

应用红外光声光谱技术及支持向量机模型测定 土壤有机质含量*

曾胤 陆宇振 杜昌文[†] 周健民

(土壤与农业可持续发展国家重点实验室(中国科学院南京土壤研究所), 南京 210008)

摘要 快速测定土壤有机质含量对作物生产和土壤肥力评价具有重要意义, 红外光声光谱技术的应用为土壤有机质快速测定提供了可能。本研究以江苏省南京市溧水区水稻土土样为材料, 探究了红外光声光谱技术在有机质测定中的应用。采用主成分分析、偏最小二乘和独立成分分析, 分别提取了土壤光谱的主成分、偏最小二乘潜变量和独立成分, 并以提取的信息输入支持向量机, 从而构建了三种支持向量机校正模型。同时, 偏最小二乘也被用于建立校正模型, 作为支持向量机模型的对照。预测结果表明, 基于独立成分的支持向量机模型效果最好, 预测相关系数 R^2 、均方根误差 $RMSEP$ 和实际测量值的标准差与光谱模型预测值标准差的比值即 RPD 值分别为 0.808、0.575 和 2.28。 F 检验表明, 该模型显著优于基于主成分的支持向量机模型, 但与基于偏最小二乘潜变量的支持向量机模型, 以及经典偏最小二乘模型没有显著差异。 t 检验表明, 各校正模型对有机质的预测结果与化学测定结果没有显著差异。因此, 红外光声光谱技术为土壤有机质的快速测定提供了新的技术手段。

关键词 土壤有机质; 红外光声光谱; 支持向量机; 定量预测

中图分类号 S158.2 **文献标识码** A

有机质是作物生长的重要营养来源^[1], 可以供给作物所需的各种营养元素(如氮、磷等)。同时, 有机质也是土壤质量的重要指标^[2]。目前, 土壤有机质测定主要采用重铬酸盐氧化法和干式燃烧法^[3]。然而, 这两种方法比较耗时、分析成本高, 且均为破坏性测试, 不利于快速原位地获取土壤有机质信息, 尤其对于大规模土壤样本分析。

红外光谱技术是一项快速无损的分析技术, 已被应用于测定包括有机质在内的各种土壤肥力参数^[4-5]。其中, 应用较多的是透射光谱、漫反射光谱、衰减全反射光谱。在光谱测定时, 透射光谱需要压片制样, 以确保样品均一可透。对于高吸收的土壤样本, 则需要用大量的卤化物对样本进行稀释, 制样比较费时, 而且得到的透射光谱一般难以定量^[6]。漫反射光谱对制样要求不高, 但受样品粒度大小影响^[7], 而导致光谱非线性变形, 有时也需要进行压片处理^[8]。衰减全反射光谱测定时, 需要加水使土壤样品呈糊状, 以确保与漫反射晶体完好

接触。然而, 这将造成强烈的水分吸收干扰, 必须在光谱处理中加以扣除^[9]。此外, 近些年, 红外光声光谱技术(Infrared photoacoustic spectroscopy, FTIR-PAS)被应用于土壤学研究的领域^[10], 该技术基于现代量热转换和光声理论, 采样特点十分独特, 是一种无辐射弛豫的直接检测, 仅对样品吸收敏感, 适合高吸收固态样品的测试, 样品需求量少, 光散射效应小, 不受样品形态限制, 还可以实现原位剖面分析^[11]。目前, 红外光声光谱技术在土壤有机质定量预测方面, 尚报道较少。

红外光谱定量分析有赖于多元校正模型。偏最小二乘(Partial least squares, PLS)是一种经典的校正方法, 能够提取光谱中与因变量相关的潜变量, 建立回归方程^[12]。然而, PLS属于线性校正方法, 对非线性问题的处理效果并不理想^[13]。由于土壤是十分复杂的体系, 土壤中各种物质基团的吸收常常出现非线性叠加。此外, 测试环境(如温度、光照等)和仪器状态的改变也造成光谱的非线性变

* 国家自然科学基金项目(41130749)资助

[†] 通讯作者, E-mail: chwdu@issas.ac.cn

作者简介: 曾胤(1988—), 男, 湖北洪湖人, 硕士研究生。E-mail: yzeng@issas.ac.cn

收稿日期: 2013-11-11; 收到修改稿日期: 2014-01-26

化^[14]。因而,选择非线性校正方法测定土壤性质参数,有望得到更好的预测结果。支持向量机(Support vector machine, SVM)是一种可用于处理非线性系统的机器学习算法^[15-16]。该方法建立在统计学习理论(Statistical learning theory, STL)、维理论(Vapnik-Chervonenkis, VC)和结构风险最小化(Structural risk minimization, SRM)原理的基础上,根据有限样本的信息,在模型复杂性和学习能力之间寻求最优折中,以使构建模型的泛化能力最强。对于非线性和高维数据空间的模式识别问题^[17-18],SVM已表现出优秀的学习能力。近些年,SVM也已成功推广至函数逼近和定量预测等领域^[19]。

