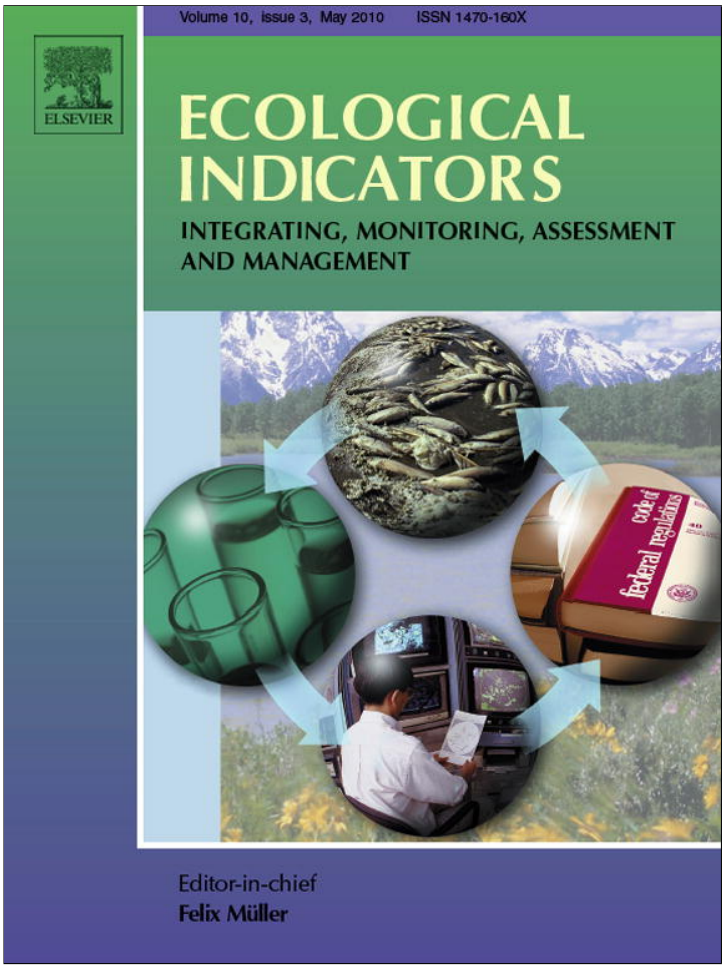

Ecological Indicators 10 (2010) 610–619

Contents lists available at ScienceDirect

Ecological Indicators

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

Mapping soil organic matter using the topographic wetness index: A comparative study based on different flow-direction algorithms and kriging methods

Tao Pei [a,b,1], Cheng-Zhi Qin [a,2], A-Xing Zhu [a,c,*], Lin Yang [a,2], Ming Luo [a,1], Baolin Li [a,1], Chenghu Zhou [a,1]

[a] State Key Laboratory of Resources and Environmental Information System, Institute of Geographical Sciences and Natural Resources Research, CAS, 11A, Datun Road Anwai, Beijing 100101, China
[b] Key Laboratory of Geographic Information Science, Ministry of Education, East China Normal University, Shanghai 200062, China
[c] Department of Geography, University of Wisconsin Madison, 550N, Park Street, Madison, WI 53706-1491, USA

ARTICLE INFO

Article history:
Received 1 January 2009
Received in revised form 22 August 2009
Accepted 13 October 2009

Keywords:
Collocated cokriging
Single-flow-direction (SFD)
Multiple-flow-direction (MFD)
Auto-correlation
Cross-correlation
Terrain variable
Digital soil mapping

ABSTRACT

Terrain attributes derived from digital elevation models have been used widely for mapping soil organic matter (SOM). Among these attributes, the topographic wetness index (TWI), an index for quantitatively indicating the balance between water accumulation and drainage conditions at the local scale, has been shown to correlate with SOM. However, TWIs used in most studies are calculated using a single-flow-direction (SFD) algorithm, which assumes that all water from a grid cell flows into only one neighboring cell. This assumption is not always valid, especially in areas with low relief where movement of water may be divergent. To overcome this SFD limitation, a multiple-flow-direction (MFD) algorithm has been developed, which distributes flow from a grid cell to several downslope neighbors. In this study we compared the effect of TWI calculations based on SFD and MFD in predictive mapping of SOM by incorporating them into different kriging methods over a 51.76 km^2 area in Nenjiang County of northeastern China. We found that the MFD-based TWI was better correlated with SOM than was the SFD-based index. We then compared the accuracies of SOM maps which were derived from MFD-based TWI and SFD-based TWI incorporated by ordinary kriging (OK), simple kriging with varying local means (SKlm), kriging with external drift (KED) and collocated cokriging (CC). The MFD-based TWI, used as a secondary variable in SKlm and CC, outperforms the SFD-based TWI. For the different kriging methods, CC (incorporating either MFD-based TWI or SFD-based TWI) showed the best performance, and OK generated a better result than SKlm and KED. Both the MFD-based TWI and SFD-based TWI proved to be incompatible with KED and SKlm due to their numerical instability caused by the rough TWI surfaces. Among all predictive methods, CC incorporating the MFD-based TWI produced the best results. This is because: (1) the MFD-based TWI is best able to indicate quantitatively soil moisture and therefore has the strongest correlation with SOM; (2) CC is capable of utilizing effectively the spatial auto-correlation of SOM and the cross-correlation between SOM and the MFD-based TWI.

1. Introduction

Terrain variables have been used widely in soil organic matter (SOM) content mapping as they can be incorporated into geostatistical methods and used as secondary variables (Bell et al., 2000; Mueller and Pierce, 2003). Terrain variables enhance SOM map quality and reduce the cost of sampling in three ways. First, terrain variables that are derived from a digital elevation model (DEM) can be acquired at a low cost. Second, terrain variables are exhaustive and spatially extensive, and provide potentially voluminous data sets, which provide relevant information at unsampled locations. The third and perhaps most important aspect is the significant correlation between terrain variables and SOM. Studies have revealed that soils with high moisture content increase SOM due to the promotion of plant growth and the slowdown of organic matter decomposition (Starr et al., 2000; Janzen et al., 2002). Since soil moisture can be modeled quantitatively by terrain indices, SOM might also exhibit significant correlation with terrain variables (Jenny, 1941; Moore et al., 1993; Janzen et al., 2002; Mueller and Pierce, 2003). The better soil moisture is quantified by a terrain variable, the stronger

* Corresponding author at: State Key Laboratory of Resources and Environmental Information System, Institute of Geographical Sciences and Natural Resources Research, CAS, 11A, Datun Road Anwai, Beijing 100101, China.
Tel.: +86 10 64888960; fax: +86 10 64889630.
E-mail addresses: peit@lreis.ac.cn (T. Pei), qincz@lreis.ac.cn (C.-Z. Qin), axing@lreis.ac.cn, axing@geography.wisc.edu (A.-X. Zhu), yanglin@lreis.ac.cn (L. Yang), luom@lreis.ac.cn (M. Luo), libl@lreis.ac.cn (B. Li), zhouch@lreis.ac.cn (C. Zhou).
[1] Tel.: +86 10 64888960; fax: +86 10 64889630.
[2] Tel.: +86 10 64889461; fax: +86 10 64889630.

1470-160X/$ – see front matter © 2009 Elsevier Ltd. All rights reserved.
doi:10.1016/j.ecolind.2009.10.005

the relation between SOM and the variable. Consequently, the higher the efficiency of the variable for the SOM mapping will be. Thus it is important to utilize improved terrain variables that may better describe soil moisture and enhance mapping quality.

Among the many terrain variables developed, the topographic wetness index (TWI) is considered a good indicator of soil moisture distribution at different landscape positions where overland flow dominates water transport processes, and may therefore show a significant relationship with SOM distribution (Beven and Kirkby, 1979). Moore et al. (1993) found significant correlations between terrain variables calculated from a DEM and measured soil attributes, and showed that slope gradient and TWI accounted for more than half of the variability of the A-horizon thickness and SOM. Bell et al. (1995) identified TWI and the depression proximity index as the most suitable predictors of the A-horizon and carbonate depth. Western et al. (1999) found that during wet periods TWI and contributing area explain up to 61% of the spatial variation in soil moisture (compared with several other terrain variables). Luca et al. (2007) proposed that terrain variables such as contributing area and TWI that incorporate upslope length could further improve SOM estimation, as these variables better reflect water redistribution. Sumfleth and Duttmann (2008) also showed that the distribution of soil carbon corresponds significantly to the TWI.

