Chin. Geogra. Sci. 2013 Vol. 23 No. 6 pp. 680–691
doi: 10.1007/s11769-013-0632-7
Springer Science Press
www.springerlink.com/content/1002-0063

Mapping Detailed Soil Property Using Small Scale Soil Type Maps and Sparse Typical Samples

ZHANG Shujie[1, 2], ZHU Axing[1, 3], LIU Wenliang[4], LIU Jing[3], YANG Lin[1]

(1. *State Key Laboratory of Resources and Environment Information System*, *Institute of Geographic Sciences and Natural Resources Research*, *Chinese Academy of Sciences*, *Beijing* 100101, *China*; 2. *University of Chinese Academy of Sciences*, *Beijing* 100049, *China*; 3. *Department of Geography*, *University of Wisconsin*, *Madison* 53706, *U.S.A.*; 4. *Institute of Remote Sensing and Digital Earth*, *Chinese Academy of Sciences*, *Beijing* 100101, *China*)

Abstract: Soil type maps at the scale of 1：1 000 000 are used extensively to provide soil spatial distribution information for soil erosion assessment and watershed management models in China. However, the soil property maps produced through conventional direct linking method usually suffer low accuracy as well as the lack of spatial details within a soil type polygon. This paper presents an effective method to produce detailed soil property map based on representative samples which were extracted from each polygon on the 1：1 000 000 soil type map. The representative sample of each polygon is defined as the location that can represent the largest area within the polygon. The representativeness of a candidate sample is determined by calculating the soil-forming environment condition similarities between the sample and other locations. Once the representative sample of each polygon has been chosen, the property values of the existing typical samples are assigned to the corresponding representative samples with the same soil type. Finally, based on these representative samples, the detailed soil property map could be produced by using existing digital soil mapping methods. The case study in XuanCheng City, Anhui Province of China, demonstrated the proposed method could produce soil property map at a higher level of spatial details and accuracy: 1) The soil organic matter (SOM) map produced based on the representative samples can not only depict the detailed spatial distribution of SOM within a soil type polygon but also largely reduce the abrupt change of soil property at the boundaries of two adjacent polygons. 2) The Root Mean Squared Error (RMSE) of the SOM map based on the representative samples is 1.61, and it is 1.37 for the SOM map produced by using conventional direct linking method. Therefore, the proposed method is an effective approach to produce spatial detailed soil property map with higher accuracy for environment simulation models.
Keywords: soil type map; existing typical sample; representative sample; detailed soil property map; digital soil mapping

Citation: Zhang Shujie, Zhu Axing, Liu Wenliang, Liu Jing, Yang Lin, 2013. Mapping detailed soil property using small scale soil type maps and sparse typical samples. *Chinese Geographical Science*, 23(6): 680–691. doi: 10.1007/s11769-013-0632-7

1 Introduction

The spatial distribution information of soil property is important input data for soil erosion simulation and water resource management (Zhu *et al*., 1996; Moriasi and Starks, 2010; Mukundan *et al*., 2010; Li *et al*., 2012). However, the published major data sources for soil property spatial information are often in the form of soil type maps with small map scales, such as 1：1 000 000 soil type map in China, and sparse typical samples (Shi *et al*., 2007; Yu *et al*., 2007). Soil property maps are often produced by linking the property values of existing typical samples to the corresponding polygon of the soil type map in China (Zhao *et al*., 2006; Zhang *et al*.,

Received date: 2012-11-28; accepted date: 2013-02-27
Foundation item: Under the auspices of Program of International Science & Technology Cooperation, Ministry of Science and Technology of China (No. 2010DFB24140), National Natural Science Foundation of China (No. 41023010, 41001298), National High Technology Research and Development Program of China (No. 2011AA120305)
Corresponding author: LIU Wenliang. E-mail: wlliu@irsa.ac.cn

2008). This method was referred to as 'conventional direct linking method' in this paper. Constrained by the small scale of the soil type maps and the limited typical samples, the level of the spatial details and the accuracy of obtained soil property map may be very low. On the other hand, the size of existing typical samples is often small and the accurate coordinates of these samples were not recorded in the Published National Soil Attribute Table, thus based on these limited samples it is impossible to produce detailed soil property maps by using current digital soil mapping methods (e.g., interpolation and regression method). However, current environment models (e.g., environmental, hydrological, ecological models) are in need of detailed soil property spatial distribution information (Band and Moore, 1995; Zhu, 1997; Liu *et al*., 2005; Chen *et al*., 2012), the soil property maps produced by using the direct linking method can not meet the need of these models.

For producing spatial detailed and higher accuracy soil property maps based on the existing small scale soil type map, many methods have been proposed in the published literatures. Brus *et al*. (2008) used Bayesian Maximum Entropy to predict soil categories in the Netherlands by estimating the probabilities of occurrence of soil categories, which is based on the 1∶50 000 soil map as soft information and 8369 soil point observations as hard information. The shortcoming of this method is that the computational burden is prohibitive when the number of soil categories larger than 10. In addition, a large number of existing samples are needed to calibrate the probability model. Vitharana *et al*. (2008) improved the 1∶20 000 soil property map in Belgium by using regression Kriging based on the relationship between electrical conductivity and depth of the Tertiary clay substratum. The relationship was modeled by an exponential curve using 60 calibration points within 0.14 km^2. This method requires a sample set with high density, and the fieldwork is very labor and cost intensive. Kempen *et al*. (2009) build a multinomial logistic regression model to quantify the relationship between ancillary covariates and soil group. The regression model was then used to update the 1∶50 000 soil groups map over 2680 km^2 area in Drenthe Province in the Netherlands. Large number of samples including 16 282 soil profile descriptions from the Dutch soil information system (DSIS) and 702 profile observations were collected to build the regression model. Yang *et al.* (2010; 2012) developed a method to upgrade the 1∶20 000 conventional soil type map in New-Bruswick by using the knowledge on soil-environment relationships. Such knowledge was extracted by relating the environment condition combinations to the soil types. Although this method does not require additional field work, it can not utilize the existing samples which do contain useful information. Yu (2012) proposed to use linking method to estimate the basic soil property distribution, and then the variation inside each polygon of the soil type map was interpolated by neighboring soil samples with stratification. At last, the soil properties of the transition areas between different soil types were re-estimated by a weighted average of estimations based on soil samples with stratification. This new method is based on the soil type map with the scale of 1∶1 000 000 and 1765 field samples in Jilin Province. However, it is usually very difficult to collect so many field samples for ordinary soil map users.

