

基于水分和粒度的土壤有机质特征波长提取与预测模型

曹永研 杨 玮 王 懂 李 浩 孟 超

(中国农业大学智慧农业系统集成研究教育部重点实验室, 北京 100083)

摘要:为减少水分、粒度对传统方式选取特征波长建立的土壤有机质预测模型的影响,本文提出新的特征波长提取方法。采集中国农业大学上庄实验站土壤样本 60 份,将样本自然风干后一分为二,一份配成 5 个粒度梯度(粒径 2~2.5 mm、1.43~2 mm、1~1.43 mm、0.6~1 mm、0~0.6 mm),另一份过 0.6 mm 筛后配成 5 个水分梯度(含水率 5%、10%、15%、20%、25%)。通过标准仪器分别获取土壤有机质含量真值和土壤光谱信息,使用随机蛙跳算法进行特征波长提取,每个水分、粒度梯度下分别选取 7 个与土壤有机质含量真值相关性较高的波长作为对应梯度下选取的特征波长,分别建立多元线性回归(MLR)、偏最小二乘(PLS)、随机森林(RF)模型,结果表明:随着含水率增高,3 种模型的建模集和预测集决定系数 R^2 基本呈减小趋势;在 2~2.5 mm 粒度梯度下,3 种模型的建模集和预测集 R^2 最低,在 0~0.6 mm 梯度下,建模集和预测集 R^2 最高,其余梯度下,建模集和预测集 R^2 接近。结合滤光片带通范围(± 15 nm),挑选出水分梯度下相同或者接近的 8 个土壤有机质特征波长,粒度梯度下选取 6 个特征波长,最终结合化学键特性在水分梯度和粒度梯度下确定的 14 个特征波长下剔除了 6 个,确定 8 个特征波长:932、999、1 083、1 191、1 316、1 356、1 583、1 626 nm。分别建立 MLR、PLS、RF 模型,结果表明:最终选取的有机质特征波长建立的 3 种模型建模集 R^2 均不低于 0.8、预测集 R^2 均不低于 0.75,其中 PLS 预测效果最佳,建模集、预测集 R^2 分别为 0.880 9、0.840 2。本研究所确定的有机质特征波长建立的模型具有更好的适用性和预测效果,相比于传统方式,一定程度上消除水分、粒度对预测的影响。

关键词: 光谱; 土壤有机质; 特征波段; 水分; 粒度

中图分类号: S237 **文献标识码:** A **文章编号:** 1000-1298(2022)S1-0241-08

Soil Organic Matter Characteristic Wavelength Extraction and Prediction Model Based on Moisture and Particle Size

CAO Yongyan YANG Wei WANG Dong LI Hao MENG Chao

(Key Laboratory of Smart Agriculture Systems, Ministry of Education, China Agricultural University, Beijing 100083, China)

Abstract: In order to reduce the influence of moisture and particle size on the soil organic matter prediction model established by the characteristic wavelengths selected in the traditional way, a method of extracting characteristic wavelengths was proposed. Sixty soil samples were collected from Shangzhuang Experimental Station of China Agricultural University, and the samples were naturally dried and divided into two, one portion was formulated into five particle size gradients (particle size of 2~2.5 mm, 1.43~2 mm, 1~1.43 mm, 0.6~1 mm, and 0~0.6 mm), the other part was sieved through 0.6 mm and formulated into five moisture gradients (5%, 10%, 15%, 20%, and 25% moisture content). The true values of soil organic matter content and soil spectral information were obtained by standard instruments, and the characteristic wavelengths were extracted by using the random frog-hopping algorithm. Totally seven wavelengths with high correlation with the true values of soil organic matter content were selected as the characteristic wavelengths under each moisture and particle size gradient, and multiple linear regression (MLR), partial least squares (PLS) and random forest (RF) models were established respectively. The results showed that the R^2 of the modeling and prediction sets of the three models basically tended to decrease as the water content increased; the R^2 of the modeling and prediction sets of

收稿日期: 2022-06-28 修回日期: 2022-07-30

基金项目: 浙江省科技计划项目(2021C02023)

作者简介: 曹永研(1998—),男,硕士生,主要从事精细农业系统集成研究,E-mail: cyyc@cau.edu.cn

通信作者: 杨玮(1981—),女,副教授,博士生导师,主要从事精细农业系统集成研究,E-mail: cauyw@cau.edu.cn

the three models was the lowest in the gradient of 2 ~ 2.5 mm, highest in the gradient of 0 ~ 0.6 mm, and close to the R^2 of the modeling and prediction sets in the rest of the gradient. Combined with the filter bandpass range of ± 15 nm, eight characteristic wavelengths of soil organic matter under moisture gradient were selected as the same or close to each other, and six characteristic wavelengths under particle size gradient were selected, and finally six wavelengths were eliminated under the 14 characteristic wavelengths determined under moisture gradient and particle size gradient by combining chemical bonding characteristics, and eight characteristic wavelengths were determined as follows: 932 nm, 999 nm, 1 083 nm, 1 191 nm, 1 316 nm, 1 356 nm, 1 583 nm, and 1 626 nm. The MLR, PLS and RF models were established respectively, and the results showed that the R^2 of the modeling set and the R^2 of the prediction set were not less than 0.8 and 0.75 for the three models established by the final selected organic matter characteristic wavelengths, and the best prediction effect was achieved by PLS, with the R^2 of the modeling set and the R^2 of the prediction set being 0.880 9 and 0.840 2, respectively. The model established had better applicability and prediction effect, and the influence of moisture and particle size on prediction was eliminated to a certain extent compared with the traditional way.

