A hierarchical Bayesian approach to hydrologic modeling

Tyler Smith

BIOL 504

Dec. 2, 2010

1. Introduction

Hydrologic watershed modeling has evolved over the past several decades, moving from tools used exclusively to solve engineering problems [cf., *Nash and Sutcliffe*, 1970] to frameworks which address scientific questions or hypotheses [e.g., *Weiler and McDonnell*, 2004]. A plethora of models exist today that attempt to conceptualize watershed functioning and the underlying processes that drive streamflow generation. These models can be considered explicit (but constrained) hypotheses of catchment behavior [*Andréassian et al.*, 2009] which may be tested through available watershed data.

Model types span from simple soil moisture accounting models to process-based models that attempt to characterize the physics of water movement through a catchment. The relative benefits and tradeoffs of different model types are well documented. In developing process-based models, inferences may be made about hydrologic processes that are not observable at the temporal/spatial scale of interest. However, physically-based models suffer from issues of over-parameterization and computational inefficiency. To make predictions about future watershed states there is an increasing desire to use simpler models that may take advantage of sophisticated optimization and uncertainty analysis algorithms. Despite this, it may be difficult (if not impossible) to relate conceptual model parameters to observable catchment characteristics affecting the reliability of models as they are extrapolated beyond calibration conditions.

In the following study, a conceptually simple hydrologic model is used to simulate watershed response (i.e., streamflow). However, due to limited temporal data at the study site a hierarchical Bayesian approach is implemented to take advantage of a number of neighboring sub-watersheds. Each of these sub-watersheds will be considered a unique case drawn from the larger geographic region of interest (i.e., the watershed that encompasses all of the sub-watersheds of interest in this study). The hierarchical approach seeks to exploit the case-specific data to improve our characterization of the true prior distribution on the hydrologic model parameters; noting that the prior distribution plays an important role in the resulting posterior distribution, especially under limited data.

The remainder of this paper is organized as follows: Section 2 will provide a description of the methods and materials used for this work, Section 3 will provide details on the applications and its results, and Section 4 will offer a discussion of the application and summary of conclusions.

2. Methods and Materials

2.1. Site and Data Description

The Tenderfoot Creek Experimental Forest (TCEF; 46°55′ N, 110°52’ W) is located at the headwaters of Tenderfoot Creek in the Little Belt Mountains of the Lewis and Clark National Forest in Montana, USA. TCEF was established in 1961 and is representative of the vast expanses of lodgepole pine (Pinus contorta) found east of the continental divide, encompassing an area of nearly 3,700 ha, in all. TCEF consists of multiple sub-watersheds with differing vegetation, topographic characteristics and silvicultural treatments. 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 sub-watersheds studied in this project were the Bubbling Creek, Stringer Creek, Spring Park, Sun Creek, and Upper Tenderfoot Creek sub-watersheds (Figure 1).

For this study, the hydrologic model was implemented under a hierarchical Bayesian framework across each sub-watershed. Due to the geographic proximity of each watershed, the same model forcing data was used for each (i.e., each sub-watershed had the same model inputs of precipitation and potential evapotranspiration). Time series data were selected for the period of study spanning from October 2007 through September 2008, using a daily time step. The data used for this study included case-specific streamflow observations from the flume at the outlet of each of the sub-watersheds (managed by the Rocky Mountain Research Station, USFS) and common meteorological observations from the Onion Park and Stringer Creek SNOTEL stations (managed by the NRCS). For additional information on available instrumentation within the greater TCEF, please refer to *Jencso et al.* [2009].

