Pedosphere ISSN 1002-0160/CN 32-1315/P 10.1016/S1002-0160(20)60016-9

SOLIM-IDW

Soil Property Mapping by Combining Spatial Distance Information into Soil Land Inference Model (SoLIM)

QIN Cheng-Zhi[1,2,3], AN Yi-Ming[1,2], LIANG Peng[1,2], ZHU A-Xing[1,3,4,5,*], YANG Lin[6]

[1]*State Key Laboratory of Resources and Environmental Information System, Institute of Geographic Sciences and Natural Resources Research, CAS, Beijing 100101 (China)*
[2]*College of Resources and Environment, University of Chinese Academy of Sciences, Beijing 100049 (China)*
[3]*Jiangsu Center for Collaborative Innovation in Geographical Information Resource Development and Application and School of Geography, Nanjing Normal University, Nanjing 210097 (China)*
[4]*Department of Geography, University of Wisconsin-Madison, Madison, WI 53706 (USA)*
[5]*Key Laboratory of Virtual Geographic Environment, Ministry of Education, Nanjing Normal University, Nanjing 210023 (China)*
[6]*School of Geographic and Oceanographic Sciences, Nanjing University, Nanjing 210093 (China)*

(Received 4 August, 2018; revised 8 January, 2019)

ABSTRACT

The Soil Land Inference Model (SoLIM) primarily proposed primarily by Zhu et al. (1997. *Soil Science Society of America Journal*, 61(2), 523–533) has been widely applied to digital soil mapping. Based on the assumption that soil property value at a location of interest will be more similar to that of a given soil sample when the environmental condition at the location of interest is more similar to that at the location from which the sample was taken, SoLIM estimates the soil property value of location of interest by using the soil property values of known samples weighted by the similarity between those samples and the location of interest in terms of an attribute domain of environmental conditions. However, current SoLIM procedure ignores information about the spatial distances between the location of interest and the soil sample locations. In this study, we propose a new method (SoLIM-IDW) which combines this spatial distance information into the SoLIM procedure to derive a soil property map. The proposed SoLIM-IDW method is based on an assumption that soil property value at a location of interest will be more similar to that of a known sample both when the environmental conditions are more similar and when the distance between the location of interest and the sample location is shorter. Our evaluation experiments on A-horizon soil organic matter mapping in two study areas with independent evaluation samples show that the proposed SoLIM-IDW method can get lower prediction errors than the original SoLIM method, multiple linear regression, geographically weighted regression, and regression-kriging with same modeling points. Future work mainly includes the determination of optimal power parameter values, or how to set the parameter properly under different application contexts.

Key Words: digital soil mapping, SoLIM, spatial distance, environmental similarity, soil property

Citation: Qin C-Z, An Y-M, Liang P, Zhu A-X, Yang L. 2019. Soil property mapping by combining spatial distance information into Soil Land Inference Model (SoLIM). *Pedosphere*. **???**(???): ???--???.

*Corresponding author. E-mail: axing@lreis.ac.cn.

INTRODUCTION

Digital soil mapping (or predictive soil mapping) is an efficient way to create predictive soil maps by developing a numerical model of the relationships between environmental covariates and soil in an area and then applying that model to a geographical database (McBratney *et al.*, 2003; Minasny and McBratney, 2016). Although the numerical model of these relationships is often developed in a statistical or geostatistical manner (such as regression and kriging) (Brus and de Gruijter, 1993; Fotheringham *et al*., 1996; Goovaerts, 1999; Heuvelink and Pebesma, 1999; McBratney *et al.*, 2003; Hengl *et al*., 2007; Lark, 2012), such a method requires a large number of soil samples which sufficiently represent the soil-environment relationship across the study area (Zhu *et al*., 2018). This means large budget requirement and highly labor-intensive field work. Meanwhile, such a method requires the represented soil-environment relationship to be stable across the entire study area (or, the stationarity assumption in geostatistics), which often cannot be met due to the complexity of soil distribution (Zhu *et al*., 2018).

The Soil Land Inference Model (SoLIM) method was originally proposed by Zhu *et al.* (1997) to overcome above mentioned disadvantages of statistical or geostatistical models for digital soil mapping (Zhu *et al*., 2018). SoLIM has been successfully applied to predictive soil mapping using environmental covariates related to soil property distribution and a few purposive samples or even ad-hoc soil samples with very limited representativeness of the soil-environment relationship in a study area (Zhu *et al.*, 2010; Qin *et al.*, 2010; Zhu *et al.*, 2015). Based on the assumption that soil property value at a location of interest will be more similar to that of a given soil sample when the environmental condition at the location of interest is more similar to that at the location from which the sample was taken, SoLIM estimates the soil property value of location of interest by using the soil property values of known samples weighted by the similarity between those samples and the location of interest in terms of an attribute domain of environmental conditions (such as slope gradient, topographic wetness index, and parent material).

$$V_i = \frac{\sum_{k=1}^{N}(S_{i,k} \times V_k)}{\sum_{k=1}^{N} S_{i,k}} \tag{1}$$

where V_i is the soil property value predicted at the location of interest i, V_k is the soil property value of the soil sample k, N is the count of soil samples, and $S_{i,k}$ is the similarity in terms of environmental conditions between i and k, which is normally computed to be the minimum among the individual similarities in terms of individual environmental attributes related to the spatial variation of soil (Zhu and Band, 1994). The estimation uncertainty at each location can also be provided by SoLIM (Zhu *et al.*, 1997), which is useful for directing the subsequent application of the resultant soil property map (e.g., Li *et al*., 2016; Zhang *et al.*, 2016).