Therefore, the existing methods investigated in the previous studies for updating soil maps heavily rely on a large amount of samples (existing samples or new additional field samples), and the map scales used in most of the aforementioned studies are much more detailed compared to the scale of soil type map in China (1∶1 000 000). On the other hand, because the number of existing typical samples in China is often very small and the accurate coordinate were not recorded in the Published National Soil Attribute Table, these typical samples can not be used directly for current digital soil mapping methods. Therefore, the above existing methods may be limited to improve Chinese soil property maps. This paper proposed an alternative method to produce detailed soil property map with higher accuracy using 1∶1 000 000 soil type map and sparse existing typical samples that were stored in the Published National Soil Attribute Table. The new method is based on the representative samples which are extracted from each polygon on the soil type map, and it is very appropriate when the soil map user does not have enough field samples to apply existing digital soil mapping methods. The case study in southeast of Anhui Province was used to illustrate the effectiveness of the proposed method.

2 Material and Methods

2.1 Study area

Xuancheng City (30°33′31″–31°18′38″N, 118°28′08″–119°38′40″E) is in the southeast of Anhui Province in

China, which includes seven counties and three of them are chosen as the study area of this paper, that is, Xuanzhou District, Langxi County and Guangde County (Fig. 1). It is about 5900 km^2 with the highest elevation of 1039 m. This area is in the transition zone from the southern mountainous area in Anhui to the Middle and Lower Yangtze River Plain. Located in the subtropical monsoon climate zone, the area is hot and rainy in summer, while it is cold and dry in winter (Zhao *et al.*, 2007; Sun *et al.*, 2008). Average annual temperature varies from 11.6℃ to 15.8℃ and annual precipitation is 1240–1780 mm. The parent materials are mainly composed of mild clay-silt-gravel (34%) formed during Quaternary system and sandstone (15%). The landform is characterized by hill and plain, with some mountains in the south. The characteristic vegetation types are evergreen coniferous forest and deciduous broad-leaved forest, most of which are secondary forest or human-made forest. The major land use type is farmland with rice as the stable crop. The whole study area experienced severe human activities. The main soil types are red soil and paddy soil, accounting for 46.5% and 39.7%, respectively.

2.2 Data and processing

The data for mapping soil organic matter (SOM) include the 1 ∶ 1 000 000 soil type map in the study area and 106 typical samples in Anhui Province. The typical samples are stored in the Published National Soil Attribute Table. Each typical sample contains 13 kinds of soil properties, including soil type, location (province), depth (cm), gravel (%), coarse sand (%), fine sand (%), silt (%), clay (%), soil organic matter (%), pH, N (%), P (%), K (%). There are 18 soil subgroups according to the 1 ∶ 1 000 000 soil type map in the study area, and there are 26 subgroups with the 106 typical samples in Anhui Province. As the *X* and *Y* coordinates of each typical sample were not recorded in the Published National Soil Attribute Table, the samples which were located in the study area can not be determined. So, all the samples in Anhui Province would be used in this paper. These are the all public soil data in this study area, which were provided by Data Center for Resources and Environmental Sciences, Institute of Geographic Sciences and Natural Resources Research, Chinese Academy of Sciences.

There are 55 independent samples which were used to validate the accuracy of produced soil organic matter maps. The spatial distribution of these validation samples is regular with 10 km spacing distance, and these validation samples were collected in October 2011. October is harvest time for the study area, so the soil prop-

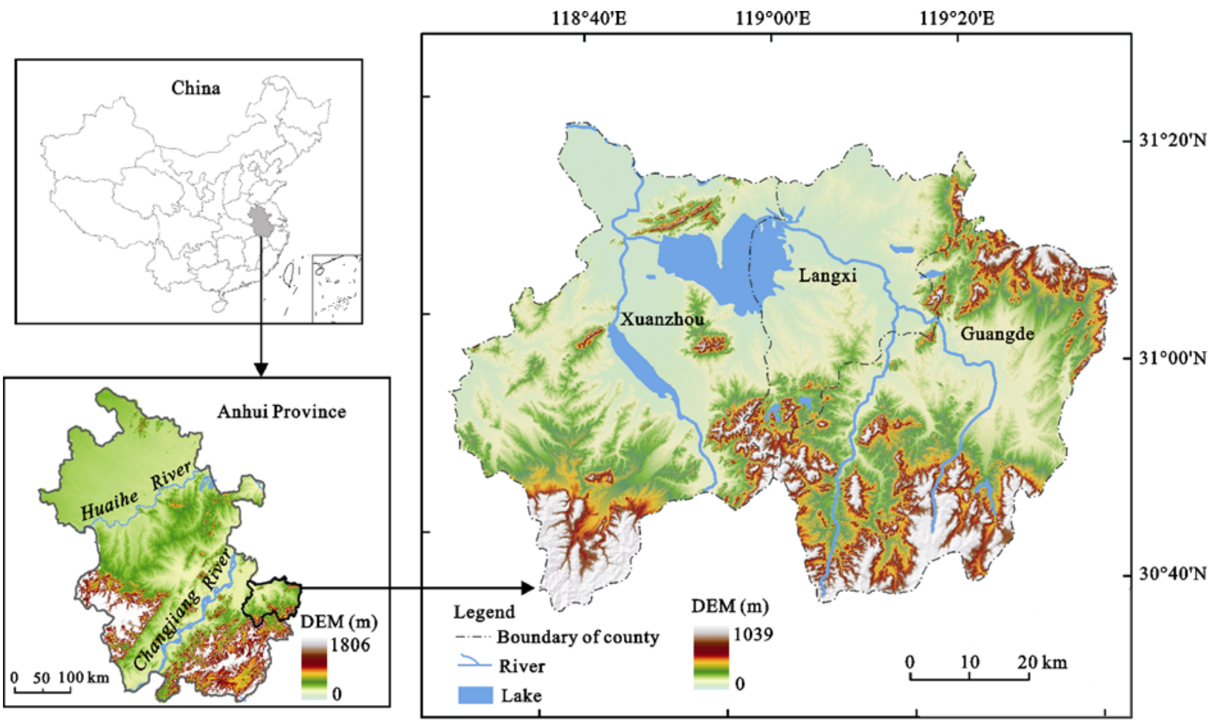

Fig. 1 Location of study area

erty values were less affected by human activities.

2.3 Methods

2.3.1 Overall description of proposed soil property mapping method

Soil property values of existing typical samples in the Published National Soil Attribute Table are assumed to be typical for corresponding soil type, and each polygon of soil type map is associated to a soil type. So, it is reasonable to assign the soil properties of these typical samples to the representative samples of the polygons with the same soil type name. If the representative samples of all the polygons are extracted and soil properties are assigned, soil property maps could be produced by using some existing digital soil mapping methods, such as regression model, Kriging interpolation, Soil-Land Inference Model (SoLIM), individual representativeness methods, *etc*. (Goovaerts, 1999; McBratney *et al*., 2000; Qin *et al*., 2012; Yang *et al*., 2012; Zhang *et al*., 2012; Liu *et al*., 2013). The flowchart of this overall methodology has been shown as Fig. 2. This method based on representative samples can overcome the shortcomings of conventional direct linking method, which can not only describe the detailed spatial distribution of soil property within a soil type polygon but also largely reduce the abrupt change of soil property at the boundaries of two adjacent polygons.

