Geoderma 330 (2018) 177–192

Contents lists available at ScienceDirect

Geoderma

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

A mixed model for landscape soil organic carbon prediction across continuous profile depth in the mountainous subtropics

Moritz Laub[a], Sergey Blagodatsky[a,b,*], Rong Lang[a,c,e], Xueqing Yang[a,d,e], Georg Cadisch[a]

[a] Institute of Agricultural Sciences in the Tropics (Hans-Ruthenberg-Institute), University of Hohenheim, Garbenstrasse 13, 70599 Stuttgart, Germany
[b] Institute of Physicochemical and Biological Problems in Soil Science, Russian Academy of Sciences, 142290 Pushchino, Russia
[c] Center for Mountain Ecosystem Studies, Kunming Institute of Botany, Chinese Academy of Sciences, Heilongtan, Lanhei Road 132, 650201 Kunming, China
[d] Key Laboratory of Economic Plants and Biotechnology, Kunming Institute of Botany, Chinese Academy of Sciences, Kunming 650201, China
[e] World Agroforestry Centre (ICRAF), China, East Asia Office c.o. Kunming Institute of Botany, Heilongtan, Lanhei Road No. 132, Kunming 650201, China

ARTICLE INFO

Handling Editor: A.B. McBratney

Keywords:
Digital soil mapping
Cost-constrained conditioned Latin hypercube sampling
Cokriging
Regression kriging
Upscaling

ABSTRACT

Due to the spatial variability of soil resources in rapidly changing landscapes, such as rubber expansion areas in mountainous South East Asia, landscape based soil organic carbon (SOC) stock assessments need new approaches to obtain cost effective high-resolution soil maps. 3D modelling presents the opportunity to model changes of soil properties with soil depth and in space in one single model. While most 3D models make use of spatial autocorrelation to create soil maps, it might be feasible for upscaling to neglect the spatial autocorrelation and only model autocorrelation within the soil profiles. We propose a "mixed model over continuous depth" (MMCD), which uses a linear and quadratic term to model changes of soil properties with depth and predicts the spatial distribution of soil properties at the landscape level. As the study area of 43 km^2 in South West China was subject to multiple constraints such as sparse road networks, steep terrain, and poor infrastructure, we applied the cost-constrained conditioned Latin hypercube sampling (CCLHS) scheme for soil sampling at 120 locations to a depth of 1 m. The MMCD provides information on the most important drivers of selected soil properties, and their relative importance. In this study, SOC was strongly linked to an interaction of elevation with mean horizon depth ($p < 0.001$) and to the land use type ($p < 0.001$). An iterative leave-one-third-out evaluation was performed to compare the MMCD to several established 2D and 3D mapping approaches. The MMCD proofed to be as powerful as these established techniques, with an overall modelling efficiency (EF) of 0.72. All tested models had a strong decrease of accuracy with depth, from an EF of about 0.8 in the topsoil to 0.2 at 0.8 to 1 m subsoil depth. The MMCD was further used to model highly unbalanced SOC density data with 120 independent topsoil observations and only 11 locations with subsoil observations (EF of 0.75), where the computed prediction intervals (95%) accurately covered the range of legacy measurements. Our approach allowed upscaling of SOC density predictions to the surrounding larger nature reserve of 270 km^2. The resulting MMCD and 3D maps revealed that on average, 15 and 10% of SOC stocks are expected in the 0.6 to 0.8 m and 0.8 to 1 m soil depth intervals, respectively. The combination of CCLHS and MMCD is particularly suitable for mountainous subtropical areas with poor road networks. However, this approach requires a strong relationship of the soil property of interest with explanatory environmental covariates, as it does not consider spatial autocorrelation for soil mapping. The advantage of this restriction is that it is easy to apply to highly unbalanced datasets and easy to upscale, given that the environmental covariates in the surrounding area are similar to the calibration area.

