

## Inter–Annual Modeling and Seasonal Forecasting of Intermountain Snowpack Dynamics

James B Odei\*

Mevin B Hooten\*

Jiming Jin<sup>†</sup>

### Abstract

Due to 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 seasonal forecasts of snowpack would allow water resources to be more effectively managed throughout the entire water year. Many available models either require delicate calibrations, are inappropriate at certain scales, or only correspond to either the temporal or spatial domain. We present a data-based statistical model that characterizes seasonal snow water equivalent 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. We use SNOTEL data to implement and validate this model.

**Key Words:** Bayesian Model, Empirical Orthogonal Functions, SNOTEL, Snow Water Equivalent, Spatio-Temporal Model.

### 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 region of the semi-arid Western United States (WUS) comprises 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.

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 (Source: [ncdc.info@noaa.gov](mailto:ncdc.info@noaa.gov)), 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 based on 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 they perform poorly during the snowmelt stage due to inadequate representation of snowmelt regimes (Jin et al., 1999) and spring climate forcing (Saha et al., 2006).

With the growing concern about the effects of change in climate on snowpack accumulation (Gleick, 1987; Lettenmaier and Sheer, 1991), and high intensity of societal seasonal water supply requirements (Carroll and Cressie, 1996), accurate forecasting of water supplies has become increasingly urgent. Previous hydrologic simulation and spatial statistical models 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

\*Utah State University, Dept. of Mathematical Sciences, 3900 Old Main Hill, Logan, UT 84322

<sup>†</sup>Utah State University, Dept. of Watershed Sciences, 5210 Old Main Hill, Logan, UT 84322

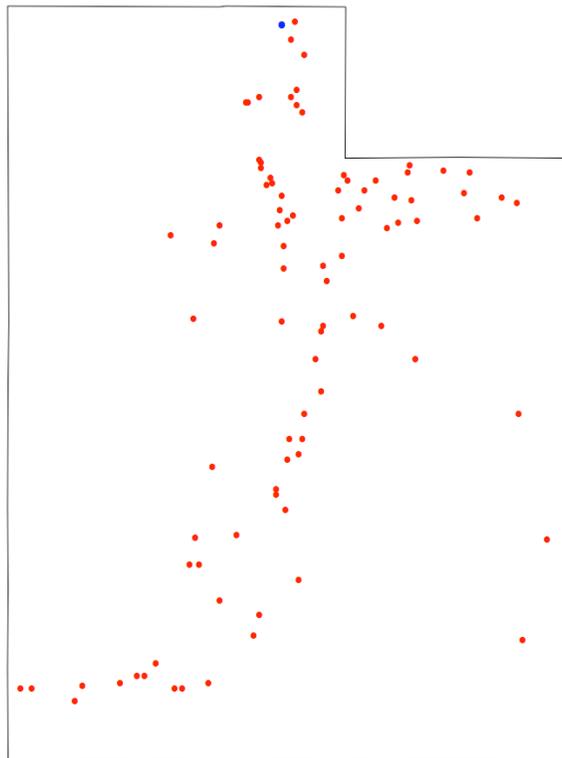
flood outlooks, and flood forecasts (Day, 1985; Hudlow, 1988) and in situations where areas have no observed measurements available, the NWS spatial statistical model uses ground-based and airborne snow water equivalent (SWE) data to estimate the SWE (Day, 1990; Carroll et al., 1995) across the WUS river basins on a 30 arc second grid (Carroll and Cressie, 1996).

Despite these recent successes in snow modeling, there still remains a need for accurate seasonal forecasts. The physical models, utilized in a conventional manner, impose a significant challenge in forecasting these systems, however, we can effectively address such issues by implementing statistically based snow models that rely heavily on observational data. In this paper we present a nested time-series statistical model for SWE data arising from the SNOTEL monitoring stations. Though we consider only the temporal aspect in this preliminary analysis, the approach can be easily extended to the spatial setting.

