

基于PCA的土壤含水量高光谱反演模型比较研究

张磊¹, 汪泓^{1*}, 董俊红², 肖玫军³, 贾煜¹

¹贵州大学矿业学院, 贵州 贵阳

²北京威特空间技术有限公司, 北京

³贵州省山地资源研究所, 贵州 贵阳

收稿日期: 2021年10月5日; 录用日期: 2021年10月26日; 发布日期: 2021年11月8日

摘要

土壤含水量(SMC)是农作物生长过程中的重要指标之一, 快速测定土壤水分状况对于农业发展极为重要。针对高光谱波段信息冗余和共线性特点以及光谱测定中无关信息和噪声带来的影响, 对原始光谱进行倒数对数一阶微分变换预处理, 利用主成分分析(PCA)对光谱信息进行降维, 提取出12个主成分作为自变量; 利用偏最小二乘、支持向量机、随机森林、BP神经网络建立SMC预测模型并进行综合对比分析。结果表明: 利用PCA对光谱信息降维后的主成分建立的模型均有良好的预测能力, 选取的主成分累积贡献率达到92%以上; 通过对光谱变换前后模型的综合对比可以发现PCA-RF模型精度最高(R^2 分别为0.8362, 0.9630, RPD分别为2.4229、3.7019)。

关键词

高光谱, 土壤水含量, 主成分分析, 倒数对数一阶微分变换

Comparison of Hyperspectral Soil Water Content Inversion Models Based on PCA

Lei Zhang¹, Hong Wang^{1*}, Junhong Dong², Jiujun Xiao³, Yu Jia¹

¹College of Mining, Guizhou University, Guiyang Guizhou

²Beijing Witt Space Technology Co. Ltd., Beijing

³Guizhou Institute of Mountain Resources, Guiyang Guizhou

Received: Oct. 5th, 2021; accepted: Oct. 26th, 2021; published: Nov. 8th, 2021

*通讯作者。

文章引用: 张磊, 汪泓, 董俊红, 肖玫军, 贾煜. 基于 PCA 的土壤含水量高光谱反演模型比较研究[J]. 应用数学进展, 2021, 10(11): 3659-3667. DOI: 10.12677/aam.2021.1011387

Abstract

Soil moisture content (SMC) is one of the important indexes in the process of crop growth. How to quickly measure soil moisture is very important. In the previous SMC water content inversion research, the characteristic band and linear empirical model are mostly used to construct the inversion model, and the characteristics of hyperspectral band information redundancy and collinearity are not fully considered, so the prediction ability of the model needs to be improved. In order to reduce the influence of irrelevant information and noise in spectral determination, the original spectrum was preprocessed by reciprocal logarithm first-order differential transformation; In order to weaken the redundancy and collinearity of hyperspectral band information, principal component analysis (PCA) is used to reduce the dimension of the original spectral information, and 12 principal components are extracted as independent variables; The SMC prediction model is established by using partial least squares, support vector machine, random forest and BP neural network. The results show that PCA has good prediction ability for the principal component model after spectral dimensionality reduction, and the cumulative contribution rate of the selected principal component is 99.99%; Through the comprehensive comparison of the models before and after spectral transformation, it can be found that the accuracy of PCA-RF model is the highest (R^2 is 0.8362 and 0.9630 respectively, RPD is 2.4229 and 3.7019 respectively).

Keywords

Soil Water Content, Principal Component Analysis, First Order Differential Transformation of Reciprocal Logarithm

Copyright © 2021 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

土壤含水量(soil moisture content, SMC)是影响农作物生长、发育和墒情监测的重要因素之一,也是农业灌溉、管理决策中重要的基础数据[1]。通过快速、准确的 SMC 监测可以判定出农作物所需求的水分指标,对农业生产建设具有建设性的指导意义,传统的 SMC 监测方法费时费力、成本高昂,迫切需要采用一个高效准确可信度高的监测方式[2]。高光谱遥感具有光谱分辨率高、波段连续性强、信息量丰富、无损零污染等特点[3] [4]。

