

SVM 和 BP 检测滨海湿地土壤有机质

张 森¹, 卢 霞^{1*}, 聂格格², 李昱蓉¹, 邵亚婷¹, 田燕芹¹, 范礼强¹, 张钰娟¹

1. 江苏海洋大学测绘与海洋信息学院, 江苏 连云港 222005
2. 河南财经政法大学资源与环境学院, 河南 郑州 450046

摘 要 近年来,虽然随着高光谱技术的出现可以快速获取土壤中的养分含量,但不同的土壤类型对估算的精度会有很大的差异。滨海湿地土壤类型受海洋环境影响较大,其高光谱反射率与内陆土壤类型的表现会有所不同,也就造成了同样的估算模型在反演滨海湿地土壤的养分含量时,反演精度的降低,随着近年来海洋资源的开发与滨海湿地生态恢复工作的不断推进,探索一种合适的估算模型来快速准确的获取土壤中的养分含量变得更加紧迫。该研究旨在验证利用可见-近红外高光谱反射率构建非线性模型来反演滨海湿地土壤类型中有机质(soil organic matter, SOM)含量的可行性。以江苏省盐城大丰麋鹿国家级自然保护区的第三核心区土壤作为研究对象,将土壤样本的光谱反射率进行5点 Savitzky-Golay(S-G)平滑滤波处理,再进行一阶微分 R' 、倒数的一阶微分 $(1/R)'$ 、倒数的二阶微分 $(1/R)''$ 、对数的一阶微分 $(\lg R)'$ 四种微分变换后,应用相关系数和显著性水平($p < 0.01$)提取土壤有机质含量的敏感波段,利用台湾大学林智仁教授开发的MATLAB软件中的LIBSVM工具包构建SVM(support vector machine)支持向量机估算模型,并利用MATLAB2018b软件中自带的BP(back propagation)反向传播神经网络构建估算模型,最后利用决定系数 R^2 和均方根误差RMSE进行模型的预测精度验证。结果表明:原始光谱通过5点S-G平滑滤波、微分变换与相关系数法可以较好的提取出有效波段,其中基于 $(1/R)'$ 光谱变换提取的滨海湿地土壤有机质特征波段为498~501, 1180~1182, 1946, 1947和2323~2326 nm;对比发现SVM的估算精度优于BP神经网络;利用光谱的 $(1/R)'$ 微分形式构建的SVM模型估算滨海湿地土壤SOM含量的精度最高,决定系数 R^2 与RMSE分别为0.93和0.23,并且均通过了 $p < 0.01$ 的显著性检验。因此利用高光谱构建SVM非线性模型来快速估算滨海湿地土壤中的养分含量具有一定的可行性。

关键词 滨海湿地; S-G滤波; SVM支持向量机; BP神经网络

中图分类号: O433.5 **文献标识码:** A **DOI:** 10.3964/j.issn.1000-0593(2020)02-0556-06

引 言

土壤中的有机质是湿地土壤成分中的重要组成部分,是判别湿地中土壤肥力大小的重要指标^[1]。因此快速测定湿地土壤中有机的含量,对于维护湿地生态平衡,可持续发展具有重要的意义。传统检验土壤中养分含量的方法主要是基于化学分析,费时耗力,不能满足快速、低成本的现实需求,而光谱分析正好弥补了传统式化学分析方法的不足,为估算土壤中的各种养分含量提供了有力的技术支持。

目前利用光谱分析来估算土壤中养分含量的研究主要分为三个步骤:(1)将测得的原始光谱数据进行光谱预处理,以消除或减弱原有光谱信息中的噪声,放大有用信息,比较常见的光谱预处理方法有连续投影法(SPA)^[2-3]、Savitzky-Golay卷积平滑算法^[4]、微分算法^[5]、指数变换^[6]等;(2)将预处理后的光谱提取特征波段,常用的提取方法主要有相关系数法^[7]、逐步回归法、遗传算法^[8]等;(3)将特征波段的光谱数据和对应的土壤理化数据进行建模,常用的建模方法主要分为线性与非线性建模,线性建模主要有线性回归^[9]、偏最小二乘法回归^[10]等;非线性建模主要有BP神经网络

