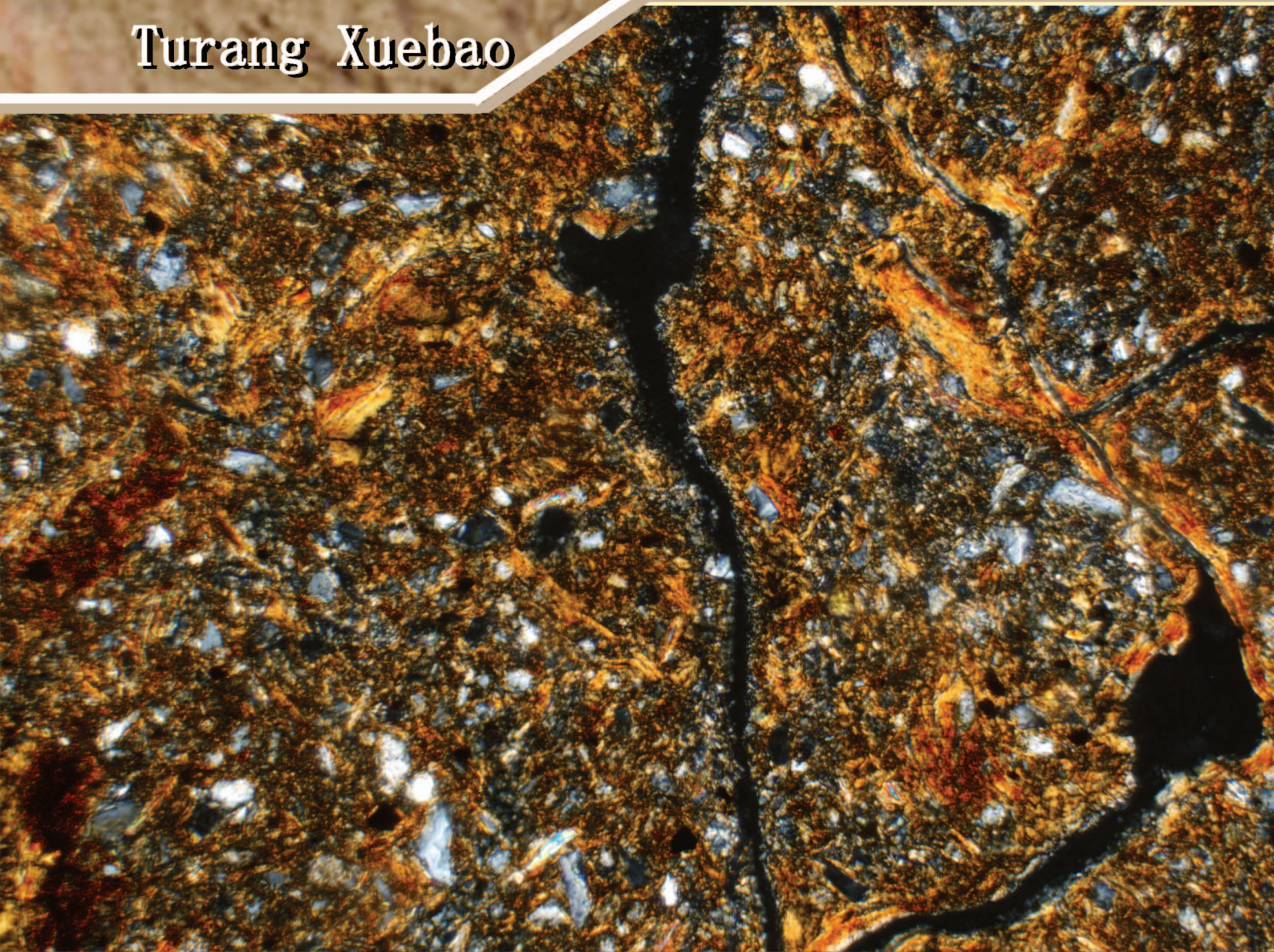


ISSN 0564-3929

Acta Pedologica Sinica 土壤学报

Turang Xuebao



中国土壤学会 主办
科学出版社 出版

2015

第 52 卷 第 2 期

Vol.52 No.2

《土壤学报》编辑委员会

主 编: 史学正

执行编委: (按姓氏笔画为序)

丁维新	巨晓棠	王敬国	王朝辉	史 舟	宇万太	朱永官
李永涛	李芳柏	李保国	李 航	吴金水	沈其荣	张玉龙
张甘霖	张福锁	陈德明	邵明安	杨劲松	杨明义	杨林章
林先贵	依艳丽	周东美	周健民	金继运	逢焕成	胡 锋
施卫明	骆永明	赵小敏	贾仲君	徐国华	徐明岗	徐建明
崔中利	常志州	黄巧云	章明奎	蒋 新	彭新华	雷 梅
窦 森	廖宗文	蔡祖聪	蔡崇法	潘根兴	魏朝富	

编辑部主任: 陈德明

责任编辑: 汪枏生 卢 萍 檀满枝

土 壤 学 报

Turang Xuebao

(双月刊, 1948年创刊)

第52卷 第2期 2015年3月

ACTA PEDOLOGICA SINICA

(Bimonthly, Started in 1948)

Vol. 52 No. 2 Mar., 2015

编 辑 《土壤学报》编辑委员会
地址: 南京市北京东路71号 邮政编码: 210008
电话: 025 - 86881237
E-mail: actapedo@issas.ac.cn

主 编 史学正
主 管 中国科学院
主 办 中国土壤学会
承 办 中国科学院南京土壤研究所

出 版 科 学 出 版 社
地址: 北京东黄城根北街16号 邮政编码: 100717

印刷装订 北京中科印刷有限公司
总 发 行 科 学 出 版 社
地址: 北京东黄城根北街16号 邮政编码: 100717
电话: 010 - 64017032
E-mail: journal@mail.sciencep.com

国外发行 中国国际图书贸易总公司
地址: 北京399信箱 邮政编码: 100044

Edited by Editorial Board of Acta Pedologica Sinica
Add: 71 East Beijing Road, Nanjing 210008, China
Tel: 025 - 86881237
E-mail: actapedo@issas.ac.cn

Editor-in-Chief Shi Xuezheng
Superintended by Chinese Academy of Sciences
Sponsored by Soil Science Society of China
Undertaken by Institute of Soil Science,
Chinese Academy of Sciences

Published by Science Press
Add: 16 Donghuangchenggen North Street,
Beijing 100717, China

Printed by Beijing Zhongke Printing Limited Company
Distributed by Science Press
Add: 16 Donghuangchenggen North Street,
Beijing 100717, China
Tel: 010 - 64017032
E-mail: journal@mail.sciencep.com

Foreign China International Book Trading Corporation
Add: P. O. Box 399, Beijing 100044, China

国内统一刊号: CN 32-1119/P

国内邮发代号: 2-560

国外发行代号: BM45

定价: 60.00元

国 内 外 公 开 发 行



ISSN 0564-3929



9 770564 392156

DOI: 10. 11766/trxb201311290572

基于局部加权回归的土壤全氮含量可见-近红外 光谱反演*

陈颂超¹ 冯来磊¹ 李 硕¹ 纪文君¹ 史 舟^{1,2†}

(1 浙江大学农业遥感与信息技术应用研究所, 杭州 310058)

(2 浙江大学唐仲英传感材料及应用研究中心, 杭州 310027)

