Environmetrics Research Article

Received: 16 December 2013, Revised: 30 March 2014, Accepted: 31 March 2014, Published online in Wiley Online Library: 8 May 2014

(wileyonlinelibrary.com) DOI: 10.1002/ENV.2275

A Bayesian hierarchical model for forecasting intermountain snow dynamics

James B. Odei*, Jürgen Symanzik and Mevin B. Hooten

Because of a continual increase in the demand for water as well as an ongoing regional drought, there is an imminent need to monitor and forecast water resources in the Western United States. In particular, water resources in the Intermountain West rely heavily on snow water storage. Thus, the need to improve seasonal forecasts of snowpack and considering new techniques would allow water resources to be more effectively managed throughout the entire water-year. Many available models used in forecasting snow water equivalent (SWE) measurements require delicate calibrations. In contrast to the physical SWE models most commonly used for forecasting, we offer a statistical model. We present a data-based statistical model that characterizes seasonal SWE in terms of a nested time series, with the large scale focusing on the inter-annual periodicity of dominant signals and the small scale accommodating seasonal noise and autocorrelation. This model provides a framework for independently estimating mainly the temporal dynamics of SWE for the various snow telemetry sites. We use snow telemetry data from 10 stations in Utah over 34 water-years to implement and validate this model. Copyright © 2014 John Wiley & Sons, Ltd.

Keywords: empirical orthogonal functions; time series; water-year; snow water equivalent

1. INTRODUCTION

Snow is considered to play significant roles in climate, terrestrial biosphere, and hydrology. As an integral part of the annual water budget, during the winter snowpack, water is accumulated and then released in the spring snowmelt season (Ichii *et al.*, 2007).

The Intermountain West region of the semi-arid Western United States (WUS) composes of varying ecological and economic systems (Bailey, 1995) where over 75% of water resources result from snowmelt water. In this region, many different systems including snow courses, snow telemetry (SNOTEL), aerial markers, and airborne gamma radiation (Cowles *et al.*, 2002) are used to measure the amount of water in the snow. The term Intermountain West generally composes of six states including Utah. We use the term "intermountain" because our underlying principle can be extended to other states from this region.

A report from the National Climatic Data Center indicated that since 1998, multi-year droughts have seriously affected the supply of water in the Southwest (`http://www.wcc.nrcs.usda.gov/snotel/Utah/utah.html`), and these droughts are some of the major risks residents and ecosystems of this region are facing (Mote *et al.*, 2005). Thus, effective long-term management decisions must be made on the basis of predictions concerning accumulation and melt of intermountain snow, a major water resource for the region. Modeling can offer a useful tool to perform such prediction, and physically-based numerical computer models are broadly used to generate snow and related forecasts (Jin *et al.*, 1999). The majority of these physical models are able to produce reasonable forecasts during the snow accumulation phase, but some perform poorly during the snowmelt stage because of inadequate representation of snowmelt regimes (Jin *et al.*, 1999; Jin and Wen, 2012) and spring climate forcing (Saha *et al.*, 2006).

With the growing concern about the effects of change in climate on snowpack accumulation (Lapp *et al.*, 2005; Mote *et al.*, 2005; Mote, 2006) and societal seasonal water supply requirements (Miller *et al.*, 2003; Van Rheenen *et al.*, 2004; Brekke *et al.*, 2004; Stewart *et al.*, 2005), accurate forecasting of water supplies has become increasingly urgent. Previous hydrologic simulation models and a spatial statistical model developed by the National Weather Service (NWS) enables us to address some of these questions. The NWS hydrologic simulation model generates extended streamflow predictions, water supply and spring flood outlooks, and flood forecasts (Day, 1985; Hudlow, 1998). With the spatial model, estimation of snow water equivalent (SWE) across the WUS river basins are obtained on a 30-arc second grid (Carroll and Cressie, 1996). An important part of the NWS spatial model is that its covariance function produces the covariance between the SWE at two sites. In situations where areas have no observed measurements available, the NWS spatial statistical model uses ground-based and airborne SWE data (Day, 1990; Carroll *et al.*, 1995, 1999).

Despite some successes in statistical snow modeling, there still remains a need for accurate seasonal forecasts. For example, Seidou *et al.* (2006) used Bayesian models for the interpolation of field measurements of SWE. Cowles *et al.* (2002) presented a Bayesian

* *Correspondence to: James B. Odei, Department of Mathematics & Statistics, Utah State University, 3900 Old Main Hill, Logan, UT 84322-3900, U.S.A. E-mail: jamesbeg.odei@aggiemail.usu.edu*

Department of Mathematics & Statistics, Utah State University, 3900 Old Main Hill, Logan, UT 84322-3900, U.S.A.

spatio-temporal analysis of the combined SWE data from four sources (snow courses, SNOTEL, aerial markers, and airborne gamma radiation) that allow for systematic differences in their accuracy and reliability. Their work also focused on identifying long-term temporal trends spatially. Additionally, Leisenring and Moradkhani (2011) described a statistical approach to "calibrating" a mechanistic snow model. This model was used to produce longer range forecasts, but it did not allow that some of the current water-year had already occurred. There are also alternative approaches for estimating SWE when it is not measured directly, for example, by linking it to auxiliary data sources such as snow depth and bulk snow density (Jonas *et al.*, 2009).

Given the need for improved forecasts, we offer a new modeling direction, which is complimentary to other efforts. Some of the benefits of statistical models are as follows: they are simple, very efficient in computation, and easy to use. Most importantly, statistically-based snow models rely heavily on observational data, which greatly increase their reliability in forecasting. All these benefits are missing in physical models. In this article, we present a nested time-series statistical model for SWE data arising from the SNOTEL monitoring stations. Our objective here is to use an empirical Bayesian hierarchical statistical model to analyze our economically important variable: distribution of SWE. Our hierarchical model incorporates both spatial structure and measurement uncertainty.

Section 2 provides an overview on how the SWE data were obtained, dimension reduction, and a presentation on the formulation of the hierarchical model. In Section 3, we look at results after implementing the model from Section 2. We briefly discuss some important general assessments in Section 4 and state our conclusions in Section 5. Appendix A provides an overview of the R (R Core Team, 2012) code used for the modeling. The Appendix B provides the derivations of the full conditional distributions in this article.

2. MATERIALS AND METHODS

2.1. Snow telemetry data sources

In this article, we provide a proof-of-concept application of our model to mountain ranges in the state of Utah. This state has active SNOTEL monitoring stations that provide various snow and environmental measurements (i.e., SWE, air temperature, precipitation, snow depth, and soil moisture/temperature). In this article, 10 SNOTEL stations, each with 34 water-years (1979–2012) and 365 days per water-year yielding 124,100 ($34 \times 10 \times 365$) total SWE observations, were used. For simplicity, February 29 in leap years was ignored. These SNOTEL systems, installed, operated, and maintained by the Natural Resources Conservation Service, are designed to collect snowpack and other related climatic data in the WUS and Alaska. An illustration of the state-wide active SNOTEL sites of Utah is shown in Figure 1.

In near real time, SNOTEL uses meteor burst communications technology to communicate data where very high frequency radio signals are reflected at a steep angle off the band of ionized meteors existing above the earth surface from about 50 to 75 miles (NRCS, 2009).

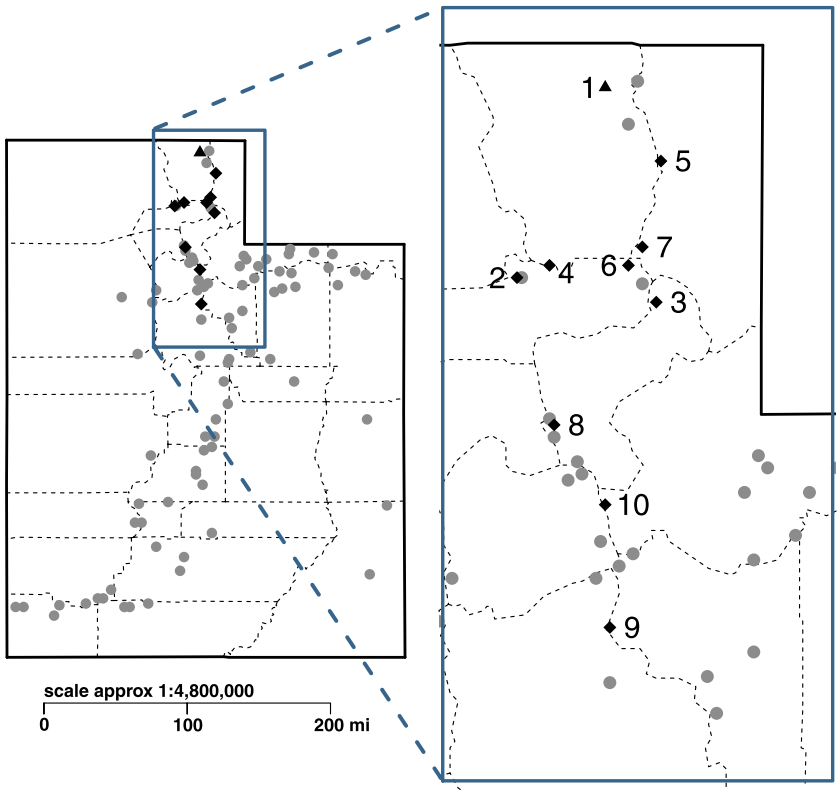

Figure 1. Locations of 90 active snow telemetry (SNOTEL) sites in Utah. The black-shaded diamonds and the triangle are the 10 sites with at least 30 years of SNOTEL measurements selected for model validation. The dashed lines represent county boundaries. The sites are as follows: (1) Tony Grove, (2) Ben Lomond Peak, (3) Horse Ridge, (4) Little Bear, (5) Bug Lake, (6) Dry Bread Pond, (7) Monte Cristo, (8) Farmington, (9) Timpanogos Divide, and (10) Parley's Summit. The Tony Grove SNOTEL site is further discussed in Section 2.1

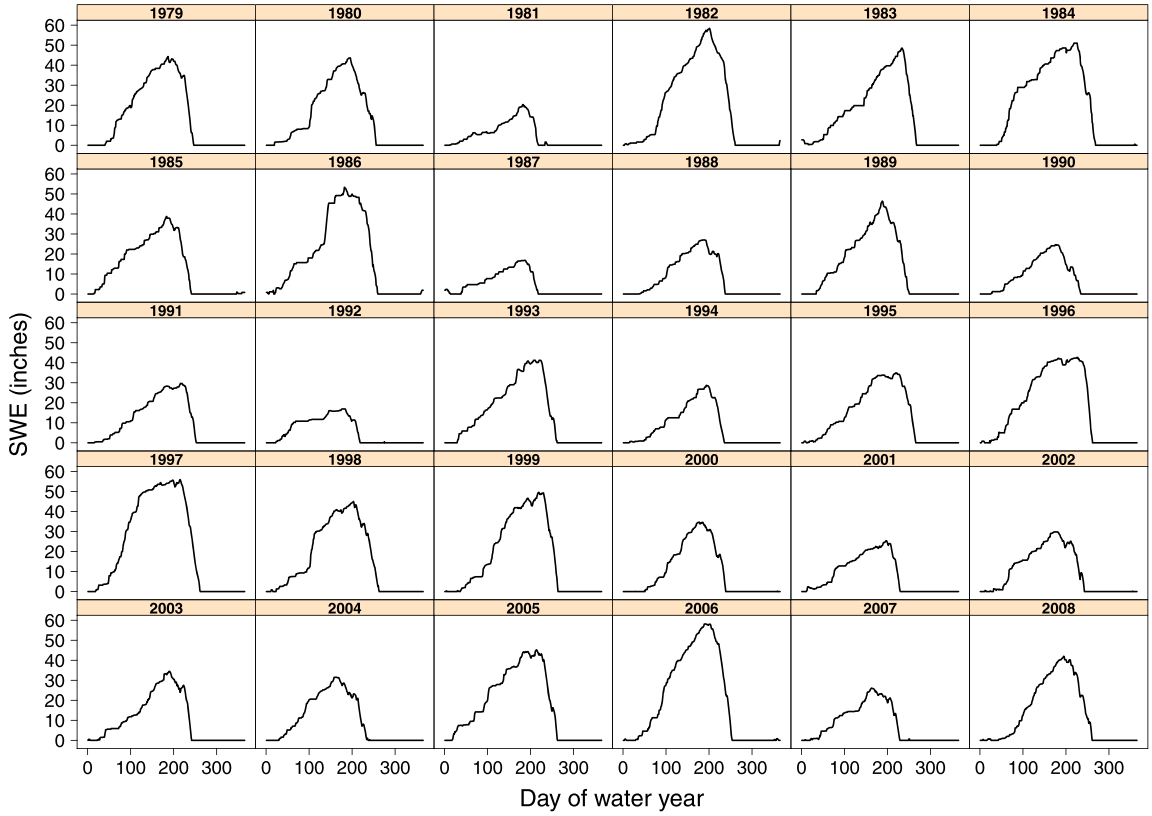

Figure 2. Plots of snow water equivalent measurements (in inches) for the Tony Grove site (1979–2008). Each of these plots represents the amount of water stored during the water-year