However, information regarding the spatial distances between the location of interest and the soil sample locations is ignored in the current SoLIM inference method, which estimates soil property values at each interest location based on the attribute domain of environmental covariates. This spatial distance is often important to consider in spatial analysis and mapping of geographical phenomena, and Tobler's first law of geography has been widely accepted to highlight this point (Tobler, 1970; De Smith *et al.*, 2007). Could the spatial distances between locations of interest and soil sample locations be considered in SoLIM inference to improve the accuracy of the resultant soil property map? In this paper, we study this issue and explore potential improvements which can be made by taking into consideration the spatial distances between locations of interest and soil sample locations in SoLIM inference.

MATERIALS AND METHODS

Basic idea

In order to revise the current SoLIM method so that it considers the spatial distances between each location of interest and soil sample locations during predictive soil property mapping, we first propose the following assumption: soil property value at a location of interest will be more similar to that of a known sample both when the environmental conditions are more similar and when the distance between the location of interest and the sample location is shorter. This assumption is a simple and reasonable extension of the assumption used in current SoLIM, combined with Tobler's first law of geography.

Note that one classic method of spatial estimation at unvisited locations from a set of value-known locations is the inverse distance weighting (IDW), which is based purely on Tobler's first law of geography and has no additional requirements regarding either the spatial distribution or count of sample locations. This way of considering the spatial distance between locations of interest and sample locations could be combined with the current SoLIM inference. Such a combination can result in a revised SoLIM (so-called SoLIM-IDW) which is based on the assumption proposed above and can take into consideration the spatial distances between locations of interest and soil sample locations, thus eliminating the problem of ignoring spatial distance information in the current SoLIM inference method.

Design of the SoLIM-IDW method

In accordance with the basic idea above, the SoLIM-IDW method proposed in this paper is designed to revise current SoLIM inference to use the following equation:

$$V_i = \frac{\sum_{k=1}^{N}(D_{i,k} \times S_{i,k} \times V_k)}{\sum_{k=1}^{N}(D_{i,k} \times S_{i,k})}\ ,\ D_{i,k} = 1/\left(d_{i,k}\right)^r \quad (2)$$

where $d_{i,k}$ is the distance between the location of interest i and the soil sample location k ($k = 1..N$), $D_{i,k}$ is the weight function based on $d_{i,k}$ (as what is normally used in IDW), and r is the power parameter which is a positive real number selected according to the principle of inverse distance weighting. $S_{i,k}$ (i.e., the similarity in terms of environmental conditions between i and k) is the same as that used in current SoLIM inference. Thus, the original SoLIM inference function (Eq. (1)) is the special case of Eq. (2) where $r = 0$.

Study areas and data

The proposed method was evaluated by application to two study areas (Fig. 1). The first is Heshan farm at a watershed scale (about 60 km^2), and the second is Xuancheng county at a regional scale (about 5900 km^2).

Fig.1（见文末）

Fig. 1 Maps of study areas. a) the Heshan farm case; b) the Xuanchen county case.

Heshan case

The Heshan farm in the Heilongjiang province of northeastern China has with very low relief. This study area (about 60 km^2) has a total relief of about 100 m and an average slope gradient of about 2°. The parent materials are mainly silt loam loess and fluvial deposits in the valley. In the Heshan farm, main types of soils at subgroup level in the Chinese soil taxonomy system (Chinese Soil Taxonomy Research Group 2001) include Mollic Bori-Udic Cambosols, Typic Hapli-Udic Isohumosols, Typic Bori-Udic Cambosols, Lithic Udi-Orthic Primosols, Pachic Stagni-Udic Isohumosols, and Fibric Histic-Typic Haplic Stagnic Gleyosols (Zhu *et al*., 2010). Soybeans and wheat are the main agricultural products in this area, which has been cultivated for more than 40 years.

The proposed method was used for digital soil mapping of A-horizon soil organic matter (SOM) content (%) in this area. Four environmental variables (i.e., slope gradient, profile curvature, horizontal curvature, and topographic wetness index) were adopted and calculated based on a DEM with a resolution of 10 m, as was done in the application of SoLIM in same area (Zhu *et al.*, 2010).

In this case, a total of 39 points were used as modeling samples, including 29 points from an integrative hierarchical stepwise sampling method (Yang *et al.*, 2013) and 10 points from subjective sampling at summit, steeper slope, and valley locations (Zhu *et al.*, 2010; Yang *et al.*, 2013). Another 44 points from a regular sampling grid (1100 m * 740 m) were used as independent evaluation samples (Fig. 1a). SOM content of each soil sample was measured by the Walkley–Black wet oxidation method (Nelson and Sommers, 1982).

Xuancheng case

The Xuancheng county in China's Anhui province is about 5900 km^2 and has a total relief of about 1058 m and an average slope gradient of about 4°. The northwestern part of this area is mainly low relief, while other parts are mainly mountainous. The parent materials in this area are complex, including Quaternary clay-silt-gravel, sandstone, shale, conglomerate, pyroclastic rocks, limestone, granite and granodiorite (Yang *et al.*, 2017). There are five main soil orders in this area, i.e., Semi-hydromorphic soils, Primitive soils, Anthropogenic soils, Eluvial soils, and Ferro-allitic soils (National Soil Survey Office, 1992).