Key words: spectrum; soil organic matter; characteristic wavelength; water content; particle size

0 引言

土壤有机质是衡量土壤肥力和土地退化程度的一项重要指标,在环境保护、农业可持续发展中占据重要地位^[1-3]。有机质不仅可以为作物生长提供营养物质,而且可以通过元素结合的方式降低土壤黏性,增加土壤微生物活性,使土壤变得松软,更适宜作物生长^[4-5]。除此之外,有机质中的微生物还可以降解土壤有害物质,保护土壤生态平衡^[6-8],因此准确测量土壤有机质含量,对于农业的精细化管理有重要意义。传统实验室测量土壤有机质含量的方法虽然精度高,但实验过程繁琐、周期长、成本高、存在安全隐患、并且会对环境造成污染,难以进行普及和推广,主要应用于科学研究^[9-11]。因此,寻找一种快速、精准、无损检测土壤有机质含量的方法具有重要意义。近红外光谱技术(NIR)具有快速、无损、无污染等特点^[12-13],目前被用于土壤有机质、土壤全氮、土壤水分等土壤养分方面的研究。目前,光谱技术在土壤中的应用已经成为研究热点,但 NIR 光谱法波段数目众多,光谱信息重叠严重^[14-15],如何筛选出有效的特征波长是提高模型预测精度和简化问题的关键^[16]。

特征波长的选取是光谱研究的关键步骤^[17]。ZHANG 等^[18]将未经自然风干和过筛的土壤,使用小波分析联合去包络线算法确定 6 个全氮特征波长,最终模型预测集决定系数 R^2 为 0.810,预测集均方根误差 (PMSEP) 为 0.053 mg/kg (实验室条件下),但模型受水分、粒度干扰严重,难以在大田下精准预测全氮含量;崔玉露等^[19]将土壤自然风干,过 0.9 mm 筛后,使用不同算法对土壤有机质特征波长进行提取,发现随机蛙跳和 CARS 筛选的波段建模精度更高 (实验室条件下),但在大田环境下,

模型预测效果一般;李冠稳等^[20]将土壤在室内自然风干、研磨、过 100 目筛后,利用 CARS 算法进行有机质特征波长筛选,最终选取 51 个有机质特征波长,模型预测集 R^2 达到 0.958,PMSEP 为 4.78 g/kg,但由于波长数目过多,光谱信息重叠严重,很难直接应用于农田环境下精准预测有机质含量;唐海涛等^[21]将不同类型的土壤经过风干,研磨,过 2 mm 筛后,利用 CARS 算法确定有机质特征波长,最终得出沼泽土的有机质 (SOM) 预测精度最高,但所建立的模型受水分、粒度影响严重,预测有机质效果差 (大田环境下);王炜超等^[22]将土壤自然风干,过 60 目筛后,利用 UVE-CARS 联合算法获取全氮特征波长,模型预测集 R^2 为 0.832 9,PMSEP 为 0.203 3 g/kg (实验室条件下),但在大田环境下模型预测精度下降。

根据上述研究可知,在寻找特征波长前,一般先经过土壤自然风干、过筛等预处理,最大程度上排除水分、粒度对其影响,实验室条件下用其特征波长建立的模型精度虽高,但模型很难适用于水分、粒度不统一的复杂大田环境。基于上述情况,本研究将土样进行不同水分、粒度梯度的预处理后进行有机质特征波长的筛选,通过找出不同水分、粒度梯度下相同或者接近的波长作为土壤有机质的特征波长,以减少水分、粒度对预测模型影响,使模型具有更好的适用性和准确性。

1 材料与方法

1.1 土壤样本采集

本实验的土壤样本取自北京市海淀区中国农业大学上庄实验田 (北纬 40.143 957 38° ~ 40.143 436 2°,东经 116.215 406 6° ~ 116.215 015 2°),为了使土样具有更好的代表性,对农田进行有机肥变量施肥,确保

实验数据梯度良好,经过 15 d 的自然沉淀,在不同施肥区进行土壤样本的采集。土壤样品的采集步骤为:①将约 1 500 m²农田按变量施肥区划分为 6 列,每列采取 10 个样品,共 60 个样本。②去除土壤表层的杂草与干土。③取距离土壤表层 5 ~ 20 cm 的土壤,每个样本采集 1.5 ~ 2.0 kg。为避免人为操作上的误差,将所采集的土样放入提前编好序号的牛皮袋,密封带回实验室。图 1 为土壤样本采集地点。

