HYDROLOGICAL PROCESSES
Hydrol. Process. **24**, 1285–1295 (2010)
Published online 26 January 2010 in Wiley InterScience
(www.interscience.wiley.com) DOI: 10.1002/hyp.7590

Spatial interpolation of snow water equivalency using surface observations and remotely sensed images of snow-covered area

Brian J. Harshburger,* Karen S. Humes, Von P. Walden, Troy R. Blandford, Brandon C. Moore and Raymond J. Dezzani

Department of Geography, University of Idaho, Moscow, ID 83844-3021, USA

Abstract:

As demand for water continues to escalate in the western Unites States, so does the need for accurate monitoring of the snowpack in mountainous areas. In this study, we describe a simple methodology for generating gridded-estimates of snow water equivalency (SWE) using both surface observations of SWE and remotely sensed estimates of snow-covered area (SCA). Multiple regression was used to quantify the relationship between physiographic variables (elevation, slope, aspect, clear-sky solar radiation, etc.) and SWE as measured at a number of sites in a mountainous basin in south-central Idaho (Big Wood River Basin). The elevation of the snowline, obtained from the SCA estimates, was used to constrain the predicted SWE values. The results from the analysis are encouraging and compare well to those found in previous studies, which often utilized more sophisticated spatial interpolation techniques. Cross-validation results indicate that the spatial interpolation method produces accurate SWE estimates [mean $R^2 = 0{\cdot}82$, mean mean absolute error (MAE) = 4·34 cm, mean root mean squared error (RMSE) = 5·29 cm]. The basin examined in this study is typical of many mid-elevation mountainous basins throughout the western United States, in terms of the distribution of topographic variables, as well as the number and characteristics of sites at which the necessary ground data are available. Thus, there is high potential for this methodology to be successfully applied to other mountainous basins. Copyright © 2010 John Wiley & Sons, Ltd.

KEY WORDS snow water equivalency; multivariate regression; snow-covered area; spatial distribution; water resources; forecasting

Received 3 July 2007; Accepted 20 November 2009

INTRODUCTION

In the mountainous regions of the western United States, approximately 50–70% of the annual precipitation falls as snow during the winter season and contributes to runoff during the spring (Serreze *et al*., 1999). Consequently, knowledge of the magnitude of the seasonal snowpack [snow water equivalency (SWE)] is important in predicting water availability (runoff). In addition, spatially distributed (gridded) estimates of SWE are important because some areas (e.g. portions of watersheds) contribute more runoff than others (Balk and Elder, 2000). These spatial estimates also serve as an important input for distributed, physically based snowmelt models (Bloschl *et al*., 1991; Luce *et al*., 1998).

Recent emphasis in snow distribution modelling has focussed on statistical relationships between snow properties (i.e. depth and SWE) and terrain characteristics (Balk and Elder, 2000). Elevation, slope, aspect, and other landscape characteristics have been used successfully to model the spatial distribution of SWE (Woo and Marsh, 1978; Elder *et al*., 1991). In recent decades, remote sensing has also been used to produce spatially distributed estimates of SWE (McManamon *et al*., 1993; Cline *et al*., 1998; Wilson *et al*., 1999; Mote *et al*., 2003). However, these estimates are often expensive, difficult to obtain, and are not suitable for operational use in mountainous terrain (Elder *et al*., 1998). The interpolation of ground-based point measurements, therefore, becomes necessary to explain and understand the spatial distribution of SWE (Balk and Elder, 2000). This has led to the use of statistical and/or geostatistical methods to generate spatially distributed values of SWE.

At present, the most widely used ground-based estimates of SWE come from snow course and SNOwpack TELemetry (SNOTEL) sites. Interpolated SWE values, using these data, have been obtained using multivariate linear regression (Chang and Li, 2000) and by regressing SWE with elevation (Daly *et al*., 2000). These studies, however, did not take into account the presence or absence of snow cover, which can lead to interpolated SWE values (>0) in areas where snow is non-existent. As a result, remotely sensed estimates of snow-covered area (SCA) have been used, along with simple regression techniques, to obtain gridded SWE estimates (Fassnacht *et al*., 2003; Molotch *et al*., 2004). There are two types of SCA data products, fractional and binary. Fractional mapping techniques used by Fassnacht *et al*. (2003) and Molotch *et al*. (2004) provide information regarding the

* Correspondence to: Brian J. Harshburger, Department of Geography, University of Idaho, Moscow, ID 83844-3021, USA.
E-mail: bharshburger@vandals.uidaho.edu

percentage (ranging from 0% to 100%) of each pixel that is covered with snow. Binary mapping techniques, on the other hand, use a set of thresholds to determine whether a pixel contains snow (SCA = 100%) or no snow (SCA = 0%) (Elder *et al*., 1998). Although fractional SCA data are more desirable from a spatial resolution standpoint, binary data have been available on an operational basis for a much longer time and have been more widely validated. Salomonson and Appel (2004) developed a simple regression approach for fractional snow products, which was ultimately used in version 5 of the Moderate resolution Imaging Spectroradiometer (MODIS) operational snow cover products released by the National Snow and Ice Data Center (NSIDC) in 2007. These data, however, were not used in this study because the data had not yet been widely validated. We attempted to validate these data for our study area, however, found the data to be inconsistent and to contain undocumented values. Given the maturity of the fractional versus binary products at this time, the priority in this study was the assessment of a new methodology for the generation of spatially distributed SWE estimates using binary SCA data.

In recent years, binary regression trees have been used successfully to obtain interpolated SWE values from surface observations (Elder *et al*., 1998; Balk and Elder, 2000; Erxleben *et al*., 2002; Winstral *et al*., 2002; Molotch *et al*., 2005; Molotch and Bales, 2006), however, this method requires large amounts of data and is better suited for small basins (Fassnacht *et al*., 2003).

Kriging and cokriging methods have also been used successfully to create spatially distributed SWE estimates (Carroll and Cressie, 1996; Ling *et al*., 1996; Erxleben *et al*., 2002; Molotch *et al*., 2005). In addition, these methods have been used to interpolate the residuals obtained using other statistical techniques (Hosang and Dettwiler, 1991; Balk and Elder, 2000; Erxleben *et al*., 2002; Molotch *et al*., 2005). Kriging and cokriging methods, however, may be inappropriate if the data (or residuals) lack spatial dependence (Balk and Elder, 2000). Furthermore, it is often difficult to fit a suitable model (i.e. linear and Gaussian) to the semivariogram as a result of the temporal and spatial variability in SWE data (Fassnacht *et al*., 2003).

The objective of the work presented here is to evaluate the performance of a simple multiple linear regression technique to generate gridded estimates of SWE, utilizing data products that are widely available for most mountainous watersheds in the western United States (snow course and SNOTEL observations of SWE, binary remotely sensed estimates of SCA, and physiographic variables). Cross-validation is used to assess the model performance in an example watershed in the intermountain west.