The proposed method was used for digital soil mapping of A-horizon soil organic carbon (SOC) (g kg^{-1}) in this area. The environmental variables used in this case include slope gradient, profile curvature, horizontal curvature, topographic wetness index, annual average precipitation, annual average temperature, and parent material (Yang *et al.*, 2017). All environmental variables have a resolution of 90 m, where topographic attributes were calculated based on the SRTM DEM.

In this case, 59 points from a multi-grade representative sampling method (Yang *et al.*, 2017) were used as modeling samples, while another 58 points from a regular sampling grid of 10 km * 10 km were used as independent evaluation samples (Fig. 1b). SOC of soil samples were measured by the dichromate oxidation method (external heat applied) (Nelson and Sommers, 1982; Zeng *et al*., 2016).

Evaluation experiment

The evaluation in this study was focused on the comparison between SoLIM and SoLIM-IDW. In order to facilitate comparison to SoLIM (i.e., SoLIM-IDW with $r = 0$), the SoLIM-IDW method proposed in this paper was evaluated in two different ways across several different *r* values (i.e., 0.25, 0.5, 0.75, 1, 1.5, 2, 2.5, 3). The first evaluation utilized quantitative statistics of the prediction errors based on independent evaluation samples, including the root mean squared error (RMSE), mean error (ME), and mean absolute error (MAE). The second was a qualitative comparison between the map resulting from the SoLIM-IDW method with the lowest RMSE and the map resulting from SoLIM. Note that currently there is no theoretical justification on the highest *r* value. With a very high *r* value, only those points being extremely close to the location of interest will influence the prediction. In this study we set the highest *r* value under test to be 3, which was proved to be large enough by the following experimental results.

While Zhu *et al*. (2015)'s evaluation in the Heshan farm has shown that the SoLIM method performed better than multiple linear regression (MLR) under different sample scenarios, in this study the performance of MLR using the same modeling samples as those for SoLIM-IDW was also compared with that of SoLIM-IDW based on independent evaluation samples. There are some methods widely used for digital soil mapping, which consider both attribute distance of environmental covariates and spatial distances, such as geographically weighted regression (GWR; Brunsdon *et al*., 1996; Fotheringham *et al*., 1996) and regression-kriging (RK; Odeh *et al*., 1995; Hengl *et al*., 2007). Although the count of modeling samples in the study area is too limited to well fit the modeling of GWR and RK (Hengl *et al*., 2007), in this study GWR and RK were also compared with SoLIM-IDW based on the same modeling samples and the same independent evaluation samples. MLR, GWR, and RK were conducted by packages in *R*. Furthermore, a rudimentary prediction which assumed the mean soil property value of modeling samples (i.e., 5.299% for Heshan case, and 11.315 g kg^{-1} for Xuancheng case) to be the predicted value for each independent evaluation sample, or "MeanValuePredict" for short, was also compared with the tested methods.

Note that in the Heshan case, the A-horizon SOM values of 39 modeling samples show an obvious right-tailed distribution (i.e., min = 2.56%, max = 32.64%, mean = 5.30%, and standard deviation = 4.72%) due to a modeling sample with very high SOM value. This situation will impact the quality of the MLR modeled directly with original values of this dataset of modeling samples. To relieve such adverse impact, it is necessary for some kind of numerical transformations before use the SOM values for fitting the MLR, such as taking the square root (Wang *et al*., 2013), the logarithm (Bostan *et al*., 2012; Song *et al*., 2016), the reciprocal, or the Box-Cox transformation. By try-and-error, the reciprocal of SOM values of modeling samples in the Heshan obeys the normal distribution, by the Shapito-Wilk normality test (P = 0.41, larger than 0.05) which is suitable for this small sample test. Thus the reciprocal transformation of SOM values of modeling samples was used before building MLR, GWR, and RK with the stepwise variable selection in the Heshan case. For modeling GWR in this case, the weighting function was calibrated based on the Gaussian function, while the bandwidth was optimized to be 2195.4 m in this case based on the Akaike information criterion (AIC) (Burnham and Anderson, 2004).

For building MLR in the Xuancheng case, the parent material was treated as eight dummy variables, each for one parent material type in this area, with a value of 1 or 0 (i.e., belonging to the corresponding parent material type, or not). In the Xuancheng case, the stepwise variable selection cannot effectively remove those highly-correlated environmental variables. Thus the principal component analysis (PCA) was adopted to relieve the collinearity issue among environmental variables, during building MLR, GWR, and RK in the Xuancheng case. Because the contribution rate to total variation from first principal component reached over 99%, the first principal component and dummy variables were used to build MLR. For modeling GWR in this case, the weighting function was still calibrated based on the Gaussian function, while the bandwidth was optimized to be 39171.1 m based on AIC. During RK modeling in the Xuancheng case, the variogram function fitted with the 59 modeling samples was with a range of 10557.9 and a partial sill of 2.289 based on a spherical model and was still not stable. This situation is similar to that for RK in the Heshan case, although the sample number in the Xuancheng case is larger than the minimum number of modeling samples (i.e., 50) recommended by Hengl *et al*. (2007).

Due to the fact that the modeling samples were collected from purposive sampling and with a limited number in both cases, we did not apply the maps from the MLR, GWR, and RK to qualitative comparison with the map from the SoLIM-IDW method.

