remote sensing

MDPI

Type of the Paper (Article)

Application of Spectrally Derived Soil Type as Ancillary Data to Improve the Estimation of Soil Organic Carbon by Using the Chinese Soil Vis-NIR Spectral Library

Supplementary Material

Yi Liu [1,2], Zhou Shi [3], Ganlin Zhang [2], Yiyun Chen [1,3,4,*], Shuo Li [5], Yongshen Hong [1], Tiezhu Shi [6], Junjie Wang [6] and Yaolin Liu [1,*]

[1] School of Resource and Environment Science, Wuhan University, 129 Luoyu Road, Wuhan 430079, China; liuyi2010@whu.edu.cn (Y.L.); hys@whu.edu.cn (Y.H.)
[2] State Key Laboratory of Soil and Sustainable Agriculture, Institute of Soil Science, Chinese Academy of Sciences, Nanjing 210008, China; glzhang@issas.ac.cn
[3] Institute of Applied Remote Sensing and Information Technology, College of Environmental and Resource Sciences, Zhejiang University, Hangzhou 310058, China; shizhou@zju.edu.cn
[4] Collaborative Innovation Center of Geospatial Technology, Wuhan University, Wuhan 430079, China
[5] The college of Urban & Environmental Science, Central China Normal University, 152 Luoyu Road, Wuhan 430079, China; shuoguoguo@gmail.com
[6] Key Laboratory for Geo-Environmental Monitoring of Coastal Zone of the National Administration of Surveying, Mapping and GeoInformation & Shenzhen Key Laboratory of Spatial Smart Sensing and Services, Shenzhen University, Shenzhen 518060, China; tiezhushi@szu.edu.cn (T.S.); wjjlight@whu.edu.cn (J.W.)
* Correspondence: chenyy@whu.edu.cn (Y.C.); liuyaolin2010@163.com (Y.L.)

Received: 18 September 2018; Accepted: 03 November 2018; Published: 06 November 2018

Prepared for: Remote sensing

These 8 pages contain 2 tables and 2 figures, as described in the Table of Contents.

Table of Contents

1. Maps showing the spatial location of some samples

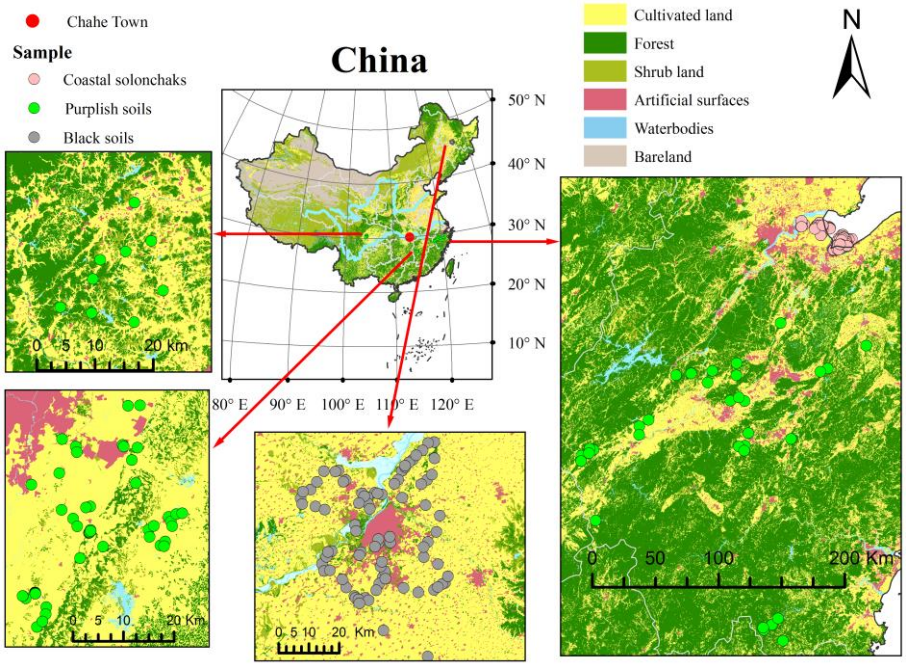

Figure S1 Location of the soil library with 515 samples in China. The location of samples from Meadow soils and Chernozems is unavailable. And some samples from Coastal solonchaks, Purplish soils and Black soils are also available.

