

Soil Lead Spatial Prediction Using a Fuzzy Soil-landscape Model

Qiang Wang^a, Lili Zhang^b, Youhua Ma^{c*}, Wenhui Yue^d, Runhe Shi^d

^aSchool of Resources & Environment, Anhui Agricultural University, Hefei 230036, China

^bAnhui vocational college of Grain Engineering, Hefei 230011, China

^cNew Rural Development Research Institute, Anhui Agricultural University, Hefei 230036, China

^dKey Laboratory of Geographic Information Science Ministry Of Education, Shanghai 200241, China
28104@ahau.edu.cn

Soil lead pollution does great harm to the environment. The relation between soil lead and environmental covariates, however, is complicated. A fuzzy logic approach with expert knowledge has proven to be successful in mapping spatial variation of soil lead. The reasons are the limited availability of expert knowledge and the ignorance of soil type and land use of the non-soil area when predicting soil lead in current fuzzy clustering methodology. This paper incorporates the soil type map, the land use map and the DEM data to construct a soil-landscape model for soil lead prediction. It compares a fuzzy C-means classifier that includes expert judgment with conditional regressive model. Prediction efficiency was evaluated in the Three Gorges area of China using the root mean square error (RMSE) and the agreement coefficient (AC) of predictions at validation points. The result indicates that the soil-landscape model constructed by the fuzzy membership functions with fuzzy c-means method and the conventional soil map were able to produce good quality soil lead spatial information.

1. Introduction

Soil lead pollution does great harm to the environment (Chen et al., 2015). Conventional surveying methods of soil lead require a large budget to be devoted to soil surveyors, a large number of samples and a long time, which has been a limitation to physically-based and geostatistical methods (Grunwald, 2016). The soil-landscape model overcomes those disadvantages (Ziadat et al., 2015). The model originates from soil scientists, who regarded the soil body as a product of several soil forming factors (Jafari et al., 2014). These factors consider climate, time, organisms, topography, and parent materials (Pinto, et al., 2016).

The sustainable utilization of soil is closely related to the healthy and stable development of modern society, and plays an important role in the development of modern civilization (Kumar and Singh, 2016). As far as ecological environment is concerned, soil is not only the basic component of ecosystem, but also an important component of natural environment factors (Ließ et al., 2016). As for human activities, soil is the basic data of agricultural production activities, and is also a precious natural resource for human survival (Desai et al., 2015). In order to fully analyze the spatial distribution of soil resources and its relationship with the surrounding environment, soil science came into being. In the late nineteenth century, Russian natural geographer and soil scientist proposed soil genesis in the study of soil formation and laid the foundation for the development of modern soil science (Siqueira et al., 2015). Soil classification, as a soil information carrier, plays a very important role in the study of soil science (Stielstra et al., 2015). It not only determines the type of soil property and the type of environmental data involved in the study, but also determines the accuracy and accuracy of soil type discrimination (Van et al., 2016). Because of the change of application demand and technical means, the application fields related to soil have strict requirements on the accuracy and accuracy of soil information (Hosseini et al., 2015). With the development of the soil, relevant disciplines and integration theory, a new method of innovation and advanced technology has been introduced to the study of soil science, which promotes the application of new technologies and new methods in soil information acquisition (Herman and Braun, 2016). These developments have deepened the study of soil, while the ways and expressions of soil information

acquisition are changing gradually with the change of social demands (Bowles et al., 2015). The traditional soil survey method to obtain the soil information by empirical and descriptive strong gradually developed into a wide use of modern science and technology (Yang et al., 2015), such as soil remote sensing and geographic information system (GIS, soil spatial information processing technology (Li et al., 2015). These new technologies will transform soil analysis from manual investigations to standardization, mechanization, and intelligence, and establish a quantitative soil landscape model (Nair, et al., 2016).

A reasonable sampling layout is the basis for accurate acquisition of representative soil landscape models. In order to make the soil sampling point more representative, the researchers optimized the soil sampling space layout. By reducing the error from the data source, the spatial layout of the final soil properties is more reflective of the actual situation. In the process of obtaining global soil information by soil information of samples, the establishment of soil landscape model is one of the most important links in soil science research. In China, the second national soil survey map of soil type included the expert knowledge from the ground investigation (Jiang et al., 2014). For using the expert knowledge, we consider the fuzzy logic approach with expert knowledge and compare it with the conditionally regressive model. The aim of this paper is to present a simple method to predict soil lead by constructing a membership function based on fuzzy c-means that may be employed for future environmental management and regularization. These are obtained after classifying the conventional soil map using expert knowledge and a land use map. This method will be compared with the linear regressive model approach.

