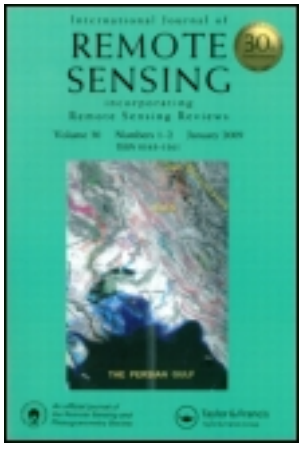

International Journal of Remote Sensing

Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/tres20

A comparison of SNOTEL and AMSR-E snow water equivalent data sets in western US watersheds

Cody L. Moser [a], Glenn A. Tootle [a], Abdoul A. Oubeidillah [a] & Venkat Lakshmi [b]

[a] Department of Civil and Environmental Engineering, University of Tennessee, Knoxville, TN, USA

[b] Department of Geological Sciences, University of South Carolina, Columbia, SC, USA

Available online: 22 Jul 2011

To cite this article: Cody L. Moser, Glenn A. Tootle, Abdoul A. Oubeidillah & Venkat Lakshmi (2011): A comparison of SNOTEL and AMSR-E snow water equivalent data sets in western US watersheds, International Journal of Remote Sensing, 32:21, 6611-6629

To link to this article: http://dx.doi.org/10.1080/01431161.2010.512936

International Journal of Remote Sensing
Vol. 32, No. 21, 10 November 2011, 6611–6629

Taylor & Francis
Taylor & Francis Group

A comparison of SNOTEL and AMSR-E snow water equivalent data sets in western US watersheds

CODY L. MOSER*†, GLENN A. TOOTLE†, ABDOUL A. OUBEIDILLAH† and VENKAT LAKSHMI‡

†Department of Civil and Environmental Engineering, University of Tennessee, Knoxville, TN, USA

‡Department of Geological Sciences, University of South Carolina, Columbia, SC, USA

(*Received 15 September 2009; in final form 12 July 2010*)

It is a consensus among earth scientists that climate change will result in an increased frequency of extreme events (e.g. floods, droughts). Streamflow forecasts and flood/drought analyses, given this high variability in the climatic driver (snowpack), are vital in the western USA. However, the ability to produce accurate forecasts and analyses is dependent upon the quality of these predictors. Run-off and stream volume analysis in the region is currently based upon *in situ* telemetry snow data products. Recent satellite deployments offer an alternative data source of regional snowpack. The proposed research investigates and compares remotely sensed snow water equivalent (SWE) data sets in western US watersheds in which snowpack is the primary driver of streamflow. Watersheds investigated include the North Platte, Upper Green and Upper Colorado. SWE data sets incorporated are *in situ* snowpack telemetry (SNOTEL) sites and the advanced microwave scanning radiometer-earth observing system (AMSR-E) aboard NASA's Aqua satellite. The time period analysed is 2003-2008, coincident with the deployment of the NASA Aqua satellite. Bivariate techniques between data sets are performed to provide valuable information on the time series of the snow products. Multivariate techniques including principal component analysis (PCA) and singular value decomposition (SVD) are also applied to determine similarities and differences between the data sets and investigate regional snowpack behaviours. Given the challenges (including costs, operation and maintenance) of deploying SNOTEL stations, the objective of the research is to determine whether remotely sensed SWE data provide a comparable option to *in situ* data sets. Correlation analysis resulted in only 11 of the 84 SNOTEL sites investigated being significant at 90% or greater with a corresponding AMSR-E cell. Agreement between SWE products was found to increase in lower elevation areas and later in the snowpack season. Two distinct snow regions were found to behave similarly between both data sets using a rotated PCA approach. Additionally, SVD linked both data products with streamflow in the region and found similar behaviour among data sets. However, when comparing SNOTEL data with the corresponding satellite cell, there was a consistent bias in the absolute magnitude (SWE) of the data sets. The streamflow forecasting results conclude regions that have few (or zero) land-based weather stations can incorporate the AMSR-E SWE product into a streamflow forecast model and obtain accurate values.

*Corresponding author. Email: cmoser3@utk.edu

International Journal of Remote Sensing
ISSN 0143-1161 print/ISSN 1366-5901 online © 2011 Taylor & Francis
http://www.tandf.co.uk/journals
http://dx.doi.org/10.1080/01431161.2010.512936

1. Introduction

Snowpack in mountainous regions of the western USA provides 50–80% of the water supply in the region (Natural Resources Conservation Service 2006). However, the West is being affected by a changed climate more than any other part of the United States, outside Alaska (Natural Resources Defense Council 2008). Warmer temperatures have major impacts on snowpack in the region. Increased temperatures cause more precipitation in the form of rain, rather than snow. A decrease in snowpack affects downstream users who expect water in the demand season (April–July). This snowpack acts as a freshwater reservoir and is a critical component for summer water supply in arid and semi-arid downstream regions.

Snow water equivalent (SWE) is the amount of water contained within a snowpack. It is the equivalent depth of water that would result if the entire snowpack was melted instantaneously. SWE data are important for an array of reasons; streamflow forecasts along with flood and drought analysis are important uses of SWE data sets. Accurate measurements of this snowpack are vital in understanding the hydrologic variability in the western USA. Effective management of limited water supplies is a critical component of the sustainability of populations throughout the western USA (Pagano *et al*. 2004).

Historically, SWE data sets have been collected for more than a century. SWE was first measured manually using a snow course. A snow course is a permanent site where manual measurements of snow depth and SWE are taken by trained observers. Generally, the courses were approximately 300 m long and situated in small meadows protected from the wind. Technological advances have allowed the National Resource Conservation Service (NRCS) to use SNOwpack TELemetry (SNOTEL) sites. The NRCS currently operates and maintains over 750 automated SNOTEL stations in 13 western states. SNOTEL uses meteor burst technology. Therefore, the data, as well as related reports and forecasts, can be made available in near real time. SNOTEL sites are generally located in remote watersheds and provide SWE data via a pressure-sensing snow pillow. Additionally, SNOTEL sites have a storage precipitation gauge and an air-temperature sensor, and they can accommodate 64 channels of data and will accept analogue, parallel or serial digital sensors (NRCS 2006). The snow-pillow device is coupled with a pressure transducer to measure snow water content. Before being placed into a database, an initial screening process is performed to investigate potential data errors. The area covered by SNOTEL was estimated at over 2×10^6 km^2 when the project was launched (Barton and Burke 1977).

The advanced microwave scanning radiometer-earth observing system (AMSR-E) instrument aboard NASA's Aqua satellite was launched in May 2002. The AMSR-E instrument measures land, oceanic and atmospheric parameters for the investigation of global and water energy cycles. The AMSR-E operational snow-mapping algorithm uses an empirical relationship to estimate SWE from surface brightness temperature (Chang *et al*. 1987). The microwave brightness temperature emitted from a snow cover is related to the snow mass, which can be represented by the combined snow density and depth (Kelly *et al*. 2003). Development of the microwave snow depth and SWE algorithm is based on experiences gained using the Nimbus-7 Scanning Multichannel Microwave Radiometer (SMMR) and Defense Meteorological Satellite Program (DMSP) Special Sensor Microwave/Imager (SSM/I) data. Snow crystals are effective scatterers of microwave radiation. When microwave radiation passes

through the snowpack, its intensity is modified by the snow crystals (Chang and Rango 2000). In general, the larger the SWE value, the more snow crystals are available to scatter microwave energy away from the sensor. The intensity of microwave radiation emitted from a snowpack depends on the physical temperature, grain size, density and the underlying surface conditions of the snowpack (Chang and Rango 2000). Challenges involved with snow-cover retrieval algorithms include forest density, liquid precipitation and complex topography. Chang and Rango (2000) provide additional information about remotely sensed snow extent, depth and water equivalent.

