Computers and Electronics in Agriculture 142 (2017) 273–282

Contents lists available at ScienceDirect

Computers and Electronics in Agriculture

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

Three-dimensional stochastic simulations of soil clay and its response to sampling density

Zhang Shiwen [a], Zhang Lanlan [b], Li Zishuang [c], Wang Qingyun [d], Cui Hongbiao [a], Sun Zhongxiang [e], Ge Chang [a], Liu Huiling [a], Huang Yuanfang [e,*]

[a] College of Earth and Environmental Sciences, Anhui University of Science and Technology, Huainan, China
[b] Colleges of Agriculture, University Science and Technology of Anhui, Chuzhou, China
[c] Dezhou Academy of Agricultural Sciences, Dezhou, Shangdong, China
[d] School of Resources and Environment, Anhui Agricultural University, Hefei, China
[e] College of Resources and Environment, China Agricultural University, Beijing, China

ARTICLE INFO

Article history:
Received 22 December 2016
Received in revised form 28 August 2017
Accepted 31 August 2017
Available online 25 September 2017

Keywords:
Soil clay
Three-dimensional stochastic simulations
Sampling density
Response
sGsim

ABSTRACT

Clay is an active component in the mechanical composition of soil. The quantitative study of the spatial distribution of soil clay content is crucial to soil microecological research and agricultural or environmental management. The main purpose of the paper was to simulate soil clay content in three dimensions and reveal its response to sampling density based on sequential Gaussian simulation. The results showed the following: (1) With a reduction in samples, especially in the A horizon, spatial correlation was relatively enhanced and randomness weakened. (2) The spatial distribution of soil clay showed soil had high clay content in the mid-eastern region of the Haidian District, Beijing, and clay content was generally low in the other areas; (3) With a decrease of sampling density, the simulated spatial distribution of clay became gradually more homogeneous. The stochastic simulation results for two kinds of sampling densities, i.e., SD_1 and SD_4 were closer to the original measured values; the general distribution was discrete and could more accurately reflect the local volatility of the original data distribution; (4) Considering the root mean square error (RMSE), accurate plot, standard error map, and quartile deviation map, SD_1 had the best effect of three-dimensional stochastic simulation, followed by SD_4; (5) For soil clay, three-dimensional sampling can be applicable to reduce samples required in the lower horizon in order to reduce the sampling workload.

1. Introduction

The clay content of soil has a significant effect on soil C mineralization, organic matter decomposition, and migration of soil microorganisms (Wang et al., 2003; He et al., 2009). Soil physics tells us that soil colloid dispersion decreases with increasing clay content because of increased aggregate strength (Kjaergaard et al., 2002). In addition, the clay content of soil is an important input parameter for models of soil water and salt transport, or for carbon and nitrogen circulation models such as Hydrus and De-Nitrification-Decomposition (DNDC) models; in addition, clay content is an important indicator used in soil pedotransfer functions such as the estimation of soil water capacity (Bagheri et al., 2005). Therefore, it is necessary to understand the distribution of soil clay content. Some effort has been devoted to the assessment of soil textures on a regional scale through one-dimensional simulation (Li et al., 1997, 1999; Zhang and Li, 2008; Zhang et al., 2013, 2014) and two-dimensional simulation at field (Weigand et al., 2001). However, the three-dimensional spatial distribution simulation of the clay content of soil, which may play a significant role in soil management and agricultural production, is still rare in the literature.

In recent years, some studies on three-dimensional spatial simulation of soil properties, such as soil strength (Castrignano et al., 2002), soil nitrate-nitrogen (Van Meirvenne et al., 2003), soil organic matter or soil carbon storage (Minasny et al., 2006, 2014; Meersmans et al., 2008; Mishra et al., 2009), soil texture horizons (He et al., 2009) and soil salinity (Li et al., 2013a,b) have been carried out based on the three-dimensional geostatistics and profile distribution function method. The application of the conditional stochastic simulation method to soil science and other disciplines has attracted increased attention after the defects of the kriging

* Corresponding author.
E-mail addresses: mamin1190@126.com (Z. Shiwen), yfhuang@cau.edu.cn (H. Yuanfang).

http://dx.doi.org/10.1016/j.compag.2017.08.031
0168-1699/

method were overcome so that the spatial structure of the original variables can be reproduced very well, and the uncertainty of the prediction results of simulation were quantifiable. Scale (e.g., sampling magnitude, sampling size, and sampling support) is frequently used in the study of geography recently, which is the spatio-temporal dimension of the study object or process, the spatio-temporal unit for information collection and processing, or the change pattern determined by spatio-temporal scope (Chen and Cai, 2010). The spatial variability of soil properties is the scale function, and autocorrelation of the same variable is very different at different scales. However, the response of a three-dimensional simulation of the clay content in soil to sample density, which will provide a basis for the optimization of sampling plan and reducing sampling cost, has still rarely been reported.

In this study, a size 55 m × 60 m field was selected as the study area. The following research topics were examined in the article under five sampling densities.

