Use of a state-space approach to predict soil water storage at the hillslope scale on the Loess Plateau, China

Liangxia Duan a,b, Mingbin Huang b,⁎, Luodan Zhang a,b

a College of Resources and Environment, Northwest A & F University, Yangling 712100, China
b State Key Laboratory of Soil Erosion and Dryland Farming on the Loess Plateau, Institute of Soil and Water Conservation, Northwest A & F University, Yangling, 712100, China

Abstract

Soil water storage is a critical variable controlling hydrological and biological processes. The precise estimation of soil water storage in diverse soil layers is fundamental to understanding hydro-biological processes and efficiently managing water resources. The objectives of this study were to evaluate the effects of topography (elevation) and soil properties (clay, silt, sand content, median grain size, and fractal dimension) on soil water storage and then to estimate soil water storage using a state-space approach. The soil water storage values of three soil layers (0–1, 1–2, and 2–3 m) were measured from May to December 2014 at 70 locations along two 187 m long transects on a hillslope of the Loess Plateau, China. Samples from various depths were also collected to determine soil properties. The best state-space approach explained 98.8% of the total variation in soil water storage, while the best classical linear regression equation only explained 64.2%. The state-space approach using any combination of variables described the spatial pattern of soil water storage much better than equivalent linear regression equations. Elevation and clay content were identified as the most effective combination for soil water storage estimation in the state-space approach, and were used to effectively predict the soil water storage spatial pattern along the second transect. The state-space approach is thus a useful tool that is recommended for predicting soil water storage spatial patterns at the hillslope scale using topography and soil properties.

© 2015 Elsevier B.V. All rights reserved.

1. Introduction

Soil water storage is a key state variable for understanding hydrologic and climatic processes, such as partitioning of precipitation and snowmelt into infiltration and runoff, percolation, and evapotranspiration (Famiglietti et al., 1998; Hu et al., 2010b; Western et al., 2004). It also has an important role in maintaining ecosystem health by regulating the transport of sediment and chemicals to environmentally sensitive areas (e.g., surface water and groundwater) (Biswas and Si, 2011a; Sun, 1986). Additionally, soil water storage is the principal limiting factor in arid and semiarid ecosystems because it controls the transpiration demand of the plant community (Gao et al., 2013a; Hu et al., 2009; Hu et al., 2010a). Therefore, understanding soil water storage and its spatial patterns is a prerequisite for improving hydrologic and climatic models (Biswas, 2014; Western et al., 2002). Furthermore, accurate estimation of soil water storage can provide essential information for the rational management of water resources and the successful restoration of vegetation on the Loess Plateau, China (Gao et al., 2013b).

Soil water storage is well recognized as being spatially variable (Jia et al., 2013) and influenced by many factors, such as topography, soil properties, vegetation, and meteorological conditions (Biswas, 2014; Brocca et al., 2010; Gómez-Plaza et al., 2001; Western et al., 1999). In particular, soil water storage is greatly affected by topography, e.g., elevation, slope, and curvature. Work at the field scale by Charpentier and Groffman (1992) revealed that the spatial variation in soil water content increased with increasing topographic heterogeneity. Tomer and Anderson (1995) demonstrated that topography was the major controlling factor and that combinations of elevation, slope, and curvature could explain 51–77% of the variability in soil water storage across a sand plain hillslope. Penna et al. (2009) reported that slope and topographic wetness index were the best univariate spatial predictors of soil moisture at a hillslope scale.

Soil texture, which determines the water-holding capacity, is another important factor affecting soil water storage. Vachaud et al. (1985) indicated that the spatial patterns of soil moisture at the field scale could in large part be explained by variability in soil texture. da Silva et al. (2001) identified soil texture as a major factor controlling soil moisture, with clay content positively correlated with soil moisture. Mohanty and Skaggs (2001) demonstrated that fields with sandy loam soils had better soil moisture stability over time than those containing silt loam soils. Soil fractal dimension, reflecting the soil structure and particle-size distribution, also likely affects soil water storage (Arya and Paris, 1981; Tyler and Wheatcraft, 1992). In addition, soil water storage is closely related to vegetation and precipitation. Canopy cover, root characteristics, and litter depth can influence runoff, interception, evapotranspiration,
and deep percolation, and thus affect soil moisture dynamics (Gómez-Plaza et al., 2001; Jacobs et al., 2004). Soil moisture is highly correlated with precipitation, increasing sharply following rainfall events and decreasing slowly over periods without rainfall (Gao et al., 2013a). Although precipitation exerts an extensive effect on soil moisture, especially during storms, soil texture is more important than precipitation in quantifying this influence (Yoo et al., 1998).