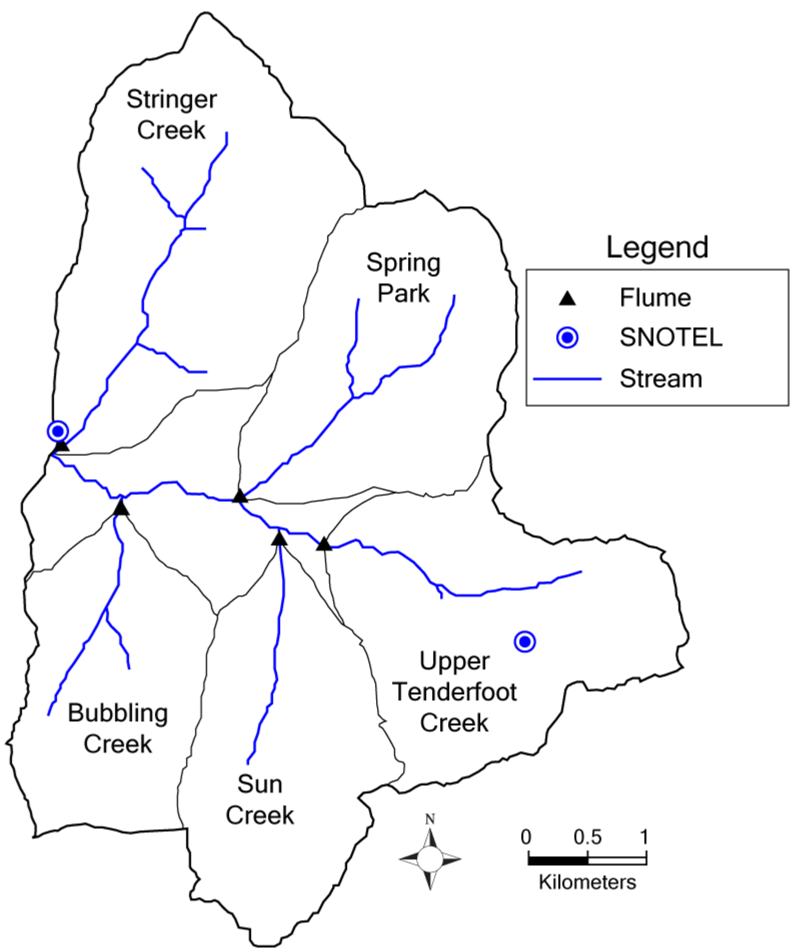

Figure 1. Tenderfoot Creek Experimental Forest with five sub-watersheds used in the study labeled, with streamflow gauging locations (flumes) and meteorological recording stations (SNOTELs) as shown.

2.2. Model Description

This study implemented a model based on the probability distributed model (PDM, Figure 2), first developed by Moore [1985]. The PDM is a conceptual rainfall-runoff model that seeks to balance model structural parsimony with watershed physical complexities. As a conceptual model, the PDM is concerned only in "the frequency of occurrence of hydrological variables of certain magnitudes over the basin without regard to the location of a particular occurrence within the basin'' [Moore, 1985, p. 274].

Soil absorption capacity controls the runoff produced by the model on the basis of the spatial variability of soil capacities across the watershed. Water in excess of the soil capacity is routed to the surface storage component, while infiltrated water eventually enters the subsurface storage component; the combination of the outflows of the two storage components comprises the total outflow of the watershed.

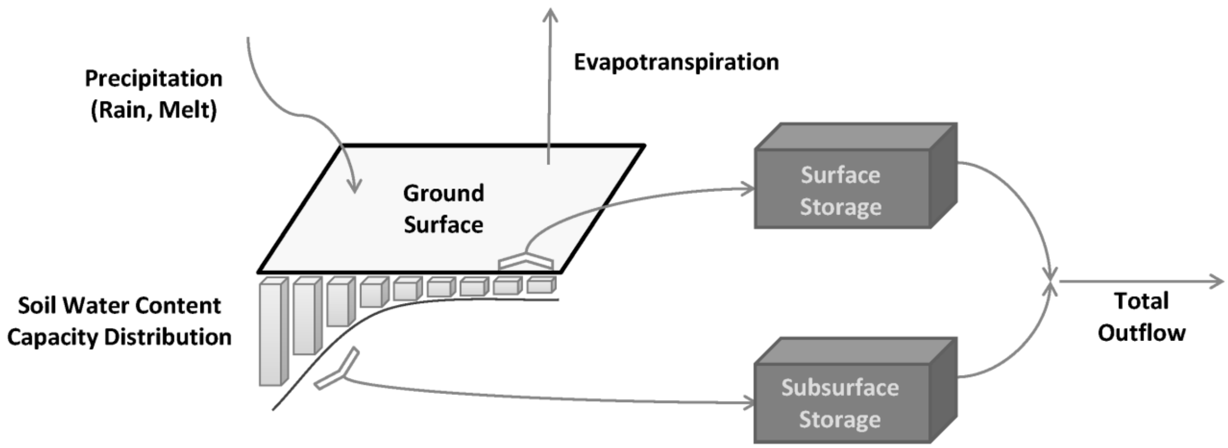

Figure 2. Schematic diagram of the probability distributed model (PDM).

