doi: 10.6046/gtzyyg.2020095
引用格式: 吴倩,姜琦刚,史鹏飞,等. 基于高光谱的土壤碳酸钙含量估算模型研究[J]. 国土资源遥感,2021,33(1):138 - 144.
(Wu Q,Jiang Q G,Shi P F,et. al. The estimation of soil calcium carbonate content based on Hyperspectral data [J]. Remote Sensing for Land and Resources,2021,33(1):138 - 144.)

基于高光谱的土壤碳酸钙含量估算模型研究

吴 倩, 姜琦刚, 史鹏飞, 张莉莉
(吉林大学地球探测科学与技术学院,长春　130026)

摘要: 土壤中碳酸钙的含量是土壤分类及肥力评价的重要依据。研究选取陕西省黄土高原区的 78 个黄绵土土壤样品为研究对象,在分析碳酸钙含量的基础上,采用 SVC HR - 1024i 便携式光谱仪获取土壤样品的可见光、近红外与短波红外高光谱反射率(350 ~ 2 500 nm)数据,对原始光谱曲线分别进行一阶微分、二阶微分和连续统去除 3 种数学变换,运用相关分析法和连续投影算法分别进行敏感波段的选取,采用随机森林回归建立土壤碳酸钙的估算模型。结果表明: 黄绵土的光谱曲线特征相似,在 1 440 nm,1 900 nm 和 2 200 nm 等处均存在明显的吸收特征,且碳酸钙含量与光谱反射率呈现正相关态势; 基于二阶微分与连续投影算法的随机森林估算模型精度最高,验证集 R^2 为 0.82,相对分析误差(residual predictive deviation,RPD)值为 2.37。

关键词: 碳酸钙含量; 黄绵土; 高光谱; 连续投影算法; 随机森林回归
中图法分类号: TP 79　**文献标志码**: A　**文章编号**: 1001 - 070X(2021)01 - 0138 - 07

0 引言

土壤是自然界不可或缺的资源[1],它为作物的生长提供了一定的养分和水分。碳酸钙($CaCO_3$)是土壤中碳酸盐的主要成分,对土壤物理、化学性质、碳的沉积、土壤酸碱度、土壤胶体性状等的研究都有重要意义[2-3]。近年来,随着遥感科学的发展,高光谱遥感应用的深入,人们越来越体会到定量遥感的必要性,基于高光谱的土壤理化性质快速、定量、准确的模型构建成为农业遥感研究领域研究的重点[4]。土壤的光谱反射特性是土壤基本特性之一,与土壤中的各种理化性质有着十分密切的关系,土壤中各种物质特殊的光谱特征以及土壤组成物质的多样性都会在光谱曲线上体现,碳酸钙的含量对土壤光谱反射特性具有一定的影响,Gaffey[5]研究发现碳酸盐矿物由于碳酸根的振动过程,在可见 - 近红外(350 ~ 2 550 nm)光谱区域有 7 个较强的吸收特征; 张婷婷[6]研究认为土壤中碳酸盐含量对土壤的光谱特征有一定的影响,在 2 230 ~ 2 270 nm 之间有明显的吸收特征; Stenberg[7]研究认为土壤碳酸盐在 2 335 nm 处有较强的吸收特征; 以往研究表明碳酸钙的可见光、近红外(VNIR,400 ~ 1 200 nm)、短波红外(SWIR,1 200 ~ 2 500 nm)光谱可作为土壤碳酸钙含量估算的理论依据,是一种物理性的无损伤和低成本的研究方法,可以用来区分土壤类型和评估大量土壤特性。为了更好地理解电磁波与地表特征之间的相互作用,国内外学者将经典的数学物理理论与遥感实践相结合,建立了近百种定量遥感模型,这些模型大体可分为统计模型、物理模型、半经验模型 3 种。利用高光谱数据估测土壤理化性质的模型有很多,近年来的研究中大多运用统计模型,建立遥感数据与地面观测数据之间的线性回归方程,统计模型简便易行,一般仅包含 3 ~ 6 个参数。以往研究中多运用偏最小二乘回归、多元线性同步回归、支持向量机等方法,且研究内容大多集中在对土壤有机质含量、土壤水分、土壤游离铁含量的估算建模上。本次研究选取随机森林回归算法(random forest regression,RFR)对土壤碳酸钙含量进行估算,随机森林回归是一种集成学习模型,其优化了组合机器学习和 CART 决策树算法,并且对异常值和有噪声的数据具有一定的容忍度[8],目前更多的应用于数据分类和回归分析[9]。

