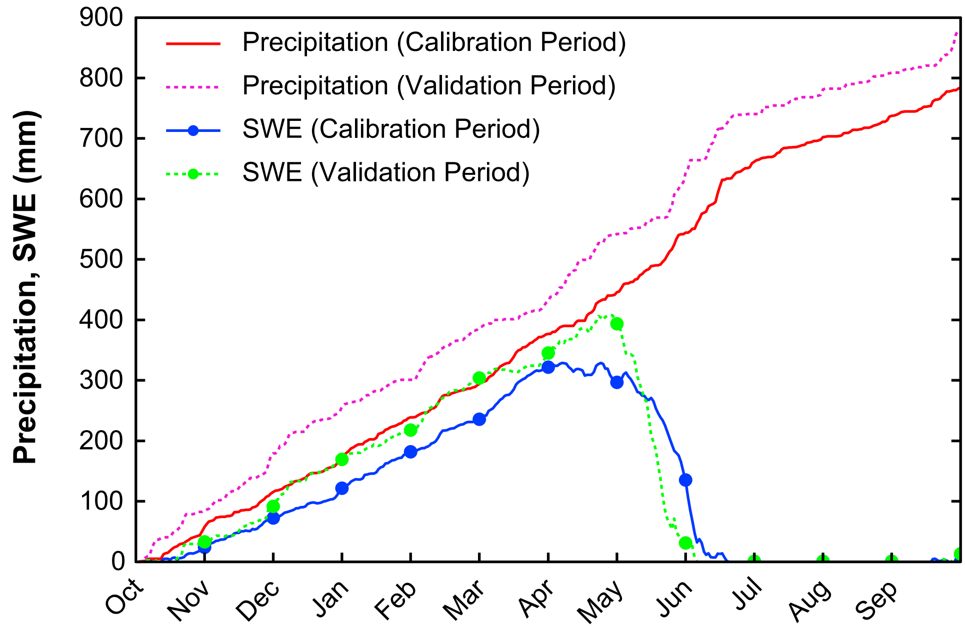

Fig. 3. Mean accumulated precipitation and snow water equivalents for the calibration period (solid lines) and the 'validation' period (dashed lines).

examples are apparent from Table 1 and Fig. 6, it should be noted that it is difficult to definitively link conceptual model processes to real watershed processes.

5.2. Assessment metric: uncertainty

The uncertainty associated with any single model structure is of substantial importance to the understanding of the risk associated with its predictions (Chiang et al., 2007). An appropriate assessment of uncertainty has the ability to temper or intensify the weight of the performance-based assessment strategies discussed previously, by introducing further criteria for selecting the optimal structure. Uncertainty can be manifested in a variety of ways, from parameter uncertainty to input uncertainty to structural uncertainty.

The benefit of implementing a Bayesian approach is in its ability to characterize the parameter uncertainty via parameter posterior distributions. These distributions on each parameter have the potential to offer insights into the behaviors of each. For example, it may be found that a certain parameter has a relatively diffuse posterior distribution and may be