The relevance of TWI for SOM mapping can be explained by considering its definition (Eq. (1)):

$$\mathrm{TWI} = \ln\left(\frac{a}{\tan\beta}\right) \tag{1}$$

where α is the specific catchment area (SCA), and $\tan\beta$ is the local slope gradient. SCA indicates the potential flow accumulation to a specific location, and $\tan\beta$ reflects the local drainage potential (Beven and Kirkby, 1979; Quinn et al., 1995). The combination of SCA and local slope gradient represents the balance between water accumulation and drainage conditions at the local scale, which reflects soil moisture and SOM distribution.

The TWI can be computed using different algorithms. There are two significant approaches to determine SCA: the single-flow-direction (SFD) algorithm and the multiple-flow-direction (MFD) algorithm. SFD assumes that all water from a grid cell flows into only one neighboring cell (that with the lowest relative elevation), and MFD assumes that flow from the current position drains into more than one downslope neighboring cell (Wolock and McCabe, 1995). Previous studies have shown that MFD performs significantly better than SFD in constructing the spatial distribution of SCA or TWI (Quinn et al., 1991; Freeman, 1991; Fairfield and Leymarie, 1991; Bertolo, 2000). Although the significant differences between SCAs and TWIs resulting from different computation approaches have been investigated, few studies have characterized explicitly the effects of different TWI algorithms on SOM mapping.

To compare the effect of different TWI algorithms on SOM mapping, we incorporated TWIs based on these algorithms into different kriging methods. Geostatistical methods are some of the most widely used tools in estimating soil properties because they utilize the auto-correlation of the primary variable and the cross-correlation between the primary variable and secondary variables. There are two ways to combine the secondary variables into a kriging system. The first is to treat the primary variable as the summation of the local mean and the residual value. The local mean can be modeled with a regression function between the primary and the secondary variable, and the residual value, acquired by subtracting the local mean from the original data, can be kriged by simple kriging (Raspa et al., 1997; Hengl et al., 2007). The second approach is to use the cokriging method in which the secondary variable(s) is (are) embedded directly into the cokriging system (Journel and Huijbregts, 1978; Vauclin et al., 1983). Comparisons between these two methods have been made (Knotters et al., 1995; Goovaerts, 2000), showing that quality of estimation is dependant on not only the kriging strategy but also the performances of the variables, that is, the auto-correlation of primary variable and the cross-correlation between variables. In this regard, we think it is necessary to use different kriging methods to determine which combination (between different kriging methods and secondary variables) is the most efficient for SOM mapping.

In this paper, we compare of impact of MFD-based TWI and SFD-based TWI on SOM mapping by implementing different kriging methods in which the TWIs are employed as secondary variables. We first compute correlation coefficients between SOM and eight secondary terrain variables and identify that both SFD- and MFD-based TWI are correlated significantly with SOM. We then map SOM by incorporating both TWIs into different kriging methods: ordinary kriging (OK), simple kriging with varying local means (SKlm), kriging with an external drift (KED) and collocated cokriging (CC). Finally we evaluate the results generated by various combinations of selected secondary variables and kriging methods by means of cross-validation.

The remainder of the paper is arranged in five sections. Section 2 describes the research area and the data set. Section 3 presents the descriptions of correlation coefficients, all kriging methods and the evaluation criteria. In Section 4, the correlations between the derived secondary variables and SOM are compared and mapping results using different kriging strategies are presented. The impact of the MFD-based TWI on SOM mapping is analyzed through the comparison between various combinations in Section 5. Concluding remarks are presented in Section 6.

2. Study area and data

2.1. Study area

The study area is located in Heshan County, Heilongjiang Province, Northeastern China (48°53′24″–48°59′24″N, 125°8′24″–125°16′12″E) (Fig. 1a). It is situated in the Laolaihe watershed, which has a total area 51.76 km^2. Mollisols are the dominant soil class in this area (Zhang et al., 2007; Yang, 2007). Clay content ranges from 5% to 20%, and the texture is loam to clay loam. The area has low relief, with elevation ranging from 278 m to 362 m above sea level and a slope gradient below 5%. The average annual temperature at the site is 12.2 °C, and the average annual precipitation is between 400 and 600 mm. Overland flow dominates soil water redistribution in spring and summer. This is due to frozen subsoil does not thawing until late April preventing significant infiltration of spring precipitation and there being abundant rainfall in summer, both of which lead to overland flow (Zhang et al., 2007; Yang, 2007). The area is managed under conventional agricultural production and the predominant crops are soybean and spring wheat.

A 10 m-resolution DEM was derived from a 1:10,000 scale topographic map (published by Chinese Bureau of Surveying and Mapping, 1987) using ArcGIS 8.3. We collected 54 soil samples at different topographic locations in July 2005 (Fig. 1b). We dug 1.2 m-depth profiles at each location, and collected SOM samples from the A-horizon. The locations of the samples were determined by GPS handset, with a positional error $<$10 m. Soil thickness in the area varies from 0.30 m to 1.25 m. The samples were analyzed for SOM, N, P, K and particle size distribution.

2.2. Computation of TWIs and other related terrain variables

We derived eight terrain variables from the DEM: (1) profile curvature (P_cuv), (2) plan curvature (F_cuv), (3) horizontal

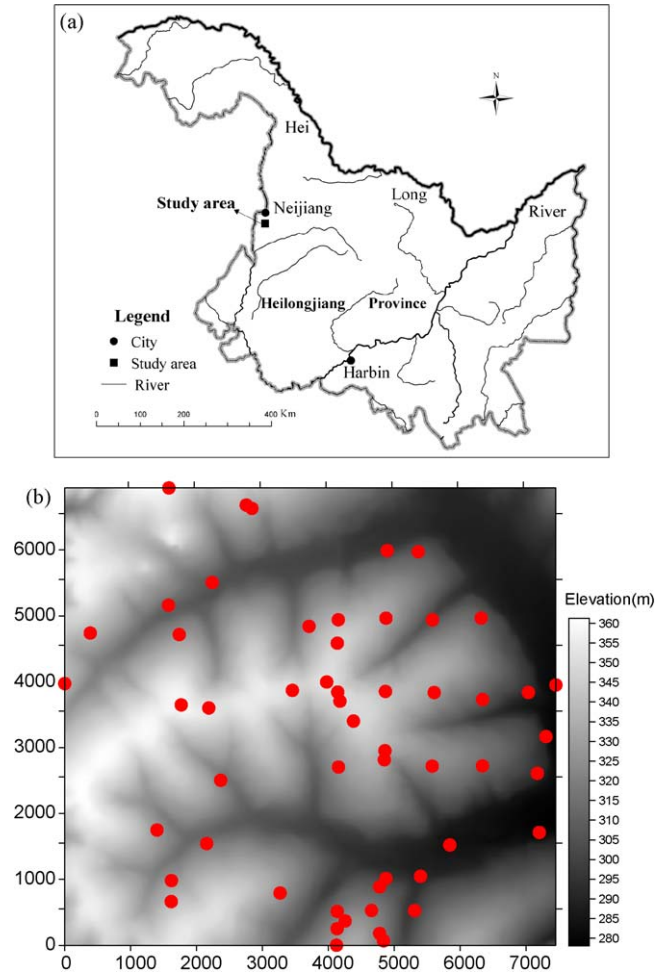

Fig. 1. Research area and locations of samples. (a) Location of the research area; (b) locations of samples and the DEM of the research area (unit of *x*- and *y*-axis: meter).

curvature (H_cuv), (4) slope, (5) α_{D8} (SCA computed with the D8 algorithm), (6) α_{MFD} (SCA computed with the MFD algorithm), (7) TWI_{D8} (TWI based on the D8 algorithm), (8) TWI_{MFD} (TWI based on the MFD algorithm). Variables (1)–(3) were computed according to Shary et al. (2002), (4) was computed according to Zevenbergen and Thorne (1987). Variables (1)–(4) were calculated using ArcGIS 8.3. Computations of the remaining terrain variables were implemented using the software SimDTA 1.0, which was developed by the second author, and their definitions are described as follows.

α_{D8} is an SCA index based on the D8 (SFD) algorithm, which assumes that water can only flow in the steepest direction of descent (O'Callaghan and Mark, 1984).