研究选取的土壤类型为黄绵土,黄绵土质地疏松,呈浅棕色,剖面发育不明显。广泛分布于甘肃、陕西、山西等水土流失较严重地区,是陕西省分布面

收稿日期: 2020 - 04 - 07; **修订日期**: 2020 - 06 - 07
基金项目: 中国地质调查项目"全国地球关键带遥感地质调查"(编号: DD20190536)资助。
第一作者: 吴 倩(1997 -),女,硕士研究生,主要从事遥感制图及其技术应用研究。Email: 1148545835@ qq. com。
通信作者: 姜琦刚(1964 -),男,教授,博士生导师,主要从事遥感地学研究。Email: jiangqigang@ jlu. edu. cn。

积最大的土壤类型,其肥力低下,主要分布在坡度大、植被较少地区,是造成土壤养分有效性低、pH 高,影响土壤结构及加速土壤侵蚀的重要因素之一,因此基于高光谱的土壤碳酸钙含量估算可以准确快速地获取土壤属性信息以及为水土流失地区提供治理等的理论参考,对于评价成土因素以及对土壤性质的研究具有重要意义。

1　材料与方法

1.1　研究区概况

研究区主要位于陕西省北部的黄土高原地区。107°15′41″ ~ 111°14′40″E,34°12′21″ ~ 39°29′17″N,黄土高原属于干旱区与半干旱区;在气候分区上,该地区横跨温带和暖温带,年平均降水量约在 300 ~ 700 mm 之间,长期的水土流失导致生态环境较脆弱。陕西省土壤类型众多,本次研究的土壤类型为黄土高原上分布面积最大的黄绵土,研究共采集了 78 个土样,其中 28 个位于陕西省最北部的榆林市,41 个位于延安市,9 个位于渭南市。图 1 为研究区地理位置及采样点分布。

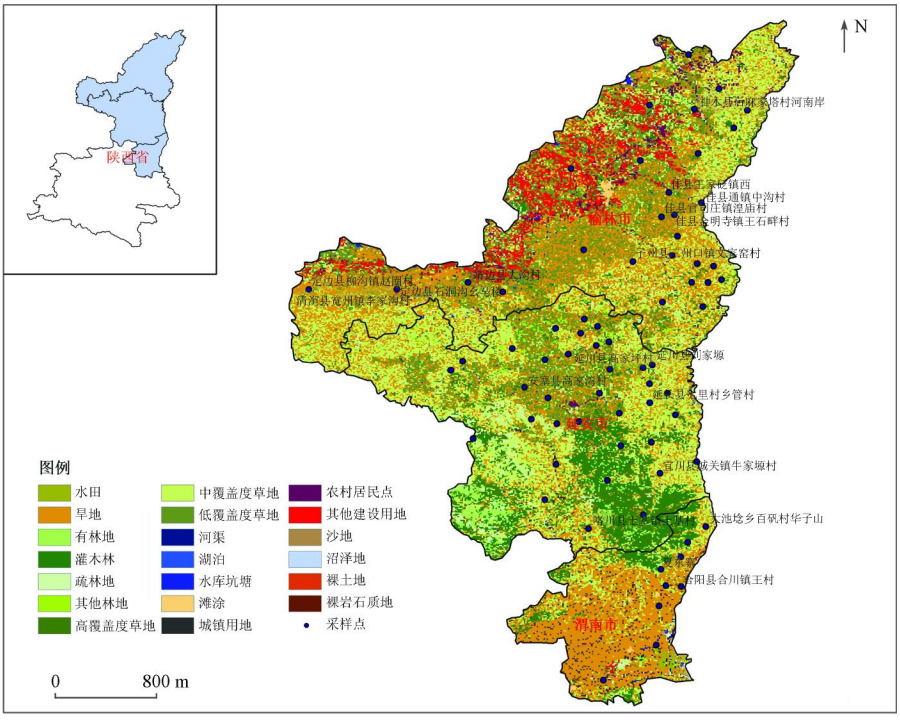


图 1　研究区地理位置及采样点分布

Fig. 1　Location of the study area and the distribution of sampling sites

1.2　样本采集与分析

根据陕西省黄绵土空间分布特征,于 2016 年 7—8 月期间在榆林市、延安市、渭南市共采集具有典型代表性的 78 个黄绵土土壤样品,对土壤样本进行风干、碾磨及过 1 mm 孔筛等操作后,采用四分法,将每个样品分为 2 份,一份利用气量法进行实验室土壤碳酸钙含量测定,另一份用作土壤反射光谱测定。

1.3　光谱测定

土壤 VNIR/SWIR 反射光谱是在暗室中控制辐照度条件测量的,测定采用了美国 Spectra Vista 公司生产的 SVC HR - 1024i 便携式光谱仪进行,其波长范围为 350 ~ 2 500 nm。在 0.6 ~ 1 m 之间放置 1 盏 50 W 白炽灯,设置 60°方位角和 15°天顶角,这种配置确保了由于阴影造成的最小扰动,使测量结果不那么依赖于粗糙度。用 8°透镜提供直径为 0.13 m 的光斑。测量中利用白板进行校正,获得了 3 个连续的和 50 个单独反射测量系列,噪声水平低于 0.001,光源的照射方向与垂直照射之间的夹角为 30°,测量土样在 4 个方向上的光谱曲线。