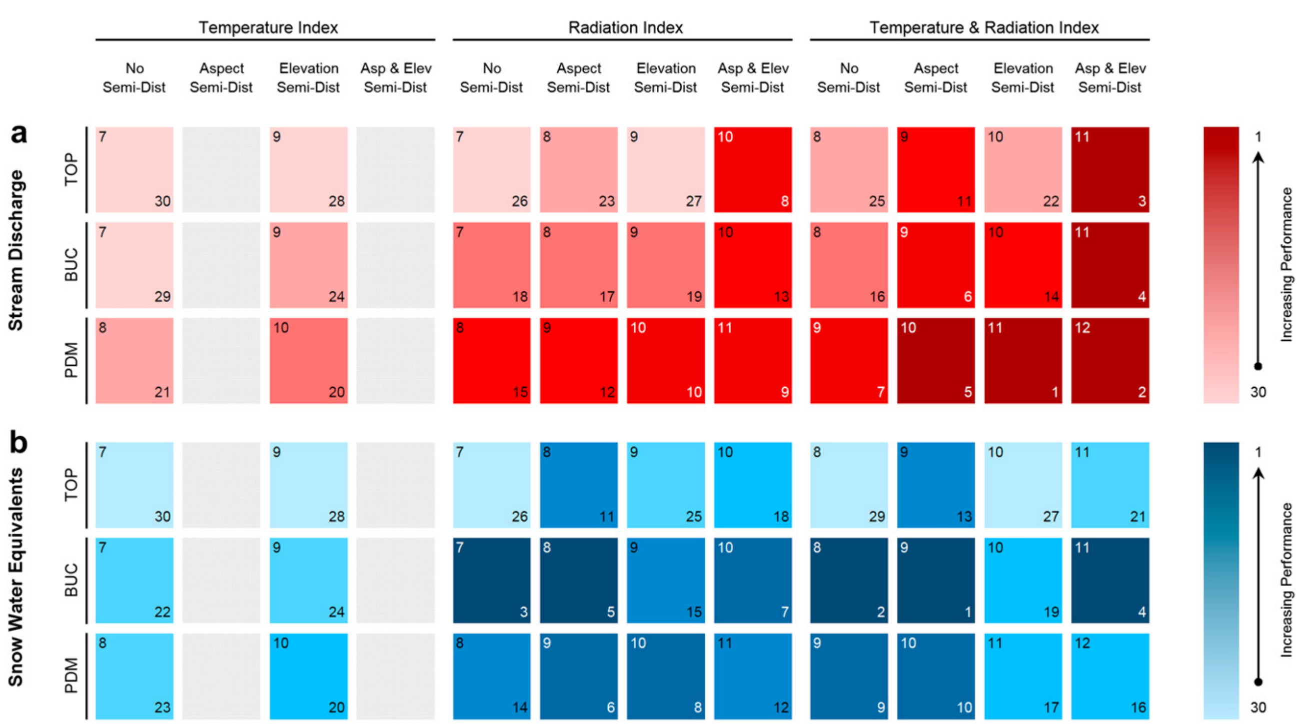

Fig. 4. Calibration results for each modular structure defined within the modeling framework, for (a) stream discharge (ranked based on BIC) and (b) snow water equivalents (ranked based on MSE). Model dimension is given in the upper left corner of each cell and model rank is given in the lower right corner of each cell.

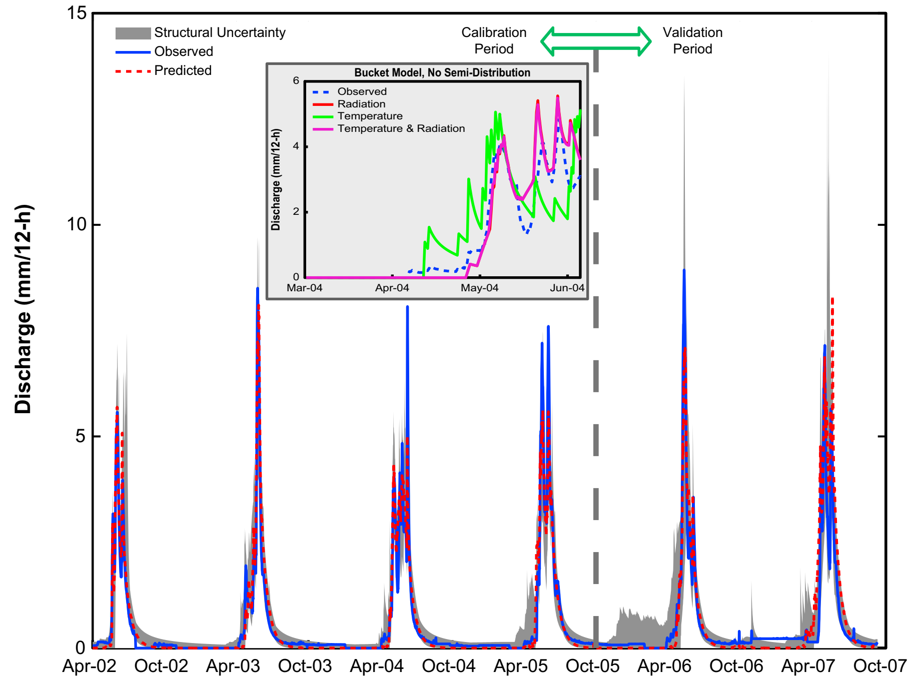

Fig. 5. Modeling framework structural uncertainty and streamflow simulation for "optimal" modular structure: Probability Distributed Model with temperature and radiation index snowmelt accounting, semi-distributed by elevation. Simulations for both the calibration and validation periods are shown. **Inset** shows three model simulations corresponding to different conceptualizations of snowmelt.