本研究以江苏省溧水区水稻土样本为实验材料,采用红外光声光谱技术,结合支持向量机建立土壤有机质的定量校正模型。SVM的输入形式直

接影响模型效果。由于全谱输入计算量过大,且易发生模型过拟合。本文分别采用主成分分析(Principal component analysis, PCA)、偏最小二乘(Partial least squares, PLS)和独立成分分析(Independent component analysis, ICA)提取土壤光谱信息,然后以提取信息输入支持向量机建模。此外,经典偏最小二乘方法也应用有机质预测,作为支持向量机模型的对照。本研究,旨在为红外光声光谱技术在土壤有机质快速预测中的应用打下基础。

1 材料与方法

1.1 土壤样品与光声光谱采集

土壤样品为水稻土,采自于江苏省南京市溧水区水稻田。土壤采集以农户田块为单位随机采集,200个土壤样品的采样点分布如图1所示。

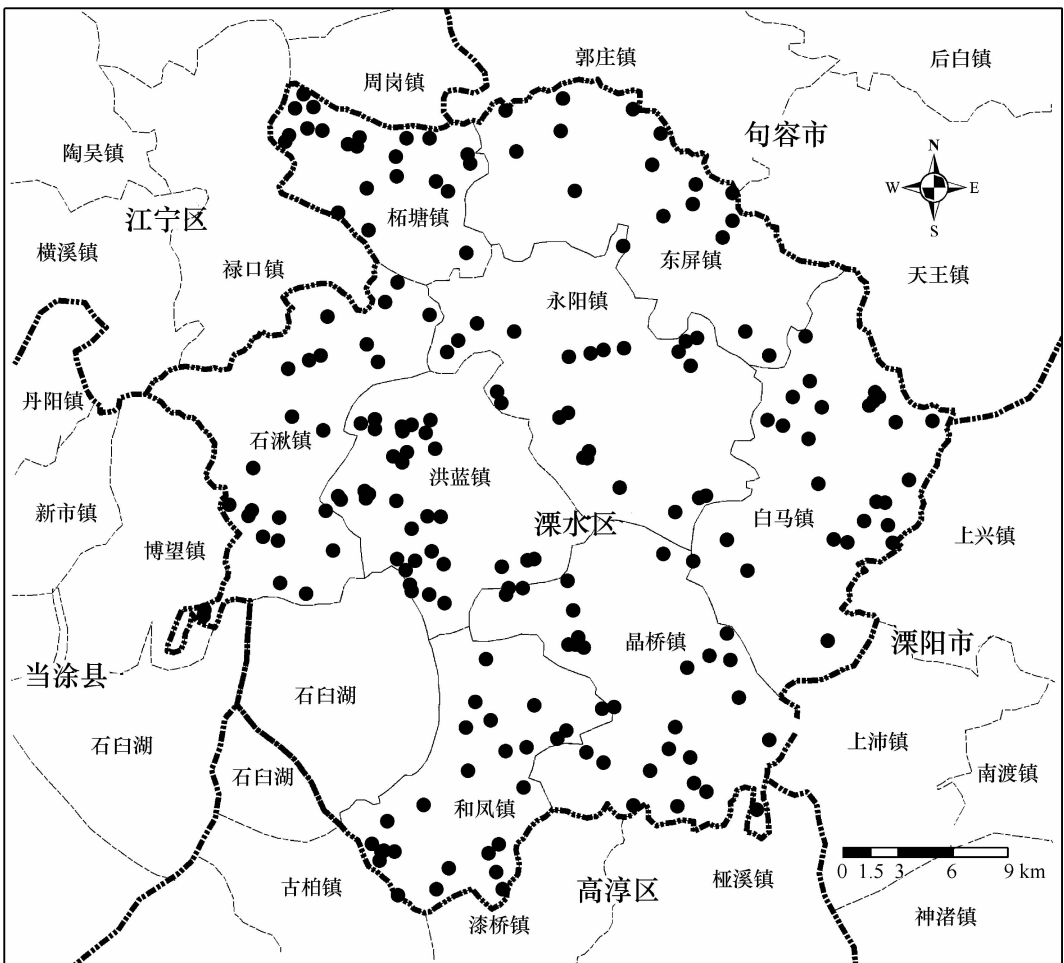


图1 土壤采样点分布图

Fig. 1 Distribution of sampling sites

土壤风干后过 2 mm 筛,采用重铬酸钾氧化法测定土壤有机质含量。土壤样品的光声光谱采集于傅里叶变换红外光谱仪 Nicolet 6700 (Thermo Elemental, USA), 配合光声附件 PA300 (MTEC Photoacoustics, USA)。将样品置于光声池(直径 5 mm、高 3 mm),用干燥的氦气连续吹扫 10 s (10 ml min^{-1}),光谱扫描范围为 $600 \sim 4\,000 \text{ cm}^{-1}$,扫描分辨率 4 cm^{-1} ,动镜速率 0.32 cm s^{-1} ,32 次扫