利用高光谱技术来快速监测土壤含水量已成为近年来一个研究热点[5]。Tao Song 等[6]利用高光谱数据分析与 SMC 的敏感波段,建立单线性回归模型预测土壤水含量; Gao Y 等[7]利用了多元线性回归(SMLR)和偏最小二乘法(PLSR)等线性回归方法建立了土壤含水量反演模型,并发现在一阶土壤反射率的基础上模型精度较高; Xia K 等[8]通过相关系数方法挑选出敏感特征波段,将特征波段组合成光谱指数建立了偏最小二乘法(PLSR)对研究区土壤含水量进行反演。目前国内外在选择建立土壤含水量反演模型的方式上大多利用特征波段或光谱指数建立线性回归模型,虽然取得了较好的效果,但其实高光谱数据具有波段信息冗余、共线性、噪声的问题,针对这一问题,大多数学者只是通过相关性分析的方法挑选出几个敏感特征波段,丢失了大量重要的信息,且仅用几个波段与土壤某一个属性之间的关系不具有得到最优预测解的可能性[9]。因此选择一种可以充分代表波段信息又减少了输入变量的方法极为重要。当

前研究中建立土壤含水量模型多采用多元线性回归或偏最小二乘回归法等线性回归模型, 较少使用非线性模型来预测土壤含水量[10]。因此简单地使用线性模型不能够充分说明它们之间的关系。

本文以贵阳市花溪区采集的农田土壤样品为研究对象, 针对高光谱土壤水分反演中波段信息噪声、共线性、冗余等问题, 利用 PCA 算法对光谱信息降维, 以获取的主成分为自变量建立多种土壤含水量反演模型, 对比筛选出最有效的土壤含水量反演模型, 验证了基于 PCA 算法实现土壤含水量高精度反演的可行性, 以期为实现高光谱影像快速、精确土壤水分含量区域反演及监测提供科学技术基础和参考。

2. 试验材料与方法

2.1. 研究区概况与土样采集

以贵州省贵阳市花溪区(106°27'~106°352'E, 26°11'~26°34'N)为研究区, 花溪区地貌主要为山地、丘陵, 处云贵高原东斜坡和苗岭山脉中段, 为典型的喀斯特地质区域, 总面积大约 964 km², 平均海拔高度为 1100 m 左右, 属于亚热带湿润温和型气候, 年平均气温在 15.5℃, 年平均降水 1215.70 mm。根据研究区耕地的分布状况和地理条件合理布设 50 个采样点, 使用直径、深度分别为 7.2 cm、5 cm 的环刀采集黄壤土壤样品, 采集时剔除表层杂草、碎石等, 保持土壤原状, 同时使用 GPS 记录采样点坐标, 将环刀密封编号后称重记录。

2.2. 土壤光谱与含水量测定

研究选用 ASD FieldSpec4 地物光谱仪在暗室中测定土样反射率光谱, 该仪器波段范围为 350~2500 nm, 采样间隔分为 1.4 nm (波段 350~1000 nm)、2 nm (波段 1000~2500 nm), 经重采样后变为 1 nm 间隔。在暗室中将土样置于实验黑色器皿中并刮平土壤表面, 光谱测定使用高探头测量, 测定光谱时需要贴紧土壤表面, 测量前需要校准标准白板的参数并去除暗电流, 在土壤表面 4 个不同位置测量, 在每个位置上采集 5 条光谱曲线, 共采集 20 条光谱曲线, 利用 ViewSpecPro 求其平均值作为土样的反射率光谱数据, 由于 350~399、2451~2500 nm 范围光谱在测定时噪声比较大, 故剔除这两个范围波段后作为原始数据。待光谱测量完毕之后, 将土样放入铝盒中称重, 记为 w_a , 并记录铝盒编号。将土样放入烘干机中进行烘干实验, 时间为 24 h, 温度为 110℃, 待土样烘干之后, 分别测定土样加铝盒及单独铝盒的重量, 记为 w_b 、 w_c , SMC 的计算公式为:

$$SMC = \frac{w_a - w_c}{w_b - w_c} \times 100\% \quad (1)$$

2.3. 光谱数据预处理与主成分分析

通过光谱对数微分变换可以去除或减少土样表面粗糙度不一致、观测光照强度变化对目标光谱产生的影响, 能够增强光谱特征、去除背景噪声[11], 故对原始数据进行倒数对数的一阶导数变换。计算公式如下:

$$\log'(1/R(\lambda_i)) = \frac{\log(1/R(\lambda_{i+1})) - \log(1/R(\lambda_{i-1}))}{(\lambda_{i+1} - \lambda_{i-1})} \quad (2)$$