(1) Three-dimensional spatial variability of soil clay and its response to different sampling densities through nested variogram analysis and sequential Gaussian simulation (sGsim).
(2) The simulation effect and its response to sampling density under the same simulation method (i.e., sGsim) through comparison of three-dimensional simulation accuracy, local and spatial uncertainty.

2. Materials and methods

2.1. Soil sampling and data processing

The 55 m × 60 m size study area was in the Haidian District, Beijing, and was divided into a grid of 156 subplots of a size of 5 m × 5 m. Soil samples were collected to a 1-m soil depth in the 100 randomly selected grid points (Fig. 1). Each sample hole was re-packed using soil from the natural soil horizons (i.e. A horizon 0–30, B horizon 30–60, C horizon 60–100 cm). A total of 300 samples were obtained from the A, B, and C horizons (i.e., 100 for each horizon). The soil clay content was determined by the Pipette method. To analyze and reveal the influence of different sampling densities on spatial variability, five sampling densities were artificially created. The sampling density SD1 is 100 sampling points per layer. In 100 samples of each layer as the base, two kinds of sampling densities (denoted as SD_2, and SD_3) were obtained by respectively reducing 20% and 40% of the sample numbers per layer. Meanwhile, two kinds of sampling densities (denoted as SD_4 and SD_5) were created from the beginning of the A horizon with 100 samples and other levels decline by 1/5, to the C horizon where the sample number was 40. To minimize the effect of sample point distribution on the spatial simulation, the structures of the five sampling densities were consistent in random selection.

2.2. Theory and methods

2.2.1. Sequential Gaussian simulation (sGsim)

The sequential approach is the most straightforward method for simulating a multivariate Gaussian field. Each value is simulated sequentially according to its normal conditional cumulative distribution function, which must be determined at each location to be simulated. The conditioning data comprise all original data and all previously simulated values within the neighborhood of the point being simulated.

The first part of a SGS is to check if the known data are normally distributed, and if not, to apply normal-score transformation on them (Deutsch and Journel, 1998). Then, one calculates the variogram of the transformed data according to equation 1. The next part is the simulation process itself.(1) One chooses at random a point to estimate inside the unknown data set. (2) Apply the kriging procedure, using the known data set and the modelized variogram, to estimate the mean $\hat{Z}(x_i$ and variance $\sigma_k^2(x_i)$ of this value. These two parameters are then considered respectively as the mean and variance of the local Gaussian probability density function of the point. (3) One randomly chooses a value for the unknown point, following the law N $(\hat{Z}(x_i, \sigma_k^2(x_i))$. (4) The simulated point is then considered as a known point and will be used to simulate the next randomly chose unknown point. (5) Repeat from step 1, until there are no more unknown points (Goovaerts, 1997; Remy et al., 2009; Lima, 2005; Zhang et al., 2017). The fundamental steps in SGS algorithm were depicted in Fig. 2.

2.2.2. Effect of spatial simulation and its response to sampling density

We evaluated and analyzed three-dimensional simulation efforts and their response to sampling density from the simulation accuracy and uncertainty.

2.2.2.1. The accuracy of spatial simulation and its response to the sampling density. The cross-validation root mean squared error (RMSE) was calculated to assess the simulation performance and its response to different sampling densities. The RMSE value indicates the accuracy of prediction; a smaller RMSE indicates the prediction results are relatively more accurate (Zhang et al., 2012, 2013; Ye et al., 2015).

$$\mathrm{RMSE} = \sqrt{\frac{1}{n}\sum_{j=1}^{n}(z(x_j) - \bar{z}(x_j))^2} \quad (1)$$

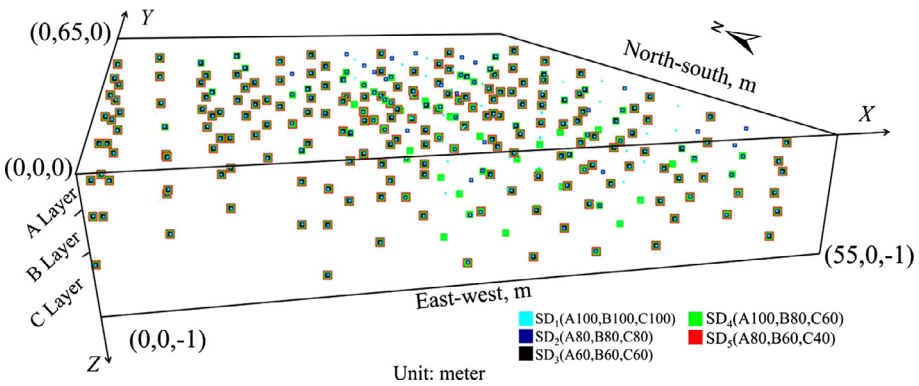

Fig. 1. Three-dimensional map of the sampling points D_1, SD_2, SD_3, SD_4, and SD_5 respectively represent five sampling densities. A(0–30 cm), B(30–60 cm), and C(60–100 cm) are three levels of soil. 100, 80, 60 and 40 represent the samples of each soil layer.