Geoderma 285 (2017) 280–292

Contents lists available at ScienceDirect

Geoderma

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

Comparisons of spatial and non-spatial models for predicting soil carbon content based on visible and near-infrared spectral technology

Long Guo [a], Chang Zhao [b], Haitao Zhang [a,*], Yiyun Chen [c,d,f], M. Linderman [b], Qing Zhang [a], Yaolin Liu [c,d,e,f]

[a] College of Resources and Environment, Huazhong Agricultural University, Wuhan 430070, China
[b] Geographical and Sustainability Sciences, The University of Iowa, Iowa City 52246, USA
[c] Suzhou Institute of Wuhan University, Suzhou 215123, China
[d] School of Resource and Environmental Science, Wuhan University, Wuhan 430079, China
[e] Collaborative Innovation Platform for Geospatial Information Technology, Wuhan University, Wuhan 430079, China
[f] Key Laboratory of Geographic Information System of the Ministry of Education, Wuhan University, Wuhan 430079, China

ARTICLE INFO

Article history:
Received 26 February 2016
Received in revised form 25 September 2016
Accepted 12 October 2016
Available online 18 October 2016

Keywords:
Spatial dependence
Partial least squares regression
Spatial models
Visible and near-infrared diffuse reflectance spectroscopy

ABSTRACT

Visible and near-infrared (VNIR) reflectance spectroscopy is a rapid, non-destructive, and cost-effective method for predicting soil properties. Partial least squares regression (PLSR) is a common method used to predict soil properties based on VNIR reflectance spectra. However, PLSR ignores the spatial autocorrelation of soil properties and the assumption of linear regression models, in which explanatory variables and model residuals should be independently and identically distributed. In this study, PLSR, partial least squares–geographically weighted regression (PLS–GWR), partial least squares regression Kriging (PLSRK), and partial least squares–geographically weighted regression Kriging (PLS–GWRK) were constructed to predict soil organic matter (SOM) based on soil spectral reflectance. In addition, this study explores the influence of the spatial non-stationarity of explanatory variables on prediction accuracy. Among the aforementioned models, PLSR was used as a reference model; PLS–GWR considered the spatial autocorrelation of SOM and its auxiliary variables; PLSRK and PLS–GWRK considered the spatial dependence of the model residuals to ensure the usability of PLSR and PLS–GWR. A total of 256 topsoil samples (0–30 cm) were collected from Chahe Town, located in Jianghan Plain, China, and the reflectance spectra (400–2350 nm) of soil were used. The prediction capabilities of the models were evaluated using the coefficient of determination (R^2), the root-mean-square error (RMSE), and the ratio of performance to inter-quartile range (RPIQ). The evaluation indices showed that PLS–GWRK was the optimal model for predicting SOM using VNIR spectra. PLS–GWRK has the lowest values of $RMSE_C$ [0.109 ln $(g \cdot kg^{-1})$] and $RMSE_P$ [0.223 ln $(g \cdot kg^{-1})$] and the highest values of $R^2{}_C$ (0.933), $R^2{}_P$ (0.653), and RPIQ (3.015). PLS–GWR result showed that the spatial dependence of SOM and principal components could improve prediction accuracy compared with the PLSR result. The result of PLSRK showed that the spatial dependence of the model residuals could influence the prediction accuracy of PLSR. The PLS–GWRK approach explicitly addressed the spatial dependency and spatial non-stationarity issues for interpolating SOM at regional scale.

1. Introduction

Soils, which are natural expressions of minerals, organic matter, gases, liquids, and a myriad of organisms, continuously undergo development through physical, chemical, and biological processes (Jenny, 1941; Lyell, 1989). Soil organic matter (SOM) is a complex mixture that is significantly involved in soil fertility, biological productivity, and agricultural sustainable development (Overstreet and DeJong-Huges, 2009). Apparent changes in SOM will influence crop growth conditions, global climate, and the atmospheric concentrations of greenhouse gases (GHGs) (Shaver et al., 1992). The formation, variation, and decomposition of SOM are influenced by various spatial and temporal factors (Parton et al., 1987), which result in the uncertainty, complexity, and heterogeneity of SOM spatial distribution across different landscapes (Bai et al., 2005; Liu et al., 2014). Therefore, the rapid and accurate mapping of SOM spatial distribution has gained considerable interests to elucidate local variations in SOM for land managers and farmers.