## 2. Material and Methods

### 2.1 SNOTEL Data Sources

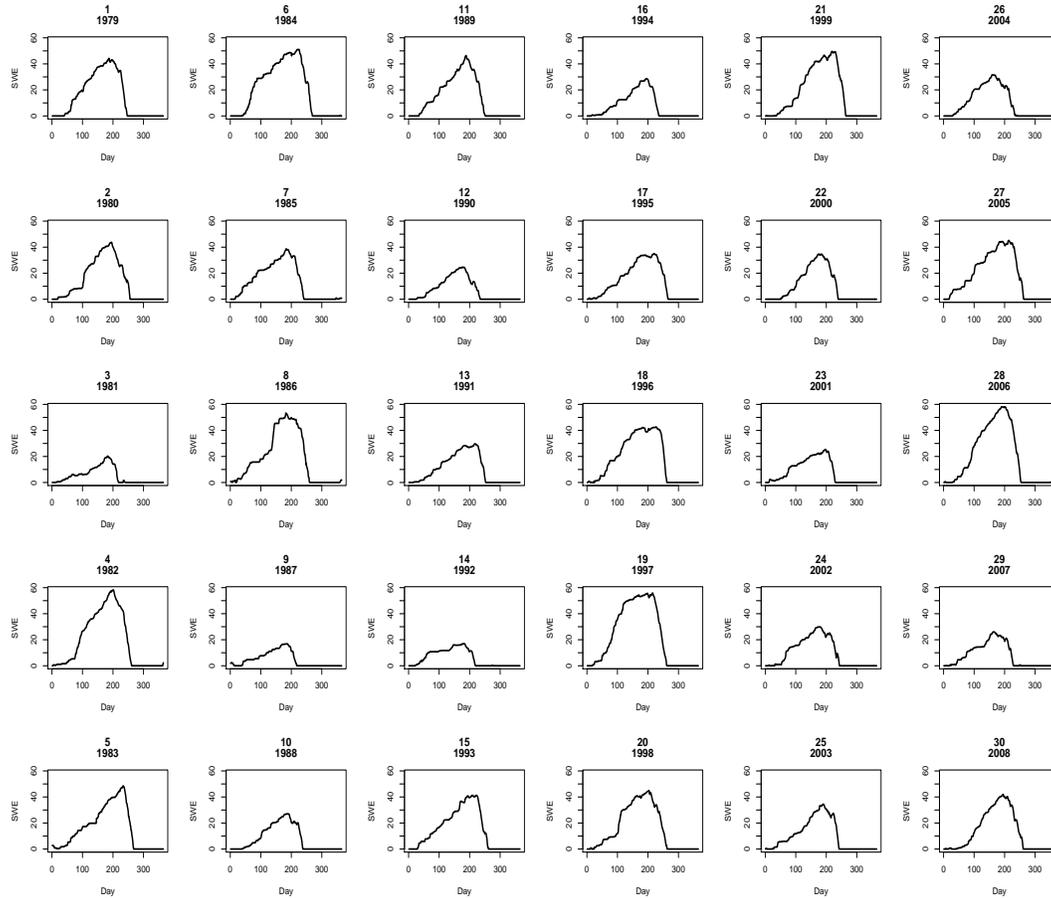
For the initial proof of concept portion of the study we focus our research efforts on the mountain ranges in Utah. This state has more than ninety snowpack telemetry (SNOTEL, NRCS 2008) monitoring stations that provide various snow and environmental measurements (i.e., swe, temperature, precipitation, snow depth) to be used in our modeling efforts. An illustration of the state-wide SNOTEL sites is shown in Figure 1.



**Figure 1:** Locations of active SNOTEL sites in Utah.

The plots of SWE measurements for Tony Grove SNOTEL site from 1979 to 2008 can be seen in Figure 2. A distinct feature of these data is the overall snow accumulation in a given year. For example, years with large (e.g., 1997) and small amounts of snow (e.g., 1981, 1987) are readily apparent. Also, there were years with

a relatively large amount of snow water storage at the beginning of the water year but not much snow water storage at the end. Beside these observations, there are other dominant signals that cannot be seen but nonetheless could be modeled.



**Figure 2:** Plots of SWE measurements for Tony Grove Site (1979–2008). Each of these plots represents the amount of water stored.

## 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., state-space models, Shumway and Stoffer, 2006) is to project the process onto a lower dimensional manifold for further modeling. This technique, sometimes referred to as fixed- or low-rank approximation (Wikle, 2009), 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. In this way, the data can be fully utilized to learn about an underlying process which is governed by a parsimonious set of parameters. Many forms of basis functions are possible (e.g., wavelets, Fourier, splines, empirical orthogonal functions) and their utility varies depending on the goals of the study.