A comparison of *in situ* SWE with satellite-based SWE is the first step in determining the similarities and differences between the data sets. Three outcomes are possible: (1) the data sets are closely related, (2) the data sets have little similarity, or (3) the data sets are inconsistent. If satellite-based SWE is closely related to *in situ* SWE, steps can be taken to incorporate satellite data in streamflow, flood and drought forecasts and analyses. Additionally, and if required, new land-based climate stations can be effectively positioned based on conclusive results from satellite data. However, if the data sets have little similarity between them, the reliability and consistency of both data sets comes into question, particularly satellite data. Finally, the data sets may be inconsistent (e.g. highly related in one region and unrelated in another region). This is possible because both data sets have instrumentation and measurement limitations, and the topography, as well as the climate variability of the region, may limit the accuracy of SWE measurements.

There are advantages and disadvantages associated with each data set. Snow metamorphism, forest cover, liquid precipitation and complex landscapes all affect the microwave emission characteristics making it difficult to extract accurately values of snow properties with satellite instrumentation (Andreadis and Lettenmaier 2006). However, these factors are much less of an issue with land-based stations. Satellite data provide global coverage, while *in situ* stations only provide point coverage. Installation, operation and maintenance of land-based sensors are costly, but it is also expensive to deploy satellites with advanced instrumentation. Both land-based and satellite-based data are gathered and published in near real time. This assists water-supply forecasters and managers in making timely decisions for efficient allocation of water. Between the two products, there is a significant difference in the period of record in which data has been collected (i.e. launch date). Digital land-based sensors that measure SWE were first installed around 1980; instruments aboard satellites that measure SWE were launched as early as 2002.

Although AMSR-E data sets are relatively new (i.e. short period of record), there are numerous applications these data sets can be used for, and a large amount of research has incorporated these data into studies. The value of AMSR-E soil moisture data was shown in Bindlish *et al*. (2009), and flood forecasts were improved with success. The work of Reichle *et al*. (2007) compared AMSR-E and the SMMR soil moisture data sets. Spatial and temporal resolution of sea surface temperatures (SSTs) using AMSR-E data was dramatically improved in Reynolds *et al*. (2007). The research of Narayan and Lakshmi (2008) fused soil moisture estimates from the AMSR-E instrument with backscatter data and produced a higher spatial resolution of soil moisture variability. Additional research has utilized AMSR-E sea ice data (Cavalieri *et al*. 2006, Comiso and Steffen 2008) and moderate resolution imaging spectroradiometer (MODIS) data (King *et al*. 1992, Zarco-Tejada *et al*. 2003).

The objective and contribution of this research is to compare land-based SWE point data and satellite-based SWE spatial data within the western USA. This will provide comprehensive and valuable information about the validity of each data set. While satellite products have been compared to other satellite products, few studies have been carried out comparing satellite observations with land-based SNOTEL data. This is most likely attributed to the lack of an intense spatial coverage of SNOTEL sites given the resolution of satellite data (i.e. there is generally one SNOTEL gauge in a grid of 25 km^2). While this is a challenge/limitation, there is an interest in comparing both the spatial and temporal variability and the magnitude of satellite data to *in situ* data. Surface observations, such as snow courses and automated *in situ* measurement devices like snow pillows, are unable to capture fully the considerable spatial and temporal variability in snow properties over large areas (Andreadis and Lettenmaier 2006). Ultimately, one would hypothesize that the use of satellite-generated spatial SWE data will result in an improved representation of basin-wide hydrology compared to *in situ* data. The incorporation of satellite data in hydrologic-based forecasts is a natural 'next step', as displayed in the movement from snow course (manually measured snowpack data) to SNOTEL (snowpack data provided by remote sensors). Research and analysis of global hydrology may be based strictly on satellite data in the near future. This would increase the efficiency, as well as decrease the costs of collecting hydroclimatic data.

2. Watershed descriptions

Three critical western US watersheds that receive extensive snowfall are included within the scope of this research. Basins include the North Platte, Upper Green and Upper Colorado. The headwaters of the North Platte River are located in northern Colorado. The watershed is bound on all sides by mountain ranges: the Rawah and Never Summer ranges to the east, Rabbit Ears range to the south and Park Range to the west (Daniels 2007). From northern Colorado, the North Platte River continues north into Wyoming and then flows east into Nebraska. Understanding the hydrology of the North Platte River headwaters is critically important for water-resource planning in the Rocky Mountains and Great Plains regions (Daniels 2007).

The Upper Green River originates in western Wyoming and is the primary tributary of the Colorado River. The headwaters begin in the Wind River Mountains. From Wyoming, the Green River flows south into Utah and Colorado, where it flows into the Colorado River. The Upper Colorado River Basin originates in the mountains of central Colorado. The Colorado River flows south into Nevada and Arizona. The Colorado River is the major source of water for the driest part of the country. Upwards of 30×10^6 Americans across seven states now depend on it for agricultural, municipal, industrial and hydroelectric needs – and the basin is among the fastest-growing areas in the country (NRDC 2008). The locations of all watershed regions are located in figure 1.

3. Data sources

3.1 *SNOTEL*

Four states containing 84 SNOTEL sites are used in this study. The distribution is as follows: Wyoming (36), Colorado (35), Idaho (7) and Utah (6). Snowpack on the first

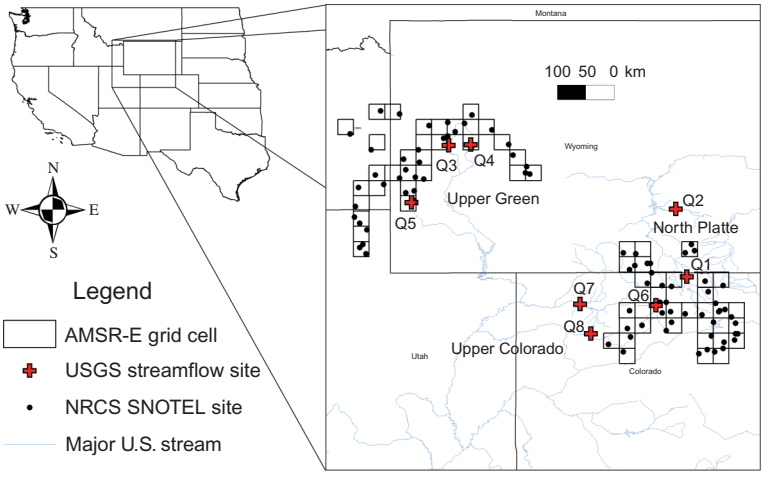

Figure 1. Location map showing the three regions investigated and all data used (SNOTEL stations, AMSR-E grid cells and streamflow stations).

day of the month (January–April) from 2003 to 2008 is used in this study. Additionally, SWE on the first day of April is a good indicator of the water content in the maximum seasonal snowpack and for the water supply for the coming months (Woodhouse 2003) and is the primary value of SWE used in this study. The SNOTEL sites used in this study are shown in figure 1.

