



Contents lists available at ScienceDirect

Geoderma

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

Mapping soil organic matter in small low-relief catchments using fuzzy slope position information

Qin Cheng-Zhi^{a,*}, Zhu A-Xing^{a,b}, Qiu Wei-Li^c, Lu Yan-Jun^{c,a}, Li Bao-Lin^a, Pei Tao^a

^a State Key Laboratory of Resources and Environmental Information System, Institute of Geographic Sciences and Natural Resources Research, Chinese Academy of Sciences, 11A Datun Road, Anwai, Beijing 100101, PR China

^b Department of Geography, University of Wisconsin Madison, 550N, Park Street, Madison, WI 53706-1491, USA

^c School of Geography, Beijing Normal University, Beijing 100875, China

ARTICLE INFO

Article history:

Received 15 April 2010

Received in revised form 7 June 2011

Accepted 14 June 2011

Available online xxxx

Keywords:

Fuzzy slope position

Digital soil mapping

Soil organic matter

DEM

Directed sampling

ABSTRACT

Spatial transitions between slope positions (landform positions) are often gradual. Various methods have been developed to quantify the transitions using fuzzy slope positions. However, few studies have used the quantitative information on fuzzy slope positions in digital soil mapping or other terrain-related geographic modeling. This paper examines the use of such information for mapping soil organic matter content (SOM) within a purposive (or directed) sampling framework for predictive soil mapping. First, a five slope position system (i.e., ridge, shoulder slope, back slope, foot slope, channel) was adopted and the fuzzy slope positions were derived through an approach based on typical slope position locations. The typical slope position locations were extracted using a set of rules based on terrain attributes and domain knowledge. Secondly, the fuzzy slope positions were used to direct purposive sampling, which determined the typical SOM value for each slope position type. Typical SOM values were then combined with fuzzy slope position data to map the spatial variation of SOM using a weighted-average model – the fuzzy slope position weighted (FSPW) model – to predict the spatial distribution of SOM for two soil layers at depths of 10–15 cm and 35–40 cm in a low-relief watershed in north-eastern China. The study area comprised two portions: an area of about 4 km² used for model development, and an area of about 60 km² for model extrapolation and validation. Evaluation results show that our FSPW model produces a better prediction of the SOM than that provided by a multiple linear regression (MLR) model. Quantitative measures in the model-development area, including correlation coefficient, mean absolute error, and root mean square of error, show that the performance of the FSPW model with five modeling points from purposive sampling compares favorably with MLR results for 48 modeling points. Evidence from the quantitative assessment based on a validation set of 102 sample points in the model-extrapolation area shows that the FSPW model performs better than the MLR model, which suggests that information on fuzzy slope position was useful in aiding digital soil mapping over the area.

© 2011 Elsevier B.V. All rights reserved.

1. Introduction

Purposive sampling (or directed sampling) is a potentially effective way of mapping soil spatial variation, particularly for areas where existing soil data is limited (e.g., soil maps, field knowledge and field samples) (Pocknee et al., 1996; Zhu et al., 2008, 2010a). Purposive sampling assumes that the soil spatial variation of an area can be captured by the property values at some typical (key) locations (Zhu et al., 2010a). The soil property value at a location can be

predicted to be the average of the typical soil property values at these typical locations weighted by the similarity of the location to the typical locations (e.g., Zhu et al., 2010a). Under this purposive sampling assumption relatively few field samples need to be collected for model development, whereas many field samples are often required for statistical approaches (e.g., Gessler et al., 1995; Moore et al., 1993), or for geostatistical approaches (e.g., Lark, 2000; see McBratney et al., 2000 for overview). Zhu et al. (2008) and (2010b) showed that the purposive sampling approach to predicting detailed spatial variation of soil properties has obvious advantages from the perspectives of data requirement, model simplicity and accuracy of prediction.

The design of purposive sampling is generally related to a process of identifying environmental combinations (Zhu et al., 2008,

* Corresponding author. Tel.: +86 10 64889777; fax: +86 10 64889630.
E-mail address: qincz@reis.ac.cn (C.-Z. Qin).

2010a). This is based on the classical concept of soil–environment relationships (Hudson, 1992; Jenny, 1941). Unique soil properties can be associated with unique combinations (or configurations) of environmental factors such as terrain, climate or parent material. It is also assumed that changes between environmental combinations are gradual and co-vary with soil (Zhu et al., 2010a). Results from this identification process provide not only the number of environmental combinations (the number of field samples needed) but also the spatial locations of these samples (Zhu et al., 2010a).

Among the environmental factors, terrain is perhaps the one most frequently used for identifying environmental combinations. Originating from Jenny (1941)'s state-factor model of soil, the SCORPAN model was developed to formulate the notion that soil is a function of environmental factors (McBratney et al., 2003):

$$S = f(s, c, o, r, p, a, n) \quad (1)$$

where S is soil class or property at a point; s is some other, or previously measured, soil information at a point; c is climate factor; o is organism factor (vegetation, fauna or human activities); r is terrain factor; p is parent material factor; a is time factor; and n relates to spatial position.

Almost all previous studies on digital soil mapping have used terrain factor as an important, or even the sole predictive factor (see McBratney et al., 2003 for overview). Quantification of the terrain factor in current predictive soil mapping is often based solely on topographic attributes – elevation, slope gradient, profile curvature, contour curvature, topographic wetness index, etc. (Bell et al., 2000; Gessler et al., 1995; McBratney et al., 2003; Moore et al., 1993; Zhu et al., 2010a).

Spatial gradation of slope positions (or landform positions) plays an important role when identifying environmental combinations with terrain factor. The transitions between slope positions over space (e.g. from back slope to foot slope) are often gradual. These transitions, or spatial gradations, capture the transition of earth surface processes in space. Although there is a relationship between slope positions and topographic attributes, the spatial gradation of slope positions cannot be fully captured by topographic attributes such as slope gradient, curvature, etc. alone (Qin et al., 2009). This is because a slope position is a geographical object with a fuzzy boundary, and conveys qualitative and spatial contextual information as well as local geometric information. Locations on different slope positions might have the same topographic attributes, yet be associated with different geographical processes.

The quantification of the spatial gradation of slope positions (referred to as fuzzy slope position hereafter) is assumed to be important in digital soil mapping and other terrain-related geographical or ecological modeling at fine scale (MacMillan et al., 2000; Schmidt and Hewitt, 2004). Many researchers have discussed the relationship between the fuzzy slope positions and the transition of geographical processes, anticipating their usefulness in digital soil mapping (MacMillan et al., 2000; Qin et al., 2009; Ventura and Irvin, 2000). However, currently the fuzzy slope positions are often first converted to 'crisp' slope positions before being used in predictive soil mapping, if used at all. In this way, the quantitative information on spatial gradation of slope positions is lost.

Using purposive sampling as an example in this paper, we show the usefulness of information on spatial gradation of slope positions in digital soil mapping. We illustrate this through the prediction of the spatial distribution of soil organic matter (SOM) in small, low-relief catchments at a finer scale. The research issue is whether or not the information on spatial gradation of slope positions can benefit the

purposive sampling approach in predicting soil mapping at a finer scale.