Therefore, the most important question becomes how to extract the location of the representative sample of each polygon on the soil type map. According to the process of producing soil type maps, each polygon of soil map is related to a certain landscape unit. The landscape unit can be characterized by using a group of environment covariates (e.g. elevation, slope, curvature, vegetation, climate, *etc*). The representative sample can be extracted through analyzing the environment similarity between any two locations within a polygon. Figure 3 shows how to determine whether a location can represent another location within a polygon. The soil-forming environment condition at location A and location B is characterized by their elevation, slope, curvature, *etc*. If the environment condition similarity between A and B is

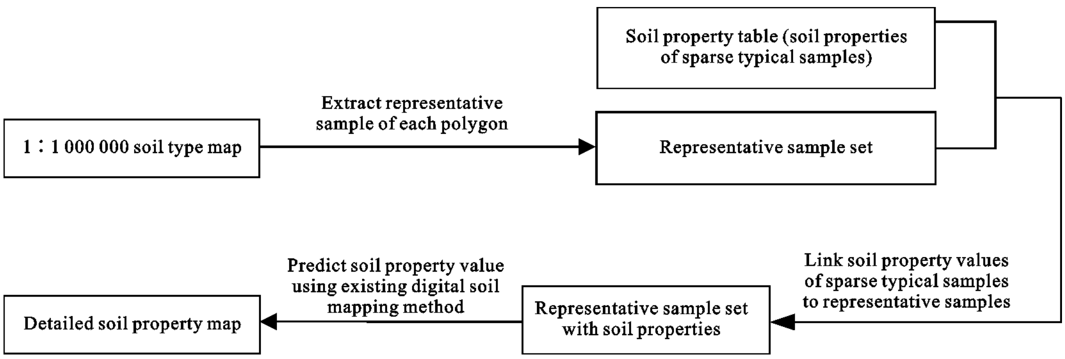

Fig. 2 Flowchart of soil property mapping method based on representative samples

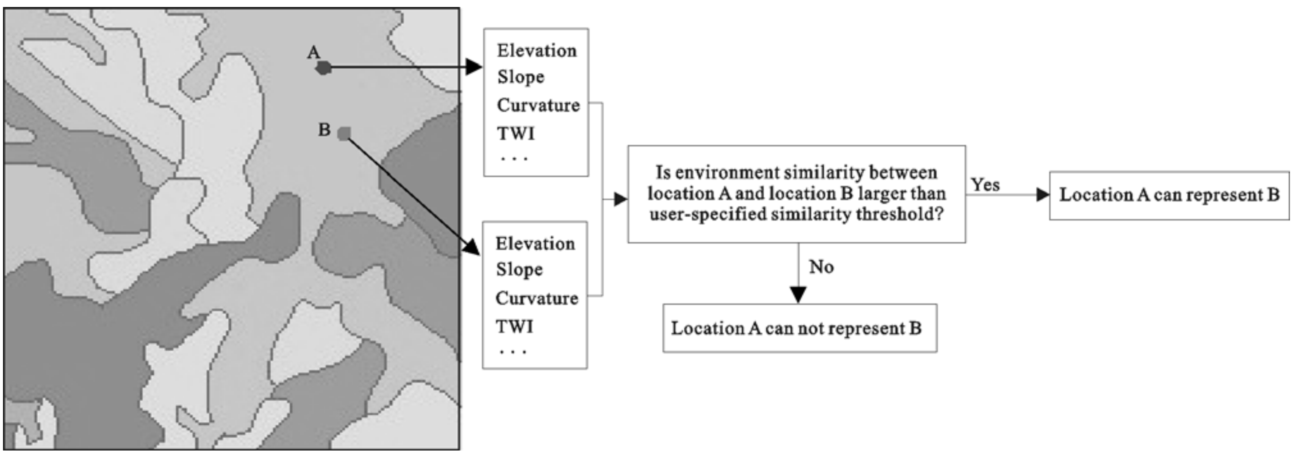

Fig. 3 Scheme of determining whether location A can represent location B. TWI is topographic wetness index

larger than a user-specified similarity threshold, the soil-forming environment condition of location A is regarded to be representative of location B, and the soil property of location B can be predicted through the soil property of location A. Otherwise, location B can not be represented by location A. In such a way, for a given location, the area which is represented by the location can be calculated within the soil type polygon. Iterate this procedure at each location in the soil type polygon, the area that each location represents can be determined (see 2.3.3 for details). The location that can represent the largest area in the polygon will be chosen as the representative sample for the polygon. Note that there may be several suitable representative locations identified by this method, because it is possible that the representing areas of several candidate representative locations are same size; but under our assumptions, any one of them can be selected.

2.3.2 Selection of environmental covariates

In order to quantify the representativeness of a certain location, it is important to select appropriate environment covariates for calculating soil-forming environment condition similarity between two locations. Covariates could be selected according to soil factor equation 'CLORPT' (Jenny, 1941), which can be described as $S = f(Cl, O, R, P, T\ldots)$. Cl represents climate conditions; O is organism; R is topography; P stands for parent material; and T is time. McBratney *et al*. (2003) summarized the specific auxiliary environment covariates used in the field of digital soil mapping.

Under the assumption that parent material and climate condition are homogeneous over each soil type polygon, the topographic factors are main covariates to indicate the variation of the spatial distribution of soil property. Slope, profile curvature, planform curvature, topographic wetness index (TWI) were selected as covariates (Fig. 4). All these covariates were derived from DEM with 90 m resolution (http://glcf.umiacs.umd.edu/). Slope, planform curvature and profile curvature were calculated by 3dMapper software (http://www.TerrainAnalytics.com). TWI was calculated by using a modified Multiple Flow Direction (MFD) algorithm (Qin *et al*., 2007). Combining all the values of environmental covariates, a 'environment covariates vector' was used to depict the environmental condition at each location (grid) (Zhu *et al*., 1996).

$$\vec{e}_{ij} = (e_{ij}^1, e_{ij}^2, \ldots, e_{ij}^m) \tag{1}$$

where $\vec{e}_{ij}$ is the environment covariates vector at the location (i, j); m is the number of covariates.