STUDY AREA

This study was conducted in the Big Wood River Basin located in south-central Idaho (Figure 1). Only a subsection of the basin, located upstream of all dams and diversions, was used in the analysis. The study area has a total area of 1625 km^2, with a minimum elevation of 1615 m and a maximum elevation of 3628 m. Approximately 46% of the study area is covered with grassland, shrubland, and agricultural areas. Deciduous, evergreen, and mixed forests cover approximately 43% of the study area; the remaining 11% is covered by bare rock (10%), wetlands (<1%), and urban/residential areas (<1%). The study area is surrounded by the Smoky Mountains to the west and the Boulder and Pioneer Mountains to the north and east. Draining south into the Snake River, runoff from the basin provides a source of water for the agricultural industry. The annual average precipitation in the Big Wood River Basin varies from 26 cm in the lower elevation areas [Idaho Department of Environmental Quality (IDEQ), 2001] to approximately 114 cm in

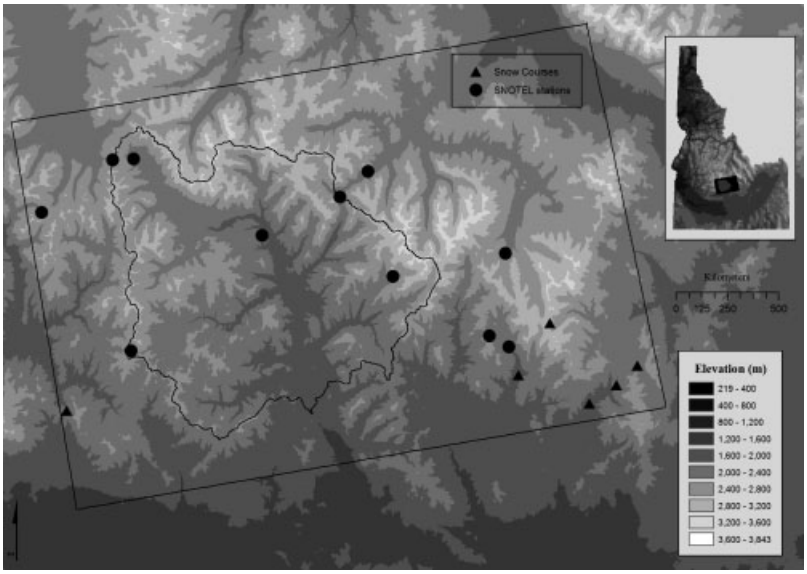

Figure 1. Snow course (triangles) and SNOTEL (dots) sites located in and around the Big Wood River Basin, Idaho

the higher elevation areas. About 60% of the basin's precipitation comes in the form of snow during the months of November–March (IDEQ, 2001).

DATA

SWE data

The Natural Resources Conservation Service (NRCS) has conducted snow surveys in the western United States since 1935 (Chang and Li, 2000). Snow course measurements are generally taken on or near the first day of every month during the snow accumulation and melt seasons (January–June). Measurements are taken most frequently on March 1 and April 1, however, the frequency of the measurements is influenced by the nature of the snowpack (snowpack condition), difficulty of access, and costs (e.g. labour, supplies, etc.). SWE data are collected at snow courses using a snow tube and are subjected to systematic bias because the measurements tend to underestimate SWE due to sampling difficulties associated with ice lenses, ground ice, and depth hoar (Carroll, 1987). The size of the sampling cutter can also influence accuracy (Goodison *et al*., 1981).

In 1977, the NRCS developed the SNOTEL data network. The SNOTEL network provides SWE measurements at 83 stations across the state of Idaho. Each SNOTEL data station is equipped with automatic measuring and remote communication devices (Chang and Li, 2000). The SNOTEL network provides hourly data; however, daily SWE values were used in this study. In comparison to snow course data, SNOTEL and SWE measurements are obtained automatically using snow pillows (Johnson and Schaefer, 2002; Johnson, 2004; Dressler *et al*., 2006). Snow pillows (typically 0·37 m^2 in area) are envelopes of stainless steel or synthetic rubber (hypalon), which contain an antifreeze solution. As the snow accumulates on the pillow, pressure is exerted on the antifreeze solution (caused by the weight of the overlying snowpack). Automatic measuring devices are then used to convert the weight of the snowpack into an electrical reading of the SWE (Johnson and Schaefer, 2002). There are several sources of error in snow pillow measurements, however, most errors are generally attributed to snow bridging, which occurs when the snow over a snow pillow is partially or fully supported by the surrounding snow or, conversely, when the snow pillow supports some of the weight of the surrounding snow (Johnson and Schaefer, 2002). In addition, errors are common during the transition from subfreezing snow temperatures to an isothermal snow cover with a temperature of 0 °C (CDWR, 1976; Johnson, 2004). The SNOTEL data, used in this study, are quality controlled, but there is no official estimate of the error by the NRCS. Consultation with snow survey personnel, familiar with the collection of data in Idaho, estimates the error at 6–8% of the measured SWE.

SWE data collected at 11 SNOTEL stations and 6 snow courses, located within and around the study area, are used in this analysis (Figure 1). A list of the stations and their attributes can be found in Table I. Stations were selected to maximize at both the spatial and elevational coverage. SWE data collected on March 1 and April 1 for 3 years (2003–2005) were used in the analysis. These dates (March 1 and April 1) were selected because snow course SWE measurements are usually taken on these dates.

SCA data

The MOD10A1 operational MODIS daily snow cover product, obtained from the NSIDC, is used to constrain the spatial interpolation of SWE to the boundary of the SCA. The MODIS acquires images in 36 spectral bands between 0·405 and 14·385 μm. In the MODIS snow mapping algorithm, snow is distinguished from other surfaces using the normalized difference snow index (NDSI) (Klein and Barnett, 2003). NDSI values are calculated using MODIS bands 4 (0·545–0·5·65 μm) and 6 (1·628–1·652 μm) from the following equation:

$$\mathrm{NDSI} = \left(\frac{\rho_4 - \rho_6}{\rho_4 + \rho_6}\right) \tag{1}$$

where ρ_4 and ρ_6 are reflectance values in bands 4 and 6. The MOD10A1 snow cover product is generated daily for most locations on the Earth's surface at a spatial resolution of 500 m. SCA data are obtained for the same dates as the SWE data.