α_{MFD} is an SCA index based on the MFD algorithm proposed by Quinn et al. (1991). The algorithm's key feature is the determination of the flow-partitioning proportion of the *i*th downslope neighboring cell (d_i), which is calculated by

$$d_i = \frac{(\tan\beta_i)^p \times L_i}{\sum_{j=1}^{8} (\tan\beta_j)^p \times L_j} \quad (2)$$

where $\tan\beta_i$ is the slope gradient from the center cell to the *i*th neighboring cell, p is the weighting of flow partition, and L_i is the geometric weighting of the *i*th neighboring cell (Quinn et al., 1991). Note that the α_{MFD} approaches asymptotically the α_{D8} as p increases. The choice of the value of p has been discussed extensively (Quinn et al., 1991, 1995; Freeman, 1991; Qin et al., 2007). In this paper, we let $p = 1$, as proposed by Quinn et al. (1991), when calculating α_{MFD} and TWI_{MFD}.

From these two definitions of SCA, two TWIs are calculated based on the following equations:

$$TWI_{D8} = \ln\left(\frac{a_{D8}}{\tan\beta}\right) \quad (3)$$

$$TWI_{MFD} = \ln\left(\frac{a_{MFD}}{\tan\beta}\right) \quad (4)$$

2.3. Data preparation

Data should be preprocessed to identify outliers before being used in SOM mapping. Outliers are those values that differ significantly from the values of surrounding locations, and can alter significantly the results of the statistical analysis. Different methods for identifying and processing outliers have been proposed (Anselin, 1995; Getis and Ord, 1996). Due to our study area being small we use a global threshold ($\bar{z} + 2s$) to identify outliers, where $\bar{z}$ is the mean of the sampled values, and s is their standard deviation. Values exceeding the threshold are treated as outliers and assigned the threshold value. The advantage of this method is that we not only keep the processed outliers (i.e. retain their relatively high values) but also ensure the stability of the semi-variogram that is used extensively in this study.

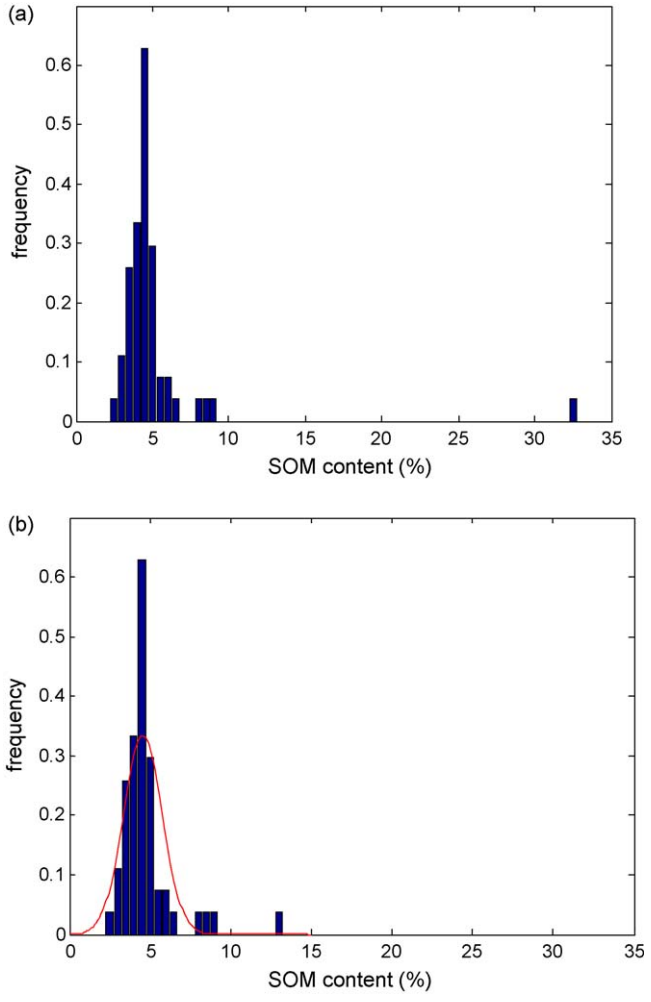

Fig. 2. Histograms of SOM data. (a) The raw data of SOM and (b) the processed data of SOM.

Table 1
Summary statistics of raw and processed SOM data.

Variables (unit)	Means	Std.	Min	Max	Normally distributed
SOM (raw data) (%)	5.12	4.02	2.59	32.64	No
SOM (processed) (%)	4.76	1.72	2.59	13.15	Yes

One SOM value was identified as an outlier (Fig. 2) and we reduced its value to the corresponding threshold. The processed SOM data were then tested for normality, which was confirmed (Fig. 2). Statistics of the SOM data before and after preprocessing are presented in Table 1. No outliers were found in the eight secondary variables, and their normality was also confirmed. Variables were standardized for correlation analysis and further estimation.

3. Methods

3.1. Correlation coefficient

We used Pearson correlation coefficients to measure the correlation between two variables. For variables x_i and y_i (i = 1,2, ..., N), the Pearson correlation coefficient between them is defined as

$$r = \frac{\sum_{i=1}^{N}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{N}(x_i - \bar{x})^2}\sqrt{\sum_{i=1}^{N}(y_i - \bar{y})^2}} \tag{5}$$

where $\bar{x}$ and $\bar{y}$ are the means of x_i and y_i, respectively (Rodgers and Nicewander, 1988).

3.2. Ordinary kriging (OK)

OK is a univariate estimator expressed as

$$z^*_{\mathrm{OK}}(x) = \sum_{\alpha=1}^{n(x)} \lambda_\alpha^{\mathrm{OK}}(x) z(x_\alpha) \quad \text{with} \quad \sum_{\alpha=1}^{n(x)} \lambda_\alpha^{\mathrm{OK}}(x) = 1. \tag{6}$$

where $z(x_\alpha)$ is the observation sampled at x_α, $n(x)$ is the number of neighboring observations, $\lambda_\alpha^{\mathrm{OK}}(x)$ is the weight for the observation of x_α. The weight $\lambda_\alpha^{\mathrm{OK}}(x)$ can be estimated by solving the following system of $n(x)+1$ equations (Journel and Huijbregts, 1978).

$$\begin{cases} \sum_{\beta=1}^{n(x)} \lambda_\beta^{\mathrm{OK}}(x)\gamma(x_\alpha - x_\beta) - \mu(x) = \gamma(x_\alpha - x) \quad \alpha = 1, \ldots, n(x) \\ \sum_{\beta=1}^{n(x)} \lambda_\beta^{\mathrm{OK}}(x) = 1 \end{cases} \tag{7}$$

where $\mu(x)$ is the Lagrange parameter, and $\gamma(\cdot)$ is the semi-variogram of the variable to be estimated.

3.3. Simple kriging with varying local means (SKlm)

The main idea of SKlm (some authors refer to it as regression kriging, e.g. Hengl et al., 2007) is to replace the known stationary mean in the simple kriging estimate by known varying local means, which can be derived from the secondary variable (Goovaerts, 1997). The SKlm estimator is

$$z^*_{\mathrm{SKlm}}(x) = m^*_{\mathrm{SK}}(x) + \sum_{\alpha=1}^{n(x)} \lambda_\alpha^{\mathrm{SK}}(x)[z(x_\alpha) - m^*_{\mathrm{SK}}(x_\alpha)] \tag{8}$$

where $m^*_{\mathrm{SK}}(x)$ is the local mean, $\lambda_\alpha^{\mathrm{SK}}(x)$ is the weight for the primary variable at x_α. $m^*_{\mathrm{SK}}(x)$ can be estimated through a regression function (usually linear) of the secondary variable: $m^*_{\mathrm{SK}}(x) = a_0(x) + a_1(x)y(x)$. $a_0(x)$ and $a_1(x)$ here are constant over the research area. $\lambda_\alpha^{\mathrm{SK}}(x)$ can be estimated by solving the following simple kriging (SK) system:

$$\sum_{\beta=1}^{n(x)} \lambda_\beta^{\mathrm{SK}}(x) C_{\mathrm{R}}^{\mathrm{SK}}(x_\alpha - x_\beta) = C_{\mathrm{R}}^{\mathrm{SK}}(x_\alpha - x), \quad \alpha = 1, \ldots, n(x) \tag{9}$$

where $C_{\mathrm{R}}^{\mathrm{SK}}(h)$ is the covariance function of the residue $R(x)$ ($R(x) = Z(x) - m^*_{\mathrm{SK}}(x)$).