The hydrologic model then is comprised of six effective parameters: maximum soil storage capacity (CMAX), spatial variability within the watershed (B), rate of drainage into subsurface storage (KB), fraction of subsurface storage released to outflow (TRES1), fraction of surface storage released to outflow (TRES2), and soil storage threshold for subsurface inflow (CF).

2.3. A Hierarchical Bayesian Implementation

Given the multiple sub-watersheds which we wish to simulate and that are considered as specific cases from the ensemble of TCEF watersheds, we aim to produce an informative/non-vague common prior (i.e., a prior that is applicable across cases) that makes use of the case specific data. Such an approach can be classified as a Bayes Empirical Bayes approach, with case specific parameters ($\underline{\theta}$) and common prior parameters (γ; "hyperparameters"). Note that while we are simulating streamflow with the hydrologic model, within our Bayesian framework we are really modeling errors (as observed streamflow minus predicted streamflow) via the likelihood function. Given this, the following notation has been adopted: let i index the cases, x_i the hydrologic forcing data for case i from the ensemble of hydrologic forcing data $\underline{x}$, and the errors for case i as ε_i from the ensemble of errors $\underline{\varepsilon}$.

The likelihood for case i is then

$$\varepsilon_i \sim \mathcal{L}_i(\varepsilon_i | \theta_i) \tag{1}$$

The likelihood function chosen for analysis (for each case) was a normal likelihood assuming the errors are Gaussian, independent, and homoscedastic. Although the likelihood function is not required to be the same for each case, it was set this way for this application.

The function relating the cases to the common prior (referred to hereafter as the "linking function") is defined such that

$$\theta \sim g(\theta|\gamma) \tag{2}$$

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

$$\gamma \sim \pi(\gamma) \tag{3}$$

which was chosen to be vague.

Given these components, the problem can be analyzed under a joint inference on all parameters and hyperparameters (and without any particular attention being paid to the specific case). In doing so, the posterior becomes

$$p(\underline{\theta}, \gamma|\underline{\varepsilon}) \propto \pi(\gamma) \prod_i \{\mathcal{L}_i(\varepsilon_i|\theta_i) \cdot g(\theta_i|\gamma)\} \tag{4}$$

where the contribution of the posterior for any case depends on the likelihood for the case specific parameters and the probability of the case specific parameters which is conditional on the common hyperparameters.

2.4. Markov chain Monte Carlo Algorithm

The adaptive Metropolis algorithm is a modification to the standard random walk, Metropolis algorithm and was used in this study to carry out the estimation of the posterior distribution. The key attribute of the AM algorithm is its continuous adaptation toward the target distribution via its calculation of the covariance of the proposal distribution using all previous states. Utilizing this attribute, the proposal distribution is updated using the information gained from the posterior distribution thus far. At step i, *Haario et al.* [2001] consider a multivariate normal proposal with mean given by the current value and covariance matrix C_i. The covariance C_i has a fixed value C_0 for the first i_0 iterations and is updated subsequently as

$$C_i = \begin{cases} C_0 & i \leq i_0 \\ s_d \cdot \mathrm{cov}(\theta_0, \cdots, \theta_{i-1}) & i > i_0 \end{cases} \tag{5}$$

where s_d is a scaling parameter depending on the dimensionality, d, of θ, the parameter set, to ensure reasonable acceptance rates of the proposed states. As a basic guideline, *Haario et al.* [2001] suggest choosing s_d for a model of a given dimension as $2.4^2/d$. An initial, arbitrary covariance, C_0, must be defined for the proposal covariance to be calculated. The steps involved in implementing the AM algorithm are discussed by *Marshall et al.* [2004].