Table I. Attributes of the snow course and SNOTEL stations used in the analysis. Stations are listed by elevation from lowest to highest

Station	Type	Elevation (m)	Latitude (°N)	Longitude (°W)
Telfer Ranch	Snow course	1780	43·53	113·77
Iron Mine Creek	Snow course	1921	43·56	113·72
Muldoon	Snow course	1927	43·57	113·91
Chocolate Gulch	SNOTEL	1963	43·77	114·42
Garfield R.S.	SNOTEL	1999	43·61	113·93
Couch Summit	Snow course	2085	43·52	114·80
Dry Fork	Snow course	2201	43·58	113·68
Stickney Mill	SNOTEL	2265	43·86	114·21
Galena Pillow	SNOTEL	2268	43·88	114·67
Hyndman	SNOTEL	2268	43·71	114·16
Swede Peak	SNOTEL	2329	43·62	113·97
Bear Canyon	SNOTEL	2408	43·74	113·94
Lost-wood Divide	SNOTEL	2408	43·82	114·26
Dollarhide Summit	SNOTEL	2566	43·60	114·67
Galena Summit	SNOTEL	2676	43·87	114·71
Vienna Mine	SNOTEL	2731	43·80	114·85
Fishpole Lake	Snow course	2835	43·64	113·85

METHODS

The snowline elevation is determined using the MODIS snow cover grids and a digital elevation model (DEM). The DEM (30 m spatial resolution) was obtained from the United States Geological Survey (USGS). Since the snowline elevation tends to be highly variable, the median snowline is used in this study. This methodology is consistent with the World Meteorological Organization's (WMO) definition, which states that the snowline is the elevation with approximately 50% snow coverage (Seidel *et al*., 1997; Kleindienst *et al*., 2000). To compute the snowline elevation, the snow cover grid had to first be converted to a point file. Each point then represents the middle of each grid cell. The elevation for each point was then determined by sampling (overlaying) the DEM. Next, the snow cover point file was split into two separate point files, one for snow-covered points and the other for non-snow-covered points. A 600-m buffer was then created around each snow-covered point and all non-snow-covered points that fell within those buffers were selected. This buffer size was selected because the snow cover imagery had a spatial resolution of 500 m and in order to select adjacent cells the buffer size needed to be larger than this value. Finally, the median snowline elevation was determined using the elevations of those selected points. This method was used individually for each snow cover image.

After the snowline elevation is determined, the SWE values are plotted against elevation. Linear regression is then used to fit a line to the data, and the residuals are obtained. Figure 2 shows a sample regression fit for data from the Big Wood River Basin. This process is used to remove the elevation trend in the SWE data. The residuals are then tested for spatial dependence using Moran's I analysis.

Moran's I is one of the most common measures of spatial autocorrelation (dependence) and is used to describe the overall spatial relationship between individual samples (Moran, 1950). Values for the Moran's I typically vary between -1 and 1. A Z-value is calculated to determine if the calculated I values are significant. Z values greater than 1·96 indicate that there is significant clustering of the data ($p = 0{\cdot}05$). On the other hand, values below $-1{\cdot}96$ indicate that there is a uniform distribution ($p = 0{\cdot}05$). Z values between $-1{\cdot}96$ and 1·96 indicate that there is a random distribution of values (complete spatial randomness). For this analysis, inverse distance weighting squared (IDW^2) is used to determine the spatial weights for the Moran's I analysis. Geoda, which is a widely used spatial statistics software program, was used to calculate the Moran's I values.

The results from the Moran's I analysis (Table II) indicate that there was no significant spatial dependence in the residuals for any of the 6 days tested (i.e. complete spatial randomness). As a result, spatial interpolation methods that rely on spatial dependence in the sample data (i.e. kriging, inverse distance weighting, etc.) were not used to interpolate the residual values.

Due to the lack of spatial dependence in the detrended residuals, two regression equations were used to model

Table II. Results from the Moran's I analysis for the 6 days analysed for the Big Wood Basin

Date	Moran's I	Z value
1March 2003	−0·04	0·0
1 March 2004	−0·09	−0·2
1 March 2005	0·13	0·8
1 April 2003	−0·12	−0·1
1 April 2004	−0·55	−1·0
1 April 2005	0·21	0·8

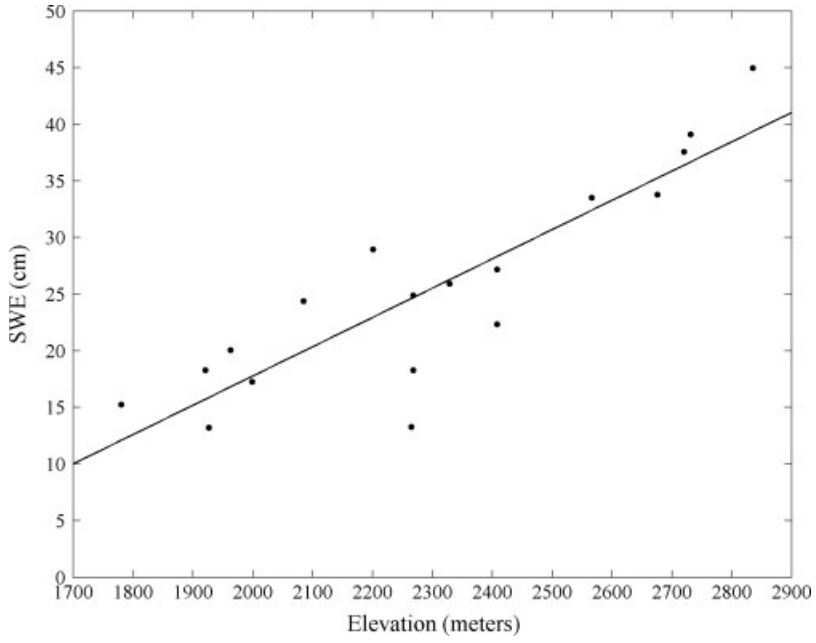

Figure 2. Example plot of SWE values (snow course and SNOTEL) plotted versus elevation for the Big Wood River Basin

the SWE values located above the snowline. Linear regression (elevation vs SWE) is used to estimate the SWE values between the snowline and the elevation of the station located directly above it. This step is completed to ensure that the interpolated SWE values approach 0 cm at the snowline elevation. Stepwise multivariate linear regression is then used to interpolate the SWE values at elevations higher than the station located directly above the snowline. The snowline was not included in the multivariate regression because it only has one physiographic attribute, i.e. elevation. The stepwise multivariate linear regression was conducted using SPSS, version 13·0.

The independent variables used in the multivariate linear regression are: elevation, slope, degrees of northness and eastness, ruggedness, latitude, longitude, clear-sky solar radiation, and vegetation cover type. Using ArcGIS 9·0, the topographic variables, elevation (Figure 3a) and slope (Figure 3b), are obtained from the 30-m DEM. In addition to the elevation and slope, an aspect was also derived from the DEM and was used to compute the degrees of northness and eastness. The degree of northness (Figure 3c) is defined as the cosine of the aspect. Values ranged from -1 to 1, with a value of -1 representing a slope facing directly south and a value of 1 indicating a slope facing directly north. The degree of eastness (Figure 3d) is defined as the sine of the aspect. Values also ranged from -1 to 1, with -1 representing an east-facing slope and 1 signifying a west-facing slope.