式中 λ_i 、 λ_{i-1} 、 λ_{i+1} 表示各相邻波长。

主成分分析(Principal components analysis, PCA)是一种可靠性强、使用最广泛的数据降维算法[12], PCA 的思想是将 m 维特征向量映射到 n 维特征向量上, 即将高维特征转换为少数几个主成分, 且主成分之间互不相关, 并使其具有最大信息量, 从而达到简化数据的目的。假定包含有 m 个样本、 n 个特征指标的数据集, 则其可以构成矩阵:

$$X = \begin{pmatrix} x_{11} & \cdots & x_{1n} \\ \vdots & \ddots & \vdots \\ x_{m1} & \cdots & x_{mn} \end{pmatrix} = (x_1, x_2, \dots, x_n) \quad (3)$$

PCA 主要步骤如下:

1) 首先将矩阵 X 进行标准化变换, 其均值为 $\bar{x}_n = \frac{1}{m} \sum_{i=1}^m x_{in}$, 标准差为 $S_n = \sqrt{\left(\sum_{i=1}^m (x_{in} - \bar{x}_n)^2 / (n-1) \right)}$,

标准化公式为:

$$X_{mn} = \frac{x_{mn} - \bar{x}_n}{S_n} \quad (4)$$

原始矩阵经过标准化得到矩阵 $X = (X_1, X_2, \dots, X_n)$ 。

2) 计算矩阵 X 的协方差矩阵及其对应的特征值和特征向量, 若满足

$$\begin{cases} B_1 = a_{11}X_1 + a_{12}X_2 + \cdots + a_{1n}X_n \\ B_2 = a_{21}X_1 + a_{22}X_2 + \cdots + a_{2n}X_n \\ \vdots \\ B_m = a_{m1}X_1 + a_{m2}X_2 + \cdots + a_{mn}X_n \end{cases} \quad (5)$$

且 $a_{11}^2 + a_{12}^2 + \cdots + a_{1n}^2 = 1$ 、 $\text{Var}(B_1)$ 值最大, 则称 B_1 为原数据的第一主成分; 若 $a_{21}^2 + a_{22}^2 + \cdots + a_{2n}^2 = 1$, $(a_{21}, a_{22}, \dots, a_{2n})$ 与 $(a_{11}, a_{12}, \dots, a_{1n})$ 向量垂直, 且 $\text{Var}(B_2)$ 值最大, 则称 B_2 为原数据的第二主成分; 可以由此推出第三、第四, ……等 n 个主成分。通常选取累计方差贡献率达到 90% 以上的前 h 个主成分, 可以有效减少主成分的个数, 同时保证了信息量的可靠性, 并将选取的 h 个主成分作为模型的输入变量。

2.4. 土壤含水量反演模型的建立与验证

研究采用线性模型 PLSR 和非线性模型 RF、BPN、SVM 综合分析对比 SMC 预测能力。通过对反射率光谱以及光谱变换后进行主成分分析筛选出主成分, 选其作为自变量、SMC 为因变量建立模型估算土壤含水量。PLSR 是一种在高光谱遥感模型中广泛使用的新型统计方法, 它可以很好的解决高光谱反演中自变量数据之间存在的多重共线性的问题[13]。BPN 神经网络是一种多层的前馈神经网络, 主要包含输入层、隐含层和输出层, 可以满足复杂的多维函数映射需求, 仅需通过训练数据进行自身的训练, 得到某种映射关系, 就可以给出最接近期望值的结果[14]。RF 是以决策树为基学习器构建 Bagging 集成的基础上, 进一步构建无数个小决策数引入随机属性选择, 使得最终集成的泛化性能可通过个体学习器之间的差异度的增加而进一步提升[15]。SVM 是一种基于统计学的模式识别方法, 它在解决小样本、非线性、高维数据集的回归问题上有许多独特的优点[16]。