1.4　数据预处理

为了提高光谱数据的质量,研究利用 ENVI5.3 对建模前的数据进行光谱数据库的建立以及 10 nm 间隔的重采样,重采样处理后每个土壤样本的光谱数据曲线包含 216 个波段。重采样后的光谱曲线依然存在一定量噪声,为了提高预测模型的精度,采用 Savitzky - Golay 卷积平滑法(S - G)对数据进行进一步的处理[10],平滑去噪后,对土壤碳酸钙光谱数据进行数学变换,得到原始光谱(R)的一阶微分(First order differential,R″)数据,二阶微分(Second

order differential, R″) 数据,连续统去除(Continuum Removal, CR)数据。

1.5 敏感波段选取

采用相关分析法(analysis of correlation, CA)与连续投影算法(successive projections algorithm, SPA)分别进行土壤碳酸钙含量与光谱反射率的相关性分析以及敏感波段的筛选,并依据建模结果对两种方法进行评价。

CA 是指对两个或多个具备相关性的变量元素进行分析,从而衡量两个变量因素的相关密切程度[10]。需要将 4 种光谱处理方法与所测样本的碳酸钙含量的真实值逐个波长进行相关性分析,并进行 $p=0.01$ 显著性检验,根据相关性的高低来确定敏感波段,相关性越高,波段越敏感。相关分析采用 SPSS 中的相关性函数进行,通过显著性检验的波段作为接下来的建模的敏感波段。相关系数公式如下:

$$r_i = \frac{\sum_{n=1}^{N}(x_{ni}-\overline{x_i})(y_n-\bar{y})}{\sqrt{\sum_{n=1}^{N}(x_{ni}-\overline{x_i})^2\sum_{n=1}^{N}(y_n-\bar{y})^2}}, \quad (1)$$

式中:r_i 为土壤碳酸钙含量与光谱反射率相关系数;i 为波段序号;N 为土壤样本的个数,x_{ni} 为第 n 个土壤样本第 i 波段的光谱反射率值;$\overline{x_i}$ 为 N 个土壤样本在 i 波段光谱反射率的平均值;y_n 为第 n 个土壤样本的碳酸钙含量实测值;$\bar{y}$ 为土壤样本碳酸钙含量实测值的平均值[11]。

SPA 在多自变量校正领域是一种采用前向选取的特征波段变量选择的方法[12]。能够有效地减少信息重叠,使变量之间的共线性达到最小,大大减少了建模变量的个数,有效提高建模的效率。实现用较少的信息量代表多数样本的光谱信息的效果[10],SPA 以均方根误差(ROOT mean square error, RMSE)为评价指标,将 RMSE 最小值下的波长个数确定为敏感波段数[13-14],连续投影算法在 Matlab2018a 中实现。

1.6 模型建立与精度验证方法

随机森林(RFR)是一种集成算法(Ensemble Learning),这种算法优化了组合机器学习和 CART 决策树算法[15]。随机森林有很多的优点,首先它能在决策类别时,去评估变量的重要性,其次它对于数据集较多的输入变量的处理精度较好,随机森林的两个应用主要是分类与回归。近几年随机森林在回归分析方面的应用也越来越广泛。该算法的估算精度比较高,树形结构自由生长,能够有效避免其他估算模型中的过度拟合现象,是对传统的决策树模型的一种优化,对数据量较大的数据集比较适用[15]。随机森林回归模型的构建运用 R. 3. 4. 1 中的 Random Forest 包进行。

研究选取土壤碳酸钙含量实测值与估测值的均方根误差(RMSE),决定系数(R^2)以及相对分析误差(residual predictive deviation, RPD)3 个参数对模型进行精度验证,均方根误差与相对分析误差越小,决定系数越大,模型的精度越高。

2 结果与分析

2.1 土壤碳酸钙描述性统计

首先对全部 78 个土壤碳酸钙样本进行异常值分析,将样本中碳酸钙含量不在 ±3 倍标准差中的极度异常的 4 个样本剔除,从剩余的 74 个样本中选择出 52 个样本用于建模,22 个样本用于验证,表 1 为土壤样本碳酸钙含量的统计结果,样本总体碳酸钙含量范围为 11.109 ~ 175.077 $g \cdot kg^{-1}$,其中建模集中的 52 个碳酸钙含量范围为 11.109 ~ 173.509 $g \cdot kg^{-1}$,验证样本碳酸钙含量范围为 11.236 ~ 175.077 $g \cdot kg^{-1}$,样本的选取基本上涵盖了总体样本中的数据范围。建模集与验证集的偏度为 0.46 和 0.25,均在 0 值附近,基本服从正态分布,建模集、验证集与总体样本的偏度差异不大,样本适合用于建立模型。

