Journal of Hydrology 516 (2014) 304–316

Contents lists available at ScienceDirect

Journal of Hydrology

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

Evaluation of sampling techniques to characterize topographically-dependent variability for soil moisture downscaling

Kevin L. Werbylo, Jeffrey D. Niemann *

Department of Civil and Environmental Engineering, Colorado State University, Campus Delivery 1372, Fort Collins, CO 80523-1372, United States

ARTICLE INFO

Article history:
Available online 21 January 2014

Keywords:
Soil moisture
Spatial variation
Downscaling
Efficient sampling
Conditioned Latin hypercube sampling
Stratified random sampling

SUMMARY

Downscaling methods have been proposed to estimate catchment-scale soil moisture patterns from coarser resolution patterns. These methods usually infer the fine-scale variability in soil moisture using variations in ancillary variables like topographic attributes that have relationships to soil moisture. Previously, such relationships have been observed in catchments using soil moisture observations taken on uniform grids at hundreds of locations on multiple dates, but collecting data in this manner limits the applicability of this approach. The objective of this paper is to evaluate the effectiveness of two strategic sampling techniques for characterizing the relationships between topographic attributes and soil moisture for the purpose of constraining downscaling methods. The strategic sampling methods are conditioned Latin hypercube sampling (cLHS) and stratified random sampling (SRS). Each sampling method is used to select a limited number of locations or dates for soil moisture monitoring at three catchments with detailed soil moisture datasets. These samples are then used to calibrate two available downscaling methods, and the effectiveness of the sampling methods is evaluated by the ability of the downscaling methods to reproduce the known soil moisture patterns. cLHS outperforms random sampling in almost every case considered. SRS usually performs better than cLHS when very few locations are sampled, but it can perform worse than random sampling for intermediate and large numbers of locations.

1. Introduction

Many hydrologic processes are influenced by patterns of volumetric water content in the soil (soil moisture). Specifically at the catchment scale, spatial patterns of soil moisture are closely related to spatial patterns of erosion (Fitzjohn et al., 1998), crop yield (Green and Erskine, 2004), and the magnitude and timing of runoff production (Western, 2001). Yet, most available methods do not estimate soil moisture at resolutions suitable for catchment-scale applications (e.g., grid cells with a 10–50 m linear dimension). For example, neutron-emission methods (Shuttleworth et al., 2010; Zreda et al., 2008) and microwave remote-sensing methods (Njoku et al., 2003; Kerr et al., 2001; Entekhabi et al., 2010) estimate spatial patterns of soil moisture at resolutions ranging from 700 m to 60 km.

Various methods have been proposed to downscale coarse-resolution soil moisture estimates. The initial and final resolutions of the soil moisture patterns are important to the design of these methods because different factors control spatial variations in soil moisture at different scales (Western et al., 2002). Methods described by Merlin et al. (2006), Kim and Barros (2002), Mascaro et al. (2010, 2011), Pellenq et al. (2003) and Temimi et al. (2010) all focus on producing soil moisture patterns at resolutions of 90 m or coarser. Kaheil et al. (2008) proposed a method to downscale soil moisture patterns to a 50 m resolution using sparse ground observations. Wilson et al. (2005) used fine-resolution topographic attributes, fine-resolution (10–40 m) in situ soil moisture observations, and a single spatial average soil moisture value on each date to estimate soil moisture patterns with resolutions of 10–40 m. Similarly, Perry and Niemann (2007) and Busch et al. (2012) explored a method using empirical orthogonal function (EOF) analysis, fine-resolution (5–15 m) topographic attributes, fine-resolution (10–40 m) in situ soil moisture observations, and a single spatial average soil moisture value on each date to estimate soil moisture patterns with resolutions of 10–40 m. Recently, Coleman and Niemann (2013) proposed a conceptual model known as the Equilibrium Moisture from Topography (EMT) model to estimate soil moisture patterns at resolutions of 10–40 m using fine-resolution (5–15 m) topographic attributes, fine-resolution (10–40 m) soil moisture observations, and a single spatial average soil moisture value on each date.

Several of these downscaling methods infer the fine-scale variability of soil moisture from its relationship to available ancillary variables. Soil moisture patterns have been shown to be correlated with spatial patterns of topography, vegetation, soil texture, and

* Corresponding author. Tel.: +1 970 491 3517.
E-mail addresses: Kevin.Werbylo13@alumni.colostate.edu (K.L. Werbylo), jniemann@engr.colostate.edu (J.D. Niemann).

0022-1694/$ - see front matter © 2014 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.jhydrol.2014.01.030

combinations of these variables (Cantón et al., 2004; Gómez-Plaza et al., 2001; Gutiérrez-Jurado et al., 2006; Western et al., 1999). At the catchment-scale, topography has been a widely-used ancillary data source (Busch et al., 2012; Coleman and Niemann, 2013; Perry and Niemann, 2007; Wilson et al., 2005) because of its known influence on soil moisture patterns at this scale (Famiglietti et al., 1998; Western and Grayson, 1999) and its nearly global availability at fine resolutions (Welch et al., 1998). Many of these downscaling methods also use in situ soil moisture observations to characterize the relationships between the variations in soil moisture and the variations in ancillary data. These relationships are commonly obtained through linear regression (Busch et al., 2012; Perry and Niemann, 2007; Wilson et al., 2005) or parameter calibration (Coleman and Niemann, 2013). Busch et al. (2012) found that such relationships are catchment-specific, which implies that soil moisture observations need be collected from the catchments where the downscaling method will be applied or the relationships need to be inferred from knowledge of the physical characteristics of the catchment. In the development of most catchment-scale downscaling methods, the in situ soil moisture observations have been collected on uniform grids, which contain hundreds of points on multiple dates (Busch et al., 2012; Coleman and Niemann, 2013; Perry and Niemann, 2007; Wilson et al., 2005). Collecting data in this manner is expensive and time-consuming, which limits the applicability of such downscaling methods.