3.2 *AMSR-E (NASA)*

AMSR-E data are mapped globally onto 25 km^2 Equal-Area Scalable Earth (EASE) Grids. Additional information about the EASE grid system is located in Brodzik and Knowles (2002). This study breaks the gridded data down further into 0.25° longitude by 0.25° latitude cells. Available AMSR-E SWE data sets include daily, five-day maximum and monthly average. The five-day maximum SWE value nearest the first of the month (January–April) from Kelly *et al*. (2004) is used in this research. It should be noted that AMSR-E data experiences numerous validation stages. Data sets used within this work are in the transitional validation stage. The transitional validation stage is the period between the beta and the validated stages. Figure 1 illustrates the AMSR-E grid cells used in this study.

3.3 *Streamflow*

Within the United States, the United States Geological Survey (USGS) collects surface-water data that describe stream levels, streamflow (discharge), reservoir and lake levels, surface-water quality and rainfall. Automatic recorders and manual measurements collect data. Slack and Landwehr (1992) identified a Hydro-Climatic Data Network (HCDN) of stream gauges as being relatively free of significant human influences and, therefore, appropriate for climate studies. Streamflow measurements from eight of these gauges are incorporated in this study. The average streamflow in the region for the months of April–July (AMJJ) from 2003 to 2008 is the value of interest that water managers are most interested in, and is used in this study. Table 1

Table 1. Streamflow sites included in the study: North Platte (2), Upper Green (3) and Upper Colorado (3).

Station ID	Stream	Basin	Drainage area (km^2)	Elevation (m)	Symbol
06620000	North Platte	NP	3706	2381	Q1
06635000	Medicine Bow	NP	6055	1955	Q2
09188500	Green	UG	1212	2276	Q3
09196500	Pine Creek	UG	196	2271	Q4
09223000	Hams Fork	UG	332	2272	Q5
09239500	Yampa	UC	1471	2041	Q6
09251000	Yampa	UC	8832	1798	Q7
09304500	White	UC	1955	1920	Q8

contains physical characteristics for the streamflow stations utilized in this study. The locations of these stations are shown in figure 1.

4. Methods

Several methods are typically used to determine the relationship between two spatial-temporal arrays of data such as climate variability (e.g. snowpack) and streamflow. Common methods include correlation analysis, principal components analysis (PCA) and singular value decomposition (SVD) (Soukup *et al*. 2009). Past research has applied an Ensemble Kalman Filter (EnKF) (e.g. Andreadis and Lettenmaier 2006, Durand and Margulis 2007) to correct (i.e. recover) the AMSR-E SWE data. The work presented does not try to correct the AMSR-E data; it only applies bivariate (correlation) and an array of multivariate (PCA and SVD) approaches to analyse the raw AMSR-E SWE data in comparison to an *in situ* data set. A preliminary analysis on the AMSR-E SWE data set is provided in Andreadis and Lettenmaier (2006). The presented work provides a more in-depth investigation on the validity of the AMSR-E snowpack product.

4.1 *Principal components analysis*

PCA is an important tool for identifying patterns in a multivariate data set (Syed *et al*. 2004). It is a widely used technique in meteorology and climatology and can be used to reduce the size of data sets without losing information (Baeriswyl and Rebetez 1997). PCA is a statistical technique that restructures a set of intercorrelated variables into an equal number of uncorrelated variables. Each new variable (principal component) is a different linear combination of the original variables, and there are no problems with multicollinearity. Figure 2 illustrates the necessary steps involved in PCA.

PCA using a varimax rotation procedure is applied in the current research to determine regions in which both data sets may behave similarly. Use of a varimax rotation results in easier interpretation of the principal component factor loadings, and a similar technique was applied in Timilsena and Piechota (2008). Based on the results of the loading matrix, the SNOTEL and AMSR-E sites that are highly correlated with the particular principal components can be identified. The relevance of PCA for the current analysis is mainly due to two reasons: (1) it can represent the variance of a scalar field with comparatively few independent coefficients, and (2) it can remove redundant variables in a multivariate data set. PCA has been extensively used in meteorological

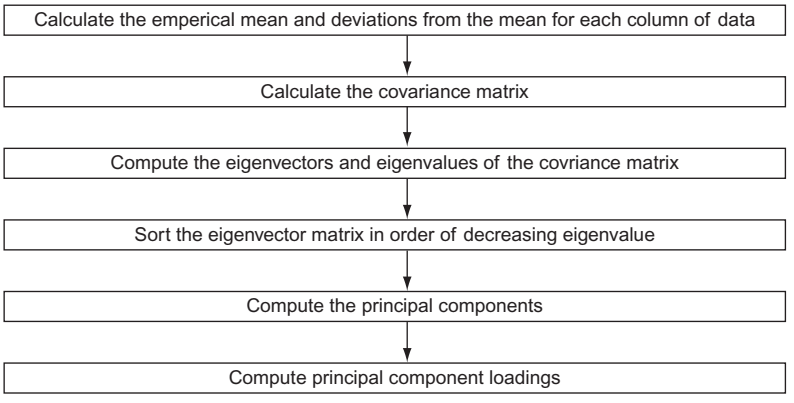

Figure 2. Flowchart illustrating the steps involved in PCA.

studies towards the establishment of the patterns, trends and modes of inter-annual to inter-decadal variability in geophysical fields (Kidson 1975, Horrel 1981, Kawamura 1994, Widmann and Schar 1997, Sengupta and Boyle 1998, Basalirwa *et al.* 1999).

The PCA method was applied to each snowpack data set separately (i.e. the PCA method was performed on the 84 SNOTEL sites and then applied to the 84 AMSR-E sites). The time period investigated was from 1 April 2003 to 2008. Retained factor loadings were investigated to identify regional patterns where SNOTEL sites and AMSR-E grid cells agree or differ. Factor loadings are on a scale of -1.0 to $+1.0$; higher values (positive or negative) signify factor representation. Loadings $>+0.90$ and <-0.90 were retained in this research. It is hypothesized that similar SWE patterns between the data sets will be discovered based on a regional scale.

4.2 *Singular value decomposition*

SVD is utilized to link both SWE data sets with streamflow in the region. It is widely known that snowpack is the primary driver streamflow in the western USA. Figure 3 shows the relationship between 1 April snowpack and streamflow in the Upper Green Basin. A single SNOTEL station and corresponding AMSR-E grid cell are plotted against streamflow (Q3) from 2003 to 2008. All plotted values are standardized. Similar relationships are found in the North Platte and Upper Colorado basins.

SVD is a powerful statistical tool for identifying coupled relationships between two spatial-temporal fields (Tootle *et al*. 2008). Bretherton *et al*. (1992) evaluated several statistical methods and concluded SVD was simple to perform and preferable for general use. Wallace *et al*. (1992) determined that SVD isolates the most important modes of variability. SVD is a widely used statistical approach to identify relationships between SSTs and hydroclimatic patterns (e.g. Uvo *et al*. 1998, Rajagopalan *et al*. 2000, Wang and Ting 2000, Shabbar and Skinner 2004). While Bretherton *et al*. (1992) provides a detailed discussion of the theory of SVD, a brief description of SVD as applied in the current research is shown in figure 4.

