Springer Environmental Science and Engineering

Gan-Lin Zhang
Dick Brus
Feng Liu
Xiao-Dong Song
Philippe Lagacherie *Editors*

Digital Soil Mapping Across Paradigms, Scales and Boundaries

Springer

Gan-Lin Zhang · Dick Brus · Feng Liu
Xiao-Dong Song · Philippe Lagacherie
Editors

Digital Soil Mapping Across Paradigms, Scales and Boundaries

Springer

Editors
Gan-Lin Zhang
Institute of Soil Science
Chinese Academy of Sciences
Nanjing
China

Xiao-Dong Song
Institute of Soil Science
Chinese Academy of Sciences
Nanjing
China

Dick Brus
Soil Science
Alterra
Wageningen
The Netherlands

Philippe Lagacherie
National Institute for Agricultural Research
Paris
France

Feng Liu
Institute of Soil Science
Chinese Academy of Sciences
Nanjing
China

ISSN 2194-3214 ISSN 2194-3222 (electronic)
Springer Environmental Science and Engineering
ISBN 978-981-10-0414-8 ISBN 978-981-10-0415-5 (eBook)
DOI 10.1007/978-981-10-0415-5

Library of Congress Control Number: 2015960825

Printed on acid-free paper

This Springer imprint is published by SpringerNature
The registered company is Springer Science+Business Media Singapore Pte Ltd.

Chapter 12
Spatial Assessment of Soil Organic Carbon Using Bayesian Maximum Entropy and Partial Least Square Regression Model

Bei Zhang and Sabine Grunwald

Abstract There has been great interest in the estimation of soil carbon over the last decade to address critical environmental, agronomic, and sociopolitical issues. Soil proximal sensing has shown much potential for soil carbon assessment. Visible/near-infrared diffuse reflectance spectroscopy (VNIRS) has been introduced as a complementary data source in digital soil mapping due to its cost effectiveness. However, in many studies, the uncertainty in soil modeling using VNIRS has not been explicitly taken into account. Bayesian maximum entropy (BME) is a modern geostatistical method that incorporates auxiliary/soft data within a theoretical sound framework. Our objective was to employ VNIR data and BME to spatially estimate soil organic carbon (SOC). Another objective was to compare the performance to estimate SOC using BME to classical geostatistical methods. A total of 1012 soil samples from Florida, USA, were employed from a database that included pairs of SOC measurements derived by dry combustion and hyperspectral data with 1-nm resolution in the VNIR spectral range (350–2500 nm). Partial least square regression (PLSR) was used to model the relationship between VNIR data and SOC. For spatial estimations of SOC, we employed BME using "hard" (SOC measurements from the laboratory) and interval "soft" data (predictions of VNIR–PLSR model). For the purpose of comparison, ordinary kriging (OK) was used with only the hard data set (OK1) and the SOC estimates derived from the VNIRS–PLSR model (OK2) at point locations. Both BME and OK2 show distinctly different pathways of assimilating vague ("soft") data into the spatial modeling process. The three spatial estimation methods (BME, OK1, and OK2) were examined using the independent validation set by calculating bias, root mean square error (RMSE), residual prediction deviation (RPD), and ratio of performance to inter-quartile distance (RPIQ). The preliminary

B. Zhang
College of Resources and Environment, Huazhong Agricultural University, 430070 Wuhan, China

B. Zhang · S. Grunwald (✉)
Soil and Water Science Department, University of Florida, 2181 McCarty Hall, PO Box 110290, Gainesville, FL 32611, USA
e-mail: sabgru@ufl.edu

G.-L. Zhang et al. (eds.), *Digital Soil Mapping Across Paradigms, Scales and Boundaries*, Springer Environmental Science and Engineering,
DOI 10.1007/978-981-10-0415-5_12

results show that BME performed generally as well as OK1, which may be due to the data splitting effects. However, both BME and OK1 were better than OK2. As BME can take advantage of data from the PLSR model, it offers the possibility to reduce the amount of laboratory-measured samples to map across a region. OK2 performed worse than OK1, which showed that using vague data into kriging leads to higher uncertainties. In this case, data from the VNIRS model may not help to improve the performance of predictions in kriging. These results underpin the potential of the BME approach in digital soil mapping.