摘 要 全氮是土壤肥力的重要指标,对作物产量具有决定性作用,采用土壤可见-近红外(Vis-NIR)光谱预测技术及时获取土壤全氮含量信息具有重要意义。采用来自 5 省的 450 个土壤本来验证局部加权回归方法(LWR)结合 Vis-NIR 光谱技术预测大面积土壤全氮含量的适用性。结果表明,LWR 模型的预测效果优于偏最小二乘回归(PLSR)、人工神经网络(ANN)和支持向量机(SVM),选取主成分数为 5,相似样本为 40 时,模型验证的决定系数(R_p^2)为 0.83,均方根误差($RMSE_p$)为 0.25 g kg^{-1} ,测定值标准偏差与标准预测误差的比值(RPD)达到 2.41。LWR 从建模集中选取与验证样本相似的土样作为局部建模样本,降低了差别大的样本对模型的干扰,从而提高了模型的预测能力。因此,LWR 建模方法通过大范围、大样本土壤光谱数据进行大尺度区域的全氮等土壤属性预测时能够发挥更好的作用。

关键词 土壤光谱;全氮预测;局部建模;偏最小二乘法;局部加权回归

中图分类号 S127/TP79 **文献标识码** A

氮是构成一切生命体的重要元素。在作物生产中,作物对氮素的需求量较大,土壤供氮不足会引起农产品产量和品质的下降,而其过量施用会造成江湖水体富营养化、地下水硝态氮积累和毒害等^[1]。因此,快速准确地获取土壤的氮素分布信息尤为重要。土壤可见-近红外(Vis-NIR)光谱技术具有快速、简便、无损、环保的特点,单条光谱曲线包含了多种理化属性信息,节省了大量人力、物力,为解决环境资源问题和促进精准农业的发展具有重要意义^[2]。近年来,借助可见-近红外漫反射光谱仪分析土壤氮素得到了广泛的关注。

国内外研究者在使用 Vis-NIR 技术获取小范围土壤氮素信息方面已经取得了一定的成果。Reeves 和 McCarty^[3]在预测马里兰州两块试验田的 180 个土壤样本全氮含量的研究中,光谱反射率经过一阶微分预处理后以 1 100 ~ 2 300 nm 为变量建立 PLSR 模型,该模型能很好地预测全氮含量,交叉验证的 R^2 为 0.92。Kuang 和 Mouazen^[4]通过 PLSR 模型对

捷克、德国 2 个农场的 198 个土样分别建立全氮预测模型,模型预测效果较好, R_p^2 分别为 0.93、0.75。徐永明等^[5]对天津南部城区 70 个样点的土壤光谱进行研究,基于波段深度的 PLSR 模型能预测土壤全氮含量($R_p^2 = 0.81$)。卢艳丽等^[6]对一块农田中的 110 个黑土土样进行研究发现,采用 550 nm 和 450 nm 组成的归一化光谱指数能较好地预测土壤全氮含量($R_p^2 = 0.83$)。

随着大样本土壤光谱数据的不断出现^[7],各类非线性预测模型如人工神经网络(ANN)、支持向量机(SVM)、局部加权回归(LWR)等方法逐渐被用于土壤属性的光谱预测建模,尤其是与土壤光谱数据库相结合的局部建模预测模型引起国际同行的极大关注。Pierna 和 Dardenne^[8]分别使用局部建模和最小二乘回归(PLSR)预测比利时土壤全氮含量,局部建模预测效果($R_p^2 = 0.80$)优于 PLSR($R_p^2 = 0.56$)。Christy 等^[9]使用 NIR(900 ~ 1 700 nm)光谱预测爱荷华州中部农田土壤水分、总碳、全氮和 pH

* 国家自然科学基金项目(41271234 和 40871100)、国家“863”计划项目(2013AA102301)、教育部新世纪优秀人才支持计划项目(NCET-10-0694)资助

† 通讯作者,E-mail:shizhou@zju.edu.cn

作者简介:陈颂超(1990—),男,浙江上虞人,硕士研究生,主要从事土壤高光谱研究。E-mail:chensongchao@zju.edu.cn

收稿日期:2013-11-29;收到修改稿日期:2014-11-26

等土壤属性,通过 LWR 建立的模型预测效果较好, R_p^2 分别为 0.82、0.87、0.86 和 0.72。Genot 等^[10]建立了比利时瓦隆地区土壤有机碳、全氮的 LOCAL 模型,该模型具有很好的预测精度 ($R_p^2 > 0.90$)。

国内学者也陆续开展了使用非线性模型预测全氮等土壤属性的研究,但更多使用的是 ANN^[3,11] 和 SVM^[3,12] 等模型,局部建模研究较少。我国地域宽广且土壤类型多样,建立应用性强的土壤光谱预测模型是当前土壤 Vis-NIR 技术走向实用化的重要研究方向。LWR 是一种局部建模方法,它从光谱库中选取光谱特征相近的样本建立局部模型,提高了模型的预测能力。土壤光谱数据越全面,基于大样本土壤光谱数据的局部模型预测效果就越好。

本研究分析了浙江、吉林、云南、海南和甘肃 5 个省 450 个土壤样本,主要研究目的:(1)通过参数优选建立预测全氮的最佳 LWR 模型;(2)比较 LWR 模型与 PLSR、SVM、ANN 等线性、非线性模型对土壤全氮预测的精度。

1 材料与方法

1.1 样品采集

本研究使用的数据来源于浙江大学牵头组建的中国土壤可见-近红外光谱数据库^[7]。该光谱数据库自 2006 年开始收集全国不同地区的土壤样品,

均为 0~20 cm 的表层土壤,由各合作单位野外采集,至撰稿日期已从 20 个省共收集到 20 多个土类共计 4 935 个土壤样本,初步建立了全国范围的土壤可见-近红外光谱数据库。本研究使用的土壤样本来自浙江、云南、吉林、海南、甘肃 5 省,共计 20 个土壤类型的 450 个土样,其中包括中国典型土壤如水稻土、砖红壤、滨海盐土、草甸土、黑钙土、风砂土等(表 1)。土样经过风干、研磨并过 2 mm 孔筛,通过四分法分成两份,分别用于理化分析和光谱测试。采用半微量凯氏定氮法测定土壤全氮含量。

由于样本来自不同省份,所以全氮含量差异很大,最小值为 0.28 g kg^{-1} ,最大值为 3.65 g kg^{-1} ,平均含量为 1.24 g kg^{-1} (表 2)。

根据 Kennard-Stone 算法选取全体样本的 3/4 作为建模集(337 个),剩余 1/4(113 个)的样本作为验证集来评判模型预测精度。Kennard-Stone 算法根据光谱主成分空间的欧氏距离(Euclid Distance)选择样本,即先寻找全体样本空间中欧氏距离最远的两个样本,归入建模集。再依次计算全体样本中每个剩余样本到建模集样本的距离,选取每个剩余样本的最短距离,将这些剩余样本最短距离中的最长距离所对应的样本选入建模集。重复上一个步骤,直至建模集中样本的数量和所需建模集样本数量一致。从表 2 中土壤样本全氮含量分布特征来看,验证样本的标准差和变异系数与全体样本全氮