In this study, the data are broken down into two geographic regions (north-west and south-east), and SVD between snowpack and streamflow is performed separately in each region. The north-west region contains the Upper Green River Basin. The

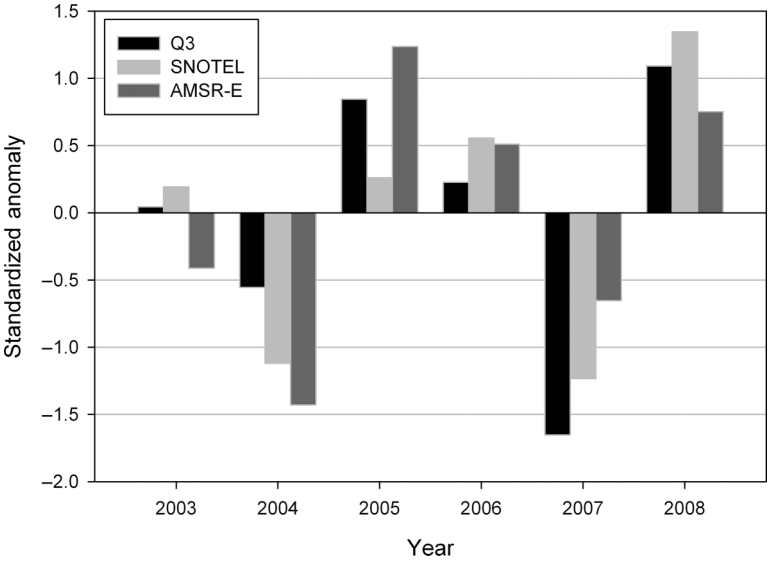

Figure 3. April–May–June–July streamflow versus 1 April snowpack (2003–2008). Standardized SWE from a single SNOTEL station and its corresponding AMSR-E grid cell are plotted against standardized streamflow (Q3).

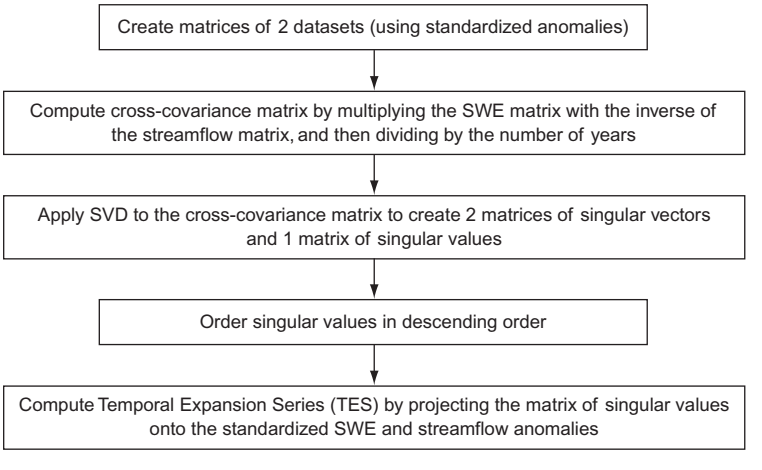

Figure 4. Flowchart illustrating the steps involved in SVD.

North Platte and Upper Colorado River Basins are within the south-east region. The north-west region contains 38 SNOTEL stations and their corresponding AMSR-E grid cells. Three streamflow stations (Q3, Q4 and Q5) are also located in the north-west region (see figure 1). The south-east contains the remaining 46 SNOTEL/AMSR-E sites along with Q1, Q2, Q6, Q7 and Q8. Similarly to Uvo *et al*. (1998), Rajagopalan *et al*. (2000) and Soukup *et al*. (2009), a significance level of 90% is applied.

4.3 *Application – streamflow forecasting*

The objective of streamflow forecasting is to predict accurately the volume of water during the high-demand season (i.e. April–July in this region). Accurate forecasts of seasonal streamflow volumes assist a broad array of natural-resource decision makers, and Water Supply Outlooks (WSOs) are currently issued jointly by the National Resource Conservation Service (NRCS 2006), the National Weather Service (NWS) and local cooperating agencies (Pagano *et al*. 2004).

Three USGS streamflow stations are forecasted in this study: Q1 in the North Platte River Basin, Q4 in the Upper Green River Basin and Q8 in the Upper Colorado River Basin. Two predictor screening steps are performed to find appropriate predictors. First, SNOTEL stations and AMSR-E cells that are $\geq$95% significant ($n = 6$, $r = 0.81$) with streamflow in the region are determined. Additionally, a search radius is applied. While Dressler *et al*. (2006) used a 200 km radius, climate variability that influences streamflow in the region was estimated to be approximately 100 km (figure 1). Therefore, all SNOTEL stations and AMSR-E grid cells that are at least 95% significant and within a 100 km radius of the streamflow station being forecasted are used as initial predictors in each forecast model.

NRCS forecasters rely on a statistical principal component regression technique to predict future streamflow using information about current SWE, precipitation, base flow and climate indices (Garen 1992). The NWS is increasingly engaged in dynamic simulations of streamflow by initializing a conceptual hydrologic model with current soil moisture and snowpack conditions (Day 1985). The presented work only incorporates SWE from the two data sources being investigated (SNOTEL and AMSR-E) as predictors for comparison purposes.

The most satisfactory and statistically rigorous way to deal with intercorrelation is the use of principal components regression (Garen 1992), and it is the method used to forecast streamflow in this study. Garen's 1992 method for selecting the principal components to include into the forecast model is followed and applied in this research. The method Garen follows is a three step process: (1) components are added to the model one at a time in sequence, beginning with the one having the largest eigenvalue and progressing in order of decreasing eigenvalue; (2) when the first component with a non-significant regression coefficient is found, the components retained are the ones in sequence up to, but not including, the non-significant one; and (3) regression coefficients must have the same algebraic sign as their correlations with the dependant variable.

Forecast validation and verification statistics calculated included R^2, R^2-adjusted, R^2-predicted (R^2_{pred}), PRESS and the Durbin–Watson statistic. R^2 measures the proportion of variation in the response that is accounted for by the predictor variables; a higher R^2 indicates a better fit of the model to the data. The R^2-adjusted statistic has an adjustment that prevents the model from appearing better simply due to adding marginally important predictor terms. The Durbin–Watson statistic was used to check for autocorrelation in residuals.

Forecast accuracy of each model (SNOTEL vs. AMSR-E) was evaluated using the R^2-R^2_{pred} statistic. R^2_{pred} is calculated from the Predicted REsidual Sums of Squares (PRESS) statistic. PRESS is based upon a leave-one-out cross-validation in which a single year or observation is removed when fitting the model. As a result, the prediction errors are independent of the predicted value at the removed observation (Garen 1992). For selecting a model when the primary interest is in prediction, the model with

the smaller PRESS (higher R^2_{pred}) is preferable (Montgomery *et al*. 2006). The idea is that if an area has poor spatial coverage of SNOTEL/weather stations, AMSR-E SWE data collected from satellites may help improve overall forecast accuracy. For the research purposes of this article (i.e. the comparison of two data sets), regions were chosen that had intensive coverage of land-based stations. When an area does not have this luxury, the use of satellite data to forecast streamflow presents an alternative option in identifying important predictor variables.

5. Results

5.1 *Bivariate statistics*

Correlation (r) values between SNOTEL SWE recorded on the ground and AMSR-E SWE recorded by satellite were determined to analyse the differences between the SWE data sets. This was accomplished by finding the AMSR-E grid cell that encompassed each SNOTEL station. Similarly to the multivariate approach applied, a 90% significance level was selected. This step investigated the bivariate similarities and differences of each data set. Furthermore, sites found to be significant were investigated in more depth (i.e. elevation) to determine possible reasons for similarities or differences between data sets.