2. Basic idea

The basic idea of this research is that slope positions reflect the integrative effect of earth surface processes, and that the spatial gradation of slope positions can be used to capture the transitional nature of such processes over a slope. Information on the spatial gradation of slope positions would then be comprehensive and indicative of spatial variation of soil properties, especially for small, low-relief areas.

Fuzzy slope positions can play two roles to support the development of a purposive sampling approach to predicting spatial distribution of soil properties in small, low-relief areas, given similar non-terrain environmental factors (e.g., geology, land use). First, slope positions can be treated as environmental combinations, so that areas with high membership values of slope positions can be targeted in order to obtain typical soil property values of environmental combinations. This is acceptable, since the slope positions comprehensively reflect the terrain conditions that assert significant effect on the spatial distribution of soil properties along a slope. In this way the required number of field samples might be minimized because the number of slope position types in a small, low-relief area is often very limited.

Second, fuzzy slope positions can be used to weight the typical soil property values in particular slope positions to predict the spatial distribution of the soil properties, by the weighted average model used in Zhu et al. (2010a). This means that soils associated with similar environmental combinations have similar properties, a position that is supported by soil–landscape model theory, in particular as advocated in the SCORPAN model (McBratney et al., 2003). Zhu et al. (2010b) have shown that the weighted-average model works well over areas where the soil–environment relationship is nonlinear.

3. Method

3.1. Quantification of spatial gradation of slope positions

With the aim of minimizing the number of field samples required in a soil survey, we have adopted a slope position classification system consisting of only five slope positions: ridge, shoulder slope, back slope, foot slope, and channel. This system forms a sequence of segments from top to bottom of a slope. If necessary, the system can be further subdivided to include the convexity and concavity of the surface shape along the contour (Qin et al., 2009).

Information on fuzzy slope positions is derived through an approach based on typical locations of slope positions (Qin et al., 2009). This approach is in two parts: the first is to extract prototypes for each slope position in the study area by formulating a set of rules based on terrain attributes and domain knowledge. It should be noted that the number of prototypes for a given slope position type is often too large to be treated as locations for field sampling. The second is to compute the similarity of a given location to the prototypes for slope positions. The detailed steps for deriving the fuzzy slope positions can be found in Qin et al. (2009).

3.2. Purposive sampling based on fuzzy slope positions

The derived fuzzy slope positions are used to determine the typical soil property value for each slope position type. In this step, samples are taken at locations with high fuzzy membership values (e.g., close to 1, given the range of membership is the interval [0, 1]) in one slope

position type, and with very low fuzzy membership value (e.g., close to 0) in other slope position types. With the aim of minimizing the number of field samples, we can then set a single field sampling point per slope position type, using above rule.

3.3. Estimating soil property values using fuzzy slope positions

In this step the weighted average model adopted from Zhu (1997) was used to predict the spatial distribution of soil properties using

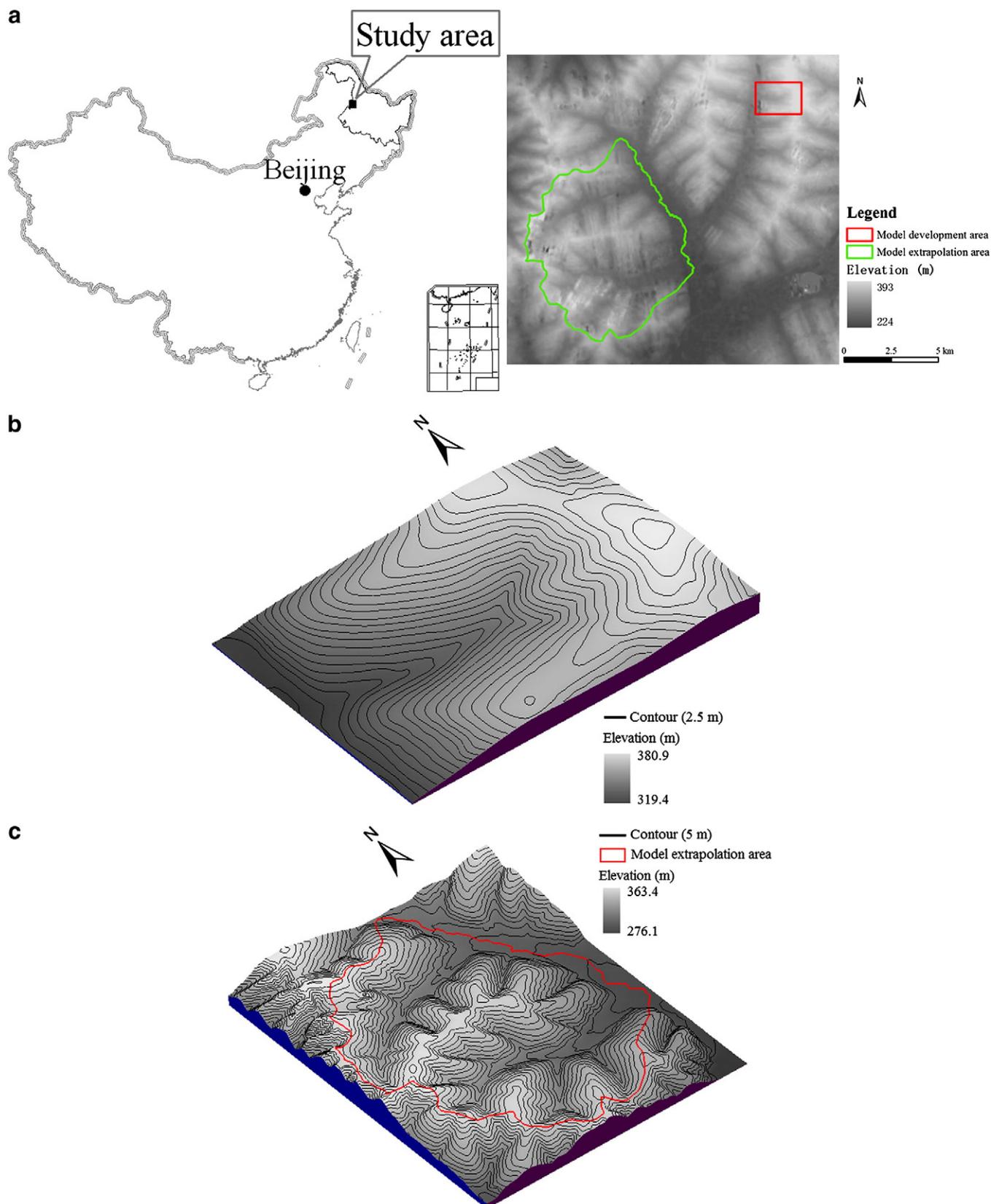


Fig. 1. Map of the study area. a) location of the study area; b) topography of the model-development area; c) topography of the model-extrapolation area.

fuzzy slope positions ('fuzzy slope position weighted', FSPW). The predicted value of a soil property at a given location is the weighted average of the typical soil property value at each slope position and the fuzzy membership values of the location in these slope positions, given by:

$$V_{ij} = \frac{\sum_{k=1}^n S_{ij}^k V^k}{\sum_{k=1}^n S_{ij}^k} \quad (2)$$

where V_{ij} is the predictive soil property value at location (i,j) ; S_{ij}^k is the membership value in the k -th type of slope position for location (i,j) ; V^k is the typical soil property value for the k -th type of slope position; and n is the number of slope position types.