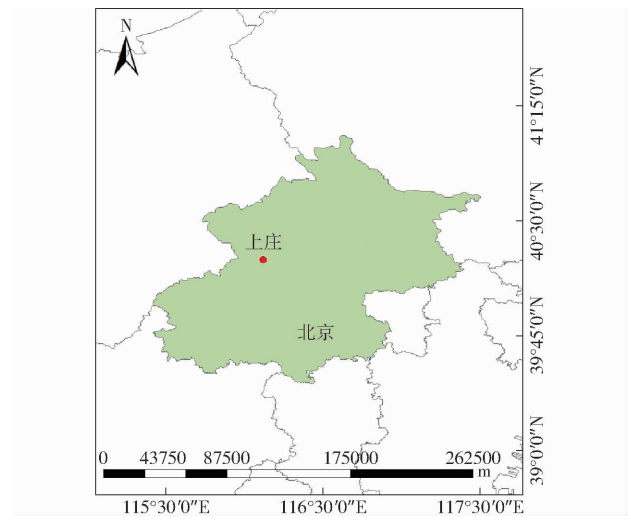


图 1 土壤样本采集地点
Fig.1 Soil sample collection sites

1.2 土壤样本制备

传统方式在寻找有机质特征波长时需经过自然风干、过筛等土壤样本预处理(最大程度排除水分、粒度影响),然后采集土壤样本的光谱数据,进而寻找有机质的特征波长,这种情况下特征波长所建立的模型,适合没有水分且颗粒较为统一的土壤样本,但农田的土壤中存在水分,且土壤颗粒大小不一,因此上述模型的适用性显著降低。基于上述原因,改变了土壤样本的预处理方式:将 60 个样本自然风干后,一分为二。通过人为胁迫的方式,将其中一组土壤样本过 5 种孔径筛子后配成 5 个不同的土壤粒度梯度(粒径 2 ~ 2.5 mm、1.43 ~ 2 mm、1 ~ 1.43 mm、0.6 ~ 1 mm、0 ~ 0.6 mm),60 个样本经人为胁迫后得到 300 个样本。另一组样本过 0.6 mm 筛后,加入不同体积的去离子水人工胁迫为 5 个水分梯度(含水率 5%、10%、15%、20%、25%),同样得到 300 个样本。

1.3 土壤样本光谱数据采集

所使用的光谱仪为美国海洋光学公司生产的微型近红外光谱仪(NIRQuest512 型),该光谱仪使用的检测器是 512 像素 InGaAs 线性检测器,光谱检测范围为 900 ~ 1 700 nm,光谱分辨率为 3.1 nm,入射狭缝为 25 μm,满足实验需求。光谱数据采集方法

为:将光谱仪按操作方法设置完成,预热 15 min 后开始数据采集。将积分球平整放到待测土样上。在软件页面保存所采集数据。图 2 为 NIRQuest512 型近红外光谱仪的实物图。



图 2 NIRQuest512 型近红外光谱仪
Fig.2 NIRQuest512 near infrared spectrometer

1.4 土壤有机质含量测定

土壤有机质含量的测定步骤为:将自然风干后的 60 个样本过 0.6 mm 筛。将称量后的土样放入坩埚中。将盛有土样的坩埚放入马弗炉。将马弗炉温度设置为 660℃,灼烧 4 ~ 6 h。通过计算灼烧前后的土壤样本质量差与灼烧前土样质量之比,得到土壤有机质含量真值。本次实验所测量的土壤有机质含量(质量比)介于 39.23 ~ 53.03 g/kg 之间,平均值为 46.35 g/kg,标准差为 3.53 g/kg,样本含量分布梯度明显,符合华北土壤有机质含量区间,可用于土壤有机质模型建立。

2 数据处理

2.1 光谱预处理

在进行光谱数据采集时,由于仪器本身和外界环境噪声等会对测量带来影响,为了消除外界因素对光谱数据的干扰,在 Matlab 软件中,使用 SG 卷积平滑算法进行滤波,使得光谱曲线变得更加光滑。以有机质含量 40.23 g/kg(不同水分、粒度梯度土样)为例,对比 SG 卷积前后的差异性,如图 3、4 所示,图 3a、4a 光谱曲线中存在锯齿形状,包含了一定的噪声信号,经过 SG 卷积后,光谱曲线变得光滑,消除了噪声的影响,如图 3b、4b 所示。

为进一步提高光谱数据和有机质含量真值的相关性,将 SG 平滑处理后的光谱数据进行多元散射校正(MSC)处理,多元散射校正可以有效解决光的散射带来的影响,提高光谱数据的可靠性^[23-25]。

2.2 特征波长选取

在土壤有机质特征波长提取中,随机蛙跳算法所确定的特征波长建立的模型精度较高^[19],因此本文采用随机蛙跳算法进行特征波长的提取。有机质特征波长的提取流程:①在每个水分梯度下提取 7 个特征波长(与有机质含量真值相关性最高的),5 个水分梯度下共提取 35 个特征波长,结合滤波片的带通范围(中心波长 ± 15 nm),对 35 个特征波长再次提取,挑选出相同或邻近的波长,最终确定 8 个特

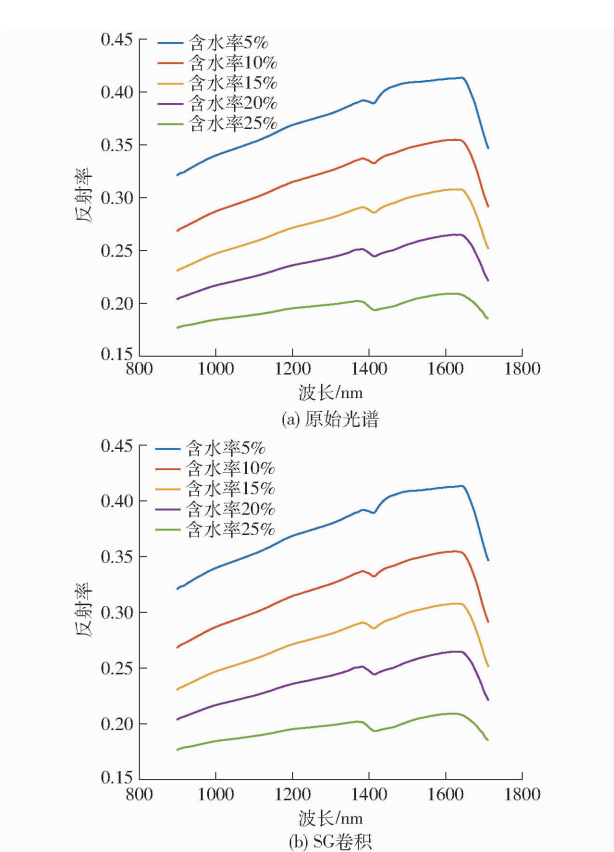


图3 不同含水率原始光谱与预处理后光谱
Fig.3 Original and pretreated spectra with different moisture contents