unnecessary to the structure's success. Identification of characteristic posterior distributions (e.g., uniform, flat, peaked, bimodal, threshold, etc.) is an important tool in model checking and model identifiability under a Bayesian inferential approach; this is similar to the use of dotty plots in the GLUE methodology (e.g., Wagener and Kollat, 2007). 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.

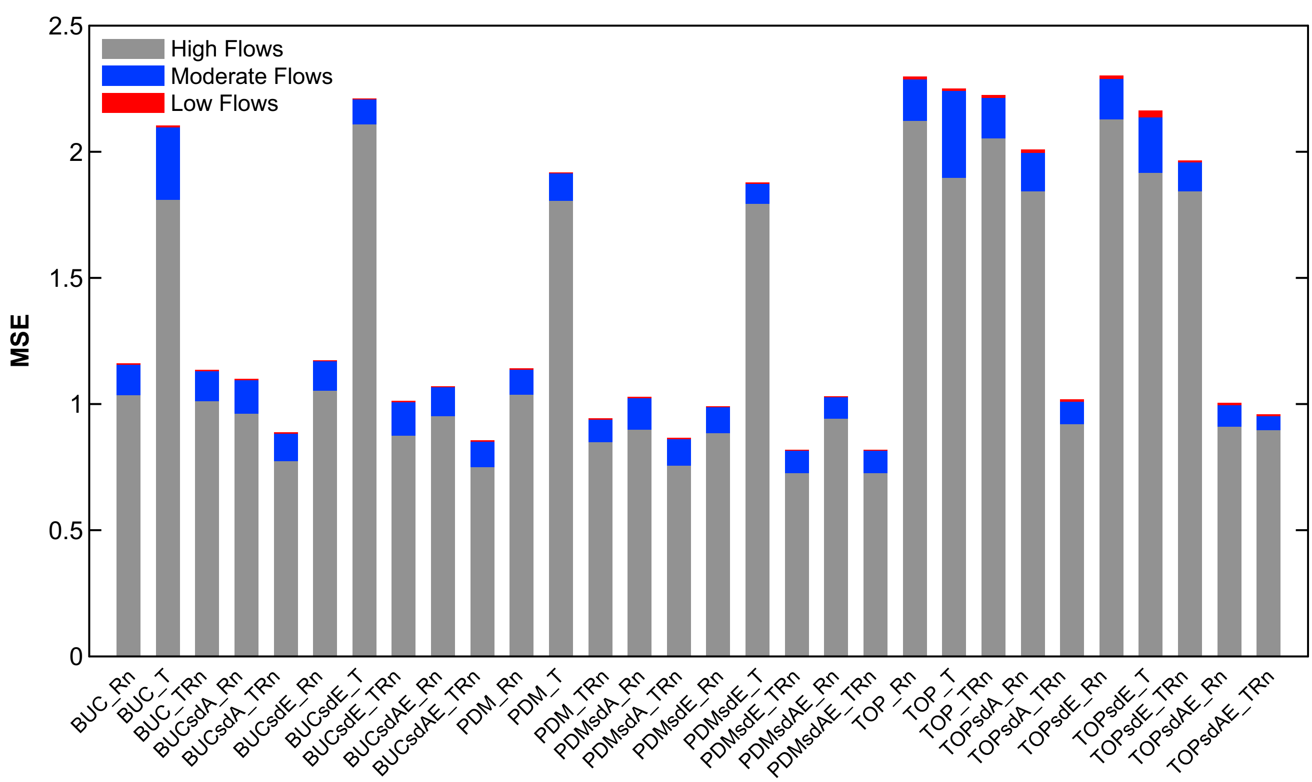

Fig. 6. Mean squared error by flow category (low, moderate, high) for each of the thirty modular structures, over the calibration period.