RESULTS AND DISCUSSION

Heshan case

Table I shows that in the Heshan case, the SoLIM-IDW method obtained lower error, in terms of RMSE, than the SoLIM method, GWR, and MeanValuePredict when $r = 0.25{\sim}1$. As r values increased, the performance of SoLIM-IDW first increased then decreased, with the lowest RMSE being produced by the SoLIM-IDW when $r = 0.75$. The SoLIM-IDW with $r = 0.5$ produced the second lowest RMSE, which is very close to that from the SoLIM-IDW with $r = 0.75$. RMSE values from MLR and RK were very similar and were even larger than RMSE from MeanValuePredict in this case study.

TABLE I

Quantitative evaluation of the SoLIM-IDW method based on independent sample set.

Method		Heshan case: A-horizon SOM			Xuancheng case: A-horizon SOC		
		RMSE	ME	MAE	RMSE	ME	MAE
		%	%	%	g kg^{-1}	g kg^{-1}	g kg^{-1}
SoLIM (i.e., SoLIM-IDW with *r*=0)		1.22	-0.74	1.00	3.82	-0.40	3.19
SoLIM-IDW	*r* = 0.25	1.15	-0.68	0.94	3.80	-0.45	3.16
	r = 0.5	1.11	-0.64	0.90	3.81	-0.50	3.14
	r = 0.75	1.10	-0.61	0.88	3.86	-0.55	3.15
	r = 1	1.17	-0.59	0.90	3.93	-0.58	3.19
	r = 1.5	1.66	-0.63	1.04	4.11	-0.59	3.26
	r = 2	2.44	-0.74	1.25	4.25	-0.56	3.32
	r = 2.5	3.13	-0.85	1.43	4.35	-0.51	3.41
	r = 3	3.60	-0.93	1.57	4.41	-0.45	3.47
MLR		2.50	-0.45	1.31	4.02	-0.70	3.37
GWR		1.24	0.11	0.83	4.02	-0.57	3.32
RK		2.51	-0.45	1.31	3.99	-0.69	3.38
MeanValuePredict		1.36	-0.95	1.16	4.31	-0.18	3.43

Fig. 2 shows that the map of A-horizon SOM (%) estimated by the SoLIM-IDW with $r = 0.75$ in the Heshan case has a spatial pattern which is similar to that estimated by SoLIM. The range of predicted values from SoLIM-IDW with $r = 0.75$ in the Heshan case is [2.97%, 29.38%], which is much wider than that generated by SoLIM (i.e., [4.80%, 6.16%]) and much closer to that of the actual soil samples (i.e., [2.25%, 32.64%]). The main differences between the spatial patterns shown in the maps resulting from the two methods is that there is a “bull’s eye” pattern in the map generated by SoLIM-IDW (i.e., obvious higher SOM prediction near the high-value modeling sample located in channel). This phenomenon occurs often and is a consequence of the characteristics of IDW. In this case, areas in the flat channel have high soil moisture and rich humus, and thus are reasonable areas for higher SOM prediction.

Fig.2（见文末）

Fig. 2 Map of the A-horizon SOM (%) in the Heshan case as estimated by a) SoLIM-IDW with *r* = 0.75, and b) SoLIM. (While the prediction value range from SoLIM in this case is 4.80% -- 6.16%, the legend used for the result map from SoLIM is revised to be same as that of SoLIM-IDW for consistency of comparison.)

Xuancheng case

As shown in Table I, the SoLIM-IDW method with $r = 0.25{\sim}0.5$ produced lower RMSE than SoLIM method in the Xuancheng case. The SoLIM-IDW method with $r = 0.25{\sim}1$ as well as the SoLIM method performed better than RK, GWR, and MLR, while MeanValuePredict performed worst. SoLIM-IDW produced the lowest RMSE at $r = 0.25$, which is very close to the second lowest RMSE produced by the SoLIM-IDW with $r = 0.5$. Its performance decreased, meaning that the RMSE increased, as r values increased from there. In both case studies, the SoLIM-IDW method produced lower RMSE than the SoLIM method for at least some of the r values tested, though the SoLIM-IDW method produced the lowest error under a different value of r in each case. Thus, the proposed SoLIM-IDW method can improve on the original SoLIM by taking into consideration the spatial distances between locations of interest and soil sample locations.

Fig. 3 shows that the map of A-horizon SOC (g kg^{-1}) estimated by the SoLIM-IDW method with $r = 0.25$ in the Xuancheng case has a spatial pattern similar to that estimated by SoLIM, which is a phenomenon also seen in the Heshan case. There are crisp blocks in the maps of A-horizon SOC generated by both SoLIM-IDW and SoLIM. This is due to the effect of a nominal (categorical) environmental variable (i.e., parent material) in the inference. The range of values predicted by SoLIM-IDW with $r = 0.25$ in the Xuancheng case is 5.81--21.91, which is wider than that predicted by SoLIM (i.e., [7.85, 20.68]) and closer to that of the actual soil samples (i.e., 2.54--27.23). Much as with the Heshan case study, there are "bull's eyes" in the result map generated by SoLIM-IDW, which is due to the characteristics of IDW and the local effect of modeling samples.