2.3.3 Calculation of representative samples

In order to select the optimal representative samples, the representing area of each location should be calculated within a soil type polygon. A traversing algorithm was performed to achieve this goal. First of all, the polygon-based soil type map was converted into grid-based data with the same spatial resolution as the selected topographic environmental covariates. For a given location (i, j), calculate the similarities between this location and all the other locations within the polygon (Fig. 5). If the similarity between location (i, j) and a location (a, b) is larger than a user-defined similarity threshold (e.g., 0.8), the location (i, j) is regarded to be representative of (a, b), which was colored green in Fig. 5. Otherwise, location (i, j) can not represent location (a, b), which was colored red in Fig. 5. The total of the grids with green color is the representing area of the location (i, j).

Environment condition similarity calculation was carried out on two levels, individual environment variable level and location level, which was demonstrated as Equation (2).

$$S_{i,j}^{a,b} = \mathop{P}_{v=1}^{m}(E^v(e_{i,j}^v, e_{a,b}^v)) \tag{2}$$

where $S_{i,j}^{a,b}$ represents the similarity between location (i, j) and location (a, b) $(a = 1, 2, 3, \ldots, n; b = 1, 2, 3, \ldots, n)$; $e_{i,j}^v$ is the value of the vth environment covariate on location (i, j); $e_{a,b}^v$ is the value of the vth covariate associate with the location (a, b); m is the number of covariates. E is the function for calculating similarity on individual environmental variable level. The E function is associated with the data type of environment covariate. For example, E is usually Boolean function when the covariate is nominal or ordinal type (e.g., parent material). E function may be Euclidean distance, Mahalanobis distance, Gower distance, *etc*. when the covariate is interval or ration type (e.g., temperature, slope) (Shi *et al*., 2004). In this paper, Gower distance was adopted to calculate individual covariate similarity.

P is the function for calculating similarity on location level. The relationship between different environment covariates should be considered when choosing the form of P function. When there is no prior knowledge about the relationship, weighted-average approach is a safe

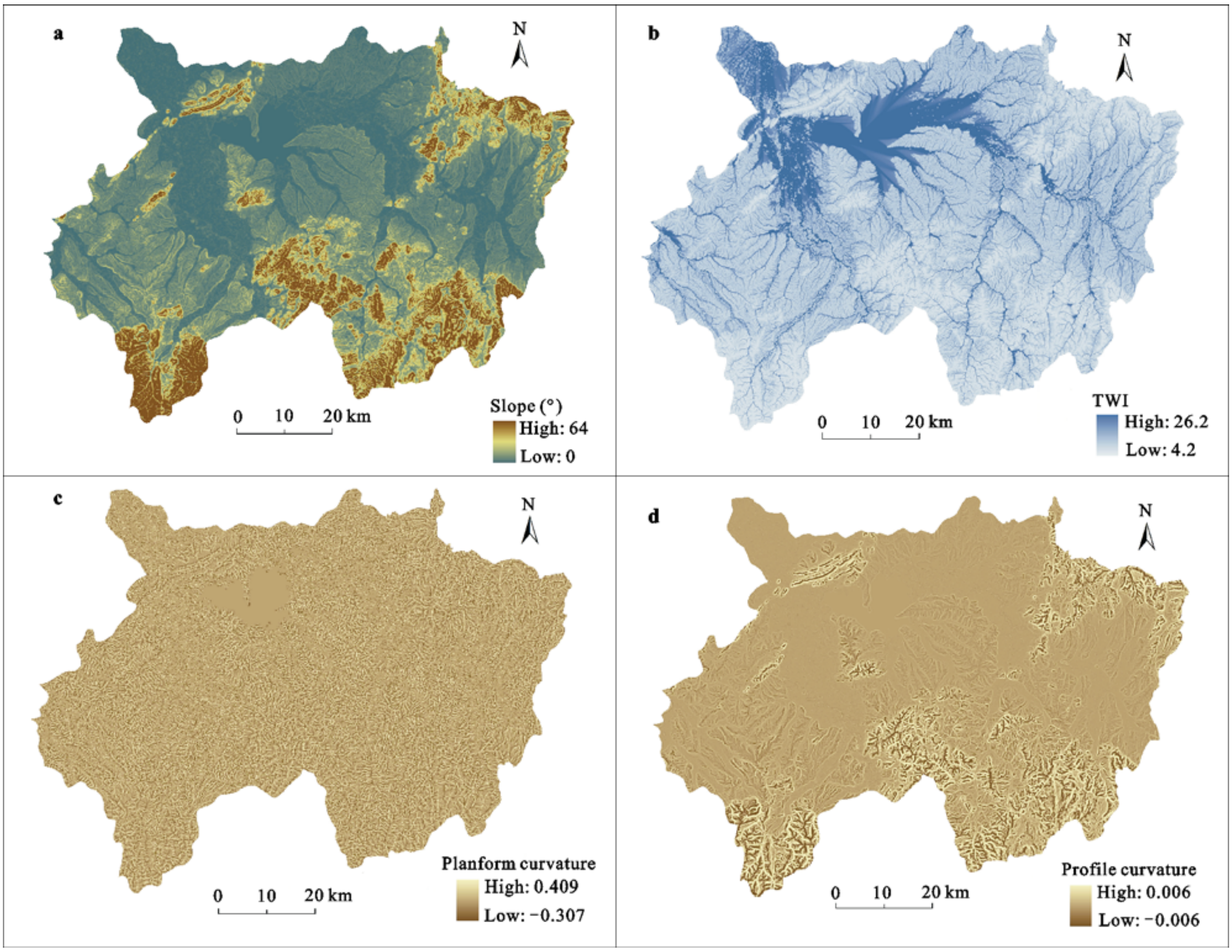

Fig. 4 Topographic covariates of slope (a), topographic wetness index (TWI) (b), planform curvature (c) and profile curvature (d)

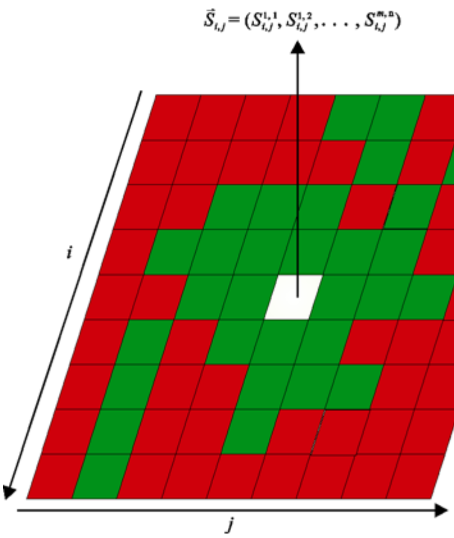

Fig. 5 Calculation of representing area of location (i, j). Location (i, j) can represent green locations, while can not represent red ones

choice (McBratney *et al*., 2003). In this paper, the limiting factor approach was adopted to integrate the similarities on individual environment covariate level (Zhu *et al*., 1996).