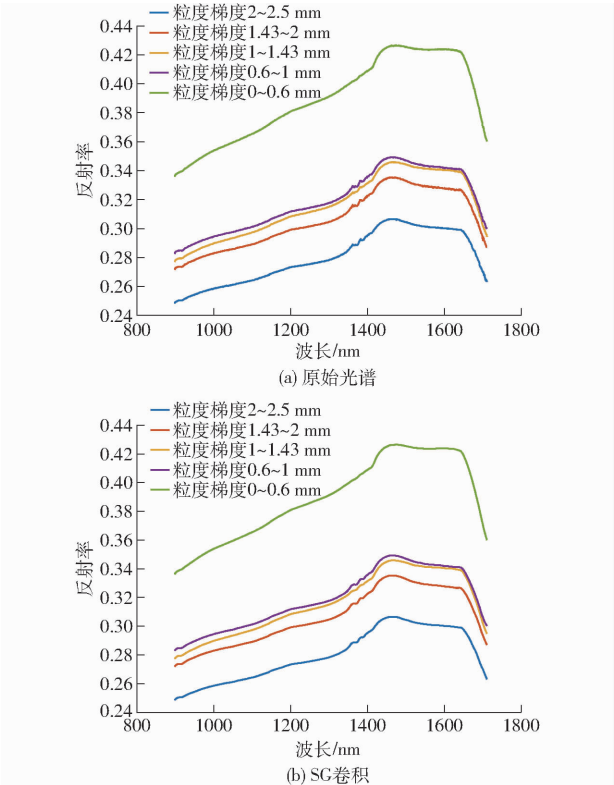


图4 不同粒度梯度下原始光谱与预处理后光谱
Fig.4 Original spectra and pretreated spectra under different particle size gradients

骤①,最终确定6个特征波长。③为排除全氮对预测有机质含量的影响,将水分梯度和粒度梯度下得到的14个特征波长,结合全氮主要官能团的特征吸收区间进行剔除,最终确定8个特征波长。

以粒度1.43~2 mm、含水率15%的土样为例,将SG+MSC预处理后的光谱数据和有机质含量真值作为参数输入随机蛙跳算法中,运行结果如图5所示,从图5a、5b分别提取7个与有机质含量真值相关性最高的波长作为对应梯度下所提取的特征波长,将其余水分、粒度梯度下的样本均按照上述过程寻找对应梯度下的有机质特征波长。表1为所有水分梯度下所特征波长,表2为所有粒度梯度下所找特征波长。

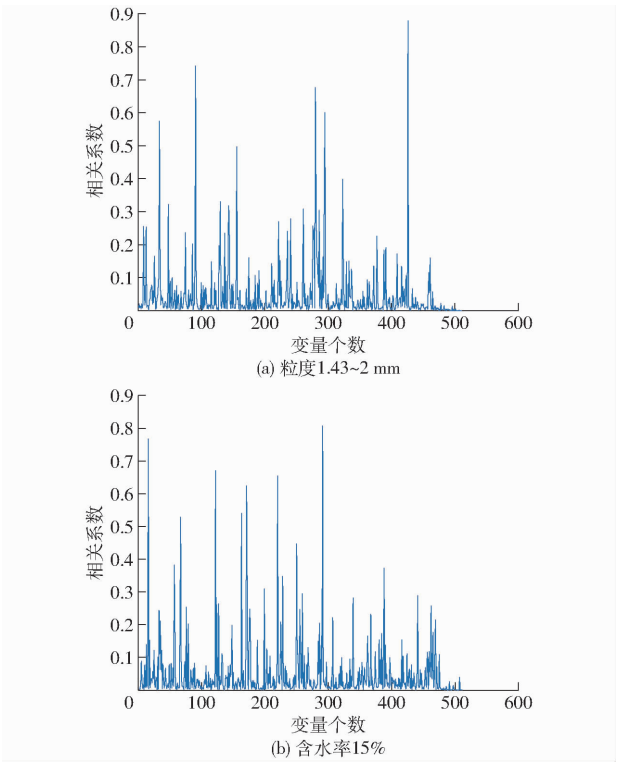


图5 不同梯度下随机蛙跳运行结果
Fig.5 Random frog-hopping running results under different gradients

表1 不同水分梯度下所确定有机质特征波长

Tab.1 Characteristic wavelengths of organic matter determined under different moisture gradients	
水分梯度/%	特征波长/nm
5	1 130、1 306、1 360、1 316、1 624、1 308、1 444
10	1 583、1 062、1 425、1 582、1 114、1 191、1 152
15	1 547、1 316、1 036、1 585、1 387、918、1 319
20	1 048、1 469、1 518、1 007、1 327、1 629、991
25	999、1 320、1 191、1 444、1 086、1 035、1 510

征波长。②粒度梯度下的特征波长的提取方法如步骤①,最终确定6个特征波长。③为排除全氮对预测有机质含量的影响,将水分梯度和粒度梯度下得到的14个特征波长,结合全氮主要官能团的特征吸收区间进行剔除,最终确定8个特征波长。