描的平均光谱记为样品光谱,用碳黑背景光谱对样品光声信号强度进行标准化处理。

1.2 数据处理与模型评价

原始光谱采用 Savitzky-Golay^[20] 23 点一次多项式进行预处理,滤除高频噪声。预处理后的光谱利用 Kennard-Stone 法^[21],将土壤样本按照 3:1 划分为校正集和预测集。校正集和预测集样本分别为 150 和 50 个。土壤有机质含量的统计参数见表 1。

表 1 校正集和预测集中有机质含量的统计参数

Table 1 Statistics of organic matter content in calibration and prediction sets

样本集 Sample set	最大值 Maximum (g kg^{-1})	最小值 Minimum (g kg^{-1})	均值 Mean (g kg^{-1})	标准差 Standard deviation (g kg^{-1})	变异系数 Variation coefficient (%)
校正集 Calibration set	25.60	16.20	21.57	1.69	7.83
预测集 Prediction set	24.80	18.60	20.65	1.31	6.34

全谱数据直接输入支持向量机,计算量过大,且模型预测效果不够理想。本文利用 PCA、PLS 和 ICA,分别提取出校正集样本光谱的主成分(Principal component, PC)、偏最小二乘潜变量(Latent variable, LV)和独立成分(Independent component, IC),作为支持向量机输入,从而建立三种校正模型:PC-SVM、LV-SVM 和 IC-SVM。SVM 类型采用 nu-SVM,核函数为非线性的径向基(Radial basis function, RBF)函数。选择合适的 SVM 核参数,即惩罚参数 c 和 RBF 核函数 γ ,和支持向量机输入成分 PC、LV 和 IC 的个数,对模型预测能力影响很大。本文采用二维网格搜索算法,并结合交叉验证均方根误差(Root-mean-square error of cross validation, $RMSECV$) 确定最佳的 SVM 核参数组合。PC 和 LV 的个数也由 $RMSECV$ 确定。交叉验证采用留十法交叉验证(Leave-ten-out cross validation)^[22]。由于 ICA 提取出的 IC 并不是按照变异信息量进行排序的,目前还没有明确的标准确定 IC 的数目^[23],一般推荐将全部潜在的 IC 均提取出来^[24]。所以,为避免信息丢失,本文将校正集样本光谱的全部 IC 直接作为支持向量机的输入。

作为支持向量模型的对照,经典偏最小二乘方法也被用作有机质定量预测。偏最小二乘回归的最优潜变量数也由最小的 $RMSECV$ 对应的潜变量数确定。偏最小二乘和支持向量机的建模数据,均

采用中心化预处理。

采用实测值和预测值之间的线性相关系数(R^2)以及预测均方根误差(Root-mean-square error of prediction, $RMSEP$)来比较各模型的预测性能^[25]。实际测量值的标准差(Standard deviation, SD)与光谱模型预测值标准差($RMSEP$)的比值即 RPD (Ratio of performance to standard deviation)值用以评价模型综合预测性能^[26]。根据 Viscarra Rossel 建议^[27],在土壤光谱分析中, $RPD < 1.0$,说明模型很差,没有实用性; $1.0 < RPD < 1.5$,模型较差,应用范围很有限; $1.4 < RPD < 1.8$,模型一般,可用于定性评估; $1.8 < RPD < 2.0$,模型良好,可用于定量预测; $2.0 < RPD < 2.5$,模型很好; $RPD > 2.5$,模型优秀。同时,为进一步评价校正模型的预测能力,应用 F 检验比较两两模型间的差异显著性,两模型比较时, F 值由较大的 $RMSEP$ 的平方除以较小的 $RMSEP$ 的平方得到^[28]。此外,应用配对 t 检验评价光谱方法与常规的有机质化学测定方法之间差异的显著性^[29]。

1.3 算法实现

PLS_Toolbox 6.2 (Eigenvector Research Inc., USA) 用于计算光谱主成分和偏最小二乘潜变量。FastICA toolbox 2.5 (Department of Information and Computer Science, Helsinki University of Technology) 用于提取光谱独立成分。LIBSVM 工具箱用于支持