Several studies have considered more efficient sampling techniques to observe catchment-scale soil moisture, but these studies have focused on capturing catchment-average conditions. In particular, many researchers have attempted to estimate a catchment-wide spatial average soil moisture using point-scale observations from a limited number of locations (Brocca et al., 2009; Martinez-Fernandez and Ceballos, 2005; Grayson and Western, 1998). Similarly, other researchers have attempted to validate coarse-resolution remote sensing estimates by upscaling point soil moisture observations (Cosh et al., 2008, 2006; Crow et al., 2012, 2005).

Sampling techniques have also been proposed to efficiently capture the variability of catchment conditions, but such techniques have not been applied to soil moisture. Conditioned Latin hypercube sampling (cLHS) (Minasny and McBratney, 2006) and stratified random sampling (SRS) (Avery and Burkhart, 2001) aim to determine monitoring locations for the variable of interest based on knowledge of ancillary variables. The goal of both methods is to identify sampling locations that represent a diverse set of values for the ancillary variables so that the sampling is less likely to be redundant. cLHS and SRS are similar in that they each divide the observed range of each ancillary variable into bins and then select the observation locations from the locations within each bin. These methods differ in how the bins are determined. cLHS divides the range of each ancillary variable into equally probable bins such that each bin contains the same number of observations. For SRS, different methods have been used to determine the bins (McKenzie and Ryan, 1999; Worsham et al., 2012). Here, we focus on the case where SRS divides the range of each ancillary variable into bins that cover an equal fraction of the observed range, regardless of the number of observations within each bin. Both cLHS and SRS are potentially more efficient than uniform or random sampling because they aim to reduce redundancy in the information gathered at the sampling locations. Minasny and McBratney (2006) evaluated cLHS in the context of soil mapping and found that sample histograms created from cLHS better replicate the known histograms of topographic, vegetative, and land use ancillary variables than those created from random sampling and a stratified sampling method. Recently, Worsham et al. (2012) evaluated the use of cLHS and a SRS method by their ability to improve spatial estimates of soil carbon content. Both methods outperform random sampling when sample sizes are limited, but cLHS does not consistently outperform SRS in that context. The SRS method they used stratifies the landscape into units based on soil type and land use data. Samples are then selected randomly from each spatially-contiguous unit in order to sample across the ranges of the ancillary variables as well as the spatial extent of the region. McKenzie and Ryan (1999) also used an SRS method (Brus and de Gruijter, 1997) with climate and topographic ancillary variables to make spatial predictions of soil depth, total phosphorus, and total carbon. The SRS method they used only focuses on adequately covering the ranges of the ancillary variables (not the spatial extent of the area of interest).

The objective of the present paper is to assess the effectiveness of two strategic sampling techniques at identifying the relationships between topographic attributes and soil moisture for catchment-scale downscaling applications. Two strategic sampling techniques are considered: the cLHS method proposed by Minasny and McBratney (2006) and an SRS method that is similar but not identical to the SRS technique used by McKenzie and Ryan, (1999). These sampling methods are coupled with two downscaling methods: the EMT model (Coleman and Niemann, 2013) and the EOF method (Busch et al., 2012). The ancillary variables that are required by these downscaling methods (various topographic attributes) are used by the sampling techniques to identify locations where the soil moisture should be monitored. Then, the soil moisture values at the monitored locations are used to define the relationships between the topographic attributes and soil moisture in the downscaling methods. The downscaling methods are then used to produce estimates of the catchment-scale soil moisture patterns from the spatial-average soil moisture and topographic attributes at each catchment. Ultimately, the performance of the sampling methods is evaluated by the ability of the two downscaling techniques to reproduce the actual catchment-scale soil moisture patterns at three application catchments (Tarrawarra, Satellite Station, and Cache la Poudre) when supplied with data from the sampling methods. As a secondary objective in this study, the EMT model and EOF method are compared under a variety of the data-limited conditions.

2. Methodology

2.1. Sampling methods

The cLHS method proposed in Minasny and McBratney (2006) can be summarized as follows. To start, the values of the ancillary variables at all locations on the desired fine-resolution grid within the region of interest are organized into a matrix X of size N by K where N is the number of locations and K is the number of ancillary variables observed at each location. Any row in X corresponds to a location in the catchment, and each column contains the values for a particular ancillary variable. In the present application, the ancillary variables are various topographic attributes that are required by the downscaling methods (discussed in more detail later). Using the values in each column of X, the ancillary variables are divided into n bins where n is the number of desired samples (i.e. locations). For a given ancillary variable, the limits for the bins are defined so that each bin contains an equal number of observed values. Fig. 1a displays a hypothetical example where a single ancillary variable is used and the range of the ancillary variable has been divided into 3 bins in this manner. n locations are then randomly selected from X producing a matrix x of size n by K that contains the values of the ancillary variables at the selected locations. A particular row of x represents one of the selected locations. An associated matrix η of size n by K is then created. An element of η is associated with a particular bin number (1 to n) and a