Keywords Bayesian maximum entropy · Digital soil mapping · Soil carbon · Visible/near-infrared reflectance spectroscopy

12.1 Introduction

Soil carbon is considered as the largest pool of carbon in terrestrial ecosystems (Lal 2004) with multiple environmental cobenefits including fertility, productivity, and soil health that influence many agronomical, environmental, and political issues (Lacoste et al. 2014). Mapping the spatial distribution of soil carbon at a variety of spatial and temporal scales has been of great interest to address needs (Grunwald 2009; Minasny et al. 2013). A variety of methods have been used in soil carbon mapping, such as regression kriging (Vasques et al. 2010), geographically weighted regression (Zhang et al. 2011), and random forest (Wiesmeier et al. 2010). Visible/near-infrared reflectance spectroscopy (VNIRS) has been established as an alternative to more costly laboratory measurements to characterize soil properties. It is rapid and nondestructive and requires less sample preparation with less or no chemical reagents (McCarty et al. 2002; Viscarra Rossel et al. 2006; Brown 2007; Vasques et al. 2008). Modeling the quantitative relationships between soil attributes and spectral characteristics requires sophisticated statistical techniques (Viscarra Rossel et al. 2006). A variety of regression methods have been used for modeling soil VNIRS, such as principal component regression (PCR), partial least squares regression (PLSR), multiple linear regression (MLR), and artificial neural network (ANN) (Mouazen et al. 2010; Rossel and Behrens 2010). Among those techniques, PLSR is the most widely used (Brown et al. 2005; Vasques et al. 2008; Volkan Bilgili et al. 2010). The algorithm of PLSR is computationally faster than other methods; models are more interpretable and are relatively insensitive to over-fitting (Brodský et al. 2013). However, Brodský et al. (2013) found that PLSR modeling can cause uncertainty in the map of spatial prediction. More importantly, the uncertainty from spatial estimation by kriging can be substantial. Consequently, using VNIRS data directly in the kriging process may be not a good choice. The geostatistical methods that can incorporate auxiliary variables, such as regression kriging (RK) using VNIRS data to estimate soil properties (Ge et al. 2007), might be an alternative approach. However, if the relationship between auxiliary variable and target variable is not constant in all parts of the study area, the predictions might be even worse than just using plain kriging (Hengl

et al. 2007). It is critical to note that RK and PLSR fail to incorporate the prediction uncertainty explicitly into the modeling process. To explicitly incorporate soil spectral data into the modeling process to predict soil properties has been underexplored, but will be addressed in this study.

Bayesian maximum entropy (BME) proposed by Christakos (1990, 2000) is a modern geostatistical approach, which can integrate data with uncertainty into the modeling process, aiming to improve predictive capabilities compared with traditional estimation methods. In this framework, the term "hard data" refers to the most precise and accurate data with current instrumentation (e.g., soil analytical laboratory measurements), while "soft data" may represent varying levels of uncertain observations related to the target variables. The latter may be estimates of soil carbon derived from spectral data. Intervals of values or probability density functions are two ways to represent soft data. BME has been successfully applied in soil science (Bogaert and D'Or 2002; Douaik et al. 2004, 2005), environmental risk assessment (Lee 2005; Yu et al. 2009; Bogaert et al. 2009), environmental health (Puangthongthub et al. 2007; Money et al. 2009; Lee et al. 2009; Pang et al. 2010), and climate research (Lee et al. 2008).

Different kinds of soft data have been used in soil science, such as legacy soil map and raw measurement data (Bogaert and D'Or 2002; Douaik et al. 2004). There are no studies yet that have incorporated soil VNIRS data into the BME framework to improve predictions of soil properties.

The aims of this research were to: (i) investigate the performance of BME spatial estimation for SOC combined with VNIRS data, (ii) assess the performance of BME with soft data derived from VNIRS–PLSR models, and (iii) compare the accuracy of BME spatial estimation with traditional ordinary kriging as a reference.