从表1、2可知,相同算法在不同水分、粒度下所提取的土壤样本的特征波长不同,其原因是不同水分、粒度下土壤样本反射率不同,导致所提取的波长

表 2 不同粒度梯度下所确定有机质特征波长

Tab.2 Characteristic wavelengths of organic matter determined at different particle size gradients	
粒度梯度/mm	特征波长/nm
2 ~ 2.5	1 106、1 077、1 088、1 056、1 046、1 015、1 143
1.43 ~ 2	1 355、1 611、1 502、932、975、1 377、1 233
1 ~ 1.43	1 014、1 191、1 357、1 075、1 093、1 358、1 057
0.6 ~ 1	926、1 025、1 280、1 041、911、932、1 344
0 ~ 0.6(传统方式)	1 027、955、1 160、1 505、1 416、1 268、1 316

不同,为减少不同样本反射率带来的误差,结合滤波片带通范围(± 15 nm),从不同水分梯度中找出它们相同或者接近的特征波长作为水分梯度下的有机质特征波长,提取的特征波长为:999、1 039、1 191、1 316、1 444、1 514、1 583、1 626 nm,共 8 个有机质特征波长。同理,从不同粒度梯度下提取相同或者接近的波长作为粒度梯度下的有机质特征波长,提取的特征波长为:932、1 020、1 050、1 083、1 356、1 503 nm,共 6 个有机质特征波长。

不同水分、粒度下所找有机质的特征波长共 14 个,因数目过多,不利于设备的开发,要进一步缩减特征波长数目。近红外光谱反演土壤有机质、土壤全氮等土壤养分的原理是:利用主要官能团在特定波段区间产生特定吸收峰,吸收峰的高低代表对应土壤养分含量。由于近红外光谱波段数目众多,导致光谱信息重叠严重,为准确预测有机质含量,从化学键的角度去除干扰。通过查阅文献可知,土壤全氮中主要含氮官能团(N—H、C—N、N—N)键在近红外光谱区的特征吸收波段区间为 1 020 ~ 1 060 nm、1 420 ~ 1 569 nm^[26-27],为了排除全氮对有机质预测的影响,结合全氮的主要官能团的特征吸收区间,将上述 14 个有机质特征波长进行剔除,最终确定 932、999、1 083、1 191、1 316、1 356、1 583、1 626 nm 作为最终有机质的特征波长。

3 结果与讨论

3.1 不同水分、粒度梯度下提取的特征波长对土壤有机质模型建立的影响

为了分析不同水分、粒度梯度下提取的特征波长对土壤有机质预测模型的影响,本节分别验证水分梯度和粒度梯度对土壤有机质预测模型的影响程度。

将每个水分梯度下所提取的 7 个有机质特征波长分别输入多元线性回归(MLR)、偏最小二乘(PLS)、随机森林(RF)算法中建立土壤有机质预测模型,结果如表 3 所示。

由表 3 可知,随着水分含量的增加,MLR、PLS、

表 3 不同水分梯度下土壤有机质模型预测结果

Tab.3 Prediction results of soil organic matter model under different moisture gradients					
水分梯度/%	建模算法	建模集		预测集	
		R^2	RMSEC/(g·kg ⁻¹)	R^2	RMSEP/(g·kg ⁻¹)
5	MLR	0.66	0.305	0.60	0.432
	PLS	0.70	0.501	0.65	0.574
	RF	0.60	0.641	0.54	0.571
10	MLR	0.60	0.606	0.55	0.721
	PLS	0.65	0.523	0.59	0.581
	RF	0.55	0.762	0.50	0.801
15	MLR	0.57	0.706	0.52	0.766
	PLS	0.60	0.562	0.55	0.612
	RF	0.49	0.886	0.43	0.932
20	MLR	0.36	0.856	0.30	0.879
	PLS	0.50	0.594	0.45	0.643
	RF	0.40	0.896	0.30	0.967
25	MLR	0.45	0.841	0.34	0.887
	PLS	0.40	0.601	0.35	0.667
	RF	0.42	0.954	0.32	0.976

RF 建模集、预测集 R^2 基本呈降低趋势,其原因可能是水分中含有和土壤有机质相同的官能团 O—H,水分含量越高,对预测有机质干扰越大。通过分析建模集、预测集 R^2 可知:PLS 算法建立的预测模型最优,其次是 MLR,RF 建模效果最差。

同理,分别将每个粒度下所提取的 7 个有机质特征波长分别输入到多元线性回归、偏最小二乘、随机森林算法中建立土壤有机质预测模型,结果如表 4 所示。

表 4 不同粒度梯度下土壤有机质模型预测结果

Tab.4 Prediction of soil organic matter model under different particle size gradients					
粒度梯度/mm	建模算法	建模集		预测集	
		R^2	RMSEC/(g·kg ⁻¹)	R^2	RMSEP/(g·kg ⁻¹)
2 ~ 2.5	MLR	0.39	0.703	0.35	0.803
	PLS	0.48	0.593	0.44	0.608
	RF	0.45	0.665	0.40	0.755
1.43 ~ 2	MLR	0.59	0.638	0.52	0.785
	PLS	0.63	0.557	0.54	0.554
	RF	0.62	0.590	0.53	0.652
1 ~ 1.43	MLR	0.56	0.590	0.52	0.766
	PLS	0.62	0.512	0.57	0.542
	RF	0.60	0.607	0.54	0.625
0.6 ~ 1	MLR	0.60	0.557	0.57	0.607
	PLS	0.64	0.485	0.58	0.501
	RF	0.63	0.580	0.56	0.616
0 ~ 0.6 (传统方式)	MLR	0.75	0.501	0.70	0.563
	PLS	0.82	0.432	0.75	0.453
	RF	0.78	0.540	0.72	0.601