In the spatio-temporal literature, principal components are commonly used for such dimension reduction and are referred to as empirical orthogonal functions

(EOFs). The two most common ways to obtain these basis functions are to decompose the space-time covariance matrix spectrally (eigen decomposition) or to use the singular value decomposition on a matrix of the data directly. In this case, we take the latter approach.

Applying SVD to our SNOTEL data, we can decompose the SWE data (over  $n$ -sites) in matrix form, say  $\mathbf{Y}_{nT \times m} = [\mathbf{y}_{11}, \dots, \mathbf{y}_{1T}, \dots, \mathbf{y}_{j1}, \dots, \mathbf{y}_{jT}, \dots, \mathbf{y}_{n1}, \dots, \mathbf{y}_{nT}]'$ , as:

$$\mathbf{Y} - \mathbf{M} = \tilde{\mathbf{U}}\tilde{\mathbf{D}}\tilde{\mathbf{V}}' \approx \mathbf{U}\mathbf{D}\mathbf{V}', \quad (1)$$

where  $n = 90$ ,  $T = 30$ , and  $m = 365$ . The  $\mathbf{U}\mathbf{D}\mathbf{V}'$  representation on the right hand side of (1) is a truncation of the  $\tilde{\mathbf{U}}\tilde{\mathbf{D}}\tilde{\mathbf{V}}'$  decomposition based on the first  $q$  important signals. The matrix  $\mathbf{M}$  has the same dimension as  $\mathbf{Y}$  ( $nT \times m$ ) with each row made up of the average SWE for each day over the 90 sites for the past 30 years. The matrix of orthonormal vectors  $\mathbf{V}$  is of dimension  $m \times q$ , and contains the dominant signals while the matrix of left singular vectors,  $\mathbf{U}$ , has dimension  $nT \times q$  and contains the times series and noting that  $q$  is much smaller than  $m$ . The singular values are contained in  $\mathbf{D}$ , a diagonal matrix of dimension  $q \times q$  with the eigenvalues on the diagonal. The approximation in (1) then, is the mechanism for reducing dimensionality in the parameter space. Since  $\mathbf{U}$  contains the set of time series corresponding to each signal in  $\mathbf{V}$  for all sites,

$$\mathbf{U} = [\mathbf{u}_{11}, \dots, \mathbf{u}_{1T}, \dots, \mathbf{u}_{j1}, \dots, \mathbf{u}_{jT}, \dots, \mathbf{u}_{n1}, \dots, \mathbf{u}_{nT}]', \quad (2)$$

we can obtain approximations of the SWE for a particular year  $t$ , at site  $j$  as

$$\mathbf{y}_{jt} \approx \boldsymbol{\mu} + \mathbf{V}\mathbf{D}\mathbf{u}_{jt}, \quad (3)$$

where  $\boldsymbol{\mu}$  is the average SWE for each day over the 90 sites for the past 30 years. This general equation (3) provides motivation for a hierarchical statistical model.

### 2.3 Nested Time–Series Models

In a more general sense, by considering the spatial dependence among the SNOTEL sites, we can specify a model for all of the SWE data by partitioning the temporal structure into a large and small temporal resolution. Thus, a nested time-series model with spatial structure results.

#### 2.3.1 General Statistical Model

Using a Bayesian approach, we propose a hierarchical statistical model for SWE. For prediction (forecasting), we model  $\mathbf{y}_{jt}$  (for all  $j, t$ ) as:

$$\mathbf{y}_{jt} = \boldsymbol{\mu} + \mathbf{V}\mathbf{D}\mathbf{u}_{jt} + \boldsymbol{\varepsilon}_{jt}, \quad \boldsymbol{\varepsilon}_{jt} \sim N(\mathbf{0}, \boldsymbol{\Sigma}_y) \quad (4)$$

where

$$\mathbf{u}_{jt} = A_t \mathbf{u}_{j(t-1)} + \boldsymbol{\eta}_{jt}, \quad \boldsymbol{\eta}_{jt} \sim N(\mathbf{0}, \sigma_u^2 \mathbf{I}) \quad (5)$$

with the covariance between any two sites with respect to the  $k^{\text{th}}$  signal defined as

$$\text{cov}(\eta_{kit}, \eta_{kjt}) = \sigma_k^2 \exp \left\{ \frac{-d_{ij}}{\phi_k} \right\}, \quad (6)$$

and

$$\text{cov}(\eta_{kit}, \eta_{ljt}) = 0, \quad \forall i, j, t, \tau, \quad k \neq l. \quad (7)$$

The covariance matrix,  $\boldsymbol{\Sigma}_y$  handles the small scale temporal SWE correlation which allows for latent autocorrelation in the days of the water year. The propagation matrix  $A_t$  then handles the large scale temporal dynamics. 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, or, more generally, allowed to vary in space.