12.2 Materials and Methods

12.2.1 Study Area

The study area is the State of Florida located in the southeastern Coastal Plain, USA, extending over six and one-half degrees of latitude (24.55–31.00 N, 80.03–87.63 W). The prevalent climate in Florida is humid subtropical, while the southern part has a tropical climate (add reference). The majority of soils in Florida are Spodosols (32 %), Entisols (22 %), Ultisols (19 %), Alfisols (13 %), and Histosols (11 %) (Natural Resources Conservation Service 2006). Land use and land cover are composed of wetlands (28 %), pinelands (18 %), and urban and barren lands (15 %), while agriculture, rangelands, and improved pasture occupy 9 %, 9 %, and 8 % of this state, respectively (Florida Fish and Wildlife Conservation Commission 2003). Florida is characterized by relatively flat topography with gentle slope varying from 0 to 5 % in most parts of the region. Only less than 1 % of the state has moderate slopes of 5–19 % (US Geological Survey 1999).

12.2.2 Data Preparation

We used soil data from a previous project "Rapid Assessment of Trajectory Modeling of Changes in Soil Carbon across a Southeastern Landscape" (courtesy of the soil database maintained by Dr. Grunwald's Pedometrics, Landscape Analysis, and GIS Laboratory). A detailed description of sampling design, laboratory analysis, and spectral scanning were provided by Xiong et al. (2014). Briefly, a total of 1012 soil samples were collected from March 2008 to August 2009 using a random-stratified sampling design based on land use—soil order strata. At each site, four 20 × 5.8 cm soil cores were collected within a 2-m-diameter area. These four soil samples were bulked in the field and then placed in a cooler until they could be transported to laboratory. SOC was analyzed in the laboratory using dry combustion (Shimadzu TOC-V/SSM-5000). For preliminary analysis, SOC values higher than the 75 % quantile of the whole data were removed. This pretreatment reduced the number of sample to 759.

Spectral data were derived from scanning of soil samples in the laboratory in the VNIR spectral range (350–2500 nm) at 1-nm intervals. Each sample was scanned four times. The average of these four scans was computed for every single sample. Three preprocessing methods were applied to the spectra data. First, the reflectance curves were smoothed across a moving window of 9 nm by using the Savitzky–Golay algorithm with a third-order polynomial. Then, to reduce the complexity of the data, the averages of reflectance values were taken across a 10-nm window. Last, the second-order derivate was applied with a 4 polynomials and 7-nm window. Those 3 steps reduced each of the spectral curves to 215 values.

12.2.3 Data Analyses

The BME approach was employed that provides a systematic and rigorous way to incorporate soft data in addition to hard data into the modeling process. According to Christakos (1990), BME balances two requirements: high prior information about the spatial variability and high posterior probability about the estimated map. The first requirement uses a variety of sources of prior information and involves the maximization of an entropy function. The second requirement leads to the maximization of a so-called Bayes' function.

To implement BME and kriging, all soil samples were randomly divided into four groups. The first one (model set) included laboratory-measured SOC and scanned spectral data and was used for establishing the VNIRS model using PLSR. By using the VNIRS–PLSR model, predictions of SOC were derived with spectral data from the second group (soft data set) to acquire prediction values and deviations. With the prediction outcomes of the VNIRS–PLSR model, each individual soft interval can be obtained. Specifically, for each sample in the soft data set, the upper limit of the interval equaled the SOC prediction value plus one deviation, and

the lower limit was set to the SOC predication value minus deviation. The last two groups were hard data and independent validation data, respectively. Although both of them were laboratory-measured SOC, the hard data set was only used for calibration and the other set was used to validate the performance of geostatistical estimations. Soft interval data and hard data were both integrated into the BME estimation process. The model set, soft data set, hard data set, and validation set included 190, 380, 114, and 75 samples, respectively.

For spatial estimations of SOC, we employed BME using "hard" (SOC measurements from the laboratory) and "soft" data (VNIR data). The BME analysis included three main stages (Christakos 1990; Douaik et al. 2005):

Prior stage: with the goal to maximize the information content using generalized knowledge which was implemented using the model set (i.e., pairs of laboratory-measured SOC and VNIR data) and PLSR to estimate the prior probability density function (PDF).
Meta-prior stage: By using the VNIRS–PLSR model, predictions of SOC were derived with spectral data from the second group (i.e., the soft data set) to acquire prediction values and deviations. With the prediction outcomes of the VNIRS–PLSR model, each individual soft interval can be obtained. Specifically, for each sample in the soft data set, the upper limit of the interval equaled the SOC prediction value plus deviation and the lower limit was set to the SOC predication value minus deviation.
Posterior stage: aiming to maximize the posterior PDF through updating of the prior PDF by taking into account the hard data set. The posterior and the prior PDFs are related through the conditional probability law based on Bayes' theorem.