表 1 土壤样本碳酸钙含量的统计结果

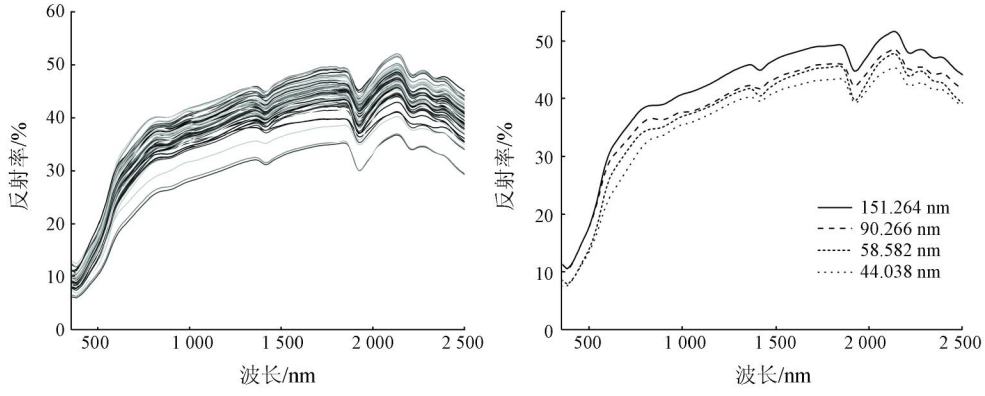
Tab. 1 Statistics of calcium carbonate in soil samples

样本集	数目	最小值/($g \cdot kg^{-1}$)	最大值/($g \cdot kg^{-1}$)	平均值/($g \cdot kg^{-1}$)	标准差	偏度	变异系数/%
建模样本	52	11.109	173.509	85.365	28.262	0.457	44.82
验证样本	22	11.236	175.077	87.607 8	33.798	0.253	49.99
总样本	74	11.109	175.077	86.362	30.536	0.354	46.94

2.2 土壤碳酸钙含量与光谱相关性分析

图 2(a)为全部碳酸钙样本波长与反射率之间的原始光谱图,由图中可以看出,不同碳酸钙含量的土壤反射率曲线总体趋势基本相同,深度有所不同,吸收特征基本相同,在 1 400 nm 和 1 900 nm 和 2 200 nm 处有明显的吸收特征,这与前人分析结果相似,通常认为是由 OH, AL - OH 等引起的[16-17],除此之外,近红外波段的反射率总体大于可见光部分的反射率;如图 2(b)所示,随着碳酸钙含量的增加,光谱反射率也随之增加,两者呈现正相关态势,且近红