Using the similarity calculating method described above, the representing area of each location within a polygon can be calculated under the condition that the similarity threshold is 0.8. In this paper, this similarity threshold was determined subjectively by a soil survey expert. When the soil-forming environment condition similarity between two locations is larger than 0.8 (the maximum value of similarity is 1.0), one location could be considered to be representative of another. The location which can represent the largest area was chosen as the representative sample for the polygon. Traversing all the polygons of the soil type map, the representative sample set for all soil polygons was obtained (Fig. 6).

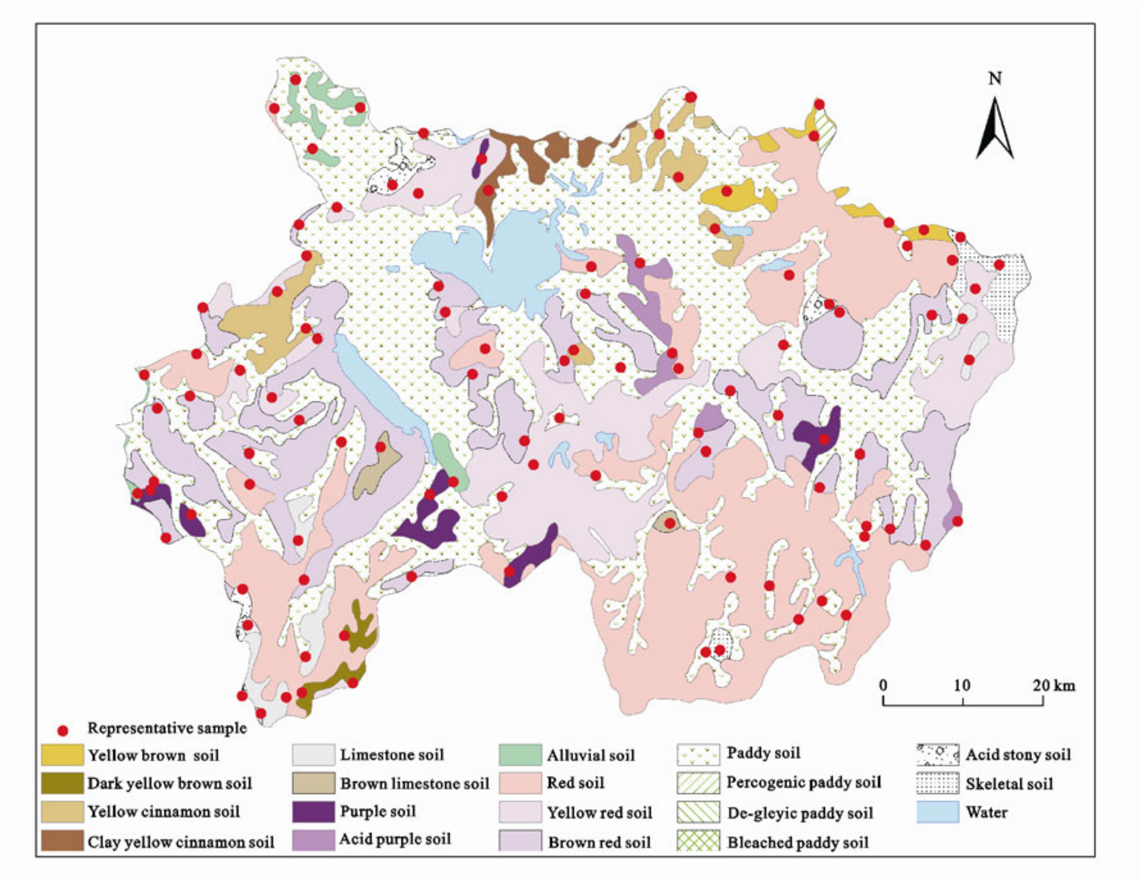

Fig. 6 Spatial distribution of representative samples for 1∶1 000 000 soil type map in study area

2.3.4 Linking soil property value of existing typical samples to representative samples

In the soil attribute database there are 106 existing typical samples with 28 soil subgroups in Anhui Province. As the X and Y coordinates were not recorded in the property table, the samples which were located in the study area can not be determined. So, all these typical samples rather than only the samples located in the study area were used to link the soil properties of typical samples to the representative samples with the same soil type. However, the soil property table in China is not complete. In fact, according to the 1∶1 000 000 soil type map, there are 18 soil subgroups but only 10 soil subgroups of them exist in the Anhui Province soil property table. Thus, the following three criterions were adopted to resolve this problem: 1) If the soil subgroup in this study area exists in the soil property table of Anhui Province, link the average value of the soil properties of the typical samples to the representative samples with the same soil subgroup. 2) If the soil subgroup in this study area does not exist in the soil property table of Anhui Province, but typical samples with the same soil group exist in the soil property table of Anhui Province, then link the average value of the soil properties of the typical samples to the representative samples with the same soil group. 3) If neither the soil subgroup nor the soil group in this study area exist in the soil property table of Anhui Province, but typical samples with the same soil subgroup exist in the soil property table of adjacent provinces, link the average value of the soil properties of the typical samples in the adjacent provinces to the representative samples with the same soil subgroup. According to these three criterions, all the representative samples are assigned the typical soil properties value of the corresponding soil subgroup or group.

3 Results and Evaluation

3.1 Mapping spatial distribution of soil organic matter

Based on these representative samples extracted from each polygon in the soil type map, the spatial distribution of Soil organic matter (SOM) could be mapped by using existing digital soil mapping method (such as regression model, Kriging interpolation model, individual representativeness method, *etc*.). In this paper, the individual representative method (Liu's method) was selected to produce SOM map (Liu *et al*., 2013). The basic theory for Liu's method is the soil factor equation of Dokuchaeiv and the soil-land model described by Hudson (1992), and this method can make full use of auxiliary soil formative environment factors to improve the accuracy of soil map. Besides topographic covariates (slope, planform curvature, profile curvature and TWI), climate and parent materials condition also affect the soil developing over a large area. In this process of predicting soil organic matter over the whole study area, climate data (average annual precipitation and average annual temperature) and parent material data (geology type data) was added to the previous four topographic covariates. The climate data with 1 km resolution were provided by Laboratory of Ecosystem Network Observation and Modeling, Institute of Geographic Sciences and Natural Resources Research, Chinese Academy of Sciences (CAS). Geology Map at the scale of 1∶500 000 was used to describe the characteristics of parent materials, which was provided by the Institute of Soil Science, CAS. Based on the characteristics of rock, the 27 geology types were re-classed into eight ones. Obviously, the resolution of the climate and parent material data is coarser than that of the topographic covariates. However, in the field the variation of the climate and parent material condition is also much smaller than the variation of topographic condition. On the other hand, the climate and parent material condition affects the formation and development of soil at large scale, and at small scale the topographic condition is the main factor. So, the climate and parent material data were resampled into 90 m resolution in this paper.