Heterogeneity in controlling factors and their combinations can create high spatial variability in soil water storage. To better understand this spatial variability, quantifying the relationships between soil water storage and its potential influencing factors is very important. Therefore, a large volume of research has been done to estimate soil water storage based on easily measured variables at various scales (Gao et al., 2013b; Penna et al., 2009; Qiu et al., 2003; Tomer and Anderson, 1995; Western et al., 1999). However, the analytical approach used in such studies is based on classical statistics. Such methods consider variables to be spatially independent of each other, with no spatial structure, and thus can generate erroneous or misleading results (Nielsen and Alemi, 1989). In contrast, the autoregressive state-space approach, by means of the Kalman filter, provides opportunities for suitable identification of the spatial process of a variable by taking into account spatial associations (Timm et al., 2003a; Wendroth et al., 2003).

The state-space approach has proven more effective than classical statistical methods for identifying localized variation (Stevenson et al., 2001; Timm et al., 2004), and has been widely applied to estimate soil properties and vegetation yield and generate reliable predictions (Cassel et al., 2000; Li et al., 2001; Nielsen and Alemi, 1989; Stevenson et al., 2001; Timm et al., 2003a; Wendroth et al., 1992; Wendroth et al., 1999; Wendroth et al., 2003). The state-space approach has received increasing attention in the past ten years. For example, Joschko et al. (2006) successfully estimated earthworm biodiversity based on pH and total nitrogen values along a regional scale transect. Jia et al. (2011) described the spatial distribution of total net primary productivity of managed grasslands in a small catchment. Liu et al. (2012) and She et al. (2014) both reported that the state-space approach described spatial variation in soil organic carbon much better than the equivalent linear regression equation. Aquino et al. (2015) reported on the effect of land leveling on the spatial relationships of soil properties in a 1 ha lowland area. However, few studies have quantitatively predicted soil water storage using a state-space approach and previous state-space approaches have not been validated. Although Morkoc et al. (1985) studied the spatial association of soil moisture using a state-space approach, they only focused on the uppermost 5 cm. Thus, there is a strong need to identify the spatial variability of soil water storage in diverse soil layers in the profile from 0 to 3 m using a state-space approach.

The Loess Plateau, a severely eroded area, has received considerable and extended attention in China. Afforestation was a useful way to prevent soil erosion and hence the “Grain for Green” project was implemented in 1999 to plant trees and convert slope cropland to forest, shrub, and grassland (Fu et al., 2006). Black locust (Robinia pseudoacacia L.) is a promising tree for reforestation due to its fast growth, superior drought tolerance, and extensive cover area in the Loess Plateau as a non-native tree species (Qiu et al., 2010). However, afforestation of the area has produced some negative effects, such as the emergence of non-native tree species (Qiu et al., 2010). The prevailing landform is loessial tableland and gullyland, covering 35 and 65% of the watershed, respectively. This area is characterized by a continental monsoon climate with a mean temperature of 9.2 °C. The mean annual precipitation is 582.3 mm, more than 58.2% of which falls between July and September. The groundwater table is about 50–80 m below surface and agricultural production on the tableland mainly relies on natural rainfall (no irrigation). The soil, derived from wind-deposited loess, belongs to the loessial soil group according to the FAO-UNESCO soil classification system. Dominant plant species in this region include wheatgrass (Agropyron cristatum L.), green bristlegrass (Setaria viridis L.), black locust (R. pseudoacacia L.), and Chinese arborvitae (Platycladus orientalis L.).