表 1 土壤样点分布及类型

Table 1 Distribution of soil sampling sites and soil types

地点 Locations	样本数 Number of samples	土壤类型 Soil types
浙江省 Zhejiang Province	104	水稻土、滨海盐土
吉林省 Jilin Province	251	黑土、黑钙土、草甸土、冲积土、风砂土、沼泽土
甘肃省 Gansu Province	42	黄绵土、黑垆土、灰钙土、黑钙土、灌漠土、褐土、淤积土、山地黄麻土、栗钙土、潮土、灌耕土、淡灰钙土
海南省 Hainan Province	24	砖红壤
云南省 Yunnan Province	29	风砂土

表 2 土壤全氮含量统计特征

Table 2 Statistics of total nitrogen contents in soil

样本集 Dataset	样本数 Number of samples	平均数 Mean (g kg^{-1})	最小值 Minimum (g kg^{-1})	最大值 Maximum (g kg^{-1})	标准差 SD	变异系数 CV (%)
全体样本 Total dataset	450	1.24	0.28	3.65	0.57	46.0
验证样本 Validation dataset	113	1.27	0.28	3.61	0.59	46.5

含量的空间分布相一致,具有代表性。

1.2 光谱数据采集与预处理

光谱测量采用美国 ASD 公司的 FieldSpec Pro FR 型光谱仪进行土壤测试,波长范围为 350 ~ 2 500 nm,光谱分辨率为 3 nm (350 ~ 1 000 nm) 和 10 nm (1 000 ~ 2 500 nm),采样间隔为 1.4 nm (350 ~ 1 000 nm) 和 2 nm (1 000 ~ 2 500 nm)。光谱测量在一个能控制光照条件的暗室内进行。光源为 50W 的卤素灯,距土壤表面 30 cm,天顶角为 30°,采用 25° 视场角的探头,其位于土壤样本表面的正上方 15 cm 处。测试之前先进行白板校正。每个土样采集 10 条光谱曲线,取其各波段平均值作为该土样的实际反射光谱数据。

每个土壤样本的光谱去除噪声较大的边缘波段(350 ~ 399 nm, 2 451 ~ 2 500 nm)。保留 400 ~ 2 450 nm 的反射光谱数据用于后续研究。为减少研究数据受实验室光学环境场差异和磨样过筛的影响,对所有的光谱数据均尝试采用 Savitzky-Golay 平滑加一阶微分或二阶微分变换、多元散射校正 (Multiplicative Scatter Correction, MSC)、标准正态变量变换 (Standard Normal Variate, SNV) 等处理方法。通过对后期预测效果的比较,最终选用 Savitzky-Golay 平滑 (2 阶 11 窗口) 加一阶微分方法。

1.3 建模方法

在局部加权回归 LWR 中,首先确定用于每个局部建模集的主成分数 (C) 和样本数 (A),然后通过主成分分析来压缩预处理后的光谱信息。对于验证集中每一个预测样本,在建模集中选择与它最相似的 A 个样本,样本相似程度通过主成分得分的空间距离测量来定义。最后使用加权最小二乘回归来计算最优回归系数^[13]。

样本相似度指标和加权函数是 LWR 中的关键参数,设置如下:

空间距离测量使用欧氏距离 d 值来度量:

$$d = \sqrt{\sum (a_i - b_i)^2}, i = 1, 2, \dots, n \quad (1)$$

加权最小二乘回归中加权函数 $W(d)$ 采用立方加权函数法:

$$W(d) = (1 - d^3)^3, 0 \leq d \leq 1 \quad (2)$$

偏最小二乘回归是一种基于因子分析的非参数线性回归方法,常用于光谱定量研究。该方法集主成分分析、典型相关分析和线性回归分析于一体,实现了数据结构的简化,解决了自变量之间多重相关的问题,适合用于光谱分析这种自变量较多

的情况。本研究采用的 PLSR 模型是基于 NIPALS (Nonlinear Iterative Partial Least Squares) 算法得到。

人工神经网络是一种模仿人脑基本结构和功能来处理非线性高维与高阶性知识信息问题方法^[14],基本结构是由输入层,输出层和隐藏层三部分组成,能模拟任何非线性输入变量和输出变量的关系,具有很强的自组织,自适应学习和记忆联想能力,能够实现许多常规方法难以达到的精度^[2]。

支持向量机的基本思想是将输入变量通过非线性映射映射到一个高维特征空间中,然后在这个高维空间构建最优分类面。核函数是 SVM 的核心,适当的核函数可以得到高精度的高维空间的分类函数。SVM 常用的核函数有线性核函数、多项式核函数、径向基函数和 Sigmoid 核函数 4 种^[2]。

1.4 数据分析

本研究中光谱反射率的 Savitzky-Golay 平滑加一阶微分、主成分分析、PLSR 建模验证由 Unscrambler X 10.1 实现, Kennard-Stone 算法和 LWR 建模验证由 R 软件实现, SVM 和 ANN 建模验证由 Matlab 2011b 实现。

2 结果与讨论

2.1 不同全氮含量土壤样本的光谱曲线特征

土壤氮素的形态多样,包括有机态氮和无机态氮,具体又包括氨基酸态氮、氨基糖态氮、铵态氮、酸解性未知态氮和非酸解性氮等。因此,土壤氮素的 Vis-NIR 吸收和反射光谱特性较为复杂。土壤全氮预测的机理是通过其与有机质的相关性来间接预测还是基于特征吸收的直接预测在国际上存在着争议,当全氮和有机质相关性好时全氮预测效果好,当这种相关不存在时,全氮建模可能是基于氮的特征吸收^[15]。但可以肯定的是土壤全氮和有机质密切相关,因为在表层土壤中 90% 左右的全氮是有机态氮。本研究中土壤全氮和有机质的相关系数达到 0.87,该结果印证了上述观点。有机质中 NH、CH 和 CO 等官能团的特征波段出现在中红外区域,而在 Vis-NIR 区域存在的是倍频峰和合频峰^[16]。由于不同倍频峰和合频峰相互重叠,直接获取 Vis-NIR 的有效信息比较困难^[15],目前的研究主要通过化学计量学中的多元统计模型提取光谱中的信息,并以此为基础建模预测土壤有机质、全氮等属性。从图 1 中不同全氮含量水平的反射光谱来看,随着土壤全氮含量的增加,Vis-NIR 光谱反射

率降低,这与前期研究中有有机质含量越高光谱反射率越低的结论相近^[2],同样表明了全氮与有机质密切相关。

一阶微分处理能够清晰反映光谱细微变化,因此进一步分析在全波段范围(400~2 450 nm)土壤全氮含量与一阶微分转换后光谱反射率的相关性(图2)。全氮含量与一阶微分光谱相关系数较高的峰值出现在820、1 400、1 430、1 630、1 800、1 930、2 100、2 200和2 300 nm等几个波段,这些波段可能是全氮预测的重要建模波段。Stenberg等^[15]认为1 100、1 600、1 700~1 800、2 000和2 200~2 400 nm附近的波段对有机碳和全氮的建模至关重要,史舟等^[7]研究表明580、820、1 400和2 200 nm是不同地区不同类型土壤共有的有机质重要建模波段,张娟娟等^[17]认为750~920和1 360~1 420 nm可能是不同类型土壤的有机质敏感波段。本研究中全氮建模的重要波段与上述研究的波段相重叠,且由于