Table 1
Mean squared error performance rank for each modular structure analyzed for the low, moderate, and high flow conditions, as well as overall performance.[a]

Modular structure	Performance rank				Modular structure	Performance rank				Modular structure	Performance rank			
	Low	Mod.	High	Overall		Low	Mod.	High	Overall		Low	Mod.	High	Overall
BUC_Rn	17	20	17	19	PDM_Rn	11	9	18	15	TOP_Rn	27	27	29	27
BUC_T	22	29	22	29	PDM_T	8	14	21	21	TOP_T	23	30	25	30
BUC_TRn	18	19	16	16	PDM_TRn	10	8	6	7	TOP_TRn	26	25	27	25
BUCsdA_Rn	15	22	15	17	PDMsdA_Rn	16	21	10	12	TOPsdA_Rn	29	24	24	23
BUCsdA_TRn	12	15	5	6	PDMsdA_TRn	14	13	4	5	TOPsdA_TRn	25	7	12	11
BUCsdE_Rn	1	18	19	18	PDMsdE_Rn	3	12	8	10	TOPsdE_Rn	28	26	30	26
BUCsdE_T	2	11	28	24	PDMsdE_T	13	2	20	20	TOPsdE_T	30	28	26	28
BUCsdE_TRn	19	23	7	14	PDMsdE_TRn	7	6	2	2	TOPsdE_TRn	21	16	23	22
BUCsdAE_Rn	4	17	14	13	PDMsdAE_Rn	6	3	13	8	TOPsdAE_Rn	24	4	11	9
BUCsdAE_TRn	9	10	3	4	PDMsdAE_TRn	5	5	1	1	TOPsdAE_TRn	20	1	9	3

[a] Where BUC refers to the bucket model, PDM refers to the Probability Distributed Model, TOP refers to TOPMODEL, sdA refers to aspect semi-distribution, sdE refers to elevation semi-distribution, sdAE refers to aspect and elevation semi-distribution, Rn refers to radiation index snowmelt accounting, T refers to temperature index snowmelt accounting, and TRn refers to temperature and radiation index snowmelt accounting.

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 model structure (at every time step), the overall structural uncertainty of the modeling framework can be investigated. Fig. 5 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; the Probability Distributed Model, semi-distributed by elevation with temperature and radiation index snowmelt accounting (*PDMsdE_TRn*) based on BIC model ranking. The uncertainty in the validation period tends to be greater than during the calibration period, which is consistent with expectations. However, individual model uncertainties were generally consistent in both calibration and validation periods (i.e., the parts of the hydrograph that were most uncertain during calibration tended to also be most uncertain for the validation period).

Beyond the assessment of structural uncertainty for the complete modeling framework, 95% prediction intervals were computed for each model for the validation period. Evaluation of uncertainty in this manner enables the use of two important statistics: the reliability and sharpness associated with the interval. The reliability statistic reflects the percent of the observed data that are encapsulated by the interval, while the sharpness statistic is a measure of the mean width of the interval (Yadav et al., 2007). Here, the reliability of the modular structures ranged from 90.3% to 96.5% with a mean of 93.7%. This result provides additional confirmation of the use of Eq. (3) as an appropriate likelihood function because the prediction intervals on average contain the appropriate percentage of observed data. The modular structures featured sharpness values that varied between approximately 1.4 (mm/12-h) to 2.5 (mm/12-h) with a mean value of 1.9 (mm/12-h).

5.3. *Assessment metric: realism*

The assessment of precipitation-runoff models based on one or more internal system process 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; Uhlenbrook et al., 1999). The process of applying conceptual model structures complicates the ability to link model variables to actual watershed states because of the intrinsically aggregated nature of the true spatial variability; however, while difficult, studies have attempted to relate conceptual models to watershed processes (Seibert and McDonnell, 2002).

Accepting the limitations of the forcing data (precipitation, streamflow), snow water equivalents data was used as a first step toward checking the model's realism. This data represents an "independent" 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.