2. Materials and methods

2.1 Study area and data source

The study area is the Tongling County in the Anhui province of China (Figure 1). A land use map was prepared by visual interpretation of SPOT images from 2009 based on detailed field knowledge. The DEM is generated from the contour lines with a GDEM (Global Digital Elevation Model) derived from interpretation of the Advanced Space borne Thermal Emission and Reflection Radiometer (ASTER) image with a 30m resolution. ELE (elevation), SL (slope), PR (profile curvature), PC (plan curvature), MC (mean curvature), AS (aspect), and FAA (flow accumulation area) were derived from the two DEMs using standard commands in the Arc/Info GRID module. The TWI (terrain wetness index) for each pixel was calculated as $TWI = \ln(A/\tan(SL))$ by the D-8 method. Soil lead field survey work was conducted in July 2013. Measurements were made following the selection of sample locations that the similarity value great than 0.85, and the inductively coupled plasma-mass spectrometry (ICP-MS) was performed to determine the Lead concentration. In total, 142 soil samples were retrieved and their locations are shown in figure 1.

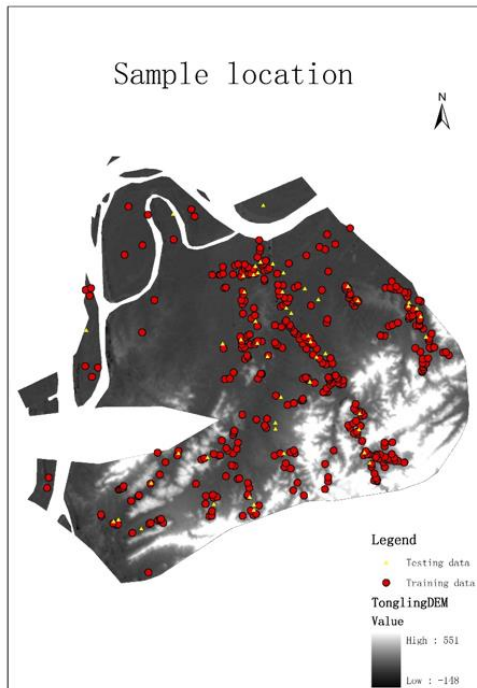


Figure 1: Sample locations positioned on the DEM of the study area

2.2 Prediction methods

The fuzzy c-means classifier (FCM) was used to identify the environment classes that existed in the environmental data set. The FCM classifier first divides an environmental database into a given number of classes and then computes the membership of each object to each of these classes. Those unique combinations can be regarded as the prototype by the case study method to establish the fuzzy membership function for soil properties prediction (Brungard, et al., 2015). Expert knowledge extracted from the conventional soil map, the fuzzy c-means method, locally specific knowledge was incorporated to predict soil lead by establishing the fuzzy membership function in this paper.

There are thus four main components that need to be adhered to replicate this approach. The first step is to extract different soil subareas extracting by the soil order based on expert knowledge. The second step involves generation of terrain attributes. The third step includes the classification by the fuzzy c-means methods performed by the FCM software and the optimal class selection by the partition coefficient. The fourth step contains the fuzzy membership function establishing by those class prototype and the prediction of the soil lead by the fuzzy membership.

For the statistical model, we considered both the multivariate linear regression model and the Regressive model. In the absence of spatial correlation, the two models provide identical results. We selected a stepwise approach to identify the optimal DEM variables. The explanatory variables of the multiple linear regression soil landscape models were ELE, SL, AS, PR, PL and TWI and the response variable was the soil lead.

For both prediction maps, validation was performed using the same dataset that is independent from the calibration data. To determine the accuracy of the final predictions with their corresponding observed values, we use the root mean square error (RMSE) and agreement Coefficient (AC).