外波段的增长速度明显高于可见光部分的增长速度,反射率曲线在于 2 350 nm 处有一吸收谷,这主要是由于碳酸根的影响作用引起的[16]。

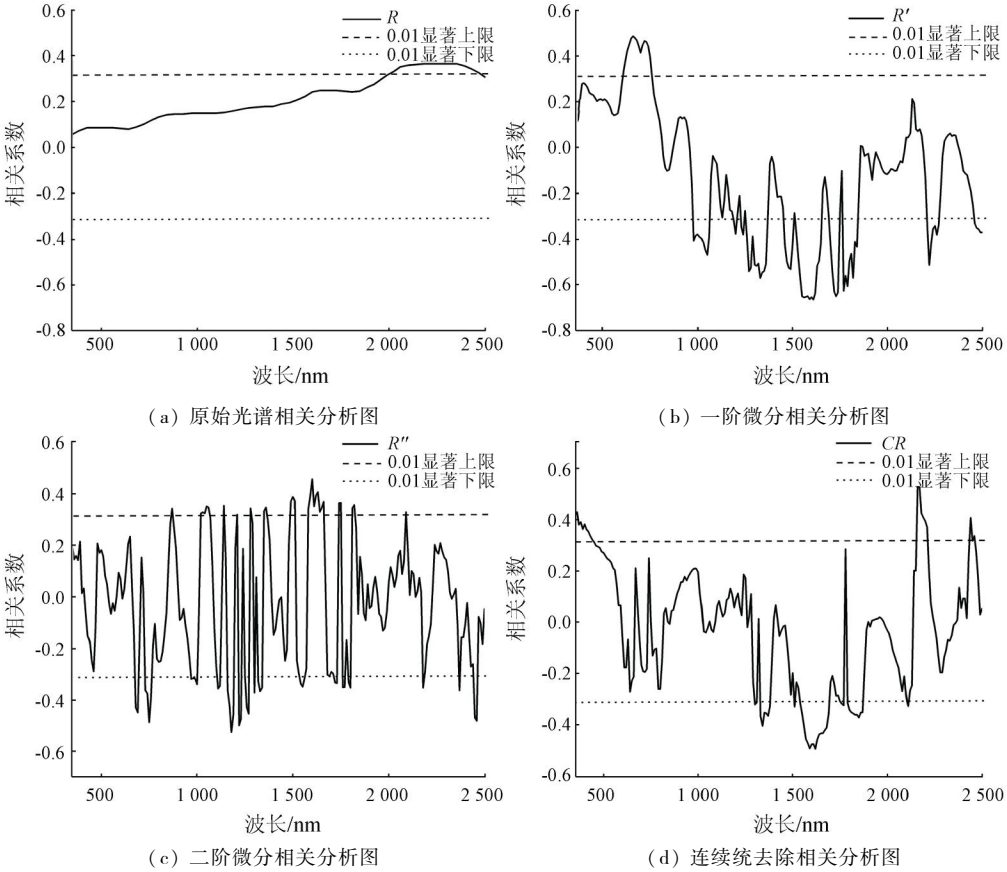


(a) 全部样本反射率图　　(b) 部分样本反射率图

图 2　研究样本波长反射率图

Fig. 2　Reflectance maps of calcium carbonate content of samples

图 3 为基于相关分析算法的敏感选波段筛选图,由图 3 可以看出,原始光谱曲线在全波段的相关性较差,相关系数最高值在 2 340 nm 左右,相关系数仅为 0.37,曲线整体呈现正相关的态势。经过一阶微分、二阶微分、连续统去除后,光谱曲线与土壤碳酸钙含量的相关性得到了显著的提高,从而可以进一步提高模型的预测能力。其中一阶微分相关系数最大值在 1 610 nm 处达到了 -0.66,二阶微分相关系数在 2 230 nm 处达到了 0.56,微分运算虽然可以较好地消除背景信号,但高阶微分在应用中往往表现出不稳定性,在应用中存在一定的问题,以往研究发现,在光谱分析中随着阶数的增加,相关系数呈现先增加后减少的趋势,并在二阶处达到最大值[18],因此,研究采用一阶与二阶微分。连续统去除相关系数在 2 150 nm 处为到 0.53,均达到中度相关水平,4 种数学处理方法分别有 54,82,54,44 个波段入选建模波段。