2. Materials and methods

2.1. Study area

The study was conducted in the Wangdonggou watershed (35°12′–35°16′N, 107°40′–107°42′E; elevation 946 to 1226 m above sea level, area 8.3 km²), located in Changwu County, Shaanxi Province, China. The prevailing landform is loessial tableland and gullyland, covering 35 and 65% of the watershed, respectively. This area is characterized by a continental monsoon climate with a mean temperature of 9.2 °C. The mean annual precipitation is 582.3 mm, more than 58.2% of which falls between July and September. The groundwater table is about 50–80 m below surface and agricultural production on the tableland mainly relies on natural rainfall (no irrigation). The soil, derived from wind-deposited loess, belongs to the loessial soil group according to the FAO-UNESCO soil classification system. Dominant plant species in this region include wheatgrass (Agropyron cristatum L.), green bristlegrass (Setaria viridis L.), black locust (R. pseudoacacia L.), and Chinese arborvitae (Platycladus orientalis L.).

2.2. Experimental design

After a detailed field survey, one typical hillslope covered with black locust was selected as the study site. Two 187 m long transects were laid out along the hillslope with a mean slope of 36.4%. A total of 35 locations were selected for the installation of access tubes for soil water measurements and sampling along each of two transects (A and B) (Fig. 1). The distance between sampling locations was 5.5 m, while the distance between the two transects was 10 m. The elevation of each location was measured using differential kinematic GPS. Because the state-space approach is designed for variables taken in one dimension, observations for each transect were conducted starting from the 0–1 m soil layer at the first location and ending in the 2–3 m soil layer at the 35th location (Fig. 1) (Wendroth et al., 2003). Transects A and B both had 105 monitoring positions. The variables obtained from transect A were used to establish the state-space approach, and observations from transect B were used to validate the state-space approach.

2.3. Soil sampling and measurements

One 3-m soil core per location along both transects was taken using a soil auger (5 cm in diameter). Fifteen distributed samples were collected at 0.2 m increments from each soil core for soil particle analysis. Each soil sample was air-dried and passed through a 2-mm sieve. Soil particle sizes were measured using a Mastersizer 2000 (Malvern Instruments, Malvern, England) with three replicates. An access tube was installed at each location to measure soil water content to a depth of 3.0 m with a neutron probe (CNC-503B DR, ChaoNeng, China) calibrated using standard methods (Hauser, 1984; Huang and Gallichand, 2006). Soil water measurements were conducted nine times from May to December 2014 at each location, at increments of 0.1 and 0.2 m in the 0–1 and 1–3 m soil layers, respectively. If rainfall event happened, the measurement was conducted on the fourth day after rainfall event. The values of soil water storage (mm) for the 0–1, 1–2, and 2–3 m soil layers were calculated from the measured soil water content. To better apply the state-space approach, the mean soil water storage value of all nine measurements was used for analyses at each location.

The mean soil properties (e.g., clay, silt, sand content, soil median grain size, and fractal dimension) in the three soil layers were calculated to identify the spatial variability of soil water storage in diverse soil
layers. The median soil grain size was calculated using the particle size distribution. The soil fractal dimension \( D \) was calculated using (Tyler and Wheatcraft, 1992)

\[
\frac{M(\delta < d_i)}{M_T} = \left( \frac{d_i}{d_{\text{max}}} \right)^{3-D}
\]

where \( \delta \) is the soil particle size; \( M \) is the cumulative mass of particles of the \( i \)th size class when \( \delta \) is less than \( d_i \); \( M_T \) is the total mass; \( d_i \) is the mean particle diameter \((\text{mm})\) of the \( i \)th size class; and \( d_{\text{max}} \) is the mean diameter of the largest particle. The mean particle diameter was taken as the arithmetic mean of the upper and lower sieve sizes. Taking the logarithm of both sides of Eq. (1), we derived Eq. (2) to solve for \( D \):

\[
D = 3 - \frac{\log(M(\delta < d_i) / M_T)}{\log(d_i / d_{\text{max}})}
\]

2.4. The state-space approach

State-space approach is a multivariate autoregressive technique adopted from applied time series analysis that can be used to quantify the spatial coincidence of a set of variables (Morkoc et al., 1985; Shumway, 1988). In the analysis, the state of a variable or set of variables, observed at location \( i \), is related to its state at location \( i - h \), where \( h \) ranges from 1 to \( n - 1 \) (\( n \) refers to the total sampling numbers) (Timm et al., 2003a; Wendroth et al., 2003). The state-space approach generally consists of two equations (state equation and observation equation) to represent the relationship between the input and output of a dynamic system (Shumway, 1988). The state equation is calculated using:

\[
Z_i = \Phi Z_{i-1} + \omega_i,
\]

where \( Z_i \) is the state vector (of a set of \( p \) variables) at location \( i \); \( \Phi \) is a \( p \times p \) matrix of state coefficients that indicates spatial regression; and \( \omega_i \) is the uncorrelated zero mean model error. This is the usual structure of a common autoregressive model in which the coefficients of the matrix \( \Phi \) could be calculated by multiple regressions, taking \( Z_i \) as the dependent variable and \( Z_{i-1} \) as the independent variable (Timm et al., 2003b). In the case of the state-space approach, however, the true state of the variable is considered “embedded” in the observation equation according to:

\[
Y_i = MZ_i + \nu_i
\]

where the observation vector \( Y_i \) is related to the state vector \( Z_i \) by an observation matrix \( M \) and an observation noise vector \( \nu_i \), also considered to be zero-mean, uncorrelated, and normally distributed. The noises \( \omega_i \) and \( \nu_i \) are assumed to be independent of one other (Shumway, 1988).

Prior to the state-space approach, the observed data were scaled to remove differences in the order of magnitude among variables using:

\[
z_i = [Z_i - (m - 2s)] / 4s
\]

where \( z_i \) is a scaled value (dimensionless) with a mean of 0.5 and a standard deviation of 0.25 (Wendroth et al., 2003), and \( m \) and \( s \) are the mean and standard deviation of the \( Z_i \) data, respectively. This transformation enables state coefficients of \( \Phi \) having magnitudes directly proportional to their contribution to each state variable to be used in the state-space approach (Jia et al., 2011; Timm et al., 2003a).

2.5. Statistical analysis

Descriptive statistical parameters (e.g., minimum, maximum, mean, standard deviation, coefficient of variation) were calculated to illustrate the basic trends in soil water storage and its influencing factors. Pearson correlation analysis was applied to determine the strength of linear dependence between the scaled soil water storage (SSWS) and the other variables. Relationships between SSWS and the combination of its influencing factors were modeled by classical linear regression equations. All classical statistics were performed using SPSS 17.0 software (SPSS Inc., 2008).

SSWS values were also estimated by the state-space approach. Estimated values and the coefficients of the matrix \( \Phi \) of the state equation (Eq. (3)) were obtained using Applied Statistical Time Series Analysis (ASTSA) software developed by Shumway (1988). The coefficient of
determination ($R^2$) and root mean square error (RMSE) were used to evaluate the performance of both the classical linear regression and state-space approaches. The RMSE was calculated using

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (O_i - P_i)^2},$$

where $O_i$ and $P_i$ are the observed and predicted values of SSWS, respectively.

3. Results and discussion

3.1. Spatial patterns of soil water storage and influencing factors

Table 1 presents descriptive statistics representing soil water storage data and potential influencing variables across the 105 sampling locations (transect A). Soil water storage varied from 121.93 to 258.18 mm (mean 159.23 mm), which was within the range suggested by Liu and Shao (2014) for the Loess Plateau. The coefficient of variation (CV) is an index that shows the overall variation or heterogeneity of a given variable. The CV values for soil water storage and other variables ranged from 0.8% (fractal dimension) to 41.0% (sand content) (Table 1). According to Nielsen and Bouma (1985), the elevation, silt content, and fractal dimension demonstrated weak variability (CV ≤ 10%), while soil water storage, clay, sand content, and soil median grain size had moderate variability (10% < CV ≤ 100%). Moderate degrees of soil water storage variability at the hillslope scale have been reported by other researchers (Biswas and Si, 2011b; Gao and Shao, 2012; Jia et al., 2013; Liu and Shao, 2014; Tomer and Anderson, 1995). Pearson correlation analysis showed that soil water storage was positively correlated with elevation, clay content, and fractal dimension and negatively correlated with silt content ($P < 0.01$, Table 1). These results are consistent with the findings of da Silva et al. (2001) with respect to soil moisture increasing with clay content. No statistically significant relationships were found between soil water storage and sand content or soil median grain size ($P > 0.05$, Table 1). Thus, the elevation, clay, silt content, and fractal dimension were expected to estimate soil water storage by both state-space and classical statistical approaches.

