

基于多尺度 SNV-CWT 特征的黑土有机质、水分、总铁及 pH 值估测

谭 洋, 姜琦刚*, 刘骅欣, 刘 斌, 高 鑫, 张 博

吉林大学地球探测科学与技术学院, 吉林 长春 130026

摘 要 田间土壤属性复杂且随时间变化, 快速精准地获得多种土壤理化指标数据对指导精细农业操作具有重要意义。为避免土壤水分带来的干扰, 基于光谱技术的土壤成分含量预测需在土壤样本干燥的情况下进行光谱测量, 然而土壤水分同样是指导农业生产的重要指标。为同时预测黑土区土壤有机质(SOM)、水分(SMC)、总铁(Fe)和 pH 值, 提出测量湿土土壤样本的可见-近红外光谱, 并采用标准正态变量变换(SNV)-连续小波变换(CWT)法分解光谱反射率, 逐样本进行 SNV 后, 以 Mexh 为小波基函数进行 10 个尺度($2^1, 2^2, \dots, 2^{10}$)的分解, 并与常用光谱处理方法进行对比, 包括高斯滤波(GS)、一阶导数(FD)、连续统去除(CR), 数学变换等 7 种方法。将 74 个样本数据划分为两组, 其中 50 个作为建模集, 24 个作为验证集。经 SNV-CWT 变换后, 每个尺度的小波系数与每个目标变量间置信度小于 0.05 的波段作为随机森林(RF)预测模型的输入变量, 以各尺度验证模型精度为标准确定每个预测目标的最佳分解尺度; 通过计算最佳尺度小波系数与土壤成分间的皮尔森相关系数(PCC), 基于模型的相关系数(MBC)和灰色关联度(GRD), 判断各属性的特征波段, 且分别以三种相关系数作为指标, 以过滤式筛选法建立不同属性的 RF 估测模型。结果表明: 与 7 种常用的处理方法相比, SNV-CWT 分解后四种土壤成分的预测精度均有提高, SOM, SMC, Fe 和 pH 对应的最佳分解尺度分别为 7, 8, 1 和 10。在以多维特征作为输入变量的情况下, SOM 与 SMC 的验证模型决定系数(R^2)即可达到 0.90 和 0.93。三种分析方法中以 MBC 计算的相关系数为波段筛选指标建立的模型精度最佳, 其中 SOM 与 SMC 的 R^2 均为 0.94, 且 $Fe(R^2=0.67, Mse=0.01\%, RPD=1.76)$ 与 $pH(R^2=0.80, Mse=0.1, RPD=2.24)$ 的模型精度具有大幅度提高, 可应用于多种土壤理化指标数据的提取与监测。

关键词 黑土区; 近红外光谱; 数据处理; 有机质; 水分; 总铁; pH

中图分类号: TP79 **文献标识码:** A **DOI:** 10.3964/j.issn.1000-0593(2021)11-3424-07

引 言

土壤是由矿物质、有机质(soil organic matter, SOM)、水分(soil moisture content, SMC)和空气等多种成分构成的复杂系统。土壤成分检测和分析可用于提高土壤条件评估, 指导精细化生产的农业活动。其中, SOM 和 SMC 是土壤及农作物生长的重要因素, 关系土壤肥力, 影响农作物长势; 总铁(Fe)是土壤矿物中的主要元素之一, 包括 Fe^{2+} , Fe^{3+} , 复合其他化学成分的铁, 如 EDTA, 以及各种铁的氧化物, 是植物生命活动必须的微量元素; 土壤 pH 值描述了碱饱和和水平, 与微量元素的有效性、以及与潜在有毒元素, 如铝, 有密切关系。传统基于样本的化学分析方法成本高、检测周

期长, 反射率光谱法已成为快速、可靠地获取土壤组成信息的一种替代方法。基于不同波长反射率与土壤成分的响应关系, 可见光-近红外反射光谱已被广泛应用于预测不同土地利用类型土壤的 SOM、SMC^[1-2]等。

在光谱数据处理中, 选择一种适合的信号处理方法至关重要, 有效的光谱处理方法可以极大地提高估测模型的精度。对于不同土壤成分的提取, 通常可采用不同的数据处理方法获得理想的精度^[3-4], 如利用多元散射校正(multiple scatter correction, MSC)方法提高总铁、游离铁和无定形铁的估算精度, 平滑滤波(savitzky golay, SG)结合 MSC 提高土壤 pH 和其他土壤成分的估算精度^[5], 分数阶导数处理提高盐渍化土壤 SMC 估测精度^[6], 利用离散小波变换提高土壤重金属元素的估测精度^[7]。大部分研究为提高土壤成分的