The physical and chemical analyses of soil properties are costly, time-consuming, and require a large number of soil samples to describe soil properties at broad scales in detail (Conforti et al., 2015). Consequently, advanced technologies and emerging approaches, such as spectrum technologies and geographical statistical models, have been used in research on soil properties (Harris et al., 2010; Kumar et al., 2012b; Viscarra Rossel and Chen, 2011). Field and laboratory experiments

* Corresponding author.
E-mail address: zht@mail.hzau.edu.cn (H. Zhang).

http://dx.doi.org/10.1016/j.geoderma.2016.10.010

have also shown that visible (Vis, 400–700 nm) and near-infrared (NIR, 700–2500 nm) diffuse reflectance spectroscopy (DRS) is an efficient tool for the rapid and economical prediction of soil properties (Nocita et al., 2014). The Vis and NIR (VNIR) spectrum contains integrative soil information, such as color, amount of water, and particle size, whereas organic matter content and composition can be determined by using wavelength-specific electromagnetic radiation absorptions. These parameters are primarily associated with the resulting vibrational energy transitions of the dominant molecular bonds in the absorption of soils in VNIR spectral regions (Vasques et al., 2008; Viscarra Rossel et al., 2006). Moreover, VNIR DRS can predict SOM because of its distinct absorption features in VNIR regions that are caused by various chemical bonds, such as C—H, C—C, C═C, C—N, and O—H (Jia et al., 2014; Vohland et al., 2014; Wang et al., 2013a). Predicting soil properties requires empirical equations based on the relationship between reflectance spectra and reference soil properties in the spectral library.

Partial least squares regression (PLSR), as one of the classical and most commonly used models, is frequently applied to solve the redundancy and multicollinearity of spectral reflectance and to predict soil properties in the soil-spectral field (Ben-Dor et al., 1997; Chang et al., 2001; Dalal and Henry, 1986; McCarty et al., 2002). PLSR is a non-spatial regression model, and thus, the spatial structure of soil samples and their auxiliary variables are not considered when it is applied to predict soil attributes. However, many studies have demonstrated that significant spatial correlation among samples may be expected, particularly when soil samples are collected at field scales or at small intervals (Wan et al., 2011; Wang et al., 2013b; Zhang et al., 2011). Meanwhile, theory of regionalized variable, including the random aspect and the structured aspect of the spatial variables as the fundamental basis of geostatistics, can provide adequate information and theoretical foundation in expressing structural properties and predicting soil properties (Matheron, 1963). Thus, the spatial characteristics of soil properties and their auxiliary variables should be considered when constructing prediction models.

Rivoirard (2002) noted that the key assumption of the regression model was that no spatial dependence existed between the auxiliary variable and the residuals of the linear regression of the target variable on auxiliary variables. That is, soil property observations and their auxiliary variables should be independent of each other to guarantee the optimality of the prediction model and follow the foundation assumption of the regression model (Conforti et al., 2015). However, spatial dependence among soil samples has not received significant attention when estimating soil properties through VNIR DRS, particularly in regional studies (Ge et al., 2007; Gogé et al., 2014). Thus, neglecting the issue of spatial dependence can render VNIR DRS models as suboptimal.