The spatial patterns of soil water storage and pertinent variables across the 105 sampling locations are shown in Fig. 2. Most variables (except elevation) exhibited remarkable point-to-point fluctuations compared to overall variation, as indicated by the CV. Elevation had an obvious spatial distribution trend along the transect (Fig. 2), which caused a strong spatial dependence (Jia et al., 2011). The spatial distribution of soil water storage was characterized by the localized variation and, therefore, might be better represented by a local model (e.g., a state-space approach) rather than space-independent models (e.g., standard multiple regression) (Liu et al., 2012; Timm et al., 2003a).

3.2. Autocorrelation and cross-correlation functions

Before estimation of SSWS by the state-space approach, spatial correlation structures among variables should be evaluated (Li et al., 2001). Hence, autocorrelation and cross-correlation functions were calculated as per methods detailed elsewhere (Aquino et al., 2015; Timm et al., 2003b). The autocorrelation function (ACF) identifies how a variable is spatially correlated with itself and thus provides information about the lag distance over which a measured value is related to its neighbors (Aquino et al., 2015; She et al., 2014). Using a $t$-test and a 95% confidence interval, spatial dependence between adjacent observations of SSWS was shown to be significant up to 7 lags (Fig. 3a). Elevation and silt content were both strongly autocorrelated within ranges of 12 and 11 lags, respectively (Fig. 3a, d). The ACFs of clay content and fractal dimension also manifested significant spatial autocorrelations up to 6 lags (Fig. 3c, e). If the autocorrelation coefficient is higher than the confidence interval in at least 1 lag, the state-space approach can be carried out (Cassel et al., 2000; Joschko et al., 2006; Timm et al.,...
Consequently, the sampling distance used in the present study for SSWS, elevation, clay, silt content, and fractal dimension was sufficient to identify their spatial representativity.

The cross-correlation function (CCF) can be used to quantify the degree of linear association between pairs of variables separated by distance (Cassel et al., 2000; Shumway et al., 1989). It can evaluate the correlation structure of their spatial distributions and further provide deep insight into the spatial covariance structure (Aquino et al., 2015; Timm et al., 2003a). Similar to ACF, the CCF was calculated to determine the spatial correlation structure between SSWS and elevation, clay, silt content, and fractal dimension, respectively (Fig. 4). At the 95% confidence level, SSWS was positively cross-correlated with elevation (Fig. 4a), clay content (Fig. 4b), and fractal dimension (Fig. 4d) and negatively cross-correlated with silt content (Fig. 4c). T-test results at the 5% probability level indicated the cross-correlation coefficients between SSWS and elevation had a strong spatial dependence (Fig. 4a). Spatial dependence between SSWS and silt content was significant up to 6 lags (Fig. 4c). The cross-correlogram between SSWS and clay content, however, showed weak spatial dependence (Fig. 4b). A similar result was found between SSWS and fractal dimension (Fig. 4d). The CCFs for SSWS and its significant controlling variables were sufficient to describe their distributions along the transect using the state-space approach. The results of ACF and CCF analyses indicated that elevation, clay, silt content, and fractal dimension are appropriate for estimating SSWS using a state-space approach.

3.3. State-space approaches of SSWS estimation

State-space approaches can be used to quantify how strongly SSWS at position $i$ is spatially based on itself and other variables at position $i-1$. This means that if the first SSWS value along the transect is known, all ensuing SSWS values can be calculated using the previous one and possible permutations of four variables (i.e., elevation, clay, silt content, fractal dimension). Table 2 presents all of the autoregressive state-space approaches and $R^2$ and RMSE values. To systematically compare the prediction results, the state-space approaches were divided into four groups: univariate, bivariate, trivariate, and quadrivariate equations.

For the four univariate state-space approaches, the best performance was found for the equation that included silt content ($R^2 = 0.932$, $RMSE = 0.068$). Use of elevation, clay content, or fractal dimension also predicted SSWS well, explaining 92.7, 92.1, and 90.3% of the total variation of SSWS, respectively (Table 2). Examining the univariate equations showed that the SSWS value at position $i-1$ contributed 72.8–83.7% to the estimated SSWS at location $i$, while elevation, clay, silt content, or fractal dimension only contributed 10.6–21.5%
All variables have been scaled using Eq. (5).

**Table 2**
State-space equations of scaled soil water storage (SSWS) using elevation (E), clay content (Wclay), silt content (Wsilt), and fractal dimension (D).