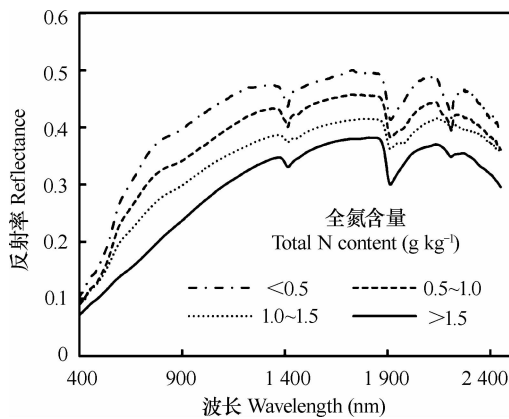


图1 不同全氮含量土壤光谱反射率曲线
Fig. 1 Reflectance spectra of soils in different total nitrogen content

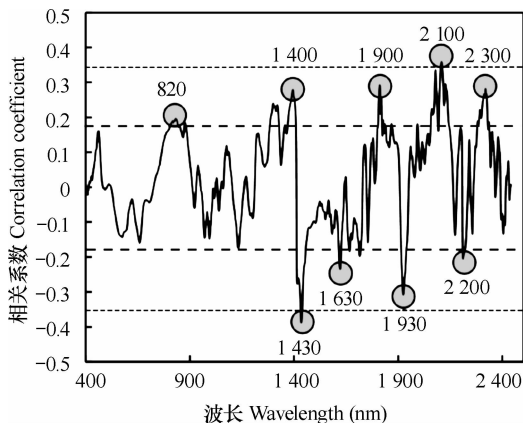


图2 土壤全氮含量与一阶微分光谱的相关系数曲线
Fig. 2 Correlation of soil total nitrogen content with the first derivatives of the reflectance

土壤全氮和有机质具有较好的相关性,基于 Vis-NIR 和多元统计模型的土壤全氮预测具有一定的可行性。

2.2 LWR 建模参数优选

主成分数(PCs)和相似样本数是局部加权回归方法 LWR 中重要的两个参数。Næs 等^[13]研究认为 35 个相似样本和 3 个主成分可以作为今后 LWR 应用的一个参考,Ramirez-Lopez 等^[18]通过局部建模研究黏土和 pH 时认为 50 个左右的相似样本是最佳的样本数。因此,本研究将主成分数和相似样本数的范围分别设定为 2~10 和 20~100 之间,在不同主成分数和相似样本数的组合下比较模型预测的均方根误差(图3)。

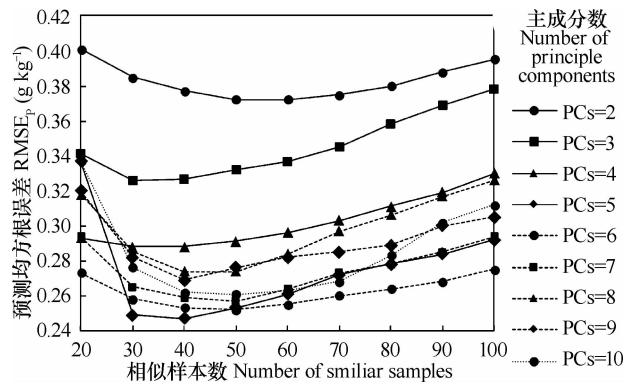


图3 不同主成分数、相似样本数 LWR 建模的 $RMSE_p$
Fig. 3 $RMSE_p$ fitted with the LWR model relative to number of principal components and number of similar samples

由图3可知,当主成分数、相似样本数不同时,LWR 模型的 $RMSE_p$ 相差较大,当主成分数为 5、6 且相似样本数为 30、40、50、60 时,模型效果较好($RMSE_p \leq 0.26 \text{ g kg}^{-1}$)。当主成分数为 5 且相似样本数为 40 时,LWR 模型的预测效果最佳, $RMSE_p$ 达到最小为 0.25 g kg^{-1} 。当主成分数过低时($PCs \leq 3$),这些主成分中的信息(累计特征值 $\leq 73\%$)无法较好地反映完整的光谱信息,因此验证模型的预测效果较差($RMSE_p \geq 0.32 \text{ g kg}^{-1}$);当主成分数过高时($PCs \geq 8$),过多的主成分信息(累计特征值 $\geq 88\%$)使模型出现过拟合,预测效果变差。因此,在本研究中 LWR 模型推荐的主成分数为 5,相似样本数为 40。

研究区域和土壤样本的空间异质性决定了 LWR 模型中主成分数和相似样本数这两个重要参数随着建模样本的变化而变化。Shi 等^[19]在预测中国 2 782 个土壤样本有机质的研究中,当主成分数和相似样本分别优化为 10 (累计特征值 =

97.77%) 和 50 时 LWR 预测效果最佳, $RMSE_p$ 为 5.18 g kg^{-1} , RPD 为 2.88。Hergarten 等^[20] 收集来自塔吉克斯坦 6 个地区农田的 725 个土壤样本, 70% 土壤样本作为建模集, 剩余的样本为验证集, 当近似样本数为 170 个时, LWR 模型预测有机碳的效果最佳, RPD 为 1.63。因此, 在不同研究区域应用 LWR 时需要同时对主成分数和相似样本数进行优选。

2.3 LWR 建模预测结果对比

首先将 LWR 模型验证结果与线性模型 PLSR 的验证结果进行对比。本研究中, PLSR 预测模型的最佳因子数为 7。图 4a 和图 4d 分别为实测土壤全氮含量和 PLSR、LWR 预测结果的检验。比较表明, 对于跨省的土壤样本全氮含量的预测, LWR 模型的验证结果远远好于 PLSR 模型, 其验证 R_p^2 由 0.63 提高至 0.83, $RMSE_p$ 由 0.36 g kg^{-1} 降低至 0.25 g kg^{-1} ,

RPD 由 1.63 增加至 2.41。此外, LWR 模型的拟合曲线相比 PLSR 模型更加接近 1:1 线, 这说明 LWR 模型相比 PLSR 模型能较好解决高值低估现象。

国内采用 PLSR 方法进行土壤全氮光谱预测多集中在小尺度、小样本的研究, 如任红艳等^[21] 将扬州真武镇的 146 个土壤样本过 0.25 mm 孔径筛, 以 106 个土壤样本建立 PLSR 全氮预测模型 (剩余 40 个样本作为验证集), 预测模型的验证结果: R_p^2 为 0.92, RPD 达 3.49。李颖等^[22] 将北京一块试验田的 72 个土壤样本反射率经一阶微分和 SNV 处理后, 以 3:1 建立建模集和验证集, PLSR 模型对土壤全氮的预测效果较好, 模型验证 R^2 为 0.95, RPD 达 4.68。上述研究表明 PLSR 模型在小尺度研究区域内对全氮预测效果较好, 当研究区域扩大至跨省的尺度时, 由于土壤类型增多, PLSR 模型就会受到局

