Pedosphere **27**(2): 344–357, 2017
doi:10.1016/S1002-0160(17)60322-9
ISSN 1002-0160/CN 32-1315/P

Published by Elsevier B.V. and Science Press

PEDOSPHERE

www.elsevier.com/locate/pedosphere

Regional Soil Mapping Using Multi-Grade Representative Sampling and a Fuzzy Membership-Based Mapping Approach

YANG Lin[1,2], A-Xing ZHU[2,3,4,5,6,*], ZHAO Yuguo[1], LI Decheng[1], ZHANG Ganlin[1], ZHANG Shujie[7] and Lawrence E. BAND[8]

[1] *State Key Laboratory of Soil and Sustainable Agriculture, Institute of Soil Science, Chinese Academy of Sciences, Nanjing 210008 (China)*
[2] *State Key Laboratory of Resources and Environment Information System, Institute of Geographical Sciences and Resources Research, Chinese Academy of Sciences, Beijing 100101 (China)*
[3] *Key Laboratory of Virtual Geographic Environment (Nanjing Normal University), Ministry of Education, Nanjing 210023 (China)*
[4] *Jiangsu Center for Collaborative Innovation in Geographical Information Resource Development and Application, School of Geography, Nanjing Normal University, Nanjing 210023 (China)*
[5] *State Key Laboratory Cultivation Base of Geographical Environment Evolution (Jiangsu Province), Nanjing 210023 (China)*
[6] *Department of Geography, University of Wisconsin-Madison, Madison WI 53706 (USA)*
[7] *China Academy of Urban Planning & Design, Beijing 100037 (China)*
[8] *Institute for the Environment, University of North Carolina, Chapel Hill, Chapel Hill NC 27599 (USA)*

(Received April 15, 2016; revised January 22, 2017)

ABSTRACT

High-resolution and detailed regional soil spatial distribution information is increasingly needed for ecological modeling and land resource management. For areas with no point data, regional soil mapping includes two steps: soil sampling and soil mapping. Because sampling over a large area is costly, efficient sampling strategies are required. A multi-grade representative sampling strategy, which designs a small number of representative samples with different representative grades to depict soil spatial variations at different scales, could be a potentially efficient sampling strategy for regional soil mapping. Additionally, a suitable soil mapping approach is needed to map regional soil variations based on a small number of samples. In this study, the multi-grade representative sampling strategy was applied and a fuzzy membership-weighted soil mapping approach was developed to map soil sand percentage and soil organic carbon (SOC) at 0–20 and 20–40 cm depths in a study area of 5 900 km^2 in Anhui Province of China. First, geographical sub-areas were delineated using a parent lithology data layer. Next, fuzzy *c*-means clustering was applied to two climate and four terrain variables in each stratum. The clustering results (environmental cluster chains) were used to locate representative samples. Evaluations based on an independent validation sample set showed that the addition of samples with lower representativeness generally led to a decrease of root mean square error (RMSE). The declining rates of RMSE with the addition of samples slowed down for 20–40 cm depth, but fluctuated for 0–20 cm depth. The predicted SOC maps based on the representative samples exhibited higher accuracy, especially for soil depth 20–40 cm, as compared to those based on legacy soil data. Multi-grade representative sampling could be an effective sampling strategy at a regional scale. This sampling strategy, combined with the fuzzy membership-based mapping approach, could be an optional effective framework for regional soil property mapping. A more detailed and accurate soil parent material map and the addition of environmental variables representing human activities would improve mapping accuracy.

Key Words: fuzzy clustering, parent lithology, representative grade, sampling strategy, soil spatial variations

Citation: Yang L, Zhu A X, Zhao Y G, Li D C, Zhang G L, Zhang S J, Band L E. 2017. Regional soil mapping using multi-grade representative sampling and a fuzzy membership-based mapping approach. *Pedosphere*. **27**(2): 344–357.

Regional soil information is increasingly needed for ecological modeling, land resource management, and policy-making applications at a regional or national scale (Paustian *et al*., 1997; Heinemann *et al*., 2002; Liu *et al*., 2011; Poggio *et al*., 2013). For example, global and regional models that address climate change, land degradation, and food security require soil spatial distribution data as input parameters. At present, soil property maps derived from legacy (conventional) soil maps at small scales are still the major source of soil spatial data for those models (Taghizadeh-Mehrjardi *et al*., 2014). However, the spatial resolutions and details of these soil property maps are not compatible with other model input data derived from detailed terrain analyses and remote sensing techniques (Zhu *et al*., 2001; Taghizadeh-Mehrjardi *et al*., 2014), which will greatly impact modeling accuracy. Therefore, there is an increasing need to derive high-resolution and detailed regional soil spatial information.