收稿日期: 2018-12-08, **修订日期:** 2019-04-22

基金项目: 国家自然科学基金项目(41506106), 海岸带地理环境监测国家测绘地理信息局重点实验室开放基金项目(GE-2017-003), 2019年江苏省研究生实践创新计划项目(SJCX19_0972), 江苏省海洋技术专业建设项目(PPZY2015B116), 江苏高校优势学科建设工程项目资助

作者简介: 张 森, 1995年生, 江苏海洋大学测绘与海洋信息学院硕士研究生 e-mail: zhangsen.henan@foxmail.com

* 通讯联系人 e-mail: luxia1210@163.com

络^[11]、最小二乘支持向量机(LS-SVM)^[12]等。已有部分研究发现利用非线性模型来估算土壤养分含量的精度要高于线性模型,如 Bao 等^[13]在综合分析了不同地貌土壤有机质含量与相应光谱反射率之间的关系的基础上,对 PLS 与 PLS-SVM 两种建模方法预测矿区土壤有机质含量的准确度做了比较,结果表明 PLS-SVM 比 PLS 更能准确的预测。章海亮等将光谱进行平滑、标准归一化、多元散射校正和一阶导数处理后,用偏最小二乘和最小二乘支持向量机建立校正模型,最后比较得出连续投影可见/短波近红外光谱利用最小二乘支持向量机建模是一种精确的土壤有机质和速效钾的测定方法。然而针对滨海湿地土壤类型的非线性高光谱估算研究相对较少^[14]。滨海湿地作为一种陆地与水域之间的生态系统,受海洋环境的影响较大,表现出独特的土壤特性。因此本文以江苏省大丰麋鹿野牧区滨海湿地土壤为研究对象,利用非线性模型 SVM 支持向量机的建模方法并对比 BP 神经网络算法来估算滨海湿地土壤有机质的含量,以此来验证滨海湿地土壤有机质含量的非线性高光谱估算模型的可行性。

1 实验部分

1.1 研究区概况

江苏省大丰麋鹿国家级自然保护区坐落于中国四大湿地(南黄海湿地、青藏高原湿地、东北三江平原湿地、鄱阳湖湿地)之一的南黄海湿地上,核心区面积 4 万亩,是世界上最大的麋鹿自然保护区(32°59′—33°03′N, 120°47′—120°53′E),该区域主要为暖温带大陆季风性气候,海洋性和季风性特征显著^[15];地势平坦,为平原盐渍沼泽,黄河口沉积物成土母质;第三核心区植被茂密,分布着互花草草、碱蓬、芦苇等植被群落,土壤类型主要为潮盐土与草甸滨海盐土亚类,表层土壤含盐量在 0.04%~1.13%之间^[16]。

1.2 样本采集

依据研究区土壤类型和植被群落分布特征,采用规则网格法(1 000 m×1 000 m)划分研究区,采样方案如图 1 所示。每个网格采用对角线采样法采集 0~20 cm 的表层土壤,之后均匀混合,作为该网格的土壤样本,共采集 34 个土壤样本。