Fig.3（见文末）

Fig. 3 Map of the A-horizon SOC (g kg^{-1}) in the Xuancheng case as estimated by a) SoLIM-IDW with $r = 0.25$, and a) SoLIM. (While the prediction value range from SoLIM in this case is 7.85--20.68, the legend used for the result map from SoLIM is revised to be same as that of SoLIM-IDW for consistency of comparison.)

Discussion

Above results of two case studies showed that the SoLIM-IDW with $r = 0.5$ consistently produced a RMSE being very close to the lowest RMSE in both case studies. Thus $r = 0.5$ could be used with the proposed SoLIM-IDW for now, before further studies on determining the optimal r values, or how to set the parameter r properly under different application contexts (such as study area characteristics, and the spatial resolution of soil mapping) in an adaptive manner.

Note that soils can change quickly in a very short distance due to the subtle change of environmental conditions, which cannot be captured by the environmental dataset at the resolution used with the digital soil mapping. For such change of soils, the proposed method and other existing digital soil mapping methods cannot work.

CONCLUSIONS

Current SoLIM inference utilizes an attribute domain of environmental covariates but ignores the spatial distances between the location of interest and the soil samples. In this paper, we propose a new method, SoLIM-IDW, which takes that spatial distance into consideration during the SoLIM-based predictive soil property mapping process. The evaluation experiments for each of two case areas show that the proposed SoLIM-IDW method performed better than the original SoLIM, MLR, GWR, and RK, while the error produced by SoLIM-IDW was minimized under different r values in each of the two cases. Currently the power parameter $r = 0.5$ is suggested based on the results from two case studies, under which the SoLIM-IDW consistently produced a RMSE being very close to the lowest. Future work mainly includes the determination of optimal power parameter (r) values under different application contexts in an adaptive manner.

ACKNOWLEDGEMENT

This study was funded by the National Natural Science Foundation of China (No. 41431177, 41422109, 41871300), National Basic Research Program of China (No. 2015CB954102), PAPD, and Outstanding Innovation Team in Colleges and Universities in Jiangsu Province. We thank the support from the Innovation Project of LREIS (No. O88RA20CYA). Supports to A-Xing Zhu through the Vilas Associate

Award, the Hammel Faculty Fellow Award, and the Manasse Chair Professorship from the University of Wisconsin-Madison are greatly appreciated.

REFERENCES

Bostan P A, Heuvelink G B M and Akyurek S Z. 2012. Comparison of regression and kriging techniques for mapping the average annual precipitation of Turkey. *Int J Appl Earth Obs.* **19**: 115–126.

Brunsdon C, Fotheringham A S and Charlton M E. 1996. Geographically weighted regression: a method for exploring spatial nonstationarity. *Geogr Anal.* **28**: 281–298.

Brus D J and de Gruijter J J. 1993. Design-based versus model-based estimates of spatial means: theory and application in environmental soil science. *Environmetrics*. **4**: 123–152.

Burnham K P, Anderson D R. 2004. Multimodel inference: understanding AIC and BIC in model selection. *Sociol Method Res.* **33**: 261–304.

Chinese Soil Taxonomy Research Group, 2001. *Keys to Chinese Soil Taxonomy, 3rd Edition*. University of Science and Technology of China Press, Hefei.

De Smith M J, Goodchild M F and Longley P. 2007. *Geospatial Analysis: A Comprehensive Guide to Principles, Techniques and Software Tools*. Troubador Publishing Ltd.

Fotheringham A S, Charlton M and Brunsdon C.1996. The geography of parameter space: An investigation of spatial non-stationarity. *Int J Geogr Inf Sci.* **10**: 605–627.

Goovaerts P. 1999. Geostatistics in soil science: state-of-the-art and perspectives. *Geoderma*. **89**: 1–45.

Hengl T, Heuvelink G B M and Rossiter D G. 2007. About regression-kriging: from equations to case studies. *Comput Geosci.* **33**: 1301–1315.

Heuvelink G B M and Pebesma E J. 1999. Spatial aggregation and soil process modelling. *Geoderma*. **89**: 47–65.

Lark R M. 2012. Towards soil geostatistics. *Spat Stat.* **1**: 92–99.

Li Y, Zhu A-X, Shi Z, Liu J and Du F. 2016. Supplemental sampling for digital soil mapping based on prediction uncertainty from both the feature domain and the spatial domain. *Geoderma*. **284**: 73–84.

McBratney A B, Mendonça Santos M L and Minasny B. 2003. On digital soil mapping. *Geoderma*. **117**: 3–52.

Minasny B and McBratney A B. 2016. Digital soil mapping: a brief history and some lessons. *Geoderma*. **264**: 301–311.

National Soil Survey Office. 1992. *Soil Census and Technology of China* (in Chinese). Agricultural Press, Beijing.

Nelson D W and Sommers L E. 1982. Total carbon, organic carbon, and organic matter. In: Page A L, editor, *Methods of Soil Analysis, Part 2. Chemical and Microbiological Properties. Agronomy Monograph No. 9. 2nd Ed*. ASA and SSSA, Madison, WI. p. 539–579. doi:10.2134/agronmonogr9.2.2ed.c29.

Odeh I, McBratney A and Chittleborough D. 1995. Further results on prediction of soil properties from terrain attributes: heterotopic cokriging and regression kriging. *Geoderma*. **67**: 215–226.

Qin C-Z, Zhu A-X, Qiu W-L, Lu Y-J, Li B-L and Pei T. 2012. Mapping soil organic matter in small low-relief catchments using fuzzy slope position information. *Geoderma*. **171-172**: 64–74.

