Ecological Indicators 93 (2018) 562–571

Contents lists available at ScienceDirect

Ecological Indicators

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

Original Articles

# Effects of sampling density on interpolation accuracy for farmland soil organic matter concentration in a large region of complex topography

Jun Long[a,b], Yaling Liu[c], Shihe Xing[a,b,*], Longxia Qiu[a], Qian Huang[a], Biqing Zhou[a], Jinquan Shen[d], Liming Zhang[a,b,*]

[a] University Key Lab of Soil Ecosystem Health and Regulation in Fujian, College of Resources and Environment, Fujian Agriculture and Forestry University, Fuzhou 350002, Fujian Province, China

[b] National Engineering Research Center of Juncao, Fujian Agriculture and Forestry University, Fuzhou 350002, Fujian Province, China

[c] Pacific Northwest National Laboratory, Joint Global Change Research Institute, University Research Court, College Park, Maryland 5825, USA

[d] Fujian Cropland Construction and Soil and Fertilizer Station, Fuzhou 350003, Fujian Province, China

ARTICLE INFO

*Keywords:*
Complex topography
Interpolation accuracy
Sampling density
Farmland soil organic matter
Spatial variation

ABSTRACT

Sampling density significantly affects the estimation of soil organic matter (SOM) concentration because it influences the interpolation accuracy. High sampling density may ensure adequate estimation, but it is costly. Low number of samples may underrepresent spatial variation and generate unacceptable predictions. Identifying a reasonable sampling density is challenging, especially where the topography is complex and characterized by slope-rich terrain. Here we addressed this challenge by taking a large region of complex topography as study area. The region had a total area of $1.24 \times 10^5$ km$^2$ and can be separated into three typical landforms, namely, hill-mountain, valley-basin, and plain-platform. Out of 235,309 sampling sites, 188,247 were randomly selected as training sites, on which 20 sampling densities were designed and ordinary kriging was interpolated. The remaining 47,602 sites were used as testing sites to calculate the accuracies of SOM concentration predictions at different sampling densities in the entire region of complex topography and its various landforms. Overall, the prediction accuracy was positively correlated with the sampling density ($R^2 \geq 0.98$). Specifically, with increasing sampling density, accuracy improved slowly at first then rapidly. However, the tipping point at which prediction accuracy significantly improved with the increases of sampling density varied among the areas. These sampling densities were 0.10, 0.11, 0.10, and 0.09 samples per hectare for the entire region, valley-basin, hill-mountain, and plain-platform, respectively. Further comparisons showed that valley-basin was the landform that had the best performance in interpolation accuracy, followed by hill-mountain, entire region and plain-platform. Their normalized root mean square error (NRMSE) values were 23.86%–28.91%, 24.22%–29.54%, 25.32%–30.77%, and 31.21%–37.76%, respectively. Moreover, interpolation accuracy was more sensitive to sampling density in simple topography (flat regions such as plain-platform) than in complex landforms (slope-rich terrains like hill-mountain, and valley-basin). These variations in the relationships between interpolation accuracy and sample density suggest that topography must be considered when designing a scientific sampling density. More importantly, when a high level of interpolation accuracy and low sampling costs are required in regions of similar complex topography, our findings may help optimize soil sampling density.

## 1. Introduction

Soil organic matter (SOM) affects environmental processes like soil fertility and carbon sequestration (Dolan et al., 2006; Maia et al., 2010; Panagos et al., 2013). Estimating the spatial and temporal variations in farmland SOM concentration helps maintain food production and mitigate global warming (Ferreyra et al., 2002; Dai et al., 2014; Zhao et al., 2014; Ottoy et al., 2017). However, soil properties are heterogeneous and SOM concentrations in different landforms may vary greatly (Lugato et al., 2010; Yu et al., 2011; Ajami et al., 2016). Therefore, accurate estimation of SOM concentration often requires scientific interpolation and large numbers of soil samples. Nevertheless, high sampling densities may be costly, whereas low sampling densities usually underrepresent spatial variations and may generate unacceptable predictions (McBratney and Webster, 1983; Bogunovic et al., 2014; Bhunia et al., 2016). Furthermore, the quantitative

* Corresponding authors at: University Key Lab of Soil Ecosystem Health and Regulation in Fujian, College of Resources and Environment, Fujian Agriculture and Forestry University, Fuzhou 350002, Fujian Province, China.
*E-mail addresses:* fafuxsh@126.com (S. Xing), lmzhang_1979@163.com (L. Zhang).

https://doi.org/10.1016/j.ecolind.2018.05.044
Received 14 February 2018; Received in revised form 8 May 2018; Accepted 18 May 2018

relationship between interpolation accuracy and sampling density is unclear or do not exist at all. These issues are even more apparent in complex topographies characterized by slope-rich terrain. Therefore, it is necessary to analyze the effects of sampling density on interpolation accuracy in complex topographies, to ensure prediction accuracy and limit sampling cost.

Sampling density substantially influences SOM concentration estimation by affecting interpolation accuracy. Wang et al. (2015) showed that 40 sampling sites sufficed to interpolate soil organic carbon (SOC) in a 0.44 $km^2$ coal mining area on the Loess Plateau. The study included 78 sampling sites and six sampling densities. Miller et al. (2016) studied an experimental area in the Uckermark District in the northeast German plain to estimate the SOC prediction accuracies of spatial modeling methods at two scales (meso-scale: 80 sampling points on 6 ha; macro-scale: 28 sampling points on 65 ha). Studies showed that the relationship between interpolation accuracy and sampling density in different landforms can vary significantly (Yu et al., 2011). Bourennane et al. (2000) designed five sampling densities based on 219 observation sites in a 380-ha area in the southwestern Parisian Basin. They found that the interpolation accuracy of universal kriging with external drift gradually improved with sampling density. However, Sahrawat et al. (2008) reported that the median or mean SOC values detected by using different sampling sizes did not differ significantly. Their conclusion was based on 114 soil samples collected from a gently sloping watershed (< 1% slope) ~500 ha in area located in the Indian semi-arid tropics. In another study, interpolation accuracies were compared at three sampling densities in a hilly area and were based on 254 soil samples. Results showed that increasing sampling density and selecting a reasonable interpolation method helped achieve an accurate estimate of the spatial variations in SOC (Zhang et al., 2015).