Block updating (sampling all parameters concurrently) is utilized in the AM algorithm, enhancing the computational efficiency and reducing run time. While the AM algorithm has many beneficial traits, it can potentially experience difficulties with initialization (sampling an appropriate starting parameter set from a place of high posterior density) and in exploring the parameter space if the parameters are considerably non-Gaussian, given the proposal distribution is a multivariate Gaussian. Although the adaptive Metropolis algorithm is not a true Markov chain because of the adaptive component, results establishing the validity and the ergodic properties of the approach have been proven [*Haario et al.*, 2001].

3. Application & Results

The hierarchical model structure was run through the AM algorithm and the values of each of the parameters were estimated. Recall that for each of the five cases (sub-watersheds), the PDM model was applied to produce estimates of streamflow and required six parameters to be estimated (per case). Additionally, the likelihood function featured an unknown variance parameter that was also estimated on a per case basis. Finally, the hyperparameters associated with the common prior were also estimated; this represented an additional 13 parameters. In total, 48 parameters (30+5+13) were estimated by the algorithm under the hierarchical Bayesian framework prescribed previously.

The hyperparameters are used to define the true, but unknown common prior on each hydrologic model parameter and the likelihood function parameter. Table 1 provides details on the form of these priors, as well as physical constraints associated with each of the parameters.

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

Parameter	Value Range	Prior	Hyperparameters
CMAX	> 0 (> CF)	CMAX $\sim Gamma(A_1, A_2)$	A_1 (shape); A_2 (scale)
B	> 0	B $\sim Gamma(B_1, B_2)$	B_1 (shape); B_2 (scale)
KB	0 < KB < 1	KB $\sim Beta(C_1, C_2)$	C_1 (shape); C_2 (shape)
TRES1	0 < TRES1 < 1 (< TRES2)	TRES1 $\sim Beta(D_1, D_2)$	D_1 (shape); D_2 (shape)
TRES2	0 < TRES2 < 1 (> TRES1)	TRES2 $\sim Beta(E_1, E_2)$	E_1 (shape); E_2 (shape)
CF	> 0 (< CMAX)	CF $\sim Gamma(F_1, F_2)$	F_1 (shape); F_2 (scale)
VARP (likelihood variance)	> 0	VARP $\sim Exp(G_1)$	G_1 (rate)

An attempt to aid in the reduction of initialization issues associated with the AM algorithm was carried out, where each case was run independently (with vague priors) from random parameter starting values to identify potential regions of high posterior density (neglecting the influence of priors and parameter interactions present in the hierarchical implementation). The values derived from these pilot runs for each case were then used as the initial starting points (for the hydrologic model parameters) for the full hierarchical implementation.

The full hierarchical implementation was then set to run for several consecutive runs of 100,000 iterations (to allow progress to be checked at reasonable time intervals). Following these initial runs of 100,000 iterations, it became apparent that the initialization issues that were hoped to be avoided by the initial pilot runs were ultimately not. Recall that the AM algorithm requires an initial covariance matrix to be supplied prior to the adaptive calculation of the covariance matrix on iterations above some threshold (5% of total was used here). If this initial covariance matrix is not appropriate (i.e., there are no MCMC jumps made in the initial 5% of iterations), when the adaptive calculation of covariance begins the covariances tend to be very small (on the order of $1x10^{-10}$) as a consequence. While the jump rate was maintained at an acceptable level (≈ 25% acceptance), these jumps were occurring over extremely small parameter spaces (e.g., a parameter with a value on the order of $1x10^{2}$ was only moving in the third decimal place). Figure 3 shows a typical trace of one parameter chain (hydrologic parameter CMAX from Bubbling Creek case).