Correlation analysis resulted in only 11 of the 84 (13%) SNOTEL sites being significant with its corresponding AMSR-E SWE grid cell at 90% from 1 April 2003 to 2008. The average elevation of all 84 SNOTEL sites used is 2737 m. In the 11 cases that had 90% significance, the average elevation was 2471 m, while the remaining 73 SNOTEL sites had an average elevation of 2777 m. Also, there were no significant sites located above 3048 m (10 000 ft), while 18 were present in the overall sample. These results conclude elevation has an effect on satellite-obtained SWE from the AMSR-E instrument and confirm the findings of Andreadis and Lettenmaier (2006). At higher elevations, snow accumulation patterns are different and the AMSR-E instrument has increased difficulty in capturing snowpack across complex terrains that contain large variations in elevation. Distribution of significant sites based on elevation is presented in table 2.

Table 3 provides first of the month (1 January 2003–1 April 2008) time-series statistics for the three regions investigated in this study. While the North Platte region contains the highest recorded SNOTEL values for all months, the highest recorded

Table 2. Distribution of significant (90%) AMSR-E grid cells with a focus on different elevation ranges.

Elevation range (m)	Total number of AMSR-E cells	Number of significant AMSR-E cells	%
2000–2200	5	2	40.0
2200–2400	9	3	33.3
2400–2600	15	2	13.3
2600–2800	18	2	11.1
2800–3000	19	2	10.5
3000+	18	0	0

Notes: Results are based on correlation analysis. We concluded that the AMSR-E instrument captures snowpack more accurately in low elevation regions compared to high elevation regions.

Table 3. Time-series breakdown comparing first of the month (1 January–1April) bivariate statistics between SNOTEL and AMSR-E data sets in the three regions from 2003 to 2008.

2003–2008 (mm)	SNOTEL	AMSR-E	SNOTEL	AMSR-E	SNOTEL	AMSR-E	SNOTEL	AMSR-E
North Platte region	Jan		Feb		Mar		Apr	
Average	247	44	347	67	454	82	541	71
Standard deviation	112	17	156	24	186	34	230	41
R^2	0.19		0.14		0.40		0.31	
Upper Green region								
Average	188	48	279	70	360	86	419	85
Standard deviation	87	20	137	26	165	34	210	57
R^2	0.29		0.30		0.33		0.37	
Upper Colorado region								
Average	183	44	259	61	341	70	403	68
Standard deviation	75	15	103	23	126	28	165	38
R^2	0.18		0.34		0.33		0.31	

AMSR-E SWE values are within the Upper Green region. Snowpack standard deviation (i.e. variability) is lowest for all months in the Upper Colorado basin. The R^2 statistic was calculated between data sets for the time series described above to detect if both data sets contained similar snowpack patterns. Based on the R^2 values, the region in which both snowpack data sets are most consistent is the Upper Green, while the least consistent snowpack region is the North Platte. It was suspected that the North Platte region SWE values may be inconsistent because it receives considerably more snowfall (i.e. based on SNOTEL) compared to the other two regions and that this 'deeper' snowpack may be causing problems for the AMSR-E instrument. However, this is not the case because R^2 values late in the snowpack season (March and April) are greater than early season R^2 values. Furthermore, for the months of January and February, R^2 values are smaller across all three of the regions compared with March and April. This result was unexpected (i.e. the opposite was expected) because it was hypothesized that the AMSR-E instrument would be better at capturing an earlier snowpack (Andreadis and Lettenmaier 2006) compared to snowpack that has accumulated throughout the season.

The complex terrain of the region may provide a possible explanation for this unexpected outcome. Snow has accumulated later in the snowpack season creating smoother underlying topography and a more consistent snowpack. This may result in fewer problems for the AMSR-E instrument to capture more accurately the snowpack depth. Results suggest that the AMSR-E instrument may experience problems when attempting to capture an early snowpack due to the reflectivity characteristics of the snow particles, while SNOTEL does not have this setback. Consequently, the SWE products have higher agreement during the later snowpack season in this region. Additionally, this unexpected result is likely due to the low quality of the AMSR-E product and the current algorithm being used to estimate SWE. This is substantiated in section 6 with regards to magnitude differences between AMSR-E and SNOTEL. This will most likely be corrected and improved with the continuing development of the AMSR-E SWE algorithm. Nonetheless, these statistics provide valuable information about the reliability of the AMSR-E data set.

5.2 *Principal components analysis*

Two regions were identified in which both data sets behave similarly (figure 5). Region 1 is in the Upper Snake/Upper Green River Basins. The first principal component factor (Region 1) retained 17 SNOTEL sites and five AMSR-E grid cells. Eight significant SNOTEL sites identified were within significant AMSR-E cells. Three of the 17 SNOTEL sites retained in factor one showed up outside of the region (i.e. anomalies). There is also one AMSR-E grid cell anomaly located outside of Region 1. One reason deviations from the trend may occur is because each principal component explains the maximum possible variance in the sample. Retaining factor loadings greater than the chosen value (i.e. 0.90) would result in fewer anomalies, while a decrease in the cut-off value to retain for factor loadings would result in more anomalies.

Region 2 is in the Upper Colorado/North Platte River Basins. The second principal component factor (Region 2) retained seven SNOTEL sites and 12 AMSR-E grid cells. Three significant SNOTEL sites identified were within a significant AMSR-E cell. Anomalies were not present in Region 2; however, the spread of AMSR-E grid cells in Region 2 is greater than that of SNOTEL sites. This suggests that AMSR-E SWE data have more inconsistency compared to SNOTEL sites in the Upper Colorado/North

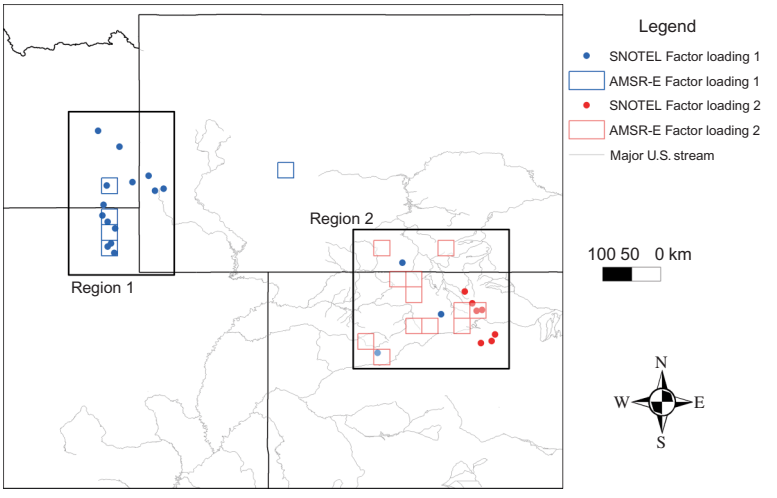

Figure 5. PCA regions map. SNOTEL stations and AMSR-E grid cells that were retained are shown. Factor loading 1 corresponds to region 1, while factor loading 2 corresponds to region 2.

Platte region. The data sets agree more in Region 1 than in Region 2, as shown by the number of AMSR-E grid cells that encompass SNOTEL sites (see figure 5). Furthermore, factor loading three contained only one significant SNOTEL site and three AMSR-E grid cells.