模型的精度评价指标选用决定系数(R^2)、均方根误差(RMSE)、相对分析误差(RPD), 通常越好的模型, 其 R^2 和 RPD 越高, 而 RMSE 越低[17]。 R^2 的值越接近 1 越代表模型的估测效果好, RPD 的值大于 2 时代表模型具有非常好的估测能力, RPD 的值在区间[1.4, 2]时代表模型具有较好的估测能力, RPD 的值小于 1.4 时代表模型不具备估测能力。

3. 结果与分析

3.1. 土壤含水量概况

从表 1 可以看出, 样本划分为 37 个训练集和 13 个验证集, 训练集、验证集 SMC 的平均值分别为

0.156、0.154，总样本的 SMC 均值为 0.155，且变异系数为 0.407，介于训练集 0.385 和验证集 0.455 之间，属于中等变异。

Table 1. Description and statistics of SMC

表 1. 土壤含水量描述统计

Sample type	Number of samples	Minimum value	Maximum value	Mean value	Standard deviation	CV
Training set	37	0.0365	0.281	0.156	0.060	0.385
Validation set	13	0.0526	0.244	0.154	0.070	0.455
Whole set	50	0.0365	0.28	0.155	0.063	0.407

3.2. 光谱变换与 PCA 主成分分析

土壤样本的原始反射率光谱曲线如图 1(a)所示，所有的光谱曲线走势大致一致且总体呈下降趋势，在可见光波段范围内反射率快速变化，在近红外波段后整体变化速率较为缓慢，可以发现土壤水分在 1400、1900、2200 nm 波段附近存在明显的吸收峰。图 1(b)为经过倒数对数的一阶导数变换后的光谱曲线，所有的光谱曲线走势相同且在小范围波段发生剧烈变化，在 1400、1900 nm 波段附近形成较明显的波峰。

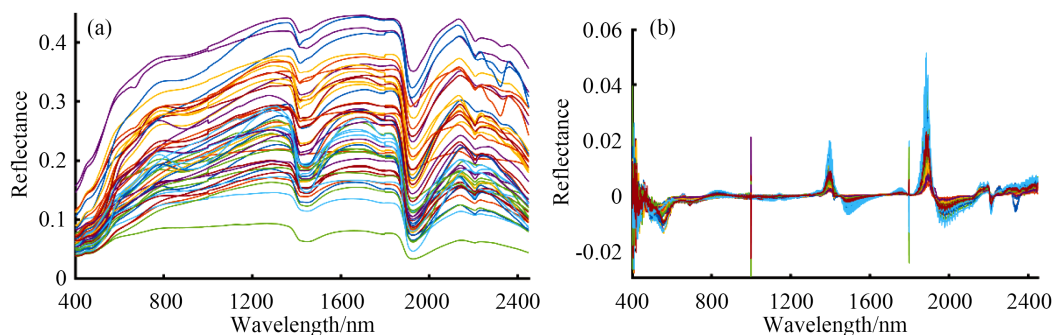


Figure 1. (a) The original reflectance spectrum curve; (b) The first derivative reflectance spectral curve of reciprocal logarithm

图 1. (a) 原始反射率光谱曲线；(b) 倒数对数的一阶导数反射率光谱曲线

研究利用 PCA 算法对光谱 2050 个波段降维并筛选主成分。通过对土样反射率数据以及对数微分预处理后的数据进行主成分分析后，按照通常研究中选取主成分累积贡献率达到 90% 以上的要求，选出前 12 个主成分分量，如表 2 所示。选取了 PCA 降维后前 12 个主成分，光谱原始反射率以及倒数对数一阶导数反射率曲线的各主成分累积贡献率分别为 99.996%、92.061%。将 PCA 降维后获取的 12 个主成分作为自变量输入模型。