<table>
<thead>
<tr>
<th>State-space equations</th>
<th>$R^2$</th>
<th>RMSE</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Univariate equations</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(SSWS) = 0.728(SSWS)$_i - 1$ + 0.215(E)$_i - 1$ + $w_i$</td>
<td>0.927**</td>
<td>0.070</td>
</tr>
<tr>
<td>(SSWS) = 0.787(SSWS)$_i - 1$ + 0.155(Wclay)$_i - 1$ + $w_i$</td>
<td>0.921**</td>
<td>0.072</td>
</tr>
<tr>
<td>(SSWS) = 0.837(SSWS)$_i - 1$ + 0.108(Wsilt)$_i - 1$ + $w_i$</td>
<td>0.932**</td>
<td>0.068</td>
</tr>
<tr>
<td>(SSWS) = 0.787(SSWS)$_i - 1$ + 0.155(D)$_i - 1$ + $w_i$</td>
<td>0.903**</td>
<td>0.080</td>
</tr>
<tr>
<td><strong>Bivariate equations</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(SSWS) = 0.652(SSWS)$_i - 1$ + 0.192(E)$_i - 1$ + 0.123(Wclay)$_i - 1$ + $w_i$</td>
<td>0.988**</td>
<td>0.029</td>
</tr>
<tr>
<td>(SSWS) = 0.720(SSWS)$_i - 1$ + 0.192(E)$_i - 1$ + 0.040(Wsilt)$_i - 1$ + $w_i$</td>
<td>0.934**</td>
<td>0.066</td>
</tr>
<tr>
<td>(SSWS) = 0.652(SSWS)$_i - 1$ + 0.191(E)$_i - 1$ + 0.123(D)$_i - 1$ + $w_i$</td>
<td>0.986**</td>
<td>0.032</td>
</tr>
<tr>
<td>(SSWS) = 0.755(SSWS)$_i - 1$ + 0.130(Wclay)$_i - 1$ + 0.077(Wsilt)$_i - 1$ + $w_i$</td>
<td>0.895**</td>
<td>0.082</td>
</tr>
<tr>
<td>(SSWS) = 0.762(SSWS)$_i - 1$ + 0.094(Wclay)$_i - 1$ + 0.094(D)$_i - 1$ + $w_i$</td>
<td>0.886**</td>
<td>0.086</td>
</tr>
<tr>
<td>(SSWS) = 0.755(SSWS)$_i - 1$ + 0.077(Wsilt)$_i - 1$ + 0.130(D)$_i - 1$ + $w_i$</td>
<td>0.895**</td>
<td>0.083</td>
</tr>
<tr>
<td><strong>Trivariate equations</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(SSWS) = 0.651(SSWS)$_i - 1$ + 0.183(E)$_i - 1$ + 0.119(Wclay)$_i - 1$ + 0.017(Wsilt)$_i - 1$ + $w_i$</td>
<td>0.977**</td>
<td>0.040</td>
</tr>
<tr>
<td>(SSWS) = 0.635(SSWS)$_i - 1$ + 0.186(E)$_i - 1$ + 0.075(Wclay)$_i - 1$ + 0.074(D)$_i - 1$ + $w_i$</td>
<td>0.923**</td>
<td>0.071</td>
</tr>
<tr>
<td>(SSWS) = 0.651(SSWS)$_i - 1$ + 0.183(E)$_i - 1$ + 0.075(Wsilt)$_i - 1$ + 0.074(D)$_i - 1$ + $w_i$</td>
<td>0.988**</td>
<td>0.030</td>
</tr>
<tr>
<td>(SSWS) = 0.768(SSWS)$_i - 1$ + 0.119(D)$_i - 1$ + 0.080(Wclay)$_i - 1$ + 0.070(Wsilt)$_i - 1$ + 0.080(D)$_i - 1$ + $w_i$</td>
<td>0.888**</td>
<td>0.085</td>
</tr>
<tr>
<td><strong>Quadrivariate equation</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(SSWS) = 0.635(SSWS)$_i - 1$ + 0.180(E)$_i - 1$ + 0.074(Wclay)$_i - 1$ + 0.011(Wsilt)$_i - 1$ + 0.073(D)$_i - 1$ + $w_i$</td>
<td>0.979**</td>
<td>0.038</td>
</tr>
</tbody>
</table>

All variables have been scaled using Eq. (5). $w_i$: model noises; $R^2$: coefficient of determination; RMSE: root mean square error; **: $P < 0.01$.
the bivariate, trivariate, and quadrivariate equations all slightly underestimated SSWS when the observed SSWS was large. The data in Table 2 and Fig. 5 indicated elevation and clay content were the best combination (bivariate approach; $R^2 = 0.988$) for representing the SSWS spatial pattern.

