Geoderma 335 (2019) 27–34

Contents lists available at ScienceDirect

Geoderma

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

On the use of the critical event concept for quantifying soil moisture dynamics

Tiejun Wang[a,*], Shailesh Kumar Singh[b], András Bárdossy[c]

[a] Institute of Surface-Earth System Science, Tianjin University, Weijin Road 92, Tianjin 300072, PR China
[b] National Institute of Water and Atmospheric Research, Christchurch, New Zealand
[c] Institute of Hydraulic Engineering, University of Stuttgart, Stuttgart D-70569, Germany

ARTICLE INFO

Handling Editor: A.B. McBratney

Keywords:
Soil moisture
Critical event
Depth function
AWDN

ABSTRACT

Soil moisture is a key state variable in terrestrial water cycles, which links various land surface and hydrological processes. Owing to the significant spatiotemporal variability in soil moisture, collecting sufficient soil moisture data for relevant studies can be an astonishingly difficult task. Here, a statistical method based on the concept of depth functions was used to define Critical Events (CE) of soil moisture, which was hypothesized to contain disproportionally more system information on soil moisture dynamics. To test the feasibility of applying the CE concept for quantifying soil moisture dynamics, a long-term soil moisture dataset was retrieved from the Automated Weather Data Network (AWDN) located in the continental United States. The method of Temporal Stability Analysis (TSA) was adopted to examine the information embedded in soil moisture data that were collected under different conditions. The results showed that similar information on the relative as well as absolute wetness conditions at the AWDN sites was extracted from the entire time series and CEs of soil moisture, the latter of which contained much less soil moisture observations. Finally, the field data revealed that the occurrence of CEs of soil moisture in the study area was mainly affected by soil depth and interannual variability in precipitation. The number of CEs of soil moisture tended to be larger at deeper soil depths as well as in either dry or wet years than in normal years, suggesting that more soil moisture measurements under those conditions were needed to provide adequate information on soil moisture systems.

1. Introduction

Designing appropriate strategies for monitoring hydrological processes is critical for hydrological studies and water resources management (Woods et al., 2001; Brocca et al., 2010; Zacharias et al., 2011). As a key state variable, soil moisture links various land surface and hydrological processes; however, owing to the significant spatiotemporal variability in soil moisture, collecting sufficient soil moisture data for relevant uses can be an astonishingly difficult task, especially in areas with little logistical support. Although an array of techniques have been developed to measure in-situ soil moisture contents with reasonable accuracy (see the review by Robinson et al. (2008)), available automated soil moisture monitoring sites around the globe (Crow et al., 2012; Ochsner et al., 2013) are still less than those for monitoring other hydrological processes (e.g., precipitation-*P* and streamflow), mostly due to their high installation and operational costs. Recent developments of remote sensing techniques can partially overcome this problem by providing large spatial coverages of soil moisture (e.g., Kerr, 2007; Franz et al., 2015; Cosh et al., 2016); however, technical obstacles still exist for using remotely sensed soil moisture data for various research and application purposes (e.g., limited penetration depth and interference of cloudiness).

To enhance the representativeness of soil moisture measured in the field, various techniques have been developed, most notably based on the method of Temporal Stability Analysis (TSA). Derived from the field observation of a temporal persistence in the spatial pattern of soil moisture, Vachaud et al. (1985) introduced the concept of TSA to identify representative sites for monitoring soil moisture. Since the seminal work of Vachaud et al. (1985), the TSA method has been extensively used to study soil moisture spatiotemporal variability from field to regional scales (e.g., Grayson and Western, 1998; Martinez-Fernandez and Ceballos, 2003; Guber et al., 2008; Zhao et al., 2010; Wang and Franz, 2015; Wang et al., 2017a). It should be noted that in some of the field studies, soil moisture data were manually collected with considerable time gaps between consecutive measurements. Therefore, knowledge of the frequency and timing for sampling soil

* Corresponding author.
E-mail address: tiejun.wang@tju.edu.cn (T. Wang).

https://doi.org/10.1016/j.geoderma.2018.08.013
Received 29 August 2017; Received in revised form 26 February 2018; Accepted 7 August 2018

moisture is desirable in order to provide adequate information on soil moisture systems. However, the TSA method is not suitable for providing such information, in part due to the lack of the understanding of the controls on soil moisture temporal dynamics (Vanderlinden et al., 2012). For instance, based on the TSA method, Martinez-Fernandez and Ceballos (2003) showed that soil moisture was temporally more stable under dry conditions in a semiarid catchment in Spain; whereas, Zhao et al. (2010) found that soil moisture was temporally more stable under wet conditions in a semiarid steppe in China. As such, investigations into the representativeness of soil moisture data that are collected discontinuously over time are still warranted for the purposes of hydrological studies and water resources management.

In this study, the concept of Critical Event (CE) initially proposed by Singh and Bárdossy (2012) was used to investigate the information embedded in soil moisture data that were collected under different conditions. The rationale behind the CE concept is that information contained in a time series (e.g., P and streamflow) is not homogeneous (Wagener et al., 2003). Observations from certain time periods (i.e., CE) with more embedded information are more useful for understanding hydrological systems (Xia et al., 2004; Seibert and Beven, 2009; Singh, 2010). Most statistical methods, such as principle component analysis, intend to describe the mean behavior of a set of observations. In contrast, the CE concept intends to identify states that are unusual. Those unusual states, which can be identified with the help of the concept of a statistical depth, provide more insights to the behavior of hydrological systems under study (Singh and Bárdossy, 2012). In Singh and Bárdossy (2012), the authors identified CEs for P and streamflow in the Neckar catchment in Germany, which contained about 6% of the total data points. The authors showed that the identified CEs were enough to calibrate hydrological models with similar model performances to those that were calibrated using the entire dataset. Singh et al. (2012) applied a CE-based method to improve the calibration of a physically based hydrological model. They showed that the CE-based calibrated model performed even better during the validation period. The CE concept has also been proven to be useful for selecting information-rich data for training data hungry artificial neutral network models (Singh et al., 2014). Moreover, Singh and Bárdossy (2012) advocated to use the CE method for guiding field collections of hydrological data (e.g., timing and frequency), particularly in ungauged catchments.

