Catena 141 (2016) 46–55

Contents lists available at ScienceDirect

Catena

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

Three-dimensional geostatistical modeling of soil organic carbon: A case study in the Qilian Mountains, China

D.J. Brus [a,b,*], Ren-Min Yang [a], Gan-Lin Zhang [a]

[a] State Key Laboratory of Soil and Sustainable Agriculture, Institute of Soil Science, Chinese Academy of Sciences, Nanjing 210008, China

[b] Alterra, Wageningen University and Research Centre, Environmental Science Group, P.O. Box 47, 6700 AA Wageningen, The Netherlands

ARTICLE INFO

Article history:
Received 5 October 2015
Received in revised form 15 February 2016
Accepted 18 February 2016
Available online 1 March 2016

Keywords:
Digital soil
Mapping kriging with an external drift
REML
Linear mixed model
Sum-metric variogram

ABSTRACT

For mapping soil properties in three dimensions the simplest option is to choose a series of depth intervals, and to calibrate a two-dimensional (2-D) model for each interval. The alternative is to calibrate a full three dimensional (3-D) model that describes the variation in lateral and vertical direction. In 3-D modeling we must anticipate possible changes with depth of the effects of environmental covariates on the soil property of interest. This can be achieved by including interactions between the environmental covariates and depth. Also we must anticipate possible non-stationarity of the residual variance with depth. This can be achieved by fitting a 3-D correlation function, and multiplying the correlation between two points by the residual standard deviations at these two points that are a function of depth. In this paper various 3-D models of the natural logarithms of SOC are compared with 2-D depth-interval specific models. Five environmental covariates are used as predictors in modeling the lateral trend. In the 3-D models also depth was used as a predictor, either categorical, with categories equal to the depth intervals (3-Dcat), or continuous (3-Dcon). The covariance of the residuals in 3-D is modeled by a sum-metric covariance function. Both stationary and non-stationary variance models are fitted. In the non-stationary variance models the residual standard deviations are modeled either as a stepwise function or as a linear function of depth. In the 2-D models the regression coefficients differed largely between the depth intervals. In the 3-Dcat model extreme values for the regression coefficients were leveled out, and in the 3-Dcon model only the coefficients of NDVI and aspect changed with depth. The 3-Dcon model with a residual standard deviation that is a stepwise function of depth had the largest residual log-likelihood and smallest AIC among all 3-D models. Based on the cross-validation root mean squared error (RMSE) there was no single best model. Based on the mean and median of the standardized squared error (MSSE, MedSSE) the 2-D models outperformed all 3-D models. Overestimation of the prediction error variance by the kriging variance was less strong with the non-stationary variance models compared to the stationary variance models. 3-D modeling is required for realistic geostatistical simulation in spatial uncertainty analyses.

1. Introduction

In the past decades numerous papers have been published on the mapping of soil organic carbon (SOC)(see Minasny et al., 2013 for a review). This volley of papers illustrates the importance attached to an accurate three-dimensional picture of carbon in soil, in order to assess the role of soil as a terrestrial sink of carbon (Kirschbaum, 1995; Lal, 2004). A common problem in mapping SOC using legacy soil data is that the sampled depth intervals are not constant. This is because pedologist preferred to collect bulked soil samples from pedogenic horizons. As a consequence the reported SOC concentrations are average concentrations of depth intervals that differ between the sampling locations. To use these data in three dimensional (3-D) mapping the following three-step approach is very common(see for instance Malone et al., 2009; Liu et al., 2013; Adhikari et al., 2014). In the first step at the sampling locations an equal-area smoothing spline function is fitted to describe the variation of SOC with depth. In the second step, the fitted function is used to compute average SOC concentrations for the depth intervals of interest, e.g. 0–5, 5–15, 15–30, 30–60, 60–100, and 100–200 cm, the depth intervals used in the GlobalSoilMap project (Arrouays et al., 2014). Then in the third step these average concentrations are mapped one-by-one. For instance, to map the average concentration in the first depth interval 0–5 cm only the 'observed' average concentrations of this depth interval are used; the 'observed' average concentrations of other depth intervals are not exploited. Although the final picture of SOC concentration has three dimensions, this picture is constructed with two-dimensional (2-D) models. A drawback of this

* Corresponding author at: State Key Laboratory of Soil and Sustainable Agriculture, Institute of Soil Science, Chinese Academy of Sciences, Nanjing 210008, China.
E-mail address: dick.brus@wur.nl (D.J. Brus).

http://dx.doi.org/10.1016/j.catena.2016.02.016
0341-8162/

2-D modeling approach per depth interval is that predictions can be suboptimal in situations where SOC at a specific depth is clearly correlated with SOC at another location and depth. Also, it can result in unrealistic predicted SOC profiles (vertical distributions).

An alternative for the approach with multiple 2-D models is a 3-D approach, in which the variation in three dimensions is described by a single model. In this paper we will explore the potentials of 3-D geostatistical modeling of SOC in a case study in the Qilian Mountains in China. Lateral variation in the mean of SOC is modeled as a linear combination of environmental covariates such as terrain attributes and climate variables. Vertical variation is modeled by including depth as a predictor, either as a categorical or as a continuous predictor. To anticipate possible changes of the effects of the environmental covariates with depth, interactions between the environmental covariates on the one side and depth on the other side are included in the model.