向量机建模。其余程序为自行编写。所有数据处理工作均在 Matlab 上完成。

2 结果与讨论

2.1 土壤红外光声光谱特征

不同土壤样本的光谱曲线变化比较一致(图2)。在4000~600 cm⁻¹波数范围内,光声光谱的吸收信息丰富,总体表现为三大吸收区间:高频区(3000~3900 cm⁻¹);中频区(1400~1900 cm⁻¹);低频区

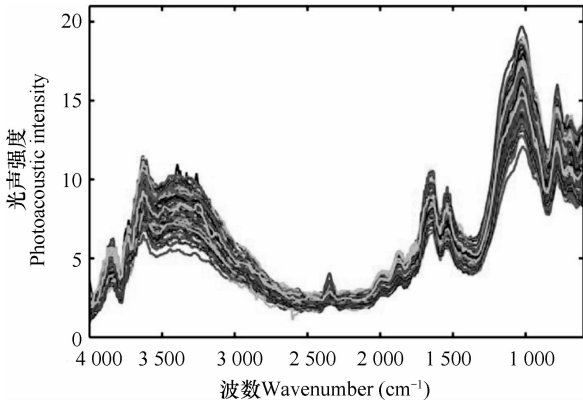


图2 土壤样本的红外光声光谱图

Fig. 2 Infrared photoacoustic spectra of soil samples

(1200~700 cm⁻¹)。土壤的光声光谱特征与其组分紧密相关。在3630 cm⁻¹处的尖峰主要由土壤中高岭石的O-H伸缩吸收导致^[30]。在3100~3500 cm⁻¹处表现明显的宽峰,说明土壤样本中可能含有蒙脱石或者蛭石^[31]。1640 cm⁻¹处的尖峰对应于C=O伸缩振动(酰胺I谱带),1540 cm⁻¹处尖峰主要对应于酰胺II谱带吸收峰(N-H弯曲振动和C-N伸缩振动的异相组合吸收)^[32],这两个尖峰均与土壤有机质有关;1024 cm⁻¹处表现出最强的吸收峰,主要由土壤中黏土矿物的Si-O伸缩振动和多糖的C-O伸缩振动吸收引起;在783 cm⁻¹的吸收峰则由土壤中芳香烃有机质的C-H面外弯曲振动引起^[33]。

由上述分析可以看出,土壤的光声光谱中包含了有机质的吸收信息,这为应用多元校正的方法测定有机质含量提供了理论基础。

2.2 基于PLS模型的土壤有机质预测

首先采用偏最小二乘方法建立PLS模型,对土壤有机质进行预测。PLS的最优潜变量数为12,预测均方根误差RMSEP和RPD分别为0.650和2.02(见表2)。图3a为该模型对预测样本测定结果的散点图。图中斜线为参考线,对应于完全准确预测

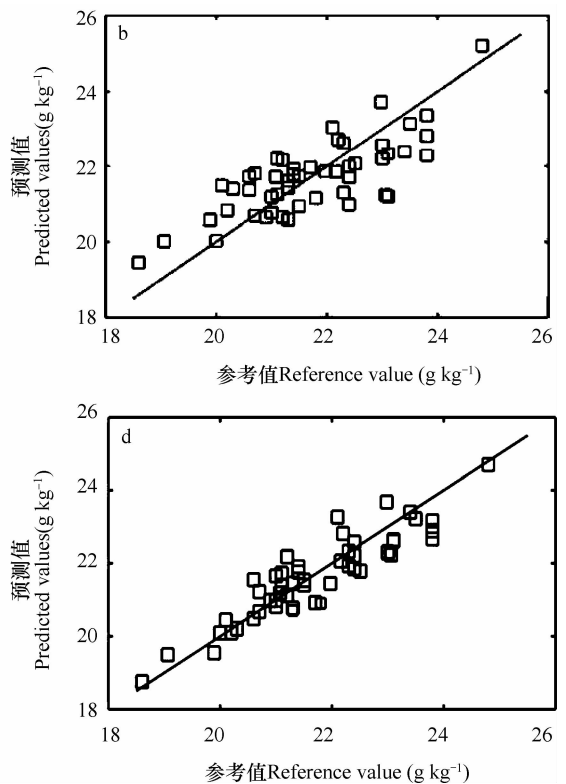
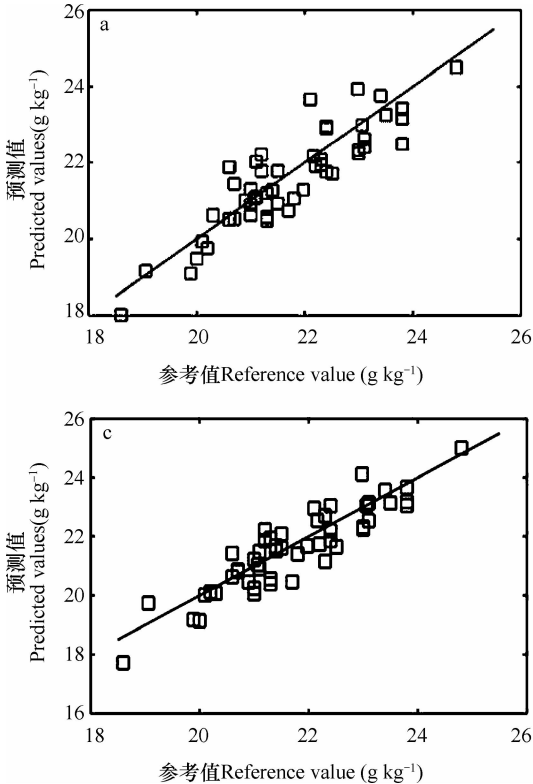