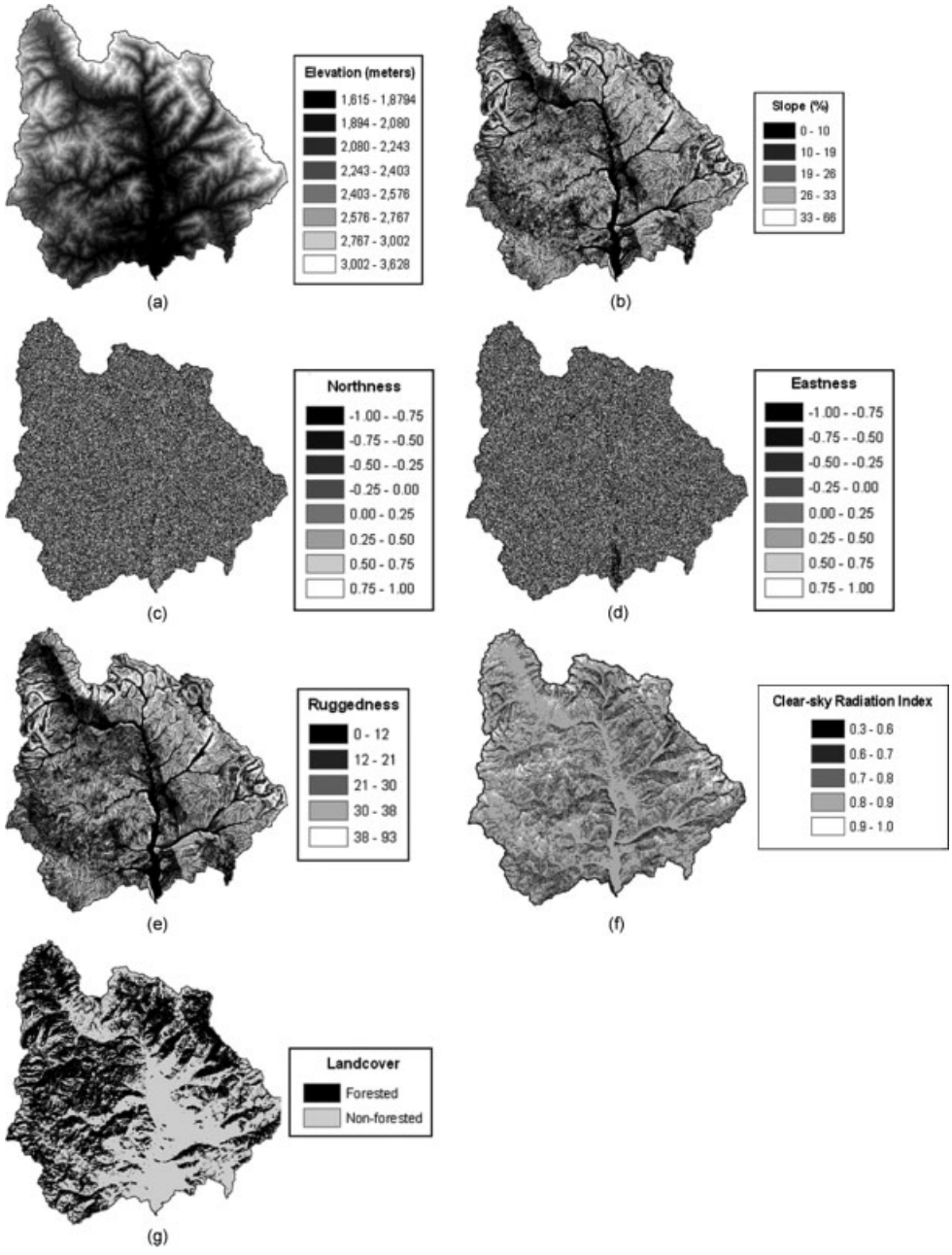

Figure 3. Physiographic variables used in the multivariate linear regression analysis: (a) elevation (m), (b) slope (%), (c) degree of northness, (d) degree of eastness, (e) ruggedness, (f) clear-sky radiation index values, and (g) land cover

The ruggedness (Figure 3e) variable is defined as the standard deviation of the elevation within a 0·21 km by 0·21 km window. High ruggedness values are indicative of areas with a large range in elevation. Small values are found in areas where the elevation is relatively constant.

Clear-sky solar radiation is calculated using Solar Analyst, which is an extension for ArcView 3·3. The Solar Analyst generates an upward-looking hemispherical viewshed (fisheye photograph) using 32 directions for each location on a 30-m resolution DEM (Lewkowicz and Ednie, 2004). Merging the hemispherical viewsheds that are calculated for each cell (on the DEM), Solar Analyst produces a clear-sky solar radiation density map (in MJ/m^2). Average clear-sky solar radiation is calculated for each 30-m pixel located within the study area from March 1 to April 1 for each of the 3 years used in the analysis (Figure 3f). Finally, the values are converted into fractional values by dividing them by the maximum observed clear-sky solar radiation value for that day.

The type of vegetation cover (Figure 3g) is obtained from land cover data collected during the USGS National Land Cover Characterization Project (2001). These data were collected across all the 50 states and Puerto Rico using Landsat 5 and 7 data. For this study, the vegetation cover is classified as forested or non-forested because these were the dominant land cover types in the study area. A value of 1 was assigned to the forest-covered pixels and a value of 0 was assigned to the non-forest-covered pixels.

Since a stepwise variable selection method was used in the analysis, the variance inflation factor (VIF) was used as a diagnostic tool to test for multicollinearity (Neter *et al*., 1996). The VIF measures the dependence between a predictor and the other independent variables (Helsel and Hirsch, 1992; Neter *et al*., 1996). A VIF score greater than 4 indicates that multicollinearity might be a problem, however, serious problems exist when the VIF is greater than 10 (Neter *et al*., 1996). Additionally, a correlation matrix (Table III) was constructed to assess the dependence between the predictor variables. If the selected regression equation was found to violate the multicollinearity assumption then further analysis would be necessary to determine which variables should be added or removed from the equation. The residuals from each equation were also tested for normality. In this study, however, the assumptions of collinearity and normality were not violated.

Each of the steps described in the Section on Methods was performed for each day separately. This is due to the fact that the snowline elevation changed from day to day and the relationships between the independent variables and dependent variable were not constant over time.

Cross-validation is used to compare the estimated (predicted) SWE values with the measured values. This is accomplished by removing each SWE data point and then using the remaining observations to estimate the data value. This procedure is repeated for every SWE observation used to generate each multivariate linear regression equation. The predicted values were subtracted from the observed values to obtain the residuals. The resulting residuals were then evaluated to assess the performance of the models. The residuals from the cross-validation procedure were used to compute the mean MAE, root mean squared error (RMSE), and coefficient of determination (R^2).