收稿日期: 2021-01-25, 修订日期: 2021-04-09

基金项目: 中国地质调查局二级项目: 全国地球关键带遥感地质调查项目(DD20190536), 国家自然科学基金项目(41804098)资助

作者简介: 谭 洋, 1990 年生, 吉林大学地球探测科学与技术学院博士研究生 e-mail: tanyang331@163.com

* 通讯作者 e-mail: jiang_qigang@163.com

预测精度选择烘干样本测量反射率,或利用外部参数正交化法^[8]等去除水分影响。且常规土壤光谱数据处理方法大多着重于突出单一估测目标的变化特征,削弱其他目标的影响^[9]。为同时提高多种土壤成分的预测精度,需要减弱颗粒大小不均和粒子表面带来的散射影响对光谱反射率测量带来的误差,并增强不同成分响应特征的细节信息。

标准正态变量变换(standard normal variable, SNV)针对每一样本的光谱数据,降低样本表面的非特异性散射噪声。多尺度连续小波变换(continuous wavelet transform, CWT)在不同程度上放大光谱数据的细节特征^[10],可有效探测化学元素,树冠水分^[11],叶片的氯离子浓度^[12]等。本文提出使用 SNV-CWT 多尺度分解法处理湿土样本高光谱反射率数据,以随机森林(RF)预测模型的精度为标准,对比 7 种常用处理方法,确定不同成分的最佳分解尺度。利用三种方法计算和分析光谱分解前后与各成分相关关系,并实现各成分特征波段的提取与建模分析。

1 实验部分

1.1 研究区概况

东北黑土区主要分布在吉林省和黑龙江省,少量分布在辽宁省北部和内蒙古东部,如图 1 所示。根据以往研究,黑土区土壤肥沃,SOM 含量丰富,且自南向北递增,极少部分 SOM 含量少于 2%。采样点位于松辽平原北部的黑土带内,主要分布在黑龙江省中部,海伦市(44°04′—44°45′N, 125°26′—126°31′E)和吉林省中部,德惠市(47°00′—47°41′N, 126°25′—127°06′E)。

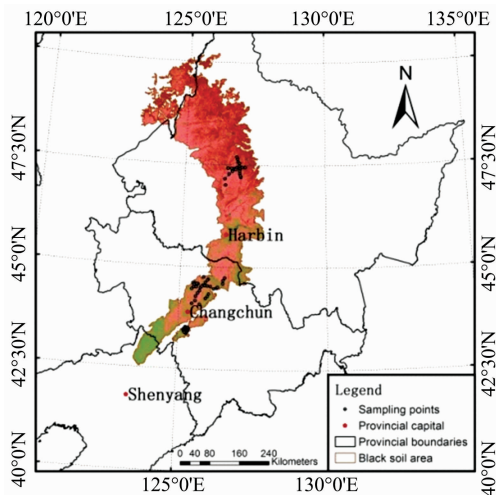


图 1 研究区及采样点分布

Fig. 1 Study area and sampling regions

1.2 土壤样本采集、成分测定及光谱测量

土壤样本采集于 2017 年 5 月 4 日至 9 日,正值春耕时期,土壤表面冰雪完全消融,且无植被覆盖。共有 74 个采样点,每个采样点在 5 m×5 m 范围采集 5 份土壤,采集深度为 0~15 cm,所有样本均被独立装入自封袋并标记样本号,采样点中心地理坐标由便携式 GPS 记录。

在实验室条件下将每个采样点的 5 份土壤混合均匀,过 2 mm 筛剔除碎石子残留秸秆等杂质。每份样本共取出 100 g 土壤分别用于测定各土壤成分含量,其中,利用重力法测定 SMC 含量,重铬酸钾加热法测定 SOM 含量,点位法测定土壤 pH 值,等离子光谱仪测定 Fe 含量。土壤样本成分含量统计如表 1 所示。其中 SOM 含量介于 2.0%~6.3%,平均值约为 3.5%,整体含量较高;土壤样本 SMC 近似于野外土壤水分含量,值域分布区间较大,介于 3.8%~24.4%;Fe 含量介于 2.3%~3.4%之间,且标准偏差(SD)与变异系数(CV)均较小(SD=2.12, CV=7.81),总体含量差异不明显;pH 值介于 4.7~8.1 之间,均值为 5.7,标准偏差较小(SD=0.77),总体偏离平均值较小。

表 1 研究区土壤样本成分值统计

Table 1 Statistical values of soil component from study area

Contents	Max	Min	Mean	SD	CV/%
SOM/(g·kg ⁻¹)	63.36	19.87	34.86	12.65	36.29
SMC/(g·kg ⁻¹)	244.61	38.02	105.25	46.82	44.54
Fe/(g·kg ⁻¹)	33.63	23.02	27.16	2.12	7.81
pH	8.09	4.73	5.74	0.77	13.41

室内光谱测量在黑暗的实验室进行,将过筛土壤样本置于深度 1 cm 的铝制测试盒,刮平表面,用 90 W 的卤素灯照向表面。采用 ASD FieldSpec3 光谱仪测定光谱反射率,每个样本在不同位置共计测量 20 条光谱曲线。利用 ViewSpecPro 软件计算平均光谱作为该样本的原始反射光谱曲线。