Table 2. The principal component contribution rate of soil spectrum

表 2. 土壤光谱主成分贡献率

Component	Original spectrum		Component	The first derivative spectrum of reciprocal logarithm	
	Contribution rate/%	Cumulative contribution rate/%		Contribution rate/%	Cumulative contribution rate/%
1	94.506	94.506	1	54.746	54.746
2	3.222	97.728	2	9.907	64.653

Continued

3	1.583	99.311	3	5.256	69.909
4	0.438	99.749	4	4.222	74.131
5	0.096	99.846	5	3.718	77.849
6	0.087	99.933	6	3.273	81.122
7	0.025	99.958	7	2.597	83.719
8	0.016	99.974	8	2.436	86.155
9	0.011	99.985	9	1.911	88.066
10	0.005	99.991	10	1.522	89.588
11	0.003	99.994	11	1.322	90.910
12	0.002	99.996	12	1.151	92.061

3.3. SMC 反演模型建立与检验

3.3.1. 原始光谱建模

利用 PCA 对原始光谱数据提取主成分光谱特征, 将 12 个主成分作自变量, SMC 作为因变量, 得到 PCA-PLSR、PCA-SVM、PCA-BPN、PCA-RF 四种 SMC 反演模型, 模型精度评价如图 2(a)所示。由图 2(a)、表 3 可得: PCA-RF 的决定系数($R^2 = 0.8362$)最高, RPD 值达到 2.4229, 表明了 PCA-RF 模型的预测能力非常好, 是优异的 SMC 预测模型; PCA-BPN 的预测能力次之, 同样具有良好的预测能力, R^2 和 RPD 分别为 0.7844、2.1835; 而 PCA-SVM 和 PLSR 的预测能力稍差一些, R^2 分别为 0.7646、0.7234, RPD 值都大于 1.8, 具有较好的预测能力。客观的说, 通过 PCA 主成分建立的 SMC 估测模型具有良好的预测效果。

3.3.2. 倒数对数的一阶导数光谱建模

基于 PCA 对倒数对数的一阶微分光谱进行 4 种反演模型构建[图 2(b)], 其模型精度评价如表 3 所示。根据图 2(b)以及表 3 可知, PCA-RF 模型的反演能力仍是最好的, 其 R^2 高达 0.9630, RPD 值为 3.7019, 表明其具有非常好的预测精度; PCA-BPN 次之, 其 R^2 为 0.8642, RPD 为 2.7942, 模型的精度也很好; PCA-SVM 的 R^2 为 0.8571, RPD 为 2.7624; 而 PCA-PLSR 反演精度最低, 其 R^2 为 0.7341, RPD 为 1.9095。

3.3.3. 模型对比分析

根据图 3 中不同模型的拟合效果可知, 4 种模型的中 PCA-RF 模型在光谱对数微分变换前后的精度都是最高的, PLSR 精度最低, 在进行对数微分变换后各模型预测能力都得到了提升, 其中 PCA-RF 提升效果最为明显, 其 R^2 值提高了 15.16%; RPD 值提高了 52.79%; 其次为 PCA-SVM 模型, 其 R^2 值提高了 12.1%; RPD 值提高了 42.8%; PCA-BPN 模型, 其 R^2 值提高了 10.17%; RPD 值提高了 27.97%; 而 PCA-PLSR 模型提升不明显。

通过各模型的预测精度对比可以发现, 非线性模型 PCA-BPN、PCA-RF、PCA-SVM 的预测能力优于线性模型 PCA-PLSR, 光谱对数微分变换可以有效的提升模型的预测能力, 其能够降低光谱测定中背景噪声、无关信息带来的影响。

4. 讨论

光谱数学变换可以极大的消除光谱测定中不同背景、噪声带来的影响[17]。本文选用对数微分变换算法对原始光谱进行预处理, 利用 PCA 算法对变换前后的数据降维提取主成分, 结合 RF、BPN、SVM、

