Quantitative Analysis of Soil Organic Matter Using Vis---NIR Spectroscopy

Tianyin Cao Hui Tan

School of Automotive Engineering, Hunan Vocational College of Finance and Industry, Hengyang, Hunan, 421002, China

Abstract

In response to the demands of modern agriculture and "precision agriculture," soil organic matter has emerged as a critical indicator of soil fertility, necessitating the development of rapid and environmentally friendly detection technologies. Soil organic matter primarily consists of carbohydrates and nitrogen-containing compounds. In this study, a rapid analytical model for soil organic matter was established using soil visible–near-infrared (Vis-NIR) spectroscopy. To ensure the objectivity of the model, 50 soil samples were randomly selected to form a validation set, while the remaining samples constituted the modeling set. Within the modeling set, a calibration subset (52 samples) and a prediction subset (51 samples) were partitioned based on the similarity of the chemical value distributions among the samples. The moving-window partial least squares (MW-PLS) method was employed to optimize the wavelength combinations. The optimal MW-PLS model utilized a wavelength range of 2122–2480 nm, with 7 latent variables (LVs) and 180 wavelengths. In the independent validation, the optimal model yielded a root mean square error (SEPV) of 0.268 g·kg^{-1} and a correlation coefficient (R^2, V) of 0.928. The results demonstrate a high correlation and low error between predicted and measured values, confirming the feasibility of employing rapid and green Vis-NIR spectroscopy for soil organic matter analysis. This technique holds promising potential for large-scale soil monitoring applications.

Keywords

visible–near-infrared spectroscopy; sliding-window partial least squares; optimal wavelength combination; soil organic matter

基于Vis-NIR光谱技术的土壤有机质定量分析

曹添银 谭辉

湖南财经工业职业技术学院汽车工程学院，中国 · 湖南 衡阳 421002

摘要

随着现代农业以及“精准农业”的要求，土壤有机质作为土壤肥沃度的重要评价指标，急需一种快速绿色的检测技术。土壤有机质主要由碳水、含氮化合物等组成。本研究，基于土壤可见-近红外（Vis-NIR）光谱对土壤有机质建立了快速分析模型。为了模型的客观性，从153个土壤样品中，随机选出检验集（50），剩余的为建模集，基于样品化学值分布的相似性进行建模集中定标集（52）与预测值（51）的划分。移动窗口偏最小二乘法（MW-PLS）被用于优选波长组合。最优MW-PLS模型对应的波长组合、潜变量数（LV）、波长个数分别为2122-2480 nm、7、180。独立检验中，最优模型的均方根误差（SEPV）、相关系数（RP, V）分别为0.268（g kg^{-1}）、0.928。结果表明，预测值与化学值之间的具有高的相关性和低的误差；通过快速绿色的Vis-NIR光谱技术对土壤有机质分析的可行性。Vis-NIR光谱技术有望应用到大规模的土壤检测中。

关键词

可见-近红外光谱分析；移动窗口偏最小二乘；波长组合优选；土壤有机质

1 引言

土壤是农业生态环境的根基，土壤中有机质的含量在一定范围内，决定着土壤的肥力，从而影响着农作物的产量[1]。传统的测定方法为灼失量法、重铬酸钾法，需要化学试剂、费时、专业操作性性、不利于快速在线检测。而随着“精准农业”的要求，亟需一种快速实时精准的测量方法。可见-近红外（Vis-NIR）光谱具有简便、快速分析多种成分复杂样品的优势，在农业领域具有重要应用潜力。近年来，国内外已有相关的土壤近红外（NIR）光谱研究[2-4]。NIR光谱具有重叠、平坦、缺乏解释性，组分吸收峰不显著等特点。而土壤是具有多组分的复杂复合物，由于存在其他未知组分的干扰，单纯采用目标组分的吸收波长对多组分复杂样品进行光谱预测，其效果往

【基金项目】2025年度湖南省教育厅科学研究项目“Vis-NIR光谱结合化学计量学用于汽车机油多品牌鉴别与掺杂判别分析”（项目编号：25C0910）。

【作者简介】曹添银（1988-），男，中国湖南衡东人，硕士，讲师，从事职业教育研究。

往不佳。选择建模效果好的适当的波长组合是近红外光谱分析的核心关键问题。

偏最小二乘法（PLS）是经典的光谱分析方法，它结合了主成分分析、典型相关分析、多元线性回归分析等数学优化技术。潜变量数是PLS的一个重要参数指标，其表示样品信息的光谱综合变量个数，合理选择PLS潜变量数对于充分利用光谱信息和消除噪声非常重要[5]。

