Remote Sensing of Environment 215 (2018) 44–56

Contents lists available at ScienceDirect

Remote Sensing of Environment

journal homepage: www.elsevier.com/locate/rse

Spatial snow water equivalent estimation for mountainous areas using wireless-sensor networks and remote-sensing products

Zeshi Zheng[a,*], Noah P. Molotch[b], Carlos A. Oroza[a], Martha H. Conklin[c], Roger C. Bales[a,c]

[a] Department of Civil and Environmental Engineering, University of California, Berkeley, CA, USA
[b] Institute of Arctic and Alpine Research, Department of Geography, University of Colorado, Boulder, CO, USA
[c] Sierra Nevada Research Institute, University of California, Merced, CA, USA

ARTICLE INFO

Keywords:
Mountain hydrology
Snow water equivalent
Wireless-sensor networks
Sierra Nevada
k-Nearest neighbors
Reconstructions
MODIS
Lidar

ABSTRACT

We developed an approach to estimate snow water equivalent (SWE) through interpolation of spatially representative point measurements using a *k*-nearest neighbors (*k*-NN) algorithm and historical spatial SWE data. It accurately reproduced measured SWE, using different data sources for training and evaluation. In the central-Sierra American River basin, we used a *k*-NN algorithm to interpolate data from continuous snow-depth measurements in 10 sensor clusters by fusing them with 14 years of daily 500-m resolution SWE-reconstruction maps. Accurate SWE estimation over the melt season shows the potential for providing daily, near real-time distributed snowmelt estimates. Further south, in the Merced-Tuolumne basins, we evaluated the potential of *k*-NN approach to improve real-time SWE estimates. Lacking dense ground-measurement networks, we simulated *k*-NN interpolation of sensor data using selected pixels of a bi-weekly Lidar-derived snow water equivalent product. *k*-NN extrapolations underestimate the Lidar-derived SWE, with a maximum bias of −10 cm at elevations below 3000 m and +15 cm above 3000 m. This bias was reduced by using a Gaussian-process regression model to spatially distribute residuals. Using as few as 10 scenes of Lidar-derived SWE from 2014 as training data in the *k*-NN to estimate the 2016 spatial SWE, both RMSEs and MAEs were reduced from around 20–25 cm to 10–15 cm comparing to using SWE reconstructions as training data. We found that the spatial accuracy of the historical data is more important for learning the spatial distribution of SWE than the number of historical scenes available. Blending continuous spatially representative ground-based sensors with a historical library of SWE reconstructions over the same basin can provide real-time spatial SWE maps that accurately represents Lidar-measured snow depth; and the estimates can be improved by using historical Lidar scans instead of SWE reconstructions.

1. Introduction

In the state of California, ecosystem processes and water supplied for agricultural and urban uses depend on the snowpack in the Sierra Nevada as the primary source of spring and summer streamflow (Bales et al., 2006). As the prediction of water availability and flood peaks depend in part on snowpack conditions, accurate knowledge of the snowpack can assist decision making for water resources management (California Department of Water Resources, 2013).

Current decision making for water management in California during the snowmelt season relies on ground measurements in the Sierra Nevada, which include continuous snow-pillow and snow-depth sensor measurements, and monthly manual snow surveys (Molotch and Bales, 2005). Ground stations are sparsely placed in the mountains compared to the spatial scale of each watershed. Therefore, the measurements may not be representative of physiographic features required to capture spatial variability of snow depth and snow water equivalent (SWE), either at the site or basin scale. Satellite-based remote sensing, such as MODIS and Landsat, has been used to map snow coverage at regional to global scales. However, they provide only pixel-wise fractional snow-coverage information, with no direct information on snow depth or SWE (Dozier et al., 2008; Molotch and Margulis, 2008; Painter et al., 2009; Raleigh et al., 2013; Rittger et al., 2013; Rosenthal and Dozier, 1996). A modeled snow-data product that is commonly used in the Continental United States is the Snow Data Assimilation System (SNODAS), which integrates snow information from both satellite and ground stations, providing daily snow depth and snow water equivalent information at 1-km^2 resolution (Barrett, 2003). Recent work validating the SNODAS spatial product with Lidar suggested that the performance of SNODAS in the Tuolumne River basin is less accurate than 3-D

* Corresponding author.
E-mail address: zeshi.z@berkeley.edu (Z. Zheng).

https://doi.org/10.1016/j.rse.2018.05.029
Received 17 October 2017; Received in revised form 20 March 2018; Accepted 29 May 2018

(x,y,elevation) bilinear interpolation of ground stations (Bair et al., 2016).

Snow-coverage information and modeled spatial land-surface meteorological data can be used to back-integrate SWE from the snow melt-out date to the date of maximum SWE at the beginning of the snowmelt season. This technique has been applied across several mountain ranges and is referred to as the SWE-reconstruction technique (Bair et al., 2016; Guan et al., 2013; Margulis et al., 2016; Rittger, 2012). Although SWE reconstruction captures both temporal variability and spatial variability, it can only be done at the end of the season when the daily energy inputs and snow covered area are known (Cline et al., 1998).