Most of the previously reported studies were conducted in small regions of single or simple topography. Therefore, the effects of sampling density on interpolation accuracy in large regions of complex, slope-rich topography remain unclear. In addition, the numbers of sampling densities designed in previous studies were relatively small (generally ≤ 6). Consequently, these studies did not elucidate variations in the relationship between interpolation accuracy and sampling density. Does interpolation accuracy respond to sampling density in complex topography in the same way as it does in simple topography? Does interpolation accuracy improve at a constant rate with sampling density increases or is there a quantitative relationship between them? These issues must be resolved in order to estimate SOM concentration accurately. The objectives of the present study, therefore, were to analyze the effects of sampling density and topographic variations on interpolation accuracy, and to examine the changes in the relationship between interpolation accuracy and sampling density in an entire region of complex topography and its various landforms. To that end, a large region of complex topography was used as the study area. It was characterized by slope-rich terrain and included nine cities. Moreover, 20 different sampling densities were designed and they were based on 235,309 sampling sites. Also, the interpolation accuracy for farmland SOM at different sampling densities in the entire study site was investigated.

## 2. Materials and methods

### 2.1. Study area

The large region of complex topography investigated in the present study has a subtropical oceanic monsoon climate and is located at 115°50′–120°40′E and 23°33′–28°20′N. It includes nine cities (Fig. 1). Under the influence of the Neocathaysian structural system, this region is characterized by slope-rich terrain, which can be separated into three typical landforms, namely, hill-mountain, valley-basin, and plain-platform (Zhang, 2002; Xing, 2003). The hill-mountain areas are distributed widely throughout the entire study area. Their elevations are mainly > 50 m and their slopes are > 6°. The valley-basin areas are embedded in the hill-mountain areas in a branch or bead configuration. They have elevations mostly > 50 m and slopes ≤6°. The plain-platform areas are distributed primarily in the coastal areas and their elevations are ≤50 m. More detailed geographical and climatic conditions of these three landforms are presented in Table 1. As of 2008, the farmland in the study area covered $1.33 \times 10^6$ ha. Forty-eight percent of it was situated in the hill-mountain areas, 27% in the valley-basin areas, and 25% in the plain-platform areas (Jiao, 2016).

### 2.2. Sampling sites

The 235,309 soil sampling site datasets were obtained from a project in which soil was tested for formulated fertilization. This project was sponsored by the Ministry of Agriculture of the People's Republic of China in 2008. Sampling site locations were determined using data from the Second Soil Survey of China conducted from 1980 to 1999, which was the most detailed and comprehensive survey to have ever been conducted in China (Zhao et al., 2006). However, the sampling density in 2008 was substantially higher than that in the Second Soil Survey. On average, each sampling site covered 5.65 ha. Moreover, 46%, 30%, and 24% of the sampling sites were located on the farmland in the hill-mountain, valley-basin, and plain-platform areas, respectively, which were commensurate with the relative proportions of farmland in each topographic region.

The sampling network was designed as follows. First, the farmlands were separated into several sampling units according to soil type, land use, and administrative division. Soil properties within the same sampling unit were essentially equal. Second, the sampling sites were located at the center of each sampling unit. Each sample consisted of 15–20 cores delineating an "s" shape around the sampling site. In this study, rice was the main farmland crop. The soil samples were taken from topsoil (0–15 cm) with a stainless-steel corer. Samples were collected in autumn 2008 after harvest and before soil preparation and base fertilizer application. SOM was determined by oxidation with potassium dichromate ($K_2Cr_2O_7$) and oil bath heating (Jing et al., 2014). Samples were first air-dried, cleansed of animal (Araneidae, Coleoptera, Lithobiomorpha, Staphylinidae, Oligochaeta, and Hymenoptera) and plant residues, ground, and sieved through a 0.25-mm mesh.

The sampling sites database was established as follows. First, sampling site latitudes and longitudes were recorded with a global positioning system (GPS; accurate to ~0.1″), imported into ArcGIS v. 9.3 (Environmental Systems Research Institute, Inc., Redlands, CA, USA), and converted into the Geodetic Coordinate System Beijing 1954 (Fig. 2a). Second, sample attributes data including location, soil type, farming system, SOM concentration, pH, total nitrogen, total phosphorus, total potassium, soil texture, and other parameters were linked to the aforementioned map. The database was then established according to the corresponding data structures and coding rules (Zhang et al., 2008).

### 2.3. Methods

#### 2.3.1. Sampling density design

Using the "Create Subsets" tool of ArcGIS v. 9.3, 80% of the 235,309 sampling sites (188,247; Fig. 2b) were randomly selected as training sites to perform interpolation and produce prediction maps. The remaining 20% (47,602; Fig. 2c) were used as testing sites to calculate interpolation accuracy (Maroju, 2007). To examine the effects of sampling density on interpolation accuracy, 20 subsets were created from the set of 188,247 training sites. The sampling sites of these 20 subsets were randomly selected with the "Create Subsets" tool. They were chosen in 5% increments of the total training sites from 9365 to 188,247 and evenly distributed spatially. For example, the 9365 training sites (5%) subset was sparse, whereas the 188,247 training