1.3 光谱数据处理

光谱数据首先以原始光谱反射率为输入矩阵,进行标准正态变量变换。逐一将光谱反射率数据 $X = (X_1, X_2, \dots, X_p)$ 在 p 波长处转换为 $Z = (Z_1, Z_2, \dots, Z_p)$, 其中

$$Z_i = (X_i - m) / s \quad (1)$$

式(1)中, m 和 s 分别表示 X_1, X_2, \dots, X_p 的均值和标准差^[13]。以变换后的数据 $Z_{n \times p}$ (n 为样本数)作为输入矩阵进行多尺度连续小波变换。光谱反射率的连续小波变换可以表示为小波基函数 $\Psi(\lambda)$ 经过比例缩放和平移所产生的一系列系数,用式(2)表示为^[12]

$$W_f(a, b) = \langle f, \Psi_{a,b} \rangle = \int_{-\infty}^{+\infty} f(\lambda) \frac{1}{\sqrt{a}} \Psi\left(\frac{\lambda-b}{a}\right) d\lambda \quad (2)$$

式(2)中, $W_f(a, b)$ 为小波系数, a 和 b 均为实数, a 为比例因子(小波基的宽度), b 为平移因子(小波基的位置)。小波系数 $W_f(a, b)$ 为小波基与光谱反射率在不同尺度下的关系。

参与对比的数据处理方法包括一阶导数(first derivative, FD), 二阶导数(second derivative, SD), 连续同去除(continuum removal, CR), 标准正态变量变换(SNV), 多元散射校正(MSC), 高斯滤波法(Gaussian filter, GS), 以及对数变换[logarithm transform, $\log(1/R)$]。其中 FD 和 SD 变换, 首先对离散光谱反射率进行具有 11 个平滑点的二阶多项式拟合, CR 通过调用 Tripack R 包中的 Delaunay 三角测量的凸包实现光谱拟合; 其余处理方法均利用 Python 实现, 且均以湿土样本的原始光谱反射率为输入数据。为减少反

射率边缘的噪声,原始光谱反射率仅保留 400~2 500 nm 范围的值。为了可视化不同处理方法的差异,所有的信号变换

都绘制在图 2 中,其中 SNV-CWT 处理方法绘制的是 10 尺度(2^{10})的分解系数。

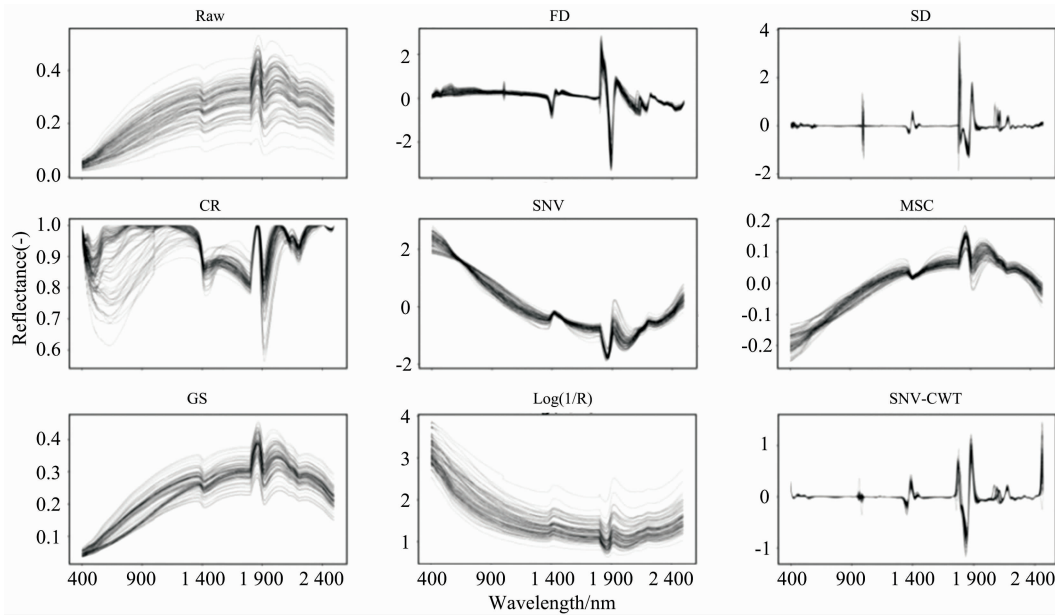


图 2 经过不同处理方法的光谱曲线

Fig. 2 Reflectance spectra under different transformation methods

1.4 特征分析、建模与精度验证

皮尔森相关系数(Pearson correlation coefficients, PCC)常用来衡量特征与目标变量间的线性相关程度,定义为两个变量之间的协方差与标准差的商。基于模型的系数(model based coefficients, MBC)是一种基于模型的相关关系指标,与传统的线性方法相比,可以在建模之前判断特征变量与目标变量间基于模型的非线性关系。灰色关联分析(grey relation analysis, GRA)是一种多因素统计分析方法,通过计算参考数列和比较数列变化曲线几何形状相似程度判断因素间的关联程度,由此方法计算得出的灰色关联度(GRD)可作为判定特征数据与相关因素序列的相关性指标。研究中采用三种方法对 SNV-CWT 处理前后的光谱数据进行特征分析,判断不同土壤属性的响应波段范围。

