

# 土壤中红外光谱库支持下的局部建模集优化

沈佳丽<sup>1</sup>, 陈颂超<sup>2,3</sup>, 洪永胜<sup>3</sup>, 李硕<sup>1,4\*</sup>

(1. 华中师范大学地理过程分析与模拟湖北省重点实验室, 湖北 武汉 430079;

2. 浙江大学杭州国际科创中心, 浙江 杭州 311200;

3. 浙江大学农业遥感与信息技术应用研究所, 浙江 杭州 310058;

4. 农业农村部光谱检测重点实验室, 浙江 杭州 310058)

**摘要:** 土壤中红外(MIR)光谱能快速、无污染、低成本地估算土壤有机碳等理化属性。随着各种尺度土壤光谱库的建立, 使用其进行快速土壤分析引起广泛关注, 但光谱库的通用模型在局部尺度上的预测效果不理想。开发“局部化”光谱建模方法是提高土壤光谱库性能的有效途径。本文提出了一种新的方法, 通过光谱相似度计算和建模子集构建, 旨在从库中快速建立最优局部建模集以提高预测精度。比较了欧氏、马氏、余弦三种距离算法衡量待测样本与库样本之间的相似度并生成距离矩阵; 使用连续统去除法从距离矩阵中提取库容曲线中的特征点。利用偏最小二乘回归建立土壤 MIR 光谱与有机碳含量间的定量关系。结果表明, 三种距离算法结合连续统去除得到的第一特征点均可得到较佳的预测精度。马氏距离不仅模型精度最高( $R^2 = 0.764$ ,  $RMSE = 1.021\%$ )而且用到的库样本数最少(14% 库容)。本方法可改善 MIR 光谱分析的成本效率并能提高局部尺度的预测能力。

**关键词:** 土壤碳; 相似度; 距离矩阵; 连续统去除; 偏最小二乘回归

**中图分类号:** TP79; S151.9 **文献标识码:** A

## Novel local calibration optimization from soil mid-infrared spectral library

SHEN Jia-Li<sup>1</sup>, CHEN Song-Chao<sup>2,3</sup>, HONG Yong-Sheng<sup>3</sup>, LI Shuo<sup>1,4\*</sup>

(1. Key Laboratory for Geographical Process Analysis & Simulation of Hubei Province, Central China Normal University, Wuhan 430079, China;

2. ZJU-Hangzhou Global Scientific and Technological Innovation Center, Hangzhou 311200, China;

3. Institute of Remote Sensing and Information Technology, Zhejiang University, Hangzhou 310058, China;

4. Key Laboratory of Spectroscopy Sensing, Ministry of Agriculture and Rural Affairs, Hangzhou 310058, China)

**Abstract:** Soil mid-infrared (MIR) can provide a rapid, non-polluting, and cost-efficient method for estimating soil properties, such as soil organic carbon (SOC). Although there is a wide interest in using the soil spectral library (SSL) for soil analysis at various scales, the SSL with a general calibration often produces poor predictions at local scales. Therefore, developing methods to ‘localize’ the spectroscopic modelling is a reliable way to improve the use of SSL. In this study, we proposed a new approach that aims to rapidly build the optimal local model from the SSL by calculating the spectral similarity and developing the local calibration, in order to further improve the prediction accuracy. The distance matrix was constructed by three distance algorithms, namely Euclidean distance, Mahalanobis distance, and Cosine distance, which were compared and used to measure the similarity between the local samples and the SSL. The capacity curve, which was taken from the distance matrix, was used with a method called “continuum-removal” to find the feature points. Partial least-squares regression was used to build the spectroscopic models for SOC estimation. We found that for all three distance algorithms combined with the continuum-removal, the local calibration derived from the first feature point gave us a good idea of how accurate the prediction would be. The Mahalanobis distance can effective-

收稿日期: 2022-08-25, 修回日期: 2023-02-28

Received date: 2022-08-25, Revised date: 2023-02-28

基金项目: 国家自然科学基金(41601370)、农业农村部光谱检测重点实验室开放基金课题(2022ZJUGP003)和中央高校基本科研业务费专项资金(CCNU22JC022)

**Foundation items:** Supported by the Young Scientists Fund of the National Natural Science Foundation of China (41601370); the Key Laboratory of Spectroscopy Sensing, Ministry of Agriculture and Rural Affairs, P. R. China (2022ZJUGP003); the Fundamental Research Funds for the Central Universities (CCNU22JC022)

作者简介(Biography): 沈佳丽(1998-), 女, 浙江嘉兴人, 华中师范大学城市与环境科学学院硕士研究生. E-mail: shenjienie@mails.ccnu.edu.cn

\*通讯作者(Corresponding author): E-mail: shuoguo@zju.edu.cn

ly develop the optimal local calibration from the MIR SSL, which not only achieved the best accuracy ( $R^2 = 0.764$ , RMSE = 1.021%) but also used the least number of samples from SSL (14% SSL). On local scales, the approach we proposed can significantly improve both the analytical cost and the accuracy of the soil MIR technique.