For the purpose of comparison, ordinary kriging (OK) was used with only the hard data set (OK1). Then, OK was also employed using SOC estimates derived from the VNIRS–PLSR model (OK2). Both BME and OK2 show distinctly different pathways of assimilating vague data into the spatial modeling process. The three spatial estimation methods (BME, OK1, and OK2) were examined using the independent validation set by calculating bias, root mean square error (RMSE), residual prediction deviation (RPD), and ratio of performance to inter-quartile distance (RPIQ).

12.3 Results and Discussion

The descriptive statistics of SOC are reported in Table 12.1. Soil organic carbon (SOC) in Florida was highly variable with range values for all four sets with more than 17 g kg^{-1}. The means of all the four data sets were similar, except the validation set with a mean of 11.2 g kg^{-1} and median of 10.2 g kg^{-1}. This indicates a slight bias in the validation data set toward high SOC values which may have impacted the validation evaluation process. Moreover, the minimum value of the validation set was 3.9 g kg^{-1} that was substantially higher than the values in the soft

Table 12.1 Descriptive statistics of soil organic carbon in different data sets

	N	Mean	SD	Median	Min	Max	Range	Skew	Kurtosis
		g kg^{-1}							
Whole set	759	10.3	4.4	9.4	1.3	21.3	20.0	0.6	−0.5
Model set	190	9.9	4.5	8.7	3.0	21.1	18.1	0.7	−0.6
Soft set	380	10.2	4.2	9.5	1.3	20.9	19.6	0.5	−0.5
Hard set	114	10.3	4.7	9.5	1.9	20.9	19.1	0.6	−0.6
Validation set	75	11.2	4.6	10.2	3.9	21.3	17.4	0.6	−0.7

N Number of observations in each set, *SD* Standard deviation

and hard sets. This may be another reason that impacts the result of validation especially when there are estimated values less than 3.9 g kg^{-1} derived from BME and kriging. As SOC in all the data set was positively skewed, the PLSR model was built using natural logarithm-transformed SOC values.

The PLSR model built from the model dataset performed fairly well with R^2 of 0.52 and RMSE of 0.32 g kg^{-1} (Fig. 12.1). Figure 12.2 shows the predicted SOC values using the VNIRS–PLSR model and the spectral data from the soft set. This shows a good model fit. In both models (Figs. 12.1 and 12.2), residuals in the high and low SOC range were found indicating the uncertainty arising from models.

The spatial covariance derived from the hard set and soft interval data was modeled by nesting two exponential models. The sills for these two models were 13.5 and 2, and the ranges were about 1000 m and 30,000 m, respectively. The ranges of these models showed that the spatial correlation of SOC in Florida was relatively large (Fig. 12.3).

Figures 12.4, 12.5, and 12.6 show the results of spatial estimation of BME, OK1, and OK2 in Florida. The basic patterns of these three maps are quite similar, with the high values of SOC mainly located in the southeast corner of Florida consisting of highly organic soils. The range in SOC was narrower for BME than kriging, with

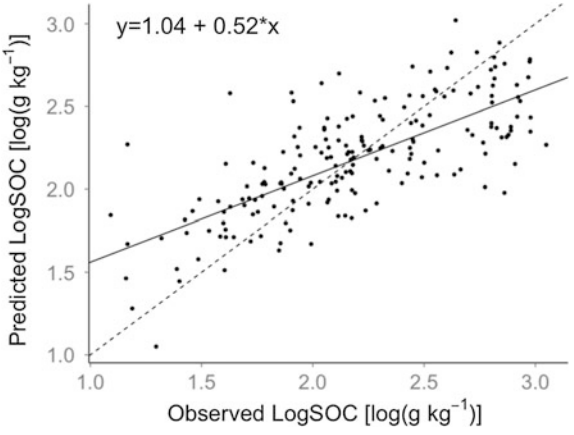

Fig. 12.1 Predicted and observed log-transformed soil organic carbon (log SOC) values derived from partial least square regression (PLSR) using the model set