### 2.3.2 Non-Spatial Statistical model

A simplified version of the general statistical model is to treat the SNOTEL sites independently. Applying the same basic modeling specification as in the general case before, we let the dynamics of SWE for one site be represented by a dimensionally reduced vector autoregression:

$$\mathbf{y}_t = \boldsymbol{\mu} + \mathbf{VDA}_t \mathbf{u}_{t-1} + \boldsymbol{\varepsilon}_t, \quad \boldsymbol{\varepsilon}_t \sim N(\mathbf{0}, \boldsymbol{\Sigma}_y) \quad (8)$$

where the covariance structure,  $\boldsymbol{\Sigma}_y$ , handles the small scale temporal (daily) SWE correlation, with the remaining parameters defined as:

$$\begin{aligned} A_t &= \text{diag}(\boldsymbol{\alpha}_t), \\ \boldsymbol{\alpha}_t &\sim N(\boldsymbol{\alpha}, \sigma_\alpha^2 \mathbf{I}), \\ \boldsymbol{\alpha} &\sim N(\mathbf{0}, \sigma_o^2 \mathbf{I}), \\ [\boldsymbol{\Sigma}_y]_{ij} &= \sigma_y^2 \exp\left\{\frac{-\delta_{ij}}{\phi}\right\}. \end{aligned}$$

The autoregression coefficients  $\boldsymbol{\alpha}_t$ , can be considered mixed effects that depend on the global coefficients,  $\boldsymbol{\alpha}$ . The  $A_t$ , then, handles the large scale temporal SWE correlation and  $\delta_{ij}$  is a measure of the temporal proximity which allows structure between neighboring days.

### 2.3.3 Implementation

For our Bayesian hierarchical statistical model, we then obtain the full conditional and posterior distributions as proportional to the product of the likelihood of the data given the latent process models and parameter models. The model was then implemented using a hybrid Metropolis-Hastings and Gibbs Sampling algorithm. The resulting samples of the posterior and posterior predictive distributions were then used to calculate posterior summary statistics.

For the purpose of forecasting we consider measurements of SWE for a water year  $\mathbf{y}_T$  partitioned into observed and unobserved SWE measurements. Thus, we can think of both as a joint multivariate normal random vector:

$$\mathbf{y}_T = \begin{bmatrix} \mathbf{y}_{Tobs.} \\ \mathbf{y}_{Tunobs.} \end{bmatrix} \sim N\left(\begin{bmatrix} \boldsymbol{\mu}_{obs.} \\ \boldsymbol{\mu}_{unobs.} \end{bmatrix}, \begin{bmatrix} \boldsymbol{\Sigma}_{obs.,obs.} & \boldsymbol{\Sigma}_{obs.,unobs.} \\ \boldsymbol{\Sigma}_{unobs.,obs.} & \boldsymbol{\Sigma}_{unobs.,unobs.} \end{bmatrix}\right). \quad (9)$$

Then, in the MCMC algorithm, we can use composition sampling to obtain samples from the posterior predictive distribution for the unobserved SWE as

$$\mathbf{y}_{Tunobs.} | \mathbf{y}_{Tobs.} \sim N(\boldsymbol{\mu}^*, \boldsymbol{\Sigma}^*), \quad (10)$$

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

$$\boldsymbol{\mu}^* = \boldsymbol{\mu}_{unobs.} + \boldsymbol{\Sigma}_{unobs.,obs.} \boldsymbol{\Sigma}_{obs.,obs.}^{-1} (\mathbf{y}_{Tobs.} - \boldsymbol{\mu}_{obs.}), \quad (11)$$

and

$$\boldsymbol{\Sigma}^* = \boldsymbol{\Sigma}_{unobs.,unobs.} - \boldsymbol{\Sigma}_{unobs.,obs.} \boldsymbol{\Sigma}_{obs.,unobs.}^{-1} \boldsymbol{\Sigma}_{obs.,unobs.}. \quad (12)$$