图3 模型预测散点图(a: PLS模型; b: PC-SVM模型; c: LV-SVM模型; d: IC-SVM模型)

Fig. 3 Scatter plot of reference versus predicted values from a: PLS model; b: PC-SVM model; c: LV-SVM model; d: IC-SVM model

注:图中斜线为参考线,对应于完全准确预测时的分布结果 Note: The solid lines in scatter plots correspond to the exact prediction results

时的分布结果,实际预测的结果沿着参考线分布。该模型 $RPD > 2$,说明红外光声光谱技术可以很好地预测有机质含量。对 PLS 模型的预测结果与化学测定值进行 t 检验,结果表明,该方法与化学测定方法之间没有显著差异(见表 4),进一步说明红外光声光谱方法可以用于测定土壤有机质含量。

2.3 基于 PC-SVM 模型的土壤有机质预测

将校正集光谱的主成分 PC 输入支持向量机,建立 PC-SVM 校正模型。当主成分数为 17 时,交叉验证均方根误差 $RMSECV$ 最小,因而选取前 17 个主成分作为支持向量机输入变量。根据网格搜索结果,支持向量机的惩罚参数 $c = 10$, RBF 核参数 $\gamma = 0.01$ 时,获得最优校正模型。用该模型对预测

样本进行测定,预测均方根误差 $RMSEP$ 和 RPD 值分别为 0.806 和 1.63(表 2)。根据 RPD 值的评价标准, $RPD < 1.8$,说明该模型还不能够应用于定量分析。然而,该模型的相对均方根预测误差只有 3.9%($RMSEP$ 除以预测样本化学值的平均值),这在实际定量分析中是可以接受的。而且,通过 t 检验也得到,PC-SVM 的预测结果与化学方法测定的结果不存在显著差异(见表 3)。

此外,PC-SVM 模型与 PLS 模型相比,预测精度有所下降,从预测散点图(图 3 b)也可以看出这一点。然而 F 检验表明,两模型预测性能并无显著性差异(表 3)。

表 2 不同支持向量机模型的土壤有机质预测结果

Table 2 Prediction of soil organic matter content by different support vector machine model

模型 Model	参数设置 Parameters	R^2	$RMSEP$ ($g\ kg^{-1}$)	RPD
PLS	LVN = 12	0.755	0.650	2.02
PC-SVM	PCN = 17, $c = 10$, $\gamma = 0.0100$	0.624	0.806	1.63
LV-SVM	LVN = 14, $c = 100$, $\gamma = 0.0010$	0.786	0.608	2.16
IC-SVM	ICN = 150, $c = 100$, $\gamma = 0.3162$	0.808	0.575	2.28

注: LVN 为偏最小二乘潜变量数; PCN 为主成分数; ICN 为独立成分数; c 为支持向量机惩罚参数; γ 为支持向量机径向基函数的核参数; $RMSEP$: 预测均方根误差; RPD 由预测样本参考值的标准差除以 $RMSEP$ 计算得到 Note: LVN: latent variable number of partial least squares; PCN: principal component number; ICN: independent component number; c : penalty parameter of SVM; γ : kernel parameter of radial basis function in support vector machines; $RMSEP$: root-mean-square error of prediction; RPD : the ration of performance to standard deviation, which is calculated by dividing $RMSEP$ with the standard deviation of reference values in prediction set

2.4 基于 LV-SVM 模型的土壤有机质预测

提取校正集偏最小二乘的潜变量作为支持向量机的输入变量,建立 LV-SVM 回归模型。支持向量机核参数 $c = 100$ 和 $\gamma = 0.001$, 偏最小二乘潜变量数为 14 时,得到最佳的校正模型。用预测样本对模型进行测试,得到 $RMSEP$ 和 RPD 分别为 0.608 和 2.16(见表 2)。根据 RPD 值,该模型预测性能很好。预测散点图如图 3c 所示。