Non-spatial linear regression models do not consider the spatial autocorrelation of soil properties and VNIR DRS. Several scholars attempted to improve the robustness and accuracy of prediction models by applying spatial models or combination of geostatistical models. Geographically weighted regression (GWR) (Fotheringham, 1997), is a local, spatial, statistical technique used to accurately analyze spatial non-stationarity and understand spatial variability in data. Although GWR has not been used to predict soil properties via VNIR DRS, this technique has been used to construct prediction models that consider environmental factors as auxiliary variables. Zhang et al. (2011) used ordinary Kriging (OK), inverse distance weighted (IDW), multiple linear regression (MLR), and GWR to predict soil organic carbon (SOC) based on rainfall, land cover, and soil type. Kumar et al. (2013) compared the model performances of MLR and GWR in predicting SOC based on terrain attributes, climate, bedrock geology, and land use data. Wang et al. (2013b) used GWR and co-Kriging (COK) to draw the map of soil total nitrogen (TN) using multiple environmental variables. All of these studies demonstrated the excellent performance of GWR in considering the spatial characteristics of input variables compared with the other selected models. Meanwhile, non-stationary GWR coefficients can show the varying influence degrees of environmental factors on soil properties in different geographical locations. This method can further indicate the importance of the spatial characteristics of input variables in constructing a steady and effective soil model. Soil spectra can reflect comprehensive soil properties, which exhibit strong spatial dependence and heterogeneity. However, few studies have explored the spatial dependence of spectral reflectance in predicting soil properties and discussed the feasibility of using spatial models to predict soil properties via VNIR DRS. Thus, GWR will be used in this study to explore the spatial non-stationarity of soil spectra in predicting SOM and model performance.

Meanwhile, the question regarding the independent and identical distribution of the model residuals has been considered when predicting soil properties via VNIR DRS. Conforti et al. (2015) used PLSR with a linear mixed effect model (LMEM) to ensure that the model residuals were independent and identically distributed based on the laboratory-based soil VNIR spectra. Their results showed that this approach could improve the prediction of SOM, and that LMEM could consider the influence of the model residuals. Ge et al. (2007) used the regression-Kriging (RK) method to account for spatial dependence among soil samples and aid in developing prediction models. The results showed that RK predicted most soil properties, including soil particle size distribution (clay and sand) and soil chemical analysis [Ca, Mg, Na, cation-exchange capacity (CEC), pH, P, and Zn] based on laboratory soil reflectance spectra. Bilgili et al. (2011) used COK and RK to predict soil properties [$CaCO^3$, SOM, Ca, K, Mg, Na, CEC, pH, electrical conductivity (EC), clay, and sand] via VNIR DRS. Both techniques improved the predicted accuracy of soil variables compared with PLSR and OK. These studies demonstrated the importance of keeping model residuals independent and the outstanding performance of comprehensive models in predicting soil properties based on soil VNIR DRS. GWR Kriging (GWRK) is an extension of GWR that combines local regressions (i.e., GWR) with the Kriging of regression residuals (Harris et al., 2010) to improve the estimations of soil properties at regional scale. Kumar et al. (2012a) applied GWRK and RK to predict SOC using multiple environmental variables, including temperature, precipitation, elevation, slope, geology, land use, and normalized difference vegetation index. Liu et al. (2015) used GWR and GWRK to predict the spatial distribution of SOC density based on nine environmental variables. These studies showed the outstanding extension of GWRK in considering model residuals from GWR. Although GWRK has been applied in ecology and environment fields by continuous or discontinuous variables, minimal research on the application of GWRK based on soil spectra can be found. Meanwhile, the spatial autocorrelation of the spectral reflectance and the spatial dependence of the model residuals should not be ignored in estimating soil properties (Bilgili et al., 2011; Conforti et al., 2015; Ge et al., 2007). Thus, in this study, RK and GWRK will be used to explore the spatial autocorrelation of the soil spectra and the spatial dependence of the model residuals in predicting SOM. To remove the multicollinearity and reduce the dimensionality of spectral reflectance, the principle components (PCs), which are transformed from soil spectra via PLSR, will be used as explanatory variables to construct PLSR Kriging (PLSRK), partial least squares GWR (PLS–GWR) model, and PLS–GWR Kriging (PLS–GWRK) model. We hope this research can fill in the gaps in the application of spatial models to predict soil properties via VNIR DRS, and guarantee the basic assumption of the regression models to construct steady and effective soil spectral models.

Jianghan Plain is a source of commodity grains, cotton, and edible oil in China. Most of the farmlands on this plain are influenced by human activities, such as irrigation, fertilizer use, and plowing, which can change the spatial characteristics of soil properties at field scale. Rapid and accurate evaluation of soil fertility is important in precise agriculture management and development. This study aims to (i) explore appropriate preprocessing methods to analyze soil spectral reflectance, (ii) explore the spatial autocorrelation of SOM and spectral reflectance, (iii) construct prediction models for SOM based on traditional