Generally, these SNOTEL sites are located remotely on high watersheds where accessibility is difficult or restricted. For maintenance purposes, access involves the use of helicopters, skiing, snowmobiles, hiking, and snowshoeing. Overall, there are six Natural Resources Conservation Service data collection offices that monitor daily site statistics. When data are received, they are converted to engineering units and screened properly for errors and then stored in a database. The data (and additional information) are made available to the public through the National Water and Climate Center website (`http://www.wcc.nrcs.usda.gov`).

Figure 2 shows plots of SWE measurements for the Tony Grove SNOTEL site from 1979 to 2008. The time of SWE measurements is in days of a water-year (1 October–30 September). For example, in the 1979 water-year, SWE measurements were taken between 1 October 1978 and 30 September 1979. Years with large (e.g., 1997) and small (e.g., 1981 and 1987) measurements of overall SWE are readily apparent in Figure 2. Also, there were years (e.g., 1982 and 2006) with relatively high SWE measurements early in the water-year and sharp decreases by the middle of the water-year. In addition to these observations, there are other factors or important (dominant) signals underlying the SWE measurements at the SNOTEL sites that cannot be seen directly but nonetheless could be modeled. We will discuss in Section 2.2 how we can capture the important signals underlying the SWE measurements.

2.2. Dimension reduction

Spatial and spatio-temporal data and processes of interest are naturally high dimensional, and when considered in a statistical context, it is common for model specifications to have a large number of parameters and state variables. A popular approach to dealing with high-dimensional parameter spaces in spatial and spatio-temporal models (e.g., Shumway and Stoffer, 2006, state-space models) is to project the process onto a lower-dimensional manifold for further modeling. This technique, sometimes referred to as fixed-rank or low-rank approximation (Wikle, 2010, Ch. 8) or fixed rank kriging in geostatistics (Cressie and Johannesson, 2008; Shi and Cressie, 2007), can be readily achieved through the use of a transformation involving a set of basis functions that map the latent lower-dimensional dynamic process to the scale at which the data are collected. It should be noted that in theory, dimension reduction procedures reveal a set of latent, underlying processes. The data can then be fully utilized to learn about an underlying process that is governed by a parsimonious set of parameters. Many forms of basis functions are possible (e.g., wavelets, Fourier, splines, and empirical orthogonal functions (EOFs)), and their utility varies depending on the goals of the study.

The predictor variables (e.g., SWE on a given date at several snowcourses) used in operational water supply forecasting are typically correlated. A satisfactory and statistically rigorous manner to deal with these inter-correlations is the use of principal component analysis (Garen, 1992). In the spatio-temporal literature, principal components are commonly used for such dimension reduction and are referred to as EOFs (Bjornsson and Venegas, 1997). The two most common ways to obtain these basis functions are to decompose the space-time

covariance matrix spectrally (eigen decomposition) (Golub and van Loan, 1996) or to use the singular value decomposition (SVD) (Golub and Reinsch, 1970) on a matrix of the data directly. In this article, we take the latter approach.

On the basis of the SNOTEL measurements, we obtain our SWE data for site s in matrix form, say $\mathbf{Y}_{T\times M} = [\mathbf{y}_{s1}, \mathbf{y}_{s2}, ..., \mathbf{y}_{sT}]'$, where $\mathbf{y}_{st}$ is of dimension 365×1 and represents SWE measurements at site s in year t for all 365 days in a water-year, $M = 365$ days and T is determined by prior water-years. For example, for 2008 water-year, we only use data from 1979 through 2007 (29 years), for 2009 water-year, we use data from 1979 through 2008 (30 years), and so. So, $T = 29$ for 2008 water-year, 30 for 2009 water-year, and so on. The sample average of SWE at site s for the past T years for day d, $\bar{y}_{..d}$ is $\frac{1}{T}\sum_{t=1}^{T} y_{std}$ where the scalar y_{std} represents the observed SWE at site s in year t for day d. The obtained SWE sample average, $\bar{y}_{..d}$, is then repeated at each site s for day d in all the T water-years to form a matrix $\bar{\mathbf{y}}$ that has the same dimension as $\mathbf{Y}$ ($T \times M$) with each row made up of the average SWE observed for each day at site s for T water-years. Thus, $\bar{\mathbf{y}}_{T\times M} = [\bar{\mathbf{y}}_{..1}, \bar{\mathbf{y}}_{..2}, ..., \bar{\mathbf{y}}_{..d}, ..., \bar{\mathbf{y}}_{..365}]$, where $\bar{\mathbf{y}}_{..d}$ is of dimension $T \times 1$ and contains T repetitions of the obtained SWE average, $\bar{y}_{..d}$, for day d.

Applying SVD to our SNOTEL data, we decompose the SWE data after removing the day-of-year sample mean for site s independently as

$$\mathbf{Y} - \bar{\mathbf{y}} = \tilde{\mathbf{U}}\tilde{\mathbf{D}}\tilde{\mathbf{V}}' \approx \mathbf{U}\mathbf{D}\mathbf{V}' \tag{1}$$

The approximation in Equation (1) is the mechanism for reducing dimensionality in the parameter space. The $\mathbf{U}\mathbf{D}\mathbf{V}'$ representation on the right hand side of Equation (1) is a truncation of the $\tilde{\mathbf{U}}\tilde{\mathbf{D}}\tilde{\mathbf{V}}'$ decomposition based on the first q important signals (EOFs). It should be noted here that q is much smaller than M. The matrix of orthonormal vectors $\mathbf{V}$ (first q columns of $\tilde{\mathbf{V}}$) is of dimension $M \times q$ and contains the important signals (EOFs). The matrix of left singular vectors, $\mathbf{U}$ (first q columns of $\tilde{\mathbf{U}}$), has dimension $T \times q$ and contains the latent times series. The singular values are contained in $\mathbf{D}$, a diagonal matrix of dimension $q \times q$ with the eigenvalues on the diagonal that is obtained from the first q eigenvalues (on the diagonal) of $\tilde{\mathbf{D}}$.

Because $\mathbf{U}$ contains the set of time series corresponding to each signal in $\mathbf{V}$ for site s,

$$\mathbf{U} = [\mathbf{u}_{s1}, \mathbf{u}_{s2}, ..., \mathbf{u}_{sT}]'$$

where each $\mathbf{u}_{st}$ is of dimension $q \times 1$, we can obtain approximations of the SWE for a particular year t, at site s, $\mathbf{y}_{st}$ of dimension $M \times 1$ by adding the day-of-year sample mean to the product of $\mathbf{V}$, $\mathbf{D}$, and the time series corresponding to that selected year as

$$\mathbf{y}_{st} \approx \bar{\mathbf{y}}_{..} + \mathbf{V}\mathbf{D}\mathbf{u}_{st} \tag{2}$$

where $\bar{\mathbf{y}}_{..}$ (also of dimension $M \times 1$) consists of the averages of SWE measured for each day in the water-year at site s for the past T years. That is, $\bar{\mathbf{y}}_{..} = [\bar{y}_{..1}, \bar{y}_{..2}, ..., \bar{y}_{..d}, ..., \bar{y}_{..M}]'$. There may be an underlying physical process that may operate on a lower-dimensional manifold. If we can parse the model into a data component that accounts for additional noise that cannot be explained and a process component represented by an underlying latent dynamic process, then we have defined a hierarchical model. For more details on this basic approach and its strengths, see Wikle and Hooten (2010). Thus, this general equation (2) provides motivation for a hierarchical statistical model.

2.3. Nested time-series models

In a more general sense, by considering the temporal dependence of SWE among the SNOTEL sites, we can specify a model for all of the SWE data by partitioning the temporal structure into a large and a small temporal resolution. The specifications of the large and small temporal resolutions are carried out using vector autoregressive and conditional autoregressive (CAR) models, respectively. Thus, a nested time-series model with a spatio-temporal structure results.

2.3.1. The Bayesian hierarchical statistical model

A general spatio-temporal statistical model was introduced in Odei *et al.* (2009). In this article, we focus on a simplified version of the general spatio-temporal statistical model that treats the SNOTEL sites independently. Using a Bayesian approach, we propose a hierarchical statistical model for SWE. Letting the dynamics of SWE for any site s be represented by a dimensionally reduced vector autoregression, for prediction (forecasting), we model $\mathbf{y}_{st}$ as follows:

$$\mathbf{y}_{st} = \bar{\mathbf{y}}_{s.} + \mathbf{V}\mathbf{D}A_{st}\mathbf{u}_{s,t-1} + \boldsymbol{\varepsilon}_{st} \tag{3}$$

with $\boldsymbol{\varepsilon}_{st} \sim N(\mathbf{0}, \boldsymbol{\Sigma}_{\mathbf{y}})$, $t = 1, 2, ..., T$, and where $\bar{\mathbf{y}}_{s.}$ (of dimension $M \times 1$) consists of the averages of SWE measured for each day in the water-year at site s for the past T years. Thus, we have $\bar{\mathbf{y}}_{s.} = [\bar{y}_{s.1}, \bar{y}_{s.2}, ..., \bar{y}_{s.d}, ..., \bar{y}_{s.M}]'$. The covariance structure, $\boldsymbol{\Sigma}_{\mathbf{y}}$, handles the small-scale temporal (daily) SWE correlation, which allows for latent autocorrelation in the days of the water-year. The latent time series for site s in the tth year, $\mathbf{u}_{st}$, in Equation (2) is postulated as the product of a propagation matrix, A_{st} and $\mathbf{u}_{s,t-1}$, the latent time series in year $t-1$ for site s. The propagation matrix A_{st} of dimension $q \times q$ handles the inter-annual scale temporal SWE correlation. This matrix could be specified in various ways depending on the desired amount of generality in the dynamics; it could also be linked to temporally varying covariates such as Pacific Decadal Oscillation and El Niño/Southern Oscillation. Here, we treat $\mathbf{y}_{st}$ as a joint multivariate normal distribution, which would be discussed further in Section 2.3.2.