**Key words:** soil carbon, similarity, distance matrix, continuum-removal, PLSR

## 引言

中红外(Mid-Infrared, MIR)光谱技术的应用有助于土壤的定性及定量分析,可以节约传统土壤分析需要耗费的大量时间,同时也可以减少分析成本。MIR技术不仅能用于土壤肥力分析,如有机质、pH等<sup>[1]</sup>,还能用于检测污染物,如重金属、农药等<sup>[2]</sup>,可为土壤改良及施肥提供重要的技术支撑。联合国粮农组织(FAO)于2017年成立了全球土壤实验室网络(GLOSOLAN),并在此框架下于2020年发起了关于土壤光谱学的倡议,旨在促进各国及地区的实验室使用光谱分析土壤<sup>[3]</sup>。

近年来,国家<sup>[4]</sup>、大陆<sup>[5]</sup>和全球尺度<sup>[6]</sup>土壤光谱库的建立,为土壤分析提供了数据基础。库包含了土壤样本的理化属性数据、光谱数据,以及坐标、土壤类型、土地利用/覆盖等信息。对于土壤的定量分析(如有机碳),MIR通常比可见近红外(visible - Near Infrared, vis - NIR)具有更佳的能力<sup>[7-10]</sup>。但当前土壤光谱库的建设以vis - NIR居多,这是由于MIR传感器对温控的要求较高、仪器昂贵,且便携式仪器选择有限。目前极少数的土壤MIR光谱库已经建成,如全球库<sup>[11]</sup>、瑞士库<sup>[12]</sup>,土壤有机碳(SOC)、pH等关键土壤属性也得到了较好的定量效果,但MIR光谱库在各国及地区的建立仍待推进。

即便有无限容量的库,全局数据建立的通用关系在应用到局部尺度时面临挑战。这是由于全局模型(Global)包含了多样的土壤变异信息,从而难以准确表征甚至会掩盖局部样本的变异,尤其当待测样本来自多个局部时<sup>[13]</sup>。开发新的算法和模型对于提高土壤光谱库的性能是一种有效途径。此外,基于相似的光谱具有相似的土壤成分这一假设,开发“局部化”光谱建模的方法也同样重要。先将少量且具代表性的待测样本入库,令其尽可能囊括待测样本集的主要特征,再从库中匹配到与待测集相似的样本集合用于建模的数据驱动策略,是当前主流。相关方法如Spiking<sup>[14]</sup>、RS-Local<sup>[15]</sup>、C-KS<sup>[16]</sup>等已对土壤vis - NIR光谱库进行了测试并取得积极效果,但尚不清楚适用于vis - NIR的方法是否同样也适用于MIR。而且当库已包含一定的局部

样本特征信息时,如何选择恰当的样本数用于建立局部模型以获得较佳的精度尚无定论。

本文探索并提出了一种从土壤MIR光谱库中快速确定局部建模集容量的方法,以期提升MIR光谱分析技术的效果,同时响应FAO发起的GLOSOLAN-Spec倡议。

## 1 材料与方法

### 1.1 数据来源

研究数据来自国际土壤参比与信息中心(International Soil Reference and Information Center)发布的全球土壤光谱库(Global Soil Spectral Library, GSSL)<sup>[17]</sup>。本研究从库中筛选出同时含有MIR光谱及SOC含量的有效数据共计677个土柱样点(不同土层样本共计3755个),采样点广泛分布在除南极洲之外的六个洲(图1)。经风干、研磨、过筛( $<100\mu\text{m}$ )后,土样的MIR光谱由Bruker VERTEX 70红外光谱仪在实验室测得,吸收光谱的波段范围为 $7498\sim 600\text{cm}^{-1}$ ,光谱分辨率 $4\text{cm}^{-1}$ 。该仪器由液氮制冷,配备金板做谱线校准。

### 1.2 光谱预处理与集合划分

土壤MIR光谱采用常用的预处理流程:先重采样至 $16\text{cm}^{-1}$ 并保留 $4000\sim 650\text{cm}^{-1}$ 波段范围,再经Savitzky - Golay平滑法(2阶3窗口)和一阶微分对光谱进行噪声消除和信号增强处理。

光谱数据集的划分使用Kennard - Stone(KS)法,以避免因验证集特征与建模集相近而令预测结果过优,当前也已成为深入研究光谱库时构建代表性样本集合的方法<sup>[18-19]</sup>。本研究基于光谱前三个主成分,使用KS法以土柱为单位从总样本中按约5:1选出3167个样本组成土壤光谱库(记作SSL);剩余588个样本做验证集,用于模型精度的检验(记作Test)。

在光谱分析时,使用主成分分析法(Principal Components Analysis, PCA)对预处理后的光谱数据进行降维,以实现少量不相关的变量来表征样本的主要特征。SSL和Test前三个主成分贡献率分别达55.2%和59.9%。

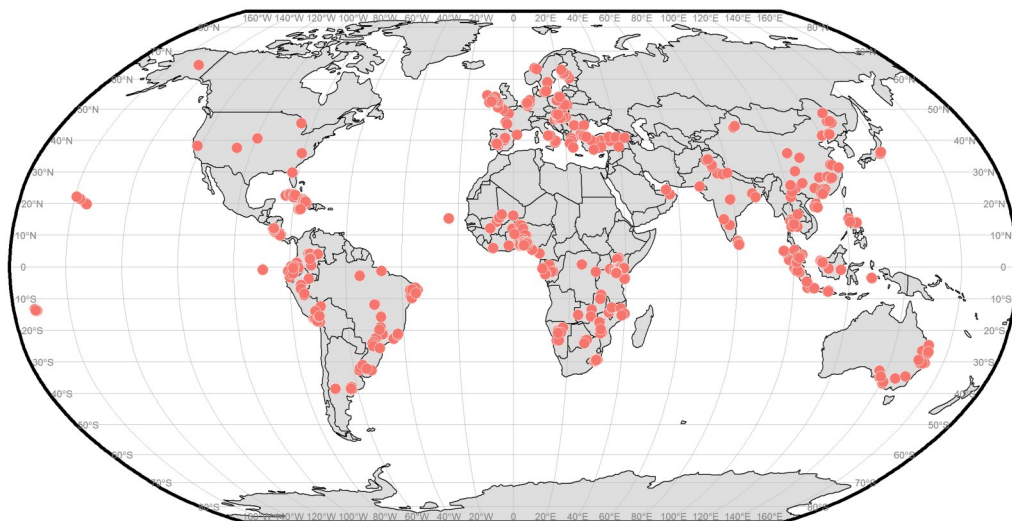


图1 全球土壤光谱库中677个包含中红外光谱和土壤有机碳的土样分布

Fig. 1 677 site locations of with MIR spectra and SOC in the Global Soil Spectral Library