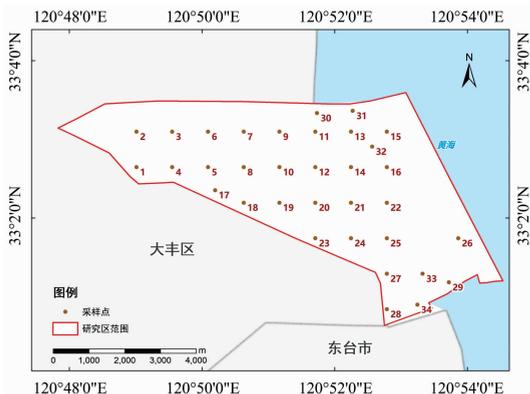


图 1 研究区土壤采样点空间分布

Fig. 1 The spatial distribution of soil sampling sites in study area

将所采集的土壤样品在室温条件下自然风干,去除石块、残根等杂物后研磨、过 80 目筛子,保存,用于待测土壤 SOM 和室内反射光谱。土壤 SOM 含量测定是用水合热重铬酸钾氧化-比色法;土壤 pH 范围为 8.02~8.94 之间,平均值在 8.4,属于碱性土壤,表 1 为 34 个采样点获得的土壤样品 SOM 的统计结果。

表 1 研究区土壤样本 SOM 含量统计

Table 1 SOM statistical results of soil samples in the study area

养分项目	最小值	最大值	标准差	均值	变异系数/%
SOM/(g·kg ⁻¹)	7	45.3	8.3	13.2	63.1

1.3 土壤反射光谱测定

土壤反射光谱利用美国 Spectra Vista 公司生产的 SVC HR-1024I 光谱仪进行测定。波段范围为 350~2 500 nm,其中在 350~1 000 nm 光谱分辨率≤3.0 nm,光谱间隔≤1.5 nm;1 000~1 900 nm 光谱分辨率≤9.5 nm,光谱间隔≤3.6 nm;1 900~2 500 nm 光谱分辨率≤9.5 nm,光谱间隔≤2.5 nm。利用 BRDF 系统建立土壤测试环境:探头垂直向下,视场角 4°,距离土壤样品(样品盒的直径为 9 cm,高度为 2 cm 的圆形玻璃器皿)表面约为 1 m,室内照明光源,为 50 W 的卤素灯,天顶角 45°。测定时,将装有土壤样品的玻璃皿放在黑色阻尼布上,保持土壤表面平整,每个土壤样本测量 5 次,取平均值作为土壤样本反射光谱;测量期间,每隔 15 min 测定白板反射光谱以便校正。

1.4 分析方法

光谱预处理主要采用 Savitzky-Golay(S-G)卷积平滑滤波与微分算法。S-G 卷积平滑滤波原理主要是采用最小二乘拟合系数来建立滤波函数,然后对每一个窗口范围内的波长数据进行多项式最小二乘拟合,其拟合的表达式可表示为

$$\hat{X}_i = a_0 + a_1\lambda_i + a_2\lambda_i^2 \quad (1)$$

其中 \hat{X}_i 是 S-G 平滑算法建立二次拟合后得到的拟合值, a_0 , a_1 和 a_2 是方程式的系数。为了研究不同的微分形式对建模精度的影响,选取一阶微分、倒数的一阶微分、倒数的二阶微分以及对数的一阶微分几种常见的微分变换形式作对比,用 R 代表 S-G 滤波后的光谱, R' 代表光谱的一阶微分形式, $(1/R)'$ 代表光谱的倒数的一阶微分形式, $(1/R)''$ 代表光谱的倒数的二阶微分形式, $[\lg(R)]'$ 代表光谱的对数的一阶微分形式,其计算方法如式(2)~式(5)

$$R'(\lambda_i) = [R(\lambda_{i+1}) - R(\lambda_{i-1})]/2\Delta\lambda \quad (2)$$

$$(1/R)'(\lambda_i) = [(1/R)(\lambda_{i+1}) - (1/R)(\lambda_{i-1})]/(2\Delta\lambda) \quad (3)$$

$$\begin{aligned} (1/R)''(\lambda_i) &= \frac{(1/R)'(\lambda_{i+1}) - (1/R)'(\lambda_{i-1})}{2\Delta\lambda} \\ &= \frac{(1/R)(\lambda_{i+1}) - 2(1/R)(\lambda_i) + (1/R)(\lambda_{i-1})}{\Delta\lambda^2} \end{aligned} \quad (4)$$

$$[\lg(R)]'(\lambda_i) = \{[\lg(R)](\lambda_{i+1}) - [\lg(R)](\lambda_{i-1})\}/2\Delta\lambda \quad (5)$$