随机森林(random forest, RF)回归是一种基于决策树的机器学习算法,采用 bootstrap 抽样技术抽取训练集,并建立每个训练集的 CART 树,构成森林,具有强大的预测能力,已被广泛应用于不同科学领域。该方法学习速度快,可以处理大量的输入变量,即使在特征变量数远大于样本数据时也可以很好地避免过拟合,且模型抗粗差能力较强。

采用决定系数(coefficient of determination, R^2)、均方差(mean squared error, Mse)和相对分析误差,即标准差与预测均方根误差比(residual predictive deviation, RPD)这 3 个指标检测与评价模型的稳定性和预测能力。其中 R^2 越接近 1 说明模型拟合程度越好;当 $RPD < 1.4$ 时,模型质量较差,无法对样本进行预测;当 $1.4 < RPD < 2$ 时,模型质量较好,可以用来进行目标变量的粗略估算;当 $2 < RPD < 2.5$ 时,模型质量很好,可以用于目标变量的定量预测; $RPD > 2.5$ 时,模型具有极好的预测能力^[14]。检验与校正模型的统计指标

(statistical indicators, SI)计算公式如式(3)~式(5)

$$Mse = \frac{\sum_{i=1}^n [F(i) - A(i)]^2}{n} \quad (3)$$

$$R^2 = 1 - \frac{\sum_{i=1}^N (F_i - A_i)^2}{\sum_{i=1}^N (F_i - \bar{A})^2} \quad (4)$$

$$RPD = \frac{SD}{\sqrt{Mse}}$$

$$\text{其中, } SD = \sqrt{\frac{\sum_{i=1}^n (F(i) - \bar{A})^2}{n-1}} \quad (5)$$

其中, SD 为标准差, $F(i)$ 为预测值, $A(i)$ 为测量值, \bar{A} 为测量值的均值, n 为样本数。

2 结果与讨论

2.1 原始数据分析

原始光谱反射率曲线在 1 400, 1 900 及 2 200 nm 附近均具有明显的拐点。湿土壤条件下的光谱反射率较干土相比,在 1 400 和 1 900 nm 附近具有较大程度的波动。如图 3 所示,两条曲线的有机质含量相近,分别为 6.0% 和 6.1%,但含水量差异较大,分别为 15.8% 和 5.4%。干土情况下两样本的光谱反射率曲线形状相似,含水量不同时,湿土样本两条光谱反射率形状具有较大差异,说明水分对反射率具有显著影响。而 Fe、pH 与反射率的最大 PCC 分别为 0.18 与 0.25,与 SOM 的最大值(0.63)相比对反射率的影响较低。

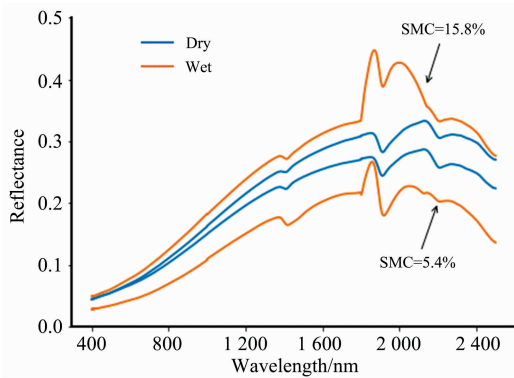


图 3 干湿土壤样本原始光谱反射率示意图

Fig. 3 Raw reflectance spectra of dry and wet samples

2.2 数据处理方法的有效性分析

以建模精度为标准验证数据处理方法的有效性，将置信度小于 0.05 的多维变量作为 RF 模型输入特征，8 种处理方法的预测精度如表 2 所示。除 $\log(1/R)$ 变换外，SOM 的建

模精度均有一定的提高，且经过 SNV-CWT 变换的模型精度可以达到 $R^2=0.90$ ， $Mse=0.15\%$ ， $PRD=3.21$ 。SMC 在原始反射率情况下，没有符合显著性水平的波段，经过不同方法处理后仅 GS 与 $\log(1/R)$ 方法下的建模精度 $RPD < 2$ ，其余 6 种处理变换均有一定的提高，且经过 SNV-CWT 变换精度为 $R^2 = 0.93$ ， $Mse = 1.37\%$ ， $PRD = 4.71$ 。常用方法中 SD，CR，SNV，MSC 和 GS 可以粗略建立 Fe 的预测模型，然而建模结果精度均不理想，经过 SNV-CWT 处理后建模精度可以得到最大的提升，但 RPD 仍小于 1.5。8 种处理方法较原始反射率相比，pH 建模精度均有不同程度的提高，且经过 CR，SNV 和 SNV-CWT 处理后，模型精度 RPD 均可大于 1.5。由表 2 可以看出，不同数据处理方法对每种成分的影响效果不一，采用 SNV-CWT 处理，可同时获得四种土壤成分大幅度提升的建模精度。与 Stenberg 等研究结果相比，SOM 和 SMC 建模精度一般在 $R^2_{cv} = 0.50$ 到 0.90 之间^[4]，经过 SNV-CWT 处理后，在以多维特征维输入变量的情况下 SOM 与 SMC 的建模精度已达到可获得的较好结果 ($R^2 = 0.90$ 和 0.93)。