For comparing the proposed soil property mapping method based on representative samples with conventional direct linking method, the spatial distribution map of soil organic matter was also produced by using conventional direct linking method. Figure 7 shows the spatial distribution maps of soil organic matter in the study area in Xuancheng City by using conventional direct linking method (Fig. 7a) and the proposed new method based on representative samples (Fig. 7b).

3.2 Spatial details level evaluation of soil organic matter map

In the soil organic matter map produced by using conventional direct linking method (Fig. 7a), the variation of soil property only occurs at the common boundaries of two adjacent polygons. Such a map generates the spatial variation of soil distribution and lack of spatial details within each soil type polygon. This polygon-based presentation does not accord with the spatial variation regular of soil property in the field.

In the soil organic matter map produced based on representative samples (Fig. 7b), the abrupt change at the boundaries of two adjacent polygons had been reduced largely. This method links the average value of typical samples to the corresponding representative

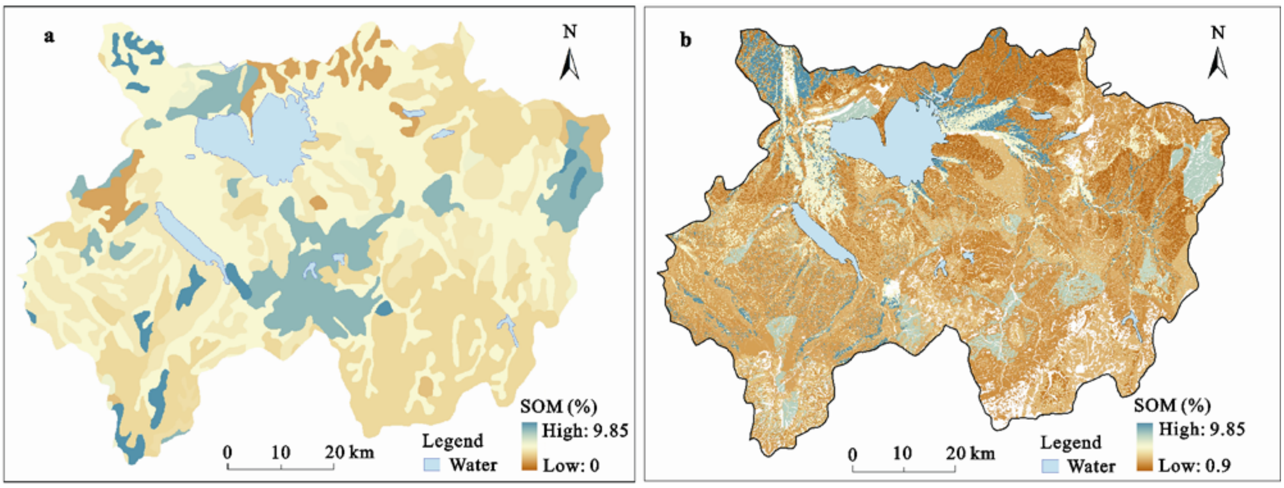

Fig. 7 Spatial distribution maps of soil organic matter (SOM) produced by conventional direct linking method (a) and representative samples method (b)

sample with the same soil type name, and then based on these representative samples extracted by the proposed method in this paper, the spatial distribution map of soil organic matter was produced by using digital soil mapping method of Liu *et al*. (2013). The spatial distribution of soil organic matter predicted based on representative samples is closely related to the distribution of the selected environment covariates, which accords to the soil-landscape model. In addition, the spatial variation of the soil organic matter is much gradual and there are no obvious boundaries of the soil distribution. This case study demonstrates the soil mapping method based on representative samples is better than conventional direct linking method, because the soil property map produced based on representative samples is higher quality on the spatial details level than that of soil map produced by using conventional direct linking method.

3.3 Accuracy evaluation of soil organic matter map

In order to evaluate the overall accuracy of the soil organic matter map produced based on representative samples, 55 individual samples were collected under regular sampling scheme with about 10 km spacing distance (Fig. 8). These regular samples can capture the spatial distribution characteristics of soil organic matter very well, which were sampled in October 2011.

Root Mean Square Error (RMSE) was calculated by comparing the observed values of the validation samples with the predicted values at these locations. RMSE of the soil organic map produced by conventional direct linking method is 1.61, while that of the soil organic matter map produced based on representative samples is 1.37. Therefore, the representative samples extracted by the method proposed in this paper can be used to produce soil property maps with higher accuracy.

4 Discussion

The most important step of this new soil mapping method based on representative samples is to extract the representative samples for each soil type polygon. The