RESULTS AND DISCUSSION

The results from the cross-validation indicate that the spatial interpolation method produces accurate spatial estimates of SWE for each of the 6 days tested (Table IV). The average R^2, MAE, and RMSE values were 0·82, 4·34 cm, and 5·29 cm, respectively. The average R^2 values were slightly higher for April 1 dates (mean = 0·83), when compared with March 1 dates (mean = 0·81). In addition, the MAE and RMSE values for March 1 dates (mean MAE = 4·95 cm, mean RMSE = 5·94 cm) were larger than those for the three April 1 dates (mean MAE = 3·73 cm, mean RMSE = 4·64 cm). This suggests similar predictability for the April 1 dates than the March 1 dates.

The most important predictor variable of SWE was elevation for each of the 6 days tested (Table IV). The sign of the coefficient was positive, which indicates that the SWE values increase with increasing elevation. Ruggedness was also identified as an important predictor in three of the 6 days tested. As it would be expected, the sign of the coefficient was positive, because areas with greater topographic variability (high-ruggedness values) tend to be located at relatively high elevations. Clear-sky

Table III. Correlation matrix for the predictor variables used in the multivariate linear regression analysis on the SWE of the Big Wood Basin. Predictor variables used in the multivariate regressions are highlighted in bold

	Elevation	Latitude	Longitude	**Solar**	Slope	**Ruggedness**	Northness	Eastness	Land cover
Elevation	1·000	0·482	−0·443	−0·151	0·620	0·614	−0·292	−0·184	−0·381
Latitude		1·000	−0·636	0·160	0·265	0·119	−0·280	−0·564	−0·315
Longitude			1·000	−0·003	−0·500	−0·467	0·122	0·570	0·494
Solar				1·000	−0·701	−0·697	−0·249	−0·043	−0·039
Slope					1·000	0·955	−0·177	−0·100	−0·128
Ruggedness						1·000	−0·136	0·008	−0·144
Northness							1·000	−0·002	−0·353
Eastness								1·000	0·239
Land cover									1·000

Table IV. Cross-validation results. *N* represents the number of stations used in each multivariate regression equation for the 6 days analysed for the Big Wood Basin. N/A is used to designate dates where the basin was determined to be 100% snow-covered using the snow cover imagery

Date	Snowline elevation (m)	*N*	R^2	MAE (cm)	RMSE (cm)	Independent variables and sign of coefficient
1 March 2003	1628	17	0·80	5·00	5·99	+Elevation
1 March 2004	N/A	17	0·91	4·45	5·14	+Elevation, +ruggedness
1 March 2005	N/A	17	0·71	5·39	6·70	+Elevation
1 April 2003	2011	12	0·90	4·35	5·28	+Elevation, +ruggedness, −solar
1 April 2004	2209	10	0·74	3·26	4·49	+Elevation, +ruggedness
1 April 2005	1956	14	0·86	3·59	4·14	+Elevation

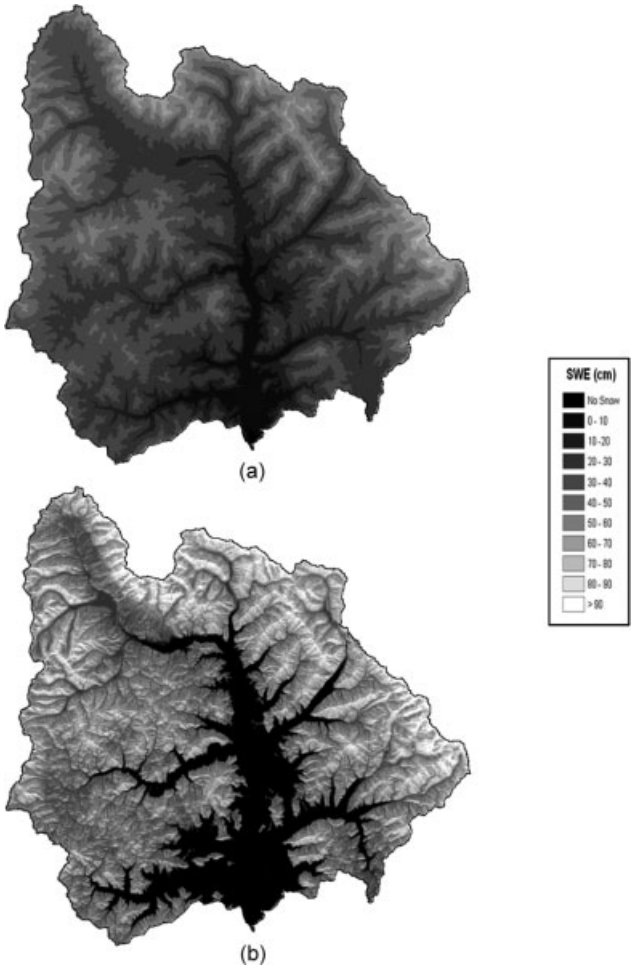

Figure 4. Gridded estimates of SWE (cm) for the Big Wood River Basin: (a) 1 March 2003 and (b) 1 April 2003

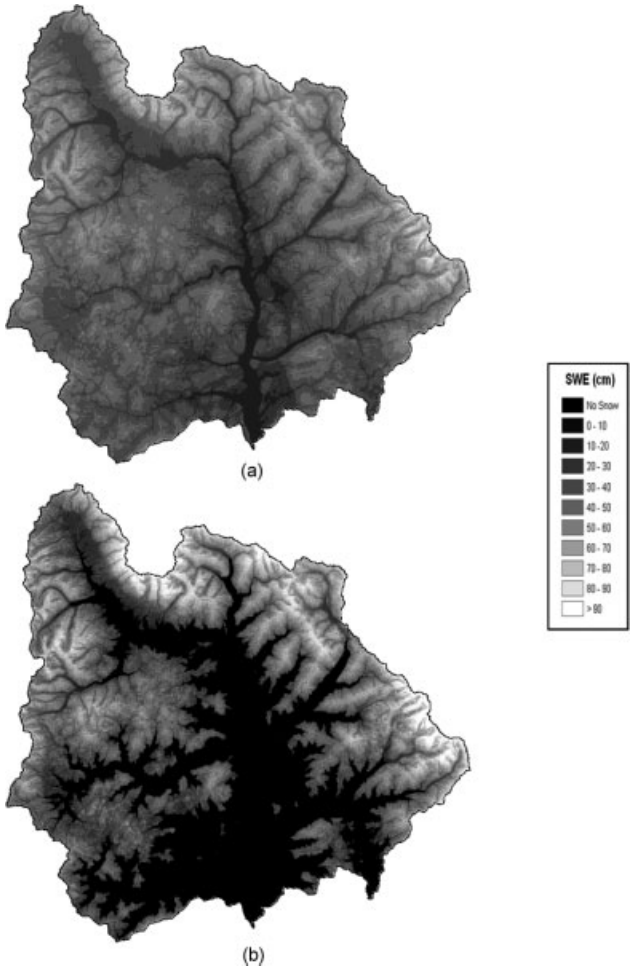

Figure 5. Gridded estimates of SWE (cm) for the Big Wood River Basin: (a) 1 March 2004 and (b) 1 April 2004

solar radiation was also identified as a predictor variable for one of the days tested. The sign of the coefficient was negative, which suggests that as the amount of solar radiation increased the amount of SWE decreased. This is because of the fact that incoming solar radiation is an important source of energy for melting the snowpack (Zuzel and Cox, 1975; Leavesley *et al*., 1983).