SSL和Test的SOC含量分布特征如表1所示,前者的均值同后者接近,但SSL的SOC含量范围跨度以及标准差值都显著高于Test。SSL包含的大量土壤变异与土壤样品的成土因素如母质、气候、地形和土地利用类型等密切相关。

表1 土壤有机碳含量的特征统计(%)

Table 1 Characteristic statistics of SOC content (%)

数据集	样本数	最小值	第1四分位数	中值	均值	第3四分位数	最大值	标准差
SSL	3167	0.00	0.21	0.47	1.19	1.18	60.00	2.75
Test	588	0.00	0.24	0.55	1.21	1.29	24.00	2.03

### 1.3 局部建模集优化

光谱经预处理后,从SSL中优化局部建模集(Local)的过程分两步:光谱相似度计算和建模子集构建(图2)。

本文考察了欧式距离(Euclidean Distances, ED)、马氏距离(Mahalanobis Distances, MD)和余弦距离(Cosine Distance, CD)三种经典方法来衡量Test样本与SSL样本之间的相似度。其中,ED和MD得到的距离值越小表示越相似,反之越相异;CD计算的是夹角余弦,值越接近0表示越相异,越接近1表示越相似<sup>[20]</sup>。计算过程如图2左下所示,Test( $n$ 个样本)中每一个待测样本都与SSL( $m$ 个样本)的所有样本进行相似度计算,得到的 $n \times m$ 个距离值经排序之后组成距离矩阵(ED和MD按升序排,CD按降

序)。随着距离矩阵行数的逐渐增加(从1到 $m$ ),累积从SSL匹配到无重复的相似样本数(记作“库容”)也随之增加。

连续统去除也称去包络线,常用于提取连续变量的局部变化特征<sup>[6]</sup>。如图2右下所示,先从距离矩阵提取库容曲线,再对该曲线进行连续统去除,然后通过设定可选阈值 $p$ 和必选阈值 $q$ 分别约束矩阵行数 $\Delta x$ 和库容 $\Delta y$ ,提取库容曲线上陡变之处的“特征点”。 $p$ 值代表距离矩阵的行间隔,相邻行( $p=1$ )包含的样本信息存在显著冗余; $q$ 值过小则有更多的变化特征被选入,反之则更少。最后,用每个特征点的位置(第 $x_i$ 行)对应其从库中累积(前 $x_i$ 行)匹配到无重复样本,组成一系列的Local集。

矩阵第一行是从库中匹配到的最相似样本,样本数位于库容曲线的起点,所以该点位不会出现在连续统去除的曲线上。本文以距离矩阵第一行作为考察的起点,将其库容对应的样本所组成的建模集记作“Local<sub>0</sub>”;以全库作为考察终点,记作“Global”。

### 1.4 建模方法及评价指标

本研究用经典的偏最小二乘回归(Partial Least-Squares Regression, PLSR)建立土壤MIR光谱与SOC含量间的定量关系模型,并用留一法交叉检验以避免模型校正过拟合<sup>[21]</sup>。评估和比较模型的性能使用决定系数(Coefficient of Determination,  $R^2$ ),评估模型的误差使用均方根误差(Root Mean Square Errors, RMSE)。较高的 $R^2$ 和较低的RMSE表