For the work presented, the SNOTEL network is assumed to be the best available *in situ* SWE product in the western USA due to the coverage (750+ stations) and technology (equipment used). It was expected that distinct SNOTEL regions would be identified during this analysis. Andreadis and Lettenmaier (2006) found small improvement when assimilating the AMSR-E data, and the improvement appeared mostly when the seasonal SWE was relatively low. In this study, however, the overall variability of the SNOTEL and the AMSR-E SWE data sets suggest strong regional similarities and close proximity during the peak snowpack season, as shown in figure 5.

5.3 *Singular value decomposition*

SVD results linking both SWE data sets with streamflow in each region (north-west and south-east) are shown in figure 6(*a*) (SNOTEL) and (*b*) (AMSR-E). For the north-west region, SVD found 14 SNOTEL sites that were significant with two of the three streamflow stations in the region. SVD produced no significant SNOTEL stations with Q5. Only Q3 was found to have a significant relationship with AMSR-E grid cells in the north-west region. Four of the 14 SNOTEL sites that were significant had a significant corresponding AMSRE-E grid cell with streamflow in the region.

In the south-east region, SNOTEL was found to be significant with all five streamflow stations (Q1, Q2, Q6, Q7 and Q8). AMSR-E was found to be significant with all of the streamflow stations except Q1. There were 22 significant SNOTEL sites and 18 significant AMSR-E cells. Of the 18 significant SNOTEL sites, nine of them had a corresponding AMSR-E grid cell that was also significant with streamflow in the region.

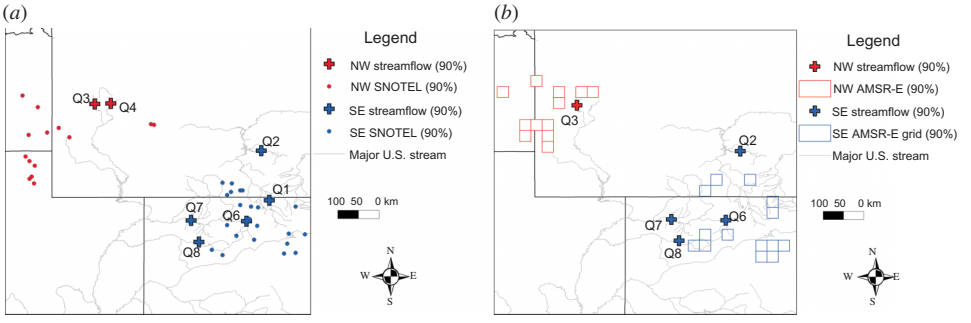

Figure 6. (*a*) SVD SNOTEL results. SNOTEL stations shown are 90% significant with the streamflow stations shown in each region. (*b*) SVD AMSR-E results. AMSR-E grid cells shown are 90% significant with the streamflow stations shown in each region.

Snowpack data acquired from SNOTEL sites are a foundation for determining stream levels in the region, and this basis was confirmed using SVD. Similarly to PCA, the results from SVD are encouraging. Results suggest there are similar relationships between both SWE data sets and streamflow in snow-driven western US watersheds. Although the SVD AMSR-E results are not nearly as strong as SNOTEL, the results proved to be better than expected.

5.4 *Streamflow forecasting*

Predictor screening resulted in a comparable number of predictors between SNOTEL sites and AMSR-E grid cells that were included in each forecast model. For Q1 forecasts, there were six predictors in the SNOTEL model and five in the AMSR-E model. Q4 and Q8 models had three SNOTEL predictors and five AMSR-E predictors after the screening process.

After performing PCA and following Garen's procedure, one principal component was significant and had the same sign as the streamflow station being forecasted for each model (SNOTEL and AMSR-E). Forecast accuracy was determined using the R^2_{pred} statistic. The forecast accuracy for Q1 and Q8 are comparable for each model. For Q1, the SNOTEL model produced an R^2_{pred} of 0.71, while the AMSR-E model's R^2_{pred} was 0.79. For Q8, the SNOTEL model's R^2_{pred} was 0.96 compared to 0.91 for AMSR-E. When forecasting Q4, the AMSR-E model produced a more accurate forecast ($R^2_{\text{pred}} = 0.90$) compared to 0.41 for the SNOTEL model. Statistical parameters determined for all forecast models are provided in table 4.

Comparing *in situ* and satellite SWE data sets was the primary purpose of the presented work. Regions that have intensive coverage of land-based stations are ideal and investigated in this study. The global coverage of collecting hydrologic parameters with satellites is a great improvement compared to point coverage. Based on the results presented, regions that have few (or zero) land-based weather stations can incorporate the AMSR-E SWE product in a streamflow forecast model and find success.

6. Discussion of magnitude differences among SWE data sets

The AMSR-E SWE data set has a range of 0–480 mm (0–19 in). In many instances, SNOTEL sites record SWE values that exceed this range within the region (see table 3).

Table 4. Streamflow forecast statistics for the North Platte, Upper Green and Upper Colorado River basins.

	Q1 – North Platte		Q4 – Upper Green		Q8 – Upper Colorado	
	SNOTEL	AMSR-E	SNOTEL	AMSR-E	SNOTEL	AMSR-E
R^2	0.83	0.90	0.89	0.97	0.98	0.95
R^2-adjusted	0.79	0.88	0.86	0.97	0.98	0.93
R^2-predicted	0.71	0.79*	0.41	0.90*	0.96*	0.91
PRESS (10^3 km^3)	15.8	11.0	0.6	0.1	1.1	2.6
Durbin–Watson	3.0	1.5	3.3	1.4	1.8	0.9

Notes: *The preferred model based on R^2_{pred}.
Forecast accuracy was measured using the R^2_{pred} parameter.

For the 1 April SWE data analysed in this study, and for at least one year from 2003 to 2008, 47 of the 84 (56%) SNOTEL sites included in this study record SWE values that exceeded the maximum AMSR-E value (>480 mm). However, there were no cases in which the AMSR-E recorded the maximum value of 480 mm. Figure 7(*a*)–(*c*) provides a box plot of first of the month (January–April) SWE values for each basin from 2003 to 2008. It is clear there are significant differences in magnitude recorded by SNOTEL and AMSR-E. Additionally, there is significantly more variability in the SNOTEL data set.

Magnitude issues limit the use of AMSR-E SWE to a narrow list of applications. In many statistical models, the magnitude of the predictor variables is irrelevant (i.e. magnitude does not affect the results due to the restructuring of variables), and this data set would be sufficient to use for analysis (i.e. streamflow forecasting). The statistical approaches utilized in this study were used in place of normalization techniques. Correlation analysis and the multivariate approaches applied within this study do not take into account absolute differences in magnitude, only relative differences. While PCA restructured the set of SWE variables, SVD used standardized anomalies of each data set. However, when integrating AMSR-E SWE in a non-multivariate-based regression model, the magnitude of the data set would impact the model's overall skill because extreme high events would not be fully captured.

Incorporating AMSR-E SWE into a physical model would result in an inaccurate basin-wide hydrologic representation, and is therefore not recommended without the use of correction factors. The development of correction factors to apply to AMSR-E SWE values is needed because this would provide a basis for incorporating AMSR-E gridded satellite SWE into physically based climate models. The work of Durand and Margulis (2007) demonstrated the application of the EnKF framework to merge synthetic remote-sensing measurements at multiple scales and frequencies with a land surface model in order to characterize SWE over the course of the accumulation season. They were successful and recovered true basin-wide SWE within a root-mean-square error (RMSE) of approximately 2 cm. Although satellite-based remote-sensing approaches show promise in this study, as well as Skofronick-Jackson *et al*. (2004) and Bindschadler *et al*. (2005), *in situ* precipitation gauges still form the primary source for estimating snowfall. The development of algorithms to map SWE is an evolutionary process, and the findings of this study provide valuable information to help increase the accuracy of the AMSR-E SWE product.