3.4. Kriging with external drift (KED)

Like the SKlm approach, KED uses secondary information to derive the local mean of the primary variable and then performs simple kriging on the corresponding residues:

$$z^*_{\mathrm{KED}}(x) = m^*_{\mathrm{KED}}(x) + \sum_{\alpha=1}^{n(x)} \lambda_\alpha^{\mathrm{KED}}(x)[z(x_\alpha) - m^*_{\mathrm{KED}}(x_\alpha)] \tag{10}$$

where $m^*_{\mathrm{KED}}(x)$ is the local mean and can be modeled as $m^*_{\mathrm{KED}}(x) = a^*_0(x) + a^*_1(x)y(x)$, $\lambda_\alpha^{\mathrm{KED}}(x)$ is the weight for the primary variable at x_α. In KED, the regression coefficients ($a^*_0(x)$ and $a^*_1(x)$) are estimated implicitly through a kriging system (Eq. (11)) and varied with different locations (Goovaerts, 1997).

$$\begin{cases} \sum_{\beta=1}^{n(x)} \lambda_\beta^{\mathrm{KED}}(x) C_{\mathrm{R}}^{\mathrm{KED}}(x_\alpha - x_\beta) + \mu_0^{\mathrm{KED}}(x) + \mu_1^{\mathrm{KED}}(x) y(x_\alpha) = C_{\mathrm{R}}^{\mathrm{KED}}(x_\alpha - x) \quad \alpha = 1, \ldots, n(x) \\ \sum_{\beta=1}^{n(x)} \lambda_\beta^{\mathrm{KED}}(x) = 1 \\ \sum_{\beta=1}^{n(x)} \lambda_\beta^{\mathrm{KED}}(x) y(x_\beta) = y(x) \end{cases} \tag{11}$$

where the $C_{\mathrm{R}}^{\mathrm{KED}}(\cdot)$ is the covariance of the residual $R(x)$ ($R(x) = Z(x) - m^*_{\mathrm{KED}}(x)$).

3.5. Collocated cokriging (CC)

CC is a cokriging method in which the only incorporated secondary datum is the one collocated with the location x. The estimator of CC is

$$z^*_{\mathrm{CC}}(x) = \sum_{\alpha_1=1}^{n_1(x)} \lambda_{\alpha_1}^{\mathrm{CC}}(x) z_1(x_{\alpha_1}) + \lambda_2^{\mathrm{CC}}(x)[z_2(x) - m_2 + m_1] \tag{12}$$

with weights ($\lambda_{\alpha_1}^{\mathrm{CC}}(x)$) of the primary variable and that ($\lambda_2^{\mathrm{CC}}(x)$) of the secondary variable summed to 1 (i.e. $\sum_{\alpha_1=1}^{n_1(x)} \lambda_{\alpha_1}^{\mathrm{CC}}(x) + \lambda_2^{\mathrm{CC}}(x) = 1$), where m_1 and m_2 are means of the primary and the secondary variable, respectively, and $n_1(x)$ is the number of neighboring observations of the primary variable. The weights in Eq. (12) are estimated by solving the following kriging system.

Table 2
Correlation coefficients between terrain variables and SOM.

	Slope	P_cuv	F_cuv	H_cuv	α_{D8}	α_{MFD}	TWI_{D8}	TWI_{MFD}
SOM	−0.118	−0.100	−0.074	−0.092	0.190	0.158	0.307[a]	0.482[a]

[a] Significant at the 0.05 level.

$$\begin{cases} \sum_{\beta_2=1}^{n_1(x)} \lambda_{\beta_1}^{CC}(x) C_{11}(x_{\alpha_1} - x_{\beta_1}) + \lambda_2^{CC}(x) C_{12}(x_{\alpha_1} - x_{\beta_1}) \\ \quad + \mu^{CC}(x) = C_{11}(x_{\alpha_1} - x) \quad \alpha_1 = 1, \ldots, n_1(x) \\ \sum_{\beta_1=1}^{n_1(x)} \lambda_{\beta_1}^{CC}(x) C_{21}(x - x_{\beta_1}) + \lambda_2^{CC}(x) C_{22}(0) + \mu^{CC}(x) = C_{21}(0) \\ \sum_{\alpha_1=1}^{n_1(x)} \lambda_{\alpha_1}^{CC}(x) + \lambda_2^{CC}(x) = 1 \end{cases} \quad (13)$$

where $C_{11}(\cdot)$ is the covariance function of the primary variable, $C_{12}(\cdot)$ is the cross covariance function between the primary variable and the secondary variable, and $C_{22}(0)$ is the covariance function of secondary variable for $|h| = 0$ (Goovaerts, 1997).

3.6. Evaluation criteria

Due to the limited number of soil samples we used cross-validation to evaluate the impacts of TWI_{D8} and TWI_{MFD} on SOM mapping. The principle of cross-validation is to remove a known observation and then re-estimate it with the remaining data. Here, we used the mean absolute error of prediction (MAE) and the mean squared error of prediction (MSE) as criteria for the comparison. MAE measures the average absolute error between the measured and predicted SOM values and is defined as: $MAE = \frac{1}{n}\sum_{\alpha=1}^{n} z^*(x_\alpha) - z(x_\alpha)$, where $z(x_\alpha)$ is the measured value of SOM, $z^*(x_\alpha)$ is the predicted value of SOM, and n is the number of samples. MSE measures the average square error between the measured and predicted SOM values and is defined as: $MSE = \frac{1}{n}\sum_{\alpha=1}^{n} (z^*(x_\alpha) - z(x_\alpha))^2$.

4. Results

4.1. Correlation between TWIs and SOM

The Pearson's correlation coefficients between the variables derived from the DEM and processed SOM measurements are listed in Table 2. Correlation coefficients for TWI_{MFD} and TWI_{D8} are the strongest and considerably higher than the others. Correlations between the TWIs and SOM were stronger than those between SCAs and SOM, which is likely due to the TWIs' better representation of soil moisture.

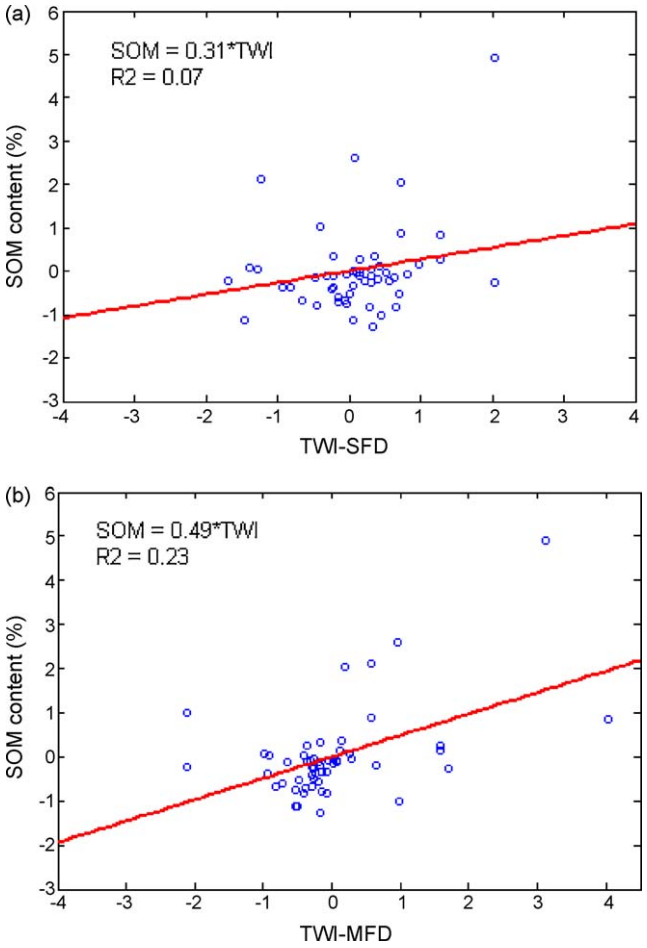

Fig. 3. Linear regression between processed SOM and TWIs. (a) TWI_{D8} and (b) TWI_{MFD} (both linear regression functions are tested to be significant at the 0.05 level).