NIR光谱是包含数百上千波长的高维数据。以当前的科学计算能力，无法采取穷举方式实现任意波长组合的优选。以上述PLS为例，遍历全谱范围内（波长数：n）所有的波长组合，共计$C_n^1+C_n^2+\cdots+C_n^n=2^n-1$个组合，分别建立PLS模型，并依据建模效果选择最优波长组合，其算法过程难以实现。因此，需采用适当的局部优化策略建模。

用于建模的局部优化策略有很多种，移动窗口偏最小二乘法（MW-PLS）是一种经典的性能良好的连续波段筛选方法[6-9]。它使用初始波长（I）和波长数（N）作为循环参数选择所有子波段，分别建立PLS模型，依据建模效果选择最优波段。在全谱范围（波长数：n）内，共计$n(n+1)/2$个连续子波段。该算法的计算复杂度不高，普通的计算机可以在短时间内完成。

定标与预测集不同的划分会使建立的模型预测效果及参数产生变动，导致结果的不稳定。在本研究中，首先随机选取部分样品作为检验集，剩余的样品作为建模集；为了获取客观、稳定、实用的模型，对于建模集的样品基于样品的化学值分布的相似性和均匀性进行定标与预测集划分，这也是为了避免评价失真。本研究使用Vis-NIR光谱结合经典的连续波长筛选MW-PLS方法对土壤有机质进行快速测定。

2 实验与方法

2.1 材料、仪器和方法

本实验收集总共153个相同类型的土壤样品，然后再干燥后研磨，并使用0.25 mm土壤筛筛分样品。通过使用标准重铬酸钾（$K_2Cr_2O_7$）氧化土壤分析方法测量每个样品的有机质含量的化学值，并用作Vis-NIR光谱分析的参考值[10]。所有样品的实际值范围为1.24至4.70 (g kg^{-1})，平均值和标准偏差分别为2.45和0.63(g kg^{-1})。

使用XDS Rapid Content™光栅光谱仪（FOSS；丹麦）配备2毫米比色皿传输附件对光谱数据进行采集；光谱扫描范围设定在400-2498 nm，光谱分辨率设置为2 nm；检测器分别为Si（400-1100 nm）和PbS（1100-2498 nm）。所有光谱均在25±1℃和46％±1％相对湿度下记录。由于单次测量存在的随机误差可能较大，因此将每个样品测量三次，最终采取三次测量的平均值进行建模分析。

2.2 样品划分和评价指标

用于定标、预测和检验集的实验设计：首先，从利用程序随机选择部分样品作为检验集，那么剩下的样品则为建模集。为了客观、有效的评价模型，再把建模集划分为定标集与预测集。依据最大最小放入定标集，次大次小放入预测集的准则把进行划分。

样品具体划分如下：随机划分建模集（103）与检验集（50）；基于样品化学值分布的相似性及均匀性进行定标集（52）与预测集（51）的划分。定标集与预测集的样品化学值分布图1所示。

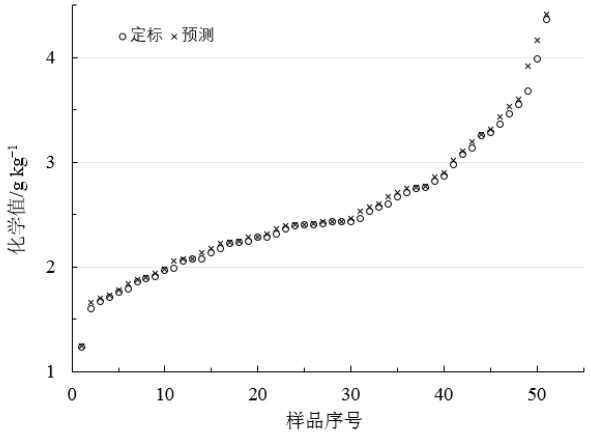


图1 定标-预测集样品化学值分布图

均方根误差（SEP）和预测的相关系数（RP）作为评价指标，SEP越小、RP越大则表明预测效果越好，分别在建模中表示为SEPM，RP, M，在验证中表示为SEPV，RP, V。