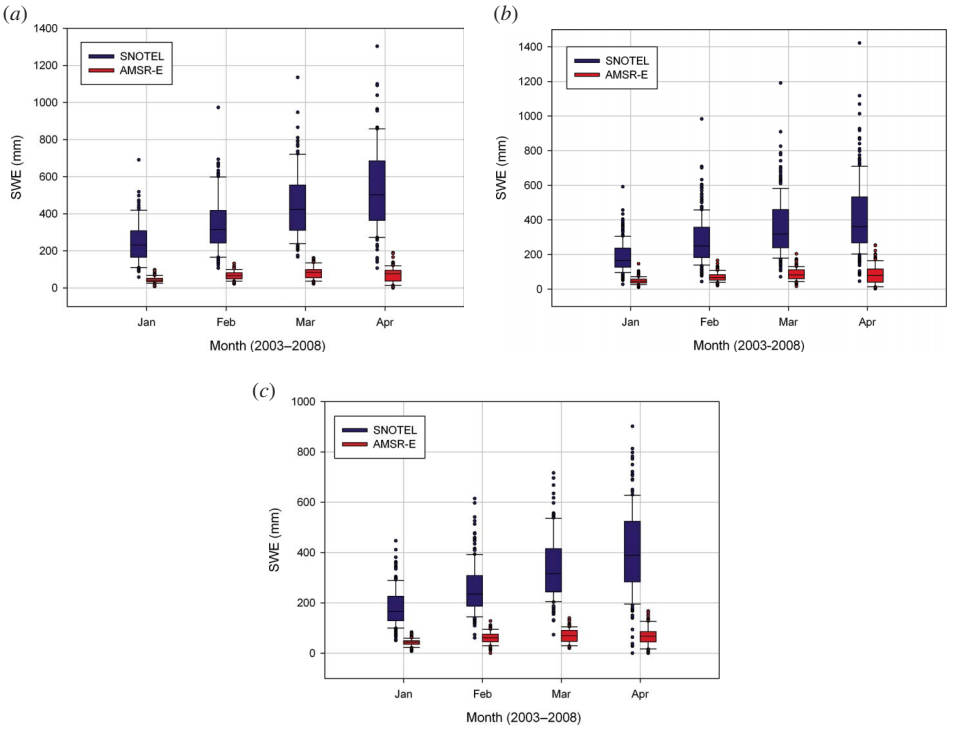

Figure 7. (*a*) Box plot of North Platte first of the month SWE values from 2003 to 2008 showing the significant difference in SWE magnitude between SNOTEL and AMSR-E data sets. (*b*) Box plot of Upper Green first of the month SWE values from 2003 to 2008 showing the significant difference in SWE magnitude between SNOTEL and AMSR-E data sets. (*c*) Box plot of Upper Colorado first of the month SWE values from 2003 to 2008 showing the significant difference in SWE magnitude between SNOTEL and AMSR-E data sets.

7. Conclusions

A thorough comparison has been made between SNOTEL and AMSR-E SWE data products. Statistical techniques applied include correlation, PCA and SVD. The relationship between SNOTEL and AMSR-E data sets was greater utilizing multivariate statistical methods compared to bivariate approaches. Bivariate approaches concluded the SWE data sets agreed more in low elevation areas and later in the snowpack season and significant magnitude differences between data sets were discovered. PCA utilizing a varimax rotation identified two distinct snowpack regions in which both data sets were consistent and compared well with one another. Additionally, similar relationships between both SWE data sets and streamflow in the region were found utilizing SVD. The incorporation of both SWE products was applied to forecast regional streamflow. Forecast results were comparable for two of the three streamflow stations (North Platte and Upper Colorado); however, the AMSR-E SWE data set produced a more skilful forecast compared to SNOTEL for the Upper Green station. SWE from NASA's Aqua satellite was found to be sufficient to use in statistically based forecast models in which magnitude did not affect results. Given the projections of climate change (e.g. increased temperatures and decreased snowpack), the ability of satellites to capture important hydrologic parameters such as SWE is vital in many earth science

applications. Incorporating additional satellite parameters such as soil moisture and vegetative cover, in addition to SWE, may be in the near future for next-generation climatologists and forecasters.

Acknowledgements

This research is sponsored by the Wyoming Water Development Commission, the USGS/State of Wyoming Water Research Program, and the University of Wyoming Office of Water Programs. Additional funding was provided by the University of Tennessee and the Oak Ridge National Laboratory JDRD program.

References

Andreadis, K.M. and Lettenmaier, D.P., 2006, Assimilating remotely sensed snow observations into a macroscale hydrology model. *Advances in Water Resources*, **29**, pp. 872–886.

Baeriswyl, P.A. and Rebetez, M., 1997, Regionalization of precipitation in Switzerland by means of principal component analysis. *Theoretical and Applied Climatology*, **58**, pp. 31–41.

Barton, M. and Burke, M., 1977, SNOTEL: an operational data acquisition system using meteor burst technology. In *Proceedings of the 45th Annual Western Snow Conference*, 18–21 April 1977, Albuquerque, NM (Fort Collins, CO: Colorado State University), pp. 75–87.

Basalirwa, C.P.K., Odiyo, J.O., Mngodo, R.J. and Mpeta, E.J., 1999, The climatological regions of Tanzania based on the rainfall characteristics. *International Journal of Climatology*, **19**, pp. 69–80.

Bindlish, R., Crow, W.T. and Jackson, T.J., 2009, Role of passive microwave remote sensing in improving flood forecasts. *IEEE Geoscience and Remote Sensing Letters*, **6**, pp. 112–116.

Bindschadler, R., Choi, H., Shuman, C. and Markus, T., 2005, Detecting and measuring new snow accumulation on ice sheets by satellite remote sensing. *Remote Sensing of Environment*, **98**, pp. 388–402.

Bretherton, C.S., Smith, C. and Wallace, J.M., 1992, An intercomparison of methods for finding coupled patterns in climate data. *Journal of Climate*, **5**, pp. 541–560.

Brodzik, M.J. and Knowles, K., 2002, EASE-Grid: a versatile set of equal-area projections and grids. In *Discrete Global Grids*, M. Goodchild and A.J. Kimerling (Eds.). Available online at: http://www.ncgia.ucsb.edu/globalgrids-book/ease_grid/ (accessed 22 February 2009).

Cavalieri, D.J., Markus, T., Hall, D.K., Gasiewski, A.J., Klein, M. and Ivanoff, A., 2006, Assessment of EOS aqua AMSR-E arctic sea ice concentrations using Landsat-7 and airborne microwave imagery. *IEEE Transactions on Geoscience and Remote Sensing*, **44**, pp. 3057–3069.

Chang, A.C., Foster, J.L. and Hall, D.K., 1987. Nimbus-7 SMMR derived global snow cover parameters. *Annals of Glaciology*, **9**, pp. 39–44.

Chang, T.C. and Rango, A., 2000, Algorithm theoretical basis document (ATBD) for the AMSR-E snow water equivalent algorithm. Available online at: http://nsidc.org/data/docs/daac/ae_swe_ease_grids/atbd-amsr-snow.pdf (accessed 18 February 2009).