TWI_{MFD} showed a stronger correlation with SOM than TWI_{D8}. This was also validated by the linear regression analysis between

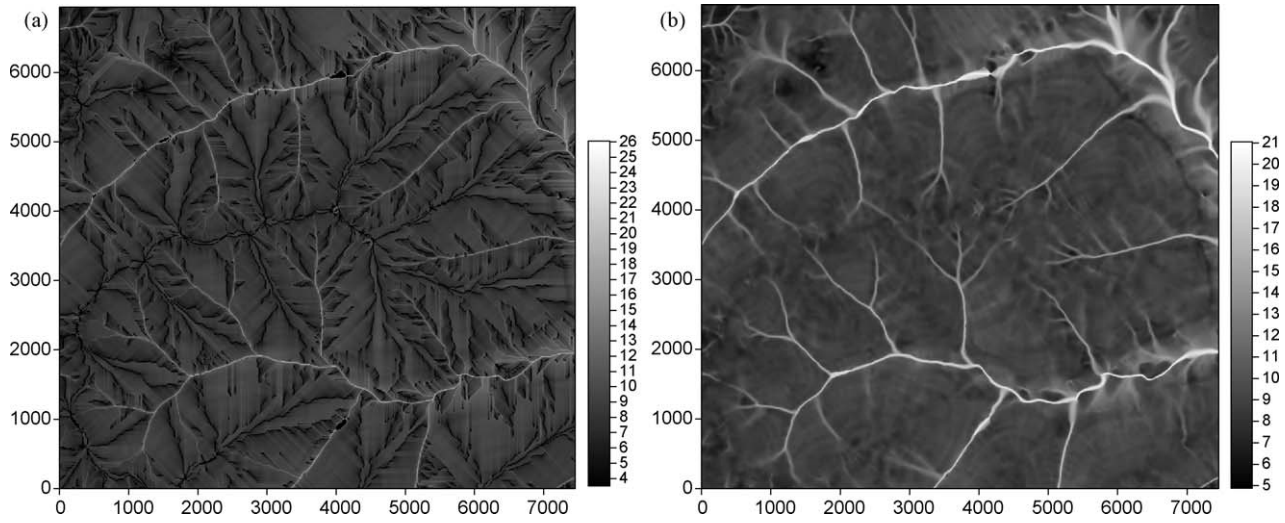

Fig. 4. Distributions of TWI_{D8} and TWI_{MFD}. (a) TWI_{D8} and (b) TWI_{MFD} (unit of x- and y-axis: meter).

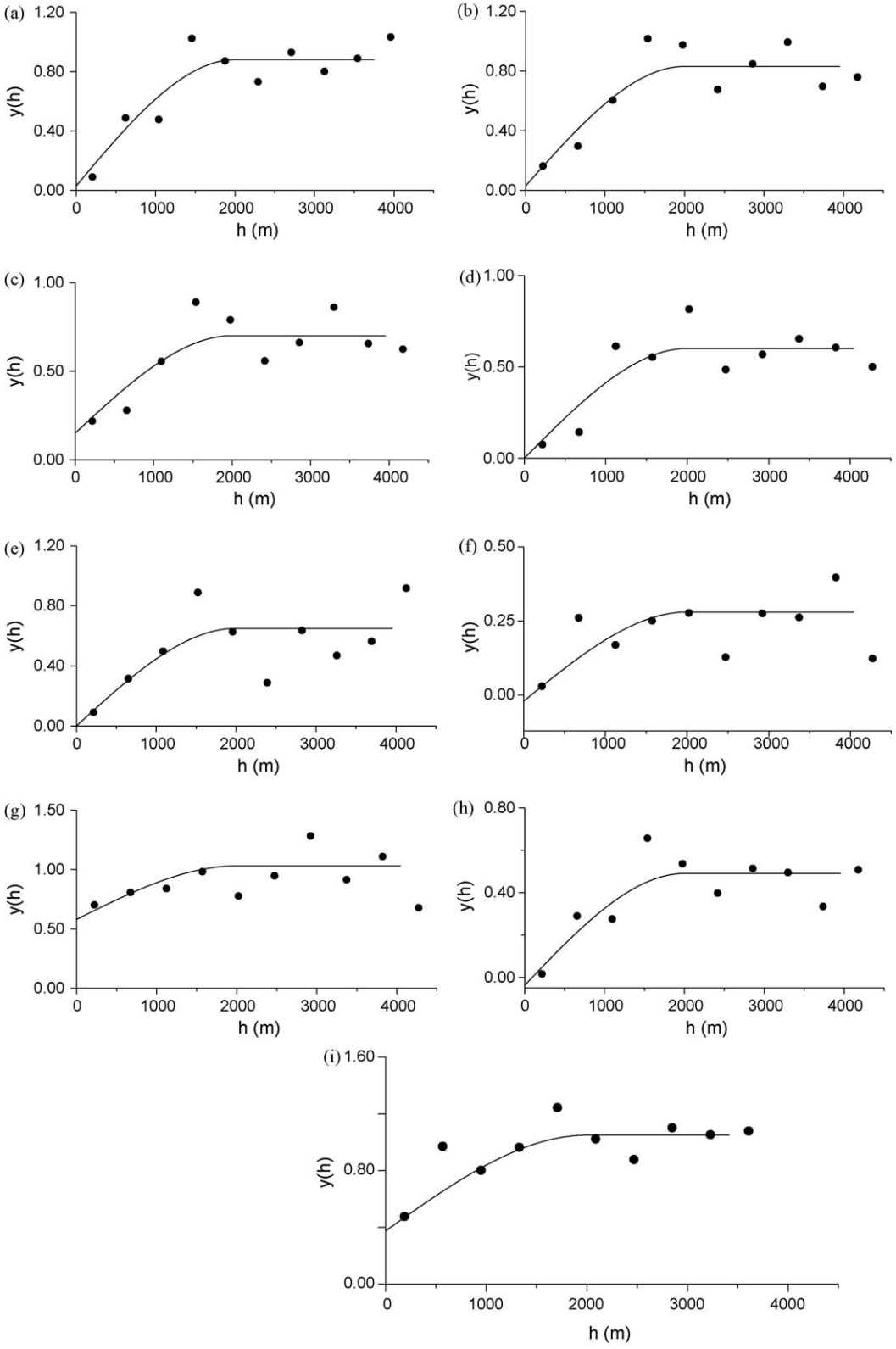

Fig. 5. Experimental (dots) and modeled (lines) semi-variograms and cross semi-variograms. (a) SOM, (b) regression residue of SOM (TWI_{D8}), (c) regression residue of SOM (TWI_{MFD}), (d) residue of SOM (TWI_{D8}), (e) residue of SOM (TWI_{MFD}), (f) SOM-TWI_{D8}, (g) TWI_{D8}, (h) SOM-TWI_{MFD}, and (i) TWI_{MFD}.

SOM and the TWIs (Fig. 3), in which the R^2 and the regression coefficient of the linear regression function between SOM and TWI_{MFD} (0.23 and 0.49) were larger than those between SOM and TWI_{D8} (0.07 and 0.31). This might reflect that water flow is more divergent in the low-relief areas and TWI_{MFD} may better represent the resulting soil moisture distribution than TWI_{D8}. To compare their performance in different SOM mapping strategies, we designated TWI_{D8} and TWI_{MFD} (their distributions are shown in Fig. 4) as exhaustive secondary variables and employed OK, SKlm, KED and CC for spatial prediction.

Table 3
Parameters of semi-variogram and cross semi-variogram models.