1. Introduction

Due to the high importance of soil organic carbon (SOC) for the global carbon cycle (Zeng et al., 2004) and soil fertility (Gregorich et al., 1994; Shukla et al., 2006), there is a need for new approaches that quantify soil organic carbon stocks with high-resolution maps, also considering deeper soil horizons (Minasny et al., 2013). This is especially the case because SOC stocks greatly vary within landscapes (Dieleman et al., 2013), between different cropping systems (Gauder et al., 2016; Hellebrand et al., 2010), and amongst different land uses

Abbreviations: CCLHS, cost-constrained conditioned Latin hypercube sampling; DEM, digital elevation model; DIMLR, depth interval-based multiple linear regression; DSM, digital soil mapping; EF, modelling efficiency; KED, kriging with external drift; MMCD, mixed model over continuous depth; (R) MSE, (root of the) mean-squared-error; SAGA, System for Automated Geoscientific Analysis

* Corresponding author at: Institute of Agricultural Sciences in the Tropics (Hans-Ruthenberg-Institute), University of Hohenheim, Garbenstrasse 13, 70599 Stuttgart, Germany.
E-mail address: Sergey.Blagodatskiy@uni-hohenheim.de (S. Blagodatsky).

https://doi.org/10.1016/j.geoderma.2018.05.020
Received 30 November 2017; Received in revised form 15 May 2018; Accepted 17 May 2018

(Guo and Gifford, 2002). Soil sampling at the landscape level for digital soil mapping (DSM) is often costly, and site access difficulties as well as hazards can be major issues (e.g. Kidd et al., 2015), especially in mountainous areas. Virtually all sampling design methods fall somewhere in between the two extremes of optimizing the sampling either towards geographic space, such as grid sampling, or the virtual hyperspace of environmental covariates, which are assumed to be correlated with the soil properties of interest. If upscaling is wanted, the good distribution of sampling points within environmental covariates is of high priority, because upscaling is especially sensitive to errors (Cambule et al., 2013). Conditioned Latin hypercube sampling (Minasny and McBratney, 2006), one of the most widely used approach of covariate hyperspace methods, has been criticized as being inflexible in field sampling (Kidd et al., 2015) due to a lack of alternative sampling points, given inaccessibility. As mountainous landscapes with their steep slopes and poor road networks represent areas that are especially difficult to sample, sampling schemes which optimize feasibility are invaluable. The cost-constrained conditioned Latin hypercube sampling (CCLHS) (Roudier et al., 2012), which builds on the method of Minasny and McBratney (2006), is one of the most suitable methods. In conditioned Latin hypercube sampling without cost function, a user-predefined number of sampling points is distributed within the feature space of environmental covariates, using simulated annealing (Minasny and McBratney, 2006). An objective function, consisting of three weighted terms, assesses how well the marginal distribution of randomly drawn sampling points represents the marginal distribution of all possible sampling points for: (1) continuous covariates, (2) categorical covariates, and (3) the correlation of covariates. For a defined number of iterations, worst fitting sampling points are randomly replaced by alternative points and then kept if they reduce the objective function. The CCLHS adds a cost term to the objective function, which reduces the effort of reaching sampling points in addition to optimizing within the marginal distribution of environmental covariates. This approach was found suitable for areas difficult to access (Silva et al., 2014); it also represents a feasible design to implement upscaling exercises.

The choice of the model to create the maps should complement the selected sampling approach and match the characteristics of the study area. A selection of methods is available, such as geostatistical kriging and cokriging (McBratney and Webster, 1983), linear and nonlinear regression models, decision trees such as Random Forest and combined approaches such as regression kriging (Odeh et al., 1995), kriging with external drift (Goovaerts, 1997), or Cubist (RuleQuest Research, 2016). Because purely geostatistical approaches are of limited use for upscaling, a model based on environmental covariates might be preferable, especially if it performs equally well as geostatistical models in the core area.