2.3 MW-PLS方法

移动窗口偏最小二乘算法 (MW-PLS)是一种基于PLS的连续型波长选择方法[6-9]。它使用初始波长（I）和波长数（N）作为循环参数来实现所有有限的单向波段中的遍历选择，使用移动窗口方法在原始光谱范围中选出多组子波段建立PLS模型，并基于预测的效果，对每一个PLS模型优选出潜变量（LV）的值。最后，将全部子波段的局部最优PLS模型的预测效果进行比较，根据建模预测均方根误差（SEPM）和建模预测相关系数（RP,M）的值以确定全局最优模型和波段。

对于具有1050个波长的400-2498nm、波长间隔为2nm的搜索范围，参数I，N，LV设定为400≦I≦2498，1≦N≦2498，1≦LV≦20，并且要求窗口不能超出光谱的范围，即需要条件I+2×(N-1)≦E得到满足。

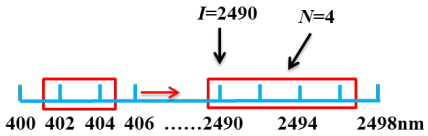


图2 MW-PLS算法原理图

3 结果与讨论

整个扫描区域（400-2498 nm）的53个土壤样品的近红外光谱如图3所示。作为对比，全谱数据首先用于建立1050波长的PLS模型，相应的SEPM，RP, M分别为0.279（g kg⁻¹）和0.937。

3.1 MW-PLS方法

为了比较，全光谱区域（400-2498 nm）首先用于MW-PLS算法的搜索范围。此外，根据物理学中的自然波段划分，可见光谱区域（400-780 nm）、短波近红外区域（780-1098 nm）、长波近红外区域（1100-2498 nm）和整个近红外区域（780-2498 nm）也用于搜索范围以执行MW-PLS算法。

使用2.3节中提到的方法，选择了五个搜索范围内的最佳MW-PLS模型。整个NIR区域（780-2498 nm）的最佳I、N和LV分别为2122 nm，180和7。相应的波段为2122-2480 nm，SEPM、RP, M分别提高到0.244（g kg⁻¹）、0.960。对于五个搜索范围，所有预测效果（SEPM, RP, M）和采用的参数总结在表1中。

从表1中可以看出，最优模型的波长主要集中在长波（1100-2498 nm）区域。在全谱区域利用MW-PLS进行波段选择时，最好的波段处于（530-2488 nm），其SEPM与全局最优模型一致，但是模型的复杂度明显更高，因此不考虑。

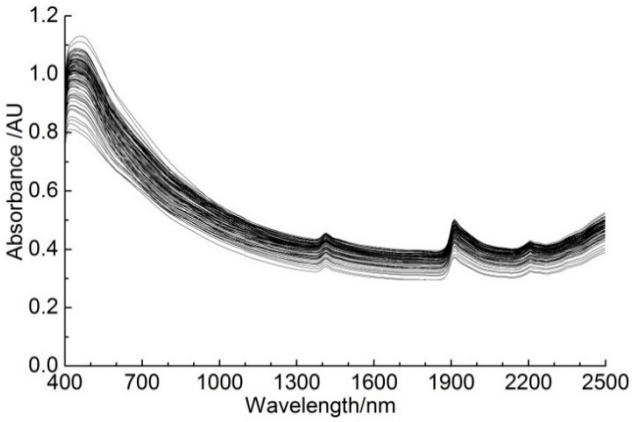


图3 土壤样品的近红外光谱图

表1 五个搜索范围内最优MW-PLS模型的预测效果

Search range	Waveband	N	LV	SEPM	RP,M
400-780	434-772	170	11	0.262	0.949
780-1100	782-1060	140	5	0.316	0.929
1100-2498	2122-2480	180	7	0.244	0.960
780-2498	2122-2480	180	7	0.244	0.960
400-2498	530-2488	980	8	0.244	0.952