The remaining parameters in our hierarchical model are defined as follows:

$$
\begin{aligned}
A_{st} &= \text{diag}(\boldsymbol{\alpha}_{st}) \\
\boldsymbol{\alpha}_{st} &\sim N\left(\boldsymbol{\alpha}, \sigma_{\alpha}^2 \mathbf{I}\right) \\
\boldsymbol{\alpha} &\sim N\left(\mathbf{0}, \sigma_{o}^2 \mathbf{I}\right) \\
\boldsymbol{\Sigma}_{\mathbf{y}} &= \sigma_{\mathbf{y}}^2 \left(\mathbf{I} - \rho \mathbf{W}\right)^{-1}
\end{aligned}
$$

The CAR model specification we used for $\mathbf{W}$ and ρ is standard in both the spatial and time-series statistical literature (Schabenberger and Gotway, 2005). $\mathbf{W}$ itself in our case is fixed and thus does not have a prior. It merely provides a feeling for which time points are neighbors of which other time points. $\mathbf{W}$ is an $M \times M$ temporal proximity matrix that allows for structure between neighboring days. The elements of the matrix $\mathbf{W}$, w_{kd} $(k, d = 1, 2, ..., M)$ are $1/2$ if the absolute difference between the kth and dth days is one and it is 0 otherwise. Furthermore, recent literature in Markov random field modeling suggests that small ρ parameters are not very useful in describing dependence structures. Thus, the recommendation has been to use either a very strong prior on ρ that puts it close to 1 or simply remove it altogether (Banerjee *et al.*, 2004). The latter choice yields an intrinsic CAR (ICAR) model that is technically improper (in the probability sense) because it does not integrate to one. However, we alleviate this problem by informing ρ to be less than one but very large (Banerjee *et al.*, 2004). Thus, ρ is estimated by sampling from a truncated normal distribution with mean and standard deviation of 0.99 and 0.01, respectively, on the interval $[-1, 1]$.

The variance component $\sigma_{\mathbf{y}}^2$ was assumed to have an inverse gamma distribution $\sigma_{\mathbf{y}}^2 \sim IG(\nu_1, \nu_2)$ with initial values of the scale (ν_1) and shape (ν_2) parameters as 3 and 0.05, respectively. Here, the autoregression coefficients $\boldsymbol{\alpha}_{st}$, can be considered mixed effects that depend on the global coefficients, $\boldsymbol{\alpha}$. In order to make our priors on the $\boldsymbol{\alpha}_{st}$ and $\boldsymbol{\alpha}$ parameters vague, we specified the initial values of σ_{α}^2 and σ_{o}^2 as 10 and 100, respectively. With these types of models, variances are notoriously sensitive to prior specifications (Gelman, 2006). Thus, how well the model behaves would depend on good choices of initial values for the variances. In this article, we chose values that were relatively diffused (spread out) and conservative for what we thought the variances might likely be. Further detailed descriptions and applications of how a hierarchical model works can be found in Davis and Seaman (2002), Suess *et al.* (2002), and Wikle *et al.* (2001).

2.3.2. *Implementation*

As part of our Bayesian hierarchical statistical modeling, given the SWE measurements for any site s, $\mathbf{y}_{st}$, we obtain the posterior distribution as proportional to the product of the likelihood of the data given the latent process models and parameter models as

$$
\left[\{\boldsymbol{\alpha}_{st}\}, \boldsymbol{\alpha}, \sigma_{\mathbf{y}}^2, \rho | \{\mathbf{z}_{st}\}\right] \propto \left(\prod_{t=1}^{T} \left[\mathbf{z}_{st} | \boldsymbol{\alpha}_{st}, \sigma_{\mathbf{y}}^2\right] \times \left[\boldsymbol{\alpha}_{st} | \boldsymbol{\alpha}\right]\right) \times [\boldsymbol{\alpha}] \times \left[\sigma_{\mathbf{y}}^2\right] \times [\rho] \tag{4}
$$

where $\mathbf{z}_{st} = \mathbf{y}_{st} - \bar{\mathbf{y}}_{s.}$. The bracket notation [.] denotes the probability distribution of the argument contained within the brackets. We also obtain the full conditional distributions of $\boldsymbol{\alpha}_{st}$, $\boldsymbol{\alpha}$, $\sigma_{\mathbf{y}}^2$, which are straightforward to sample from, and they depend on the centered response variable $\mathbf{z}_{st}$. The full conditional distribution for both $\boldsymbol{\alpha}_{st}$ and $\boldsymbol{\alpha}$ is multivariate normal and that of $\sigma_{\mathbf{y}}^2$ is inverse gamma (Appendix B).

Analytically, the posterior obtained in Equation (4) is not tractable. Thus, with the estimate of the parameter ρ, we then implemented the model using a hybrid Metropolis–Hastings and Gibbs Markov Chain Monte Carlo (MCMC) sampling algorithm (Banerjee *et al.*, 2004; Carlin and Lewis, 2009) using the R software (R Core Team, 2012). The R packages used as part of the model implementation include MCMCpack (Martin *et al.*, 2011), mvtnorm (Genz and Bretz, 2009; Genz *et al.*, 2011), msm (Jackson, 2011), lattice (Sarkar, 2008), Mass (Venables and Ripley, 2002), and spdep (Pebesma and Bivand, 2005; Bivand *et al.*, 2008).

The MCMC algorithm was run for 5000 iterations with burn-in periods of 1250 iterations. Convergence, which was assessed visually, occurred rapidly and the posterior distribution appeared well characterized. Summaries of the post-convergence MCMC samples provide posterior inference for model parameters. The resulting MCMC samples were used to calculate summary statistics for all the latent processes and model parameters $(\boldsymbol{\alpha}, \boldsymbol{\alpha}_{st}, \boldsymbol{\Sigma}_{\mathbf{y}})$. Many statistical methods (e.g., simple and multiple linear regression models, and logistic regression models) provide confidence intervals and can (in cross-validation mode) provide prediction intervals. The advantage of the simulation approach is that it provides estimates of the entire posterior distribution that permits us to obtain interval estimates as well. For example, the 2.5th and the 97.5th quantiles of the (post-convergence) sampled values for a model parameter provide a 95% interval estimate of the parameter. In the Bayesian framework, such an interval is termed credible interval or credible set to differentiate it from the confidence interval in frequentist statistics (Schabenberger and Gotway, 2005). Unlike confidence intervals, credible intervals have a direct probabilistic interpretation: a 95% credible interval defines an interval having a 0.95 posterior probability of containing the parameter of interest.

For the purpose of forecasting, we consider measurements of SWE for a water-year at site s, $\mathbf{y}_{st}$ (for $t \in \{2, 3, ..., T\}$) partitioned into observed (*obs.*) and unobserved (*unobs.*) SWE measurements. Thus, we can think of both as a joint multivariate normal random vector

$$
\begin{aligned}
\mathbf{y}_{st} &= \begin{bmatrix} \mathbf{y}_{st,obs.} \\ \mathbf{y}_{st,unobs.} \end{bmatrix} \\
&\sim N\left(\begin{bmatrix} \bar{\mathbf{y}}_{st-1.,obs.} \\ \bar{\mathbf{y}}_{st-1.,unobs.} \end{bmatrix}, \begin{bmatrix} \boldsymbol{\Sigma}_{obs.,obs.} & \boldsymbol{\Sigma}_{obs.,unobs.} \\ \boldsymbol{\Sigma}_{unobs.,obs.} & \boldsymbol{\Sigma}_{unobs.,unobs.} \end{bmatrix}\right)
\end{aligned}
$$

where for forecasting after day d, $\bar{\mathbf{y}}_{st-1.,obs.}$ is the average SWE over $t-1$ previous years corresponding to days 1 to d and $\bar{\mathbf{y}}_{st-1.,unobs.}$ is the average SWE over $t-1$ previous years corresponding to days $d+1$ to $M = 365$. Then, in the MCMC algorithm, we can use composition sampling (a technique for computing an integral) (Gelman *et al.*, 2003; Carlin and Lewis, 2009) to obtain samples from the posterior predictive distribution for the unobserved SWE as

$$\mathbf{y}_{st,unobs.}|\mathbf{y}_{st,obs.} \sim N\left(\boldsymbol{\mu}^*, \boldsymbol{\Sigma}^*\right)$$

where we use multivariate normal results to obtain the conditional mean and variance

$$\boldsymbol{\mu}^* = \bar{\mathbf{y}}_{st-1.,unobs.} + \boldsymbol{\Sigma}_{unobs.,obs.}\boldsymbol{\Sigma}^{-1}_{obs.,obs.}\left(\mathbf{y}_{st,obs.} - \bar{\mathbf{y}}_{st-1.,obs.}\right)$$

and

$$\boldsymbol{\Sigma}^* = \boldsymbol{\Sigma}_{unobs.,unobs.} - \boldsymbol{\Sigma}_{unobs.,obs.}\boldsymbol{\Sigma}^{-1}_{obs.,unobs.}\boldsymbol{\Sigma}_{obs.,unobs.}$$

2.3.3. Model validation

For the purpose of validation, the model was implemented using different forecast initiation dates/days in a water-year and selected SNOTEL sites with at least 34 years of data for our posterior predictions. The initiation dates/days were chosen such that they were 30 days apart with the first initiation day being 8 January (when we have a reasonable amount of SWE data for meaningful predictions). Posterior predictive credible intervals were also obtained for the "unobserved" hold-out SWE data. Here, the hold-out SWE data are the data obtained by splitting the original SWE data into two groups, the training and the testing sets. The choice of the train/test split was carried out using the day d after which the forecasting would be made. Thus, SWE measurements from day 1 to day d (training set) in the current water-year are considered observed, and SWE measurements from day $d+1$ to day $M = 365$ (testing set) in the current water-year are considered as unobserved when the model is implemented. The SVD is calculated on $t-1$ previous years of SWE data prior to the forecast in the current water-year.

The probability skill of the SWE forecasts is assessed by means of the ranked probability skill score (RPSS; Wilks (1995)). The RPSS is used to evaluate a model's skill in capturing categorical probabilities relative to climatology. The SWE values were divided into three categories: values below the 25th percentile, above the 75th percentile, and the rest between the 25th and 75th percentiles. The RPSS ranges from $-\infty$ to 1.0; the highest value indicating a perfect forecast and a negative value implying that the forecast has lesser accuracy as compared with that of climatology.

3. RESULTS

Our Bayesian hierarchical model presented in Section 2.3 was fitted to the SWE data described in Section 2.1. SVD was performed on the data matrix in Equation (1), the scaled SWE data ($\mathbf{Y} - \bar{\mathbf{y}}$), and truncated to obtain the important seasonal signals and inter-annual time series.