2. Indicators of PLSR model performance

To assess PLSR model performance, we calculated five indicators, namely root-mean-square error of cross-validation (RMSE_{cv}), coefficient of determination in cross-validation (R^2_{cv}), root mean square error of prediction (RMSEP), R^2_p, and residual predictive deviation (RPD).

$$\mathrm{RMSE}_{CV} = \sqrt{\frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{n}} \tag{1}$$

$$R^2_{CV} = \frac{[\mathrm{cov}(y_i, \hat{y}_i)]^2}{\mathrm{var}(y_i)\mathrm{var}(\hat{y}_i)} \tag{2}$$

where n refers to the number of calibrated samples, y_i and $\hat{y}_i$ are the measured and estimated values of the *i*th sample in calibration set using leave-one-out cross-validation, and $\bar{y}$ is the average measured value. $\mathrm{Cov}(y_i, \hat{y}_i)$ is the covariance of measured and estimated values. $\mathrm{Var}(y_i)$ and $\mathrm{var}(\hat{y}_i)$ are the variance of measured and estimated values, respectively.

$$\mathrm{RMSEP} = \sqrt{\frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{n}} \tag{3}$$

$$\mathrm{R}^2_{\mathrm{p}} = \frac{[\mathrm{cov}(y_i, \hat{y}_i)]^2}{\mathrm{var}(y_i)\mathrm{var}(\hat{y}_i)} \tag{4}$$

where n refers to the number of validated samples, y_i and $\hat{y}_i$ are the measured and estimated values of the *i*th sample in validation set, respectively, and $\bar{y}$ is the average measured value. $\mathrm{Cov}(y_i, \hat{y}_i)$ is the covariance of measured and estimated values. $\mathrm{Var}(y_i)$ and $\mathrm{var}(\hat{y}_i)$ are the variance of measured and estimated values, respectively.

$$\mathrm{RPD} = \frac{\mathrm{SD}}{\mathrm{RMSEP}} \tag{5}$$

where SD is the standard deviation of reference values in the validation set.

3. VIP analysis

The important wavelengths used in the PLSR models of estimating SOC were measured by variable importance in the projection (VIP) scores [1]. Wavelengths with VIP scores of >1 were considered the most correlative and significant for estimating SOC concentration [2-4]. As observed in previous studies, the critical wavelengths for SOC estimation occurred at 450–1050, 1410, 1890, and 2250–2380 nm (Figure 3) [5-7]. The visible region of 450–800 nm was mainly affected by iron oxides and the dark color of SOM [8]. The other regions of 800–1050, 1410, 1890, 2250–2380 nm were caused by the overtones and combinations of fundamental vibrations of organic molecules, such as C–H, N–H, S–H, C=O, and O–H [9].

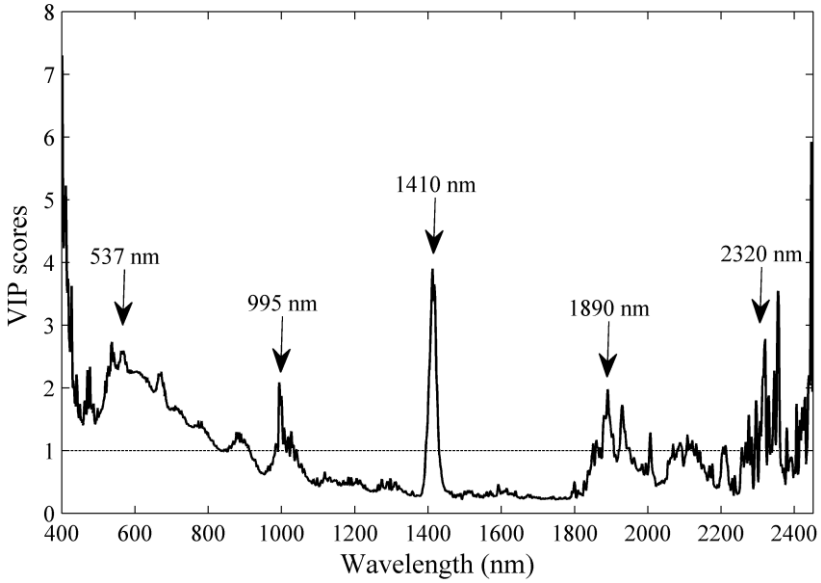

Figure S2. Variable importance projection (VIP) scores (black line) associated with the cross-validation of partial least-squares regression model for soil organic carbon concentration estimation by using laboratory spectroscopy and the entire dataset form Chinese soil spectral library. The threshold for VIP was set to 1 (horizontal dashed line).