式中 λ_i 为各波段的波长, $\Delta\lambda$ 表示波长 λ_{i+1} 到 λ_i 的间隔^[17]。特征波段的选取是利用皮尔逊相关系数法以及相关系数的显著性检验。为研究不同的非线性建模方法在利用高光谱反射

率预测滨海湿地土壤养分含量上的精度差异,选取支持向量机(support vector machine, SVM)与 BP 神经网络两种非线性模型作对比分析。

SVM 建模采用台湾大学林智仁教授开发的 LIBSVM 工具包, BP 神经网络利用 MATLAB2018b 软件中自带的工具包通过编程实现。模型验证主要使用决定系数 R^2 与均方根误差 RMSE, 其中决定系数是相关系数的平方, 是一个能够直观判断拟合优势的指标, 决定系数越接近于 1, 说明实测值与预测值的拟合程度越高, 模型的精度也就越好; 均方根误差是观测值与真值偏差的平方和观测次数比值的平方根, 它对模型的建模能力和预测能力都能做出非常有效的评价, RMSE 的值越小, 模型的反演能力越强。

2 结果与讨论

2.1 滨海湿地土壤光谱特性

由于原始光谱在 350~400 nm 之间存在很多噪声, 因此选取 400~2 400 nm 波段做分析。在 MATLAB 2014b 软件中利用 Savitzky-Golay(S-G)滤波器将滨海湿地土壤的原始光谱进行 5 点平滑滤波处理, 处理后的光谱反射率曲线如图 2 所示, 可以看出经过自然风干的 34 个土壤样本测得的光谱曲线走势具有很大的相似性, 但是由于每个土壤样本中的有机质含量等土壤理化性质的不同, 所测得的土壤样本光谱反射率在波峰、波谷、反射率强弱上还是有所不同, 这与章海亮等的研究结果相同。总体波段范围内的反射率介于 0.1~0.7 之间; 同时可以清楚的发现在 1 400 和 1 900 nm 两个波段附近有两个明显的吸收谷, 在 700 和 1 000 nm 两个波段附近有两个比较弱的吸收谷, 这主要是由土壤中的水分子振动的倍频与合频所产生的; 在 1 950~2 400 nm 波段内光谱曲线呈现波浪式状态, 主要是由于土壤样本中的少量水分与空气中的水分吸收产生的, 这与研究土壤光谱反射率特征的相关报道一致^[2-3]。总体上看土壤反射率随着波长的增加而不断增大, 其中在 400~600 nm 波段范围上升速度比较明显, 600~800 nm 波段之间呈现中等缓慢上升, 800 nm 以后波谱反射率上升比较平缓。

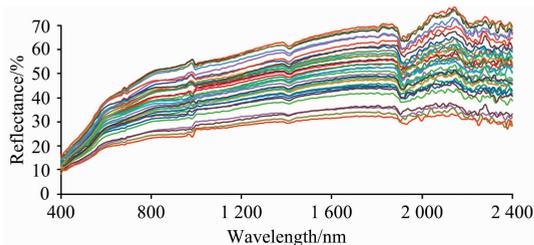


图 2 S-G 滤波后土壤样本反射率曲线

Fig. 2 Soil sample reflectance curve after S-G filtering

2.2 土壤 SOM 含量的特征波段提取

将 S-G 滤波后的光谱反射率进行四种微分变换, 并与土壤实测的 34 个土壤样本组的 SOM 含量做相关性分析与显著性分析, 如图 3 所示, 其中(a)图为一阶微分变换后的相关系数, (b)图为倒数的一阶微分后的相关系数, (c)图为倒数的