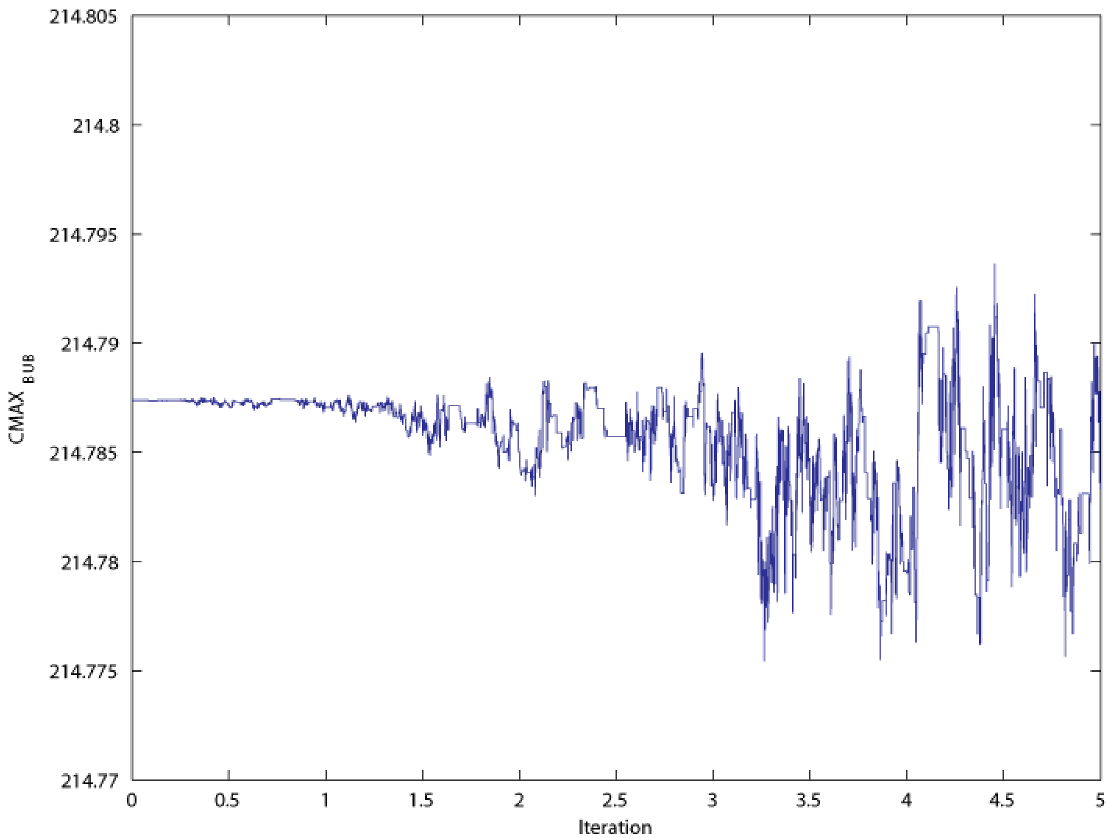

Figure 3. Plot of the parameter trace for the hydrologic parameter CMAX for the Bubbling Creek case. Note the initialization problems represented by a lack of jumping in the initial iterations and a very narrow exploration of the parameter space by successful jumps.

In all likelihood, this suggests that a few of the initial parameter variances are the cause of the initialization issues. Efforts were made to adjust the initial covariance matrix to address the initialization issues, however, the interactions and complexities posed by the number of parameters (48) has yet to be overcome. In any case, these results suggest that perhaps the AM algorithm with all parameters being sampled at once is a non-ideal technique for hierarchical models.

It was intended that the application would be setup for two concurrent runs of 500,000 iterations with starting points selected as perturbations (randomly selected within ± 20%) of the parameter set that provided the best value of the joint likelihood function from the initial 100,000 iterations runs. This step was included to test convergence of the parameter chains (albeit somewhat simplisticly). However, given the initialization issues with the shorter, exploratory runs this has yet to be done.

Under the premise that the sampling had worked (and ultimately will be working), it was intended that the relationship between the posteriors of each parameter (for each case) and the associated common prior would be highlighted by plotting the normalized histograms for each case-specific parameter on the same plot as the common prior. Such a figure would indicate the differences between the case-specific posteriors directly while also highlighting the vagueness of the common prior (i.e., is it flat and wide or peaked and narrow).