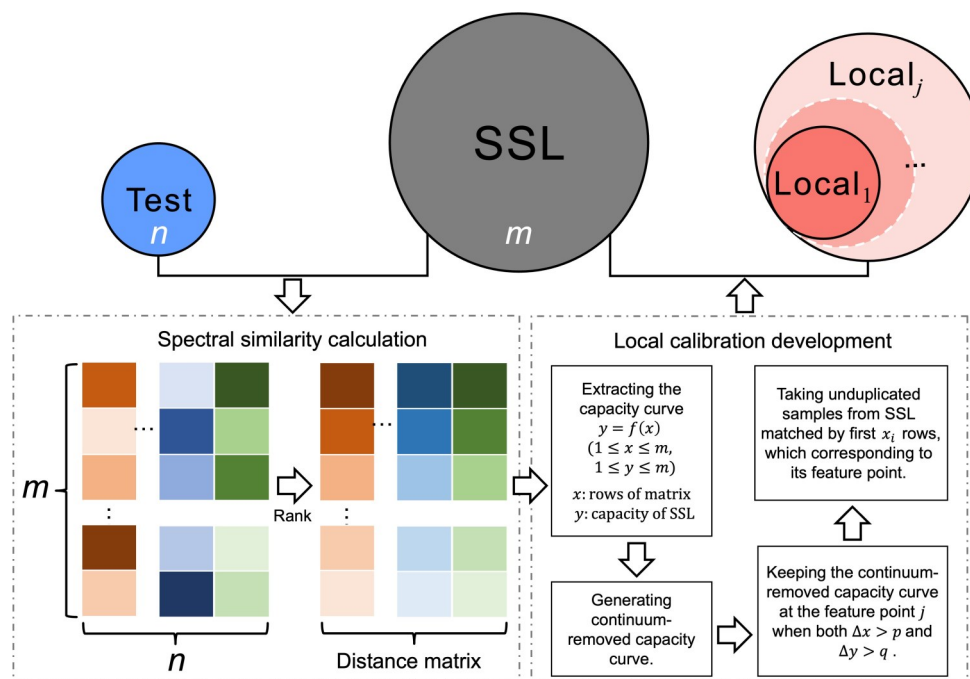


图2 局部建模集优化流程

Fig. 2 The scheme of optimizing local calibrations

明模型较好。当预测精度差异不显著时,使用更少的库容则更佳。在R软件中,光谱数据处理使用prospectr包,PLSR使用pls包。

## 2 结果与讨论

### 2.1 光谱分析

图3(a)和3(b)分别是SSL与Test的平均光谱及95%置信区间。土壤MIR吸收光谱在 $4\ 000\sim 2\ 000\ \text{cm}^{-1}$ 范围内表现剧烈,而在 $2\ 000\sim 650\ \text{cm}^{-1}$ 呈现较多峰谷但幅度较小。在 $2\ 920\ \text{cm}^{-1}$ 和 $2\ 850\ \text{cm}^{-1}$ 与 $-\text{CH}_2$ 伸缩振动有关,而在 $1\ 400\ \text{cm}^{-1}$ 附近的吸收由有机化合物引起<sup>[22]</sup>。SSL和Test的平均光谱曲线无显著差异,但前者的置信区间更宽,如 $3\ 500\ \text{cm}^{-1}$ 和 $1\ 500\sim 650\ \text{cm}^{-1}$ 附近。图3(c)和3(d)分别是SSL和Test光谱的前三个主成分特征向量曲线,共同解释了超55%的总体变异。PC1的贡献最大,且两者的PC1曲线波形相似,前述 $2\ 920\ \text{cm}^{-1}$ 、 $2\ 850\ \text{cm}^{-1}$ 和 $1\ 400\ \text{cm}^{-1}$ 附近与C有关的吸收也体现在该曲线上;PC2和PC3的曲线波形仅在部分波段表现出相似的波形。这表明,SSL与Test相似的特征信息主要保存在PC1中。

### 2.2 阈值设定与预测精度对比

随着行数的增加(图4(a)),ED、MD和CD的库容曲线均呈快速增长直至在约200行开始趋缓。放

大观察可发现三条曲线并不完全重合,并随着各行样本量的加入,曲线斜率也发生相应变化,细节如图4(b)所示。在约200行之前,ED、MD和CD存在多处高于5的峰,在600~800行范围也存在少量的高峰,峰值在850行之后均低于5。三者连续统去除曲线在2300或2400行附近陆续开始平直,表明局部建模集的库容已达全库水平。

图4(c)和4(d)展示了三种距离算法各自3167个模型预测同一Test集的精度。以图4(c)为例,三条曲线的 $R^2$ 值从第一行开始增长并在200行内波动且维持较高位,至200行附近出现陡降,随后在600~800行范围又出现了小幅跃升之后趋于平缓,最终收敛至全库的 $R^2$ 。表2显示,ED、MD和CD在第一行( $\text{Local}_0$ )的库容均达300个左右。虽然MD的RMSE值高于全库(Global)之外,但其 $R^2$ 明显高于ED、CD以及Global。总体而言,三者 $\text{Local}_0$ 的各指标均优于Global,因此,从前200行快速定位出精度同时优于 $\text{Local}_0$ 和Global的行位置是关键。

图5更详细地展现了前200行的曲线变化情况。经连续统去除后(图5(b)),ED、MD和CD的第一个峰均高于5,结合图4(b)的全局变化,本研究将图2中的必选阈值 $q$ 设置为5。连续统去除后所有大于5的“特征点”以相应颜色的圆点标于图5各曲线上。而预测精度曲线(图5(c)、5(d))显示,相邻

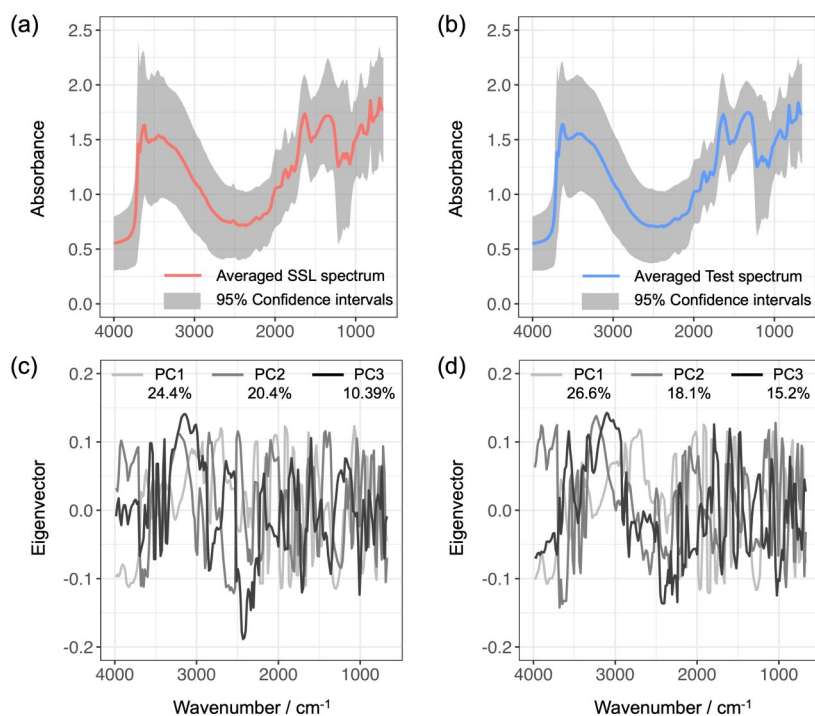


图3 SSL(a,c)和Test(b,d)的平均光谱曲线及95%置信区间

Fig. 3 Average absorbance with 95% confidence intervals of SSL (a, c) and Test (b, d)