The distributed SWE grids (Figures 4–6), produced using the spatial interpolation method, indicate that the interpolated values were successfully decreased to 0 cm at the snowline elevation. This result is significant because one of the major goals of the study was to constrain the spatial interpolation to the SCA. Multivariate regression was not used to extrapolate data in the area between the snowline and the lowest SWE observation station above the snowline; because this can result in SWE values greater than 0 at the snowline. Instead, single-variable linear regression (elevation vs SWE only) was used to interpolate SWE values between the snowline (at 0 cm) and the elevation of the lowest SWE observation station above the snowline.

The distributed SWE grids (Figures 4–6) also successfully capture the temporal variability in the observed SWE data. The distributed SWE values tend to be more

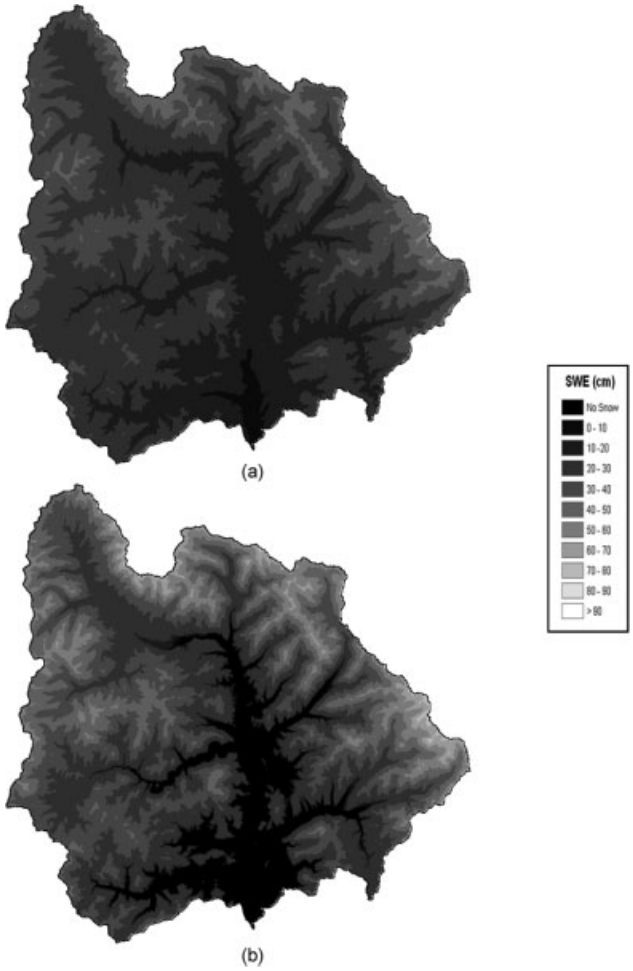

Figure 6. Gridded estimates of SWE (cm) for the Big Wood River Basin: (a) 1 March 2005 and (b) 1 April 2005

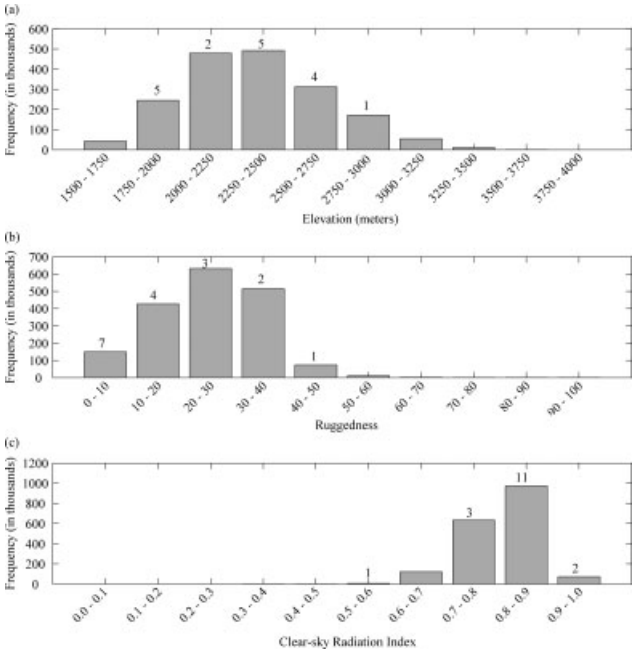

Figure 7. Histograms for the physiographic variables used in the multivariate regression equations of SWE in the Big Wood River Basin: (a) elevation, (b) ruggedness, and (c) clear-sky radiation index values

evenly distributed (less spatial variability) on March 1 (Figures 4a, 5a, and 6a) than on April 1 (Figures 4b, 5b, and 6b). This is due to the fact that the snowpack is diminishing in the lower elevations (between March 1 and April 1), but is still accumulating at the higher elevations. Thus, the SWE-elevation gradient becomes steeper around April 1. The SWE grids can also be used to identify years with lower basin-wide SWE values. As can be seen on 1 April 2005 (Figure 5b), the SWE was much lower in the low-elevation areas than it was for the other 2 years (Figures 4b and 6b).

Discussion of errors

In addition to the regression errors (Table IV) and measurement error inherent in the SWE observations, which was discussed in the Section on Data, there are several other potential sources of uncertainty in the interpolated SWE values. First, the median snowline elevation was used in the analysis that may introduce an error because the snowline is highly variable and changes spatially and temporally in response to meteorologic and other physiographic characteristics. In addition, since the MODIS snow cover grids have a spatial resolution of 500 m, errors in the determination of the snowline elevation would be expected to be larger in areas with rugged terrain. Additionally, it is often difficult for satellite sensors to detect the presence of snow cover under the forested canopy. Secondly, the use of single-variable linear regression (elevation vs SWE) to interpolate the SWE values between the snowline and the elevation of the lowest station above the snowline may introduce uncertainty in the gridded estimates, especially if the elevation change is significant.