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

	Signal						
	1	2	3	4	5	6	7
Prop. of Variance (PV)	80.68%	9.89%	3.01%	1.92%	0.95%	0.67%	0.39%
Cumulative PV	80.68%	90.57%	93.58%	95.50%	96.45%	97.11%	97.50%

### 3. Results and Discussion

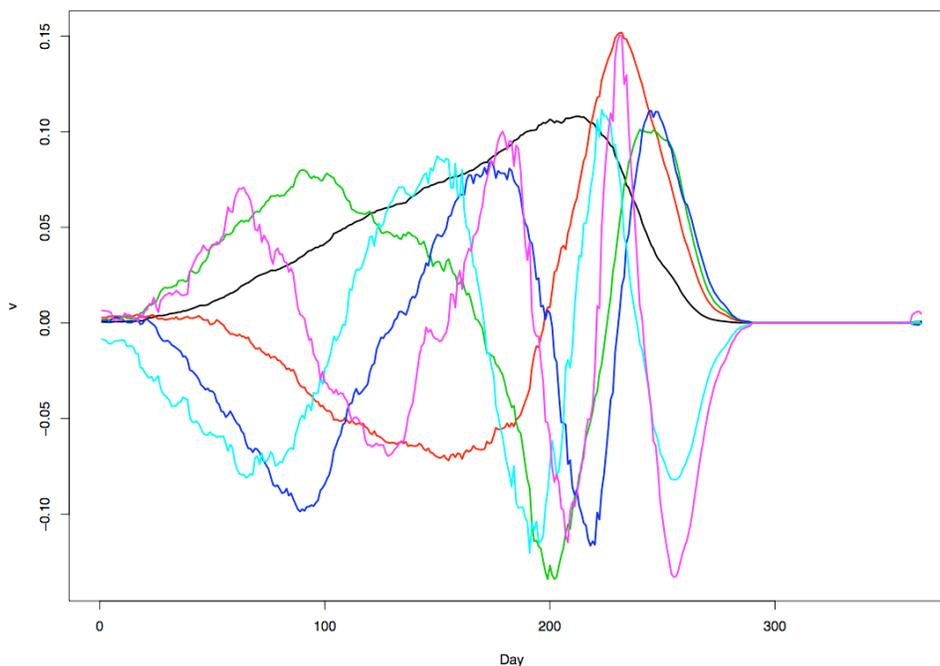
This section presents the results of fitting the model given in Section 2.3.2 to the data described in Section 2.1. SVD was performed on the matrix in (1) and truncated to obtain the dominant seasonal signals and inter-annual time-series. We consider plots of the posterior prediction of SWE as a posterior predictive mean with 50% and 95% credible intervals (see Figure 5). The initial SVD of the data matrix ( $\mathbf{Y} - \mathbf{M}$ ), showed that  $q = 6$  of the signals contributed greater than 97% of the variability in the data (see Table 1) and provided a reasonable reduction in the dimension of the parameter set to facilitate estimation while retaining the dominant system dynamics.

A plot of the first six dominant seasonal signals can be seen in Figure 3. The first dominant signal (black), which accounts for approximately 81% of the variability in the data, reveals the overall average SWE measurements with the maximum towards the end (around day 230; May 18<sup>th</sup>) of the water year which begins on October 1. The second most important signal (red), 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.

A plot of the time series for the chosen Tony Grove SNOTEL site can be seen in Figure 4. In this figure, consider the first two time series, which correspond to the expression of the first two dominant signals through the years 1979–2008. In the first time series plot (black), we observe that the smallest value occurs in years numbered 3, 9 and 14, with the largest occurring in year numbered 19 (1997). The years with the smallest values (1981, 1987, and 1992) reveals very low amount of snow and the year with the largest value (1997), had large amount of snow. In the case of the second time series (red), a higher value gives indication of large amount of snow at the beginning of the snow year and with a lower value indicating smaller amounts of snow. The red one shows a discrepancy between the middle and late portion of the year. Combining the two time series mentioned, we get 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 ( $\boldsymbol{\alpha}, \boldsymbol{\alpha}_t, \boldsymbol{\Sigma}_t$ ), we found that although most of the posterior  $\boldsymbol{\alpha}_t$  were significantly different from zero, the global autoregression coefficients  $\boldsymbol{\alpha}$  were not. This is likely due to the limited amount of inter-annual data ( $T = 30$ ). The covariance matrix for the short time scale had a very large posterior range parameter, indicating a high degree of smoothness in the daily SWE measurements.