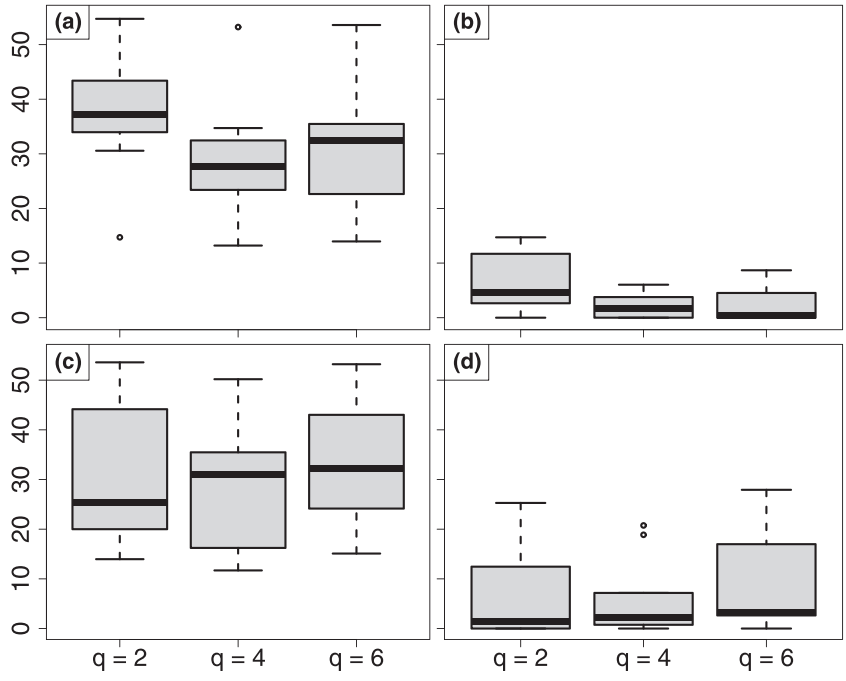

Figure 3. Boxplots for proportion of snow water equivalent that fell outside the 50% (a and c) and 95% (b and d) credible intervals of the posterior predictive distribution in forecasting as of 8 January in the 2008 water-year (a and b) and 2009 water-year (c and d) with important signals, $q = 2$, 4, and 6 over the 10 selected snow telemetry sites. The spread is given in the boxplots, in which the median, lower and upper quartiles, and minimum and maximum values are given

Table 1. Showing the proportion of variance explained by the orthogonal components of each decomposition (signal)

	Signal (g)					
	1	2	3	4	5	6
Proportion of variance (PV) (%)	80.68	9.89	3.01	1.92	0.95	0.67
Cumulative PV (%)	80.68	90.57	93.58	95.50	96.45	97.12

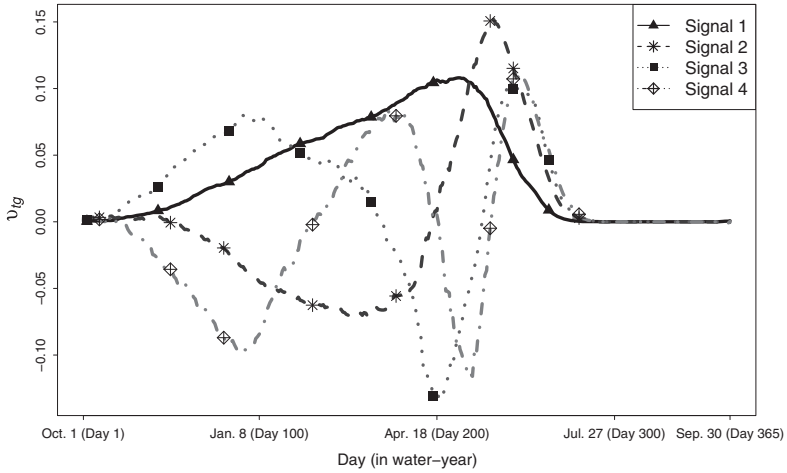

Figure 4. Plot of the orthonormal vectors, $\boldsymbol{v}_{tg}$ $(t = 1, 2, ..., 365; g = 1, 2, 3, 4)$ containing the first four annual dominant signals. These plots may vary by year to be forecasted and also depend on the data matrix set up in Equation (1)

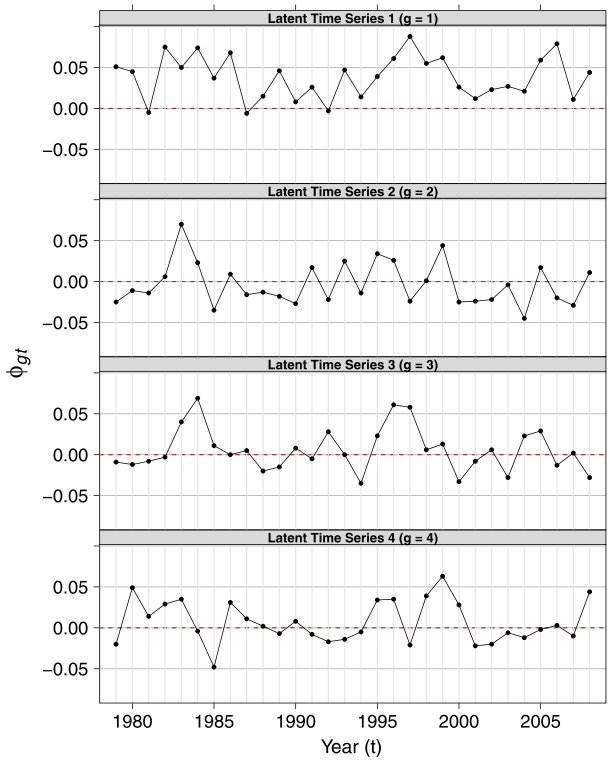

Figure 5. Plot of the left singular vectors (or latent time series for the dominant signals), $\boldsymbol{\phi}_{gt}$ $(g = 1, 2, 3, 4; t = 1, 2, ..., 30)$ containing the inter-annual time series for the Tony Grove snow telemetry site based on the four important signals (g) obtained from the snow water equivalent data reduction (1979–2008)

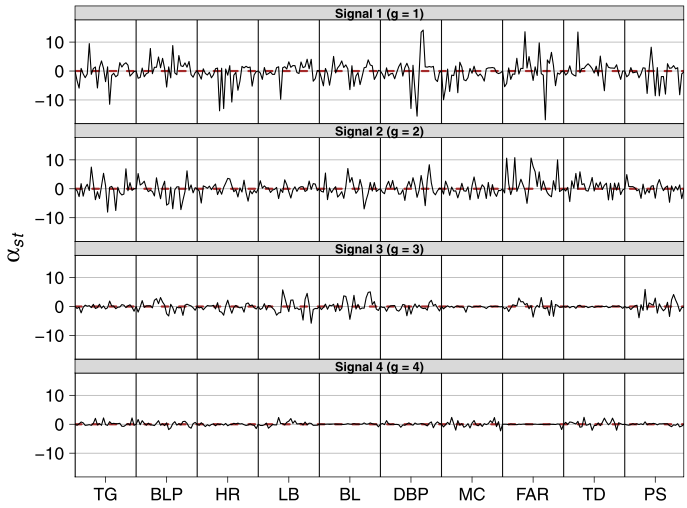

Figure 6. Plot of the estimates of the model parameter, $\boldsymbol{\alpha}_{st}$ $(s = 1, 2, ..., 10; t = 1, 2, ..., 30)$, over the 10 selected snow telemetry sites based on the four important signals (g) obtained from the snow water equivalent data reduction (1979–2008). TG=Tony Grove, BLP=Ben Lomond Peak, HR=Horse Ridge, LB=Little Bear, BL=Bug Lake, DBP=Dry Bread Pond, MC=Monte Cristo, FAR=Farmington, TD=Timpanogos Divide, and PS=Parley's Summit

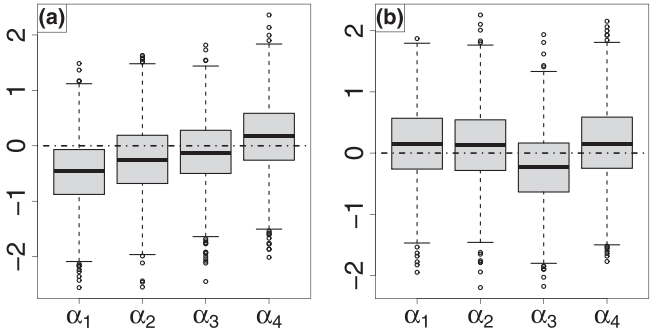

Figure 7. Boxplots for the estimates of the model parameter, $\boldsymbol{\alpha}$, based on the four important signals $(g = 1, 2, 3, 4)$ obtained from the snow water equivalent data reduction in the 2008 water-year for the (a) Tony Grove and (b) Little Bear snow telemetry sites. The spread is given in the boxplots, in which the median, lower and upper quartiles, and minimum and maximum values are given

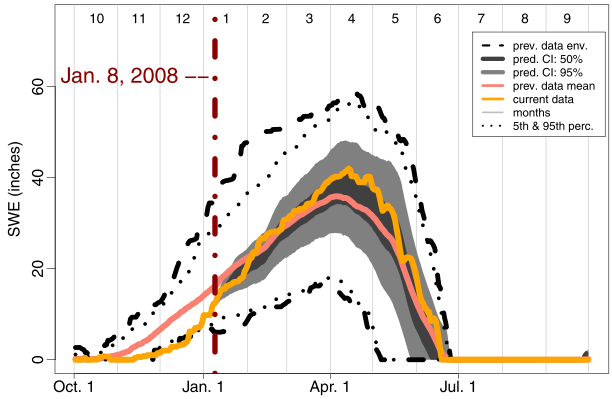

Figure 8. Observed snow water equivalents (SWEs) for 2008 water-year and posterior prediction as of 8 January 2008 (day 100; vertical brown dashed-dotted line) for the Tony Grove snow telemetry site. The horizontal axis shows the day (in water-year), and the vertical axis shows the SWE (in inches) in each plot. The previous data envelope (prev. data env., black dashed line) indicates the minimum and maximum SWEs for each day from the past 29 water-years. The dark gray-shaded region (pred. CI: 50%) and light gray-shaded region (pred. CI: 95%) are estimates of the 50% and 95% credible intervals, respectively, of the posterior predictive distribution. The solid salmon line (prev. data mean) shows the average amount of SWE based on the past water-years and the solid orange line (current data) indicates the actual observed SWE for the current water year. The dotted black line (5th & 95th perc.) represents the 5th and 95th percentile of the SWE measured for each day from the past 29 water-years

To choose the value of q for our truncation, we selected three different values (2, 4, and 6) and then constructed boxplots for the proportion of the actual SWE that fell outside the obtained 50% and 95% credible intervals for the posterior predictive distribution to forecast SWE after 100 days (January 8) with $q = 2$, 4, and 6 in the 2008 and 2009 water-years based on the 10 selected SNOTEL sites in this article. Overall, from Figure 3, with the exception of a few outliers, the boxplots for $q = 4$ showed a much shorter range of the proportion of the actual SWE that fell outside the obtained 50% and 95% credible intervals in both water-years. The initial SVD of the data matrix $(\mathbf{Y} - \bar{\mathbf{y}})$ for $T = 29$ and for $T = 30$ years showed that four of the signals contributed about 95.5% of the variability in the data (Table 1) and thus, provided a reasonable reduction in the dimension of the parameter set to facilitate estimation while retaining the dominant system dynamics. On the basis of these results, $q = 4$ was used for all of the following years.

A plot of the first four dominant seasonal signals can be seen in Figure 4. The first dominant signal (solid black line, signal 1), which accounts for approximately 81% of the variability in the data, reveals the overall average SWE measurements with the maximum around day 230 (18 May) of the water-year that begins on 1 October in the previous calendar year. The second most important signal (dashed line, signal 2), which explains an additional 10% of the variability in the data, gives an indication of the discrepancy between the middle and the end of the water year.

Figure 5 shows a latent time-series plot (say $\boldsymbol{\phi}_{gt}$) from $\mathbf{U}$ in equation (1) for the Tony Grove SNOTEL site ($s = 1$). In this figure, we consider the first four latent time series, which correspond to the expressions of the first four dominant signals through the years 1979–2008. We note here that for this site, $\boldsymbol{\phi}_{gt}$ is a $1 \times T$ matrix that consists of the gth row entries of $[\mathbf{u}_{s1}, \mathbf{u}_{s2}, \mathbf{u}_{s3}, ..., \mathbf{u}_{sT}]$. In the first latent time-series ($g = 1$) plot, we observe that the smallest value of almost zero occurs in years numbered 3 (1981), 9 (1987), and 14 (1992), and the largest value occurs in year numbered 19 (1997). The years with the smallest values indicate a very low amount of snow, and the year with the largest value had a large amount of snow. In case of the second latent time series ($g = 2$), a lower value indicates a smaller amount of snow at the beginning of the snow year (from 1 October to 30 November), and a higher value indicates a large amount of snow. This time series shows a discrepancy between the middle (from 1 February to 30 April) and late portion (between 1 May and 30 June) of the year. Combining these first two time series, we obtain a general shape of the overall amount of snow for a particular snow season with approximately 90% of the dynamics represented.