3.4. Comparison of modeling results using the two methodologies

Classical linear regression equations were also established with elevation, clay, silt content, and fractal dimension and sorted into four groups and compared with the results obtained using the space-state approach. The classical equations were all significant at $P < 0.01$ (Table 3). Among the univariate linear regressions, elevation was deemed the most suitable variable to interpret the variability in SSWS because it had relatively high $R^2$ and low RMSE values (Table 3). A similar result was found in the Pearson correlation analysis (Table 1). For the univariate state-space approaches, however, the equation using elevation had a lower $R^2$ than the equation using silt content (Table 2). Among all of the linear regression equations, the combination of all four independent controlling factors was the best model; however, it could only explain 64.2% of the total variation in SSWS. The inclusion of more variables in the classical linear regression equations could ensure better performance (Table 3), but this result was notably different from the state-space approaches for which performance was not improved by increasing the number of variables (Jia et al., 2011; Liu et al., 2012).

Comparing the results in Tables 2 and 3 indicates that all state-space approaches provide better estimations (higher $R^2$ and lower RMSE values) of SSWS than the equivalent classical linear regression equations with the same variables at the hillslope scale. This is attributed to the fact that the state-space approaches could identify spatial relationships among the variables while classical linear regression equations ignore local behavioral tendencies of observations (Jia et al., 2011; Nielsen and Alemi, 1989; Wendroth et al., 1999).

3.5. Prediction of soil water storage using the state-space approach

The best state-space approach for predicting SSWS along transect A for each group (univariate, bivariate, trivariate, quadrivariate) was applied to transect B. As shown in Fig. 6, the univariate model using silt content alone explained 94.2% of the total variation of SSWS. When the values of SSWS were low, the predicted values of SSWS were close to observed values. When the values of SSWS were great, however, the points were scattered and the predicted results were not in good agreement with the observed values. The data in Table 2 and Fig. 5 indicated elevation and clay content were the best combination (bivariate approach; $R^2 = 0.988$) for representing the SSWS spatial pattern.

![Fig. 5. The best (a) univariate, (b) bivariate, (c) trivariate, and (d) quadrivariate state-space models of scaled soil water storage (SSWS) along transect A. All variables have been scaled using Eq. (5).](image)

### Table 3

<table>
<thead>
<tr>
<th>Linear and multiple regression equations</th>
<th>$R^2$</th>
<th>RMSE</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Univariate regression</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SSWS = 0.198 + 0.605E</td>
<td>0.365**</td>
<td>0.198</td>
</tr>
<tr>
<td>SSWS = 0.317 + 0.367Wclay</td>
<td>0.135**</td>
<td>0.231</td>
</tr>
<tr>
<td>SSWS = 0.756 − 0.511Wsilt</td>
<td>0.261**</td>
<td>0.214</td>
</tr>
<tr>
<td>SSWS = 0.323 + 0.354D</td>
<td>0.126**</td>
<td>0.233</td>
</tr>
<tr>
<td><strong>Bivariate regression</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SSWS = −0.034 + 0.643E + 0.425Wclay</td>
<td>0.545**</td>
<td>0.168</td>
</tr>
<tr>
<td>SSWS = 0.454 + 0.612E − 0.520Wsilt</td>
<td>0.630**</td>
<td>0.150</td>
</tr>
<tr>
<td>SSWS = −0.024 + 0.629E + 0.408D</td>
<td>0.531**</td>
<td>0.157</td>
</tr>
<tr>
<td>SSWS = 0.787 − 0.013Wclay − 0.520Wsilt</td>
<td>0.261**</td>
<td>0.214</td>
</tr>
<tr>
<td>SSWS = 0.316 + 0.335Wsilt + 0.033D</td>
<td>0.135**</td>
<td>0.231</td>
</tr>
<tr>
<td>SSWS = 0.742 − 0.501Wsilt + 0.016D</td>
<td>0.262**</td>
<td>0.214</td>
</tr>
<tr>
<td><strong>Trivariate regression</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SSWS = 0.370 + 0.620E + 0.093Wclay − 0.453Wsilt</td>
<td>0.640**</td>
<td>0.149</td>
</tr>
<tr>
<td>SSWS = −0.034 + 0.643E + 0.414Wclay + 0.011D</td>
<td>0.545**</td>
<td>0.168</td>
</tr>
<tr>
<td>SSWS = 0.366 + 0.620E − 0.453Wsilt + 0.100D</td>
<td>0.641**</td>
<td>0.149</td>
</tr>
<tr>
<td>SSWS = 0.770 − 0.192Wclay − 0.520Wsilt + 0.181D</td>
<td>0.264**</td>
<td>0.213</td>
</tr>
<tr>
<td><strong>Quadrivariate regression</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SSWS = 0.373 + 0.619E − 0.046Wclay − 0.460Wsilt + 0.140D</td>
<td>0.642**</td>
<td>0.148</td>
</tr>
</tbody>
</table>