PLSR 建立不同的土壤含水量估测模型, 比较选出最佳 SMC 预测模型。一般研究中, 通常选择累积方差贡献率达到 90% 以上的主成分个数, 本文中选取的 12 个主成分累积方差贡献率达到了 92% 以上, 非常具有代表原始信息的能力, 同时也简化了模型输入自变量, 提高了模型运算速度。

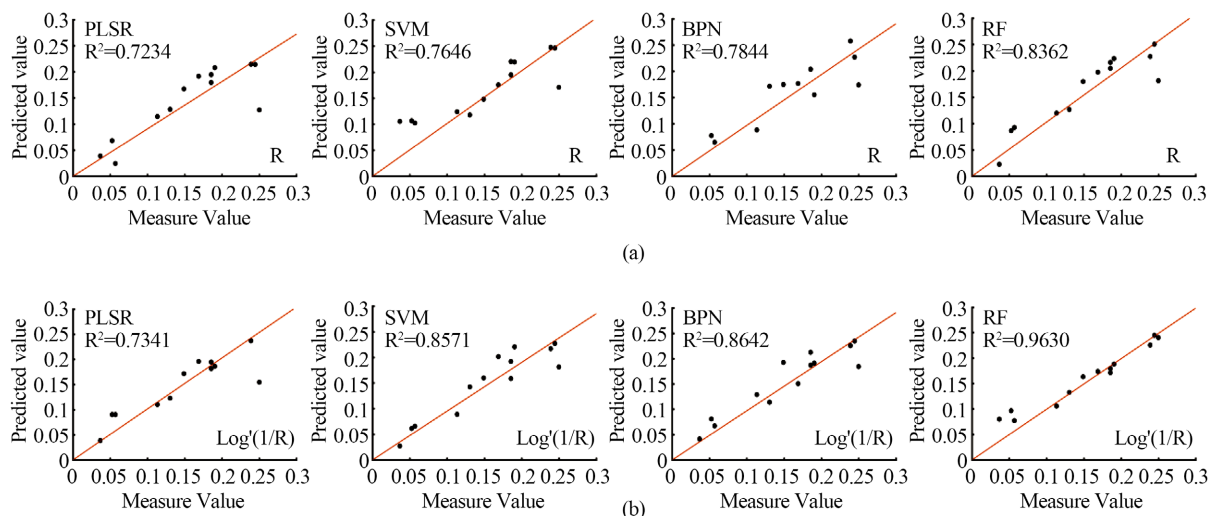


Figure 2. (a) Scatter diagram of different prediction models of original spectrum; (b) Scatter diagram of different prediction models of first-order differential spectrum of reciprocal logarithm

图 2. (a) 原始光谱不同预测模型散点图; (b) 倒数对数的一阶微分光谱不同预测模型散点图

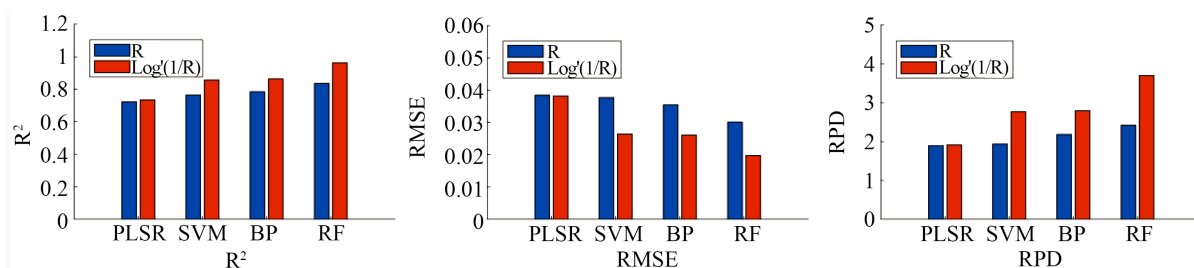


Figure 3. The original spectrum reciprocal logarithm first-order differential comparative analysis diagram

图 3. 原始光谱 - 倒数对数一阶微分对比分析图

Table 3. Accuracy analysis of different models based on original spectrum reciprocal log differential spectrum