In 3-D modeling much effort is made on 3-D modeling of the covariance of the residuals. Accurate modeling of the residual covariance is important to obtain reliable estimates of the variance of the prediction errors (kriging variance). To model the covariance of the residuals we account both for geometric anisotropy (a unit of distance in vertical direction is not comparable with such unit in lateral direction in terms of covariance) and zonal anisotropy (variances in vertical and lateral direction can be different). Very recently Li et al. (2016) compared various covariance functions commonly used in space–time geostatistics for 3-D modeling of soil salinity. We adopted a sum-metric covariance model (Heuvelink and Griffith, 2010). Besides, we account for a non-stationary residual variance in vertical direction, using the approach of Marchant et al., (2009). The residual variance is modeled as a stepwise or linear function of depth.

The following research questions were formulated:

- Do the effects of the environmental covariates change with depth, and are these changes comparable in the 2-D and 3-D models?
- Which 3-D model gives the most accurate predictions, the one with depth as a categorical predictor or the one with depth as continuous predictor?
- Can the quality of the kriging variances be improved by accounting for a non-stationary residual variance with depth?
- How different are the maps as obtained with the 2-D models and the 3-D model?
- Are predictions as obtained with the 3-D model more accurate than those with the 2-D model?

2. Study area and data

The study area is located in the Qilian Mountains, China (Fig. 1). The size of the study area is 30,193 km^2. In this area 106 locations were selected for soil sampling. The locations were selected by purposive sampling, using the method of Zhu et al. (2008). In this method representative soil sites are selected from soil-scape units constructed by fuzzy c-means classification of pixels on the basis of the soil forming factors.

At the sampling location major soil horizons were sampled, so sampling was not at fixed depth intervals. From each major soil horizon a bulked soil sample was collected, so the reported concentrations were average concentrations of soil horizons. The soil samples were dried, sieved at 2 mm and the soil organic carbon concentration was analyzed using the Walkley–Black procedure.

To obtain the SOC at fixed depth intervals a spline was fitted using Spline Tool Version 2.0 (ASRIS, 2011). This resulted into mean SOC concentrations for a series of fixed intervals, 0–5, 5–15, 5–30, 30–60, and 60–100. In this study these means were used as errorless measurements of SOC. Summary statistics of the SOC concentrations for the five depth intervals are presented in Table 1. The mean concentration decreased with depth from 4.7 mass % to 1.6; the standard deviation decreased with depth, but the coefficient of variation increased with depth. The total number of measurements in the first two depth intervals was 106; in the third, fourth and fifth depth interval 1, 6 and 31 less measurements were available, due to a soil thickness <100 cm. Fig. 2 shows the histograms of the natural log-transformed SOC concentrations per depth interval. Broadly speaking, these histograms were nicely symmetric. The log-transformed SOC concentrations in the five depth intervals were strongly correlated, see Table 2. The correlation coefficient decreased with the distance between the depth intervals.

3. Modeling

For modeling various maps with covariates were available: elevation, aspect, mean annual precipitation (MAP), mean annual temperature (MAT) and normalized difference vegetation index (NDVI). In modeling, these quantitative covariates were all scaled, so that their means were 0 and the standard deviations were 1. This improves the interpretability of the estimated regression coefficients, but does not have an effect on the coefficient of determination (R squared) and the

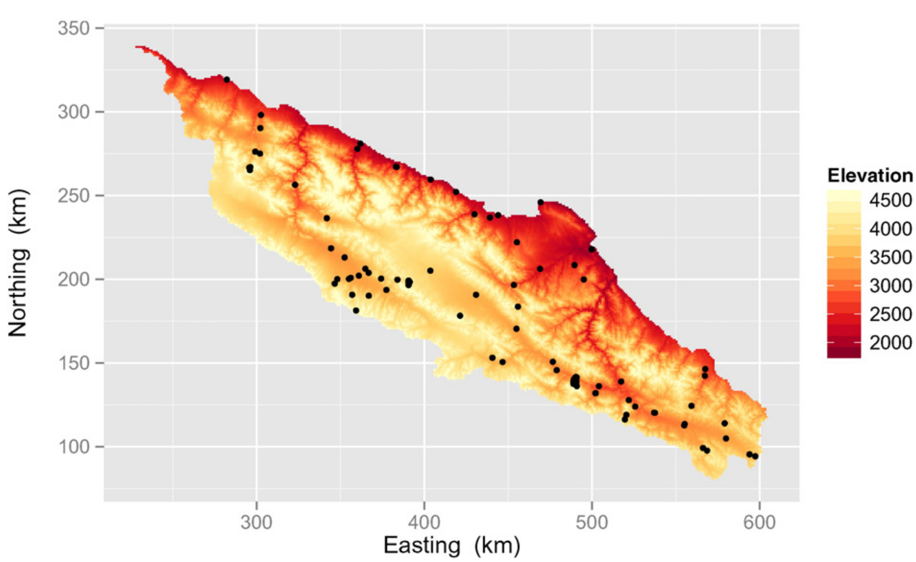

Fig. 1. The study area and the sampling locations.