In terms of the estimation of model parameters $(\boldsymbol{\alpha}, \boldsymbol{\alpha}_{st}, \boldsymbol{\Sigma}_{\mathbf{y}})$ in Section 2.3.1 based on the posterior predictive distributions obtained for the parameters of the propagation matrix $\boldsymbol{\alpha}, \boldsymbol{\alpha}_{st}$, and $\sigma^2_{\mathbf{y}}$, using composition sampling, we found that although some of the posterior $\boldsymbol{\alpha}_{st}$ were quite different from zero especially in the cases of the first two important signals ($g = 1, 2$), the global autoregression coefficients $\boldsymbol{\alpha}$ were not (Figures 6 and 7). This is likely due to the limited amount of inter-annual data ($T = 29$), or it could also be that there is no temporal autocorrelation at these time scales. The covariance matrix for the short time scale had a very large posterior range parameter $\sigma^2_{\mathbf{y}}$, indicating a high degree of smoothness in the daily SWE measurements.

The posterior predictive credible intervals always performed well in terms of capturing the "unobserved" hold-out data as part of our model validation. We obtained posterior predictions for four different initiation dates/days (8 January, 7 February, 9 March, and 8 April). The posterior predictions in the calendar days showed poor results for forecasting. This is because we have little or no snow for that period of time in the water-year (especially from 1 October to 30 November). Figure 8 shows an enlarged example of a plot for the posterior prediction of SWE as a posterior predictive mean with 50% and 95% credible intervals. Here we considered the Tony Grove SNOTEL site using the 8 January (day 100) initiation date for our posterior predictions.

The graphs of the posterior predictions for SWE based on the four different initiation dates in five consecutive water-years (2008, 2009, 2010, 2011, and 2012) with 50% and 95% credible intervals for the Tony Grove and Little Bear SNOTEL sites are displayed in Figures 9 and 10. Table 2 contains a summary of the percentages of the actual SWE measurements that fall outside the 50% and 95% credible intervals for the Tony Grove and Little Bear SNOTEL sites.

In Figure 9, we notice that in the 2011 water-year, for all the initiation dates chosen, the 95% credible interval for the posterior prediction was not able to capture well a large proportion of the actual SWE. Specifically, from the end of March toward the beginning of July, we notice that SWE measures higher than the maximum from the previous 32 water-years. Similar extreme observations were also recorded at the Timpanogos Divide, Farmington, and Monte Cristo SNOTEL sites. This resulted in under-prediction because of our model being incapable of predicting outside the previous data envelope, that is, the historic minimum and maximum SWEs. Overall, the 2011 water-year rather seemed to be an unusual year for many of the 10 selected SNOTEL sites as a result of the record high measurements of SWE. In the 2012 water-year, for the initiation dates 7 February (day 130) and 8 March (day 160), the unseemly over-prediction is due to the high amount of snow recorded in the prior months.

In Figure 10, we notice that our credible interval for the 8 January initiation date of the 2011 water-year captures only a small proportion of the actual SWE. This noticeable over-prediction was as a result of an unexpectedly high amount of SWE measured toward the end of December (above the expected average). Our obtained posterior predictions showed that the three other initiation dates in 2011 water-year yielded useful forecasts for the rest of the season. The 8 January and 9 March initiation dates of the 2012 water-year also posed some problems for our model, given the steep SWE increase in mid-January and the steep SWE decline shortly after 9 March. Overall, the 95% credible intervals capture, on average, a large proportion of the actual SWE measurements in the other water-years (and initiation dates) for this site.

The probabilistic forecast skills, expressed by RPSS using the 8 January posterior predictions, are presented in Figure 11. In the 2008 water-year, with the exception of the Farmington site, all the SNOTEL sites show a positive RPSS. In addition, for the 2009 and 2010 water-years, all 10 sites show a positive RPSS and, thus, outscore the climatology forecast. For the 2011 water-year, all the SNOTEL sites show a negative RPSS, an indication of lesser accuracy from our model compared with that of climatology forecast. This is in agreement with what was found using the predictive credible interval.

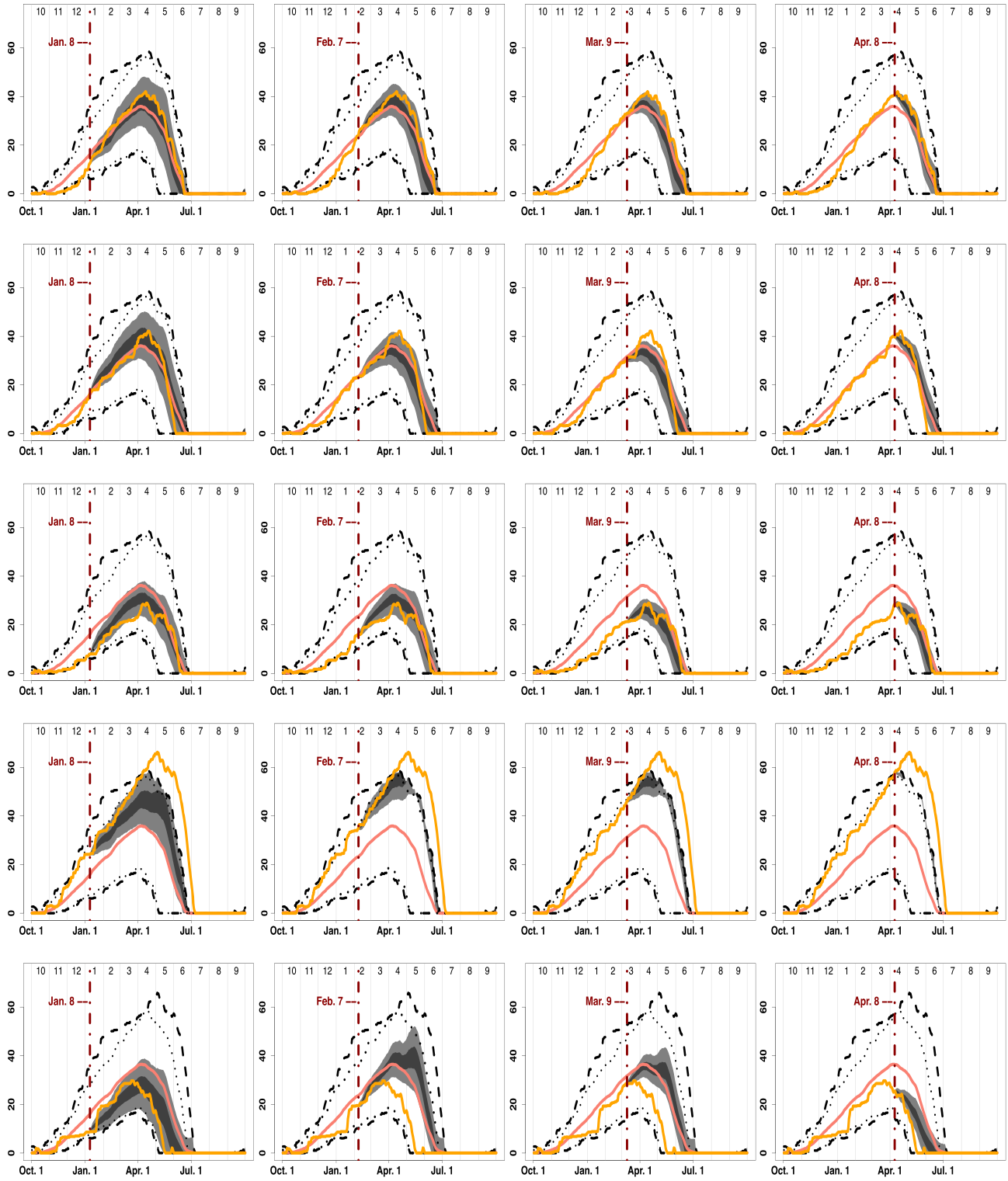

Figure 9. Observed snow water equivalents and posterior predictions for the Tony Grove snow telemetry site. Rows 1 (top) through 5 (bottom) show plots for the 2008, 2009, 2010, 2011, and 2012 water-years, respectively. Columns 1 (left) through 4 (right) show posterior predictions based on the initiation dates (8 January, 7 February, 9 March, and 8 April, respectively). See Figure 8 for additional details

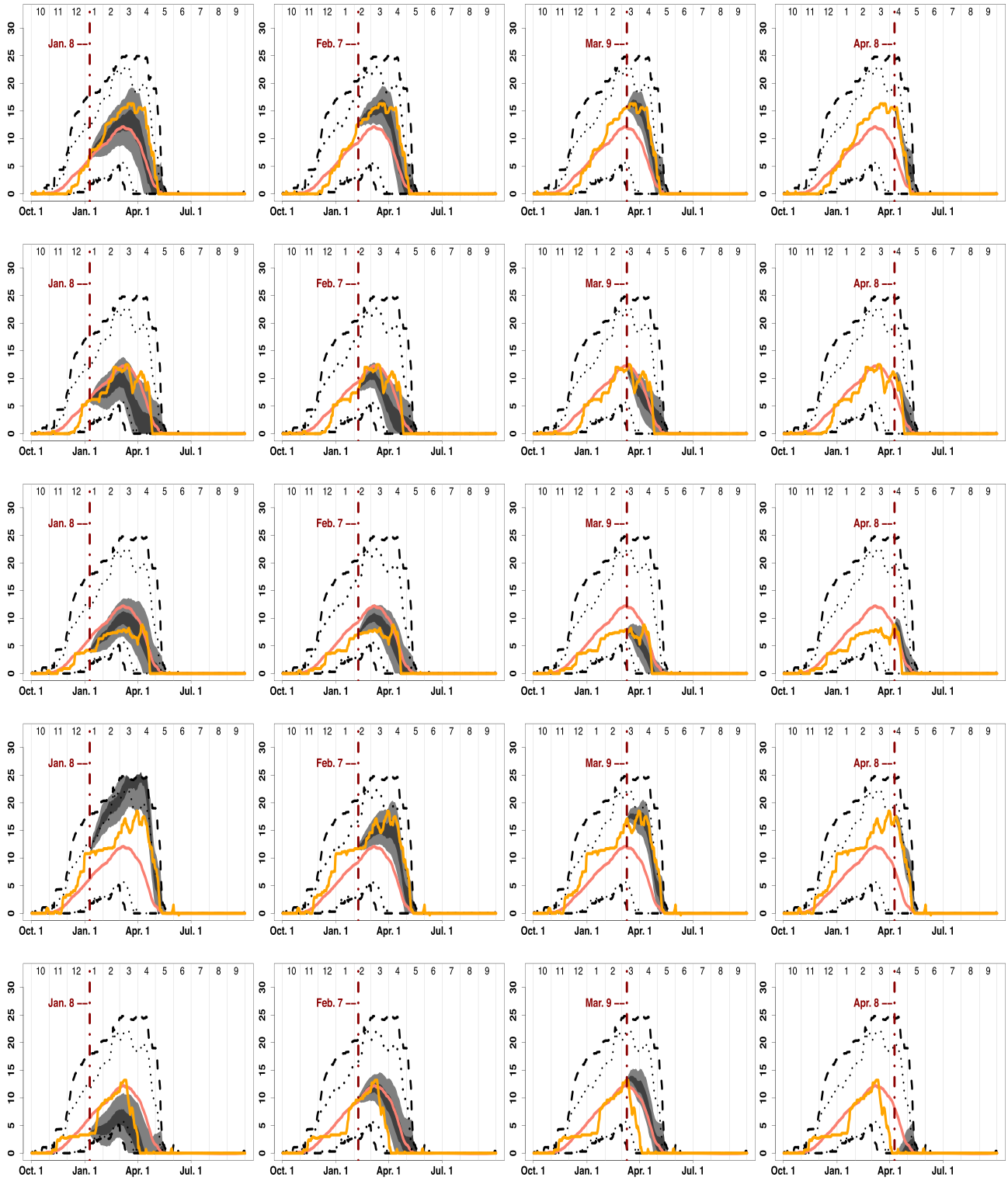

Figure 10. Observed snow water equivalents and posterior predictions for the Little Bear snow telemetry site. Rows 1 (top) through 5 (bottom) show plots for the 2008, 2009, 2010, 2011, and 2012 water-years, respectively. Columns 1 (left) through 4 (right) show posterior predictions based on the initiation dates (8 January, 7 February, 9 March, and 8 April, respectively). See Figure 8 for additional details

Table 2. Proportion of the observed SWE after the initiation dates: 8 January, 7 February, 9 March, and 8 April in the five selected water-years, which lies outside the 50% and 95% credible intervals (CI) obtained for the posterior predictive distribution for the Tony Grove and Little Bear SNOTEL sites

	Tony Grove Site								Little Bear Site							
	Percentage for 50% CI				Percentage for 95% CI				Percentage for 50% CI				Percentage for 95% CI			
Water-year	8 January	7 February	9 March	8 April	8 January	7 February	9 March	8 April	8 January	7 February	9 March	8 April	8 January	7 February	9 March	8 April
2008	25.6	29.8	32.8	31.7	0.0	1.5	11.3	1.5	34.3	19.2	16.2	13.2	7.2	5.7	0.4	1.9
2009	35.9	26.4	26.8	21.5	2.3	2.3	17.4	7.9	29.4	27.5	23.0	8.3	6.0	10.9	6.8	3.4
2010	36.6	30.2	28.7	18.1	8.3	7.5	3.8	4.9	26.0	24.2	8.7	9.8	0.4	6.0	4.5	6.8
2011	65.7	43.4	44.2	40.4	38.9	34.0	34.0	33.2	49.1	24.2	21.1	11.3	34.3	4.5	6.8	3.8
2012	40.4	43.4	36.6	18.9	4.5	30.9	29.8	12.5	26.8	27.2	28.3	7.5	20.4	7.9	16.6	1.1

SWE, snow water equivalent; SNOTEL, snow telemetry.