表 2 不同数据处理方法下的 SOM, SMC, Fe 和 pH 验证模型精度

Table 2 Statistical indicators of validation models of SOM, SMC, Fe and pH under different transformation methods

Contents	SI	Raw	FD	SD	CR	SNV	MSC	GS	Log(1/R)	SNV-CWT
SOM	R^2	0.43	0.83	0.70	0.73	0.84	0.68	0.74	0.43	0.90
	Mse	0.83	0.25	0.44	0.40	0.23	0.47	0.38	0.83	0.15
	RPD	1.46	2.51	1.85	2.01	2.71	1.76	1.97	1.37	3.21
SMC	R^2	—	0.83	0.87	0.77	0.79	0.74	0.36	0.02	0.93
	Mse	—	3.09	2.32	4.28	3.83	4.831	1.72	18.68	1.37
	RPD	—	2.79	3.49	2.45	2.34	2.34	1.26	1.13	4.71
Fe	R^2	—	—	0.20	0.17	0.14	0.09	0.27	—	0.48
	Mse	—	—	0.03	0.03	0.03	0.03	0.03	—	0.02
	RPD	—	—	1.12	1.10	1.11	1.05	1.18	—	1.39
pH	R^2	—	0.44	0.36	0.64	0.61	0.28	0.39	0.13	0.62
	Mse	—	0.27	0.32	0.18	0.19	0.36	0.30	0.88	0.19
	RPD	—	1.38	1.27	1.67	1.60	1.18	1.28	0.95	1.63

2.3 多尺度 SNV-CWT 的有效性及各成分最佳分解尺度

研究中以 Mexh 为小波基函数，进行 10 个尺度 ($2^1, 2^2, 2^3, \dots, 2^{10}$) 的 SNV-CWT 分解，将一维光谱反射率分解为 10 个尺度的小波系数。以原始光谱数据集 74×2101 维矩阵为输入，经 SNV-CWT 处理后转换为 $74 \times 10 \times 2101$ 维矩阵。针对四种土壤成分，选择每个尺度置信度小于 0.05 的小波系数作为输入变量，建立 RF 回归模型，以验证模型获得的决策系数 R^2 为标准，确定各土壤成分最佳分解尺度。如图 4 所示，SMC 的模型精度在不同尺度上整体优于 SOM，且 SOM 和 SMC 的建模精度较 Fe 和 pH 更好。这是由于 SMC 和 SOM 对光谱反射率影响起主导作用，对 Fe 与 pH 验证模型信息具有掩盖作用，且 Fe 的含量变化区间小。SOM 与 SMC 的最佳分解尺度分别为 7 和 8；Fe 与 pH 验证模型的最佳分解尺度为 1 和 10。每种土壤成分对应于不同的分解尺度有明显的增强信息，相比于单一尺度分解，多尺度 SNV-

CWT 分解对多属性探测具有更大优势。

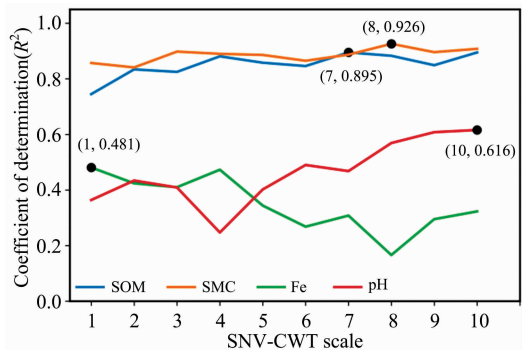


图 4 SNV-CWT 多尺度情况下的不同目标变量的模型决策系数

Fig. 4 Coefficients of determination (R^2) corresponding to different soil properties on multi SNV-CWT scales