4. Case study

4.1. Study area and data

4.1.1. Study area

The study area is located in the low-relief part of the Nenjiang watershed in north-eastern China (Fig. 1a). It consists of two portions (catchments) about 8 km apart, with similar environmental conditions: one for model development (Fig. 1b) and the other for model extrapolation (Fig. 1c). Soils in the study area are formed on deposits of silt loam loess. The parent material is the same over both areas. Land use at the time of the study was mainly soybean and wheat farming.

The model-development area is a small catchment of about 4 km². Its elevation difference is about 60 m and the average slope gradient is 2.4°. The model-extrapolation area is about 60 km² in size. The elevation difference is about 100 m and the average slope gradient is 2°

4.1.2. Digital elevation model (DEM)

The DEM was created from a 1:10,000 topographical map with the contour interval of 2.5 m. The contour lines were digitized from scanned paper maps. TopoGrid and TINLATTICE functions of ArcGIS software were combined to create the DEM on a 10 m universal transverse Mercator (UTM) grid. During the creation of the DEM, the approach proposed by Hengl et al. (2004) was applied to reducing errors in the DEM. Before the calculation of topographic attributes, a DEM pre-processing algorithm proposed by Planchon and Darboux (2001) was used to remove depressions and revise flat areas with a very gentle slope gradient (0.1% in this study), which is more suitable for low-relief areas (Qin et al., 2006).

4.1.3. Soil sampling

In the model-development area, 48 soil sampling field points were selected (Fig. 2), mainly along transects, and based on spatial transitions between slope positions. Four transects were designed across the area from east to west, the field sampling points being defined both by typical location and transitions between slope positions. The first transect started at the highest point on the ridge and extended to the start of the main channel. The second transect started at the ridge and stretched along the side ditch. The third transect extended from ridge to channel in a line perpendicular to the contours of the side slope in the southern part of the area. The fourth transect extended from the ridge to the lowest point of the main channel. Some sampling points were also located along the main channel and along a contour on the back slope. It is considered that these encompassed the major terrain variations in this area, including variations between slope positions.

Two soil samples were collected at each field point: one from a depth of 10–15 cm (referred to as the first layer) and one from a depth of 35–40 cm (referred to as the second layer). Each soil sample, weighing about 100 g, was sealed in a plastic bag for laboratory analysis. The Tyurin method was used to measure the SOM content as percentages of the samples (Bao, 2000). At one field point a second-layer SOM value was not obtained, because the material at the required depth comprised mainly gravel and coarse sand which could not be used for SOM measurement. Table 1 shows the statistics of the measured SOM values.

In the model-extrapolation area, 102 field points were sampled (Fig. 3) and the SOM content in the first (10–15 cm) layer was measured; the results were then used to evaluate the performance of the model when extrapolated to watersheds nearby. Three sampling strategies were adopted in selecting the field points (Fig. 3): regular sampling, subjective sampling, and transect sampling. A regular sampling grid 1100 m × 740 m was used to collect validation points for the overall performance of the model extrapolation. Further, we conducted subjective sampling in areas with unique characteristics and where the spatial variation of soil might not have been captured by regular sampling mostly on ridges, shoulder slopes and foot slopes (Fig. 3). We also used transect sampling to cover major environmental variations along the shortest distance from ridge to channel. Two such transects were sampled: one approximately E–W with sample spacing about 80 m, and one approximately N–S with sample spacing about 60 m (Fig. 3). With this combination of sampling strategies, major soil variation in the area was covered for evaluating the model extrapolation.

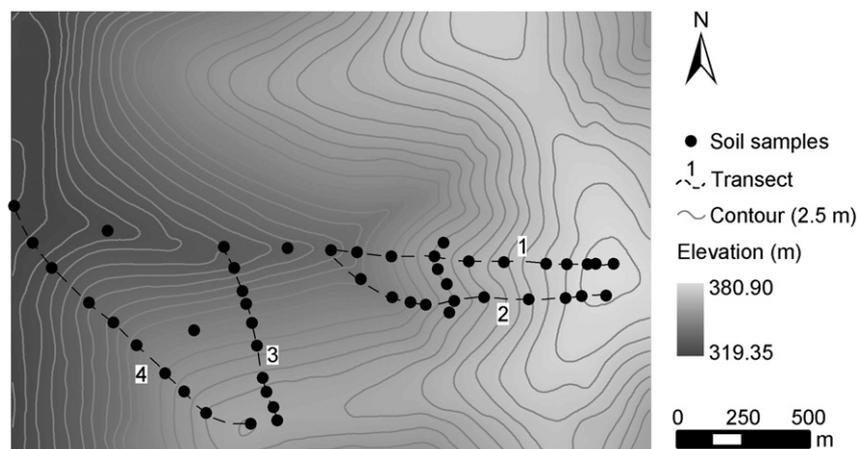


Fig. 2. Field points in the model-development area.

Table 1
Descriptive statistics of measured soil organic matter content.

| | Minimum (%) | Maximum (%) | Mean (%) | Std. Deviation (%) |
|--|-------------|-------------|----------|--------------------|
| Model-development area (first layer) (48 samples) | 0.836 | 6.900 | 3.749 | 1.502 |
| Model-development area (second layer) (47 samples) | 0.302 | 6.487 | 1.825 | 1.504 |
| Model-extrapolation area (first layer) (102 samples) | 2.245 | 9.180 | 4.336 | 1.189 |

4.2. Model implementation

4.2.1. Deriving fuzzy slope positions

As set out in Section 3, fuzzy slope positions were derived using the method of Qin et al. (2009). The rules for extracting prototypes of slope positions and deriving fuzzy slope positions in this study area were set by combining the slope position definitions of Pennock et al. (1987), MacMillan et al. (2000), and Schmidt and Hewitt (2004). The rules used for this study area (see Table 2) were specified for local topographic attributes (i.e., slope gradient, profile curvature) and a regional terrain index – the relative position index (RPI) proposed by Skidmore (1990), which approximately estimates how far a location is from ridge or valley. RPI is calculated from the Euclidean distance to the nearest valley divided by the sum of the Euclidean distances to the nearest valley and ridge.

Fig. 4 shows the similarity to the back slope type derived using these rules for the model development area. By comparing the shaded

Table 2

Parameter settings for extracting and deriving the fuzzy slope positions in the study area. Fuzzy membership function applied to deriving fuzzy slope position is one of three curve shape types obtained by adjusting the shape-controlling parameters (i.e. w_1 and w_2): 'S'-shaped, 'bell'-shaped, and 'Z'-shaped (see Qin et al., 2009). The profile curvature protocol assigns a negative value for concavity and a positive value for convexity.

| | RPI | | Profile curvature ($\times 10^{-3} \text{ m}^{-1}$) | | Slope ($^\circ$) | |
|----------------|-------------------|---------------------------|---|-------------------------|--------------------|-----------------|
| | Typical locations | Fuzzy inference | Typical locations | Fuzzy inference | Typical locations | Fuzzy inference |
| Ridge | ≥ 0.99 | 'S'; $w_1 = 0.05$ | [-0.5, 0.5] | 'bell'; $w_1 = w_2 = 1$ | | |
| Shoulder slope | [0.8, 0.9] | 'bell'; $w_1 = w_2 = 0.1$ | ≥ 0.5 | 'S'; $w_1 = 1$ | | |
| Back slope | [0.4, 0.6] | 'bell'; $w_1 = w_2 = 0.2$ | [-0.01, 0.01] | 'bell'; $w_1 = w_2 = 1$ | | |
| Foot slope | [0.1, 0.2] | 'bell'; $w_1 = w_2 = 0.1$ | ≤ -0.5 | 'Z'; $w_2 = 1$ | | |
| Channel | ≤ 0.01 | 'Z'; $w_2 = 0.1$ | [-0.5, 0.5] | 'bell'; $w_1 = w_2 = 1$ | ≤ 0.5 | 'Z'; $w_2 = 2$ |

relief map with this similarity map we can see that the derived fuzzy back slope captures the spatial gradation of the back slope well.