4. Optimal spectral pretreatment for PLS-DA and PLSR

The optimal spectral pretreatments for PLS-DA and PLSR were based on agreement rate and coefficient of determination in prediction (R_p^2), respectively. Therefore, the spectral pretreatments for PLS-DA could be different from those of PLSR.

Table S1. Spectral pretreatment for PLS-DA and PLSR.

Model	Strategy	Soil type	LVs	Pretreatment
PLSR	Strategy I	Five soil types	12	Lg+SG+MC
	Strategy II, Strategy III	Coastal solonchaks	3	SNV+MC
		Meadow soils	7	SNV+SG+MC
		Chernozems	11	SNV+SG+MC
		Black soils	16	MSC+SG+MC
		Purplish soils	4	None
PLS-DA	-	Five soil types	12	Lg+MC

5. Misclassifying six samples from Coastal solonchaks to Meadow soils and Chernozems

To study how the misclassification between two soil types with evident spectral difference affected subsequent SOC estimation, six validation samples (20%) from Coastal solonchaks were allocated to Meadow soils and Chernozems (SI Figure S2). The R_p^2 of Coastal solonchaks decreased greatly from 0.51 to 0.31 when 20% Coastal solonchaks samples were misclassified to Meadow soils and to 0.23 when 20% Coastal solonchaks samples were misclassified to Meadow soils (Table 3 and SI Table S2).

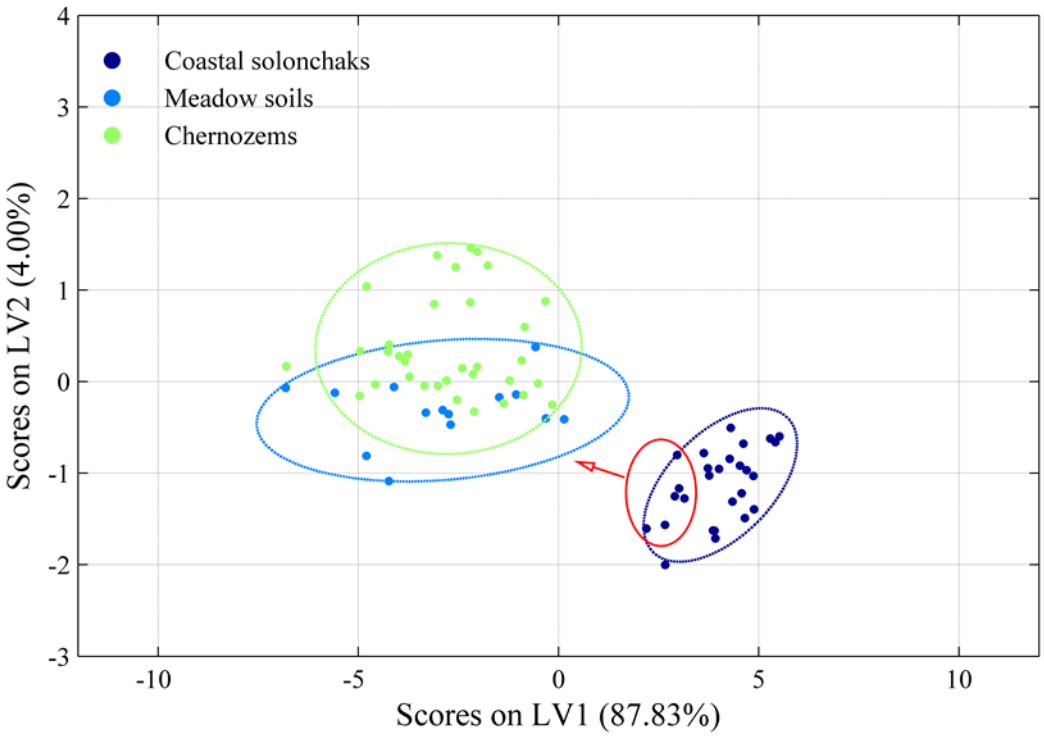

Figure S3. Scatter diagram of scores on latent variable 2 (LV2) plotted against latent variable 1 (LV1) for validation samples in partial least squares discriminant analysis (PLS-DA) models. The six samples in red ellipse were selected to be misclassified to Meadow soils and Chernozems. The other three ellipses (blue, green, and dark blue) were the 90% confidence ellipse for each soil type.