A downside of many common DSM techniques, including regression-based models, is an individual modelling of different depth intervals (e.g. the depth intervals proposed by GlobalSoilMap.net; Arrouays et al., 2014), although some new models do not require this step anymore. Pedogenetic horizons in this case have to be normalized to depth intervals (mostly with the boundaries 0, 5, 15, 30, 60, and 100 cm depth) by an equal-area spline function (Bishop et al., 1999), which can severely bias the modelling dataset for profiles with anomalies in distributions (Dorji et al., 2014), for example when buried subsoil horizons with higher SOC contents are present. The individual modelling of depth intervals also neglects any autocorrelations of horizons from the same profile. It therefore is desirable to maintain the pedogenetic horizon-based data structure in DSM and to include the change of soil properties with depth in modelling. Some recent approaches have done this, such as Orton et al. (2016) and Brus et al. (2016), combining non-stationary sum-metric variance modelling with linear interaction models to fit a 3D autocorrelation structure. However, while both these approaches perform well to create maps within a study area, their spatial autocorrelation lacks data when upscaling.

For upscaling purposes, we therefore propose to make use of mixed models and directly fit them to pedogenetic horizon data. This simple approach neglects spatial autocorrelation but models the autocorrelation of soil properties within profiles, by adding random intercepts and deviations of the depth effects at each sampling point. This improves the estimation of soil properties change with depth. To create such a model, we made use of mixed-effect models including random effects (Laird and Ware, 1982), more commonly used in design experiments (Piepho et al., 2004) than in DSM. In this "mixed model over continuous depth" (MMCD), interactions of environmental covariates with depth and depth^2 allow to model the change of soil properties with profile depth as well as the landscape with one single model.

Because no data-demanding sum-metric autocorrelation needs to be modelled, the MMCD should also be able to model situations where significantly more data are available for the topsoil than for the subsoil, but the subsoil shows less variation, as is the case for bulk or SOC density. The MMCD should well complement the CCLHS, especially in upscaling, since its calibration depends on a good distribution of sampling points within environmental covariates, which is the main idea behind all Latin hypercube sampling schemes.

In this context, we hypothesized that (i) the combination of MMCD and CCLHS is a suitable DSM approach for heterogeneous mountainous areas, and an improvement over depth interval-based models. If strong explanatory covariates exist, (ii) it should perform equally well as other recent 3D DSM approaches and in that case, the MMCD can be upscaled to the surrounding area. As several soil properties, such as bulk and SOC density, are usually most variable in the topsoil, we furthermore hypothesized that (iii) these soil properties can be effectively modelled and upscaled with a MMCD, combining abundant topsoil bulk density measurements with sparsely sampled subsoil measurements.

Within this study, we evaluated whether it is possible to create detailed estimations of SOC concentrations and densities at the landscape level and to scale these up from our 43 km^2 study area in the subtropical mountainous Yunnan to the surrounding nature reserve of 267 km^2. The main objective of this study was therefore to test the combination of the sampling scheme (CCLHS) with the interpolation model (MMCD) and to compare it to established DSM techniques, as well as to create and evaluate high-resolution estimates of SOC concentrations and densities at the landscape level. Because CCLHS was used to optimize sampling point distribution, digitally available environmental covariates, such as elevation or land use maps, were the key components used to develop the sampling design and to create digital soil property maps.

2. Materials and methods

2.1. Study area

Our study area is part of the Naban River Watershed National Nature Reserve in the prefecture Xishuangbanna (hereafter referred to as the "Naban Reserve"), Southwest China, which is subject of the Sino-German research project "Sustainable Rubber Cultivation in the Mekong Region" (SURUMER, https://surumer.uni-hohenheim.de/). Three central watersheds were selected as study sites because they represented the full range of land uses and elevation gradients and contained the most important research plots of the SURUMER project. With an area of 43 km^2, they cover 16% of the nature reserve (Fig. 1).

Cambisols, Ferralsols, Acrisols, and Hydragic Stagnic Anthrosols (World Reference Base of Soil Resources - WRB 2006) have been identified as the main soil types in a prior study (Wolff and Zhang, 2010), while during this study, also Umbrisols and Nitisols, with the WRB 2014 (IUSS Working Group, 2014), were identified (data not published). Granite was the dominant parent material in the western part of the nature reserve and phyllite in the eastern part (Wolff and Zhang, 2010), but due to Chinese confidentiality politics, more detailed information or maps of parent material could not be obtained. The study area covers diverse environmental conditions, with elevations