Comiso, J.C. and Steffen, K., 2008, Introduction to special section on large-scale characteristics of the sea ice cover from AMSR-E and other satellite sensors. *Journal of Geophysical Research*, **113**, p. C02S01.

Daniels, J.M., 2007, Flood hydrology of the North Platte River headwaters in relation to precipitation variability. *Journal of Hydrology*, **344**, pp. 70–81.

Day, G.N., 1985, Extended streamflow forecasting using NWSRFS. *Journal of Water Resources Planning and Management*, **111**, pp. 157–170.

Dressler, K.A., Leavesley, G.H., Bales, R.C. and Fassnacht, S.R., 2006, Evaluation of gridded snow water equivalent and satellite snow cover products for mountain basins in a hydrologic model. *Hydrological Processes*, **20**, pp. 673–688.

Durand, M. and Margulis, S.A., 2007, Correcting first-order errors in snow water equivalent estimates using a multifrequency, multiscale radiometric data assimilation scheme. *Journal of Geophysical Research*, **112**, p. D13121.

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

Horrel, J.D., 1981, A rotated principal component analysis of the interannual variability of the Northern Hemisphere 500 mb height field. *Monthly Weather Review*, **109**, pp. 2080–2092.

Kawamura, R., 1994, A rotated EOF analysis of global sea surface temperature variability with interannual and interdecadal scales. *Journal of Physical Oceanography*, **24**, pp. 707–715.

Kelly, R.E., Chang, A.T., Tsang, L. and Foster, J.L., 2003, A prototype AMSR-E global snow area and snow depth algorithm. *IEEE Transactions on Geoscience and Remote Sensing*, **41**, pp. 6075–6084.

Kelly, R.E.J., Chang, A.T.C., Foster, J.L. and Tedesco, M., 2004, Updated every five days. AMSR-E/Aqua 5-day 3 Global Snow Water Equivalent EASE-Grids V002, [2003–2008]. Boulder, CO: National Snow and Ice Data Center. Digital media.

Kidson, W.J., 1975, Eigenvector analysis of monthly mean surface data. *Monthly Weather Review*, **103**, pp. 107–186.

King, M.D., Kaufman, Y.J., Menzel, W.P., and Tanre, D., 1992. Remote-sensing of cloud, aerosol, and water-vapor properties from the moderate resolution imaging spectrometer (MODIS). *IEEE Transactions on Geoscience and Remote Sensing*, **30**, pp. 2–27.

Montgomery, D.C., Peck, E.A., Vining, G.G. and Myers, R.H., 2006, *Introduction to Linear Regression Analysis*, 4th ed., pp. 141–142 (Hoboken, NJ: Wiley).

Narayan, U. and Lakshmi, V., 2008, Characterizing subpixel variability of low resolution radiometer derived soil moisture using high resolution radar data. *Water Resources Research*, **44**, p. W06425.

Natural Resources Conservation Service (NRCS), 2006, Snow surveys and water supply forecasting. Available online at: http://www.wcc.nrcs.usda.gov/factpub/aib536.html (accessed 3 March 2009).

Natural Resources Defense Council (NRDC), 2008, Hotter and drier: the West's changed climate. Available online at: http://www.nrdc.org/globalWarming/west/contents.asp (accessed 26 January 2009).

Pagano, T.C., Garen, D.C. and Sorooshian, S., 2004, Evaluation of official Western US seasonal water supply outlooks, 1922–2002. *Journal of Hydrometeorology*, **5**, pp. 683–698.

Rajagopalan, B., Cook, E., Lall, U. and Ray, B.K., 2000, Spatiotemporal variability of ENSO and SST teleconnections to summer drought over the United States during the twentieth century. *Journal of Climate*, **13**, pp. 4244–4255.

Reichle, R.H., Koster, R.D., Liu, P., Mahanama, S.P.P., Njoku, E.G. and Owe, M., 2007, Comparison and assimilation of global soil moisture retrievals from the Advanced Microwave Scanning Radiometer for the Earth Observing System (AMSR-E) and the Scanning Multichannel Microwave Radiometer (SMMR). *Journal of Geophysical Research*, **112**, p. D09108.

Reynolds, R.W., Smith, T.M., Liu, C., Chelton, D.B., Casey, K.S. and Schlax, M.G., 2007, Daily high resolution blended analyses for sea surface temperature. *Journal of Climate*, **20**, pp. 5473–5496.

Sengupta, S. and Boyle, J.S., 1998, Using principal components for comparing GCM simulations. *Journal of Climate*, **11**, pp. 816–830.

Shabbar, A. and Skinner, W., 2004, Summer drought patterns in Canada and the relationship to global sea surface temperatures. *Journal of Climate*, **17**, pp. 2866–2880.

Skofronick-Jackson, G., Kim, M., Weinman, J. and Chang, D., 2004, A physical model to determine snowfall over land by microwave radiometry. *IEEE Transactions on Geoscience and Remote Sensing*, **42**, pp. 1047–1058.

Slack, J.R. and Landwehr, J.M., 1992, Hydro-Climatic Data Network: a US Geological Survey streamflow data set for the United States for the study of climate variations, 1874–1988. USGS Open-File Report, pp. 92–129, 193.

Soukup, T.L., Aziz, O.A., Tootle, G.A., Piechota, T.C. and Wulff, S.S., 2009, Long lead-time streamflow forecasting of the North Platte River incorporating oceanic-atmospheric climate variability. *Journal of Hydrology*, **368**, pp. 131–142.

Syed, T.H., Lakshmi, V., Paleologos, E., Lohmann, D., Mitchell, K. and Famiglietti, J.S., 2004, Analysis of process controls in land surface hydrological cycle of the continental United States. *Journal of Geophysical Research-Atmospheres*, **109**, p. D22105.

Timilsena, J. and Piechota, T., 2008, Regionalization and reconstruction of snow water equivalent in the upper Colorado River basin. *Journal of Hydrology*, **352**, pp. 94–106.

Tootle, G.A., Piechota, T.C. and Gutierrez, F., 2008, The relationships between Pacific and Atlantic Ocean sea surface temperatures and Colombian streamflow variability. *Journal of Hydrology*, **349**, pp. 268–276.

Uvo, C.B., Repellii, C.A., Zebiak, S.E. and Kushnir, Y., 1998, The relationship between tropical Pacific and Atlantic SST and northeast Brazil monthly precipitation. *Journal of Climate*, **11**, pp. 551–562.

Wallace, J.M., Gutzler, D.S. and Bretherton, C.S., 1992, Singular value decomposition of wintertime sea surface temperature and 500-mb height anomalies. *Journal of Climate*, **5**, pp. 561–576.

Wang, H. and Ting, M., 2000, Covariabilities of winter US precipitation and Pacific sea surface temperatures. *Journal of Climate*, **13**, pp. 3711–3719.

Widmann, M. and Schar, C., 1997, A principal component and long-term trend analysis of daily precipitation in Switzerland. *International Journal of Climatology*, **17**, pp. 1333–1356.

Woodhouse, C.A., 2003, A 431-yr reconstruction of western Colorado snowpack from tree rings. *Journal of Climate*, **16**, pp. 1551–1561.

Zarco-Tejada, P.J., Rueda, C.A. and Ustin, S.L., 2003. Water content estimation in vegetation with MODIS reflectance data and model inversion methods. *Remote Sensing of Environment*, **85**, pp. 109–124.