For the purposes of validation, the model was then implemented using different days and SNOTEL sites in a water year for our posterior predictions. The posterior predictive credible intervals always performed well in terms of capturing the “un-observed” hold-out data as part of our model validation. We also found that the earliest day in the water year that yielded useful forecasts for the rest of the season was day 100 (Jan. 1<sup>st</sup>). Graphs of the posterior predictions for the SWE of the Tony Grove site in the years 2008 and 2009 with 50% and 95% credible intervals, are displayed in Figure 5.



**Figure 3:** Plot of the first six annual dominant signals.

#### 4. Conclusion

The statistical model we have presented, provides a framework for estimating the spatio-temporal dynamics of SWE for the various SNOTEL sites independently. Using a parsimonious set of important signals (6 in this case), we can reduce the dimension of the SWE process and obtain meaningful posterior forecasts for the remainder of the water year, beginning on January 1 of that year.

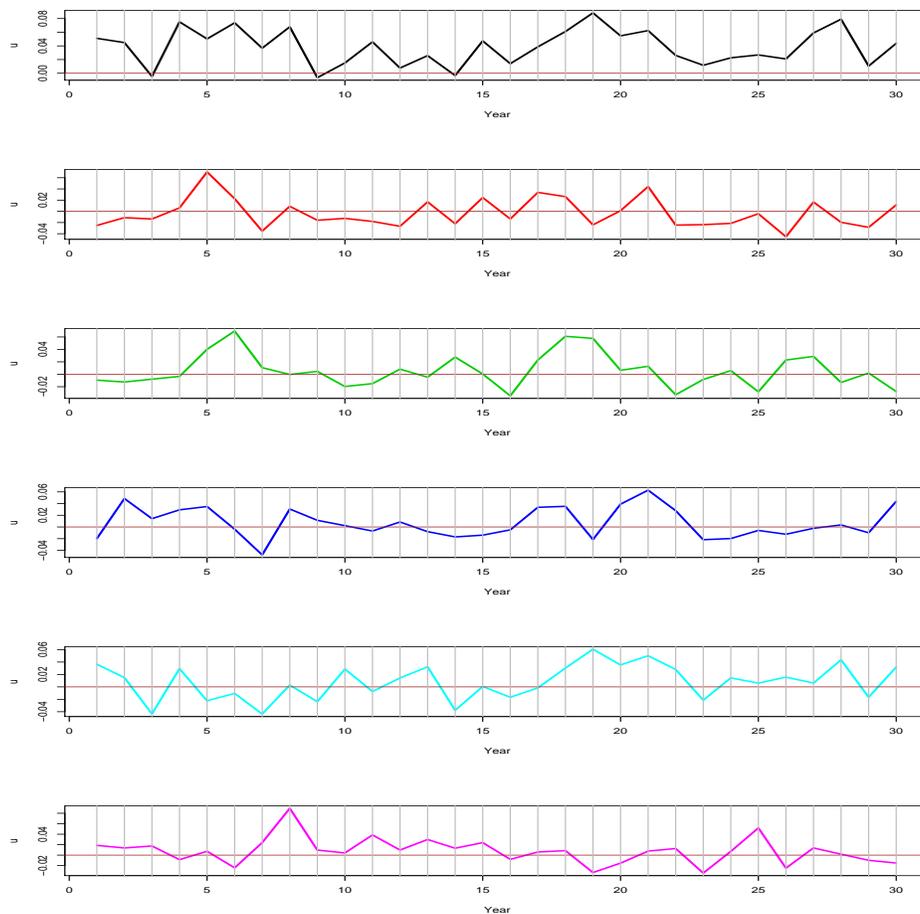
A few examples of areas where this research could have the highest impact are: reservoir level forecasting for summer water availability, irrigation potential for agriculture, avalanche research, both summer and winter forage and shelter for wildlife, water quality (temperature, sediment suspension, levels) for fisheries, both winter and summer recreation, and finally, wildfire prevalence, modeling, and prevention (Knapp, 1998).