二阶微分后的相关系数, (d)图为对数的一阶微分后的相关系数。从图 3 可以发现四种光谱变换形式不同程度上改变了原始光谱与土壤有机质含量的相关系数, 并且将有关土壤有机质的敏感波段处的反射率进一步放大, 同时可以发现在 700, 1 000, 1 400 和 1 900 nm 几处光谱吸收谷附近的相关性都较低, 这说明土壤受水分影响会降低光谱反射率与土壤有机质含量的相关系数。

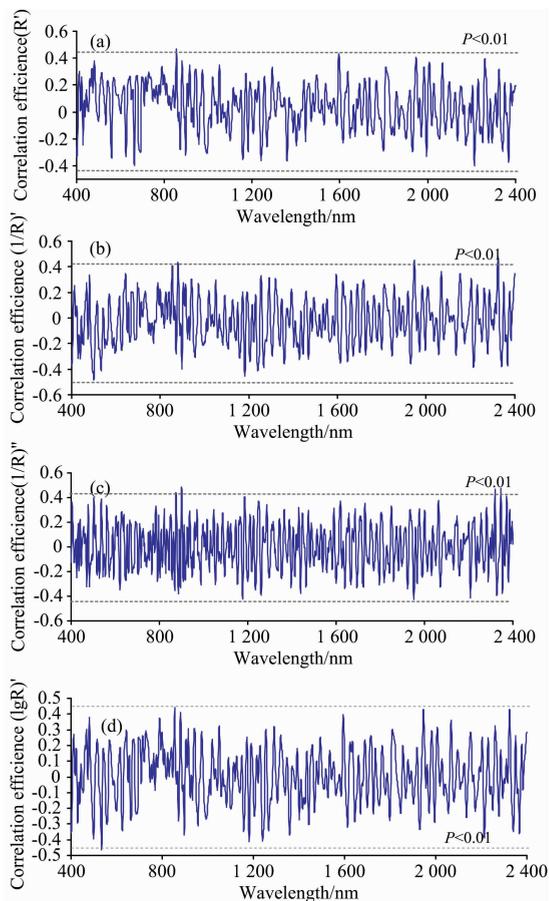


图 3 变换后的光谱与 SOM 含量的相关性

Fig. 3 Correlation between transformed spectra and SOM content

挑选出置信水平 $p < 0.01$ 的波长作为特征波段, 筛选结果如表 2 所示。由表 2 可知: 每一种微分变换后所提取出的特征波段的数量不同, 并且每一种微分变换后与对应的土壤 SOM 养分含量相关性也不同。相关性最高的是 $(1/R)'$ 光谱变换形式, 所提取到的特征波段数有 13 个, 分别为 498~501, 1 180~1 182, 1 946, 1 947 和 2 323~2 326 nm, 其中在 2 324 nm 波段附近呈现正相关性, 在 500 nm 波段附近主要呈现负相关性。

但是提取到的敏感波段与章海亮等利用连续投影法所提取的水稻土、砖红土和黄土有机质敏感波段 362, 392, 422, 437, 537, 652, 702 和 1 062 nm 有所不同。差异的原因可能是由于研究区土壤类型的不同, 前人研究结果发现大丰麋鹿国家级自然保护区第三核心区的土壤类型主要为潮滩盐土和草甸海滨盐土, 这两种土壤类型通常含盐量在 0.8%~2.0%

之间,最高甚至可达 4% 的含盐量^[18],而土壤盐分含量会影响土壤的反射率,也会影响到土壤中有机质的光谱信息^[6]。

表 2 土壤 SOM 含量的特征波段筛选

Table 2 Sensitive band deletion of soil SOM content

光谱变换	最大相关		最小相关		变量个数	敏感波段/nm ($p < 0.01$)
	对应波段/nm	相关系数	对应波段/nm	相关系数		
R'	855	0.468**	2 213	-0.402*	2	855, 854
$(1/R)'$	2 324	0.479**	500	-0.484**	13	498~501, 1 180~1 182, 1 946, 1 947, 2 323~2 326
$(\lg R)'$	855	0.441**	534	-0.465**	3	533, 534, 855
$(1/R)''$	900	0.477**	1952	-0.425*	7	874, 899, 900, 901, 2 319, 2 320, 2 346