由表 4 可知,在 0~0.6 mm 梯度下,3 种模型建模集 R^2 均在 0.75 以上,预测集 R^2 在 0.7 以上,在 2~2.5 mm 梯度下,3 种模型建模集 R^2 均在 0.5 以下,预测集均在 0.45 以下,其余梯度下,3 种模型建模集、预测集的 R^2 变化不显著,建模集 R^2 基本在 0.6 左右,预测集 R^2 基本在 0.55 左右,原因可能是当土壤颗粒特别细时(0~0.6 mm),对有机质预测影响微弱,所建立的模型精度高,当土壤颗粒特别粗时(2~2.5 mm),对有机质预测干扰大,所建立的模型精度低,剩下粒度梯度(0.6~1 mm、1~1.43 mm、1.43~2 mm)对有机质预测影响程度差别较小,所建立的模型精度接近。通过分析建模集、预测集 R^2 可知:PLS 算法所建立的预测效果最佳,其次是 RF,MLR 建模效果最差。

对比表 3、4 可知,表 3 是在不同水分梯度下建立的土壤有机质模型(过最细的筛,相当于粒度没有影响),且模型最高精度建模集 R^2 为 0.7,预测集 R^2 为 0.65,精度一般,因农田中含有颗粒不一的土壤,会导致模型精度再次下降。同理,表 4 是在不同粒度梯度下建立的有机质预测模型(含水率为 0,相当于水分没有影响),且模型最高精度建模集 R^2 为 0.82,预测集 R^2 为 0.75,因农田土壤中存在水分,会导致模型适用性降低。因农田土壤存在水分且颗粒不一,为消除两者影响,同时将水分、粒度作为研究对象。

3.2 水分、粒度梯度下综合提取的 8 个特征波长的预测模型

为验证本文选取特征波长的方式相比传统方式具有更好的适用性和准确性,将粒度 0~0.6 mm(含水率 0)的土壤样本下所挑选的 7 个有机质特征波长(传统方式)设置为对照组。为模拟大田土壤中水分和土壤颗粒不统一的情况(结合北京市中国农业大学上庄实验大田土壤含水率在 20% 以下),在上述 5 个水分、5 个粒度梯度中去除含水率 20%、25% 梯度,在其余梯度下分别抽取 10 个样本,共挑选 80 个样本,按照建模集、预测集比例为 5:3 进行划分,分别将最终确定的 8 个有机质特征波长和对照组所选择的 7 个波长输入多元线性回归、偏最小二乘、随机森林算法建立土壤有机质预测模型,最终结果如表 5 所示。

由表 5 可知,本文选取特征波长的方法相比于传统方式具有更好的预测性:MLR、PLS、RF 建模集、预测集 R^2 均高于传统方式下所建立的模型精度。本文选取特征波长的方法相比于传统方式具有更好的适用性:对比表 4 粒度(0~0.6 mm)所建立的模型精度可得,MLR 建模集、预测集 R^2 分别从 0.75、

表 5 不同特征波长选取方法下有机质模型预测结果
Tab.5 Prediction of organic matter model under different characteristic band selection methods

波段选取 方法	建模 算法	建模集		预测集	
		R^2	RMSEC/ ($\text{g}\cdot\text{kg}^{-1}$)	R^2	RMSEP/ ($\text{g}\cdot\text{kg}^{-1}$)
本文方式	MLR	0.84	0.321	0.77	0.364
	PLS	0.880 9	0.460	0.840 2	0.563
	RF	0.80	0.420	0.75	0.478
传统方式	MLR	0.68	0.621	0.60	0.674
	PLS	0.72	0.501	0.66	0.571
	RF	0.70	0.563	0.60	0.610

0.70 降低到 0.68、0.60,PLS 建模集、预测集 R^2 分别从 0.82、0.75 降低到 0.72、0.66,RF 建模集、预测集 R^2 分别从 0.78、0.72 降到 0.70、0.60,原因是在粒度 0~0.6 mm(含水率为 0)土样下寻找波长所建立的模型适合用来预测此梯度下土样土壤有机质含量,当土样存在水分且土样颗粒大小不一时,模型预测精度就会下降。本文选取波长方式所建立的 MLR、PLS、RF 中,PLS 模型精度最高,如图 6 所示(本文选取波段方式下)。

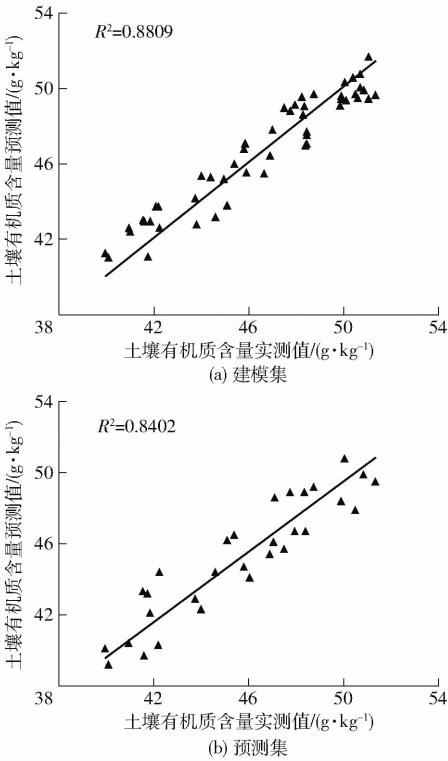


图 6 PLS 建模散点图
Fig.6 PLS modeling scatter diagrams

4 结论

(1)在不同水分、粒度梯度下寻找土壤有机质特征波长,相对于传统方式而言,此种方式所找的有机质特征波长具有更好的适用性和预测精度,一定

程度上消除了水分、粒度对有机质预测的影响,能够准确地在大田环境下预测土壤有机质含量。

(2)通过分析不同水分、粒度梯度下土壤有机质模型预测结果可知:含水率越高,对预测干扰越大;在 0 ~ 0.6 mm 梯度下对预测影响最小,在 2 ~ 2.5 mm 对预测影响最大,其余 3 个粒度梯度对预测影响接近。