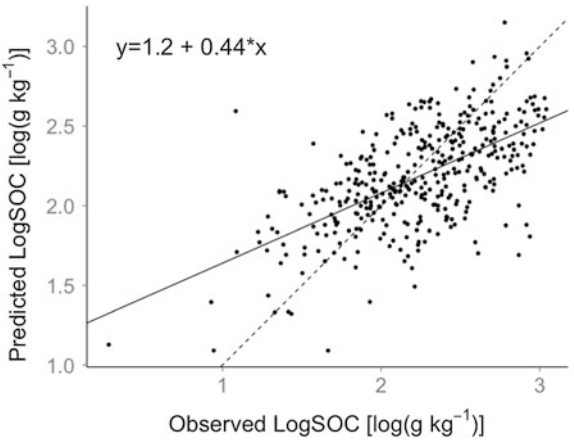

Fig. 12.2 Predicted and observed log-transformed soil organic carbon (log SOC) values derived from VNIRS–PLSR model using the soft data set

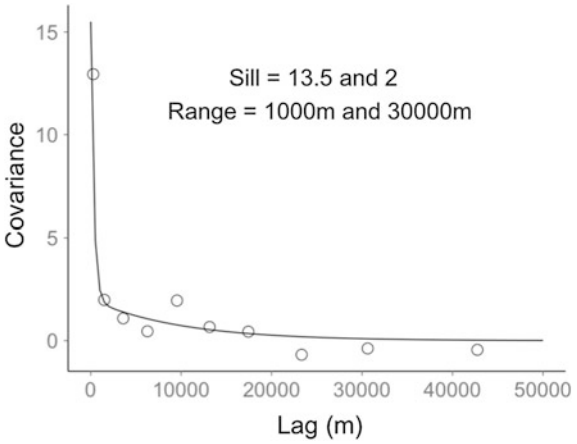

Fig. 12.3 The covariance structure of soil organic carbon (g kg^{-1}) and nested models

the maximum value only 15.4 g kg^{-1} that only covered about 75 % of the original range. This indicated that the method of BME was not sensitive to model high values. Another possible explanation may be the uncertainty associated with SOC of the PLSR model. As the goodness of fit for the model was just acceptable, the prediction SOC values may enlarge the uncertainty in BME. In contrast, the estimated SOC range by OK1 and OK2 corresponded well to the actual range in measured SOC within the State of Florida. The validation results indicate OK1 performance as well as BME (Table 12.2). The RMSE, RPD, and RPIQ of these two evaluations were almost the same. However, the SOC estimation map of BME was capable of showing more variations, whereas OK tended to smooth out SOC variation. The outcomes of OK2 were the worst, which indicate that assimilating vague data directly into kriging was not a good choice.

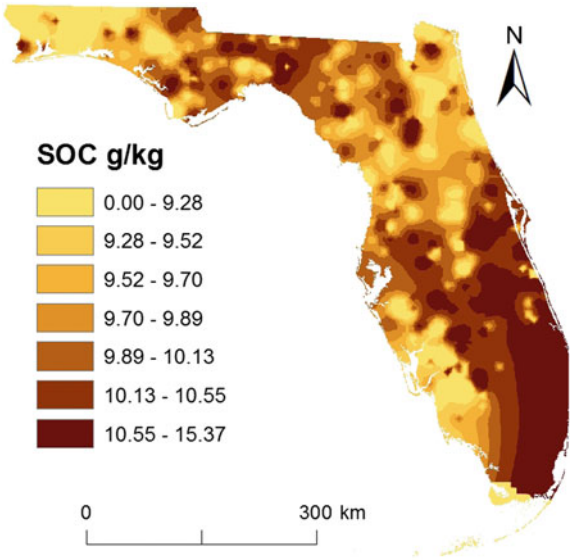

Fig. 12.4 Estimations of soil organic carbon (g kg^{-1}) using Bayesian maximum entropy