Fig. 4(b) summarizes the assessment of model realism for the calibration period. Each model's ability to simulate snow water equivalents was ranked from best to worst; the ranks are based on mean squared error rather than BIC because estimates of the log-likelihood required for the BIC are not available for auxiliary data. The results were found to indicate a pattern of general consistency to the results found for streamflow. Structures that included the radiation index or temperature and radiation index approach generally provided better predictions of snow water equivalents, while the semi-distribution method yielded insignificant patterns in relation to the overall prediction of the internal snowmelt dynamics of the Stringer Creek watershed.

6. Discussion of results

Because watershed-scale systems are extremely complex, variable, and uncertain, failure to consider multiple model structures will undoubtedly lead to Type I and/or Type II errors that can drastically inflate uncertainties already associated with the model's resulting simulations. The overall assessment of hydrologic models, then, is of crucial importance, and the consideration of multiple alternate representations of the physical system must be a fundamental component of any such assessment. The assessment of the introduced modular modeling framework proceeded through a multi-faceted approach that considered performance, uncertainty, realism, and complexity across each of the modular structures. Figs. 4, 5, and 6, along with Table 1, summarize the results of the assessment.

The interrelationship between a model's predictive power and its structural complexity was assessed using both the AIC and BIC, statistics that balance a model's performance with its number of parameters. For this study, some models had maximum log-likelihood values that were similar enough to one another that the penalty associated with having additional parameters changed the order in which the models were ranked. This study indicates that even for a fairly constrained set of modular structures the tradeoff between performance and complexity is clearly definable. In a general sense predictive performance was found to increase with increasing model dimension, although the modular structure with the best overall performance (based on BIC) was not the most complex.

Beyond using a performance-based assessment to identify the "optimal" modular structure or the relationship with structural complexity, analysis of the performance of each modular structure yielded several important results relating to the value of individual modules. The temperature index snowmelt accounting approach was found to be inferior to those containing net radiation data, a strong reflection of the importance of net radiation to the energy balance of the snowpack for the watershed test case. Further, the Probability Distributed Model component resulted in consistent improvement in performance as compared with either the bucket model or TOPMODEL components. This may reflect a more appropriate representation of the soil water content capacity distribution by the PDM than the assumption of uniform spatial variability made in the bucket model and its flexibility to handle the specific flow contributing characteristics over the rigidity introduced by TOPMODEL's use of the topographic index. Because the topographic index assumes that all locations within the watershed have the potential to contribute to streamflow, many watershed cells may be represented as active when they are not actually contributing. Recent work has shown that only a small percentage of the Stringer Creek watershed has a groundwater connection that persists throughout the year (Jencso et al., 2009). The value of semi-distributing the forcing data was less clear, perhaps due to the relatively small size of the Stringer Creek watershed. Some semi-distribution methods were found to subdue the "flashy" response seen when no semi-distribution was implemented, but little improvement to the overall model performance was often the case.

The uncertainty-based assessment focused primarily on the uncertainty of the parameters and the model predictive uncertainty (described by the statistical variance in the likelihood function) as propagated through the modular structures in the calculation of 95% prediction intervals. Two statistics were calculated for every model's prediction interval, the reliability and sharpness. Based on the narrowest mean prediction interval width, the Probability Distributed Model, semi-distributed by elevation with temperature and radiation index snowmelt accounting structure (*PDMsdE*_TRn) was selected as the least uncertain structure for the validation period; however, the reliability for this model was less than 95% suggesting that the prediction interval was too narrow. The structural uncertainty of the entire framework was also considered. By constructing framework structural uncertainty bounds, the value of the framework as a whole can be assessed (refer to Fig. 5).

Finally, a check on the realism of each modular structure in their simulation of snow water equivalents was performed. The results of the SWE simulations were found to be largely consistent with the results of the streamflow calibration (see Fig. 4(b)). Structures that contained the TOPMODEL component were further shown to be deficient in their simulation of snow water equivalents compared with the bucket and Probability Distributed models, providing consistency with the results of the assessment based on streamflows. Future work is underway to identify and link other key internal processes driving streamflow response in the Stinger Creek watershed to model realism.