Note: * $p < 0.05$, ** $p < 0.01$

2.3 模型的构建与精度验证

2.3.1 基于 SVM 支持向量机的土壤有机质含量高光谱估算模型

训练集与测试集的设置是随机选取 34 个土壤样本中 24 个光谱变换后筛选出来的特征波段数据和对应的土壤 SOM 含量值作为训练集的输入与输出,剩下的 10 个样本作为测试集的输入与输出。并利用 MATLAB 中的 mapminmax 函数将训练集与测试集的数据进行归一化处理,使其映射到 [0, 1] 区间内。在 SVM 的创建与训练中,“-t”核函数类型选择为 RBF 核函数,利用网格搜索交叉验证的方法遍历 c 与 g 的值来获取最佳 c 与 g 参数,“-s”即 SVM 类型选择为 e-SVR 类型,“-p”设置 e-SVR 类型中损失函数 p 的值为 0.01。最后利用 svmpredict 函数和已经训练好的模型预测出剩余 10 个样本的有效值,并将预测值进行反归一化处理以便更好的还原真实值。最后构建的模型验证精度如表 3 所示。由表 3 可知:土壤样本反射率 $(1/R)'$ 变换形式估算土壤有机质含量的精度最高,预测决定系数 R^2 为 0.93,预测均方根误差为 0.23。

表 3 SVM 建模方法得到的土壤 SOM 含量精度

Table 3 Soil SOM content obtained by SVM modeling

变换形式	建模精度		验证精度	
	R^2	RMSE	R^2	RMSE
R'	0.74**	0.39	0.72**	0.7
$(1/R)'$	0.68**	0.28	0.93**	0.23
$(\lg R)'$	0.89**	0.18	0.7**	0.34
$(1/R)''$	0.84**	0.32	0.84**	0.24

Note: * $p < 0.05$, ** $p < 0.01$

2.3.2 基于 BP 神经网络的土壤有机质含量高光谱估算模型

BP 神经网络的建模形式与 SVM 类似,都需要设置训练集与测试集,为方便观察对比两种建模精度,选取与 SVM

建模一样的测试集与训练集;同时与 SVM 建模相同,将训练集与测试集都进行归一化处理将其映射到 [0.1] 区间内;在创建神经网络时,训练方式选择梯度下降方法,迭代次数设置为 1 000 次,训练目标设置为 $1e-30$ 即训练的均方根误差小于 $1e-30$,神经元设置为 10 个,学习率设置为 0.01;之后的仿真测试与反归一化和 SVM 建模相同。最后构建的模型精度如表 4 所示。由表 4 可知:土壤样本反射率 $(1/R)'$ 形式估算土壤有机质含量的精度较高,其中预测决定系数 R^2 为 0.87,预测均方根误差为 0.33。

表 4 用 BP 神经网络建模方法得到的土壤 SOM 含量精度

Table 4 Soil SOM content obtained by BP

变换形式	建模精度		验证精度	
	R^2	RMSE	R^2	RMSE
R'	0.89**	0.26	0.70**	0.24
$(1/R)'$	0.83**	0.09	0.87**	0.33
$(\lg R)'$	0.95**	0.02	0.63**	0.44
$(1/R)''$	0.66*	0.06	0.77**	0.18

Note: * $p < 0.05$, ** $p < 0.01$

2.3.3 SVM 和 BP 检测土壤有机质含量的精度对比

图 4 为 SVM 与 BP 神经网络两种非线性建模方法在估算滨海湿地土壤有机质含量的精度对比,横坐标代表四种不同的光谱变换形式,纵坐标左侧代表决定系数 R^2 的值,纵坐标右侧代表 RMSE 的值。