Semi-variogram	C_0	Range (m)	C
$\gamma_{SOM}(h)$	0.03	2000	0.85
$\gamma_R^{SK}(h)$ (with TWI_{D8})	0.03	2000	0.80
$\gamma_R^{SK}(h)$ (with TWI_{MFD})	0.15	2000	0.55
$\gamma_R^{KED}(h)$ (with TWI_{D8})	0	2000	0.60
$\gamma_R^{KED}(h)$ (with TWI_{MFD})	0	2000	0.60
$\gamma_{SOM\text{-}TWI_{D8}}(h)$	−0.02	2000	0.30
$\gamma_{SOM\text{-}TWI_{MFD}}(h)$	−0.04	2000	0.53
$\gamma_{TWI_{D8}}(h)$	0.58	2000	0.45
$\gamma_{TWI_{MFD}}(h)$	0.38	2000	0.67

4.2. (Co)regionalization

4.2.1. Regionalization of OK

In Eq. (7), $\gamma(x_\alpha - x_\beta)$ and $\gamma(x_\alpha - x)$ are the coefficients and can be computed through regionalization modeling. The experimental SOM semi-variogram is shown in Fig. 5a. Due to the limited number of samples only the omnidirectional semi-variogram was computed. In previous SOM mapping studies, the spherical, Gaussian and exponential have been the most widely used models in fitting semi-variograms. Here, we use the spherical model to fit the experimental SOM semi-variogram. The regionalization SOM modeling was interactively fitted with a visual fitting tool (GSTAT 2.0)[3] (the parameters are listed in Table 3). The (co)regionalization modeling hereafter was also implemented with GSTAT 2.0.

4.2.2. Regionalizations of SKlm-TWI_{D8} and SKlm-TWI_{MFD}

$\gamma_R^{SK}(h)$ is the residual semi-variogram, which is computed by subtracting the local mean $m_{SK}^*(x)$ from the primary datum $z(x_\alpha)$. The local means were obtained by using the linear regression function between the secondary variable and SOM. In our study, the linear regression function between SOM and TWI_{D8} was estimated to be $SOM = 0.31TWI_{D8}$, and that between SOM and TWI_{MFD} to be $SOM = 0.49TWI_{MFD}$ (Fig. 3). The SOM residual semi-variograms accounting for the secondary variables (TWI_{D8} and TWI_{MFD}) are shown in Fig. 5b and c, and the parameters of each model are presented in Table 3. Due to the subtraction of local means, the residual semi-variograms display smaller sill values than those of SOM, as expected.

4.2.3. Regionalizations of KED-TWI_{D8} and KED-TWI_{MFD}

$\gamma_R^{KED}(h)$ should be inferred from pairs of primary values that are unaffected or only slightly affected by its trend (Goovaerts, 1997). In our study pairs of locations, where the difference between the collocated secondary values is less than 0.5 m_2 (m_2 is the mean of secondary variable), were selected to estimate $\gamma_R^{KED}(h)$. Residual semi-variograms of SOM accounting for the secondary variables (TWI_{D8} and TWI_{MFD}) were computed and fitted (see Fig. 5d and e). Compared with the SOM semi-variogram (Fig. 5a), the residual semi-variograms display smaller nugget and sill values (Table 3).

4.2.4. Co-regionalization of CC-TWI_{D8} and CC-TWI_{MFD}

To solve the cokriging system (Eq. (13)), the co-regionalization should be modeled for the computation of $\gamma_{11}(\cdot)$, $\gamma_{12}(\cdot)$ and $\gamma_{122}(\cdot)$. The issue is to fit the direct semi-variograms of SOM and TWIs and the cross semi-variogram between SOM and TWIs under the constraint of the authorized co-regionalization model The co-regionalization model must satisfy the permissible condition, which ensures the matrix of auto- and cross covariance models $C(\cdot)$ are positive semi-definite. The practical way to build a permissible model is to model the direct semi-variograms and cross semi-variogram with a linear combination of a set of basic models, and to ensure all principal minor determinants of the coefficients matrixes of the basic models are non-negative (Goovaerts, 1997).

We first model the co-regionalization of SOM-TWI_{D8}. According to the semi-variograms and cross semi-variograms between SOM and TWI_{D8}, two basic structures, i.e. a pure nugget model and a spherical model, are adopted. The linear combination is

$$\begin{aligned}\gamma_{SOM}(h) &= 0.03g_0(h) + 0.85g_1(h)\\ \gamma_{SOM\text{-}TWI_{D8}}(h) &= -0.02g_0(h) + 0.30g_1(h)\\ \gamma_{TWI_{D8}}(h) &= 0.58g_0(h) + 0.45g_1(h)\end{aligned} \quad (14)$$

where $g_0(h)$ is the pure nugget model, and $g_1(h)$ is the spherical model with a range of 2000 m. The cross semi-variogram of SOM-TWI_{D8} and the semi-variogram of TWI_{D8} are displayed in Fig. 5f and g. Because $\begin{vmatrix} b_{11}^1 & b_{21}^1 \\ b_{12}^1 & b_{22}^1 \end{vmatrix} = \begin{vmatrix} 0.03 & -0.02 \\ -0.02 & 0.58 \end{vmatrix} = 0.017 > 0$ and $\begin{vmatrix} b_{11}^2 & b_{21}^2 \\ b_{12}^2 & b_{22}^2 \end{vmatrix} = \begin{vmatrix} 0.85 & 0.30 \\ 0.30 & 0.45 \end{vmatrix} = 0.2925 > 0$, the linear model of the co-regionalization is a permissible one.

Similarly, we used the same linear combination (including g_0(h) and g_1(h)) to model the co-regionalization of SOM-TWI_{MFD} (Eq. (15)). The model can be proved easily to be permissible.

$$\begin{aligned}\gamma_{SOM}(h) &= 0.03g_0(h) + 0.85g_1(h)\\ \gamma_{SOM\text{-}TWI_{MFD}}(h) &= -0.04g_0(h) + 0.53g_1(h)\\ \gamma_{TWI_{MFD}}(h) &= 0.38g_0(h) + 0.67g_1(h)\end{aligned} \quad (15)$$

The cross semi-variogram of SOM-TWI_{MFD} and the semi-variogram of TWI_{MFD} are shown in Fig. 5h and i, respectively.

4.3. Mapping results

After (co)regionalization modeling, we generated SOM maps with seven mapping approaches (i.e. OK, SKlm-TWI_{D8}, SKlm-TWI_{MFD}, KED-TWI_{D8}, KED-TWI_{MFD}, CC-TWI_{D8} and CC-TWI_{MFD}) with GSLIB. The mapping results are displayed in Fig. 6. All maps show the same two high value areas in the east and northwest. Apart from this, the maps differ significantly. The OK map (Fig. 6a) shows a smooth SOM surface, whereas the SKlm, KED and CC maps are influenced significantly by the secondary variables and reveal more detail (Fig. 6b–g). These can be validated by consulting the TWI distributions (Fig. 4). The impact of TWI_{D8} is obvious in maps generated by SKlm-TWI_{D8}, KED-TWI_{D8} and CC-TWI_{D8} (Fig. 6b, d and f), while the impact of TWI_{MFD} is more pronounced in Fig. 6c, e and g.

The statistics of the estimation results show that negative values (for example, −19.31) and extremely large values (for example, 103.65) appear in the KED results (both with TWI_{D8} and TWI_{MFD}). This is because the TWIs (both TWI_{D8} and TWI_{MFD}) are raggedly regionalized variables distributed over the study area, which can be explained by their nugget values (0.38 for TWI_{MFD} and 0.58 for TWI_{D8}, which are much larger than the other values, see Table 3). The magnitude of this effect varies with the neighborhood chosen for KED (the searching radii for the primary and the secondary variable in KED were set to 2000 m as the correlation distances (i.e. range) are 2000 m in all (co)regionalization models); larger neighborhoods might be smoother. The coarse distributions make the system of KED unstable and thereby produce extreme small (negative) and large values. The black-white mingled areas in the east and northwest part of Fig. 6d and e are caused by the non-smooth predictions.

[3] GSTAT 2.0 was developed by the first author (interested readers may ask the first author for free use).

Table 4
Cross-validation results.