7. Conclusions

This study introduces a fresh implementation of a snowmelt-hydrologic modular modeling framework through the identification of dominant discharge-driving processes, the development of alternate modules representing each dominant process, and assessment of all possible modular combinations for a montane watershed. Whereas many commonly implemented assessment strategies stop at comparing objective function values, the approach of considering performance, uncertainty, and realism was enlisted in this study. Further, the issue of predictive power and model structural complexity were investigated through analysis of both the AIC and BIC statistics.

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 being that (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. The usefulness of such framework approaches to hydrologic modeling lies less in the ability to determine the ultimate "best" structure, but more in the ability to understand inter- and intra-component behavior in relation to the overall structure's predicted result.

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. Finally, we wish to thank the anonymous reviewers for their comments that greatly improved the quality of the manuscript.

References

Ahl, R.S., Woods, S.W., Zuuring, H.R., 2008. Hydrologic calibration and validation of SWAT in a snow-dominated Rocky Mountain watershed, Montana, USA. Journal of the American Water Resources Association 44 (6), 1411–1430.

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.

Atkinson, S.E., Woods, R.A., Sivapalan, M., 2002. Climate and landscape controls on water balance model complexity over changing timescales. Water Resources Research 38 (12), 1314. doi:10.1029/2002WR001487.

Bai, Y., Wagener, T., Reed, P., 2009. A top-down framework for watershed model evaluation and selection under uncertainty. Environmental Modelling and Software 24 (8), 901–916.

Baker, D.B., Richards, R.P., Loftus, T.T., Kramer, J.W., 2004. A new flashiness index: characteristics and applications to midwestern rivers and streams. Journal of the American Water Resources Association 40 (2), 503–522.

Barry, R., Prévost, M., Stein, J., Plamondon, A.P., 1990. Application of a snow cover energy and mass balance model in a Balsam Fir forest. Water Resources Research 26 (5), 1079–1092.

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., 1997. TOPMODEL: a critique. Hydrological Processes 11 (9), 1069–1085.

Beven, K., Smith, P., Freer, J., 2007. Comment on "hydrological forecasting uncertainty assessment: incoherence of the GLUE methodology" by Pietro Mantovan and Ezio Todini. Journal of Hydrology 338 (3–4), 315–318.

Beven, K.J., Kirkby, M.J., 1979. A physically based, variable contributing area model of basin hydrology. Hydrological Sciences Bulletin 24 (1), 43–69.

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

Blöschl, G., Kirnbauer, R., Gutknecht, D., 1991. Distributed snowmelt simulations in an Alpine catchment 1. Model evaluation on the basis of snow cover patterns. Water Resources Research 27 (12), 3171–3179.

Boggild, C.E., Knudby, C.J., Knudsen, M.B., Starzer, W., 1999. Snowmelt and runoff modelling of an Arctic hydrological basin in west Greenland. Hydrological Processes 13 (12–13), 1989–2002.

Braithwaite, R.J., 1995. Positive degree-day factors for ablation on the Greenland ice sheet studied by energy-balance modelling. Journal of Glaciology 41 (137), 153–160.

Braun, L.N., Brun, E., Durand, Y., Martin, E., Tourasse, P., 1994. Simulation of discharge using different methods of meteorological data distribution, basin discretization and snow modelling. Nordic Hydrology 25 (1–2), 129–144.