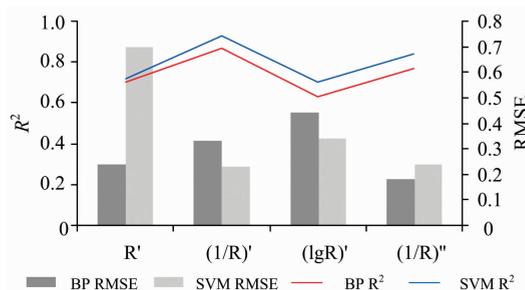


图 4 土壤 SOM 含量的建模精度对比

Fig. 4 Comparison of modeling accuracy of soil SOM content

由图 4 可知:基于决定系数 R^2 和均方根误差 RMSE 评价指标,SVM 支持向量机估算滨海湿地土壤有机质含量精度明显优于 BP 神经网络。为了更加直观的看出 SVM 模型的预测效果,用光谱变换 $(1/R)'$ 构建 SVM 模型所预测的滨海湿地土壤有机质含量与实测的含量进行对比,如图 5 所示,横坐标为实测值,纵坐标为预测值。由图 5 可以看出 SVM 预测土壤有机质含量精度较好,均匀分布于 $y=x$ 直线附近。

由以上分析结果表明:通过 SVM 支持向量机检测土壤有机质的精度要优于 BP 神经网络,这与蒋璐璐等^[19]通过最小二乘支持向量机与 BP 神经网络构建的浙江省衢州红壤和海宁青紫泥中氮磷钾含量的估算模型结果一致。通过光谱的 $(1/R)'$ 变换形式构建的 SVM 估算模型在预测滨海湿地土壤有机质的精度最高,其中预测决定系数 R^2 为 0.93,均方根

误差 RMSE 为 0.23, 这比刘雪梅^[20]通过一阶微分结合最小二乘回归方法构建的江西省遂川县枚江乡土壤有机质含量的预测精度(R^2 为 0.825 5)要高。这说明基于高光谱反射率的微分变换形式所构建的 SVM 模型在预测滨海湿地中土壤有机质含量具有一定的可行性, 但该模型是否可以预测其他地区的滨海湿地土壤有机质的含量还需进一步的验证。

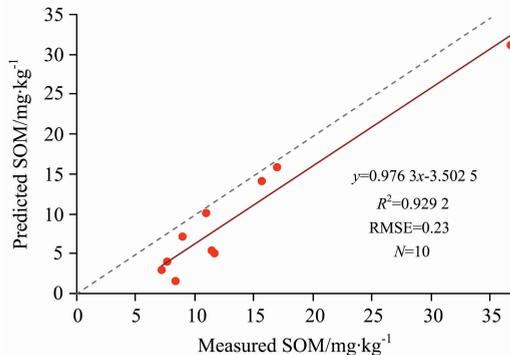


图 5 土壤 SOM 含量预测

Fig. 5 The prediction of soil SOM content

4 结 论

滨海湿地采集的 34 个土壤样本在室内测得的原始光谱进行 S-G 滤波处理与 R' , $(1/R)'$, $(1/R)''$, $(\lg R)'$ 四种微分变换, 利用相关系数法提取土壤 SOM 含量在高光谱遥感中的特征波段, 并通过 SVM 支持向量机和 BP 神经网络估算滨海湿地土壤中 SOM 含量。得出以下结论:

(1) SVM 支持向量机检测滨海湿地土壤有机质含量的精度明显优于 BP 神经网络模型, 利用可见-近红外高光谱构建的 SVM 非线性模型来快速估算滨海湿地土壤中的养分含量具有一定的可行性。

(2) 土壤有机质与光谱反射率倒数的一阶微分存在较高的相关性, 显著相关($p < 0.01$)的波段数为 13 个, 敏感波段为 498~501, 1 180~1 182, 1 946, 1 947 和 2 323~2 326 nm。