As a complement to satellite-based estimates of snow distribution, numerous statistical models have been developed to interpolate point-based snow information. Multivariate linear regression, commonly used in previous studies, can relate physiographic variables, historical SWE data, and snow covered area imagery with the observed SWE; and the accuracy is reasonably better than techniques such as inverse-distance weighting and simple kriging (Schneider and Molotch, 2016; Fassnacht et al., 2003). However, the linear-regression-based methods do not provide spatially smooth maps and the independent variables do not necessarily have a linear relationship with SWE (Zheng et al., 2016). Other than regression, one category of methods that have shown promise are nearest-neighbor-based algorithms. These algorithms are attractive because they are easy to implement, nonparametric, learning based, and can learn linear and nonlinear trends in observations (Ni and Nguyen, 2009). Simulations and estimations at either fine temporal or spatial resolutions using parametric models can be computationally intensive. Nearest-neighbor approaches have therefore become an alternative solution to many problems in spatio-temporal modeling, not only for their advantage in time complexity, but also for their superior accuracy and ability to preserve patterns from observations. The k-NN algorithm has been used for multivariate time-series simulation for weather forecasting (Rajagopalan and Lall, 1999), disaggregating meteorological time-series data to finer time scales (Prairie et al., 2007; Kalra and Ahmad, 2011), and downscaling spatial climate-model data (Gangopadhyay and Clark, 2005). The k-NN algorithm was found to be superior in preserving the spatio-temporal covariability of the observation than multivariate autoregressive approaches.

To address the issues in presently available basin-scale water-balance data, a prototype real-time observation network that includes monitoring the snow conditions is being developed for the headwater areas of the American River basin in the Sierra Nevada (Zhang et al., 2017). The system enables combining ground measurements of snow depth and historical SWE reconstruction using a k-nearest neighbors (k-NN) algorithm for real-time spatial SWE estimation (Larose, 2005).

This work documents the k-nearest neighbors spatial-SWE-estimation method and evaluates the estimates against a spatial SWE product that is derived from Lidar-measured snow depth. Three questions that motivated the present study are:

1. Does a k-NN approach for spatial SWE interpolation in mountainous regions provide accurate SWE estimates relative to other products?
2. How is the error of the k-NN estimation distributed with regard to topographic variables?
3. Is it possible to further decrease the error of the k-NN estimates by distributing the residuals spatially?

2. Methods

We applied the k-nearest neighbors (k-NN) algorithm to estimate spatial snow water equivalent (SWE) in three basins in the Sierra Nevada, California, USA (Fig. 1a, Table 1). The experiment for the American River basin focused on estimating the 2014 spatial SWE using 10 clusters of snow-depth measurements for 2014 from wireless-sensor networks, and historical SWE reconstructions based on MODIS from 2001 to 2013, aiming to evaluate the k-NN estimates temporally over the melt season. The SWE reconstructions were used by the algorithm for learning the SWE spatial distribution embedded in the data set. We did similar experiments in the Merced (2014) and Tuolumne (2014, 2016) basins using Lidar-based SWE estimates to evaluate the k-NN results spatially. For these two basins, since we have fewer sensor networks deployed, we instead selected representative pixels as hypothetical sensor-network locations based on physiographic variables using a Gaussian-mixture model; and used these Lidar-based SWE values for the k-NN experiments. In this setup we used historical SWE reconstructions, historical Lidar-derived SWE, and historical SNODAS SWE as spatial training data to explore if different data sources matter. The spatial results over the two basins were evaluated using the Lidar-derived SWE as a ground-truth data set.

2.1. American River basin analysis using wireless-sensor network data

The 10 wireless-sensor networks (Table 2) were deployed in the seasonally snow-covered region of the 5570 km^2 American basin (Fig. 1b). Each network has ten or eleven sensor nodes (Fig. 1c) that measure snow depth, temperature, relative humidity, soil moisture, and short-wave solar radiation (Zhang et al., 2017; Brun-Laguna et al., 2016). The placements were strategically selected, aiming to capture snow depth and meteorological variability from elevation gradients, south versus north-facing slopes, steep versus flat areas, and various vegetation densities. All sensors take measurements at a 15-minute intervals, and the network manager of each sensor cluster forwards the data to a central webserver (Zhang et al., 2017). Daily data averaged over each cluster were used in the current analysis.

2.2. Snow water equivalent reconstruction data

Snow water equivalent reconstruction is an existing data set that was produced by estimating historical spatial SWE for past snowmelt seasons (Guan et al., 2013; Molotch et al., 2017). The time extent of the SWE reconstructions is from 2000 to 2014 and the spatial extent covers the entire Sierra Nevada. The SWE-reconstruction method uses a snow-surface energy and mass-balance model:

$$M_p \rho L = S\downarrow(1-\alpha) + LW\downarrow + LW\uparrow + SH + LH \tag{1}$$

where M_p(m) is the potential snowmelt (assuming full snow coverage), ρ(kg/m^3) is the liquid-water density, L(kJ/kg) is the latent heat of fusion, $S\downarrow$ (J/m^2) is the subcanopy insolation, α(unitless) is snow albedo, $LW\downarrow$ (J/m^2) is the downwelling longwave radiation, $LW\uparrow$ (J/m^2) is the longwave radiation emitted from the snowpack. SH(J/m^2) and LH(J/m^2) are sensible heat exchange and latent heat exchange accordingly. We need to note that the SWE-reconstruction model did not account for precipitation that occurs during the melt period, which may introduce bias in the estimates. The potential snowmelt M_p is scaled by the fractional snow-covered area (f_{SCA}) derived from MODIS to estimate the actual daily snowmelt,

$$M = M_p \times f_{SCA}. \tag{2}$$

The time-series SWE for the season is calculated by back integrating the daily snowmelt since snow meltout:

$$\mathrm{SWE}_0 = \mathrm{SWE}_n + \sum_{j=1}^{n} M_j \tag{3}$$

where SWE$_n$ is SWE at time step n, SWE$_0$ is the initial SWE, and M_j is the actual snowmelt during time step j. The initial SWE at each model pixel can be reconstructed at the time when snow disappearance observed from the satellite ($f_{SCA} = 0$):

$$\mathrm{SWE}_0 = \sum_{j=1}^{n} M_j \quad \text{when} \quad \mathrm{SWE}_n = 0. \tag{4}$$