Finally, the streamflow hydrographs are presented in Figure 4. The information presented in these plots is representative of the hydrologic model predicted streamflow based on the "optimal"

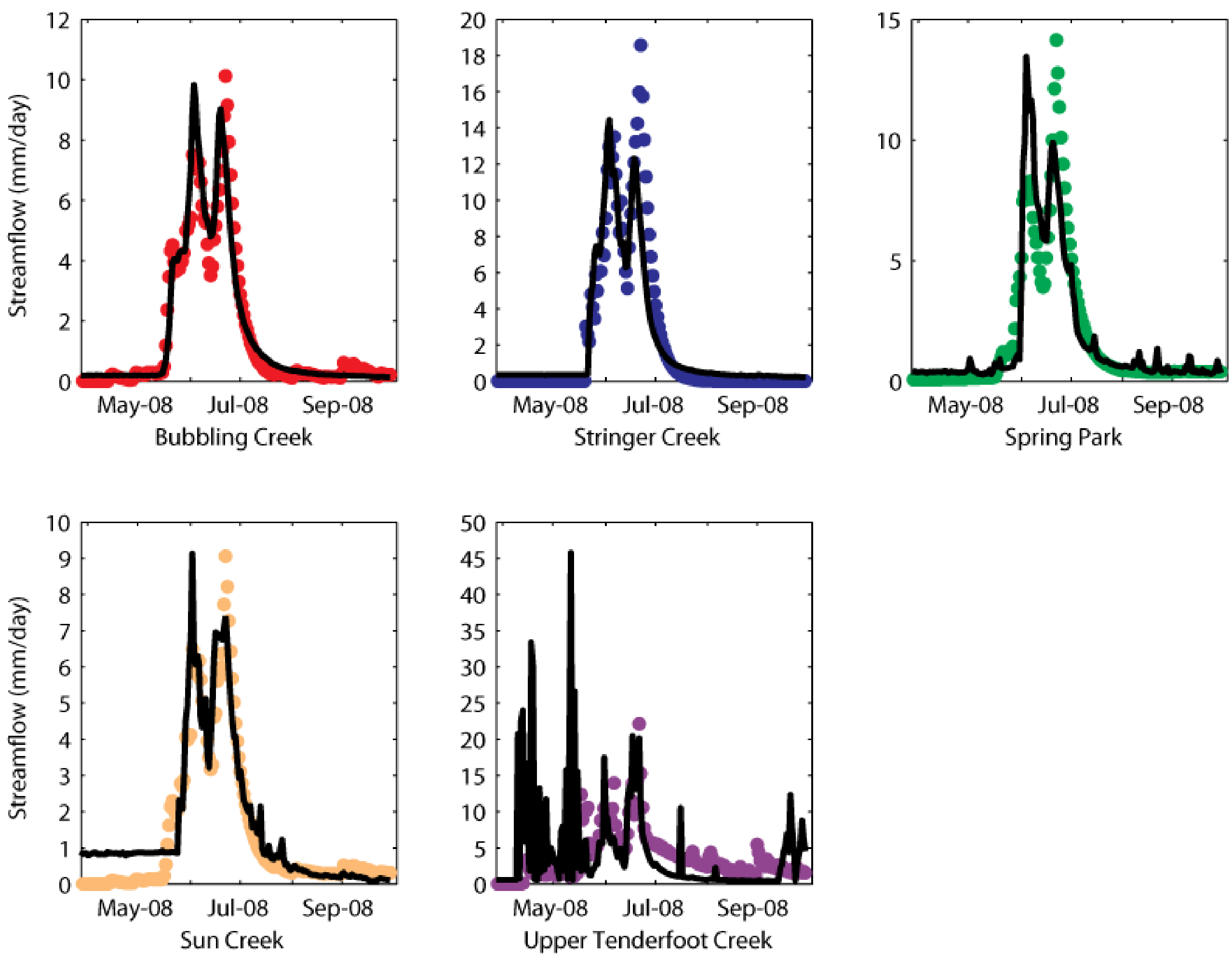

Figure 4. Plots of predicted and observed streamflow for the spring runoff period (April-September). Note the scales on the y-axis vary between subplots and that the color scheme is as follows: observed – black, Bubbling Creek – red, Stringer Creek – blue, Spring Park – green, Sun Creek – orange, and Upper Tenderfoot Creek – purple.

parameter sets. This figure reveals the differences in flow regime from sub-watershed to sub-watershed and the overall ability of the simple PDM structure to represent the observed dynamics. Note that only the spring runoff period (April-September) was used to estimate the hydrologic model parameters as this is the period of primary interest to water resources management in the region.