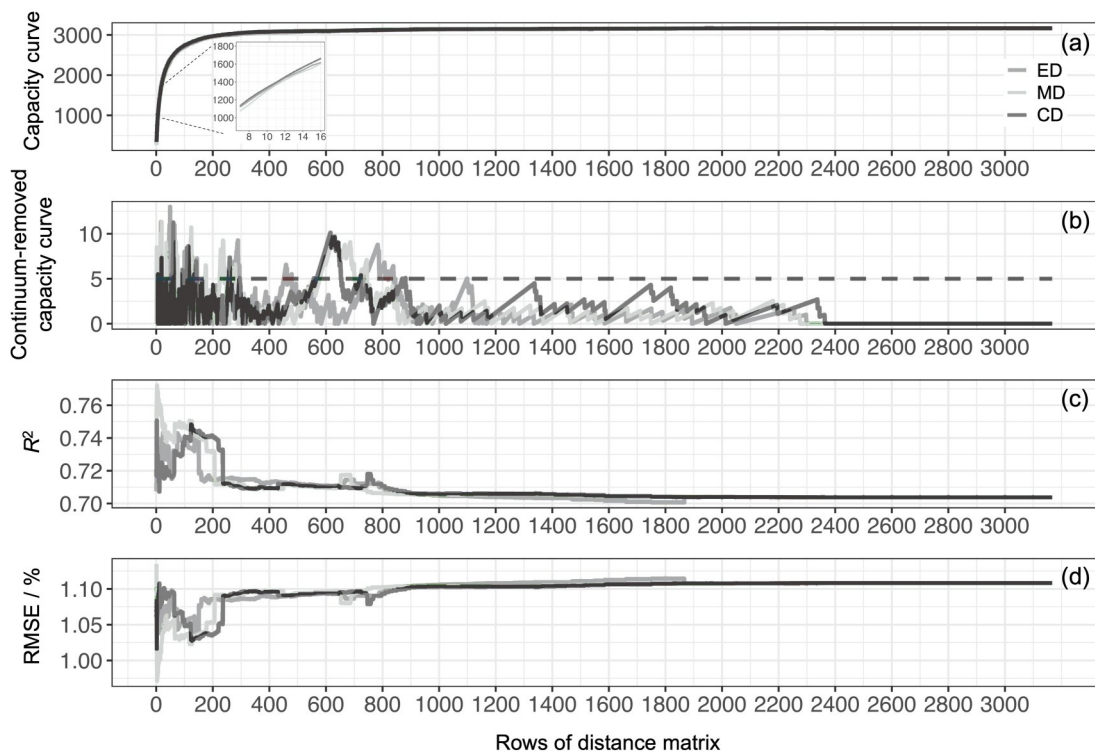


图4 距离矩阵所有行的库容(a)、连续统去除(b)及模型精度(c,d)的变化

Fig. 4 The capacity (a), continuum-removal (b) and model accuracy (c, d) for all rows of the distance matrix

行特征点的精度几乎一致。因此,本文建议将可选阈值 $p$ 设置为1,即当 $p > 1$ 时非相邻的特征点可被

保留。本研究在局部建模集的优化过程中对 $p$ 和 $q$ 的阈值设定仅供参考,在今后的研究中其普适性还

表2 距离矩阵首行、第一特征点组成的局部建模集同全库的预测精度对比

Table 2 The comparison of prediction accuracy from Local<sub>0</sub>, Local<sub>1</sub> and Global

数据集	Local <sub>0</sub>			Local <sub>1</sub>			Global		
	R <sup>2</sup>	RMSE	库容	R <sup>2</sup>	RMSE	库容	R <sup>2</sup>	RMSE	库容
ED	0.708	1.105%	333	0.733	1.060%	926	0.704		
MD	0.752	1.135%	262	0.764	1.021%	430		1.108%	
CD	0.715	1.086%	338	0.730	1.066%	1038			3167

需更多方的实验加以印证。

以图5(c)为例,ED、MD、CD的R<sup>2</sup>曲线总体在前50行和100~150行区间范围存在较高值。图5(a)显示,50行附近三者的库容已增长至约75%,150行时则已接近全库。因此,本研究继续从前50行寻找合适的局部建模集。图5(c)、5(d)显示,ED、MD和CD由第一特征点组成建模集(Local<sub>1</sub>),其精度均优于各自的Local<sub>0</sub>,即比Local<sub>0</sub>的R<sup>2</sup>更高且RMSE更低。结合表2,三者Local<sub>1</sub>的精度也均优于Global。

MD在Local<sub>1</sub>的精度优于ED和MD,且430个样本的库容相较于ED的926个和CD的1038个明显更少。而ED和CD此处的精度差异不显著,两点在图5(c)、5(d)中几乎贴在一起。因此,MD更适用于对土壤MIR光谱进行相似度的衡量,这与前人对土壤vis-NIR光谱进行的相关研究结论不同,他们认