Table S2. Model performance when six samples from Coastal solonchaks were misclassified to Meadow soils and Chernozems.

Misclassification			**Coastal solonchaks**		
From	To	Count	$\mathbf{R_p^2}$	**RMSEP**	**RPD**
Coastal solonchaks	Meadow soils	6	0.31	2.91	0.81
Coastal solonchaks	Chernozems	6	0.23	3.44	0.69

6. Stratification with different number of soil types

When the number of soil types varies from 5 to 16, stratification still could improves the performance of SOC estimation using vis-NIR spectroscopy, see Figure S4.

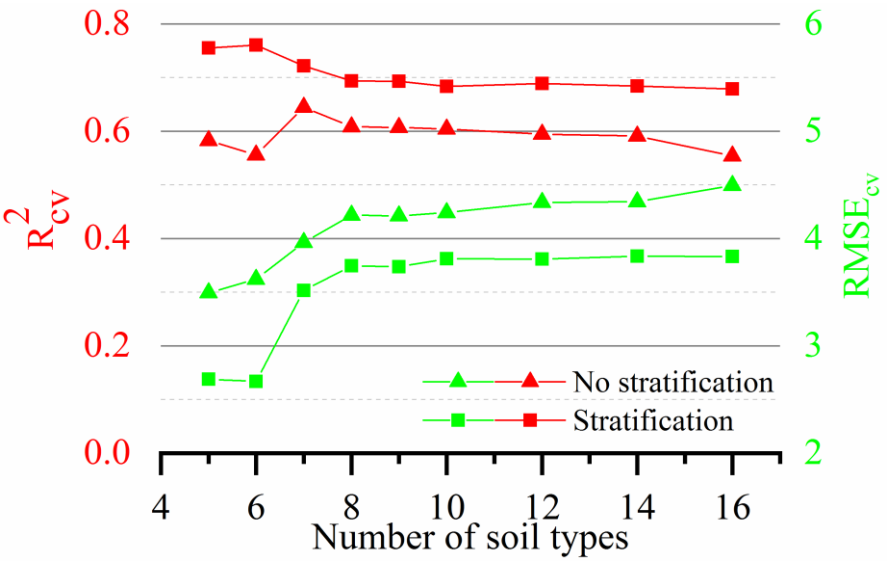

Figure S4. Performance of SOC models stratified by soil type when the number of soil type varies from 5 to 12.

Reference

1. Chong, I.G.; Jun, C.H. Performance of some variable selection methods when multicollinearity is present. *Chemometrics and Intelligent Laboratory Systems* **2005**, *78*, 103-112.
2. Trap, J.; Bureau, F.; Perez, G.; Aubert, M. Pls-regressions highlight litter quality as the major predictor of humus form shift along forest maturation. *Soil Biology and Biochemistry* **2013**, *57*, 969-971.
3. Cécillon, L.; Cassagne, N.; Czarnes, S.; Gros, R.; Brun, J.-J. Variable selection in near infrared spectra for the biological characterization of soil and earthworm casts. *Soil Biology and Biochemistry* **2008**, *40*, 1975-1979.
4. Tenenhaus, M. *La régression pls: Théorie et pratique*. Editions Technip: 1998; p 254.
5. Li, D.; Chen, X.; Peng, Z.; Chen, S.; Chen, W.; Han, L.; Li, Y. Prediction of soil organic matter content in a litchi orchard of south china using spectral indices. *Soil and Tillage Research* **2012**, *123*, 78-86.
6. Vasques, G.; Grunwald, S.; Sickman, J. Comparison of multivariate methods for inferential modeling of soil carbon using visible/near-infrared spectra. *Geoderma* **2008**, *146*, 14-25.
7. Daniel, K.; Tripathi, N.; Honda, K.; Apisit, E. Analysis of vnir (400–1100 nm) spectral signatures for estimation of soil organic matter in tropical soils of thailand. *International Journal of Remote Sensing* **2004**, *25*, 643-652.
8. Xu, D.; Ma, W.; Chen, S.; Jiang, Q.; He, K.; Shi, Z. Assessment of important soil properties related to chinese soil taxonomy based on vis–nir reflectance spectroscopy. *Computers and Electronics in Agriculture* **2018**, *144*, 1-8.
9. Stenberg, B.; Viscarra Rossel, R.A.; Mouazen, A.M.; Wetterlind, J. Chapter five-visible and near infrared spectroscopy in soil science. Advances in agronomy 2010, 107, 163-215.