	OK	Sklm (TWI_{D8})	SKlm (TWI_{MFD})	KED (TWI_{D8})	KED (TWI_{MFD})	CC (TWID8)	CC (TWI_{MFD})
MAE (%)	1.82	2.17	1.90	2.01	1.91	1.73	1.64
MSE ($\times 10^{-4}$)	19.08	22.07	18.74	19.93	20.74	17.73	16.90

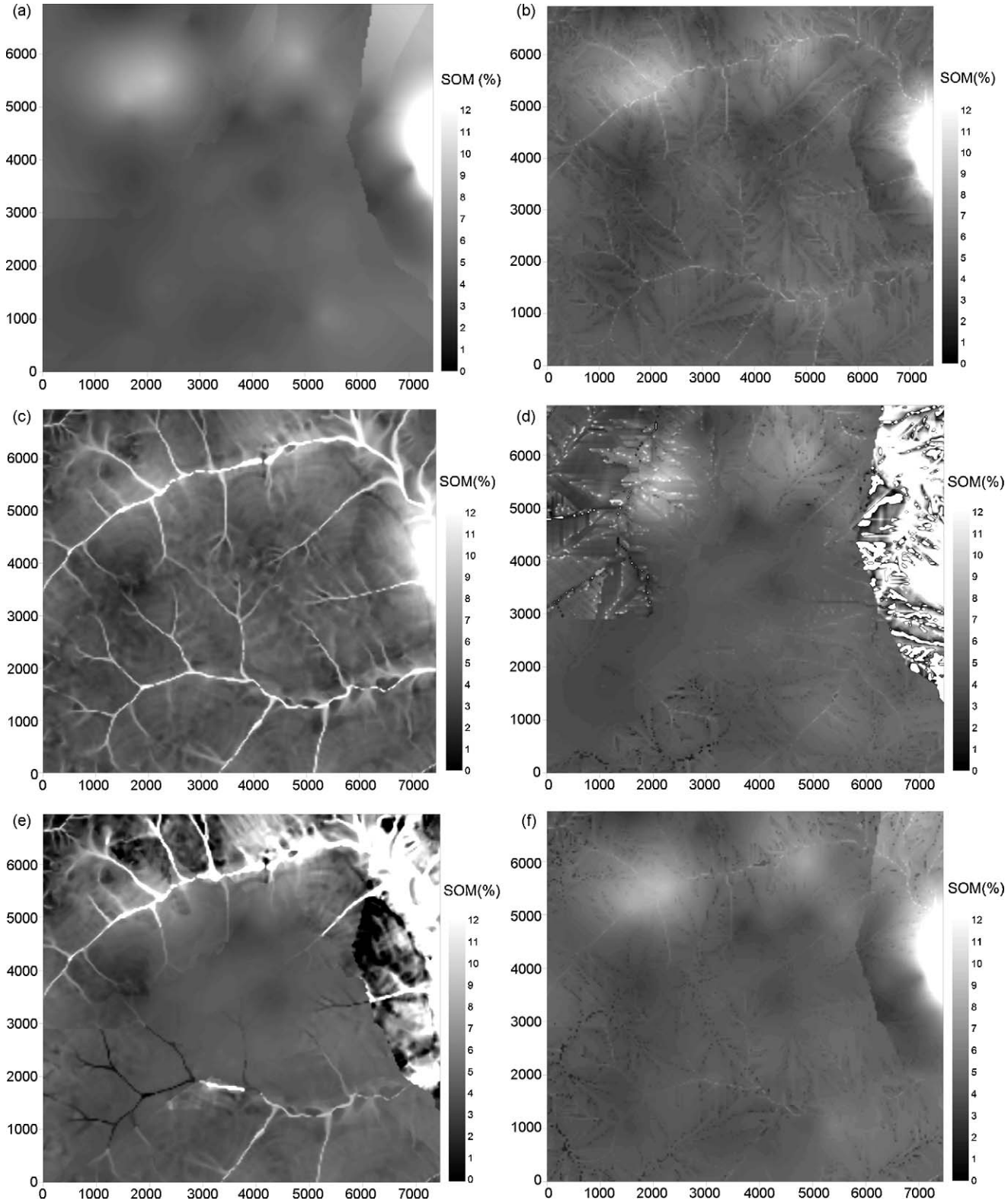

Fig. 6. SOM maps generated by different combinations of geostatistical methods and TWIs. (a) OK, (b) SKlm-TWI_{D8}, (c) SKlm-TWI_{MFD}, (d) KED-TWI_{D8}, (e) KED-TWI_{MFD}, (f) CC-TWI_{D8}, and (g) CC-TWI_{MFD} (unit of *x*- and *y*-axis: meter).

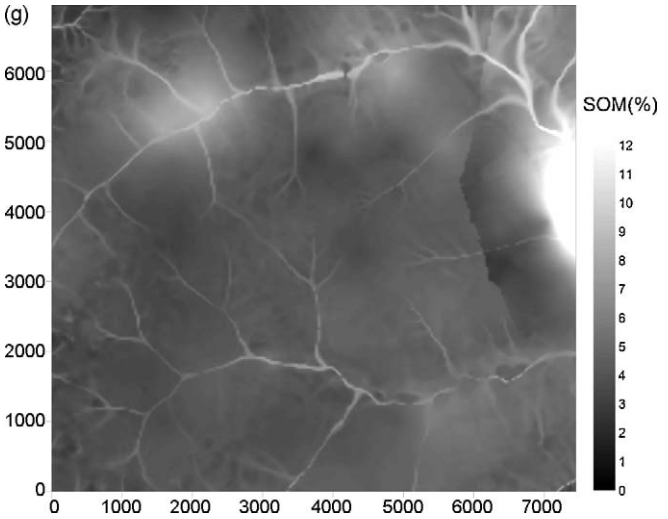

Fig. 6. (*Continued*).

5. Discussion

We first analyzed the correlation coefficients and the parameters of the (cross) semi-variograms between SOM and TWI_{MFD} to see if they reflected the ecosystem status of the study area. The small nugget value of $\gamma_{SOM}(h)$ indicates that SOM is distributed smoothly over the research area. However, its sill value, standard variance and histogram show that the variation of SOM at a large scale is significant (remembering the maximum value is over 10 times larger than the minimum). The large variation of SOM may result from soil erosion that is caused by a combination of wind, water and tillage (Yang, 2007; Wang et al., 2009). As stated in Section 3, in summer and spring, overland flow dominates water redistribution processes, and may play an important role in the soil erosion (Yang, 2007; Wang et al., 2009). The correlation coefficient between SOM and TWI_{MFD} indicates that SOM distribution is influenced by the overland flow, since TWI_{MFD} may better reflect the strength of the overland flow. In addition, the large range (2000 m) and the high sill value (0.53) of the cross semi-variogram between SOM and TWI_{MFD} also imply the influence of the overland flow on SOM redistribution. We then studied the estimation results generated using different combinations of kriging methods and TWIs. The results of the cross-validation are listed in Table 4. We first compared the impacts of TWI_{MFD} and TWI_{D8} on SOM mapping. Results of the cross-validation (Table 4) show that TWI_{MFD} outperforms TWI_{D8} for SKlm and CC, while being similar to TWI_{D8} for KED (the cross-validation KED-TWI_{MFD} produced less MAE but larger MSE). This can be explained by the difference in correlations to SOM between TWI_{MFD} and TWI_{D8}.

Regarding kriging methods, we found that results generated by SKlm and KED are no better than those by OK. OK performed better than SKlm-TWI_{D8}, SKlm-TWI_{MFD}, KED-TWI_{D8} and KED-TWI_{MFD}. This also shows that SKlm and KED may not help to reduce significantly the uncertainty of SOM mapping when secondary variables do not have strong enough correlations with the primary variable. This was supported by the correlation coefficients between SOM and TWIs and the statistics of linear regression (R^2 and the constant parameters). A possible reason is that the estimation of local mean in KED and SKlm is influenced largely by the strength of the correlation between the primary and secondary variable (Goovaerts, 1997), and the moderate correlation in this case will lead to the low accuracy of the estimation of local mean, and ultimately distort the SOM estimation. Interestingly, no matter what the secondary variable is, results generated by CC are better than those by OK, SKlm and KED (Table 4). We conclude that CC may enhance the estimation by using the supplemented information at unsampled locations even if the cross-correlation between primary and secondary variable is moderate. Among all combinations between secondary variables and kriging methods, CC combined with TWI_{MFD} produced the least MAE and MSE.

6. Conclusion

In our study, TWI_{MFD} shows the strongest correlation with SOM among the terrain variables, including the SFD-based TWI. The validation results showed that the best combination of geostatistical method and the secondary variable for SOM mapping is CC and TWI_{MFD}, which produced the lowest error of all combinations. The high performance of CC-TWI_{MFD} can be explained by two factors. First, the TWI_{MFD} is a synthesized index that shows superior capability of representing quantitatively the spatial distribution of soil moisture, and therefore displays a relatively high correlation with SOM. Second, CC is capable of utilizing effectively the spatial auto-correlation and the cross-correlation between SOM and secondary exhausted variable for the SOM mapping, even if the variable is not distributed smoothly over the research area. The performance of CC-TWI_{MFD} indicates that the combination not only reduces uncertainties but also reveals more details of SOM distribution. TWI_{MFD} is not compatible with KED because TWI_{MFD} is not distributed smoothly in the study area, which may weaken the stability of kriging systems.