In recent decades, digital soil mapping techniques

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

have been developed to predict or interpolate spatial soil variations on the base of soil samples (Skidmore *et al.*, 1991; Zhu *et al.*, 2001; Grunwald *et al.*, 2009), and regional soil mapping has been conducted recently. Mora-Vallejo *et al.* (2008) applied regression kriging for soil mapping based on 95 composite soil samples over a 13 500-km^2 study area in Kenya. The results exhibited that the variance explained by kriging model was estimated as 13% and 37%, respectively, for soil organic carbon (SOC) and clay. Mishra *et al.* (2009) predicted and mapped SOC stocks in the state of Indiana, USA at different depth intervals using profile depth distribution functions and ordinary kriging based on 414 pedons obtained from the National Soil Survey Center database. As for the validation results, the highest correlation coefficient (r) between the observed and predicted SOC values was 0.75 and the lowest was 0.34. Kumar *et al.* (2012) used a geographically weighted regression kriging (GWRK) method to estimate the SOC stock for the state of Pennsylvania, USA based on 702 georeferenced soil profiles extracted from the National Soil Survey Center database. They concluded that GWRK was the least biased and more accurate, when compared with regression kriging, based on the lowest root mean square error (RMSE) and high coefficient of determination (R^2, 0.36 *vs.* 0.23). Wang *et al.* (2013) compared the geographically weighted regression (GWR) method and ordinary cokriging (OCK) in predictive mapping of soil total nitrogen in a study area covering 1 260 km^2 using 237 calibration soil samples. The adjusted R^2 for GWR and OCK were 0.57 and 0.69, respectively. Nussbaum *et al.* (2014) estimated the SOC stocks of Swiss forest soils using a robust external-drift kriging method based on 1 033 forest soil profiles. The predictive power of their model was moderate ($R^2 = 0.34$ for SOC stock in 0–30 cm and $R^2 = 0.40$ in 0–100 cm). Most of the abovementioned regional soil mapping applications are conducted using existing legacy soil sampling points with statistical or geostatistical methods. For the areas with no point data or in need of updating, efficient derivation of regional soil distribution is a challenge.

Generally, there are two important steps for digital soil mapping (*i.e.*, soil sampling and soil mapping). Both sampling strategies and soil mapping approaches determine the accuracy of the mapping results (Brus and de Gruijter, 1997; Gregoire and Valentine, 2007; Brus and Noij, 2008; Heim *et al.*, 2009). Because field sampling is very costly at a regional scale, a cost-effective sampling design is especially important for regional soil mapping and is a prerequisite for accurate mapping results. The most commonly used sampling methods are simple random sampling and systematic sampling, especially in areas with no prior knowledge. However, these sampling strategies often require a large number of samples to obtain a desired accuracy for soil mapping at a regional scale. Another common type of sampling is geostatistics-based, which designs samples based on the spatial autocorrelation theory. However, a preliminary semi-variogram must be determined when using a geostatistics-based sampling strategy, which requires a large amount of sampling (Simbahan and Dobermann, 2006; Yang *et al.*, 2013).

Easy-to-obtain auxiliary information has been used to assist soil sampling in many recent applications to improve sampling efficiency; some examples of auxiliary data include the Latin hypercube sampling strategy proposed by Minasny and McBratney (2006), the sample optimization method to minimize spatially averaged universal kriging variance by Brus and Heuvelink (2007), and the multi-grade representative sampling strategy by Yang *et al.* (2013). Among these sampling strategies, multi-grade representative sampling is used to design typical or representative samples with different grades of representativeness to depict soil spatial variations at different scales based on the relationship between soil and its environmental covariates. It was tested to be effective for soil mapping at a watershed scale (Yang *et al.*, 2013). There are two advantages of this sampling strategy. First, it is capable of identifying a small number of representative samples which reduces sampling cost. Second, the representativeness information of the designed sample locations is an indicator of sampling order, which means that investigators can identify sample locations with higher sampling priorities and collect those samples first when sampling resources are not sufficient; they then collect samples with lower priorities when more resources become available. Due to the two advantages, multi-grade representative sampling could be a potentially efficient sampling strategy for regional soil mapping. However, soil pedogenesis in small-scale areas is usually complex. The efficiency of multi-grade representative sampling needs to be tested.

A suitable soil mapping approach is also needed for mapping regional soil properties based on an economically feasible number of representative samples when employing multi-grade representative sampling. There are numerous statistical or geostatistical mapping methods, such as regression modeling and kriging interpolations, which are widely used in digital soil map-