The main goal of this study was to apply the CE method to examine the information embedded in soil moisture data that were collected under different conditions. In particular, we attempted to answer the following questions: (1) Do CEs of soil moisture contain sufficient information for representing the entire time series? and (2) what are the primary factors that control the occurrence of CEs of soil moisture? To this end, a long-term soil moisture dataset retrieved from the widely used Automated Weather Data Network (AWDN) located in the continental United States was analyzed. To test whether CEs of soil moisture contain sufficient information, the TSA method was used. In addition, hydrometeorological and soil data were used to explore the factors that affected the occurrences of CEs of soil moisture across the AWDN sites. The results of this study offer further insights into understanding soil moisture dynamics, which is relevant for hydrological studies and water resources management.

2. Methods and materials

2.1. Concept of critical event

Examples were given by Singh and Bárdossy (2012) to demonstrate the CE concept. The authors argued that a time series of a hydrological variable might contain a large number of events with high resemblance, which could be treated as repetitions with redundant information contents. By comparison, some events (e.g., maxima and minima) can provide more information on the response of a hydrological system to changes in surrounding conditions. Based on this reasoning, CEs that contain more useful information for understanding hydrological systems can be defined if they meet certain criteria. To this end, Singh and Bárdossy (2012) proposed to use depth functions as a diagnostic tool for screening CEs from a time series of a hydrological variable. Specifically, for a one-dimensional space, extremes are defined as CEs (e.g., away from the mean). Similarly, for a multivariate dataset, its unusual states ('extremes') or CEs can be found near or on the boundary of the convex hull of the multivariate dataset in the multidimensional space, which is consistent with the concept of the statistical depth first proposed by Tukey (1975). The idea presented by Tukey (1975) was to define the center of a multivariate dataset (see Zuo and Serfling (2000) for details). More specifically, the convex hull of a set of M points in two dimensions is the smallest polygon area that encloses all M points. Mathematically, the convex hull of a set of M points in d dimensions is the intersection of all convex sets containing all M points. Conceptually, data points that are close to the center of the dataset have high depths, while those that are near the boundary of the dataset have low depths and can be treated as CEs.

Following Singh and Bárdossy (2012), the procedure for defining CEs using depth functions is briefly explained here. For a time series $X(t)$, where X is a hydrological variable (soil moisture in this study) and t is time (days in this study), $X(t)$ can be reconstructed using a sequence of d consecutive observations of $X(t)$ to form a d-dimensional dataset, X_d (the dimension d refers to the number of lag or lead time steps in X_d):

$$X_d(t) = \{[X(t-d+1), X(t-d+2), \cdots, X(t)], t = d, \cdots, T\} \quad (1)$$

where T is the total number of observations, resulting in a $d \times T$ matrix. In this study, the half-space depth function of Tukey (1975) was used to define CEs due to its simplicity and clear geometric interpretation (Zuo and Serfling, 2000). Mathematically, the half-space depth of the data point p in the d-dimensional dataset with respect to the set X_d (denoted as $D_{X_d}(p)$) is defined as:

$$D_{X_d}(p) = min(min(|\{x \in X \langle n_h, x-p\rangle > 0\}|), (|\{x \in X \langle n_h, x-p\rangle < 0\}|)) \quad (2)$$

where $\langle n_h, x-p\rangle$ is the scalar product of the d-dimensional vector, and n_h is an arbitrary unit vector in the d-dimensional space and normal to a selected hyperplane. The outer minimum is taken over all possible n_h. For each time step t, the depth of $X_d(t)$ with respect to the matrix X_d (i.e., Eq. (1)) was calculated. In particular, points on and near the boundary of the convex hull of the data cloud X had a depth of 0, those further inside had low depths, and those close to the center of the data cloud had high depths. Data depth functions have certain properties, which make them useful for providing a center-outward ordering of points in a multivariate data set, such as affine invariance (i.e., the depth of a point is independent on the underlying coordinate system), maximality at the center of the data cloud, monotonicity relative to the deepest point (i.e., depth decreases as data points move away from the center), and the depth of a point vanishing at infinity.

An example of the convex hull, hyperplane, and depth values is given in Fig. 1 for soil moisture time series with $d = 2$. In Fig. 1, the convex hull was determined by the intersection of all the convex sets. The hyperplane is the subspace, whose dimension is one less than that of its ambient space. In Fig. 1 with $d = 2$, the hyperplane is a line. The depth values were calculated based on Eq. (2) (i.e., the depth of a point is the minimum of the minimum numbers of points lying either side of a hyperplane). Fig. 2 demonstrates the depth calculation procedure in the 2-d case. First assume that we want to compute the depth of the rectangular point (highlighted in red). A hyperplane is drawn passing through the rectangular point as shown in Fig. 2a. The number of points is then counted at each side of the hyperplane. Afterwards, the hyperplane is rotated clockwise for 360 degrees (e.g., Fig. 2b–d). For each increment, the number of points lying either side of the rotated hyperplane is counted and the minimum number of the points is recorded (e.g., 4, 5, 7, and 11 as shown in Fig. 2). Finally, the depth of the