为CD更佳<sup>[20,23-24]</sup>。这说明MIR与vis-NIR不仅数据表征不同,还存在方法上的适用性差别。

值得注意的是,三者第一特征点的位置并非对应前5行R<sup>2</sup>最高点和RMSE最低点,这可能是由于该数据包含了大量的土壤变异信息,前三个主成分累计贡献也未超过80%所致。即便如此,寻找并确定Local<sub>1</sub>的过程简单、快速,且其精度优于首行样本集和全库,已达到本研究的目的。

### 2.3 主成分空间分布

以MD为例,图6中的红色点是其Local<sub>0</sub>、前4个特征点(Local<sub>1</sub>~Local<sub>4</sub>)和Global的库容在主成分空间的分布情况,蓝色点代表待测样本集Test。从Local<sub>0</sub>开始,局部建模集主成分空间的覆盖范围(以阴影表示)随着库容的增大而随之扩大并发生形变。Local<sub>0</sub>的库容占8%,其阴影范围接近Test但未完全

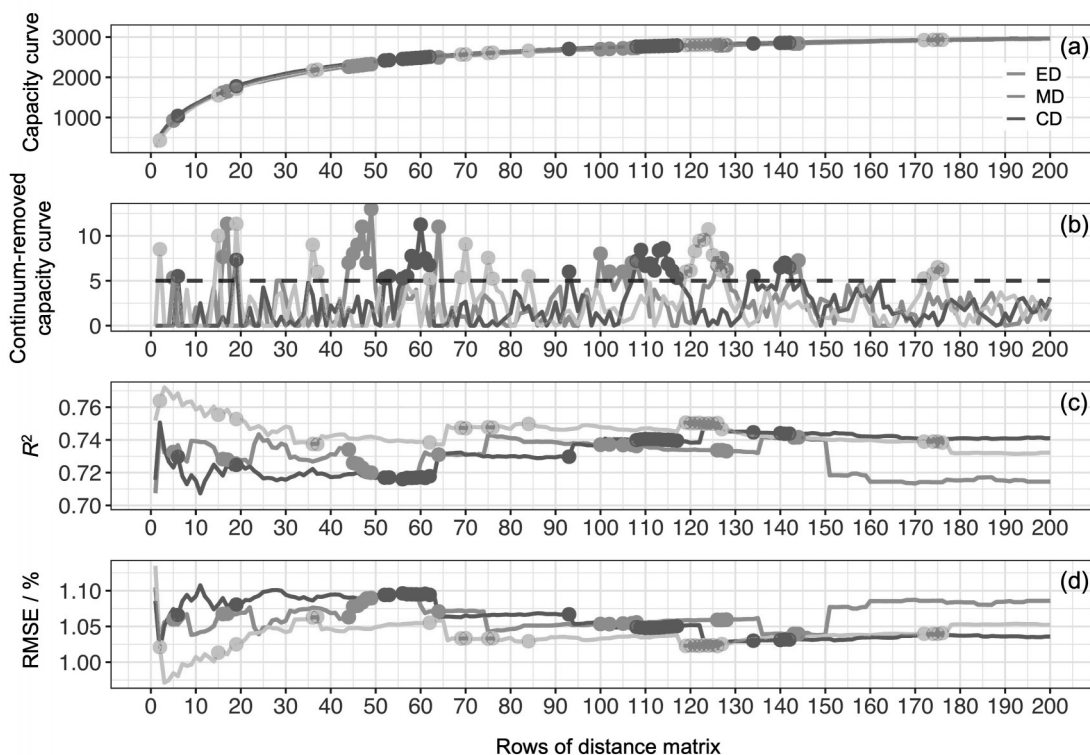


图5 距离矩阵前200行的库容(a)、连续统去除(b)和模型预测精度(c,d)的变化

Fig. 5 The capacity (a), continuum-removal (b) and model accuracy (c, d) for the first 200 rows of the distance matrix