Song X D, Brus D J, Liu F, Li D C, Zhao Y G, Yang J L and Zhang G L. 2016. Mapping soil organic carbon content by geographically weighted regression: A case study in the Heihe River Basin, China. *Geoderma*. **261**: 11–22.

Tobler W. 1970. A computer movie simulating urban growth in the Detroit region. *Econ Geogr.* **46**(Supplement): 234–240.

Wang K, Zhang C and Li W. 2013. Predictive mapping of soil total nitrogen at a regional scale: A comparison between geographically weighted regression and cokriging. *Appl Geogr*. **42**:73–85.

Yang L, Zhu A-X, Qi F, Qin C-Z, Li B-L and Pei T. 2013. An integrative hierarchical stepwise sampling strategy and its application in digital soil mapping. *Int J Geogr Inf Sci.* **27**: 1–23.

Yang L, Zhu A-X, Zhao Y, Li D, Zhang G, Zhang S and Band L E. 2017. Regional soil mapping using multi-grade representative sampling and a fuzzy membership-based mapping approach. *Pedosphere*. **27**: 344–357.

Zeng C, Yang L, Zhu A-X, Rossiter D G, Liu J, Liu J-Z, Qin C and Wang D. 2016. Mapping soil organic matter concentration at different scales using a mixed geographically weighted regression method. *Geoderma*. **281**: 69–82.

Zhang S-J, Zhu A-X, Liu J, Yang L, Qin C-Z and An Y-M. 2016. A heuristic uncertainty directed field sampling design for digital soil mapping. *Geoderma*. **267**: 123–136.

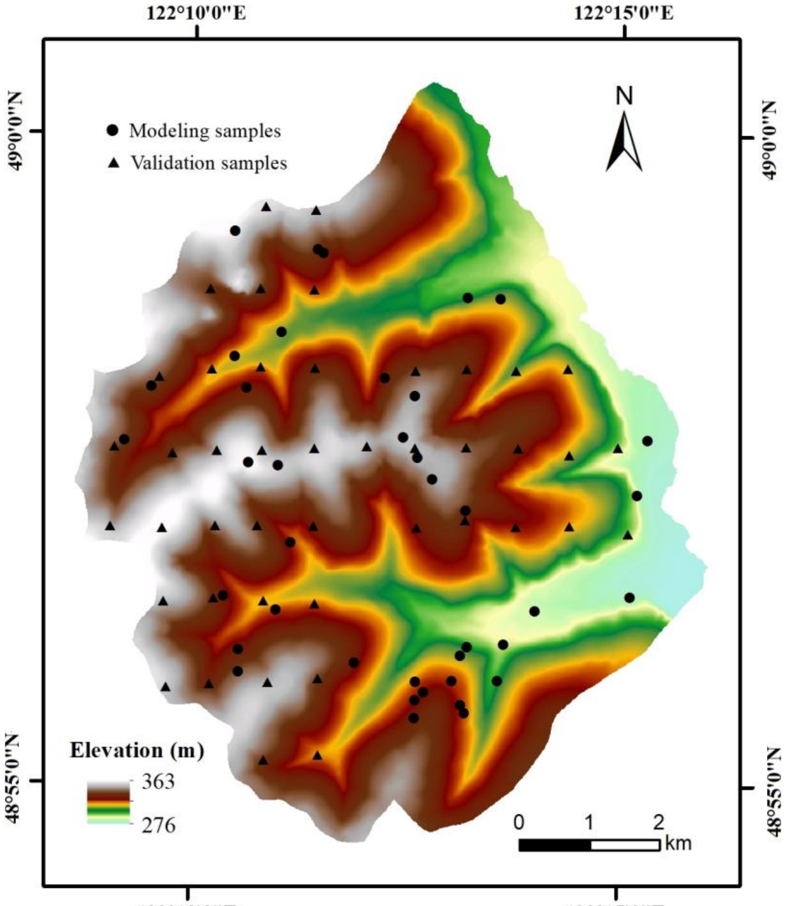
Zhu A-X and Band L E. 1994. A knowledge-based approach to data integration for soil mapping. *Can J Remote Sens*. **20**: 408–418.

Zhu A-X, Band L, Vertessy R and Dutton B. 1997. Derivation of soil properties using a soil land inference model (SoLIM). *Soil Sci Soc Am J*. **61**: 523–533.