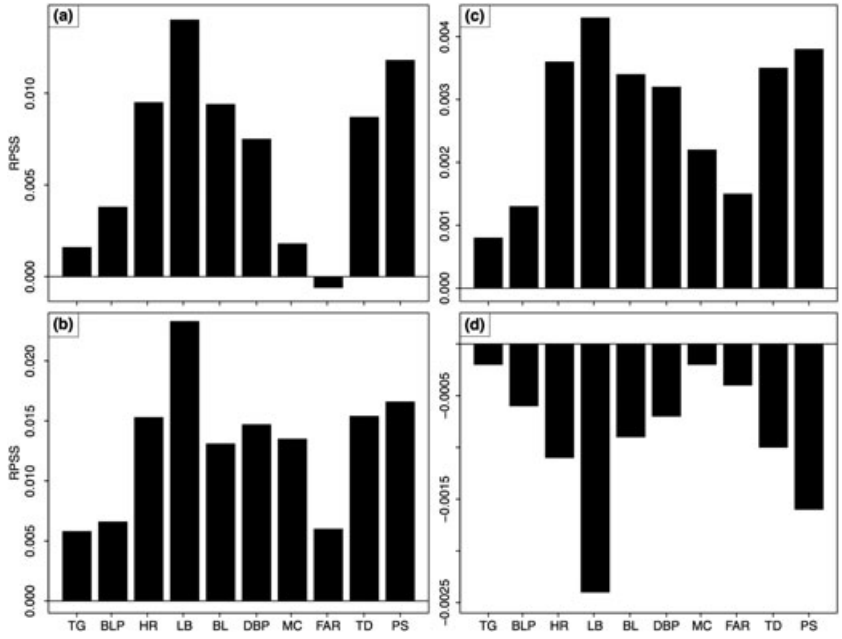

Figure 11. Ranked probability skill score for the 10 selected snow telemetry sites for the (a) 2008, (b) 2009, (c) 2010, and (d) 2011 water-years with the 8 January initiation date used in the posterior predictive forecasting. TG=Tony Grove, BLP=Ben Lomond Peak, HR=Horse Ridge, LB=Little Bear, BL=Bug Lake, DBP=Dry Bread Pond, MC=Monte Cristo, FAR=Farmington, TD=Timpanogos Divide, PS=Parley's Summit

4. GENERAL ASSESSMENTS

We used our Bayesian hierarchical model to successfully incorporate the SNOTEL SWE data in a likelihood and create posterior predictive distributions for the process of interest. Our model is comparable with other developed empirical statistical models, except that those were restricted to making statistical inferences on distributions, while our model relies heavily on observational data. Specifically, the statistical approach by itself does not explicitly incorporate knowledge about the atmospheric process but provides a heavily data-based method for prediction. Using a fully-rigorous statistical model for the phenomenological behavior of SWE over space and time, we are able to properly learn about the uncertainty in our predictions of SWE for the near-future.

A critical part of modeling efforts is evaluating model performance. A number of cross-validation approaches are available when it comes to model evaluation. However, model fit within the prediction credible interval is assumed to be homogeneous for all traditional model-fit metrics. A benefit of the Bayesian approach is that inference is made on the basis of the posterior distribution. For example, we considered plots of the posterior prediction of SWE as a posterior predictive mean with 50% and 95% credible intervals (Figure 8). The results shown in Table 2 indicate that the credible intervals for the posterior predictive of the non-spatial Bayesian hierarchical model proposed does well by capturing larger proportions of the actual SWE measured. For example, in the case of the 95% credible interval, larger proportion means 80% or more of the SWE measured. In addition, we observed that the model performed weakly for the Tony Grove site for all four initiation dates in 2011 and for the Little Bear site for the 8 January initiation date in 2011 with respect to the 95% credible interval. This was mainly due to discrepancies in the SWE measurements compared with the previous years; large amounts of snow were recorded in the month of December of the 2011 water-year for both sites. There were readily visible under-estimation, over-estimation, and data running out of the past min–max envelopes, especially in the 2011 water-year.

5. CONCLUSIONS

The statistical model presented in this article provides a framework for independently estimating the temporal dynamics of SWE for various SNOTEL sites. Using a parsimonious set of important signals, we can reduce the dimension of the SWE process and obtain meaningful posterior forecasts for the remainder of the water-year, beginning on various starting dates in a year. Except for the 2011 water-year, we noticed that our model predictive skills were positive. This is an indication that our model has some higher accuracy than that of climatology forecast. Our approach differs considerably from some of the predictive SWE models, in that, it focuses on phenomenological modeling (rather than mechanistic) and it produces within-season forecasts of SWE. In addition, our model is purely data based and does not rely on extensive mechanistic SWE models and their associated forcings and parameterizations.

Better knowledge about the uncertainty associated with snow water prediction could be put into a GIS framework to accumulate runoff volumes and provide information to farmers and land managers about future irrigation availability. Similarly, likely future water level ranges for reservoirs and lakes could be obtained in a similar fashion and provide information about fishery resources and recreational opportunities. Locally and regionally, dry conditions contribute to wildfire likelihood; thus, SWE predictions from our model could be aggregated at relevant spatial scales to provide an improved understanding of future soil moisture conditions at various time points in the late spring.

We noted that the statistical framework we presented could be modified to accommodate missing and misaligned data sources. Our model is currently not able to handle locations where historical SWE measurements contain data gaps (missing or unobserved values). However, we would be able to do meaningful forecasting if an appropriate imputation technique is first applied to fill these data gaps. This could be as simple as a moving average if only a few observations are missing or as complex as co-kriging if SWE measurements and other observations from nearby locations are available. New SWE stations are being added regularly and although this increases our spatial resolution, it does not help us learn about historical SWE patterns. These kinds of issues will not decrease our learning, but they may enhance it in other aspects (e.g., spatial information instead of temporal information) that we have not yet considered.

Our model was developed to best predict the remainder of a water-year given early information. It is able to properly account for the uncertainty in prediction so that it could be used to provide managers and scientists with an idea of how predictable a certain year might be. However, if one were interested in predicting a different quantity (like peak SWE) then a different model may need to be constructed and evaluated.

Ichii *et al.* (2007) discussed snow models in terrestrial biosphere models, and they pointed out that snow dynamics, snow depth, snow water storage, and spring run off are factors that directly affect intermountain ecosystem and economic function. Knapp (1998) discussed wildfire prevalence, modeling, and prevention. In all of these scenarios, predicted SWE might be another useful component to further enhance these models. Moreover, predicted SWE might be useful for reservoir level forecasting for summer water availability, irrigation potential for agriculture, and avalanche research.

Future work might include focusing on relationships between the SWE and other response variables of interest (water supply, fire frequency, etc.). This way, we can forecast meaningful seasonal runoffs that would be used to make sound water management decisions. The relationship between SWE and fire frequency could assist land managers with the assessment of post-fire water quality, reforestation, and animal habitat issues. For the problem of historical SWE measurements with data gaps, one could also borrow the strengths of physical process models by combining model results using a method similar to Bayesian model averaging (Hoeting *et al.*, 1999). Utilizing the components together, however, would be both advantageous and scientifically thorough and should provide better forecasts and understanding of snowpack dynamics. We will also consider the spatial dependence of all 90 active SNOTEL sites in Utah, and we will include other spatial covariates (e.g., latitude, longitude, temperature, precipitation, and elevation) in our modeling efforts. Finally, we plan to implement the general statistical model that was introduced in Odei *et al.* (2009).

REFERENCES

Baggerly KA, Berry DA. 2011. Science policy reproducible research. *Amstat News* **January 2011**(403):16–17.

Bailey RG. 1995. *Description of the Ecoregions of the United States.* 2nd Edition. Misc. Pub. No. 1391, Map scale 1:7500000, USDA Forest Service, Washington, DC, 108.

Banerjee S, Carlin BP, AE Gelfand AE. 2004. *Hierarchical Modeling and Analysis for Spatial Data.* Chapman & Hall/CRC: Boca Raton, Florida.

Bivand RS, Pebesma EJ, Gomez-Rubio V. 2008. *Applied Spatial Data Analysis With R.* Springer: New York, New York.

Bjornsson H, Venegas SA. 1997. A manual for EOF and SVD analyses of climate data. Center for Climate and Global Change Research Report No. 97-1, Montreal, Canada.

Brekke LD, Miller NL, Bashford KE, Quinn NWT, Dracup JA. 2004. Climate change impacts uncertainty for water resources in the San Joaquin River Basin, California. *Journal of the American Water Resources Association* **40**:149–164.

Carlin BP, Lewis TA. 2009. *Bayesian Methods for Data Analysis.* 3rd Edition. Chapman & Hall/CRC: Boca Raton, Florida.

Carroll SS, Carroll TR, Poston RW. 1999. Spatial modeling and prediction of snow-water equivalent using ground-based, airborne and satellite snow data. *Journal of Geophysical Research* **104**:19623–19629.

Carroll SS, Cressie N. 1996. A comparison of geostatistical methodologies used to estimate snow water equivalent. *Water Resources Bulletin* **32**(2): 267–278.

Carroll SS, Day GN, Cressie N, Carroll TR. 1995. Spatial modeling of snow water equivalent using airborne and ground-based snow data. *Environmetrics* **6**:127–139.

Cowles MK, Zimmerman DL, Christ A, McGinnis DL. 2002. Combining snow water equivalent data from multiple sources to estimate spatio-temporal trends and compare measurement systems. *Journal of Agricultural, Biological, and Environmental Statistics* **7**(4):536–557.

Cressie N, Johannesson G. 2008. Fixed rank kriging for very large spatial data sets. *Journal Of The Royal Statistical Society: Series B* **70**:209–226.

Davis SR, Seaman JW. 2002. Hierarchical Bayesian analysis of the carryover effect in two-period crossover design. *Mathematical and Computer Modeling* **36**:129–136.

Day GN. 1985. Extended streamflow forecasting using the National Weather Service River Forecast System. *Journal of Water Resource Planning and Management* **111**(2):157–170.

Day GN. 1990. A methodology for updating a conceptual snow model with snow measurements, NWS, 43, NOAA Silver Spring Maryland. Technical Report.