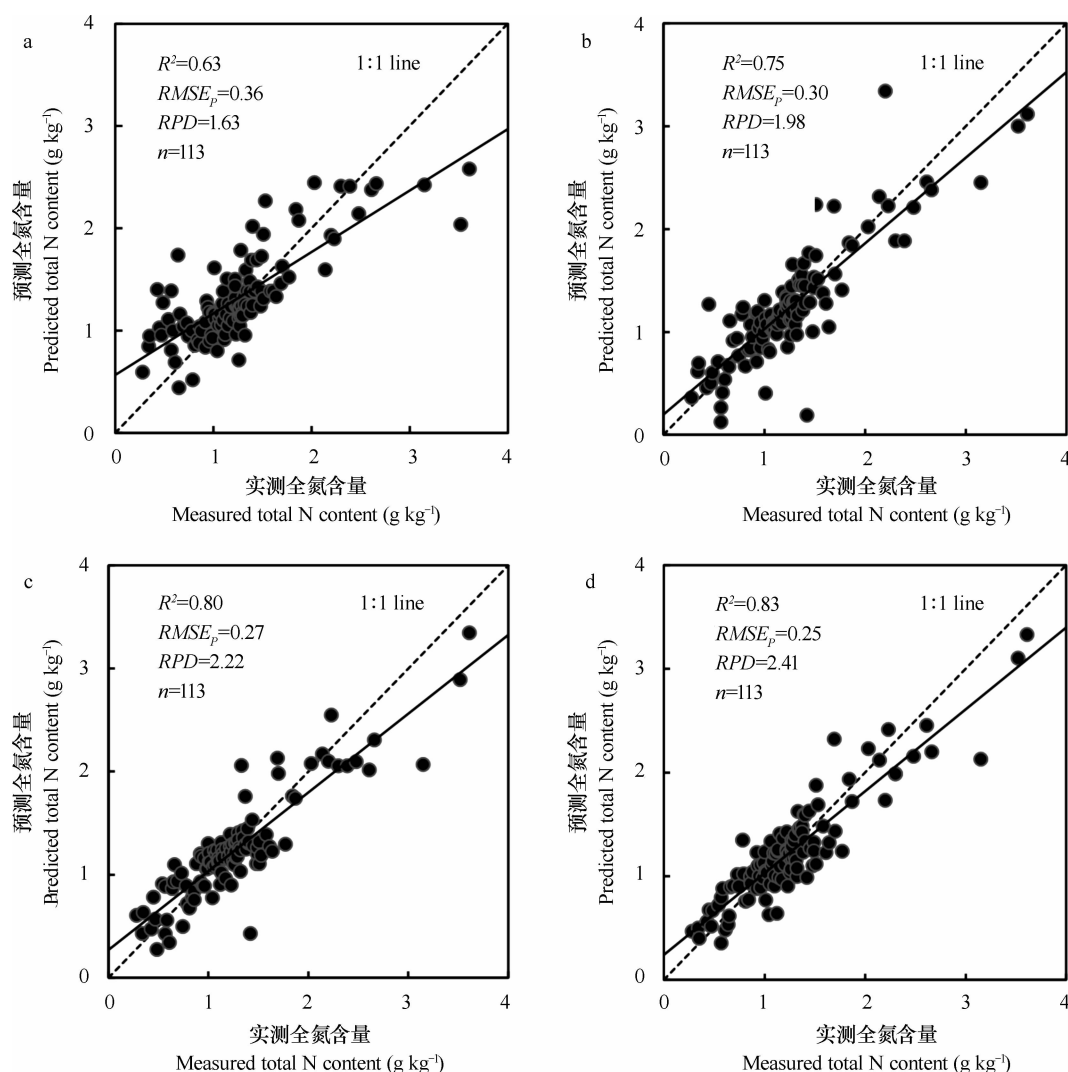


图 4 113 个样本全氮预测模型 PLSR (a)、ANN (b)、SVM (c)、LWR (d) 结果比较

Fig. 4 Comparison between PLSR (a), ANN (b), SVM (c) and LWR (d) in TN prediction of 113 soil samples

限,从而难以取得令人满意的预测精度,如张娟娟等^[23]对南京、河南、浙江、上海等地 460 个土样 350 ~ 2 500 nm 的光谱经 Savitzky-Golay 平滑加一阶微分处理后建立 PLSR 全氮预测模型,因子数为 5 时得到的最佳模型交叉验证结果 R^2 为 0.49, RPD 为 1.39。从本研究结果可见 PLSR 预测模型偏离 1:1 线情况严重,特别是对于全氮含量较高的土样存在明显的低估现象(图 4a)。

此外,本研究还进行了 LWR 与常用的 ANN、SVM 两个非线性方法的比较。研究采用同样的建模集和验证集。其中,ANN 采用三层 BP 网络模型,光谱主成分变换后的前 10 个主成分作为训练输入点,隐含层数为 1,隐含层神经元数目为 7,输出节点为土壤全氮含量;SVM 采用 ϵ -支持向量回归模型,核函数为 RBF (Radial basis function),最佳 PLS 因子数为 5。ANN 和 SVM 的建模预测结果见图 4b、图 4c,ANN 建模的验证结果 R_p^2 为 0.75, $RMSE_p$ 为 0.30 g kg^{-1} , RPD 为 1.98, SVM 的预测效果 R_p^2 为 0.80, $RMSE_p$ 为 0.27 g kg^{-1} , RPD 为 2.22。总体而言,ANN 和 SVM 的建模预测效果优于 PLSR,稍逊于 LWR。由于 Vis-NIR 能被用于识别土壤类型和预测土壤全氮等理化特性,因此 LWR 模型通过搜索相似光谱样本,即找出土壤类型相近或理化特性差异相对较小的参考样本建立土壤全氮的预测模型;而 ANN 与 SVM 等非线性模型虽有相对较好的精度,但存在可理解性差的黑箱问题。而且,在大样本土壤光谱数据的建模预测时,LWR 方法相对简单,计算效率优于 ANN 与 SVM 方法。

2.4 局部加权回归 LWR 的优势

LWR 模型明显优于 PLSR 模型的关键是因为 LWR 从建模集中选取了部分与验证集中预测样本光谱相似度高的样本来建立局部模型,减少了不相关样本给建模带来的误差。通过举例点的前 5 个主成分在欧氏空间搜索与其距离最近的 40 个相似样本,将举例点的原始光谱与 40 个相似样本原始光谱的平均值、450 个总体样本原始光谱的平均值进行光谱比较(图 5)。

图 5 中举例样本的光谱和 40 个相似样本的光谱与 450 个总体样本的光谱相比较,无论是原始光谱(Raw spectral reflectance, Raw)还是连续统去除曲线(Continuum-removed spectral reflectance, CR)相似度均较高,这表明 LWR 从 450 个总体样本中选取了与预测样本近似的局部建模集,选取的局部建模样本代表性强,减少了差异较大的样本对预测样

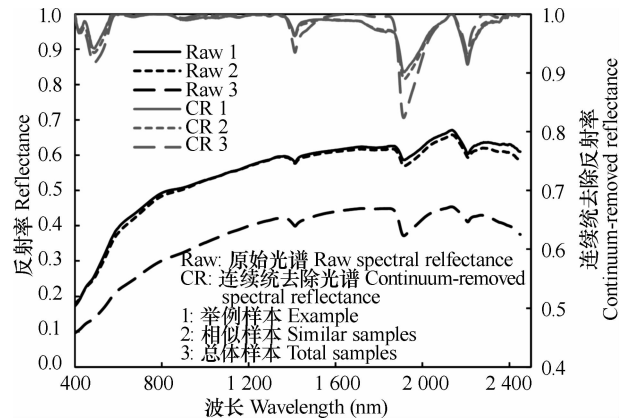


图 5 原始光谱 (Raw) 和连续统去除曲线 (CR) 比较

Fig. 5 Comparison of raw spectral reflectance (Raw) and continuum-removed spectral reflectance (CR)