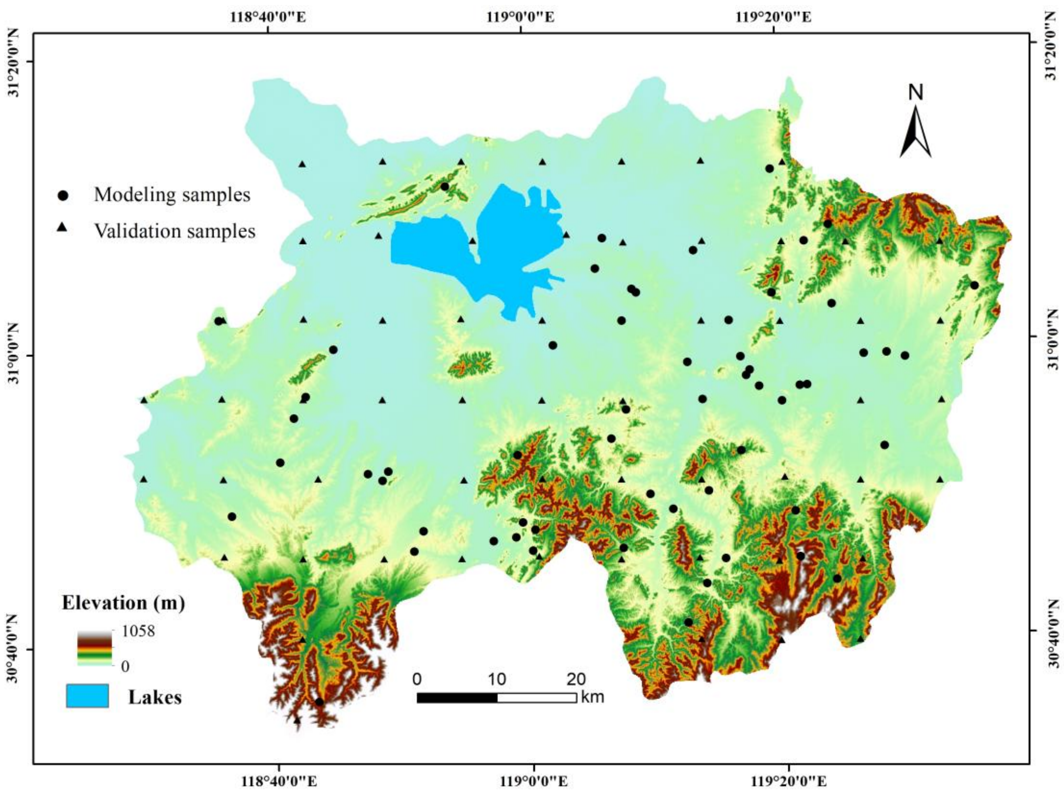
Zhu A-X, Yang L, Li B-L, Qin C-Z, Pei T and Liu B-Y. 2010. Construction of membership functions for predictive soil mapping under fuzzy logic. *Geoderma*. **155**: 164–174.

Zhu A-X, Liu J, Du F, Zhang S-J, Qin C-Z, Burt J, Behrens T and Scholten T. 2015. Predictive soil mapping with limited sample data. *Eur J Soil Sci*. **66**: 535–547.

Zhu A-X, Lv G, Liu J, Qin C-Z and Zhou C. 2018. Spatial prediction based on Third Law of Geography. *Ann GIS*. **24**: 225–240.

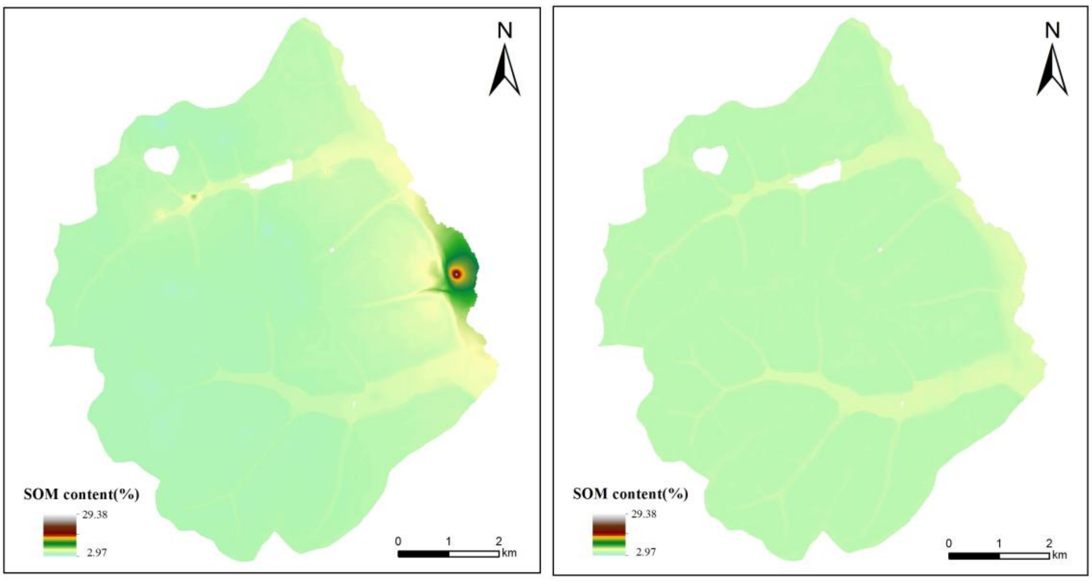


a)