4.2.2. Purposive sampling in model-development area and SOM mapping

Individual field points were used to determine the typical SOM value for each type of slope position. Thus, there are a total of five modeling points for the FSPW model. The typical SOM value for each of the five slope position types was set to be the value of a sample at which the similarity to that slope position type was as high as possible and the similarity to other slope position types as low as possible. The

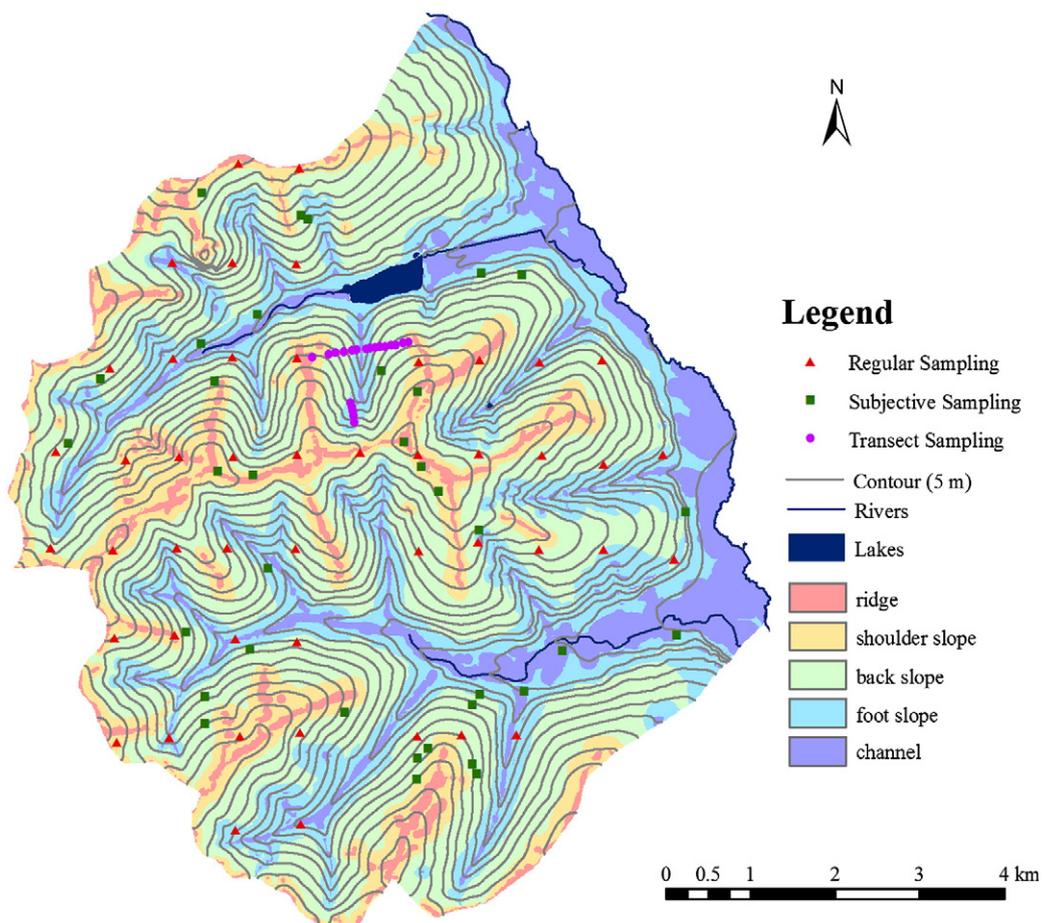


Fig. 3. Field points from different sampling strategies in the model-extrapolation area. (The base map is the hardened slope position map).

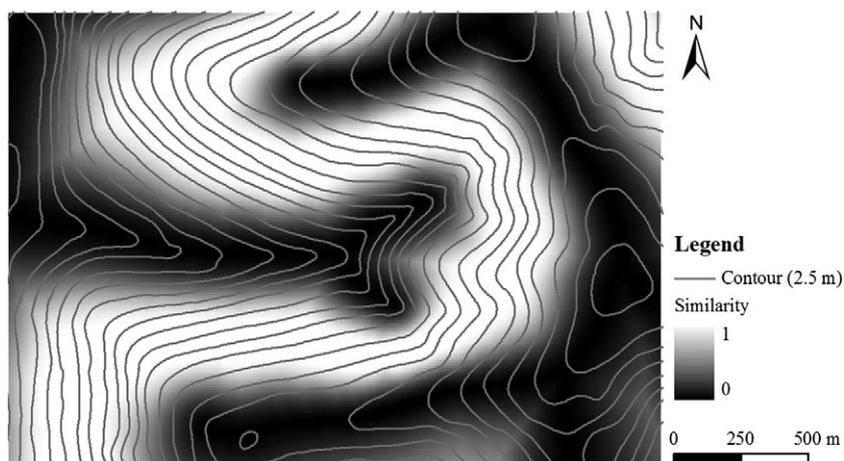


Fig. 4. Map of the similarity to the back slope type derived in the model-development area.

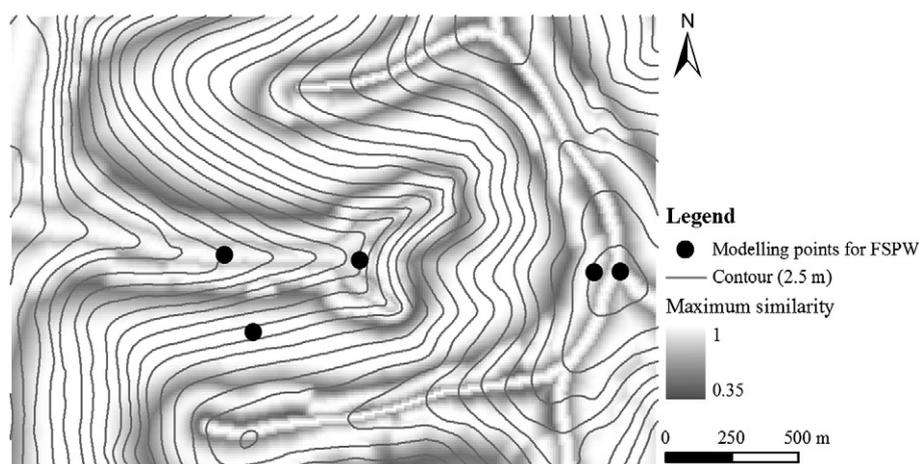


Fig. 5. Locations of the five modeling points for the FSPW model in the model-development area. (The base map shows the maximum similarity to slope positions).

similarity values for the final set of modeling points in their respective slope position types are larger than 0.97 for the model-development area. The spatial distribution of these five modeling points is shown in Fig. 5 and their corresponding SOM values are given in Table 3. The SOM map was derived by combining the fuzzy slope positions and the typical SOM value for each type of slope position calculated from Eq. (2).