4. Discussion & Conclusions

While plagued with complications related to the initialization of the AM algorithm (Figure 3), the preliminary results of the hierarchical Bayesian application to a hydrologic problem featuring five sub-watersheds located within the Tenderfoot Creek Experimental Forest highlight the potential utility of such an approach under conditions where there is limited temporal data but the temporal data is available at multiple locations in space. The Bayesian logic was attempted to be carried out by the Adaptive Metropolis algorithm in order to sample from the true posterior distribution, and the parameters from the multiple locations (cases) were related to one another via the common prior distributions on each parameter.

Difficulties in obtaining acceptable traces of the parameter chains (and thus proper movement of the chains) lead to a failure to properly complete the study under the time constraints. Efforts were made to resolve the initialization issues by manually adjusting the initial covariance matrix (set as zeros on the off-diagonal elements and parameter variances on the diagonal) through trial and error. This has led to some slight improvements in sampling but not to the point of acceptability, and is difficult to do given the interactions among the 48 parameters. Other potential solutions include changing the MCMC algorithm or the manner in which the parameters are sampled (perhaps using a Gibbs within AM routine).

This technique has been shown to be useful in gathering an understanding of the functioning of a collection of cases across a variety of disciplines and was hypothesized to be equally useful in an application to neighboring sub-watersheds. Despite the obstacles encountered in this study, this approach has the potential to be useful in applications focusing on predictions in ungauged basins [PUB, *Sivapalan et al.*, 2003], where reference watersheds are typically used to inform parameter values at the watershed of interest (which lacks measured streamflow data). Under the hierarchical Bayesian framework, if a collection of reference watersheds were available, the common prior could serve as a limiting constraint on the likely values of the transferred parameters.

Future work is planned to further investigate the use of this technique under the umbrella of predictions in ungauged basins, with particular attention being paid to the selection of an appropriate MCMC algorithm capable of dealing with the high dimensionality posed by hierarchical problems. Additionally, work is underway to assemble a larger ensemble of cases with more extensive temporal data and to extend the logic from the PDM structure to more complex, hydrologic models specifically designed for the Tenderfoot Creek Experimental Forest based on extensive field investigations.

5. References

Ahl, R. S., S. W. Woods, and H. R. Zuuring (2008), Hydrologic calibration and validation of SWAT in a snow-dominated Rocky Mountain watershed, Montana, USA, *J. Am. Water Resour. Assoc.*, *44*(6), 1411-1430.

Andréassian, V., C. Perrin, L. Berthet, N. Le Moine, J. Lerat, C. Loumagne, L. Oudin, T. Mathevet, M. H. Ramos, and A. Valréy (2009), HESS Opinions "Crash tests for a standardized evaluation of hydrological models", *Hydrol. Earth Syst. Sci. Discuss.*, *6*(3), 3669-3685.

Haario, H., E. Saksman, and J. Tamminen (2001), An adaptive Metropolis algorithm, *Bernoulli*, *7*(2), 223-242.

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

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

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

Sivapalan, M., K. Takeuchi, S. W. Franks, V. K. Gupta, H. Karambiri, V. Lakshmi, X. Liang, J. J. McDonnell, E. M. Mendiondo, P. E. O'Connell, T. Oki, J. W. Pomeroy, D. Schertzer, S. Uhlenbrook, and E. Zehe (2003), IAHS Decade on Predictions in Ungauged Basins (PUB), 2003-2012: Shaping an exciting future for the hydrological sciences, *Hydrol. Sci. J.*, *48*(6), 857-880.

Weiler, M., and J. McDonnell (2004), Virtual experiments: a new approach for improving process conceptualization in hillslope hydrology, *J. Hydrol.*, *285*(1-4), 3-18.