All variables have been scaled using Eq. (5). $R^2$: coefficient of determination; RMSE: root mean square error; **: $P < 0.01$. 

### Table 2

Classical linear regression equations of scaled soil water storage (SSWS) using elevation ($E$), clay content ($W_{clay}$), silt content ($W_{silt}$), and fractal dimension ($D$).
agreement with observed values (Fig. 6a). For the other three equations (Fig. 6b, c, d), the approaches performed better than the univariate model as indicated by higher $R^2$ and lower RMSE values, but these three models still slightly underestimated SSWS when values were great. Whether for transect A or B, the combination of elevation and clay content was the best state-space approach (highest $R^2$ and lowest RMSE). Consequently, the autoregressive state-space approach appears to provide a quantitatively meaningful means to estimate SSWS and is a powerful tool for understanding the spatial relationships between SSWS and its influencing factors at the hillslope scale. Nevertheless, SSWS values are also strongly influenced by climate, which is such an important factor that its influence may even exceed that of soil properties and topography in some circumstances. Further studies are needed to evaluate the applicability of state-space approaches for predicting SSWS under different soil and climatic types.

4. Conclusions

This study was conducted to quantify the relationships between soil water storage and its potential influencing factors and to estimate soil water storage using classical linear regression and state-space approaches at the hillslope scale on the Loess Plateau. Soil water storage values were strongly influenced by topography (elevation) and soil properties (clay, silt content, fractal dimension). The state-space approach using any combination of variables performed better than the equivalent classical linear regression equation in modeling localized variation of SSWS. The best performing state-space approach ($R^2 = 0.988$, RMSE = 0.029) was a bivariate equation including elevation and clay content. This equation was applied to independent data from a second transect and found to provide very good estimations of soil water storage, explaining 99.8% of the total variation. Hence, the state-space approach is a powerful tool for better understanding the spatial relationships between soil water storage and its influencing factors and provides a quantitatively meaningful means for accurate estimation of soil water storage in different soil depths at the hillslope scale. Future work is necessary to assess the performance of state-space models under different soil and climatic conditions.

Acknowledgments

Financial assistance for this work was provided by the National Natural Science Foundation of China (No. 41390463 and 41171186). The authors thank members of the Changwu Ecology Station, Chinese Academy of Sciences, and Ministry of Water Resources for the assistance with field measurements.

References


Fig. 6. Measured vs. estimated scaled soil water storage (SSWS) along transect B by state-space models using (a) silt content ($W_{silt}$); (b) elevation ($E$) and clay content ($W_{clay}$); (c) $E$, $W_{clay}$ and fractal dimension ($D$); and (d) $E$, $W_{ch}, W_{clay}$ and $D$ respectively.
Mohanty, B., Skaggs, T., 2001. Spatio-temporal evolution and time-stable characteristics of
Morkoc, F., Biggar, J., Nielsen, D., Rolston, D., 1985. Analysis of soil water content and tem-
Qiu, L., Zhang, X., Cheng, J., Yin, X., 2010. Effects of black locust (Robinia pseudoacacia) on