Thus, as part of ongoing work, focusing on relationships between the SWE and other response variables of interest (water supply, fire frequency, etc.). Another area we will look at, is the consideration of the spatial dependence of the 90 sites and inclusion of other spatial covariates (e.g. latitude, longitude, temperature, precipitation, elevation, etc) in our modeling.

**Acknowledgements** This research was funded by the Utah Water Initiative.

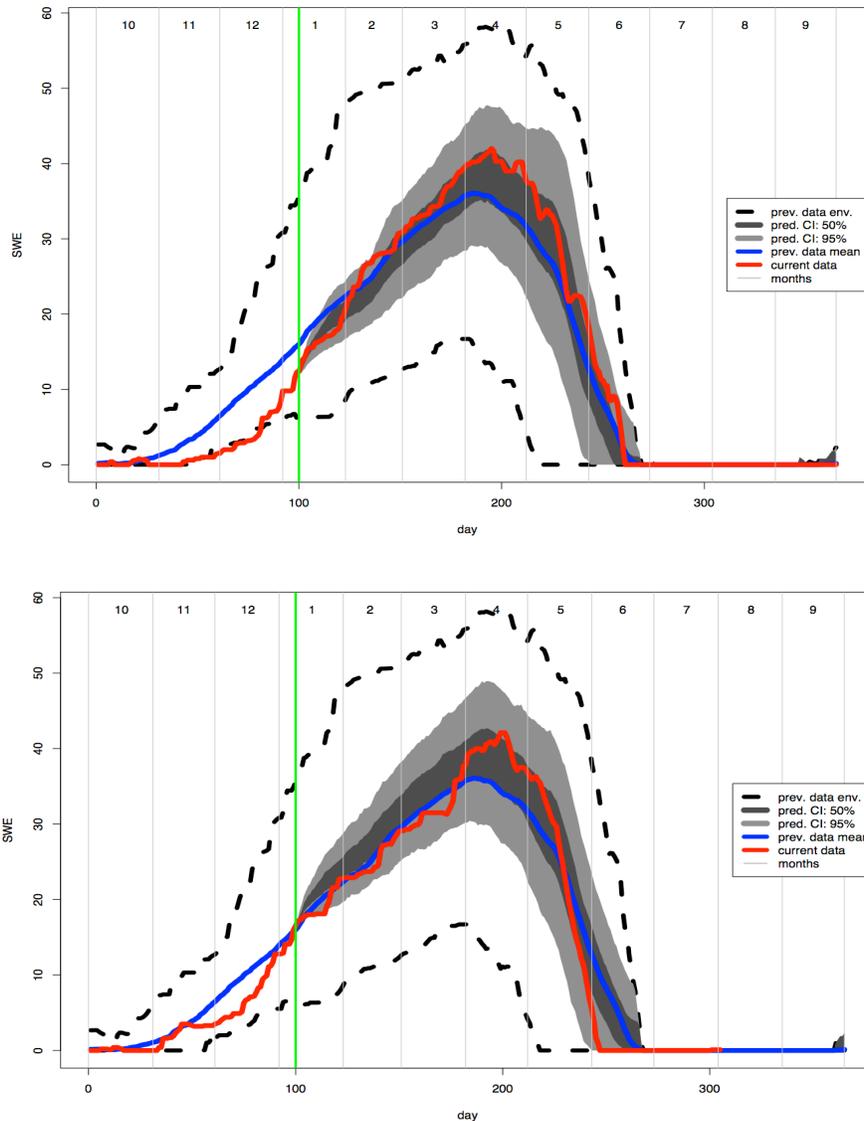
#### REFERENCES

- Bailey, R.G. (1995), "Description of the Ecoregions of the United States," (2nd ed.), Misc. Pub. No. 1391, Map scale 1:7500000, *USDA Forest Service*, pp. 108.
- Carroll, S. S., Day, G. N., Cressie, N., and Carroll, T. R. (1995), "Spatial Modeling of Snow Water Equivalent Using Airborne and Ground-Based Snow Data," *Environmetrics*, 6, 127–139.
- Carroll, S. S., and Cressie, N. (1996), "A Comparison of Geostatistical Methodologies Used To Estimate Snow Water Equivalent," *Water Resources Bulletin*, 32(2).
- Cressie, N., and Johannesson, G. (2008), "Fixed Rank Kriging for Very Large Spatial Data Sets," *Journal Of The Royal Statistical Society: Series B*, 70, 209–226.
- Cowles, M. K., Zimmerman, D. L., Christ, A., and McGinnis, D. L. (2002), "Combining Snow Water Equivalent Data from Multiple Sources to Estimate Spatio-Temporal Trends and Compare