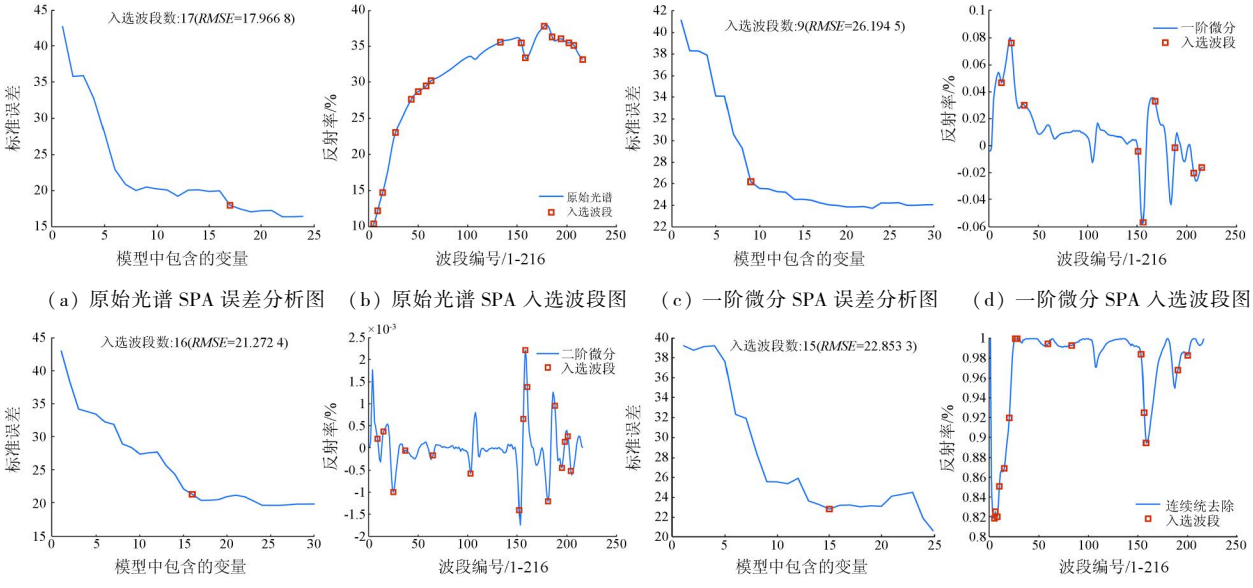
(a) 原始光谱相关分析图　　(b) 一阶微分相关分析图

(c) 二阶微分相关分析图　　(d) 连续统去除相关分析图

图 3　基于相关分析(CA)的敏感波段筛选图

Fig. 3　Sensitive band screening graph based on correlation analysis

图 4 为基于连续投影算法的敏感波段筛选图，经过连续投影算法降维处理过后筛选的敏感波段在图中用方块圈出，经 S－G 平滑后的原始光谱数据，RMSE 最小值为 17.966 8，共 350 nm，440 nm，480 nm，620 nm，770 nm，840 nm，920 nm，970 nm，1 670 nm 和 1 880 nm 等 17 个波段入选；经一阶微分后的光谱数据，RMSE 最小值为 26.194 5，共 420 nm，470 nm，620 nm，1 510 nm 和 1 530 nm 等 9 个波段入选；经二阶微分后的光谱数据，RMSE 最小值为 21.272 4，共 360 nm，420 nm，830 nm，880 nm 和 1 020 nm 等 16 个波段入选；经连续统去除后的光谱数据，RMSE 最小值为 22.853 3，共 410～430 nm，540 nm 和 570 nm 等 15 个波段入选。与相关分析法相比大大减少了模型的输入变量，可有效提高模型的计算速度。



（a）原始光谱 SPA 误差分析图　（b）原始光谱 SPA 入选波段图　（c）一阶微分 SPA 误差分析图　（d）一阶微分 SPA 入选波段图

（e）二阶微分 SPA 误差分析图　（f）二阶微分 SPA 入选波段图　（g）连续统去除 SPA 误差分析图　（h）连续统去除 SPA 入选波段图

图 4　基于连续投影算法（SPA）的敏感波段筛选图

Fig. 4　Sensitive band selection based on successive projections algorithm

2.3　建模结果分析

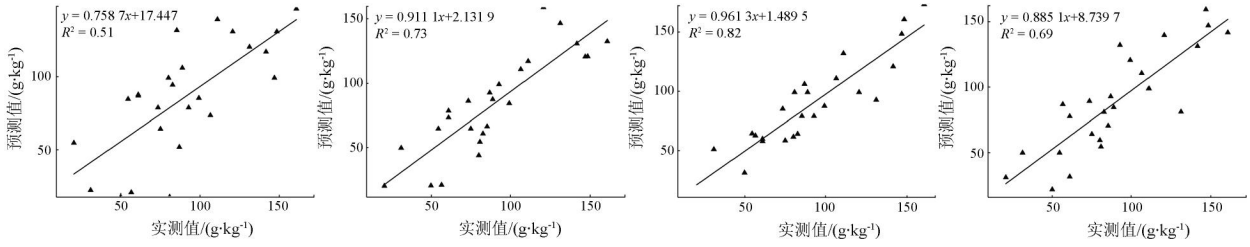
表 2 为土壤碳酸钙含量敏感波段 RFR 模型，研究表明，经数学变换的光谱为变量建立的模型精度明显高于原始光谱，数学变换有效提高了土壤碳酸钙含量与光谱反射率之间的相关性，验证集 R^2 均大于 0.58，其中二阶微分的估算精度最高，验证集 R^2 为 0.82，*RPD* 为 2.37，表明模型较为适用；从敏感波段选取方法看，除基于 CR 的建模集精度相关分析法较高于连续投影算法外，其余数学变换方法无论建模集验证集基于连续投影算法建模精度均高于相关分析法，由此看出，连续投影算法不仅变量少效率高，并且在波段筛选方面较为适用；就模型本身而言，随机森林回归整体的预测精度较好，验证了利用随机森林回归估算土壤碳酸钙含量的可行性。就模型的稳定性而言，建模集整体精度均高于验证集，模型缺乏一定的稳定性，可能原因为：①土壤碳酸钙样品普遍存在标准差较大、变异性较大的现象，碳酸钙的含量较为分散，在对碳酸钙含量进行估算研究中数据普遍存在以上问题；②获取敏感波段的方式不同。例如，洪长乔等[17]除了利用相关分析法选取的波段，还添加了前人研究结果较好的几个数据集作为自己研究的建模集之一，并分别建模进行比

表 2　土壤碳酸钙含量的敏感波段的 RFR 模型

Tab. 2　RSR models for soil calcium carbonate content based on sensitive bands

模型	波段选取方法	建模波段数	建模集		验证集		
			R^2	RMSE	R^2	RMSEp	RPD
RFR－R	CA	54	0.59	20.43	0.41	25.22	1.43
	SPA	17	0.60	20.03	0.51	22.56	1.49
RFR－R′	CA	82	0.76	19.19	0.68	19.94	1.88
	SPA	9	0.79	17.56	0.73	16.56	2.32
RFR－R″	CA	54	0.74	12.36	0.70	13.45	2.21
	SPA	16	0.89	11.25	0.82	12.79	2.37
RFR－CR	CA	44	0.71	14.28	0.58	18.62	1.83
	SPA	15	0.68	17.11	0.69	19.64	1.88