Garen DC. 1992. Improved techniques in regression-based streamflow volume forecasting. *Journal of Water Resources Planning and Management* **118**: 654–670.

Gelman A. 2006. Prior distributions for variance parameters in hierarchical models. *Bayesian Analysis* **1**(3):515–533.

Gelman A, Carlin JB, Stern HS, Rubin DB. 2003. *Bayesian Data Analysis.* 2nd Edition. Chapman & Hall/CRC: Boca Raton, Florida.

Genz A, Bretz F. 2009. *Computation of Multivariate Normal and t Probabilities.*, Lecture Notes in Statistics, vol. 195. Springer-Verlag: Heidelberg.

Genz A, Bretz F, Miwa T, Mi X, Leisch F, Scheipl F, Hothorn T. 2011. *mvtnorm: Multivariate Normal and t Distributions.* R package version 0.9-9991 http://CRAN.R-project.org/package=mvtnorm.

Golub GH, Reinsch C. 1970. Singular value decomposition and least squares solutions. *Numerische Mathematik* **14**(5):403–420.

Golub GH, van Loan CF. 1996. *Matrix Computations.* 3rd Edition. John Hopkins University Press: Baltimore, Maryland.

Hoeting JA, Madigan D, Raftery AE, Volinsky CT. 1999. Bayesian model averaging: a tutorial. *Statistical Science* **14**(4):382–417.

Hudlow MD. 1998. Technical developments in real-time operational hydrologic forecasting in the United States. *Journal of Hydrology* **102**:69–92.

Ichii K, White A, Votava P, Michaelis A, Nemani RR. 2007. Evaluation of snow models in terrestrial biosphere models using ground observation and satellite data: impact on terrestrial ecosystem processes. *Hydrological Processes* **22**(3):347–355.

Jackson CH. 2011. Multi-state models for panel data: the msm package for R. *Journal of Statistical Software* **38**(8):1–29.

Jin J, Gao X, Yang ZL, Bales RC, Sorooshian S, Dickinson RE, Sun SF, Wu GX. 1999. Comparative analyses of physically based snowmelt models for climate simulations. *Journal of Climate* **12**:2643–2657.
Jin J, Wen L. 2012. Evaluation of snowmelt simulation in the weather research and forecasting model. *Journal of Geophysical Research* **117**(D10110):1–16, DOI: 10.1029/2011JD016980.
Jonas T, Marty C, Magnusson J. 2009. Estimating the snow water equivalent from snow depth measurements in the Swiss Alps. *Journal of Hydrology* **378**(1):161–167.
Knapp PA. 1998. Spatio–temporal patterns of large grassland fires in the Intermountain West, USA. *Global Ecology and Biogeography Letters* **7**(4): 259–272.
Lapp S, Bryne J, Townshend I, Kienzle S. 2005. Climate warming impacts on snowpack accumulation in an Alpine watershed. *International Journal of Climatology* **25**:521–536.
Leisenring M, Moradkhani H. 2011. Snow water equivalent prediction using Bayesian data assimilation methods. *Stochastic Environmental Research and Risk Assessment* **25**(2):253–270.
Martin AD, Quinn KM, Park JH. 2011. MCMCpack: Markov chain Monte Carlo in R. *Journal of Statistical Software* **42**(9):22.
Miller NL, Bashford KE, Strem E. 2003. Potential impacts of climate change on California hydrology. *Journal of the American Water Resources Association* **39**:771–784.
Mote PW. 2006. Climate-driven variability and trends in mountain snowpack in Western North America. *Journal of Climate* **19**(23):521–536.
Mote PW, Hamlet AF, Clark MP, Lettenmaier DP. 2005. Declining mountain snowpack in Western North America. *Bulletin of the American Meteorological Society* **86**:39–49.
NRCS: Natural Resources Conservation Service (NRCS)—USDA. SNOTEL Data Collection
Odei JB, Hooten MB, Jin J. 2009. Inter-annual modeling and seasonal forecasting of intermountain snowpack dynamics. *2009 JSM Proceedings*: American Statistical Association, Alexandria, Virginia, *870–878,* (CD).
Pebesma EJ, Bivand RS. 2005. Classes and methods for spatial data in R. *R News* **5**(2):9–13.
R Core Team. 2012. *R: A Language and Environment for Statistical Computing.* R Foundation for Statistical Computing: Vienna, Austria. ISBN 3-900051-07-0.
Saha S, Nadiga S, Thiaw C, Wang J, Wang W, Zhang Q, Van Den Dool HM, Pan HL, Moorthi S, Behringer D, Stokes D, Pena M, Lord S, White G, Ebisuzaki W, Peng P, Xie P. 2006. The NCEP climate forecast system. *Journal of Climate* **19**(15):3483–3517.
Sarkar D. 2008. *Lattice: Multivariate Data Visualization With R.* Springer, New York: New York.
Schabenberger O, Gotway CA. 2005. *Statistical Methods for Spatial Data Analysis.* Chapman & Hall/CRC: Boca Raton, Florida.
Seidou O, Fortin V, St-Hilaire A, Favre A C, El Adlouni S, Bobée B. 2006. Estimating the snow water equivalent on the Gatineau catchment using hierarchical Bayesian modelling. *Hydrological Processes* **20**(4):839–855.
Shi T, Cressie N. 2007. Global statistical analysis of MISR aerosol data: a massive data product from NASA's terra satellite. *Environmetrics* **18**: 665–680.
Shumway RH, Stoffer DS. 2006. *Time Series Analysis and Its Applications: With R Examples.* (2nd Edition). Springer, New York: New York.
Stewart I, Cayan DR, Dettinger MD. 2005. Changes toward earlier streamflow timing across Western North America. *Journal of Climate* **18**: 1136–1155.
Suess EA, Gardner IA, Johnson WO. 2002. Hierarchical Bayesian model for prevalence inference and determination of a country's status for an animal pathogen. *Preventive Veterinary Pathogen* **55**:155–171.
Van Rheenen NT, Wood AW, Palmer RN, Lettenmaier DP. 2004. Potential implications of PCM climate change scenarios for California hydrology and water resources. *Climate Change* **62**:257–281.
Venables WN, Ripley BD. 2002. *Modern Applied Statistics With S.* Fourth Edition. Springer, New York: New York.
Wikle CK. 2010. Low-rank representations for spatial processes. In *Handbook of Spatial Statistics*, Gelfand AE, Fuentes M, Guttorp P, Diggle P (eds). Chapman and Hall/CRC: Boca Raton, Florida; 107–118.
Wikle CK, Hooten MB. 2010. A general science-based framework for dynamical spatio-temporal models. *Test* **19**:417–451.
Wikle CK, Milliff RF, Nychka D, Berliner LM. 2001. Spatio-temporal hierarchical Bayesian modeling: tropical ocean surface winds. *Journal of the American Statistical Association* **96**:382–397.
Wilks DS. 1995. *Statistical Methods in the Atmospheric Sciences: An Introduction.* Academic Press: San Diego, California, USA.

APPENDIX A. SUMMARY OF COMPUTER R CODE

The provided R code enables readers to apply the proposed methods to their own data and to reproduce all figures in this paper, a key requirement in terms of reproducible research (Baggerly and Berry, 2011). Datasets for this paper can be obtained from the freely accessible web archives of National Water and Climate Center (`http://www.wcc.nrcs.usda.gov/snow/`).

For each SNOTEL site, the "read.Snotel.R" function is used to clean up the historical and current water-year daily SWE measurements by reading the downloaded SNOTEL data in text format 1 year at a time and output vectors of values with NAs in place of blanks and "- - -". The "utah.snotel.current.data.R" and "utah.snotel.historical.data.R" functions use the "read.Snotel.R" function to put the already downloaded historical and current text files of SWE measurements for the 90 active Utah SNOTEL sites in a matrix form.

The "swe.svd.mcmc.R" function performs the SVD on the data matrix discussed in Equation (1), computes, and saves the sampling estimates of the model parameters $\boldsymbol{\alpha}, \boldsymbol{\alpha}_{st}$, and $\boldsymbol{\Sigma}_{\mathbf{y}}$ in Section 2.3.1 for a specified number of simulations. This code requires the R packages: MCMCpack (Martin *et al.*, 2011), mvtnorm (Genz and Bretz, 2009; Genz *et al.*, 2011), msm (Jackson, 2011), lattice (Sarkar, 2008), Mass (Venables and Ripley, 2002), and spdep (Pebesma and Bivand, 2005; Bivand *et al.*, 2008). It also produces samples of the posterior distributions for the unobserved SWE after a chosen day d in our forecasting of SWE. The function "plot.swe.post.R" uses output from the "swe.svd.mcmc.R" to enable us to reproduce Figures 8–10. To reproduce Figures 1–7 and 11, as well as Tables 1 and 2, we used some of the commands in the "Codes-Used.R" file. The function "prop.pred.R" is used to generate Table 2. The "invgammastrt.R" function computes the initial values of the shape and scale parameters of the inverse gamma distribution for $\sigma^2_{\mathbf{y}}$. The function "krig.swe.R" augments the historical SWE data for SNOTEL sites with missing values using a kriging approach. The text file "xcoord.txt" contains latitude and longitude coordinates of the 90 active SNOTEL sites in this article. All R code and data have been combined in a single zip archive (odei-et-al-R-Code-Data.zip) and can be downloaded from

`http://www.math.usu.edu/~symanzik/papers/odei-et-al-R-Code-Data.zip`. Our results were produced with R version 2.15.1.

APPENDIX B. DERIVATION OF THE FULL-CONDITIONAL DISTRIBUTIONS

Considering the definition of A_{st} in Section 2.3.1, we can rewrite Equation (3) as

$$\mathbf{z}_{st} = \mathbf{Q}_{s,t-1}\boldsymbol{\alpha}_{st} + \boldsymbol{\varepsilon}_{st} \tag{B1}$$

where $\mathbf{z}_{st} = \mathbf{y}_{st} - \bar{\mathbf{y}}_{s.}$ and $\mathbf{Q}_{s,t-1} = \mathbf{VD}diag(\mathbf{u}_{s,t-1})$.

The joint posterior distribution of the parameters in the full model, given the data, is given by

$$\left[\{\boldsymbol{\alpha}_{st}\}, \boldsymbol{\alpha}, \sigma^2_{\mathbf{y}}, \rho | \{\mathbf{z}_{st}\}\right] \propto \left(\prod_{t=1}^{T}\left[\mathbf{z}_{st}|\boldsymbol{\alpha}_{st}, \sigma^2_{\mathbf{y}}\right] \times [\boldsymbol{\alpha}_{st}|\boldsymbol{\alpha}]\right) \times [\boldsymbol{\alpha}] \times \left[\sigma^2_{\mathbf{y}}\right] \times [\rho] \tag{B2}$$

To use the computationally efficient Gibbs sampler, we need to analytically specify the full-conditional distribution for each parameter in the model. The derivations of these full-conditionals are given here.