Brooks, S.P., Gelman, A., 1998. General methods for monitoring convergence of iterative simulations. Journal of Computational and Graphical Statistics 7 (4), 434–455.
Brooks, S.P., Roberts, G.O., 1998. Convergence assessment techniques for Markov chain Monte Carlo. Statistics and Computing 8 (4), 319–335.
Brubaker, K., Rango, A., Kustas, W., 1996. Incorporating radiation inputs into the snowmelt runoff model. Hydrological Processes 10 (10), 1329–1343.
Cazorzi, F., Fontana, G.D., 1996. Snowmelt modelling by combining air temperature and a distributed radiation index. Journal of Hydrology 181 (1–4), 169–187.
Charbonneau, R., Lardeau, J.P., Obled, C., 1981. Problems of modelling a high mountainous drainage basin with predominant snow yields. Hydrological Sciences Bulletin 26 (4), 345–361.
Chiang, S., Tachikawa, Y., Takara, K., 2007. Hydrological model performance comparison through uncertainty recognition and quantification. Hydrological Processes 21 (9), 1179–1195.
Clark, M.P., Slater, A.G., Rupp, D.E., Woods, R.A., Vrugt, J.A., Gupta, H.V., Wagener, T., Hay, L.E., 2008. Framework for understanding structural errors (FUSE): a modular framework to diagnose differences between hydrological models. Water Resources Research 44, W00B02. doi:10.1029/2007WR006735.
Daly, C., Neilson, R.P., Phillips, D.L., 1994. A statistical-topographic model for mapping climatological precipitation over mountainous terrain. Journal of Applied Meteorology 33 (2), 140–158.
Daly, S.F., Davis, R., Ochs, E., Pangburn, T., 2000. An approach to spatially distributed snow modelling of the Sacramento and San Joaquin basins, California. Hydrological Processes 14 (18), 3257–3271.
Donnelly-Makowecki, L.M., Moore, R.D., 1999. Hierarchical testing of three rainfall-runoff models in small forested catchments. Journal of Hydrology 219 (3–4), 136–152.
Farmer, D., Sivapalan, M., Jothityangkoon, C., 2003. Climate, soil, and vegetation controls upon the variability of water balance in temperate and semiarid landscapes: downward approach to water balance analysis. Water Resources Research 39 (2), 1035. doi:10.1029/2001WR000328.
Fenicia, F., Savenije, H.H., Matgen, P., Pfister, L., 2008. Understanding catchment behavior through stepwise model concept improvement. Water Resources Research 44, W01402. doi:10.1029/2006WR005563.
Fontaine, T.A., Cruickshank, T.S., Arnold, J.G., Hotchkiss, R.H., 2002. Development of a snowfall-snowmelt routine for mountainous terrain for the soil water assessment tool (SWAT). Journal of Hydrology 262 (1–4), 209–223.
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.
Gelman, A., Rubin, D.B., 1992. Inference from iterative simulation using multiple sequences. Statistical Science 7 (4), 457–472.
Gilks, W.R., Richardson, S., Spiegelhalter, D.J. (Eds.), 1996. Markov Chain Monte Carlo in Practice. Chapman & Hall, London.
Grayson, R.B., Moore, I.D., McMahon, T.A., 1992. Physically based hydrologic modeling: 2. Is the concept realistic? Water Resources Research 28 (10), 2659–2666.
Haario, H., Laine, M., Mira, A., Saksman, E., 2006. DRAM: efficient adaptive MCMC. Statistics and Computing 16 (4), 339–354.
Hartman, M.D., Baron, J.S., Lammers, R.B., Cline, D.W., Band, L.E., Liston, G.E., Tague, C., 1999. Simulations of snow distribution and hydrology in a mountain basin. Water Resources Research 35 (5), 1587–1603.
Hock, R., 1999. A distributed temperature-index ice- and snowmelt model including potential direct solar radiation. Journal of Glaciology 45 (149), 101–111.
Hock, R., 2003. Temperature index melt modelling in mountain areas. Journal of Hydrology 282 (1–4), 104–115.
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.
Jothityangkoon, C., Sivapalan, M., Farmer, D.L., 2001. Process controls of water balance variability in a large semi-arid catchment: downward approach to hydrological model development. Journal of Hydrology 254 (1–4), 174–198.
Kirkby, M.J., et al., 1975. Hydrograph modelling strategies. In: Peel, R. (Ed.), Process in Physical and Human Geography. Heinemann, London, pp. 69–90.
Klemeš, V., 1983. Conceptualization and scale in hydrology. Journal of Hydrology 65 (1–3), 1–23.
Klemeš, V., 1986. Operational testing of hydrological simulation models. Hydrological Sciences Journal 31 (1), 13–24.