2.4 特征分析与建模

分别计算 SOM, SMC, Fe 和 pH 与原始反射率以及在最佳分解尺度系数的 PCC, MBC 和 GRD, 其中 PCC 为线性相关系数的平方, MBC 为单变量 RF 模型决策系数, GRD 计算时取分辨系数 $\rho=0.5$ 。从图 5 中可以看出, 经过 SNV-CWT 变换后, 三种方法计算得出的相关系数均较原始反射率有所提高, 且 SOM 与 SMC 相关系数均高于 Fe 与 pH 模型。这也充分解释了多维特征建模时 Fe 与 pH 的模型精度在不同尺度上均不及 SOM 与 SMC。与土壤成分对应的相关波段通过分析可分为两类: 一类是三种方法计算的相关系数曲线在相同的波长范围均有较明显的峰值, 这种波段可以认为是该成分的主要特征波段; 另一类是一种或两种分析方法计算的相关系数较高而对应的其他方法相关系数较低, 这种波段属于该成分的特征波段。SOM 在 450, 580 和 820 nm 附近的波段属于主要特征波段范围; 超过 1 400 nm, 例如 1 495, 2 210 以及 2 420 nm 左右, 属于 SOM 的特征波段。而 SMC 的特征波段主要分布在大于 1 400 nm, 从 PCC 和 MBC 曲线可以看出波段均表现出较高的相关性。1 460 和 1 950 nm 附近是明显的主要特征波段, 与前人总结的 SMC 特征波

段一致。Fe 在整个可见光近红外光谱范围区间, 并没有很明显的主要特征峰, 但较原始反射率相比, 三种分析方法计算得出的所有波段相关度有所提高, PCC 的最大值从 0.18 提升至 0.47。从 GRD 曲线看, 特征间的相关系数大小差异不明显; 从 PCC 与 MBC 曲线判断, 730, 1 360, 1 750 以及 2 110 nm 附近属于 Fe 的特征波段。pH 的特征波段分散在整个范围内, 从 PCC 与 GRD 曲线来看, 特征波段集中在 460, 920, 2 240 和 2 430 nm 附近; 从 MBC 来看, 还集中在 1 100 和 1 740 nm 附近。

总体分析, 对反射率具有主导影响的属性, 如 SOM 与 SMC, 不同分析方法均可筛选出其对应的主要特征波段。而影响较弱或含量较低的属性, 如 pH 与 Fe, 主要特征波段不明显, 分析方法不同时所对应的特征波段会有所差别。从 Fe 和 pH 的相关关系图中可以看出, 对于主要特征波段不明显的土壤成分, 不同方法计算的波段相关系数排序不同, 由此所确定的特征波段不同。因此使用三种分析方法计算波段与土壤成分的相关系数为指标, 采用过滤式方法筛选波段作为模型输入变量并进行模型精度分析。

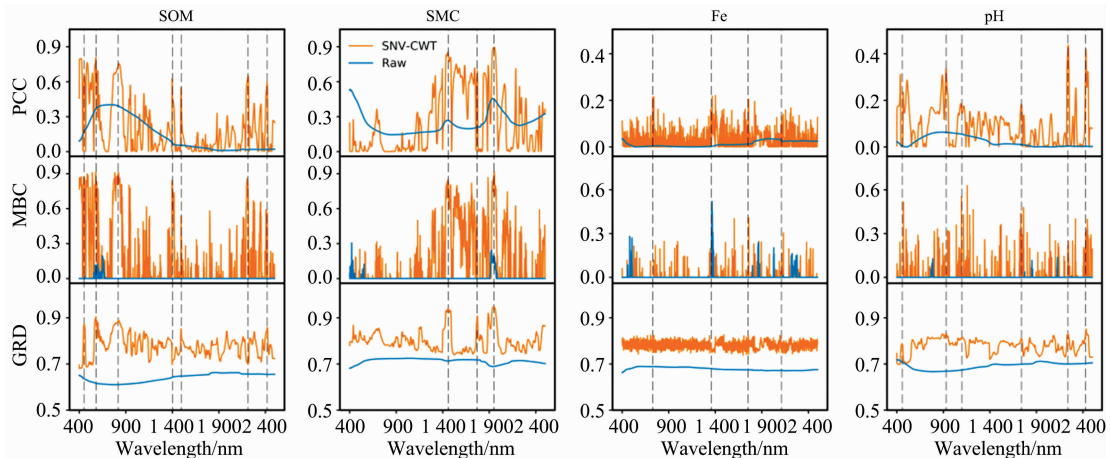


图 5 SNV-CWT 处理前后的反射率与不同成分的关系

Fig. 5 Correlations between different soil properties and reflectance before and after SNV-CWT

表 3 和表 4 分别记录了以 PCC 和 GRD 为过滤式筛选标准的具体建模指标。从表中可以看出, PCC 和 GRD 对四种成分的计算精度相似; 除 SOM 外, 其他三种成分所需的波段数目均较多, 且与多维特征建模(表 2)相比精度近似。表中 N (Numbers)表示样本数量, T (Threshold)表示相关系数阈值。

表 3 线性相关法筛选特征及建模精度

Table 3 Statistical indicators of models taking bands selected by PCC as input variables

Contents	N	T	Max	Min	R^2	Mse	RPD
SOM	3	0.80	0.81	0	0.91	0.12	3.49
SMC	99	0.75	0.90	0	0.91	1.66	4.22
Fe	75	0.11	0.22	0	0.43	0.02	1.34
pH	71	0.29	0.44	0	0.65	0.17	1.73