与 PLS 模型和 PC-SVM 模型相比, LV-SVM 模型的预测结果均有所改善。 F 检验表明, LV-SVM

和 PLS 模型间差异不显著,而 PC-SVM 与 LV 模型间差异显著(表 3)。这说明 LV-SVM 模型显著优于 PC-SVM 模型,而与 PLS 模型效果相当。可以看出,相比于主成分,偏最小二乘潜变量有更强的预测能力,能提高模型的预测精度,因为偏最小二乘潜变量不仅携带了土壤光谱信息,也包含了有机质含量的信息。支持向量机较偏最小二乘有更强的预测能力,但这种差异还没有达到显著水平。此外, t 检验也表明, LV-SVM 模型的预测结果与化学方法的测定结果之间没有显著性差异(表 4)。

表 3 不同模型对有机质预测能力的 F 检验比较

Table 3 F-test of the models for in capability of predicting soil organic matter content

模型比较 Model comparison	PLS 与 PC-SVM	PLS 与 LV-SVM	PLS 与 IC-SVM	PC-SVM 与 LV-SVM	PC-SVM 与 IC-SVM	LV-SVM 与 IC-SVM
临界 F 值 Critical F value	1.60	1.60	1.60	1.60	1.60	1.60
计算的 F 值 Calculated F value	1.54	1.14	1.28	1.76	1.96	1.12

注: 显著水平均为 0.05。计算值大于临界值,说明差异显著 Note: The significant level is 0.05. The difference is significant when calculated F value is larger than the critical F value

2.5 基于 IC-SVM 模型的土壤有机质预测

对校正集样本进行独立成分分析 (ICA), 提取出全部独立成分 (IC), 共计 150 个。以独立成分作为支持向量机输入, 建立 IC-SVM 模型。支持向量机核参数组合为 $c = 100, \gamma = 0.3162$ 时, 获得最优校正模型。模型预测结果见表 2。根据 *RPD* 值, 模型预测效果很好, 且较 LV-SVM 模型进一步改善。图 3d 为该模型对预测样本测定结果的散点图, 对比图 3a、图 3b 和图 3c 可以看出, IC-SVM 模型的预测结果沿参考线的分布最为紧凑, 因为该模型的 *RMSEP* 值最小。*F* 检验表明, IC-SVM 模型显著优于 PC-SVM 模型, 而与 LV-SVM 模型和 PLS 模型差异不显著 (见表 3)。同样地, *t* 检验表明, IC-SVM 模型

与化学方法的测定结果差异不显著 (表 4)。

用独立成分输入支持向量机能获得满意的预测结果。根据独立成分分析的假设, 从样本光谱中提取的各独立成分对应于样本组分的“纯光谱”。这些“纯光谱”较主成分和偏最小二乘潜变量更能反映样本内在的化学构成^[34], 因而以提取的独立成分进行回归分析, 可以提高模型的预测精度^[35]。然而, 尽管 IC-SVM 模型获得了最好的预测结果, 但该模型还不够简洁, 因为其支持向量机的输入变量数高达 150。这个问题, 有待进一步采用变量筛选的办法, 如遗传算法或回归系数排序法等对独立成分进行精简, 并进一步改善模型预测性能。

表 4 不同模型有机质预测值与化学测实测值的 *t* 检验比较

Table 4 *t*-test of predicted value and measured value of soil organic matter content relative to model

方法比较 Method comparison	PLS 与化学方法 PLS Versus chemical method	PC-SVM 与化学方法 PC-SVM Versus chemical method	LV-SVM 与化学方法 LV-SVM Versus chemical method	IC-SVM 与化学方法 IC-SVM Versus chemical method
临界 <i>t</i> 值 Critical <i>t</i> value	2.01	2.01	2.01	2.01
计算的 <i>t</i> 值 Calculated <i>t</i> value	1.64	0.28	1.50	1.16

注: 显著水平均为 0.05。计算值大于临界值, 说明差异显著 Note: The significant level is 0.05. The difference is significant when calculated *t* value is larger than the critical *t* value

3 结 论

红外光声光谱技术是一种尤为独特的样品信息采集技术, 可以快速无损地对样品组分进行定量分析。目前, 该技术在土壤科学中的应用还处于起步阶段。本文以南京市溧水区水稻土样本为例, 研究了红外光声光谱技术在土壤有机质快速定量预测中的应用。土壤有机质在红外光声光谱图上有丰富的吸收信息, 这些信息为有机质定量测定奠定了基础。研究表明, 利用红外光声光谱预测的土壤有机质含量与常规化学测定的结果没有显著差异, 不同输入的支持向量机校正模型均可以有效地提取出土壤光谱信息, 其中独立成分输入的支持向量机校正模型表现出较好的预测效果, 而关于土壤光谱独立成分数的筛选以精简并改善校正模型, 还有待进一步探讨。此外, 本文没有对光谱变量进行筛选, 如果建模前剔除与有机质无关的波数, 也将有望进一步提高土壤有机质预测精度。