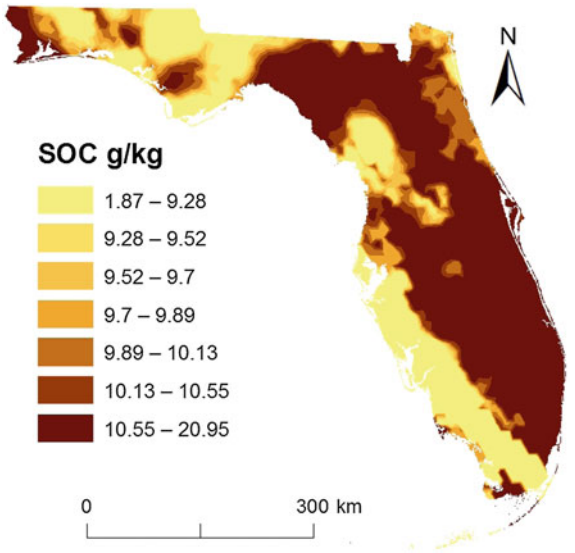

Fig. 12.5 Estimations of soil organic carbon in g kg^{-1} using ordinary kriging and the hard data set (OK1)

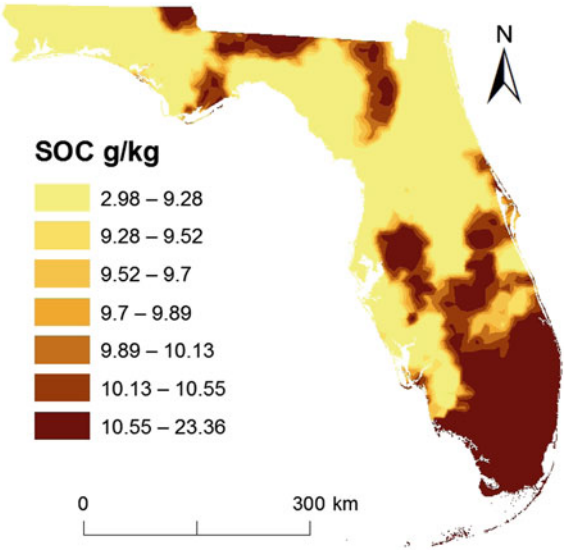

Fig. 12.6 Estimations of soil organic carbon in g kg^{-1} using ordinary kriging and the prediction from visible/near-infrared spectral (VNIRS)–partial least square regression (PLSR) model (OK2)

Table 12.2 Validation results for soil organic carbon (g kg^{-1}) derived from Bayesian maximum entropy, ordinary kriging with hard data (OK1), and ordinary kriging with predictions from partial least square regression model (OK2)

	RMSE[a]	Bias	RPD[a]	RPIQ[a]
BME	4.50	−1.27	1.02	0.63
OK1	4.45	−0.55	1.04	0.63
OK2	4.66	−1.86	0.99	0.61

[a]*RMSE* Root mean square error, *RPD* Residual prediction deviation, *RPIQ* Ratio of performance to inter-quartile distance

12.4 Conclusions

Theoretically, BME is expected to perform better than traditional univariate geostatistics (OK) because BME incorporates both—hard and soft data—into the modeling process. This was not found in the current preliminary study and needs further investigation probing into the causes. Preliminary findings suggest that the two methods, BME and kriging, performed almost the same using hard data. However, the spatial estimates of BME showed more details of SOC heterogeneity than OK1 and OK2. The VNIR spectral data used as soft data inputs in the BME modeling process possibly enhanced the capability to model SOC variability across Florida. The relatively small validation data set could not identify significant differences in performance between BME and OK in this study. Since the BME modeling process was influenced by many factors, such as preprocessing of VNIRS, the quality of the PLSR model, and parameter set during BME computing,

the accuracy of BME is expected to be improved by adjusting those factors. VNIR spectral data are easy to obtain and are poised to provide "vague" secondary data input to enhance the scarcity of hard (laboratory)-measured SOC data.

Acknowledgements We thank the following individuals for laboratory and field work to create the data sets used in this study: E. Azuaje, P. Chaikaew, A. Comerford, X. Dong, S. Moustafa, D. B. Myers, C.M. Ross, D. Sarkhot, L. Stanley, A. Stoppe, A. Quidez, and X. Xiong. We also thank co-PIs (Dr. N.B. Comerford and Dr. W.G. Harris) of the USDA-NIFA funded project "Rapid Assessment of Trajectory Modeling of Changes in Soil Carbon across a Southeastern Landscape" USDA-CSREES-NRI grant award 2007-35107-18368. And we also thank the China scholarship council and the Pedometrics, Landscape Analysis, and GIS Laboratory, Soil and Water Science Department, University of Florida, for hosting B. Zhang for a 2-year research visit.