4.3. Evaluation method

The prediction of SOM based on fuzzy slope positions was evaluated in both the model-development and model-extrapolation areas. The model-development area was used to assess the suitability of the FSPW model; the model-extrapolation area was used to

examine the portability of the developed model when extrapolated to nearby watersheds.

The performance of the FSPW model was assessed both qualitatively and quantitatively in each area. Qualitatively, the evaluation of the predicted results was based both on domain knowledge of soil science, and field knowledge. The quantitative assessment was based on a comparison between the predicted and measured values of soil properties at the evaluation points using correlation coefficient (CC), mean absolute error (MAE), and root mean square of error (RMSE), defined by:

$$CC = \frac{N \sum OP - \sum O \sum P}{\sqrt{N \sum O^2 - (\sum O)^2} \sqrt{N \sum P^2 - (\sum P)^2}} \quad (3)$$

Table 3
Similarity values and SOM measurement of the modeling points for the FSPW model.

| Sample ID | Similarity value | | | | | Slope position | SOM (%) | |
|-----------|------------------|----------------|------------|------------|---------|----------------|-------------|--------------|
| | Ridge | Shoulder slope | Back slope | Foot slope | Channel | | First layer | Second layer |
| HB8-01 | 1 | 0.43 | 0.03 | 0 | 0 | Ridge | 3.57 | 1.08 |
| HB8-03 | 0.14 | 0.97 | 0.11 | 0 | 0 | Shoulder slope | 3.50 | 1.47 |
| HB8-28 | 0 | 0 | 1 | 0 | 0 | Back slope | 3.18 | 0.98 |
| HB8-10 | 0 | 0 | 0.09 | 0.98 | 0.58 | Foot slope | 1.52 | 3.84 |
| HB8-32 | 0 | 0 | 0.03 | 0.37 | 0.98 | Channel | 6.77 | 2.81 |

Table 4

Correlation and partial correlation between SOM and topographic attributes in the model-development area.

| Topographic attribute | First-layer SOM | | Second-layer SOM | |
|---------------------------|-----------------|---------------------|------------------|---------------------|
| | Correlation | Partial correlation | Correlation | Partial correlation |
| Elevation | −0.246 | 0.102 | −0.304* | −0.051 |
| Slope gradient | −0.111 | 0.074 | 0.018 | 0.113 |
| Sin(Aspect) | −0.175 | −0.039 | −0.295* | −0.130 |
| Cos(Aspect) | −0.068 | 0.133 | −0.242 | −0.200 |
| Profile curvature | −0.212 | 0.165 | −0.383** | −0.031 |
| Horizontal curvature | −0.705** | −0.237 | −0.780** | −0.438** |
| Topographic wetness index | 0.755** | 0.433* | 0.723** | 0.240 |

* Correlation is significant at the 0.05 level (2-tailed).

** Correlation is significant at the 0.01 level (2-tailed).

$$MAE = \frac{\sum |P-O|}{N} \quad (4)$$

$$RMSE = \sqrt{\frac{\sum |P-O|^2}{N-1}} \quad (5)$$

where O and P are the observed and predicted values of a particular soil property, and N is the number of soil samples.

In both areas a multiple linear regression (MLR) model regressing SOM against a set of topographic variables was developed as a reference for assessing the FSPW method. For the development of MLR in this study we selected elevation, slope gradient, slope aspect,

profile curvature, horizontal curvature, and topographic wetness index as the predictive variables, being similar to those chosen in other studies (e.g. Bell et al., 2000; Gessler et al., 1995). According to the SCORPAN framework (McBratney et al., 2003), additional variables can be used in the production of a predictive digital soil map, but these (e.g., land cover, climate and parent material) are almost identical over the small, low-relief areas in this study. We believe that it was appropriate in the present work to use only the terrain variables for predictive soil mapping. Compared with other topographic attributes, the slope aspect is less used in predictive soil mapping. However King et al. (1999) showed that there might be a strong relationship between soil property and slope aspect, even in low-relief, gently sloping areas. We retained the slope aspect as one of the predictors in this study. The sine and cosine of the aspect were computed for developing MLR model. Slope gradient, aspect and curvatures were computed by widely-used algorithms (see Shary et al., 2002). The topographic wetness index was computed by a multiple-flow-direction-based algorithm which can be adapted to local terrain conditions (Qin et al., 2011). Table 4 shows the correlation and partial correlation between SOM and topographic attributes in the model-development area. All 48 field points were used to develop the MLR model by the stepwise variable selection method with default criteria in SPSS 11.0 (i.e., Probability-of-F-to-enter ≤ 0.050 and Probability-of-F-to-remove ≥ 0.100). For the first-layer SOM, only the topographic wetness index was selected to build the MLR model. For the second-layer SOM, only horizontal curvature was selected to build the MLR model. When more variable enters the MLR model for the second-layer SOM, the coefficient of model will get much worse.

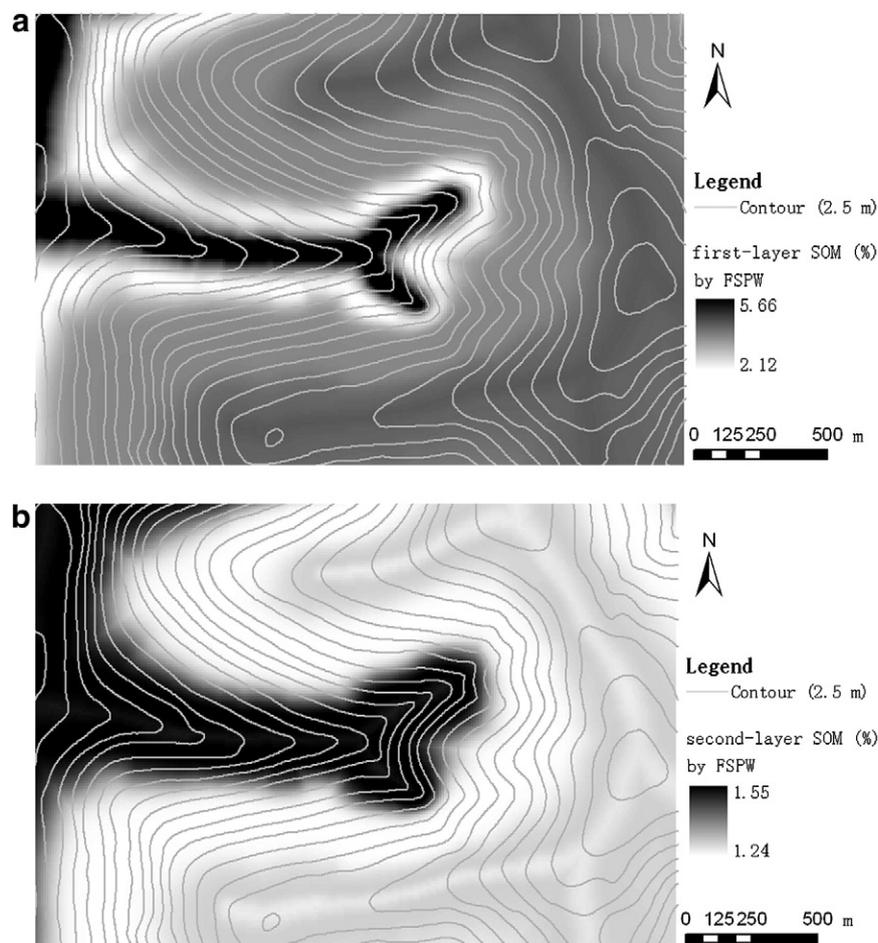


Fig. 6. Soil organic matter contents predicted by FSPW and MLR. a) SOM (10–15 cm) by FSPW; b) SOM (35–40 cm) by FSPW; c) SOM (10–15 cm) by MLR; d) SOM (35–40 cm) by MLR.