参 考 文 献

- [1] Arshad M A, Martin S. Identifying critical limits for soil quality indicators in agro-ecosystems. *Agriculture, Ecosystems & Environment*, 2002, 88(2): 153—160
- [2] Shukla M K, Lal R, Ebinger M. Determining soil quality indicators by factor analysis. *Soil & Tillage Research*, 2006, 87(2): 194—204
- [3] Palmer C J. Techniques to measure and strategies to monitor forest soil carbon//Kimble J M, Birdsey R, Heath L, et al. The potential of US forest soils to sequester carbon and mitigate the greenhouse effect. New York, USA: CRC Press Lewis Publishers, 2002: 73—90
- [4] Reeves J B III, McCarty G W, Reeves V B. Mid-Infrared diffuse reflectance spectroscopy for the quantitative analysis of agricultural soils. *Journal of Agricultural and Food Chemistry*, 2001, 49(2): 766—772
- [5] McCarty G W, Reeves J B. Comparison of near infrared and mid infrared diffuse reflectance spectroscopy for field-scale measurements of soil fertility parameters. *Soil Science*, 2006, 171(2): 94—102
- [6] Linker R. Application of FTIR spectroscopy to agricultural soils analysis//Nikolic G. Fourier transforms—New analytical approaches and FTIR. Croatia: InTech, 2010: 386—404

- [7] Nguyen T T, Janik L J, Raupach M. Diffusereflectance infrared fourier transform (DRIFT) spectroscopy in soil studies. *Australian Journal of Soil Research*, 1991, 29(1): 49—67
- [8] Culler S R. Diffuse reflectance infrared spectroscopy: Sampling techniques for qualitative/quantitative analysis of solids//Coleman P B. *Practical sampling techniques for infrared analysis*. BocaRaton, USA: CRC Press, 1993: 93—105
- [9] Linker R. Waveband selection for determination of nitrate in soil using mid-infrared attenuated total reflectance spectroscopy. *Applied Spectroscopy*, 2004, 58(11): 1277—1281
- [10] Du C W, Zhou J M. Application of infrared photoacoustic spectroscopy in soil analysis. *Applied Spectroscopy Reviews*, 2011, 46(5): 405—422
- [11] McClelland J F, Jones R W, Bajic S J. FT-IR Photoacoustic Spectroscopy//Chalmers J M, Griffiths P R. *Handbook of vibrational spectroscopy*. Chichester, England: John Wiley & Sons, 2002: 1231—1250
- [12] Wold S, Sjostrom M, Eriksson L. PLS-regression: A basic tool of chemometrics. *Chemometrics and Intelligent Laboratory Systems*, 2001, 58(2): 109—130
- [13] Walczak B, Massart D L. The radial basis functions-partial least squares approach as a flexible non-linear regression technique. *Analytica Chimica Acta*, 1996, 331(3): 177—185
- [14] Chauchard F, Cogdill R, Roussel S, et al. Application of LS-SVM to non-linear phenomena in NIR spectroscopy: Development of a robust and portable sensor for acidity prediction in grapes. *Chemometrics and Intelligent Laboratory Systems*, 2004, 71(2): 141—150
- [15] Suykens J A K, Vandewalle J. *Nonlinear modeling: Advanced black-box techniques*. Boston, USA: Kluwer Academic Publishing, 1998: 1—274
- [16] Cogdill R P, Dardenne P. Least-squares support vector machines for chemometrics: An introduction and evaluation. *Journal of Near Infrared Spectroscopy*, 2004, 12(2): 93—100
- [17] Chapelle O, Haffner P, Vladimir N V. Support vector machines for histogram-based image classification. *IEEE Transactions on Neural Networks*, 1999, 10(5): 1055—1064
- [18] Belousov A, Verzakov S, von Frese J. Applicational aspect of support vector machine. *Journal of Chemometrics*, 2002, 16(8/10): 482—489
- [19] Thissen U, Ustun B, Melssen W J, et al. Multivariate calibration with least-squares support vector machines. *Analytical Chemistry*, 2004, 76(11): 3099—3105
- [20] Savitzky A, Golay M J E. Smoothing and differentiation of data by simplified least squares procedures. *Analytical Chemistry*, 1964, 36(8): 1627—1632
- [21] Kennard R W, Stone L A. Computer-aided design of experiments. *Technometrics*, 1969, 11(1): 137—148
- [22] Zhao C H, Guo F R, Wang F L. An improved independent component regression modeling and quantitative calibration procedure. *AIChE Journal*, 2010, 56(6): 1519—1535
- [23] Lee J, Qin S J, Lee I. Fault detection and diagnosis based on modified independent component analysis. *AIChE Journal*, 2006, 52(10): 3501—3514
- [24] Browne M W. Cross-validation methods. *Journal of Mathematical Psychology*, 2000, 44(1): 108—132
- [25] Wu D, He Y, Nie P C, et al. Hybrid variable selection in visible and near-infrared spectral analysis for non-invasive quality determination of grape juice. *Analytica Chimica Acta*, 2010, 659(1/2): 229—237
- [26] Williams P C, Sobering D C. Comparison of commercial near infrared transmittance and reflectance instruments for analysis of whole grains and seeds. *Journal of Near Infrared Spectroscopy*, 1993, 1(1): 25—32
- [27] Viscarra Rossel R A, McGlynn R N, McBratney A B. Determining the composition of mineral-organic mixes using UV-vis-NIR diffuse reflectance spectroscopy. *Geoderma*, 2006, 137(1/2): 70—82
- [28] Kuzmanovski I, Aleksovska S. Optimization of artificial neural networks for prediction of the unit cell parameters in orthorhombic perovskites. Comparison with multiple linear regression. *Chemometrics and Intelligent Laboratory Systems*, 2003, 67(2): 167—174
- [29] Pontes M J C, Rocha A M J, Pimentel M F, et al. Determining the quality of insulating oils using near infrared spectroscopy and wavelength selection. *Microchemical Journal*, 2011, 98(2): 254—259
- [30] Du C W, Zhou J M. Evaluation of soil fertility using infrared spectroscopy: A review. *Environmental Chemistry Letters*, 2009, 7(2): 97—113
- [31] Madari B E, Reeves III J B, Machado P L, et al. Mid-and near-infrared spectroscopic assessment of soil compositional parameters and structural indices in two Ferralsols. *Geoderma*, 2006, 136(1/2): 245—259
- [32] Janik L J, Skjemstad J O, Shepherd K D, et al. The prediction of soil carbon fractions using mid-infrared-partial least square analysis. *Austrian Journal of Soil Research*, 2007, 45(2): 73—81
- [33] Calderón F J, Mikha M M, Vigil M F, et al. Diffuse-reflectance mid-infrared spectral properties of soils under alternative crop rotations in a semi-arid climate. *Communications in Soil Science and Plant Analysis*, 2011, 42(17): 2143—2159
- [34] Gustafsson M G. Independent component analysis yields chemically interpretable latent variables in multivariate regression. *Journal of Chemical Information And Modeling*, 2005, 45(5): 1244—1255
- [35] Kaneko H, Arakawa M, Funatsu K. Development of a new regression analysis method using independent component analysis. *Journal of Chemical Information and Modeling*, 2008, 48(3): 534—541