(3)分析最终确定的 8 个特征波长所建立模型预测结果,相比于 MLR 和 RF,PLS 算法所建立的预测模型效果最佳,建模集、预测集 R^2 分别为 0.880 9、0.840 2。

(4)在传统方式下所建立的模型,用来预测农田环境下土壤有机质含量(含有水分且颗粒大小不一)时模型精度会降低。

参 考 文 献

- [1] 韩晓增,邹文秀.我国东北黑土地保护与肥力提升的成效与建议[J].中国科学院院刊,2018,33(2):206-212.
HAN Xiaozeng,ZOU Wenxiu. Effects and suggestions of black soil protection and soil fertility increase in Northeast China[J]. Bulletin of Chinese Academy of Sciences,2018,33(2):206-212. (in Chinese)
- [2] 张晓伟,许明祥,师晨迪,等.半干旱区县域农田土壤有机碳固存速率及其影响因素——以甘肃庄浪县为例[J].植物营养与肥料学报,2012,18(5):1089-1098.
ZHANG Xiaowei, XU Mingxiang, SHI Chendi, et al. Soil organic carbon sequestration rate and its influencing factors in farmlands of semi-arid regions—a case study in Zhuanglang County, Gansu Province[J]. Journal of Plant Nutrition and Fertilizers, 2012, 18(5): 1089-1098. (in Chinese)
- [3] WEI L, YUAN Z, WANG Z, et al. Hyperspectral inversion of soil organic matter content based on a combined spectral index model[J]. Sensors, 2020, 20(10): 2777-2787.
- [4] 刘文静,李广坡,倪进治,等.土壤有机质组分中多环芳烃分配特征和生态风险[J].中国环境科学,2020,40(4):1620-1627.
LIU Wenjing, LI Guangpo, NI Jinzhi, et al. Distribution patterns and ecological risks of polycyclic aromatic hydrocarbons in different soil organic matter fractions[J]. China Environmental Science, 2020, 40(4): 1620-1627. (in Chinese)
- [5] 樊晓彦.红外吸收法测定土壤有机质含量的试验[J].科技视界,2020,86(11):215-217.
FAN Xiaoyan. Determination of soil organic matter content by infrared absorption method[J]. Science & Technology Vision, 2020, 86(11): 215-217. (in Chinese)
- [6] 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.
CHEN Yiyun, QI Tianci, HUANG Yingjing, et al. Optimization method of calibration dataset for VIS-NIR spectral inversion model of soil organic matter content[J]. Transactions of the CSAE, 2017, 33(6): 107-114. (in Chinese)
- [7] GU X, WANG Y, SUN Q, et al. Hyperspectral inversion of soil organic matter content in cultivated land based on wavelet transform[J]. Computers and Electronics in Agriculture, 2019, 167: 105053.
- [8] LIU S, SHEN H, CHEN S, et al. Estimating forest soil organic carbon content using VIS-NIR spectroscopy: implications for large-scale soil carbon spectroscopic assessment[J]. Geoderma, 2019, 348(4): 37-44.
- [9] 王海峰,张智韬,ARNON Kamieli,等.基于灰度关联-岭回归的荒漠土壤有机质含量高光谱估算[J].农业工程学报,2018,34(14):124-131.
WANG Haifeng, ZHANG Zhitao, ARNON Kamieli, et al. Hyperspectral estimation of desert soil organic matter content based on gray correlation-ridge regression model[J]. Transactions of the CSAE, 2018, 34(14): 124-131. (in Chinese)
- [10] 金秀良,徐新刚,王纪华,等.基于灰度关联分析的冬小麦叶片含水量高光谱估测[J].光谱学与光谱分析,2012,32(11):3103-3106.
JIN Xiuliang, XU Xin'gang, WANG Jihua, et al. Hyperspectral estimation of leaf water content for winter wheat based on grey relational analysis[J]. Spectroscopy and Spectral Analysis, 2012, 32(11): 3103-3106. (in Chinese)
- [11] 于雷,洪永胜,周勇,等.连续小波变换高光谱数据的土壤有机质含量反演模型构建[J].光谱学与光谱分析,2016,36(5):1428-1433.
YU Lei, HONG Yongsheng, ZHOU Yong, et al. Inversion of soil organic matter content using hyperspectral data based on continuous wavelet transformation[J]. Spectroscopy and Spectral Analysis, 2016, 36(5): 1428-1433. (in Chinese)
- [12] 勾宇轩,赵云泽,李勇,等.基于CWT-sCARS的东北旱作农田土壤有机质高光谱反演[J].农业机械学报,2022,53(3):331-337.
GOU Yuxuan, ZHAO Yunze, LI Yong, et al. Soil organic matter content in dryland farmland in Northeast China with hyperspectral reflectance based on CWT-sCARS[J]. Transactions of the Chinese Society for Agricultural Machinery, 2022, 53(3): 331-337. (in Chinese)
- [13] III J R, MCCARTY G, MIMMO T. The potential of diffuse reflectance spectroscopy for the determination of carbon inventories in soils[J]. Environmental Pollution, 2002, 116(Supp. 1): 277-284.
- [14] WANG X, ZHANG F, DING J, et al. Estimation of soil salt content (SSC) in the Ebinur Lake Wetland National Nature Reserve (ELWNNR), Northwest China, based on a Bootstrap-BP neural network model and optimal spectral indices[J].