References

Bogaert P, Christakos G, Jerrett M, Yu HL (2009) Spatiotemporal modelling of ozone distribution in the State of California. Atmospheric Environment 43:2471–2480. doi: 10.1016/j.atmosenv.2009.01.049

Bogaert P, D'Or D (2002) Estimating Soil Properties from Thematic Soil Maps. Soil Science Society of America Journal 66:1492–1500. doi: 10.2136/sssaj2002.1492

Brodský L, Vašát R, Klement A, Zádorová T, Jakšík O (2013) Uncertainty propagation in VNIR reflectance spectroscopy soil organic carbon mapping. Geoderma 199:54–63

Brown DJ (2007) Using a global VNIR soil-spectral library for local soil characterization and landscape modeling in a 2nd-order Uganda watershed. Geoderma 140:444–453. doi: 10.1016/j.geoderma.2007.04.021

Brown DJ, Bricklemyer RS, Miller PR (2005) Validation requirements for diffuse reflectance soil characterization models with a case study of VNIR soil C prediction in Montana. Geoderma 129:251–267. doi: 10.1016/j.geoderma.2005.01.001

Christakos G (1990) A Bayesian/maximum-entropy view to the spatial estimation problem. Math Geol 22:763–777. doi: 10.1007/BF00890661

Christakos G (2000) Modern Spatiotemporal Geostatistics. Oxford, New York

Douaik A, Van Meirvenne M, Tóth T (2005) Soil salinity mapping using spatio-temporal kriging and Bayesian maximum entropy with interval soft data. Geoderma 128:234–248. doi: 10.1016/j.geoderma.2005.04.006

Douaik A, Van Meirvenne M, Tóth T, Serre M (2004) Space-time mapping of soil salinity using probabilistic bayesian maximum entropy. Stoch Environ Res Risk Assess 18:219–227. doi: 10.1007/s00477-004-0177-5

Florida Fish and Wildlife Conservation Commission (FFWCC). Florida vegetation and land cover data derived from Landsat ETM+ imagery [Internet]. Tallahassee, FL; 2003. Available from: http://myfwc.com/research/gis/data-maps/terrestrial/fl-vegetation-land-cover/

Ge Y, Thomasson JA, Morgan CL, Searcy SW (2007) VNIR diffuse reflectance spectroscopy for agricultural soil property determination based on regression-kriging. Transactions of the Asabe 50:1081–1092.

Grunwald S (2009) Multi-criteria characterization of recent digital soil mapping and modeling approaches. Geoderma 152:195–207. doi: 10.1016/j.geoderma.2009.06.003

Hengl T, Heuvelink GBM, Rossiter DG (2007) About regression-kriging: From equations to case studies. Computers and Geosciences 33:1301–1315. doi: 10.1016/j.cageo.2007.05.001

Lacoste M, Minasny B, McBratney A, et al. (2014) High resolution 3D mapping of soil organic carbon in a heterogeneous agricultural landscape. Geoderma 213:296–311. doi: 10.1016/j.geoderma.2013.07.002

Lal R (2004) Soil Carbon Sequestration Impacts on Global Climate Change and Food Security. Science 304:1623–1627. doi: 10.1126/science.1097396

Lee S-J, Balling R, Gober P (2008) Bayesian Maximum Entropy Mapping and the Soft Data Problem in Urban Climate Research. Annals of the Association of American Geographers 98:309–322. doi: 10.1080/00045600701851184

Lee S-J, Yeatts KB, Serre ML (2009) A Bayesian Maximum Entropy approach to address the change of support problem in the spatial analysis of childhood asthma prevalence across North Carolina. Spatial and Spatio-temporal Epidemiology 1:49–60. doi: 10.1016/j.sste.2009.07.005

Lee SJ (2005) Models of soft data in geostatistics and their application in environmental and health mapping. Dissertation, University of North Carolina at Chapel Hill