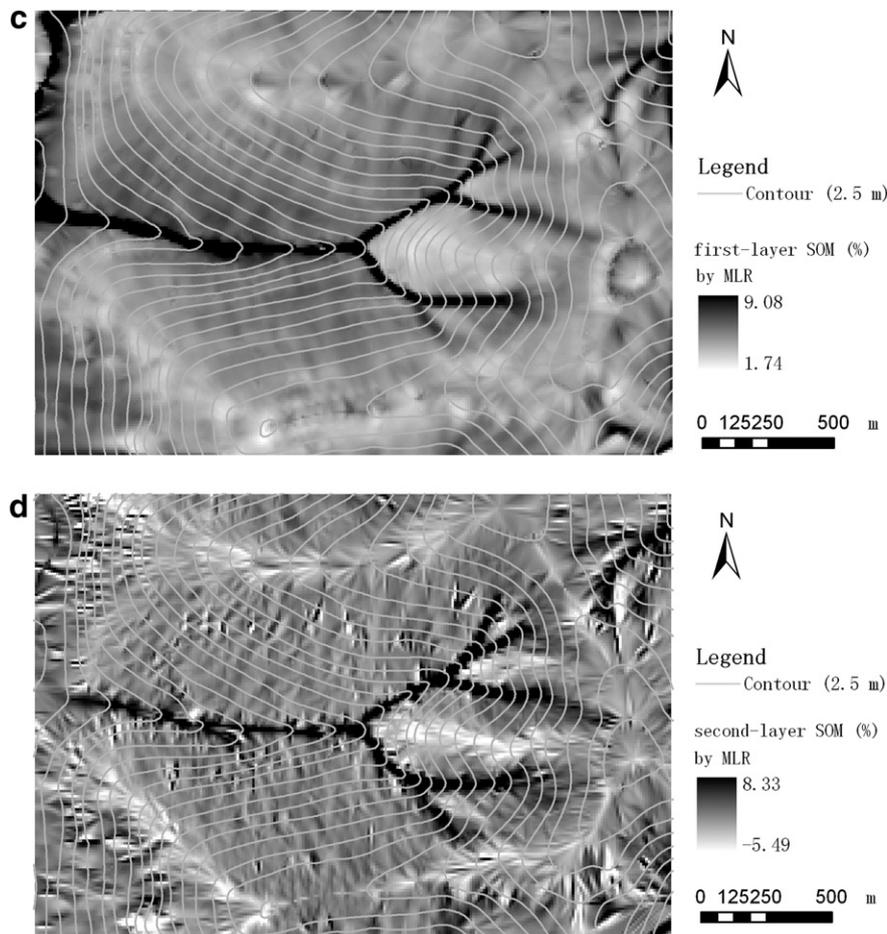


Fig. 6 (continued).

5. Results and discussion

5.1. Results and discussion for model-development area

The spatial distributions of SOM content in the first and second layers as predicted by the FSPW model are shown in Fig. 6a and b; those from the MLR model are shown in Fig. 6c and d. The FSPW model predicts a much smoother spatial transition than MLR. Smooth transition of SOM content over a small, low-relief watershed seems reasonable since it matches the smooth, low-relief terrain observed in the field. For example, the smooth decrease of first-layer SOM predicted by FSPW in the foot slope (Fig. 6a) corresponds to a more rapid change in slope gradient in the downslope direction in this area.

The MLR model predicts the spatial distribution of SOM to contain many seemingly unreasonable irregular transitions, especially along the contour, which correlate with minor features in the DEM. These minor features add unwanted and rough variation in the topographic

attributes (such as topographic wetness index and horizontal curvature). The rough variation is amplified by the MLR model which, by design, relates variation in the dependent variable to variation in the independent variables whether or not the variation makes sense in terms of changes in the actual soil properties. Such amplification is particularly significant in areas of gently varying topography where small local undulations in the terrain data produce significant apparent changes in the computed topographic attributes.

The FSPW model is less sensitive to minor features in the DEM, because the terrain information used in the FSPW model is the fuzzy slope positions, which itself has low sensitivity to minor features. In the computation of the similarity between any locations and a given slope position, there is a step wherein a minimum operator integrates the individual similarities based on individual terrain attributes (Qin et al., 2009). Thus much of the minor variability in one topographic attribute does not propagate to the fuzzy slope position data.

It should also be noted that the second-layer SOM map predicted by MLR (Fig. 6d) contains areas with obviously false value ranges (e.g.,

Table 5

Evaluation results of SOM predicted in the model-development area.

| Predictive model | Evaluation points | Organic matter (10–15 cm) | | | Organic matter (35–40 cm) | | |
|----------------------------------|--|---------------------------|------|------|---------------------------|------|------|
| | | CC | MAE | RMSE | CC | MAE | RMSE |
| FSPW (with 5 modeling points) | 43 (independent with modeling points) | 0.404** | 1.12 | 1.40 | 0.545** | 0.91 | 1.30 |
| MLR (with 48 modeling points) | 48 (same as modeling points) cross-validation | 0.755** | 0.81 | 0.98 | 0.780** | 0.65 | 0.94 |
| | | 0.645** | 0.94 | 1.24 | 0.674** | 0.77 | 1.16 |