较; Gomez 等[19]在研究拉比因河谷土壤碳酸钙含量时,利用可变重要性(VIP 分数)和 B 系数的值来确定光谱带,由此可看出敏感波段的选取方式是多样的,选择方式的不同对模型精度与稳定性会产生一定的影响。图 5 为基于 SPA 的 4 种数学变换 RFR 模型实测值与预测值拟合图,可以明显看出二阶微分的模型拟合效果较好。



(a) 原始光谱拟合曲线图　(b) 一阶微分拟合曲线图　(c) 二阶微分拟合曲线图　(d) 连续统去除拟合曲线图

图 5　基于 SPA 的四种数学变换 RFR 模型实测值与预测值验证集拟合图

Fig. 5　Fitting graph of measured value and predicted value of four mathematical transformations RFR model based on SPA

3　结论

研究采用随机森林回归的方法建立了陕西省北部黄绵土高光谱碳酸钙估算模型,旨在探讨不同光谱处理方法与不同敏感波段筛选方法模型构建的优劣以及利用随机森林回归估算碳酸钙含量的可行性,主要取得了以下 3 点结论:

1)碳酸钙较明显的吸收特征位于 1 610 nm, 2 220 nm,2 340 nm,2 270 nm 等处,且土壤碳酸钙含量与全波段光谱呈现较明显的正相关关系。

2)基于连续投影算法的模型构建精度除建模集 CR 变换外精度均高于相关分析法,表明连续投影算法在波段筛选中较为适用。

3)基于连续投影算法与二阶微分变换的随机森林回归建模精度最高,建模集 R^2 为 0.89,验证集 R^2 值为 0.82,且 RPD 值大于 2,模型可信度较高,验证了利用随机森林回归估算土壤碳酸钙含量的可行性。

参考文献(References):

[1] 赵其国,史学正. 土壤资源概论[M]. 北京:科学出版社科学,2017:15 -20.
Zhao Q G,Shi X Z. Introduction to soil resources[M]. Beijing:Science Press,2017:15 -20.

[2] 陈留美,张甘霖. 滨海沉积物发育的水稻土时间序列母质均一性判定与特性演变[J]. 土壤学报,2009,46(5):753 -763.
Chen L M,Zhang G L. Determination of parent material homogeneity and evolution of characteristics of paddy soils developed in coastal sediments[J]. Soil Journal,2009,46(5):753 -763.

[3] 张 宁,何兴东,邬 畏. 腾格里沙漠 3 种土壤有机质和碳酸钙特征[J]. 生态学报,2009,29(8):4095 -4101.
Zhang N,He X D,Wu W. Characteristics of organic matter and calcium carbonate in three soils of Tengger Desert[J]. Acta Ecologica,2009,29(8):4095 -4101.

[4] 童庆禧,张 兵,张立福. 中国高光谱遥感的前沿进展[J]. 遥感学报,2016,20(5):689 - 707.
Tong Q X,Zhang B,Zhang L F. Current progress of hyperspectral remote sensing in China[J]. Journal of Remote Sensing,2016,20(5):689 - 707.

[5] Gaffey S J. Spectral reflectance of carbonate minerals in visible and near infrared(0.35 ~ 2.55 μm):A nhydrous carbonate minerals[J]. Journal of Geophysical Research,1987,92:1429—1440.

[6] 张婷婷. 基于 PLS 模型的农业土壤成分高光谱遥感反演研究[D]. 长春:吉林大学,2010.
Zhang T T. Partial least squares modeling of hyperspectral remote sensing for mapping agricultural soil properties[D]. Changchun:Jilin University,2010.

[7] Stenberg B,Viscarra Rossel R A,Mouazen A M,et al. Visible and near infrared spectroscopy in soil science[J]. Advances in A gronomy,2010,107:163—215.

[8] Sexton J,Laake P. Standard errors for bagged and random forest estimators[J]. Computational Statistics & Data Analysis,2009,53(3):801 -811.

[9] 宋建国,高强山,李 哲. 随机森林回归在地震储层预测中的应用[J]. 石油地球物理勘探,2016,51(6):1203 -1211.
Song J G,Gao Q S,Li Z. Application of stochastic forest regression in seismic reservoir prediction[J]. Petroleum Geophysical Exploration,2016,51(6):1203 -1211.

[10] 秦倩如. 基于高光谱的陕西省土壤游离铁与碳酸钙含量预测研究[D]. 杨凌:西北农林科技大学,2019.
Qin Q R. Prediction of soil free iron and calcium carbonate content based on hyperspectral data in Shaanxi Province[D]. Yang Ling:Northwest Agricultural and Forestry University of Science and Technology,2019.