表 4 灰色关联分析法筛选特征及建模精度

Table 4 Statistical indicators of models taking bands selected by GRA as input variables

Contents	N	T	Max	Min	R^2	Mse	RPD
SOM	2	0.90	0.90	0.68	0.91	0.13	3.32
SMC	89	0.92	0.95	0.74	0.90	1.76	4.02
Fe	80	0.81	0.83	0.74	0.45	0.02	1.38
pH	98	0.82	0.85	0.70	0.65	0.17	1.75

图 6 为根据 MBC 值逐步添加特征得到的 RF 验证模型 R^2 曲线。与基于 PCC 和 GRD 指标分析的特征筛选方法相比, 此方法计算得出的相关关系更适于提高预测模型的精度。从子图中曲线的变化趋势可以看出, 对于 SOM 和 SMC, 波段数量的增加对模型精度的影响并不大, 基于第 6 尺度分

解系数的 SOM 与第 7 尺度分解系数的 SMC 验证模型 R^2 基本位于 0.9~0.95 之间,且可以在较少波段的情况

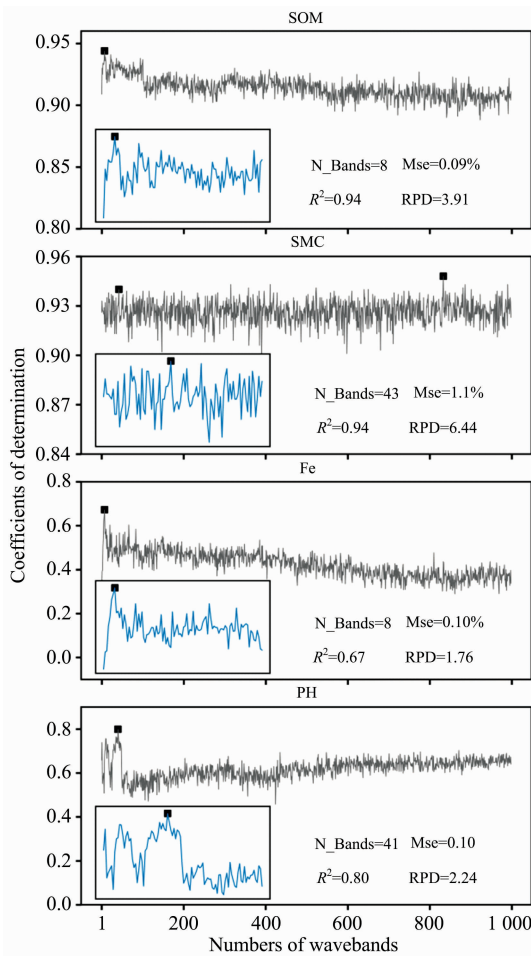


图 6 基于模型的相关关系筛选特征及建模精度
Fig. 6 Statistical indicators of models taking bands selected by MBC as input variables

下获得较优的预测模型。

Fe 的模型精度随着波段数目的增加明显降低。在可见光近红外波长范围内,由于 SOM 含量较高,在 1 400 nm 前对反射率具有主导影响,且 SMC 对光谱的吸收作用,干扰反射率曲线,使 1 400 nm 后的反射率曲线扭曲,致使 Fe 的特征波段信息被掩盖。同时由于 Fe 包含的成分复杂,易受到外界环境及有机质的影响^[5],较难估算。相关研究经过 100 目研磨及 MSC 处理后的验证模型精度 ($R^2 = 0.6$, $RPD = 0.86$) 相比,图 6 中 Fe 的建模当筛选波段数目为 8 时,最佳模型 R^2 为 0.67,此时 RPD 为 1.76 ($1.4 < 1.76 < 2$),模型较好可用于粗略估算。pH 的预测模型在入选波段数目为 41 时达到最佳,此时验证模型 $R^2 = 0.80$, $RPD = 2.24$ 。

3 结 论

多尺度 SNV-CWT 分解是一种能够有效提高多种土壤理化指标预测精度的方法。常规数据处理方法仅能在一定程度上提高某种成分的预测精度,而以 Mexh 为小波基函数的多尺度 SNV-CWT 分解方法可全面提高各种成分的信息提取质量。与 GS, FD 和 CR 等 7 种常用的处理方法相比,四种成分预测模型精度在 SNV-CWT 分解下均有提高且优于其他方法。通过对比处理前后反射率与土壤成分的 PCC, MBC 以及 GRD,可以看出多尺度分解系数能够有效地在不同尺度上放大成分的细节信息,提高土壤成分与反射率的相关系数。以 MBC 为指标进行的过滤式特征筛选为输入变量的预测模型具有最佳的精度,其中 SOM 和 SMC 的 $R^2 = 0.94$, Mse 分别为 0.09% 和 1.1%,可以获得极好的预测精度。Fe 的 $R^2 = 0.67$, $Mse = 0.01\%$, $RPD = 1.76$; pH 的 $R^2 = 0.80$, $Mse = 0.1$, $RPD = 2.24$,可以达到用于定量预测的程度。土壤属性复杂,在湿土壤情况下同时获得多种土壤成分的含量,可以大大提高土壤质量评估效率,快速进行精准农业指导。