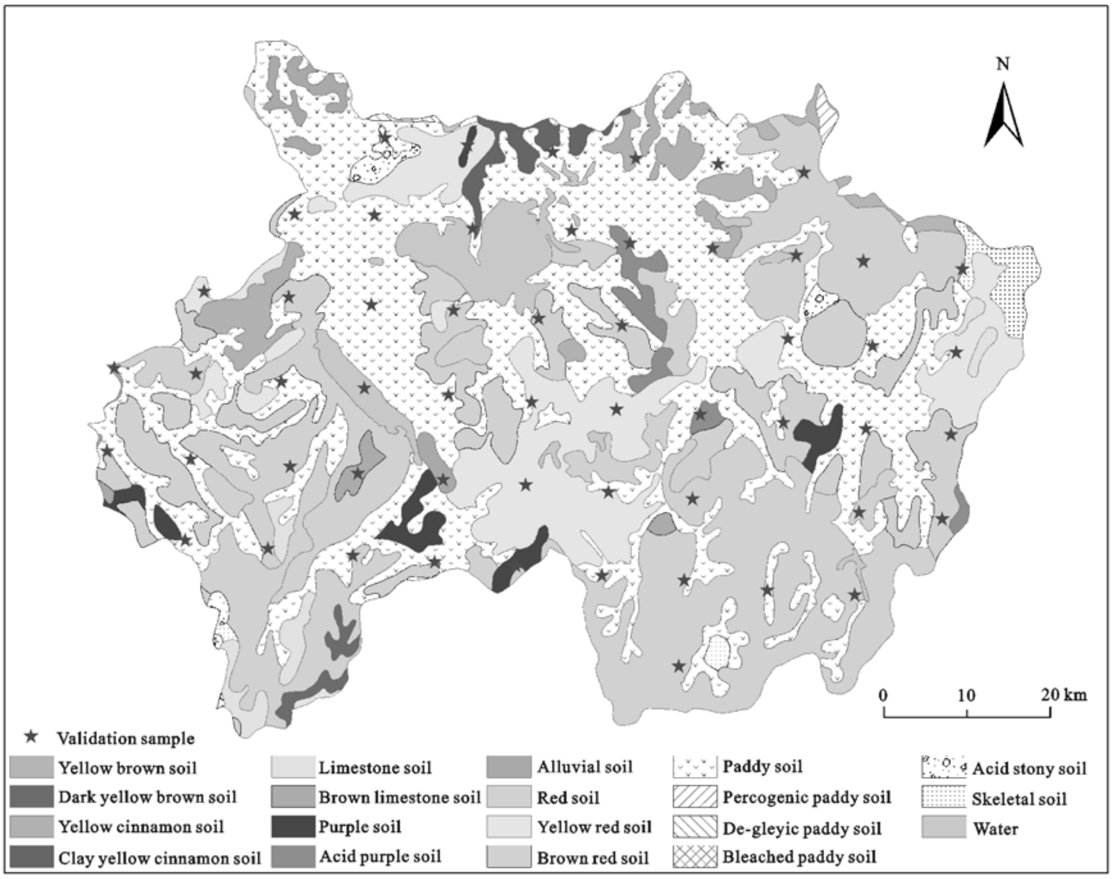

Fig. 8 Spatial distribution of regular validation samples with equal space about 10 km

representative sample of each polygon is defined as the location that can represent the largest area within the polygon, and the representing area of a representative sample is determined by calculating the soil-forming environment condition similarities between the representative sample and other locations. When the similarity is larger than the user-specified threshold, the representative sample is regarded to be representative of another location; otherwise, it can not represent the location. So, the smaller the threshold is, the larger the representing area of the representative sample is. In this paper, the similarity threshold is specified as 0.8 by a soil survey expert. In other words, when the similarity is larger than 0.8 (the maximum value of similarity is 1.0), the soil-forming environment condition on the two locations is considered to be enough similar and the soil property value could be predicted by the value on the other location. Because the similarity threshold is specified by the user, the value of the threshold may affect the location of the representative sample. In the future research, the sensitivity analysis will be done to test the variation of the locations of the representative samples.

The innovation of this paper is proposing a method to extract the representative sample of each polygon on the soil type map. Based on these representative samples, many existing digital soil mapping methods can be used to map soil property, such as regression model, Kriging interpolation, SoLIM, individual representativeness methods (Liu′s method), *etc*. In this paper, Liu′s method was chosen because it can map soil property using ad hoc sample set and can make full use of auxiliary soil formative environment factors to improve the accuracy of the soil property map. However, The shortcoming of Liu′ method is that the soil property map was not a complete map. Soil property values were predicted on the locations that can be represented by the representative samples, and ′No-data′ value was assigned to other locations that can not be represented by the representative samples. That is why, there are some white area within the study area in the soil organic matter map produced based on representative sample (Fig. 7b). The soil property values of these white areas were ′No-data′, because the similarities between these locations and the representative samples were smaller than the user-specified similarity threshold. In the future, how to design additional samples for completing the soil property map is a problem worthy of study.

The traversing algorithm was used to calculate the representing area of each location within a polygon in this paper, which was quite time consuming. For this study area with 5900 km^2, it takes about three weeks to extract the representative samples. Other optimal algorithms (such as simulated annealing algorithm, genetic algorithm) or parallel calculation method might be adopted to reduce time for computing in the future research.

For using this new soil property mapping method based on representative samples, appropriate environment covariates should be selected to characterize the soil-forming environment condition. This is because whether or not a location can represent another location is judged by the environment condition similarity between the two locations. Besides, the soil property values of the unvisited locations were also calculated based on the property values of representative samples weighted by the similarities of the unvisited location to the representative sample set. In this paper, meteorological (average annual precipitation and average annual temperature), parent material (geology type data), and topographic data (slope, curvature and TWI) were selected as covariates to characterize the soil-forming environment condition. In the future, some other covariates, such as land use, NDVI, LAI, should be used to describe the vegetation condition which also affects the process of soil formation and development.

5 Conclusions

Soil type maps at the scale of 1∶1 000 000 are used extensively to provide soil property spatial distribution information for soil erosion assessment and watershed management models in China. Soil property maps for such purposes are currently produced through linking the property values of typical samples to the polygons on the soil type map according to the same soil type name. Constrained by the small scale of the soil type maps at the scale of 1∶1 000 000 and the limited existing sparse typical samples in the Published National Soil Attribute Table, the soil property maps produced through conventional direct linking method usually suffer low accuracy as well as the lack of spatial details.

Without additional field work, this study presented an effective method to produce soil property map at a higher level of spatial details and accuracy. The data source for this new method is also the soil type map at small scale of 1∶1 000 000 and sparse existing typical samples. Unlike the conventional direct linking method, this alternative soil property mapping method first extracted the representative sample for each polygon on the soil type map. Then, link the property values of typical samples to the representative samples. The size of the representative samples is much larger than the size of the existing typical samples. Thus, based on these representative samples and auxiliary soil-formation environment factors, spatial detailed soil property map with higher accuracy could be produced.

Constrained by the polygon-based soil type map, the conventional direct linking method gives no information on spatial structure of soil property distribution within a polygon, and the variation of soil property only occurs at the common boundaries of two adjacent polygons. The new proposed method can overcome the above shortcomings, which can not only depict the detailed spatial distribution of soil property within a soil type polygon but also largely reduce the abrupt change of soil property at the boundaries of two adjacent polygons. The improvements on these two aspects make soil property spatial distribution on the map accord with the spatial variation of soil property in the field. In the study area, the RMSE of the soil organic matter map produced by conventional direct linking method is 1.61, and it is 1.37 for the soil organic matter map produced based on representative samples. This case study demonstrated that the representative samples extracted from each polygon on soil type map could be used to produce spatial detailed soil property map with higher accuracy.

Therefore, the soil property mapping method based on representative samples is an effective approach to produce soil property map for environment simulation models at a higher level of spatial details and accuracy.

Acknowledgements

The authors appreciate the Detailed Geographical Process Modeling Research Group of Institute of Geographic Sciences and Natural Resources Research, Chinese Academy of Sciences for their help in the field sampling.