Finally, there is the possibility that the SWE measurement locations (snow course and SNOTEL) may not be representative of the physiographic and snowpack conditions in the study area (Molotch and Bales, 2006). Daly *et al*. (2000) stated that most SNOTEL stations are located in forest openings that are relatively flat, wind-sheltered, and easily accessible. They also suggested that this may lead to the over-estimation of SWE. To address this uncertainty, histograms were created for the three physiographic variables identified in the multivariate regression models (Figure 7). The numbered values on the plots indicate the number of SWE stations that are found within each class. The distribution of elevation (Figure 7a), ruggedness (Figure 7b), and clear-sky radiation index (Figure 7c) values within the study are represented fairly well by the stations, with the exception of the lowest and highest values.

In addition to examining the above variables for representativeness, we recognize that there may be other variables influencing SWE that are not reliably or readily available as spatially distributed data sets to be used as input to interpolation schemes (i.e. wind, drifting potential, etc.). Snow cover persistence maps were created to

further address the representativeness of the SWE observation (SNOTEL and snow course) stations used in this project. MODIS snow cover data were used to generate snow cover persistence maps for the months of March and April. In generating these maps, values of 1–6 were assigned to each 500 m pixel based on the number of years during the period 2001–2006 where snow cover was detected (using MODIS data) during a given month. Average persistence values were computed for individual elevation zones in the basin. Five elevation zones were used in the analysis. Persistence values were also computed using ground data collected at the SWE observation stations in order to compare the snow cover persistence at the sites to those calculated for elevation zones within the basin. The results, shown in Table V, indicated that in the higher elevation zones, the average persistence values for the SWE observation stations were similar to the mean persistence values for all of the pixels in the individual elevation zones. However, the snow persistence values were greater for the SWE observation sites at the lower elevations (i.e. zones 1 and 2). This indicates that the SWE observation sites located at lower elevations are holding their snow longer than other cells within that elevation zone. This, however, is not a major limitation for the interpolation technique described here because there is often very little snow predicted by the regression in the lowest parts of the basin in March and April. Thus, the SWE grids produced here should not be severely affected by persistence differences in the lower elevations. In addition to the results described above, persistence values were also calculated for each elevation zone taking into account the density of vegetation cover. This analysis was completed to assess the influence of snow interception by the forest canopy on the interpolated SWE values. Three vegetation density classes were used, and were derived from leaf area index (LAI) data values obtained from NASA. The results indicated that vegetation density had a very minimal effect on the persistence values calculated within each elevation zone.

Table VI shows the percentage of the drainage basin that falls outside of the data range for the physiographic variables identified in the multivariate regression models. These data are organized by date and exclude the portion of the basin that was deemed non-snow-covered using the MODIS snow cover grids. The amount of uncertainty in the distributed SWE values would be expected to be highest in the portion of the basin that falls outside of the data range for the regression input variables. While this is true, however, SNOTEL and snow course data are the only widely available SWE measurements.

Table V. Average snow cover persistence values for SWE observation stations and 500 m MODIS pixels located within each elevation zone. The elevation ranges for each of the elevation zones are listed

Zone 1 (1615–1990 m)		Zone 2 (1991–2358 m)		Zone 3 (2359–2727 m)		Zone 4 (2728–3095 m)		Zone 5 (3096–3628 m)	
Satellite	SWE sites	Satellite	SWE sites	Satellite	SWE sites	Satellite	SWE sites	Satellite	SWE sites
				March					
4·70	6·00	5·23	6·00	5·79	6·00	5·99	6·00	6·00	6·00
				April					
1·23	4·50	3·44	5·29	5·42	6·00	5·93	6·00	5·99	N/A

Table VI. The percentage of the drainage basin that falls outside of the data range for the physiographic variables identified in the multivariate regression models. Data are organized by date and exclude the portion of the basin that was deemed non-snow-covered using the MODIS snow cover grids

Date	% of basin covered with snow	Independent variables	% of basin located below the range of data	% of basin located above the range of data	Total % of basin outside the range of data
1 March 2003	99·9	Elevation	7·0	9·0	16·0
1 March 2004	100·0	Elevation	7·0	9·0	16·0
		Ruggedness	3·0	7·0	10·0
		Elevation, ruggedness			22·0
1 March 2005	100·0	Elevation	7·0	9·0	16·0
1 April 2003	83·0	Elevation	0·0	11·0	11·0
		Ruggedness	1·0	9·0	10·0
		Solar	14·0	1·0	15·0
		Elevation, ruggedness, solar			26·0
1 April 2004	64·0	Elevation	0·0%	14·0%	14·0
		Ruggedness	0·1%	10·0%	10·1
		Elevation, Ruggedness			20·0
1 April 2005	87·0	Elevation	0·0	10·0	10·0

CONCLUSIONS

A new methodology was developed and tested for the creation of gridded estimates of SWE using surface observations (snow course and SNOTEL) and remotely sensed estimates of SCA. Although the methodology is rather simple, the cross-validation results indicate that the spatial interpolation method produces accurate SWE estimates (mean $R^2 = 0{\cdot}82$, mean MAE = 4·34 cm, and mean RMSE = 5·29 cm). The regression model errors (MAE and RMSE) were smaller for the April 1 dates than for the March 1 dates. Three predictor variables were identified in the multivariate linear regression (elevation, ruggedness, and clear-sky solar radiation) with elevation being the most important. Visual inspection of the SWE grids indicates that the predicted SWE values were successfully decreased to 0 cm at the snowline elevation. Since the methodology is rather simple, and the results are similar to those found using more sophisticated spatial interpolation techniques, we believe that there is potential for this method to be successfully applied to other mountainous basins in the western United States. The simplicity of the methodology should also make it relatively easy for water managers and other operational forecasters (i.e. federal agencies) to generate accurate gridded estimates of SWE in a timely manner.

ACKNOWLEDGEMENTS

The authors thank the Natural Resources Conservation Service, the National Snow and Ice Data Center, and NASA for providing us with the data required to complete this study. The authors thank Ryan Hruska for downloading the snow cover data and providing it to us. The authors also thank Devonee Harshburger for help in formatting the tables. This study was supported by the Pacific Northwest Regional Collaboratory as part of Raytheon Corporation's Synergy project, funded by NASA through NAS5-03098, Task No. 110.

REFERENCES

Balk B, Elder K. 2000. Combining binary recession-tree and geostatistical methods to estimate snow distribution in a mountain watershed. *Water Resources Research* **36**: 13–26.

Bloschl G, Kirnbauer R, Gutknecht D. 1991. Distributed snowmelt simulations in an alpine catchment. 2: parameter study and model predictions. *Water Resources Research* **27**: 3181–3188.

Carroll T. 1987. Operational airborne measurements of snow water equivalent and soil moisture using terrestrial gamma radiation in the United States. In *Large Scale Effects of Seasonal Snow Cover, Proceedings of the Vancouver Symposium*, Goodison B, Barry R, Dozier J (eds). IAHS Publication no.166, 213–223.

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

CDWR. 1976. *Snow Sensor Evaluation in the Sierra Nevada California*. California Department of Water Resources, Division of Planning: Sacramento.