We used the following software in this study: the linear regressions were done with R, soil map generation and fuzzy membership prediction were done with Arc Info, layer stacking in Erdas software, FCM software was developed in C language, whereas the partition coefficient was calculated with Matlab.

3. Results and discussions

For the fuzzy logic model, we defined a prototype as the area with the fuzzy membership above 0.85. As concerns the weighing coefficient, a value of $m=2$ is considered to have a clear-cut physical meaning that reflect the gradual changes property of soil. Fuzzy membership values below 0.85 indicate that the prototype is atypical. We notice that the number of prototypes equals 11 for Yellow-red soil, 9 for Periodical waterlogged paddy soil, 8 for Calcareous alluvial soil, 9 for Brown red soil, 7 for Ground-water paddy soil, 7 for Brown Rendzina, 8 for red soil, 8 for Acid Lithosol, 8 for Degleyed Paddy soils, 6 for Ferrosilicon yellow clay and 9 for neutral purple soil in our study area. The soil lead prediction map is the integrated result merged by the prediction results of the different soil types. The weighted average model was used to derive the soil depth map (Menezes, et al., 2016). A soil depth of each pixel at location x is computed as the equation.

$$V(x) = \frac{\sum_{k=1}^K \mu_k(x) V_k(x)}{\sum_{k=1}^K \mu_k(x)} \quad (1)$$

where $V(x)$ is the soil depth at location x ; $V_k(x)$ is the typical value of the soil depth of soil subgroup k , $k = 1, \dots, K$; $\mu_k(x)$ is the membership value of soil subgroup k at x and K is the total number of soil subgroups. The typical value of soil depth $V_k(x)$ was approximated by the measured values of typical points from the samples for the weighted average model.

The variables selection of the multiple linear regressions using the multiple regression model step by step was done using a backward analysis. Starting with the training set, all parameters are estimated. Thus, proceeding we get the final model, concerning parameters SL, ELE, AS, PR PC and TWI in turn. The final model is computed as the equation below.

$$\text{Lead} = 67.44 + \text{AS} * 0.028 + \text{ELE} * 0.073 + \text{PC} * 31.48 + \text{PR} * 65.17 + \text{SL} * 2.22 - \text{TWI} * 0.15 \quad (2)$$

Finally, the two models were used for predicting soil lead (Figure. 2 and Figure. 3). Inevitably, the regression-based map shows some unrealistic negative soil lead values, which is also unrealistic at the ground survey data.

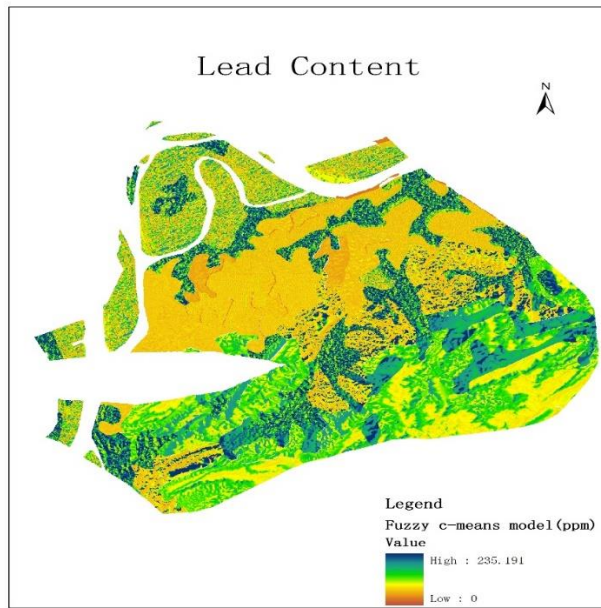


Figure 2: The soil lead map based on the fuzzy c-means model

The RMSE and AC were applied to compare the predicted map derived from the fuzzy c-means cluster approach with that obtained from the regression model. The RSME value of the same 50 validation points equals 39.829 for the fuzzy c-means approach, which is slightly lower than the value of 42.541 obtained for the linear regression model. This means that our prediction stability of soil lead values from former model is better than the latter model, despite the relatively strange values obtained. The AC value indicates a better agreement between the predicted values and the observed values at these points by the fuzzy c-means method (0.673) than by the regression model (0.415) (Table 1).