表 3. 基于原始光谱 - 倒数对数微分光谱不同模型精度分析

Model		Original spectrum			The first derivative spectrum of reciprocal logarithm		
		R ²	RMSE	RPD	R ²	RMSE	RPD
linear model	PCA-PLSR	0.7234	0.0384	1.8972	0.7341	0.0382	1.9095
	PCA-SVM	0.7646	0.0377	1.9344	0.8571	0.0264	2.7624
Non-linear model	PCA-BPN	0.7844	0.0354	2.1835	0.8642	0.0261	2.7942
	PCA-RF	0.8362	0.0301	2.4229	0.9630	0.0197	3.7019

通过对比不同 SMC 估测模型可以发现, 在对光谱数据预处理后, 除了线性模型提升不明显以外, 非线性模型的精度都得到了较大提升, 其 R² 值都高于 0.8, 表明了光谱对数微分变换处理能提高模型预测的精度。这主要是因为对数微分变换可以去除光谱测定中不同背景、噪声的影响, 特别是一阶微分可以

最大的消除一次方项噪声的影响,提高光谱分辨率和灵敏度[18]。目前在土壤水含量以及有机质反演研究中,已经证明了对数微分变换处理能有效提高模型的预测精度[19]。在本研究中也证明了光谱对数微分预处理对于模型的预测能力有很好提升效果,且非线性模型的预测能力要优于线性模型。通常情况下土壤光谱会受到土壤质地、有机质含量、土壤颗粒大小等的影响,其反射光谱会随之产生变化[20]。本次研究的土壤为单一黄壤土壤,未考虑其他土壤质地和不同有机质含量对模型的影响,所以还需要考虑在其他背景条件下模型的预测效果;研究主要选用 PCA 算法对高光谱数据进行降维,虽然取得了很好的效果,但并未考虑其他降维算法,有待进一步研究。

5. 结论

本文以贵阳市花溪区 50 个不同含水量土壤样本为研究对象,以其室内光谱和土壤含水量为基础,利用光谱变换结合 PCA 算法提取出的主成分为自变量,建立了不同的 SMC 估测模型,并对其进行综合对比分析得出:1) 经过倒数对数一阶微分变换后建立的模型预测能力得到了有效提升,非线性模型的提升效果明显优于线性模型,表明倒数对数一阶微分变换可以有效降低光谱测定中无关信息和噪声带来的影响。2) 利用 PCA 提取出的 12 个主成分建立的模型均具有很好的预测能力,表明通过 PCA 可以有效地简化数据量。3) 光谱变换后 PCA-RF 模型预测能力最优, R^2 和 RPD 分别为 0.9630、3.7019,表明 PCA-RF 模型在 SMC 估测中更为有效。