** Correlation is significant at the 0.01 level (2-tailed).

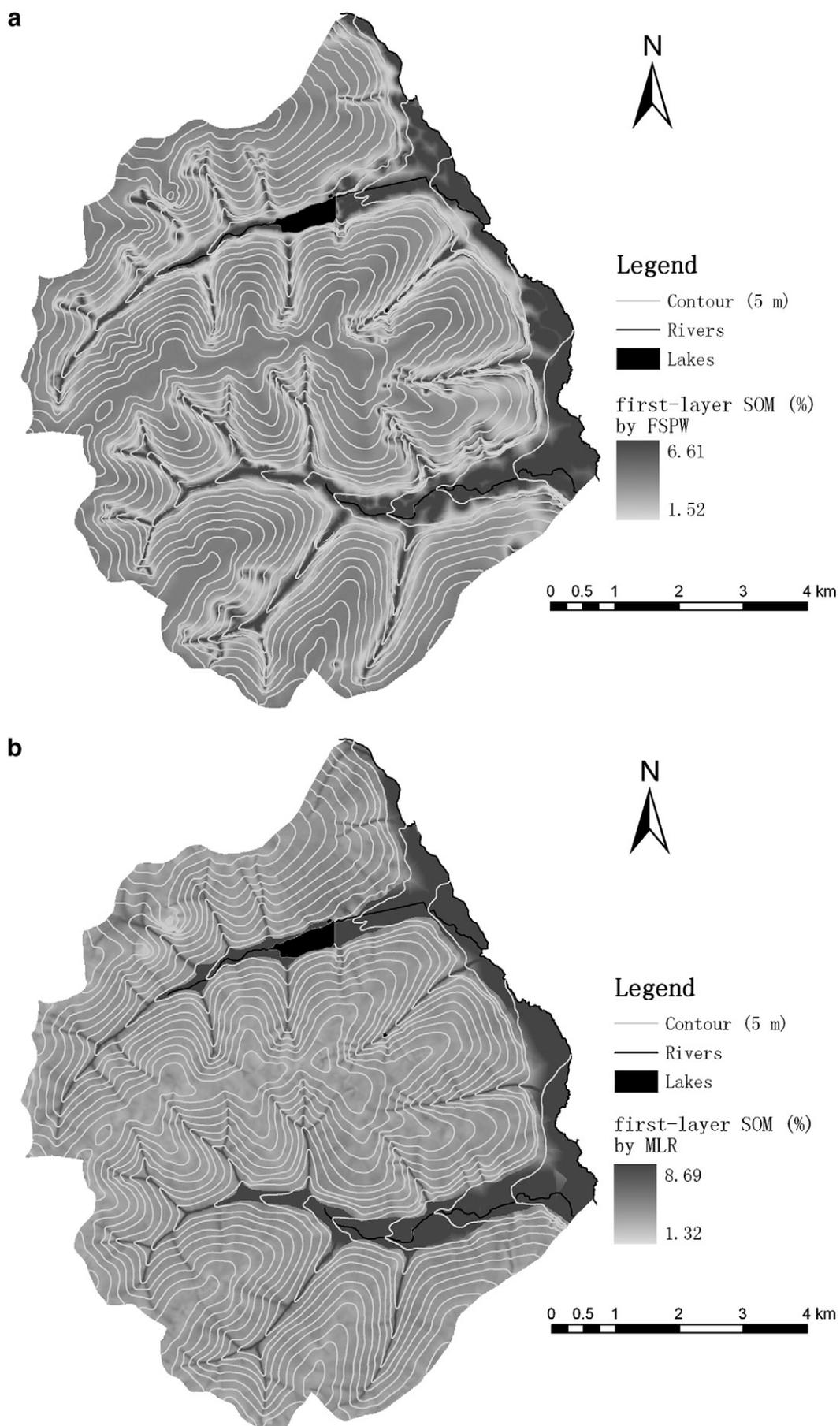


Fig. 7. First-layer SOM predicted in the model-extrapolation area. a) by FSPW; b) by MLR.

negative values). This is not the case in SOM maps predicted by the FSPW model. The negative values in the MLR results relate to the unwanted extrapolation of the regression model.

The FSPW model performance was evaluated by comparing the model development error and cross-validation error of the MLR model with the validation error of the FSPW model as observed in the model-development area. The validation error of the FSPW model is derived using the 43 points in the model-development area, that is, the 48 model development field sampling points for MLR less the five points used for developing the FSPW. Clearly, such an evaluation would seem to be biased against the FSPW model. However, since the FSPW model produces similar results as the MLR model using this evaluation, we can conclude that the FSPW performs better than the MLR model from the perspective of field sample requirements.

Evaluation results (Table 5) by evaluation points show that the correlation coefficients (CC) between predicted and measured values are significant for both FSPW and MLR models, and the predicted values from MLR show higher correlation with the observed values than do those from FSPW. The SOM prediction error from FSPW is higher than that from MLR. Considering that the number of modeling points for FSPW in this case is only five while all 48 of the evaluation points were the modeling points for the MLR model, we consider that the performance of FSPW is at least comparable with, if not better than, that of MLR for the model-development area from the perspective of the number of samples required and the level of accuracy achieved.

5.2. Results and discussion for model-extrapolation area

To assess the portability of the predictive models, both the FSPW and MLR models developed in the model-development area were applied directly to the model-extrapolation area without further tuning. Note that there is bias (about 0.6%, as shown in Table 1) between the mean of SOM samples in the first-layer in the model-development area and the SOM mean in the model-extrapolation area, although these two areas are nearby and with very similar environmental conditions (terrain condition, parent material, land use, etc.). We suspect this bias is the result of difference of other natural or human factors between these two areas. The influence on soil variation by those natural or human factors did not be directed considered in both FSPW and MLR models in this paper. So the performance evaluation in the model-extrapolation area is equal for both models.

The patterns in the predicted SOM maps and the range of predicted values (Fig. 7) are similar to those obtained in the model-development area. The statistical indices of predicted values compared to measured values for 102 independent validation points (Table 6) show that the predicted results using the FSPW model have a significant correlation (at the 0.01 level) between predicted and measured values, while those from the MLR model show little correlation.

The error in the prediction of first-layer SOM is lower with FSPW than with MLR. Tables 5 and 6 show that the error in the prediction of first-layer SOM with FSPW decreases when applied directly to the model-extrapolation area. The opposite is true of the

MLR model which suggests that FSPW is more portable than MLR in this case.

6. Conclusions

Spatial gradation of slope positions can reflect the integrative effect of earth surface processes on slopes. Although currently there are several approaches to deriving fuzzy slope positions to quantify the spatial gradation of slope positions, fuzzy slope positions are seldom used in digital soil mapping and other terrain-related geographical modeling. In this paper we present a preliminary attempt at applying fuzzy slope positions to digital soil mapping in small, low-relief areas at fine scale.

Taking prediction of spatial distribution of soil properties as an application domain, we applied the fuzzy slope position information within a purposive sampling framework for predictive soil mapping in a small, low-relief area. The fuzzy slope positions with a system of five slope positions were derived from a prototype-based approach. The fuzzy slope positions, treated as environmental combinations, were used to direct purposive sampling for determining typical soil property values in environmental combinations. Then the soil property value at a location was predicted by a weighted average model (FSPW) by combining the fuzzy slope positions with the typical soil property values of slope position types.

We selected a small, low-relief catchment in north-eastern China as a model-development area to develop the FSPW model. Soil samples were collected at five field sampling points (each point for a slope position type) to obtain the typical soil organic matter (SOM) value for each of the five slope position types. The SOM values at these five modeling points and the fuzzy slope positions were then used to predict the spatial distribution of SOM contents at depths of 10–15 cm and 35–40 cm at a resolution of 10 m. We evaluated the performance of the developed FSPW model both qualitatively and quantitatively by comparing results with those from a reference model developed based on a widely-used multiple linear regression (MLR) model. Both of the developed models were then extrapolated to a nearby watershed with similar environmental conditions to examine the portability of FSPW.