References

Band L E, Moore I D, 1995. Scale: Landscape attributes and geographical information systems. *Hydrological Processes*, 9(3–4): 401–422. doi: 10.1002/hyp.3360090312

Brus D J, Bogaert P, Heuvelink G B M, 2008. Bayesian maximum entropy prediction of soil categories using a traditional soil map as soft information. *European Journal of Soil Science*, 59(2): 166–177. doi: 10.1111/j.1365-2389.2007.00981.x

Chen L J, Zhu A X, Qin C Z *et al*., 2012. Effectiveness assessment of soil erosion critical source areas for soil and water conservation. *Journal of Resources and Ecology*, 3(2): 138–143. doi: 10.5814/j.issn.1674-764x.2012.02.005

Goovaerts P, 1999. Geostatistics in soil science: State-of-the-art and perspectives. *Geoderma*, 89(1–2): 1–45. doi: 10.1016/S0016-7061(98)00078-0

Hudson B D, 1992. The soil survey as paradigm-based science. *Soil Science Society of America Journal*, 56(3): 836–841. doi: 10.2136/sssaj1992.03615995005600030027x

Jenny H, 1941. *Factors of Soil Formation—A System of Quantitative Pedology*. New York: Dover Publications, 10–20.

Kempen B, Brus D J, Heuvelink G B M *et al*., 2009. Updating the 1:50,000 Dutch soil map using legacy data: A multinomial logistic regression approach. *Geoderma*, 151(3–4): 311–326. doi: 10.1016/j.geoderma.2009.04.023

Li R K, Zhu A X, Song X F *et al*., 2012. Effects of spatial aggregation of soil spatial information on watershed hydrological modeling. *Hydrological Processes*, 26(9): 1390–1404. doi: 10.1002/hyp.8277

Liu Jing, Zhu Axing, Zhang Shujie *et al*., 2013. Mapping soil property over large area based on the individual representativeness of samples. *Acta Pedologica Sinica*, 50(1): 12–20. (in Chinese)

Liu Y B, Batelaan O, de Smedt F *et al*., 2005. Test of a distributed modeling approach to predict flood flows in the Karst Suoimuoi catchment in Vietnam. *Environmental Geology*, 48(7): 931–940. doi: 10.1007/s00254-005-0031-1

McBratney A B, Mendonca S M L, Minasny B, 2003. On digital soil mapping. *Geoderma*, 117(1–2): 3–52. doi: 10.1016/S0016-7061(03)00223-4

McBratney A B, Odeh I O A, Bishop T F A *et al*., 2000. An overview of pedometric techniques for use in soil survey. *Geoderma*, 97(3–4): 293–327. doi: 10.1016/S0016-7061(00)00043-4

Moriasi D N, Starks P J, 2010. Effects of the resolution of soil dataset and precipitation dataset on SWAT2005 stream-flow calibration parameters and simulation accuracy. *Journal of Soil and Water Conservation*, 65(2): 163–178. doi: 10.2489/jswc.65.2.63

Mukundan R, Radcliffe D E, Risse L M, 2010. Spatial resolution of soil data and channel erosion effects on SWAT model predictions of flow and sediment. *Journal of Soil and Water Conservation*, 65(2): 92–104. doi: 10.2489/jswc.65.2.92

Qin C Z, Zhu A X, Pei T *et al*., 2007. An adaptive approach to selecting a flow-partition exponent for a multiple-flow-direction algorithm. *International Journal of Geographical Information Science*, 21(4): 443–458. doi: 10.1080/1365881060

1073240

Qin C Z, Zhu A X, Qiu W L *et al*., 2012. Mapping soil organic matter in small low-relief catchments using fuzzy slop position information. *Geoderma*, 171–172: 64–74. doi: 10.1016/j.geoderma.2011.06.006

Shi X, Zhu A X, Burt J E *et al*., 2004. A case-based reasoning approach to fuzzy soil mapping. *Soil Science Society of America Journal*, 68(3): 885–894. doi: 10.2136/sssaj2004.8850

Shi Xuezheng, Yu Dongsheng, Gao Peng *et al*., 2007. Soil information system of China (SISChina) and its application. *Soils*, 39(3): 329–333. (in Chinese)

Sun Xiaolin, Zhao Yuguo, Qin Chengzhi *et al*., 2008. Effects of DEM resolution on multi-factor linear soil-landscape models and their application in predictive soil mapping. *Acta Pedologica Sinica*, 45(5): 971–977. (in Chinese)

Vitharana U W A, Saey T, Cockx L *et al*., 2008. Upgrading a 1/20,000 soil map with an apparent electrical conductivity survey. *Geoderma*, 148(1): 107–112. doi: 10.1016/j.geoderma.2008.09.013

Yang L, Jiao Y, Fahmy S *et al*., 2012. Updating conventional soil maps using digital soil mapping. *Soil Science Society of America Journal*, 75(3): 1044–1053. doi: 10.2136/sssaj2010.0002

Yang Lin, Fahmy Sherif, Jiao You *et al*., 2010. Updating conventional soil maps using knowledge on soil-environment relationships extracted from the maps. *Acta Pedologica Sinica*, 47(6): 1039–1047. (in Chinese)

Yu D S, Shi X Z, Wang H J *et al.*, 2007. Regional patterns of soil organic carbon stocks in China. *Journal of Environmental Management*, 85(3): 680–689. doi: 10.1016/j.jenvman.2006.09.020

Yu Wanli, 2012. *Soil Property Mapping on a National Scale Based on Sparse Grid Samples*. Beijing: Institute of Geographic Sciences and Natural Resources Research, Chinese Academy of Sciences, 34–39. (in Chinese)

Zhang Shujie, Zhu Axing, Liu Jing *et al*., 2012. Sample-based digital soil mapping methods and related sampling scheme. *Soils*, 44(6): 917–923. (in Chinese)

Zhang Yong, Shi Xuezheng, Yu Dongsheng *et al*., 2008. Effects of the linkage between spatial data and attribute data on estimates of soil organic carbon. *Advances in Earth Science*, 23(8): 840–847. (in Chinese)

Zhao Liang, Zhao Yuguo, Li Decheng *et al*., 2007. Digital soil mapping by extracting quantitative relationships between soil properties and terrain factors based on fuzzy set theory. *Acta Pedologica Sinica*, 44(6): 961–967. (in Chinese)

Zhao Y C, Shi X Z, Weindorf D C *et al*., 2006. Map scale effects on soil organic carbon stock estimation in North China. *Soil Science Society of America Journal*, 70(4): 1377–1386. doi: doi: 10.2136/sssaj2004.0165

Zhu A X, Band L E, Dutton B *et al*., 1996. Automated soil inference under fuzzy logic. *Ecological Modelling*, 90(2): 123–145. doi: 10.1016/0304-3800(95)00161-1

Zhu A X, 1997. A similarity model for representing soil spatial information. *Geoderma*, 77(2–4): 217–242. doi: 10.1016/S0016-7061(97)00023-2