3.2 模型验证

将没有用于建模的检验集样品去检验全局最优MW-PLS模型（2122-2480 nm）。对于全局最优模型来说，相应的检验评估值SEPV、RP, V分别为0.268（g kg⁻¹）、0.928。对于检验集样品的土壤有机质的预测值和实测值之间的相关性如图4所示。从图中可以看出该模型的光谱预测值与实际值高度相关，这代表着模型预测精度高。

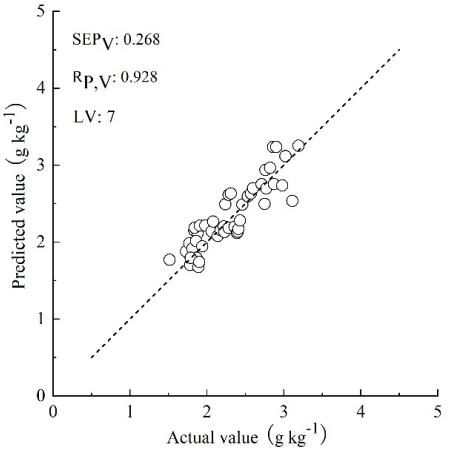


图4 最优全局MW-PLS模型检验时预测值与实际值的相关性图

4 结论

土壤中有机质的含量是判断土壤肥沃度的重要指标。土壤有机质传统的检测方法已无法满足“精准农业”的要求。Vis-NIR光谱具备快速绿色的技术优势。但由于其谱峰重叠、平坦的特点，需要和化学计量学方法结合来提升光谱的预测效果。除此之外，样品的定标预测集的划分也会影响建立的模型的参数及结果的波动。为了得到客观、合理的检验效果，基于样品化学值相似性进行定标-预测集的划分。

利用MW-PLS方法优选出了土壤有机质的最优Vis-NIR波段为1100-2498 nm，SEPM、RP, M分别为0.244（g kg⁻¹）、0.960，相比于全谱PLS结果，SEPM降低了12.5%，且模型的复杂度大幅度降低。结果表明，Vis-NIR光谱应用土壤分析的可行性，同时也可以为其他领域的光谱分析提供参考价值。

参考文献

[1] 张佳麒,王翠,王明明,等.长期秸秆还田条件下黑土有机碳的动态变化及其组分积累特征[J].土壤学报,1-13 [2024-12-26].

[2] Rossel R V, Walvoort D J, McBratney A B, et al. Visible, near infrared, mid infrared or combined diffuse reflectance spectroscopy for simultaneous assessment of various so il properties[J]. Geoderma, 2006, 131(1-2): 59-75.

[3] 陈奕云, 漆锟, 刘耀林等. 顾及土壤湿度的土壤有机质高光谱预测模型传递研究 [J]. 光谱学与光谱分析, 2015, 35(6): 1705-1708.

[4] 李美婷, 武红旗, 蒋平安等. 利用土壤的近红外光谱特征测定土壤含水量 [J]. 光谱学与光谱分析, 2012, 32(8): 2117-2121.

[5] 褚小立. 化学计量学方法与分子光谱分析技术 [M]. 北京: 化学工业出版社, 2011.

[6] J.H. Jiang, R.J. Berry, H.W. Siesler, Y. Ozaki, Wavelength interval selection in multicomponent spectral analysis by moving window partial least-squares regression with applications to mid-infrared and near-infrared spectroscopic data, Anal. Chem. 74 (2002) 3555–3565.

[7] H.Z. Chen, T. Pan, J.M. Chen, Q.P. Lu, Waveband selection for NIR spectroscopy analysis of soil organic matter b ased on SG smoothing and MWPLS methods, Chemometrics and Intelligent Laboratory Systems, 107 (1), pp. 139-146, 2011.

[8] T. Pan, J.M. Liu, J.M. Chen, G.P. Zhang, Y Zhao, Rapid determination of preliminary thalassaemia screening indicators based on near-infrared spectroscopy with wavelength selection stability, Analytical Methods, 5, pp. 4355-4362, 2013.

[9] J.M. Chen, T. Ai, T. Pan, L.J. Yao, F.G. Xia, AO-MW-PLS method applied to rapid quantification of teicoplanin wi th near-infrared spectroscopy, Journal of Innovative Optical Health Sciences, 10 (1), pp. 1650029 (13 pages), 2016.

[10] GB 9834-88, 土壤有机质的测定法（重铬酸钾氧化法）. 中华人民共和国国家标准。