基金项目

贵州省科技支撑计划项目(黔科合支撑[2021]一般 496)。

参考文献

- [1] 陈仲新, 任建强, 唐华俊, 等. 农业遥感研究应用进展与展望[J]. 遥感学报, 2016, 20(5): 748-767.
- [2] 孙越君, 郑小坡, 秦其明, 等. 不同质量含水量的土壤反射率光谱模拟模型[J]. 光谱学与光谱分析, 2015(8): 2236-2240.
- [3] 吴代晖, 范闻捷, 崔要奎, 等. 高光谱遥感监测土壤含水量研究进展[J]. 光谱学与光谱分析, 2010, 30(11): 3067-3071.
- [4] Muller, E. and Décamps, H. (2001) Modeling Soil Moisture-Reflectance. *Remote Sensing of Environment*, **76**, 173-180. [https://doi.org/10.1016/S0034-4257\(00\)00198-X](https://doi.org/10.1016/S0034-4257(00)00198-X)
- [5] Whiting, M.L. (2009) Measuring Surface Water in Soil with Light Reflectance. *Proceedings of SPIE*, **7454**, 74540D. <https://doi.org/10.1117/12.826896>
- [6] 宋韬, 鲍一丹, 何勇. 利用光谱数据快速检测土壤含水量的方法研究[J]. 光谱学与光谱分析, 2009, 29(3): 675-677.
- [7] Gou, Y., Wei, J., Li, J.-L., Han, C., Tu, Q.-Y. and Liu, C.-H. (2020) Estimating Purple-Soil Moisture Content Using Vis-Nir Spectroscopy. *Journal of Mountain Science*, **17**, 2214-2223. <https://doi.org/10.1007/s11629-019-5848-2>
- [8] Xia, K., Xia, S., Shen, Q., Yang, B., Song, Q., Xu, Y., et al. (2021) Moisture Spectral Characteristics and Hyperspectral Inversion of Fly Ash-Filled Reconstructed Soil. *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy*, **253**, Article ID: 119590. <https://doi.org/10.1016/j.saa.2021.119590>
- [9] Rinnan, Å., Van Den Berg, F., Engelsen, S.B. (2009) REVIEW of the Most Common Pre-Processing Techniques for Near-Infrared Spectra. *TrAC Trends in Analytical Chemistry*, **28**, 1201-1222. <https://doi.org/10.1016/j.trac.2009.07.007>
- [10] 郭飞, 许镇, 马宏宏, 等. 基于 PCA 的土壤 Cd 含量高光谱反演模型对比研究[J]. 光谱学与光谱分析, 2021, 41(5): 1625-1630.
- [11] Tsai, F. and Philpot, W. (1998) Derivative Analysis of Hyperspectral Data-Fluorescence in Yellow-Green. *Remote Sensing of Environment*, **66**, 41-51. [https://doi.org/10.1016/S0034-4257\(98\)00032-7](https://doi.org/10.1016/S0034-4257(98)00032-7)
- [12] Savitzky, A. (1964) Smoothing and Differentiation of Data by Simplified Least Squares Procedures. *Analytical Chemistry*, **36**, 1627-1639. <https://doi.org/10.1021/ac60214a047>
- [13] Cui, Y. and Fang, Y. (2020) Research on PCA Data Dimension Reduction Algorithm Based on Entropy Weight Me-

-
- thod. 2020 *2nd International Conference on Machine Learning, Big Data and Business Intelligence (MLBDBI)*, Taiyuan, 23-25 October 2020, 392-396. <https://doi.org/10.1109/MLBDBI51377.2020.00084>
- [14] Vasques, G.M., Grunwald, S. and Sickman, J.O. (2008) Comparison of Multivariate Methods for Inferential Modeling of Soil Carbon Using Visible/Near-Infrared Spectra. *Geoderma*, **146**, 14-25. <https://doi.org/10.1016/j.geoderma.2008.04.007>
- [15] 赵建辉, 张晨阳, 闵林, 等. 基于特征选择和 GA-BP 神经网络的多源遥感农田土壤水分反演[J]. 农业工程学报, 2021, 37(11): 112-120.
- [16] 颜文杰, 卢雯慧, 王继芬. 基于 SVM-MLP 融合模型的毒品混合物光谱识别研究[J]. 激光与光电子学进展, 2021, 58(14): 1404003. <https://doi.org/10.3788/LOP202158.1404003>
- [17] 蔡亮红, 丁建丽. 基于高光谱多尺度分解的土壤含水量反演[J]. 激光与光电子学进展, 2018, 55(1): 013001. <https://doi.org/10.3788/LOP55.013001>
- [18] Breiman, L. (2001) Random Forests. *Machine Learning*, **45**, 5-32. <https://doi.org/10.1023/A:1010933404324>
- [19] Brown, D.J., Shepherd, K.D., Walsh, M.G., Dewayne Mays, M. and Reinsch, T.G. (2006) Global Soil Characterization with VNIR Diffuse Reflectance Spectroscopy. *Geoderma*, **132**, 273-290. <https://doi.org/10.1016/j.geoderma.2005.04.025>
- [20] 喻武, 贾晓琳, 陈颂超, 周炼清, 史舟. Vis-NIR 光谱快速估测土壤可侵蚀性因子可行性分析[J]. 光谱学与光谱分析, 2018, 38(4): 1076-1081.