Due to enhanced resolution and precision in the spatial modeling of landscapes afforded by DEMs, and the different terrain variables generated from them, these terrain models have become an important information source for ecological studies. The significance of our study is to show the potential predictive power of a MFD-based TWI for mapping SOM compared to the SDF-based TWI. Our finding is applicable to various resource mapping applications, particularly for precision agriculture and monitoring the spatial distribution of nutrients and moisture conditions. The limitations of our study are twofold: (1) the comparison between the SFD-based TWI and the MFD-based TWI was implemented in a low-relief area only; (2) the sensitivity of the MFD parameters (for example, the parameter of p in the calculation of TWI_{MFD}) to the precision of SOM mapping needs to be studied further.

Acknowledgements

This study was funded through supports from grants (Project No.: 40601078 and 40501056) from National Natural Science Foundation of China, a grant of Open Research Funding Program of LGISEM0802, the National Basic Research Program of China (2007CB407207), the National Key Technology R&D Program of China (2007BAC15B01), the support by the Vilas Trustees via its Vilas Associate Program to the University of Wisconsin-Madison, and the Hamel Faculty Fellow Award from the University of Wisconsin-Madison.

References

Anselin, L., 1995. Local indicators of spatial association-LISA. Geographical Analysis 27, 93–115.

Bell, J.C., Butler, C.A., Thompson, J.A., 1995. Soil terrain modeling for site-specific agricultural management. In: Robert, P.C., Rust, R.H., Larson, W.E. (Eds.), Site-specific Management for Agricultural Systems. American Society of Agronomy, Madison, WI, pp. 209–228.

Bell, J.C., Grigal, D.F., Bates, P.C., 2000. A soil-terrain model for estimation spatial patterns of soil organic carbon. In: Wilson, J.P., Gallant, J.C. (Eds.), Terrain Analysis: Principles and Applications. John Wiley & Sons, Inc., NY, USA, pp. 295–309.

Bertolo, F., 2000. Catchment Delineation and Characterization: A Review (Tech. Rep. EUR 19563 EN) Joint Res. Cent. Eur. Commun., Ispra, Italy. (April). Available online at:.http://agrienv.jrc.it/publications/pdfs/CatchRev.pdf (accessed February 01, 2008).

Beven, K.J., Kirkby, M.J., 1979. A physically based variable contributing area model of basin hydrology. Hydrology Science Bulletin 24, 43–69.
Fairfield, J., Leymarie, P., 1991. Drainage networks from grid digital elevation models. Water Resources Research 27, 709–717.
Freeman, T.G., 1991. Calculating catchment area with divergent flow based on a regular grid. Computers & Geosciences 17, 413–422.
Getis, A., Ord, J.K., 1996. Local spatial statistics: an overview. In: Longley, P., Batty, M. (Eds.), Spatial Analysis: Modelling in a GIS Environment. GeoInformation International, Cambridge, England, pp. 261–277.
Goovaerts, P., 1997. Applied Geostatistics for Natural Resources Evaluation. Oxford University Press, New York, USA, 483 pp.
Goovaerts, P., 2000. Geostatistical approaches for incorporating elevation into the spatial interpolation of rainfall. Journal of Hydrology 228, 113–129.
Hengl, T., Heuvelink, G.B.M., Rossiter, D.G., 2007. About regression-kriging: from equations to case studies. Computers & Geosciences 33, 1301–1315.
Janzen, H.H., Ellert, B.H., Anderson, D.W., 2002. Organic matter in the landscape. In: Lal, R. (Ed.), Encyclopedia of Soil Science. Marcel Dekker, New York, USA. , Online [available]:.http://www.dekker.com/sdek/abstract~content=a713620531~db=enc (July 21, 2008).
Jenny, H., 1941. Factors of Soil Formation-a System of Quantitative Pedology. McGraw-Hill, New York, USA, 281 pp.
Journel, A.G., Huijbregts, C.J., 1978. Mining Geostatistics. Academic Press, New York, 600 pp.
Knotters, M., Brus, D.J., Voshaar, J.H.O., 1995. A comparison of kriging, Co-kriging and kriging combined with regression for spatial interpolation of horizon depth with censored observations. Geoderma 67, 227–246.
Luca, C., Si, B.C., Farrell, R.E., 2007. Upslope length improves spatial estimation of soil organic carbon content. Canadian Journal of Soil Science 87, 291–300.
Moore, I.D., Gessler, P.E., Nielsen, G.A., Peterson, G.A., 1993. Soil attribute prediction using terrain analysis. Soil Science Society of America Journal 57, 443–452.
Mueller, T.G., Pierce, F.J., 2003. Soil carbon maps: enhancing spatial estimates with simple terrain attributes at multiple scales. Soil Science Society of America Journal 67, 258–267.
O'Callaghan, J.F., Mark, D.M., 1984. The extraction of drainage networks from digital elevation data. Computer Vision, Graphics and Image Processing 28, 323–344.
Qin, C.Z., Zhu, A.-X., Pei, T., Li, B.L., Zhou, C.H., Yang, L., 2007. An adaptive approach to selecting a flow-partition exponent for a multiple-flow-direction algorithm. International Journal of Geographic Information Sciences 21, 443–458.
Quinn, P., Beven, K., Chevalier, P., Planchon, O., 1991. The prediction of hillslope flow paths for distributed hydrological modeling using digital terrain models. Hydrological Processes 5, 59–79.
Quinn, P., Beven, K.J., Lamb, R., 1995. The $\ln(a/\tan\beta)$ index: how to calculate it and how to use it within the TOPMODEL framework. Hydrological Process 9, 161–182.
Raspa, G., Tucci, M., Bruno, R., 1997. Reconstruction of rainfall fields by combining ground raingauges data with radar maps using external drift method. In: Baafi, E.Y., Schofield, N.A. (Eds.), Geostatistics Wollongong'96. Kluwer Academic, Dordrecht, pp. 941–950.
Rodgers, J.L., Nicewander, W.A., 1988. Thirteen ways to look at the correlation coefficient. The American Statistician 42, 59–66.
Shary, P.A., Sharaya, L.S., Mitusov, A.V., 2002. Fundamental quantitative methods of land surface analysis. Geoderma 107, 1–32.
Starr, G.C., Lal, R., Malone, R., Hothem, D., Owens, L., Kimble, J., 2000. Modeling soil carbon transported by water erosion processes. Land Degradation & Development 11, 83–91.
Sumfleth, K., Duttmann, R., 2008. Prediction of soil property distribution in paddy soil landscapes using terrain data and satellite information as indicators. Ecological Indicators 8, 485–501.
Vauclin, M., Vieira, S.R., Vachaud, G., Nielsen, D.R., 1983. The use of cokriging with limited field soil observations. Soil Science Society of America Journal 47, 175–184.
Wang, Z.Q., Liu, B.Y., Wang, X.Y., Gao, X.F., Liu, G., 2009. Erosion effect on the productivity of black soil in northeast China. Science China Series D-Earth Science 52, 1005–1021.
Western, A.W., Grayson, R.B., Blöschl, G., Willgoose, G.R., McMahon, T.A., 1999. Observed spatial organization of soil moisture and its relation to terrain indices. Water Resource Research 35, 797–810.
Wolock, D.M., Mccabe, G.J., 1995. Comparison of single and multiple flow direction algorithms for computing topographic parameters. Water Resources Research 31, 1315–1324.
Yang, L., 2007. Extraction of knowledge on soil-environment relationships using fuzzy *c*-means (fcm) clustering. Master Thesis. Beijing Normal University, Beijing.
Zevenbergen, L.W., Thorne, C.R., 1987. Quantitative analysis of land surface topology. Earth Surface Processes Landforms 12, 47–56.
Zhang, Y.G., Wu, Y.Q., Lin, B.Y., Zheng, Q.H., Yin, J.Y., 2007. Characteristics and factors controlling the development of ephemeral gullies in cultivated catchments of black soil region, Northeast China. Soil & Tillage Research 96, 28–41.