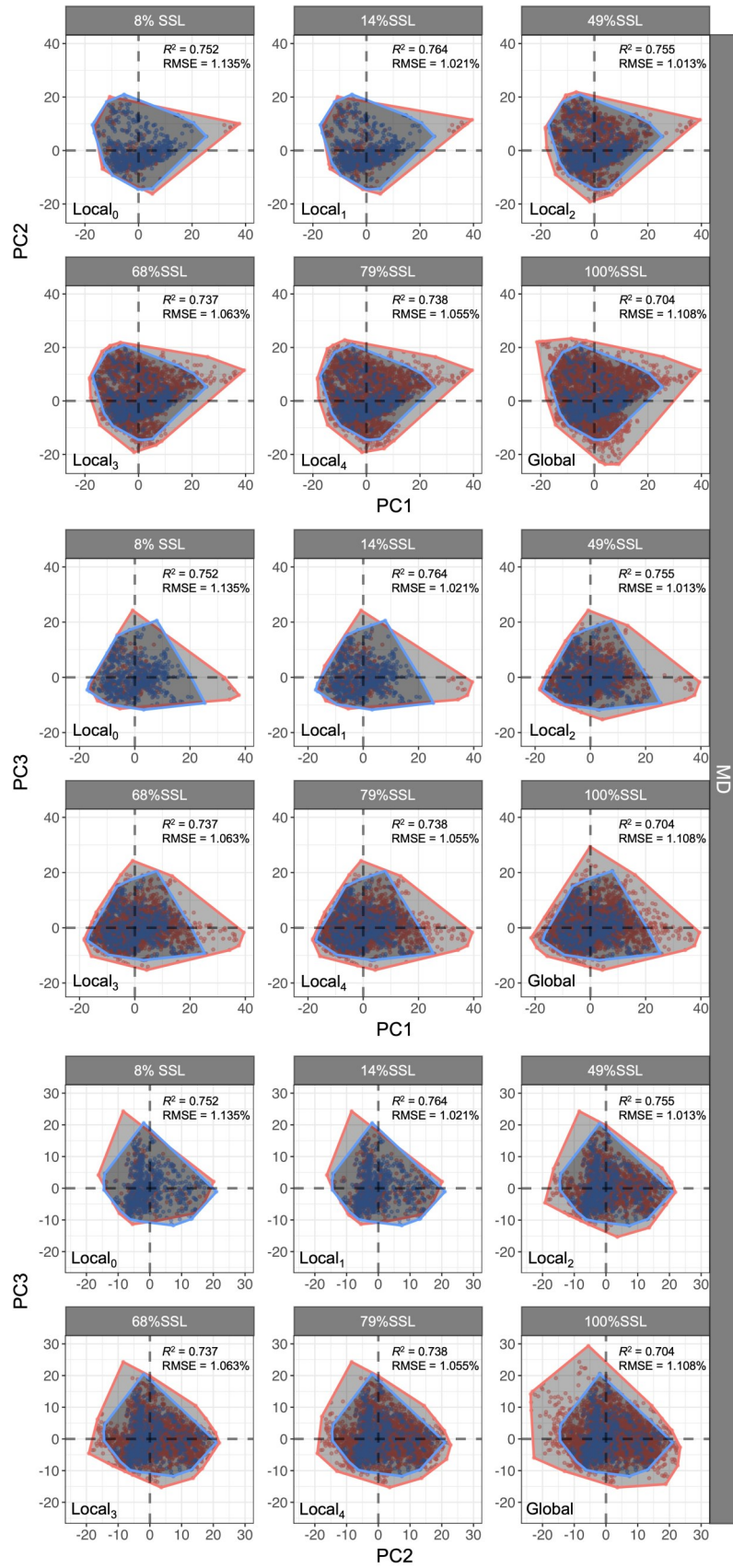


图6 基于MD法组建的五个Local集(红点)投影到Test集(蓝点)主成分空间的分布

Fig. 6 The PCA score plot from five Locals (red dots) projected onto Test (blue dots) by MD

覆盖。结合表2说明,尽管Local<sub>0</sub>是从SSL中匹配出的与Test最相似的样本,但由于Local<sub>0</sub>未能完全表征Test,预测精度虽然较高但仍未达到最佳水平,这与彭青青等的研究吻合<sup>[20]</sup>。

占库14%的Local<sub>1</sub>的阴影范围与Local<sub>0</sub>接近,但前者因样本的加入而使其范围更大,如PC12的第三象限、PC13的第二象限、PC23的第一象限都或多或少超越了Test,从而使Local<sub>1</sub>的精度优于Local<sub>0</sub>。随着库容的逐渐增大伴随着阴影范围显著超越Test,Local<sub>1</sub>之后的局部建模集相应的精度也开始下降,这种现象表明,越来越多与Test相异的样本被加入到了Local集合中<sup>[25]</sup>。因此,Local<sub>1</sub>是当前优化框架下得到的较为理想的局部建模集。

### 3 结论

本研究针对土壤MIR光谱库提出了一种快速构建最优局部建模集的方法,包含两步:光谱相似度计算和建模子集构建。通过比较ED、MD和CD三种距离算法来衡量光谱相似度并得到距离矩阵,使用连续统去除法来提取库容曲线上的特征点,并考察了每种距离算法各自3167个模型对同一个待测集合的预测和检验。结果表明:

(1) ED、MD和CD三种方法的第一特征点均可得到较佳的预测精度;

(2) 对库容曲线进行连续统去除是提取特征点的有效手段;

(3) 待测样本通过MD法从土壤MIR光谱库构建的最优局部建模集预测精度较ED和CD更优,所用库样本数最少。

### References

[1] Kumar M, Jain V. Mid-infrared spectroscopy for fast, accurate and reliable soil analysis [J]. *Trends in Analytical Chemistry*, 2018, **103**: 216 - 224.

[2] Fang H, Zhou K, Wang S, et al. Soil fertility analysis based on mid-infrared spectroscopy [J]. *International Agrophysics*, 2018, **32**(4): 553 - 563.

[3] Ge Y, Wadoux A, Peng Y. A primer on soil analysis using visible and near-infrared (vis-NIR) and mid-infrared (MIR) spectroscopy [M]. Rome, FAO. 2022.

[4] Shi Z, Wang Q L, Peng J, et al. Development of a national VNIR soil-spectral library for soil classification and prediction of organic matter concentrations [J]. *Science China: Earth Sciences*, 2014, **44**(5): 978 - 988. (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 [J]. *中国科学: 地球科学*, 2014, **44**(5): 978 - 988.

[5] Orgiazzi A, Ballabio C, Panagos P, et al. LUCAS Soil, the largest expandable soil dataset for Europe: a review [J]. *Eu-*

*ropean Journal of Soil Science*, 2018, **69**(1): 140 - 153.

[6] Viscarra Rossel R A, Behrens T, Ben-Dor E, et al. A global spectral library to characterize the world's soil [J]. *Earth-Science Reviews*, 2016, **155**: 198 - 230.

[7] Hong Y, Munnaf M A, Guerrero A, et al. Fusion of visible-to-near-infrared and mid-infrared spectroscopy to estimate soil organic carbon [J]. *Soil and Tillage Research*, 2022, **217**: 105284.