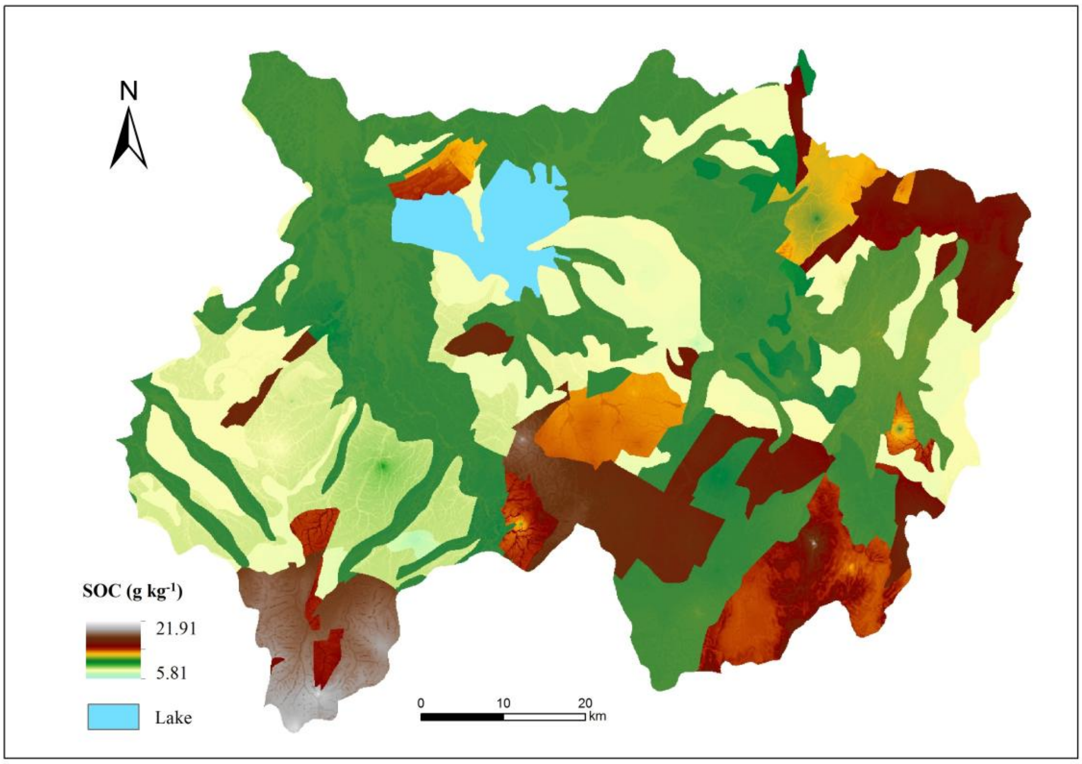
b)

Fig. 1 Maps of study areas. a) the Heshan farm case; b) the Xuanchen county case.

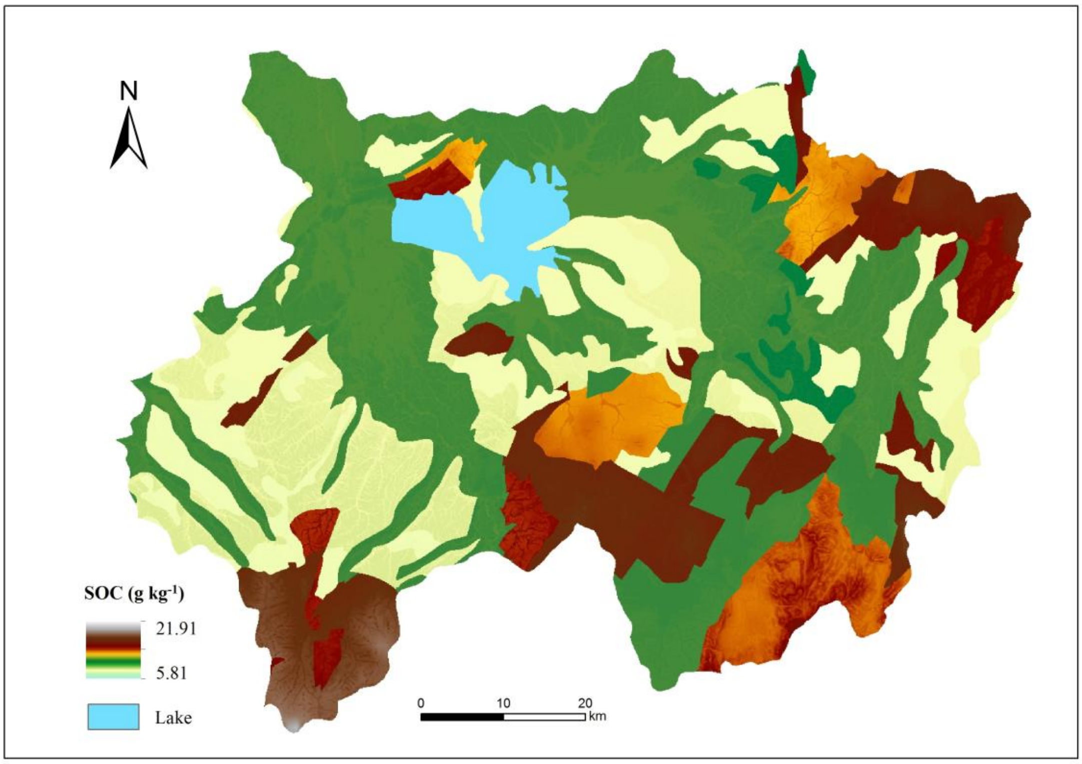


a) b)

Fig. 2 Map of the A-horizon SOM (%) in the Heshan case as estimated by a) SoLIM-IDW with r = 0.75, and b) SoLIM. (While the prediction value range from SoLIM in this case is 4.80% -- 6.16%, the legend used for the result map from SoLIM is revised to be same as that of SoLIM-IDW for consistency of comparison.)



a)



b)

Fig. 3 Map of the A-horizon SOC (g kg^{-1}) in the Xuancheng case as estimated by a) SoLIM-IDW with r = 0.25, and a) SoLIM. (While the prediction value range from SoLIM in this case is 7.85--20.68, the legend used for the result map from SoLIM is revised to be same as that of SoLIM-IDW for consistency of comparison.)