Table 1: The Comparison of the fuzzy membership function approach with the linear regression model

Number of samples (90)	RMSE	AC
Fuzzy membership approach	39.829	0.673
Multiline regression model	42.541	0.415

The fuzzy logic method can deal with large data in ordinary personal computer by transform the data to ASCII as Tongling County. The fuzzy software success to run a large data at our personal computer as ASCII or a lot of datasets can lead to the memory limitation for computational difficulties in the original image data. The soil lead result predicted by the fuzzy membership shows more realistic values than the results of the regression model. The reason is that the fuzzy c-means cluster analysis uses expert knowledge extracted from the conventional soil map and integrates it for the whole area as training data area. The prediction accuracy of the multiple linear regression models, however, determined by the range and quantity of samples, the exceptional value of prediction. The method is limited as the typical prototype sample location is dependent on the study area. The fuzzy membership value must be recalculated to require the new typical prototype for field sample when enlarging the study area. Within the same study area, the accuracy of the sample location depends upon the selection of the environmental variables, the number of environmental clusters, the classification algorithm and parameters. The principle of weighted average model in equation 2 decides the prediction value range. Therefore, the maximum prediction value cannot be more than the maximum value of ground sample and cannot be less than zero, but it is also a limitation of the weighted average model at some actual sample location whose soil lead is really equal to zero. With the development of the mathematical theory, high quality remote sensing imagery and advanced computer technology, those problems are likely to be overcome in the future. This could then promote the success of the soil lead prediction by the soil-landscape model.

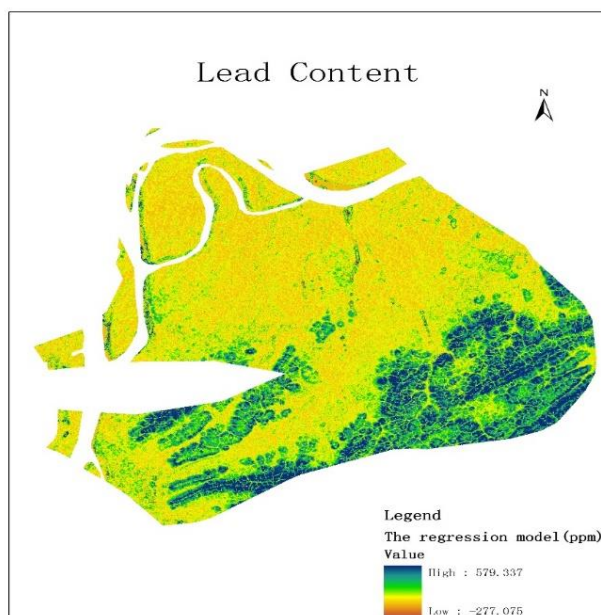


Figure 3: The soil lead map based on the regression model

4. Conclusion

This paper presents a simple approach to the construction of fuzzy membership functions from descriptive knowledge generated by fuzzy logic methods with conventional soil maps. Firstly, the c language is an effective language to accomplish the fuzzy c-means cluster in large study area for establishing soil-landscape model but data should less than two Gigabyte or the computer will halt. Secondly, the soil-landscape model constructed by the fuzzy membership functions with fuzzy c-means method and conventional soil map were able to produce good quality soil lead spatial information.