APPLYING INFRARED PHOTOACOUSTIC SPECTROSCOPY AND SUPPORT VECTOR MACHINE MODEL TO QUANTIFY SOIL ORGANIC MATTER CONTENT

Zeng Yin Lu Yuzhen Du Changwen[†] Zhou Jianmin

(State Key Laboratory of Soil and Sustainable Agriculture, Institute of Soil Science, Chinese Academy of Sciences, Nanjing 210008, China)

Abstract Fast qualification of soil organic matter (SOM) is important to crop production and evaluation of soil quality. Fourier transform infrared photoacoustic spectroscopy (FTIR-PAS) makes it feasible to quantify soil organic matter content in a rapid way. In this study, FTIR-PAS was applied to measure SOM in the soils collected from paddy fields in Lishui District of Jiangsu Province. Support vector machine (SVM) was utilized to build calibration models. Principal component analysis (PCA), partial least squares (PLS) and independent component analysis (ICA) were performed separately to extract principal components (PCs), latent variables of PLS (LVPLS) and independent components (ICs) from the soil spectra as input of support vector machine (SVM). Hence, three SVM calibration models were built up. Meanwhile, PLS was also used to form a calibration model as control. Results show that the ICs-based SVM model performed best in prediction of SOM, with correlation coefficient (R^2), root-mean-square error (RMSEP) and ratio of performance to standard deviation (RPD) being 0.808, 0.575 and 2.28 respectively. Furthermore, F-test demonstrates that this model was significantly superior over the PCs-based SVM model, but was quite similar to the LVPLS-based SVM model and the classic PLS-based SVM model. Besides, no significant difference was observed between the predictions using the calibration models and the determination using the chemical method, as was demonstrated by *t*-test. It can, therefore, be concluded that the technology of infrared photoacoustic spectroscopy can be used as a new means for rapid determination of soil organic matter content.

Key words Soil organic matter; Infrared photoacoustic spectroscopy; Support vector machines; Quantitative prediction

(责任编辑:卢萍)