- Science of the Total Environment, 2018, 615: 918 – 930.
- [15] NOWKANDEH S M, NOROOZI A A, OMAEE M. Estimating soil organic matter content from Hyperion reflectance images using PLSR, PCR, MinR and SWR models in semi-arid regions of Iran[J]. Environmental Development, 2018, 25: 23 – 32.
- [16] XIAOMI W, YIYUN C, LONG G, et al. Construction of the calibration set through multivariate analysis in visible and near-infrared prediction model for estimating soil organic matter[J]. Remote Sensing, 2017, 9(3): 201.
- [17] 张娟娟, 田永超, 姚霞, 等. 基于近红外光谱的土壤全氮含量估算模型[J]. 农业工程学报, 2012, 28(12): 183 – 188. ZHANG Juanjuan, TIAN Yongchao, YAO Xia, et al. Estimating model of soil total nitrogen content based on near-infrared spectroscopy analysis[J]. Transactions of the CSAE, 2012, 28(12): 183 – 188. (in Chinese)
- [18] ZHANG Y, LI M Z, ZHENG L H, et al. Soil nitrogen content forecasting based on real-time NIR spectroscopy[J]. Computers and Electronics in Agriculture, 2016, 124: 29 – 36.
- [19] 崔玉露, 杨玮, 王炜超, 等. 基于光谱学原理的便携式土壤有机质检测仪设计与实验[J]. 农业机械学报, 2021, 52(增刊): 323 – 328. CUI Yulu, YANG Wei, WANG Weichao, et al. Design and experiment of portable soil organic matter detector based on spectroscopy principle[J]. Transactions of the Chinese Society for Agricultural Machinery, 2021, 52(Supp.): 323 – 328. (in Chinese)
- [20] 李冠稳, 高小红, 肖能文, 等. 基于 sCARS – RF 算法的高光谱估算土壤有机质含量[J]. 发光学报, 2019, 40(8): 1030 – 1039. LI Guanwen, GAO Xiaohong, XIAO Nengwen, et al. Soil organic matter contents with hyperspectra based on sCARS and RF algorithms[J]. Chinese Journal of Luminescence, 2019, 40(8): 1030 – 1039. (in Chinese)
- [21] 唐海涛, 孟祥添, 苏循新, 等. 基于 CARS 算法的不同类型土壤有机质高光谱预测[J]. 农业工程学报, 2021, 37(2): 105 – 113. TANG Haitao, MENG Xiangtian, SU Xunxin, et al. Hyperspectral prediction on soil organic matter of different types using CARS algorithm[J]. Transactions of the CSAE, 2021, 37(2): 105 – 113. (in Chinese)
- [22] 王炜超, 杨玮, 崔玉露, 等. 基于 CatBoost 算法与图谱特征融合的土壤全氮含量预测[J]. 农业机械学报, 2021, 52(增刊): 316 – 322. WANG Weichao, YANG Wei, CUI Yulu, et al. Prediction of soil total nitrogen based on CatBoost algorithm and fusion of image spectral features[J]. Transactions of the Chinese Society for Agricultural Machinery, 2021, 52(Supp.): 316 – 322. (in Chinese)
- [23] 陈华舟, 潘涛, 陈洁梅. 多元散射校正与 Savitzky – Golay 平滑模式的组合优选应用于土壤有机质的近红外光谱分析[J]. 计算机与应用化学, 2011, 28(5): 518 – 522. CHEN Huazhou, PAN Tao, CHEN Jiemei. Combination optimization of multiple scatter correction and Savitzky – Golay smoothing modes applied to the near infrared spectroscopy analysis of soil organic matter[J]. Computers and Applied Chemistry, 2011, 28(5): 518 – 522. (in Chinese)
- [24] 王动民, 纪俊敏, 高洪智. 多元散射校正预处理波段对近红外光谱定标模型的影响[J]. 光谱学与光谱分析, 2014, 34(9): 2387 – 2390. WANG Dongmin, JI Junmin, GAO Hongzhi. The effect of MSC spectral pretreatment regions on near infrared spectroscopy calibration results[J]. Spectroscopy and Spectral Analysis, 2014, 34(9): 2387 – 2390. (in Chinese)
- [25] 罗嗣卿, 刘昊天. 混合算法优化的 BP 神经网络的大气污染度预测[J]. 黑龙江大学自然科学学报, 2020, 37(4): 389 – 394. LUO Siqing, LIU Haotian. Prediction of atmospheric pollution based on BP neural network optimized by hybrid algorithms[J]. Journal of Natural Science of Heilongjiang University, 2020, 37(4): 389 – 394. (in Chinese)
- [26] 张娟娟, 田永超, 姚霞, 等. 基于近红外光谱的土壤全氮含量估算模型[J]. 农业工程学报, 2012, 28(12): 183 – 188. ZHANG Juanjuan, TIAN Yongchao, YAO Xia, et al. Estimating model of soil total nitrogen content based on near-infrared spectroscopy analysis[J]. Transactions of the CSAE, 2012, 28(12): 183 – 188. (in Chinese)
- [27] 吴裕, 申广荣, 刘璐, 等. 土壤有机质高光谱特征及其反演研究[J]. 上海交通大学学报(农业科学版), 2019, 37(4): 37 – 44. WU Yu, SHEN Guangrong, LIU Lu, et al. Hyperspectral characteristics of soil organic matter and inversion methods[J]. Journal of Shanghai Jiaotong University(Agricultural Science), 2019, 37(4): 37 – 44. (in Chinese)