The application and evaluation showed that fuzzy slope positions effectively predict the spatial distribution of soil properties in a small, low-relief catchment at a fine scale with fewer soil samples. In the model-development area, the FSPW developed with only five modeling points produced results comparable with those of MLR using 48 modeling points, from the perspectives both of the number of field samples required and the level of accuracy achieved. In addition, the spatial distributions of the soil properties predicted by FSPW matched the smooth, low-relief terrain in the study area more closely than those predicted by MLR. Abrupt transition observed in the MLR prediction results is more related to the minor features that are very common in DEMs of low-relief areas. The FSPW model, being less sensitive to minor features in the DEM, does not produce this rough transition pattern in the prediction results. Validation of both models in the model-extrapolation area shows that the FSPW model developed in the model-development area is much more stable than the MLR model when extrapolated to areas with similar environmental conditions.

Our study has given a preliminary indication that the quantified information on spatial gradation of slope positions (measured as fuzzy membership values) can provide useful information for digital soil mapping. With the aid of fuzzy slope positions, the number of field samples needed for model development can be drastically reduced. We believe it is also potentially useful for other terrain-related geographic modeling. Further research will include sensitivity analysis of FSPW for the determination of typical soil property values on typical slope positions, together with more application assessments in areas of different terrain.

Table 6
Evaluation results of prediction of the first-layer SOM in the model-extrapolation area (with 102 validation points).

| | First-layer SOM | | |
|------|-----------------|------|------|
| | CC | MAE | RMSE |
| FSPW | 0.319** | 0.97 | 1.31 |
| MLR | 0.056 | 1.02 | 1.49 |

** Correlation is significant at the 0.01 level (2-tailed).

Acknowledgments

This study was funded by the National Natural Science Foundation of China (40971235,40971236). This study was also partly funded by the National Basic Research Program of China (2007CB407207), the Knowledge Innovation Program of the Chinese Academy of Sciences (KZCX2-YW-Q10-1-5), and the National Key Technology R&D Program of China (2007BAC15B01). Supports from the State Key Laboratory of Resources and Environmental Information Systems, and the University of Wisconsin-Madison are also appreciated. We thank three reviewers for their constructive comments on the earlier version of this paper.

References

- Bao, S.D., 2000. Soil and Agrochemistry Analysis, 3rd. China Agricultural Press, Beijing (in Chinese).
- Bell, J.C., Grigal, D.F., Bates, P.C., 2000. A soil-terrain model for estimating spatial patterns of soil organic carbon. In: Wilson, J.P., Gallant, J.C. (Eds.), *Terrain Analysis—Principles and Applications*. Wiley, New York, pp. 295–310.
- Gessler, P.E., Moore, I.D., McKenzie, N.J., Ryan, P.J., 1995. Soil-landscape modeling and spatial prediction of soil attributes. *International Journal of Geographical Information Science* 9, 421–432.
- Hengl, T., Stephan, G., Dhruba, P.S., 2004. Reduction of errors in digital terrain parameters used in soil-landscape modeling. *International Journal of Applied Earth Observation and Geoinformation* 5, 97–112.
- Hudson, B.D., 1992. The soil survey as paradigm-based science. *Soil Science Society of America Journal* 56, 836–841.
- Jenny, H., 1941. *Factors of Soil Formation, A System of Quantitative Pedology*. McGraw-Hill, New York.
- King, D., Bourennane, H., Isampert, M., Macaire, J.J., 1999. Relationship of the presence of a non-calcareous clay-loam horizon to DEM attributes in a gently sloping area. *Geoderma* 89, 95–111.
- Lark, R.M., 2000. Designing sampling grids from imprecise information on soil variability: an approach based on the fuzzy kriging variance. *Geoderma* 98, 35–59.
- MacMillan, R., Pettapiece, W., Nolan, S., Goddard, T., 2000. A generic procedure for automatically segmenting landforms into landform elements using DEMs, heuristic rules and fuzzy logic. *Fuzzy Sets and Systems* 113, 81–109.
- McBratney, A.B., Odeh, I.O.A., Bishop, T.F.A., Dunbar, M.S., Shatar, T.M., 2000. An overview of pedometric techniques for use in soil survey. *Geoderma* 97, 293–327.
- McBratney, A.B., Mendonca Santos, M.L., Minasny, B., 2003. On digital soil mapping. *Geoderma* 117, 3–52.
- Moore, I., Gessler, P., Nielsen, G., Peterson, G., 1993. Soil attribute prediction using terrain analysis. *Soil Science Society of America Journal* 57, 443–452.
- Pennock, D., Zebarth, B., de Jong, E., 1987. Landform classification and soil distribution in hummocky terrain, Saskatchewan, Canada. *Geoderma* 40, 297–315.
- Planchon, O., Darboux, F., 2001. A fast, simple and versatile algorithm to fill the depressions of digital elevation models. *Catena* 42, 159–176.
- Pocknee, S., Boydell, B.C., Green, H.M., Waters, D.J., Kvien, C.K., 1996. Directed soil sampling. In: Robert, P.C., et al. (Ed.), *Proceedings of the 3rd International Conference*, June 23–26, 1996. Minneapolis, MN. ASA-CSSA-SSSA, Madison, WI, pp. 159–168.
- Qin, C., Zhu, A.X., Li, B., Pei, T., Zhou, C., 2006. Review of multiple flow direction algorithms based on gridded digital elevation models. *Earth Science Frontiers* 13, 91–98 (in Chinese with English abstract).
- Qin, C.-Z., Zhu, A.X., Shi, X., Li, B., Pei, T., Zhou, C., 2009. The quantification of spatial gradation of slope positions. *Geomorphology* 110, 152–161.
- Qin, C.-Z., Zhu, A.X., Pei, T., Li, B.-L., Scholten, T., Behrens, T., Zhou, C.-H., 2011. An approach to computing topographic wetness index based on maximum downslope gradient. *Precision Agriculture* 12, 32–43.
- Schmidt, J., Hewitt, A., 2004. Fuzzy land element classification from DTMs based on geometry and terrain position. *Geoderma* 121, 243–256.
- Shary, P., Sharaya, L., Mitusov, A., 2002. Fundamental quantitative methods of land surface analysis. *Geoderma* 107, 1–32.
- Skidmore, A., 1990. Terrain position as mapped from a gridded digital elevation model. *International Journal of Geographical Information Systems* 4, 33–49.
- Ventura, S.J., Irvin, B.J., 2000. Automated landform classification methods for soil-landscape studies. In: Wilson, J.P., Gallant, J.C. (Eds.), *Terrain Analysis: Principles and Applications*. John Wiley & Son, New York, pp. 267–294.
- Zhu, A.X., 1997. A similarity model for representing soil spatial information. *Geoderma* 77, 217–242.
- Zhu, A.X., Yang, L., Li, B., Qin, C., English, E., Burt, J.E., Zhou, C.H., 2008. Purposive sampling for digital soil mapping for areas with limited data. In: Hartemink, A.E., McBratney, A.B., Mendonca Santos, M.L. (Eds.), *Digital Soil Mapping with Limited Data*. Springer-Verlag, New York, pp. 233–245.
- Zhu, A.X., Yang, L., Li, B., Qin, C., Pei, T., Liu, B., 2010a. Construction of quantitative relationships between soil and environment using fuzzy c-means clustering. *Geoderma* 155 (3–4), 164–174.
- Zhu, A.X., Qi, F., Moore, A., Burt, J.E., 2010b. Prediction of soil properties using fuzzy membership values. *Geoderma* 158, 199–206.