[11] 单海滨,蒋平安,颜 安,等. 基于高光谱的北疆绿洲农田灰漠土有机质反演[J]. 农业资源与环境学报,2018,35(3):276 -282.
Shan H B,Jiang P A,Yan A,et al. Hyperspectral retrieval of organic matter in grey desert soil of oasis farmland in Northern Xinjiang[J]. Journal of Agricultural Resources and Environment,2018,35(3):276 -282.

[12] Araújo M C U, Saldanha T C B, Galvão R K H, et al. The successive projections algorithm for variable selection in spectroscopic multicomponent analysis[J]. Chemometrics and Intelligent Laboratory Systems, 2001, 57(2): 65 – 73.

[13] 杨红云,周 琼,杨 珺,等. 基于高光谱的水稻叶片氮素营养诊断研究[J]. 浙江农业学报,2019,31(10):1575 – 1582.
Yang H Y, Zhou Q, Yang J, et al. Hyperspectral diagnosis of nitrogen nutrition in rice leaves[J]. Zhejiang Journal of Agriculture, 2019, 31(10): 1575 – 1582.

[14] 李勋兰,易时来,何绍兰,等. 高光谱成像技术的柚类品种鉴别研究[J]. 光谱学与光谱分析,2015,35(9):2639 – 2643.
Li X L, Yi S L, He S L, et al. Identification of pummelo cultivars based on hyperspectral imaging technology[J]. Spectroscopy and Spectral Analysis, 2015, 35(9): 2639 – 2643.

[15] 秦倩如,齐雁冰,吴 娟,等. 基于高光谱的土壤游离铁随机森林模型估算研究[J]. 土壤通报,2018,49(6):1286 – 1293.
Qin Q R, Qi Y B, Wu J, et al. Estimation of soil free iron random forest model based on hyperspectral data[J]. Soil Bulletin, 2018, 49(6): 1286 – 1293.

[16] 彭 杰,周 清,张杨珠,等. 有机质对土壤光谱特性的影响研究[J]. 土壤学报,2013,50(3):517 – 524.
Peng J, Zhou Q, Zhang Y Z, et al. Effect of soil organic m atter on spectral characteristics of soil[J]. Acta Pedologica Sinica, 2013, 50(3): 517 – 524.

[17] 洪长桥,郑 光,陈昌春. 苏北滨海土壤碳酸钙含量反射光谱估算研究[J]. 土壤学报,2016,53(5):1121 – 1129.
Hong C Q, Zheng G, Chen C C. Study on reflectance spectrum estimation of calcium carbonate content in coastal soil of Northern Jiangsu[J]. Journal of Soil Science, 2016, 53(5): 1121 – 1129.

[18] 张 东,张 飞,等. 分数阶微分在盐渍土高光谱数据预处理中的应用[J]. 农业工程学报,2014,30(24):151 – 160.
Zhang D, Zhang F, et al. Application of fractional differential in preprocessing hyperspec – tral data of saline soil[J]. Transactions of the Chinese Society of Agricultural Engineering, 2014, 30(24): 151 – 160.

[19] Gomez C, Philippe L, Guillaume C. Continuum removal versus PLSR method for clay and calcium carbonate content estimation from laboratory and airborne hyperspectral measurements[J]. Geoderma, 2008, 148(2): 141 – 148.

The estimation of soil calcium carbonate content based on Hyperspectral data

WU Qian, JIANG Qigang, SHI Pengfei, ZHANG Lili

(*College of Geo – exploration Science and Technology*, *Jilin University*, *Changchun* 130026, *China*)

Abstract: Carbonate content in soil is an important basis for soil classification and fertility evaluation. Based on an analysis of calcium carbonate content, the authors chose 78 soil samples from Loess Plateau of Shaanxi Province as the research objects. The visible near infrared hyperspectral reflectance (350 ~ 2 500 nm) data of soil samples were obtained by hyperspectral imager. Three mathematical transformations, i. e., first – order differentiation, second – order differentiation and continuum removal, were carried out on the original spectral curve, and correlation analysis was used. The method and the continuous projection algorithm were used to select the sensitive band respectively, and the Stochastic Forest regression was used to establish the estimation model of soil calcium carbonate. According to the results obtained, the spectral curve characteristics of Huangmian soil are almost the same, there are obvious absorption characteristics at 1 440 nm, 1 900 nm, 2 200 nm and so on, and the calcium carbonate content and spectral reflectance show a positive correlation trend; the accuracy of random forest estimation model based on the second – order differential and continuous projection algorithm is the highest, the validation set R^2 is 0.82, and the PRD value is 2.37.

Keywords: calcium carbonate; Huangmian soil; hyperspectral; continuous projection algorithm; random forest regression

(责任编辑: 李 瑜)