Chang K, Li Z. 2000. Modelling snow accumulation with a geographic information system. *International Journal of Geographical Information Science* **14**: 693–707.

Cline DW, Bales RC, Dozier J. 1998. Estimating the spatial distribution of snow in mountain basins using remote sensing and energy balance modeling. *Water Resources Research* **34**: 1275–1285.

Daly SF, Davis R, Ochs E, Pangburn T. 2000. An approach to spatially distributed snow modelling of the Sacramento and San Joaquin basins, California. *Hydrological Processes* **14**: 3257–3271.

Dressler KA, Fassnacht SR, Bales RC. 2006. A comparison of snow telemetry and snow course measurements in the Colorado River Basin. *Journal of Hydrometeorology* **7**: 705–712.

Elder K, Dozier J, Michaelson J. 1991. Snow accumulation and distribution in an alpine watershed. *Water Resources Research* **27**: 1541–1552.

Elder K, Rosenthal W, Davis RE. 1998. Estimating the spatial distribution of snow water equivalence in a montane watershed. *Hydrological Processes* **12**: 1793–1808.

Erxleben J, Elder K, Davis RE. 2002. Comparison of spatial interpolation methods for estimating snow distribution in the Colorado Rocky Mountains. *Hydrological Processes* **16**: 3627–3649.

Fassnacht SR, Dressler KA, Bales RC. 2003. Snow water equivalent interpolation for the Colorado River Basin from snow telemetry (SNOTEL) data. *Water Resources Research* **39**: 1208–1217. DOI: 10.1029/2002WR001512.

Goodison BE, Ferguson HL, McKay GA. 1981. Measurement and data analysis. In *The Handbook of Snow: Principles, Processes, Management and Use*, Gray DM, Male DH (eds). Pergamon Press: Toronto; 360–436.

Helsel DR, Hirsch RM. 1992. *Statistical Methods in Water Resources*. Elsevier Science: New York.

Hosang J, Dettwiler K. 1991. Evaluation of a water equivalent of snow cover map in a small catchment area using a geostatistical approach. *Hydrological Processes* **5**: 283–290.

Idaho Department of Environmental Quality. 2001. *The Big Wood River Management Plan*. IDEQ-State Office: Boise, Idaho.

Johnson JB. 2004. A theory of pressure sensor performance in snow. *Hydrological Processes* **18**: 53–64. DOI: 10.1002/hyp.1310.

Johnson JB, Schaefer G. 2002. The influence of thermal, hydrologic, and snow deformation mechanisms on snow water equivalent pressure sensor accuracy. *Hydrological Processes* **16**(18): 3529–3542. DOI: 10.1002/hyp.1236.

Klein AG, Barnett AC. 2003. Validation of daily MODIS snow cover maps of the Upper Rio Grande River Basin for the 2000–2001 snow year. *Remote Sensing of Environment* **86**: 162–176.

Kleindienst H, Wunderle S, Voigt S. 2000. Snowline analysis in the Swiss Alps based on NOAA-AVHRR satellite data. *Remote Sensing of Land Ice and Snow (Proceedings of the EARSeL Workshop)*, Dresden, Germany; 297–307.

Leavesley GH, Lichty RW, Troutman BM, Saindon LG. 1983. *Precipitation-Runoff Modeling System-Users Manual*.

Lewkowicz AG, Ednie M. 2004. Probability mapping of mountain permafrost using the BTS method, Wolf Creek, Yukon Territory, Canada. *Permafrost and Periglacial Processes* **15**: 67–80.

Ling C, Josberger EG, Thorndike AS. 1996. Mesoscale variability of the upper Colorado River snowpack. *Nordic Hydrology* **27**: 313–322.

Luce CH, Tarboton DG, Cooley KR. 1998. The influence of the spatial distribution of snow on basin-averaged snowmelt. *Hydrological Processes* **12**: 1671–1683.

McManamon A, Szeliga TL, Hartman RK, Day GN, Carroll TR. 1993. Gridded snow water equivalent estimation using ground-based and airborne snow data. *Proceedings of the 61st Western Snow Conference, 50th Eastern Snow Conference*. Quebec city, Canada; 75–81.

Molotch NP, Bales RC. 2006. SNOTEL representativeness in the Rio Grande headwaters on the basis of physiographics and remotely sensed snow cover persistence. *Hydrological Processes* **20**: 723–739. DOI: 10.1002/hyp.6128.

Molotch NP, Colee MT, Bales RC, Dozier J. 2005. Estimating the spatial distribution of snow water equivalent in an alpine basin using binary-regression tree models: the impact of digital elevation data and independent variable selection. *Hydrological Processes* **19**: 1459–1479. DOI: 10.1002/hyp.5586.

Molotch NP, Fassnacht SR, Bales RC, Helfrich SR. 2004. Estimating the distribution of snow water equivalent and snow extent beneath cloud cover in the Salt-Verde River basin, Arizona. *Hydrological Processes* **18**: 1595–1611.

Moran P. 1950. Notes on continuous stochastic phenomena. *Biometrika* **37**: 17–23.

Mote TL, Grundstein AJ, Leathers DJ, Robinson DA. 2003. A comparison of modeled, remotely sensed, and measured snow water equivalent in the northern Great Plains. *Water Resources Research* **39**: 1209.

Neter J, Kutner MH, Nachtsheim CJ, Wasserman W. 1996. *Applied Linear Statistical Models*. WCB McGraw-Hill: New York.

Salomonson VV, Appel I. 2004. Estimating fractional snow cover from MODIS using the normalized difference snow index. *Remote Sensing of Environment* **89**: 351–360.

Seidel K, Ehrler C, Martinec J, Turpin O. 1997. Derivation of statistical snowline from high-resolution snow cover mapping. In *Remote Sensing of Land Ice and Snow (Proceedings of the EARSeL Workshop)*: Freiburg, Germany; 31–36.

Serreze MC, Clark MP, Armstrong RL, McGinnis DA, Pulwarty RS. 1999. Characteristics of the western United States snowpack from snowpack telemetry (SNOTEL) data. *Water Resources Research* **35**: 2145–2160.

Wilson LL, Tsang L, Hwang JN. 1999. Mapping snow water equivalent by combining a spatially distributed snow hydrology model with passive microwave remote-sensing data. *IEEE Transactions on Geoscience and Remote Sensing* **37**: 690–704.

Winstral A, Elder K, Davis R. 2002. Spatial snow modeling of wind-redistributed snow using terrain-based parameters. *Journal of Hydrometeorology* **3**: 524–538.

Woo MK, Marsh P. 1978. Analysis of error in the determination of snow storage for small high arctic basins. *Journal of Applied Meteorology* **17**: 1537–1541.

Zuzel JF, Cox LM. 1975. Relative importance of meteorological variables in snowmelt. *Water Resources Research* **11**: 174–176.