From the joint posterior distribution in the preceding text, the full-conditional distribution for $\boldsymbol{\alpha}_{st}$, the autoregression coefficients related to the global coefficients, $\boldsymbol{\alpha}$, is

$$[\boldsymbol{\alpha}_{st}|\,.\,] \propto \left[\mathbf{z}_{st}|\boldsymbol{\alpha}_{st}, \sigma^2_{\mathbf{y}}\right] \times [\boldsymbol{\alpha}_{st}|\boldsymbol{\alpha}] \tag{B3}$$

The prior distribution of $\boldsymbol{\alpha}_{st}$ is $N\left(\boldsymbol{\alpha}, \sigma^2_{\boldsymbol{\alpha}}\mathbf{I}\right)$, and the likelihood is the process model:

$$\mathbf{z}_{st}|\boldsymbol{\alpha}_{st}, \sigma^2_{\mathbf{y}} \sim N(\mathbf{Q}_{s,t-1}\boldsymbol{\alpha}_{st}\,\boldsymbol{\Sigma}_{\mathbf{y}})$$

Further simplification yields

$$\begin{aligned}
[\boldsymbol{\alpha}_{st}|\,.\,] &\propto exp\left\{-\frac{1}{2}\left(\mathbf{z}_{st} - \mathbf{Q}_{s,t-1}\boldsymbol{\alpha}_{st}\right)' \boldsymbol{\Sigma}^{-1}_{\mathbf{y}}\left(\mathbf{z}_{st} - \mathbf{Q}_{s,t-1}\boldsymbol{\alpha}_{st}\right)\right\} \\
&\quad \times exp\left\{-\frac{1}{2}\left(\boldsymbol{\alpha}_{st} - \boldsymbol{\alpha}\right)'\left(\sigma^2_{\boldsymbol{\alpha}}\mathbf{I}\right)^{-1}\left(\boldsymbol{\alpha}_{st} - \boldsymbol{\alpha}\right)\right\} \\
&\propto exp\left\{-\frac{1}{2}\left(-2\mathbf{z}'_{st}\boldsymbol{\Sigma}^{-1}_{\mathbf{y}}\mathbf{Q}_{s,t-1}\boldsymbol{\alpha}_{st}\right)\right\} \times exp\left\{-\frac{1}{2}\left(\boldsymbol{\alpha}'_{st}\mathbf{Q}'_{s,t-1}\boldsymbol{\Sigma}^{-1}_{\mathbf{y}}\mathbf{Q}_{s,t-1}\boldsymbol{\alpha}_{st}\right)\right\} \\
&\quad \times exp\left\{-\frac{1}{2}\left(-2\boldsymbol{\alpha}'\left(\sigma^2_{\boldsymbol{\alpha}}\mathbf{I}\right)^{-1}\boldsymbol{\alpha}_{st}\right)\right\} \times exp\left\{-\frac{1}{2}\left(\boldsymbol{\alpha}'_{st}\left(\sigma^2_{\boldsymbol{\alpha}}\mathbf{I}\right)^{-1}\boldsymbol{\alpha}_{st}\right)\right\} \\
&\propto exp\left\{-\frac{1}{2}\left(-2\left(\mathbf{z}'_{st}\boldsymbol{\Sigma}^{-1}_{\mathbf{y}}\mathbf{Q}_{s,t-1} + \boldsymbol{\alpha}'\left(\sigma^2_{\boldsymbol{\alpha}}\mathbf{I}\right)^{-1}\right)\boldsymbol{\alpha}_{st}\right)\right\} \\
&\quad \times exp\left\{-\frac{1}{2}\left(\boldsymbol{\alpha}'_{st}\left(\mathbf{Q}'_{s,t-1}\boldsymbol{\Sigma}^{-1}_{\mathbf{y}}\mathbf{Q}_{s,t-1} + \left(\sigma^2_{\boldsymbol{\alpha}}\mathbf{I}\right)^{-1}\right)\boldsymbol{\alpha}_{st}\right)\right\}
\end{aligned} \tag{B4}$$

Then the full conditional distribution for $\boldsymbol{\alpha}_{st}$ is

$$\boldsymbol{\alpha}_{st} \sim N(\mathbf{A}^{-1}\mathbf{b}, \mathbf{A}^{-1}) \tag{B5}$$

where $\mathbf{A} = \mathbf{Q}'_{s,t-1}\boldsymbol{\Sigma}^{-1}_{\mathbf{y}}\mathbf{Q}_{s,t-1} + \left(\sigma^2_{\boldsymbol{\alpha}}\mathbf{I}\right)^{-1}$ and $\mathbf{b} = \left(\mathbf{z}'_{st}\boldsymbol{\Sigma}^{-1}_{\mathbf{y}}\mathbf{Q}_{s,t-1} + \boldsymbol{\alpha}'\left(\sigma^2_{\boldsymbol{\alpha}}\mathbf{I}\right)^{-1}\right)'$.

The full conditional distributions for $\boldsymbol{\alpha}$ and $\sigma^2_{\mathbf{y}}$ can be found in a similar way. Consider $\boldsymbol{\alpha}$:

$$[\boldsymbol{\alpha}|\,.\,] \propto \left(\prod_{t=1}^{T}[\boldsymbol{\alpha}_{st}|\boldsymbol{\alpha}]\right) \times [\boldsymbol{\alpha}] \tag{B6}$$

and the prior distribution of $\boldsymbol{\alpha}$ is $N(\mathbf{0}, \sigma_o^2\mathbf{I})$. Simplifying Equation (B6) yields

$$\begin{aligned}
[\boldsymbol{\alpha}|\,.] &\propto \left\{\prod_{t=1}^{T} exp\left(-\frac{1}{2}\left(\boldsymbol{\alpha}_{st}-\boldsymbol{\alpha}\right)'\left(\sigma_{\boldsymbol{\alpha}}^2\mathbf{I}\right)^{-1}\left(\boldsymbol{\alpha}_{st}-\boldsymbol{\alpha}\right)\right)\right\} \times exp\left\{-\frac{1}{2}\left(\boldsymbol{\alpha}-\mathbf{0}\right)'\left(\sigma_o^2\mathbf{I}\right)^{-1}\left(\boldsymbol{\alpha}-\mathbf{0}\right)\right\} \\
&\propto exp\left\{-\frac{1}{2}\sum_{t=1}^{T}\left(\boldsymbol{\alpha}_{st}-\boldsymbol{\alpha}\right)'\left(\sigma_{\boldsymbol{\alpha}}^2\mathbf{I}\right)^{-1}\left(\boldsymbol{\alpha}_{st}-\boldsymbol{\alpha}\right)\right\} \times exp\left\{-\frac{1}{2}\left(\boldsymbol{\alpha}'\left(\sigma_o^2\mathbf{I}\right)^{-1}\boldsymbol{\alpha}\right)\right\} \\
&\propto exp\left\{-\frac{1}{2}\left(\sum_{t=1}^{T}\left(-2\boldsymbol{\alpha}_{st}'\left(\sigma_{\boldsymbol{\alpha}}^2\mathbf{I}\right)^{-1}\boldsymbol{\alpha}\right)\right)\right\} \times exp\left\{-\frac{1}{2}\left(\sum_{t=1}^{T}\left(\boldsymbol{\alpha}'\left(\sigma_{\boldsymbol{\alpha}}^2\mathbf{I}\right)^{-1}\boldsymbol{\alpha}\right)\right)\right\} \times exp\left\{-\frac{1}{2}\left(\boldsymbol{\alpha}'\left(\sigma_o^2\mathbf{I}\right)^{-1}\boldsymbol{\alpha}\right)\right\} \\
&\propto exp\left\{-\frac{1}{2}\left(-2\left(\sum_{t=1}^{T}\boldsymbol{\alpha}_{st}'\left(\sigma_{\boldsymbol{\alpha}}^2\mathbf{I}\right)^{-1}\right)\boldsymbol{\alpha}\right)\right\} \times exp\left\{-\frac{1}{2}\left(\boldsymbol{\alpha}'\left(\sum_{t=1}^{T}\left(\sigma_{\boldsymbol{\alpha}}^2\mathbf{I}\right)^{-1}+\left(\sigma_o^2\mathbf{I}\right)^{-1}\right)\boldsymbol{\alpha}\right)\right\}
\end{aligned} \tag{B7}$$

Then the full conditional distribution for $\boldsymbol{\alpha}$ is

$$\boldsymbol{\alpha} \sim N(\mathbf{A_*}^{-1}\mathbf{b}_*, \mathbf{A_*}^{-1}) \tag{B8}$$

where $\mathbf{A}_* = \left(\sum_{t=1}^{T}\left(\sigma_{\boldsymbol{\alpha}}^2\mathbf{I}\right)^{-1}\right) + \left(\sigma_o^2\mathbf{I}\right)^{-1}$ and $\mathbf{b}_* = \left(\sum_{t=1}^{T}\boldsymbol{\alpha}_{st}'\left(\sigma_{\boldsymbol{\alpha}}^2\mathbf{I}\right)^{-1}\right)'$.

The full-conditional distribution in the case of $\sigma_{\mathbf{y}}^2$ is

$$\left[\sigma_{\mathbf{y}}^2|\,.\right] \propto \left(\prod_{t=1}^{T}\left[\mathbf{z}_{st}|\boldsymbol{\alpha}_{st}, \sigma_{\mathbf{y}}^2\right]\right) \times \left[\sigma_{\mathbf{y}}^2\right] \tag{B9}$$

and the prior distribution of $\sigma_{\mathbf{y}}^2$ is $IG(\nu_1, \nu_2)$. Here, it must be noted that

$$\left[\sigma_{\mathbf{y}}^2|\,\nu_1, \nu_2\right] = \frac{1}{\nu_1^{\nu_2}\Gamma(\nu_2)}\left(\sigma_{\mathbf{y}}^2\right)^{-(\nu_2+1)} exp\left\{-\frac{1}{\nu_1\sigma_{\mathbf{y}}^2}\right\} \tag{B10}$$

Letting $\mathbf{R}_{st} = \mathbf{z}_{st} - \mathbf{Q}_{s,t-1}\boldsymbol{\alpha}_{st}$ and further simplifying equation (B9) yields

$$\begin{aligned}
\left[\sigma_{\mathbf{y}}^2|\,.\right] &\propto \frac{1}{|\boldsymbol{\Sigma}_{\mathbf{y}}|^{\frac{T}{2}}}\left(\sigma_{\mathbf{y}}^2\right)^{-(\nu_2+1)} exp\left\{-\frac{1}{\nu_1\sigma_{\mathbf{y}}^2}\right\} \times exp\left\{-\frac{1}{2}\sum_{t=1}^{T}\left(\mathbf{R}_{st}'\boldsymbol{\Sigma}_{\mathbf{y}}^{-1}\mathbf{R}_{st}\right)\right\} \\
&= \frac{1}{\left|\sigma_{\mathbf{y}}^2\left(\mathbf{I}-\rho\mathbf{W}\right)^{-1}\right|^{\frac{T}{2}}}\left(\sigma_{\mathbf{y}}^2\right)^{-(\nu_2+1)} exp\left\{-\frac{1}{\nu_1\sigma_{\mathbf{y}}^2}\right\} \times exp\left\{-\frac{1}{2}\sum_{t=1}^{T}\left(\mathbf{R}_{st}'\left(\sigma_{\mathbf{y}}^2\left(\mathbf{I}-\rho\mathbf{W}\right)^{-1}\right)^{-1}\mathbf{R}_{st}\right)\right\} \\
&\propto \left(\sigma_{\mathbf{y}}^2\right)^{-\left(\frac{mT}{2}+\nu_2+1\right)} \times exp\left\{-\frac{1}{\sigma_{\mathbf{y}}^2}\left(\frac{1}{\nu_1}+\frac{1}{2}\sum_{t=1}^{T}\left(\mathbf{R}_{st}'\left(\mathbf{I}-\rho\mathbf{W}\right)\mathbf{R}_{st}\right)\right)\right\}
\end{aligned} \tag{B11}$$

Then the full conditional distribution for $\sigma_{\mathbf{y}}^2$ is

$$\sigma_{\mathbf{y}}^2 \sim IG(r^*, q^*) \tag{B12}$$

where $r^* = \frac{1}{\nu_1} + \frac{1}{2}\sum_{t=1}^{T}\left(\mathbf{R}_{st}'\left(\mathbf{I}-\rho\mathbf{W}\right)\mathbf{R}_{st}\right)$ and $q^* = \frac{M\times T}{2} + \nu_2$.

We have derived analytic full-conditional distributions for all parameters in the model, and the joint posterior distribution of the parameters in the model can be found using hybrid Metropolis–Hastings and Gibbs MCMC sampling algorithm (`?BCGs2003`; Carlin and Lewis, 2009).