McCarty GW, Reeves JB, Reeves VB, et al. (2002) Mid-Infrared and Near-Infrared Diffuse Reflectance Spectroscopy for Soil Carbon Measurement. Soil Science Society of America Journal 66:640–646. doi: 10.2136/sssaj2002.6400

Minasny B, McBratney AB, Malone BP, Wheeler I (2013) Chapter One – Digital Mapping of Soil Carbon. BS:AGRON 118:1–47. doi: 10.1016/B978-0-12-405942-9.00001-3

Money ES, Carter GP, Serre ML (2009) Modern Space/Time Geostatistics Using River Distances: Data Integration of Turbidity and E. coliMeasurements to Assess Fecal Contamination Along the Raritan River in New Jersey. Environ Sci Technol 43:3736–3742. doi: 10.1021/es803236j

Mouazen AM, Kuang B, De Baerdemaeker J, Ramon H (2010) Comparison among principal component, partial least squares and back propagation neural network analyses for accuracy of measurement of selected soil properties with visible and near infrared spectroscopy. Geoderma 158:23–31. doi: 10.1016/j.geoderma.2010.03.001

Natural Resources Conservation Service (NRCS), U.S. Department of Agriculture. U.S. General Soil Map (STATSGO2) [Internet]. Lincoln, NE; 2006. Available from: http://soils.usda.gov/survey/geography/statsgo/

Pang W, Christakos G, Wang J-F (2010) Comparative spatiotemporal analysis of fine particulate matter pollution. Environmetrics 21:305–317. doi: 10.1002/env.1007

Puangthongthub S, Wangwongwatana S, Kamens RM, Serre ML (2007) Modeling the space/time distribution of particulate matter in Thailand and optimizing its monitoring network. Atmospheric Environment 41:7788–7805. doi: 10.1016/j.atmosenv.2007.06.051

Rossel RAV, Behrens T (2010) Using data mining to model and interpret soil diffuse reflectance spectra. Geoderma 158:46–54. doi: 10.1016/j.geoderma.2009.12.025

United States Geological Survey (USGS). National Elevation Dataset (NED) [Internet]. Reston, VA; 1999. Available from: http://ned.usgs.gov/

Vasques GM, Grunwald S, Comerford NB, Sickman JO (2010) Regional modelling of soil carbon at multiple depths within a subtropical watershed. Geoderma 156:326–336. doi: 10.1016/j.geoderma.2010.03.002

Vasques GM, Grunwald S, Sickman JO (2008) Comparison of multivariate methods for inferential modeling of soil carbon using visible/near-infrared spectra. Geoderma 146:14–25. doi: 10.1016/j.geoderma.2008.04.007

Viscarra Rossel RA, Walvoort DJJ, Janik LJ, Skjemstad JO (2006) Visible, near infrared, mid infrared or combined diffuse reflectance spectroscopy for simultaneous assessment of various soil properties. Geoderma 131:59–75. doi: 10.1016/j.geoderma.2005.03.007

Volkan Bilgili A, van Es HM, Akbas F, et al. (2010) Visible-near infrared reflectance spectroscopy for assessment of soil properties in a semi-arid area of Turkey. Journal of Arid Environments 74:229–238. doi: 10.1016/j.jaridenv.2009.08.011

Wiesmeier M, Barthold F, Blank B, Kögel-Knabner I (2010) Digital mapping of soil organic matter stocks using Random Forest modeling in a semi-arid steppe ecosystem. Plant Soil 340:7–24. doi: 10.1007/s11104-010-0425-z

Xiong X, Grunwald S, Myers DB, et al. (2014) Holistic environmental soil-landscape modeling of soil organic carbon. Environmental Modelling & Software 57:202–215. doi: 10.1016/j.envsoft.2014.03.004

Yu HL, Chen JC, Christakos G (2009) BME estimation of residential exposure to ambient PM10 and ozone at multiple time scales. Environ Health Perspect 117:537–544. doi: 10.1289/ehp.0800089

Zhang C, Tang Y, Xu X, Kiely G (2011) Towards spatial geochemical modelling: Use of geographically weighted regression for mapping soil organic carbon contents in Ireland. Applied Geochemistry 26:1239–1248. doi: 10.1016/j.apgeochem.2011.04.014