本的干扰,较好地提高了预测效果。

土壤光谱库通过整合不同研究者采集得到的土壤光谱与属性数据,实现光谱数据的共享,为研究者在不同尺度建模预测土壤属性提供充分的数据支持。目前很多研究者致力于全球和国家级土壤 Vis-NIR 光谱数据库的建设,Brown 和 Viscarra Rossel 分别于 2006 年和 2008 年牵头建立全球土壤光谱数据库建设项目,法国、瑞典、中国等国家的土壤高光谱数据库也正在逐步建设中^[7,24-26]。对于大样本光谱库的本地化应用,类似 LWR、LOCAL 建模等局部回归方法是目前较好的方法。

中国是一个地貌复杂、土壤类型多样的国家,根据第二次土壤普查所采用的土壤发生分类系统,我国有 12 个土纲,61 个土类,235 个亚类,约 2 500 个土种,不同类型土壤由于气候、时间、母质、地形、生物等成土因素不同从而理化性质差异较大,而光谱是土壤理化性质的综合反映,因此这对推广使用土壤光谱技术是一个极大的挑战^[7]。目前,全国土壤光谱库正在逐步完善,土壤样本的类型与数量将大大增加,LWR 方法能在土壤光谱库中找到与研究区域土样相似度高的建模样本,从而提升研究区域模型的预测效果。此外,将研究区内部分样本光谱数据与土壤光谱库整合作为建模集的 Spiking 思想,结合 LWR 方法预测土壤全氮等属性是后续研究可以考虑的方向。

3 结 论

土壤全氮与一阶微分转换后光谱反射率相关系数较高的峰值出现在 820、1 400、1 430、1 630、

1 800、1 930、2 100、2 200 和 2 300 nm 等几个波段,与有机质的光谱建模重要波段相重叠。由于研究区域和土壤样本的空间异质性,局部加权回归 LWR 模型中的两个重要参数主成分数和相似样本数随着建模样本的变化而变化,因此应用 LWR 时需要对这两个参数进行优化。本研究中主成分数和相似样本数的最佳设置为 5 和 40。LWR 全氮预测模型验证集的 R_p^2 为 0.83, $RMSE_p$ 为 0.25 g kg^{-1} , RPD 达到 2.41,较好地降低了 PLSR 模型高值低估的现象,使其更接近 1:1 线。此外,LWR 模型的预测能力优于非线性的 ANN 与 SVM 模型,且不存在黑箱问题。基于大样本光谱数据的 LWR 模型能够很好预测研究区域内土壤全氮含量。随着土壤光谱库的不断完善,LWR 模型将更好地挖掘大样本数据库中的信息,使之发挥更大的作用。

参 考 文 献

- [1] 黄昌勇. 土壤学. 北京: 中国农业出版社, 2000. Huang C Y. Soil science (In Chinese). Beijing: China Agriculture Press, 2000
- [2] 纪文君, 李曦, 李成学, 等. 基于全谱数据挖掘技术的土壤有机质高光谱预测建模研究. 光谱学与光谱分析, 2012, 32(9): 2393—2398. Ji W J, Li X, Li C X, et al. Using different data mining algorithms to predict soil organic matter based on Visible-Near infrared spectroscopy (In Chinese). Spectroscopy and Spectral Analysis, 2012, 32(9): 2393—2398
- [3] Reeves III J B, McCarty G W. Quantitative analysis of agricultural soils using near infrared reflectance spectroscopy and fibre-optic probe. Journal of Near Infrared Spectroscopy, 2001, 9: 25—34
- [4] Kuang B, Mouazen A M. Calibration of visible and near infrared spectroscopy for soil analysis at the field scale on three European farms. European Journal of Soil Science, 2011, 62: 629—636
- [5] 徐永明, 蔺启忠, 王璐, 等. 基于高分辨率反射光谱的土壤营养元素估算模型. 土壤学报, 2006, 43(5): 709—716. Xu Y M, Lin Q Z, Wang L, et al. Model for estimating soil nutrient elements based on high resolution reflectance spectra (In Chinese). Acta Pedologica Sinica, 2006, 43(5): 709—716
- [6] 卢艳丽, 自由路, 王磊, 等. 黑土土壤中全氮含量的高光谱预测分析. 农业工程学报, 2010, 26(1): 256—261. Lu Y L, Bai Y L, Wang L, et al. Determination for total nitrogen content in black soil using hyperspectral data (In Chinese). Transactions of the CSAE, 2010, 26(1): 256—261
- [7] 史舟, 王乾龙, 彭杰, 等. 中国主要土壤高光谱反射特性分类与有机质光谱预测模型. 中国科学: 地球科学, 2014, 44(1): 1—11. Shi Z, Wang Q L, Peng J, et al. Development of national VNIR soil-spectral library for soil classification and the predictions of organic matter (In Chinese). Science China: Earth Science, 2014, 44(1): 1—11
- [8] Pierna J A F, Dardenne P. Soil parameter quantification by NIRS as a Chemometric challenge at 'Chimimétrie 2006'. Chemometrics and Intelligent Laboratory Systems, 2008, 91: 94—98
- [9] Christy C, Drummond P, Laird D. An on-the-go spectral reflectance sensor for soil. St. Joseph Michigan: ASAE Annual International Meeting, 2003; Paper No. 031044
- [10] Genot V, Colinet G, Bock L, et al. Near infrared reflectance spectroscopy for estimating soil characteristics valuable in the diagnosis of soil fertility. Journal of Near Infrared Spectroscopy, 2011, 19: 117—138
- [11] 郑立华, 李民赞, 潘雯, 等. 基于近红外光谱技术的土壤参数 BP 神经网络预测. 光谱学与光谱分析, 2008, 28(5): 1160—1164. Zheng L H, Li M Z, Pan L, et al. Estimation of soil organic matter and soil total nitrogen based on NIR spectroscopy and BP neural network (In Chinese). Spectroscopy and Spectral Analysis, 2008, 28(5): 1160—1164
- [12] 刘雪梅, 柳建设. 基于 LS-SVM 建模方法近红外光谱检测土壤速效 N 和速效 K 的研究. 光谱学与光谱分析, 2012, 32(11): 3019—3023. Liu X M, Liu J S. Based on the LS-SVM modeling method determination of soil available N and available K by using near-infrared spectroscopy (In Chinese). Spectroscopy and Spectral Analysis, 2012, 32(11): 3019—3023
- [13] Næs T, Isaksson T, Fean T. A user-friendly guide to multivariate calibration and classification. Chichester: NIR Publications, 2004
- [14] 金龙, 罗莹, 缪启龙, 等. 农田土壤湿度的人工神经网络预报模型研究. 土壤学报, 1998, 35(1): 25—32. Jin L, Luo Y, Miao Q L, et al. Forecast model of farmland soil moisture by artificial neural network (In Chinese). Acta Pedologica Sinica, 1998, 35(1): 25—32
- [15] Stenberg B, Viscarra Rossel R A, Mouazen A M, et al. Visible and near infrared spectroscopy in soil science. Advances in Agronomy, 2010, 107: 163—215
- [16] Guerrero C, Zornoza R, Gómez I, et al. Spiking of NIR regional models using samples from target sites: Effect of model size on prediction accuracy. Geoderma, 2010, 158: 66—77
- [17] 张娟娟, 田永超, 朱艳, 等. 不同类型土壤的光谱特征及其有机质含量预测. 中国农业科学, 2009, 42(9): 3153—3162. Zhang J J, Tian Y C, Zhu Y, et al. Spectral characteristics and estimation of organic matter contents of different soil types (In Chinese). Scientia Agricultura Sinica, 2009, 42(9): 3153—3162
- [18] Ramirez-Lopez L, Behrens T, Schmidt K, et al. Distance and similarity-search metrics for use with soil vis-NIR spectra. Geoderma, 2013, 199: 43—53
- [19] Shi Z, Wang Q L, Viscarra Rossel R A, et al. Prediction of soil organic matter using cluster-local regression and national Vis-NIR soil spectral library in China. 3rd Global Workshop on Proximal Soil Sensing. Potsdam Germany: Leibniz-Institute for Agricultural Engineering Potsdam-Bornim, 2013: 173—181
- [20] Hergarten C, Nazarmavloev F, Wolfgramm B. Building a soil spectral library for Tajikistan comparing local and global modeling approaches. 3rd Global Workshop on Proximal Soil Sensing. Potsdam Germany: Leibniz-Institute for Agricultural Engineering