References

- [1] Ben-Dor E, Patkin K, Banin A, et al. International Journal of Remote Sensing, 2002, 23(6): 1043.
- [2] Morellos A, Pantazi X E, Moshou D, et al. Biosystems Engineering, 2016, 152: 104.
- [3] Gholizadeh A, Luboš B, Mohammadmehdi S, et al. Applied Spectroscopy, 2013, 67(12): 1349.
- [4] Stenberg B, Raphael A, Viscarra R, et al. In Advances in Agronomy, 2010, 107: 163.
- [5] Wang Y, Huang T, Liu J, et al. Computers and Electronics in Agriculture, 2015, 111: 69.
- [6] Kahaer Y, Tashpolat N, Shi Q, et al. Water, 2020, 12(12): 3360.
- [7] Wang T, Jiao L, Yan C, et al. Chemometrics and Intelligent Laboratory Systems, 2019, 194: 103854.
- [8] De Santana, De Giuseppe, de Souza, et al. Microchemical Journal, 2018, 145: 1094.
- [9] Vašát, Radim, Radka Kodešová, et al. Geoderma 2017, 298: 46.
- [10] Rivard B, Feng J, Gallie A, et al. Remote Sensing of Environment, 2008, 112(6): 2850.
- [11] Cheng T, Rivard B, Sánchez G A, et al. Remote Sensing of Environment, 2010, 114(4): 899.
- [12] Blackburn G A, Ferwerda J G. Remote Sensing of Environment, 2008, 112(4): 1614.
- [13] Tom Fearn, Cecilia Riccioli, Ana Garrido-Varo, et al. Chemometrics and Intelligent Laboratory Systems, 2009, 96(1): 22.
- [14] CHEN Yi-yun, QI Tian-ci, HUANG Ying-jing, et al (陈奕云, 齐天赐, 黄颖菁, 等). Transactions of the Chinese Society of Agricultural Engineering (农业工程学报), 2017, 33(6): 107.

Estimation of Organic Matter, Moisture, Total Iron and pH From Black Soil Based on Multi Scales SNV-CWT Transformation

TAN Yang, JIANG Qi-gang*, LIU Hua-xin, LIU Bin, GAO Xin, ZHANG Bo

College of Geo-exploration Science and Technology, Jilin University, Changchun, 130026, China

Abstract Soil composition is complex and varied. Predicting the contents of soil properties fast and efficiently is important for precision agriculture. Spectra are usually measured on dried soil samples. However, soil moisture is an important indicator for the guidance of agriculture activities. In order to predict the soil organic matter (SOM), soil moisture content (SMC), total iron (Fe) and pH value, we propose to measurement VIS-NIR spectra directly on wet samples and use Standard normal variable (SNV)-Continuous wavelet transform (CWT) method on spectra. CWT method uses Mexh as wavelet filter and 10 scales after SNV on each spectrum. Seven common methods, including Gauss filter (GS), First derivative (FD), Continuous removal (CR), and Mathematical transform ($\text{Log}(1/R)$) et al were used as comparisons. All of 74 samples were divided into 50 and 24, for calibrated and validation datasets. On the coefficients of each scale after SNV-CWT, wavebands that passed 0.05 significance level were selected as RF input variables. The optimal scale for each property was confirmed based on the statistical indicators of validation models. Then the Pearson correlation coefficients (PCC), Model based coefficients (MBC) and Grey relation degree (GRD) between each property and wavelet coefficients were calculated on the optimal scales. Models were estimated by the filter screening method based on the correlation coefficients calculated by the three methods. Results showed that, accuracies of all properties were improved after SNV-CWT comparing to the 7 commonly methods. The optimal transformation scales were 7, 8, 1 and 10, corresponding to SOM, SMC, Fe and pH respectively. When taking high dimension features as input variables, the Coefficient of Determination (R^2) was reached to 0.90 and 0.93. The best analysis method was MBC. Because the models performed best when wavebands for the models were selected using MBC as a screening method, the R^2 of SOM and SMC was 0.94 and the accuracies of Fe ($R^2 = 0.67$, $\text{Mse} = 0.01\%$, $\text{RPD} = 1.76$) and pH ($R^2 = 0.80$, $\text{Mse} = 0.1$, $\text{RPD} = 2.24$) were greatly improved, methods can be used for extracting and monitoring multi soil properties.

Keywords Black soil area; Vis-Nir spectra; Data processing; Soil organic matter; Soil moisture; Total iron; pH

(Received Jan. 25, 2021; accepted Apr. 9, 2021)

* Corresponding author