[8] Chen S C, Peng J, Ji W J, et al. Study on the characterization of VNIR-MIR spectra and prediction of soil organic matter in paddy soil [J]. *Spectroscopy and Spectral Analysis*, 2016, **36**(06): 1712 - 1716. (陈颂超, 彭杰, 纪文君, 等. 水稻土可见-近红外-中红外光谱特性与有机质预测研究 [J]. *光谱学与光谱分析*, 2016, **36**(06): 1712 - 1716.

[9] Meng X X, Yu L, Zhou Y, et al. Predicting organic carbon using datafusion of visible near-infrared and middle infrared spectra by proximal soil sensing [J]. *Chinese Journal of Soil Science*, 2022, **53**(2): 301 - 307. (孟鑫鑫, 于雷, 周勇, 等. 基于可见近红外和中红外近地面光谱数据融合的土壤有机碳含量反演 [J]. *土壤通报*, 2022, **53**(2): 301 - 307.

[10] Ji W J, Shi Z, Zhou Q, et al. VIS-NIR reflectance spectroscopy of the organic matter in several types of soils [J]. *Journal of Infrared and Millimeter Waves*, 2012, **31**(3): 277 - 282. (纪文君, 史舟, 周清, 等. 几种不同类型土壤的VIS-NIR光谱特性及有机质响应波段研究 [J]. *红外与毫米波学报*, 2012, **31**(3): 277 - 282.

[11] Terhoeven-Urselmans T, Vagen T G, Spaargaren O, et al. Prediction of soil fertility properties from a globally distributed soil mid-infrared spectral library [J]. *Soil Science Society of America Journal*, 2010, **74**(5): 1792 - 1799.

[12] Baumann P, Helfenstein A, Gubler A, et al. Developing the Swiss soil spectral library for local estimation and monitoring [J]. *SOIL Discussions*, 2021: 1 - 32.

[13] Luce M S, Ziadi N, Viscarra Rossel R A. GLOBAL-LOCAL: A new approach for local predictions of soil organic carbon content using large soil spectral libraries [J]. *Geoderma*, 2022, **425**: 116048.

[14] 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 [J]. *Geoderma*, 2010, **158**(1-2): 66 - 77.

[15] Lobsey C R, Viscarra Rossel R A, Roudier P, et al. rs-local data-mines information from spectral libraries to improve local calibrations [J]. *European Journal of Soil Science*, 2017, **68**(6): 840-852.

[16] Liu Y F, Lu Y N, Guo L, et al. Construction of calibration set based on the land use types in visible and near-infrared (VIS-NIR) model for soil organic matter estimation [J]. *Acta Pedologica Sinica*, 2016, **53**(2): 332 - 341. (刘艳芳, 卢延年, 郭龙, 等. 基于地类分层的土壤有机质光谱反演校正样本集的构建 [J]. *土壤学报*, 2016, **53**(2): 332 - 341.

[17] Mid-Infrared Spectra (MIRS) from ICRAF Soil and Plant Spectroscopy Laboratory: Africa Soil Information Service (AfSIS) Phase I 2009-2013 [DB].

[18] Shen Z, Ramirez-Lopez L, Behrens T, et al. Deep transfer learning of global spectra for local soil carbon monitor-



- ing [J]. *ISPRS Journal of Photogrammetry and Remote Sensing*, 2022, **188**: 190–200.
- [19] Chen Y Y, Qi T C, Huang Y J, *et al.* Optimization method of calibration dataset for VIS–NIR spectral inversion model of soil organic matter content [J]. *Transactions of the Chinese Society of Agricultural Engineering*, 2017, **33** (6): 107 – 114. (陈奕云, 齐天赐, 黄颖菁, 等. 土壤有机质含量可见–近红外光谱反演模型校正集优选方法 [J]. *农业工程学报*), 2017, **33**(6): 107 – 114.
- [20] Peng Q Q, Chen S C, Zhou M H, *et al.* Developing of Local Model From Soil Spectral Library with Spectral Dissimilarity [J]. *Spectroscopy and Spectral Analysis*, 2022, **42** (5): 1614 – 1619. (彭青青, 陈颂超, 周明华, 等. 基于土壤光谱库和光谱相异度的局部模型构建 [J]. *光谱学与光谱分析*), 2022, **42**(5): 1614 – 1619.
- [21] Li S, Shi Z, Chen S, *et al.* In situ measurements of organic carbon in soil profiles using vis–NIR spectroscopy on the Qinghai – Tibet plateau [J]. *Environmental Science & Technology*, 2015, **49**(8): 4980 – 4987.
- [22] Li S, Viscarra Rossel R A, Webster R. The cost-effectiveness of reflectance spectroscopy for estimating soil organic carbon [J]. *European Journal of Soil Science*, 2022, **73** (1): e13202.
- [23] Ramirez–Lopez L, Behrens T, Schmidt K, *et al.* Distance and similarity–search metrics for use with soil vis – NIR spectra [J]. *Geoderma*, 2013, **199**: 43 – 53.
- [24] Li H D, Li D C, Zeng R. Estimation of soil organic carbon based on spectral similarity matching [J]. *Acta Pedologica Sinica*, 2021, **58**(5): 1224 – 1233. (李宏达, 李德成, 曾荣. 基于光谱相似性匹配的土壤有机碳估算 [J]. *土壤学报*), 2021, **58**(5): 1224 – 1233.
- [25] Guerrero C, Wetterlind J, Stenberg B, *et al.* Do we really need large spectral libraries for local scale SOC assessment with NIR spectroscopy? [J]. *Soil and Tillage Research*, 2016, **155**: 501 – 509.