- Potsdam-Bornim, 2013: 265—269
- [21] 任红艳, 史学正, 庄大方, 等. 土壤全氮含量与碳氮比的高光谱反射估测影响因素研究. 遥感技术与应用, 2012, 27(3): 372—379. Ren H Y, Shi X Z, Zhuang D F, et al. Effects on estimating soil nitrogen content and ratio of carbon to nitrogen using hyperspectral reflectance (In Chinese). Remote Sensing and Technology and Application, 2012, 27(3): 372—379
- [22] 李颀, 张小超, 苑严伟, 等. 北京典型耕作土壤养分的近红外光谱分析. 农业工程学报, 2012, 28(2): 176—179. Li J, Zhang X C, Yuan Y W, et al. Analysis of soil nutrient content based on near infrared reflectance spectroscopy in Beijing region (In Chinese). Transactions of the CSAE, 2012, 28(2): 176—179
- [23] 张娟娟, 田永超, 姚霞, 等. 基于高光谱的土壤全氮含量估测. 自然资源学报, 2011, 26(5): 881—890. Zhang J J, Tian Y C, Yao X, et al. Estimating soil total nitrogen content based on hyperspectral analysis technology (In Chinese). Journal of Natural Resources, 2011, 26(5): 881—890
- [24] Brown D J, Shepherd K D, Walsh M G, et al. Global soil characterization with VNIR diffuse reflectance spectroscopy. Geoderma, 2006, 132: 273—290
- [25] Gogé F, Gomez C, Jolivet C, et al. Which strategy is best to predict soil properties of a local site from a national Vis-NIR database? Geoderma, 2014, 213: 1—9
- [26] Wetterlind J, Stenberg B. Near infrared spectroscopy for within field soil characterization small local calibrations compared with national libraries spiked with local samples. European Journal of Soil Science, 2010, 61: 823—843

VIS-NIR SPECTRAL INVERSION FOR PREDICTION OF SOIL TOTAL NITROGEN CONTENT IN LABORATORY BASED ON LOCALLY WEIGHTED REGRESSION

Chen Songchao¹ Feng Lailei¹ Li Shuo¹ Ji Wenjun¹ Shi Zhou^{1,2*}

(1 Institute of Agricultural Remote Sensing and Information Technology Application, Zhejiang University, Hangzhou 310058, China)

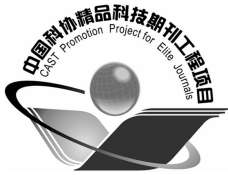
(2 Cyrus Tang Center for Sensor Material Application, Zhejiang University, Hangzhou 310027, China)

Abstract Diffuse reflectance spectroscopy within the visible and near-infrared (Vis-NIR) range is a promising way for acquisition of soil properties and digital soil mapping. The diffuse reflectance spectroscopy technique is rapid, nondestructive, environment-friendly and more efficient than the conventional analysis method. However, due to the diversity and spatial heterogeneity of soil, the prediction model based on the technique will have to face the issue of universality. Total nitrogen (TN) in soil is not only a significant index of soil fertility but also an important factor deciding crop yield. It is, therefore, essential to timely acquire the information of TN in soil. This paper introduces a method, i. e. locally weighted regression (LWR), as supplement to the use of the Vis-NIR spectrum technique in predicting TN in soil at a regional scale, and evaluates accuracy of the prediction using Vis-NIR plus LWR. To that end a total of four hundred and fifty soil samples were collected from Zhejiang, Jilin, Yunnan, Hainan and Gansu, air dried and ground to pass a 2 mm sieve. Their Vis-NIR diffuse reflectance spectra were collected using a FieldSpec Pro FR Spectrometer. The reflectance spectra in the wavelength range from 400 to 2 450 nm were denoised by Savitzky-Golay and first derivated. Three fourths of the samples were selected for calibration dataset using the Kennard-Stone algorithm and the remaining one fourth were used as validation dataset. The core of the LWR method is to select samples from the calibration dataset most spectrally similar to those in the validation dataset. The algorithm of LWR goes in three steps: to decompose and compress the spectral matrix through Principal Component Analysis and pick out local modeling subsets from the modeling dataset similar to the validation dataset by Euclidean distance. Based on the spectral distance of each sample in the local modeling subset to the validation sample, weight of the sample in the regression model is defined, by means of tri-cube weight function. The number of principle components and the number of similar samples were the crucial parameters in the LWR model, and in this study, the two parameters were optimized to be 5 and 40, respectively. The determination coefficient (R_p^2), the root mean square error ($RMSE_p$) and ratio of standard error of performance to standard deviation (RPD) was 0.63, 0.36 g kg⁻¹ and 1.63, respectively, in the PLSR model. However, the support vector machine (SVM) model and artificial neural network (ANN) model was higher than the PLSR model in prediction accuracy ($R_p^2 = 0.75 \sim 0.80$, $RMSE_p = 0.27 \sim 0.30$ g kg⁻¹, $RPD = 1.98 \sim 2.22$). Thanks to the advantages of LWR in algorithm, the LWR model reduced the interference of samples lower in similarity in local modeling, and hence increased the accuracy of TN prediction ($R_p^2 = 0.83$,

$RMSE_p = 0.25 \text{ g kg}^{-1}$, $RPD = 2.41$). The findings demonstrate that correlation coefficient between soil TN and the spectral reflectance after first-order differential transformation peaks at 820, 1 400, 1 430, 1 630, 1 800, 1 930, 2 100, 2 200 and 2 300 nm, which overlap the important bands for spectral modeling of soil organic matter. Due to the spatial heterogeneity of the study areas and soil samples, the two crucial parameters, i. e. number of similar samples and number of principle components, of the LWR model vary with the modeling datasets, so they should be optimized when LWR is used to predict TN. The LWR TN prediction model diminishes the probability of underestimating TN content as the PLSR model would, and makes the prediction closer to 1:1 line. Besides, the LWR model performs better than the non-linear ANN and SVM models in TN prediction, and does not have any black-box problem. Therefore, it can be concluded that LWR is a reliable method for prediction of soil TN content when a large spectral database is available. With the consumption of various large-scaled soil spectral libraries, LWR can be used to tap more useful information out of these soil database and bring them into full play.