Reference

- Bowles, T.M., Hollander, A.D., Steenwerth, K., Jackson, L.E., 2015, Tightly-coupled plant-soil nitrogen cycling: comparison of organic farms across an agricultural landscape. *Plos One*, 10(6), e0131888.
- Brungard, C. W., Boettinger, J. L., Duniway, M. C., Wills, S. A., & Edwards, T. C. (2015). Machine learning for predicting soil classes in three semi-arid landscapes. *Geoderma*, 239, 68-83.
- Chen, H. Y., Y. G. Teng, S. J. Lu, Y. Y. Wang and J. S. Wang. 2015. "Contamination Features and Health Risk of Soil Heavy Metals in China." *Science of the Total Environment*, 512: 143-153.
- Desai, A.R., Xu, K., Tian, H., Weishampel, P., Thom, J., Dan, B., 2015, Landscape-level terrestrial methane flux observed from a very tall tower. *Agricultural & Forest Meteorology*, 201(2), 61-75.
- Grunwald, S. (Ed.). (2016). *Environmental soil-landscape modeling: Geographic information technologies and pedometrics*. CRC Press.
- Herman, F., Braun, J., 2016, Evolution of the glacial landscape of the southern alps of New Zealand: insights from a glacial erosion model. *Journal of Geophysical Research*, 113(113), F02009.
- Hosseini, R., Newlands, N., Dean, C., Takemura, A., 2015. Statistical modeling of soil moisture, integrating satellite remote-sensing (sar) and ground-based data. *Remote Sensing*, 7(3), 2752-2780.
- Jafari, A., Khademi, H., Finke, P. A., Van de Wauw, J., & Ayoubi, S. (2014). Spatial prediction of soil great groups by boosted regression trees using a limited point dataset in an arid region, southeastern Iran. *Geoderma*, 232, 148-163.
- Jiang, X., W. X. Lu, H. Q. Zhao, Q. C. Yang and Z. P. Yang. 2014. "Potential Ecological Risk Assessment and Prediction of Soil Heavy-Metal Pollution around Coal Gangue Dump." *Natural Hazards and Earth System Sciences* 14(6): 1599-1610.
- Kumar, S., Singh, R.P., 2016, Spatial distribution of soil nutrients in a watershed of himalayan landscape using terrain attributes and geostatistical methods. *Environmental Earth Sciences*, 75(6), 1-11.
- Li, Y. Q., Zhang, S.H., Peng, Y., 2015, Soil erosion and its relationship to the spatial distribution of land use patterns in the lancang river watershed, yunnan province, china. *Agricultural Sciences*, 06(8), 823-833.

- Ließ, M., Schmidt, J., Glaser, B., 2016, Improving the spatial prediction of soil organic carbon stocks in a complex tropical mountain landscape by methodological specifications in machine learning approaches. *Plos One*, 11(4), e0153673.
- Menezes, M. D. D., Silva, S. H. G., Mello, C. R. D., Owens, P. R., & Curi, N. (2016). Spatial prediction of soil properties in two contrasting physiographic regions in Brazil. *Scientia Agricola*, 73(3), 274-285.
- Nair, S.S., Preston, B.L., King, A.W., Mei, R., 2016, Using landscape typologies to model socioecological systems: application to agriculture of the united states gulf coast. *Environmental Modelling & Software*, 79(C), 85-95.
- Pinto, L. C., de Mello, C. R., Norton, L. D., Owens, P. R., & Curi, N. (2016). Spatial prediction of soil–water transmissivity based on fuzzy logic in a Brazilian headwater watershed. *Catena*, 143, 26-34.
- Siqueira, D.S., Jr, J.M., Pereira, G.T., Teixeira, D.B., Vasconcelos, V., Júnior, O.A.C., 2015, Detailed mapping unit design based on soil–landscape relation and spatial variability of magnetic susceptibility and soil color. *Catena*, 135(341-8162), 149-162.
- Stielstra, C.M., Lohse, K.A., Chorover, J., McIntosh, J.C., Barron-Gafford, G.A., Perdrial, J.N., 2015, Climatic and landscape influences on soil moisture are primary determinants of soil carbon fluxes in seasonally snow-covered forest ecosystems. *Biogeochemistry*, 123(3), 447-465.
- Van, D.M.W.M., Temme, A.J.A.M., De Kleijn, C.M.F.J.J., Reimann, T., Heuvelink, G.B.M., Zwoliński, Z., 2016, Arctic soil development on a series of marine terraces on central spitsbergen, svalbard: a combined geochronology, fieldwork and modelling approach. *Soil Discussions*, 2(2), 1345-1391.
- Yang, J., Weisberg, P.J., Shinneman, D.J., Dilts, T.E., Earnst, S.L., Scheller, R.M., 2015, Fire modulates climate change response of simulated aspen distribution across topo-climatic gradients in a semi-arid montane landscape. *Landscape Ecology*, 30(6), 1-19.
- Ziadat, F. M., Dhanesh, Y., Shoemate, D., Srinivasan, R., Narasimhan, B., & Tech, J. (2015). Soil-Landscape Estimation and Evaluation Program (SLEEP) to predict spatial distribution of soil attributes for environmental modeling. *International Journal of Agricultural and Biological Engineering*, 8(3), 158.