**Figure 4:** Plot of the inter-annual time series for the Tony Grove site based on the six important signals obtained from the SWE data reduction (years 1979–2008).

- Measurement Systems,” *Journal of Agricultural, Biological, and Environmental Statistics*, 7(4), 536–557.
- Day, G. N. (1985), “Extended Streamflow Forecasting Using the National Weather Service River Forecast System,” *Journal of Water Resource Planning and Management*, ASCE, 111(2), 157–170.
- Day, G. N. (1990), “A Methodology for Updating a Conceptual Snow Model with Snow Measurements,” *NOAA Technical Report*, NWS, 43, Silver Spring, Maryland.
- Gleick, P. H. (1987), “Regional Hydrologic Consequences of Increases in Atmospheric  $CO_2$  and Other Trace Gases,” *Climate Change*, 10, 137–161.
- Hudlow, M. D. (1988), “Technical Developments in Real-Time Operational Hydrologic Forecasting in the United States,” *Journal of Hydrology*, 102, 69–92.
- Ichii, K., White, M. A., Votava, P., Michaelis, A., and Nemani, R. R. (2007), “Evaluation of Snow Models in Terrestrial Biosphere Models Using Ground Observation and Satellite Data: Impact on Terrestrial Ecosystem Processes,” *Hydrological Processes*, DOI: 10.1002/hyp.6616.
- Jin, J., Gao, X., Yang, Z. -L., Bales, R. C., Sorooshian, S., Dickinson, R. E., Sun, S. -F., and Wu, G. -X. (1999), “Comparative Analyses of Physically Based Snowmelt models for Climate Simulations,” *Journal of Climate*, 12, 2643–2657.
- Knapp, P. A. (1998), “Spatio-Temporal Patterns of Large Grassland Fires in the Intermountain West, USA,” *Global Ecology and Biogeography Letters*, 7, 259–272.
- Lettenmaier, D. P., and Sheer, D. P. (1991), “Climate Sensitivity of California Water Resources,” *Journal of Water Resources Planning and Management*, 117, 108–125.
- Mote, P. W., Hamlet, A. F., Clark, M. P., and Lettenmaier, D. P. (2005), “Declining Mountain Snowpack in Western North America,” *Bulletin of the American Meteorological Society*, 86, 39–49.
- Natural Resources Conservation Service (2008), SNOTEL Data Collection, Available at <http://www.wcc.nrcs.usda.gov/factpub/sntlfct.html>



**Figure 5:** Tony Grove site SWE measurements and posterior prediction as of: (Top) January 1, 2008, (Bottom) January 1, 2009.

- Saha, S., Nadiga, S., Thiaw, C., Wang, J., Wang, W., Zhang, Q., Van Den Dool, H. M., Pan, H.-L., Moorthi, S., Behringer, D., Stokes, D., Pena M., Lord, S., White, G., Ebisuzaki, W., Peng, P., and Xie, P., (2006), “The NCEP Climate Forecast System,” *Journal of Climate*, 19(15), 3483–3517.
- Shi, T., and Cressie, N. (2007), “Global Statistical Analysis of MISR Aerosol Data: A Massive Data Product from NASA’s Terra Satellite,” *Environmetrics*, 18, 665–680.
- Shumway, R. H., and Stoffer, D. S. (2006), “Time Series Analysis and Its Applications: With R Examples,” *Second Edition*, Springer, New York, NY.
- Wikle, C. K. (2009), “Low-Rank Representations for Spatial Statistics,” in Gelfand, A. E., Fuentes, M., Guttorp, P., and Diggle, P. (eds.), *Handbook of Spatial Statistics*, Chapter 1, Chapman and Hall/CRC, Boca Raton, FL. In Press.