Key words Soil spectrum; Soil TN prediction; Local modeling; Partial Least Squares Regression (PLSR); Locally weighted regression (LWR)

(责任编辑:檀满枝)



土壤学报

(Turang Xuebao)



第 52 卷 第 2 期 2015 年 3 月

目次

综述与评论

- 理论施氮量的改进及验证——兼论确定作物氮肥推荐量的方法 巨晓棠(249)
- 森林土壤氧化亚氮排放对大气氮沉降增加的响应研究进展 方华军 程淑兰 于贵瑞等(262)

研究论文

- 商丹盆地黄土母质发育土壤的微形态特征及其诊断意义 庞奖励 黄春长 查小春等(272)
- 基于土壤氮素平衡的氮肥推荐方法——以水稻为例 宁运旺 张永春(281)
- 三峡库区紫色土的碎石分布特征 王小燕 王天魏 蔡崇法等(293)
- 黄土区坡耕地土壤结皮对入渗的影响 吴秋菊 吴佳 王林华等(303)
- 基于局部加权回归的土壤全氮含量可见-近红外光谱反演 陈颂超 冯来磊 李硕等(312)
- 五价砷在中国南方两类典型土壤中的迁移特征 李璐璐 张华 周世伟等(321)
- 柠檬酸盐对黑云母表面溶解及微结构变化的影响 李继福 张文君 鲁剑巍等(329)
- 活化过硫酸钠氧化土壤对挥发性有机污染物吸附特性的影响 丁浩然 王磊 龙涛等(336)
- 退化喀斯特森林植被自然恢复中土壤有机碳 $\delta^{13}C$ 值特征 黄宗胜 喻理飞 符裕红等(345)
- 缓释复合肥对茄子产量和不同采果期品质的影响 王菲 李银科 王正银等(355)
- 氮肥和秸秆施用对稻麦轮作体系下土壤剖面 N_2O 时空分布的影响 潘晓健 刘平丽 李露等(364)
- 应用 ^{15}N 示踪法研究咸水滴灌棉田氮肥去向 张文 周广威 闵伟等(372)
- 中国南方基础工程建设中土壤生态修复技术体系与实践 徐国钢 程睿 赖庆旺等(381)
- 黑土农田施加 AM 菌剂对大豆根际菌群结构的影响 宋福强 程蛟 常伟等(390)
- 自生固氮菌对土壤钾的活化作用 张亮 袁玲 黄建国(399)
- 长期保护性耕作对纤维素降解基因 *cbh I* 多样性的影响 陈坤 李传海 朱安宁等(406)
- 模拟条件下土壤硝化作用及硝化微生物对不同水分梯度的响应 刘若萱 张丽梅 白刃等(415)
- 水力侵蚀影响下土壤有机碳和微生物数量动态变化特征 喻为 李忠武 黄金权等(423)

研究简报

- 土壤制图中土壤类型配色模型构建与应用 徐爱国 张维理 张怀志等(431)
- 土壤专题图中采样点点位标识模型的构建 张怀志 张维理 徐爱国等(440)
- 非对称混合电解质体系中恒电荷颗粒表面与 Stern 电位的测定 樊会敏 赵艳 刘新敏等(446)
- 植被重建下露天煤矿排土场边坡土壤碳储量变化 李俊超 党廷辉 薛江等(453)
- 滨海盐渍化土壤理化性质与小麦生产间的关系 谢文军 张衍鹏 张森等(461)

信息

- 第 20 届国际土壤与耕作研究组织大会 (414)
- 《土壤学报》2014 年度审稿专家名录 (467)

封面图片: 陕西丹凤盆地土壤黏化层典型微形态(由庞奖励提供)

CONTENTS

Reviews and Comments

- Improvement and validation of theoretical N rate (TNR)—Discussing the methods for N fertilizer recommendation Ju Xiaotang(261)
- Study on the responses of nitrous oxide emission to increased nitrogen deposition in forest soils; A review Fang Huajun, Cheng Shulan, Yu Guirui, et al. (270)

Research Articles

- Micromorphological features of soils derived from loess in Shangdan Basin and their significance in soil diagnosis Pang Jiangli, Huang Chunchang, Zha Xiaochun, et al. (279)
- Soil nitrogen balance based recommendation of nitrogen fertilization; A case study of rice ... Ning Yunwang, Zhang Yongchun(292)
- Spatial distribution of rock fragments in purple soil in Three Gorges Reservoir Area Wang Xiaoyan, Wang Tianwei, Cai Chongfa, et al. (301)
- Effects of soil crusts on infiltration in slope land in the Loess Area Wu Qiuju, Wu Jia, Wang Linhua, et al. (311)
- Vis-NIR spectral inversion for prediction of soil total nitrogen content in laboratory based on locally weighted regression Chen Songchao, Feng Lailei, Li Shuo, et al. (319)
- Migration of pentavalent antimony in two types of soils typical to South China Li Lulu, Zhang Hua, Zhou Shiwei, et al. (327)
- Effects of citrate dissolving surface and altering microstructure of biotite Li Jifu, Zhang Wenjun, Lu Jianwei, et al. (335)
- Effect of soil oxidation with reactivated sodium persulfate on sorption of volatile organic contaminants Ding Haoran, Wang Lei, Long Tao, et al. (343)
- Characteristics of $\delta^{13}\text{C}$ value of soil organic carbon under naturally restoring degenerative karst forest Huang Zongsheng, Yu Lifei, Fu Yuhong, et al. (353)
- Effects of slow-release compound fertilizer on yield and quality of eggplant relative to timing of harvest Wang Fei, Li Yinke, Wang Zhengyin, et al. (362)
- Spatial and temporal distributions of soil profile N_2O as affected by N fertilization and straw incorporation in the rice-wheat rotation system Pan Xiaojian, Liu Pingli, Li Lu, et al. (371)
- Fate of fertilizer N in saline water drip-irrigated cotton field using ^{15}N tracing method Zhang Wen, Zhou Guangwei, Min Wei, et al. (379)
- Soil ecology remediation technical system and its practice in infrastructure construction in South China Xu Guogang, Cheng Rui, Lai Qingwang, et al. (389)
- Impact of inoculation with AM fungi on microbial community structure in soybean rhizosphere in farmland of black soil Song Fuqiang, Cheng Jiao, Chang Wei, et al. (397)
- Mobilization of potassium in soils by *Azotobacter* Zhang Liang, Yuan Ling, Huang Jianguo(404)
- Effects of long-term conservation tillage on diversity of cellulose degradating gene *cbh I* in Fluvo-aquic soil Chen Kun, Li Chuanhai, Zhu Anning, et al. (413)
- Response of nitrification and nitrifier to change in soil moisture content under simulated conditions Liu Ruoxuan, Zhang Limei, Bai Ren, et al. (421)
- Dynamics of soil microbial population and organic carbon under water erosion Yu Wei, Li Zhongwu, Huang Jinquan, et al. (430)

Research Notes

- Construction of a color matching model for soil types and its application to soil mapping Xu Aiguo, Zhang Weili, Zhang Huaizhi, et al. (439)
- Development of soil sampling point labeling model for the matic soil maps Zhang Huaizhi, Zhang Weili, Xu Aiguo, et al. (445)
- Principle for determination of surface potential and Stern plane potential of constant charged particles in asymmetric mixed electrolyte solutions Fan Huimin, Zhao Yan, Liu Xinmin, et al. (452)
- Variability of soil organic carbon storage in dump slope of opencast coal mine under revegetation Li Junchao, Dang Tinghui, Xue Jiang, et al. (460)
- Relationships between soil physicochemical properties and wheat production in coastal saline soil Xie Wenjun, Zhang Yanpeng, Zhang Miao, et al. (466)

Cover Picture: Typical micromorphological photos in Horizon Bt of the soil profile in Danfeng Basin, Shaanxi Province, China (by Pang Jiangli)