Kuczera, G., Parent, E., 1998. Monte Carlo assessment of parameter uncertainty in conceptual catchment models: the Metropolis algorithm. Journal of Hydrology 211 (1–4), 69–85.
Kustas, W.P., Rango, A., Uijlenhoet, R., 1994. A simple energy budget algorithm for the snowmelt runoff model. Water Resources Research 30 (5), 1515–1527.
Lindström, G., Johansson, B., Persson, M., Gardelin, M., Bergström, S., 1997. Development and test of the distributed HBV-96 hydrological model. Journal of Hydrology 201 (1–4), 272–288.
Liston, G.E., Elder, K., 2006. A meteorological distribution system for high-resolution terrestrial modeling (MicroMet). Journal of Hydrometeorology 7 (2), 217–234.
Littlewood, I.G., Croke, B.F.W., Jakeman, A.J., Sivapalan, M., 2003. The role of 'top-down' modelling for prediction in ungauged basins (PUB). Hydrological Processes 17 (8), 1673–1679.
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.
Marks, D., Dozier, J., 1992. Climate and energy exchange at the snow surface in the Alpine region of the Sierra Nevada 2. Snow cover energy balance. Water Resources Research 28 (11), 3043–3054.
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.
Martinec, J., Rango, A., 1986. Parameter values for snowmelt runoff modelling. Journal of Hydrology 84 (3–4), 197–219.
Metropolis, N., Rosenbluth, A.W., Rosenbluth, N., Teller, A.H., Teller, E., 1953. Equations of state calculations by fast computing machines. Journal of Chemical Physics 21 (6), 1087–1092.
Moore, R.J., 1985. The probability-distributed principle and runoff production at point and basin scales. Hydrological Sciences Journal 30 (2), 273–297.
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.
Samanta, S., Mackay, D.S., Clayton, M.K., Kruger, E.L., Ewers, B.E., 2007. Bayesian analysis for uncertainty estimation of a canopy transpiration model. Water Resources Research 43, W04424. doi:10.1029/2006WR005028.
Schwarz, G., 1978. Estimating the dimension of a model. Annals of Statistics 6 (2), 461–464.
Seibert, J., McDonnell, J.J., 2002. On the dialog between experimentalist and modeler in catchment hydrology: use of soft data for multicriteria model calibration. Water Resources Research 38 (11), 1241. doi:10.1029/2001WR000978.
Sivapalan, M., Blöschl, G., Zhang, L., Vertessy, R., 2003. Downward approach to hydrological prediction. Hydrological Processes 17 (11), 2101–2111.
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.
Thiemann, M., Trosset, M., Gupta, H.V., Sorooshian, S., 2001. Bayesian recursive parameter estimation for hydrologic models. Water Resources Research 37 (10), 2521–2535.
Tripp, D.R., Niemann, J.D., 2008. Evaluating the parameter identifiability and structural validity of a probability-distributed model for soil moisture. Journal of Hydrology 353 (1–2), 93–108.
Uhlenbrook, S., Seibert, J., Leibundgut, C., Rodhe, A., 1999. Prediction uncertainty of conceptual rainfall-runoff models caused by problems in identifying model parameters and structure. Hydrological Sciences Journal 44 (5), 779–797.
Vrugt, J.A., Gupta, H.V., Bouten, W., Sorooshian, S., 2003. A shuffled complex evolution metropolis algorithm for optimization and uncertainty assessment of hydrologic model parameters. Water Resources Research 39 (8), 1201. doi:10.1029/2002WR001642.
Wagener, T., 2003. Evaluation of catchment models. Hydrological Processes 17 (16), 3375–3378.
Wagener, T., McIntyre, N., 2005. Identification of rainfall-runoff models for operational applications. Hydrological Sciences Journal 50 (5), 735–751.
Wagener, T., Kollat, J., 2007. Numerical and visual evaluation of hydrological and environmental models using the Monte Carlo analysis toolbox. Environmental Modelling & Software 22 (7), 1021–1033.
WMO, 1986. Intercomparison of Models for Snowmelt Runoff. in Operational Hydrology Report 23. World Meteorological Organization, Geneva.
Yadav, M., Wagener, T., Gupta, H., 2007. Regionalization of constraints on expected watershed response behavior for improved predictions in ungauged basins. Advances in Water Resources 30 (8), 1756–1774.
Ye, M., Meyer, P.D., Neuman, S.P., 2008. On model selection criteria in multimodel analysis. Water Resources Research 44, W03428. doi:10.1029/2008WR006803.