(3) 基于 SVM 构建土壤有机质的估算模型精度最高, 预测决定系数 R^2 为 0.93, 均方根误差 RMSE 为 0.23。

(4) 光谱数据的预处理与变换形式以及模型的构建方式和样本的数量都是影响最终预测精度的重要因子, 本实验由于样本数量有限并不能很好的展示样本大小对模型估算精度的影响程度, 因此未来应深入探讨土壤样本数量对估算模型的精度和均方根误差的影响程度。

References

- [1] Gabriel Y K Moinet, John E Hunt, Miko U F Kirschbaum, et al. *Soil Biology and Biochemistry*, 2018, 116: 333.
- [2] Vohland M, Emmerling C. *Soil Science*, 2011, 62(4): 598.
- [3] Wu Di, Shi Hui, Wang Songjing, et al. *Analytica Chimica Acta*, 2012, 726(0): 57.
- [4] Radim Vašát, Radka Kodešová, Aleš Klement, et al. *Geoderma*, 2017, 298: 46.
- [5] JIANG Ye-lin, WANG Rang-hui, LI Yan, et al(蒋焯林, 王让会, 李 焱, 等). *Chinese Journal of Eco-Agriculture(中国生态农业学报)*, 2016, 24(11): 1555.
- [6] GAO Deng-zhou, ZENG Cong-sheng, ZHANG Wen-long, et al(高灯州, 曾从盛, 章文龙, 等). *Chinese Journal of Ecology(生态学杂志)*, 2016, 35(4): 952.
- [7] He Yong, Xiao Shupe, Nie Pengcheng, et al. *Sensors*, 2017, 17(9): 2045.
- [8] ZOU Xiao-bo, ZHAO Jie-wen(邹小波, 赵杰文). *Acta Optica Sinica(光学学报)*, 2007, 27(7): 1316.
- [9] GAO Deng-zhou, CHEN Gui-xiang, ZHANG Wen-long, et al(高灯州, 陈桂香, 章文龙, 等). *Research and Exploration in Laboratory(实验室研究与探索)*, 2014, 33(6): 19.
- [10] Gerighausen H, Menz G, Kaufmann H, et al. *Applied & Environmental Soil Science*, 2012, 143: 487.
- [11] Tian Yongchao, Zhang Juanjuan, YAO Xiao, et al. *Geoderma*, 2013, 202-203: 161.
- [12] Kennedy Were, Dieu Tien Bui, Øystein B Dick, et al. *Ecological indicators*, 2015, 52: 394.
- [13] Bao Nisha, Wu Lixin, Ye Baoying, et al. *Geoderma*, 2017, 288: 47.
- [14] Wang Xiaoping, Zhang Fei, Kung Hsiang-te, et al. *Remote Sensing of Environment*, 2018, 218(1): 104.
- [15] LU Xia, LIN Ya-li, WU Ya-nan, et al(卢 霞, 林雅丽, 吴亚楠, 等). *Transactions of Oceanology and Limnology(海洋湖沼通报)*, 2018, 4: 74.
- [16] LIU Jin-gen, XUE Jian-hui, WANG Lei, et al(刘金根, 薛建辉, 王 磊, 等). *Chinese Journal of Ecology(生态学杂志)*, 2011, 30(8): 1793.
- [17] QIAO Juan-feng, XIONG Hei-gang, WANG Xiao-ping, et al(乔娟峰, 熊黑纲, 王小平, 等). *Agricultural Research in the Arid Areas(干旱地区农业研究)*, 2018, 36(5): 207.
- [18] DING Ning-ning, WANG Bao-song, LIANG Zhen-hai, et al(丁宁宁, 王保松, 梁珍海, 等). *Soils(土壤)*, 2011, 43(3): 487.
- [19] JIANG Lu-lu, ZHANG Yu, WANG Yan-yan, et al(蒋璐璐, 张 瑜, 王艳艳, 等). *Journal of Zhejiang University(浙江大学学报)*,

2010, 36(4): 445.

[20] LIU Xue-mei(刘雪梅). Journal of Chinese Agricultural Mechanization(中国农机化学报), 2013, 34(2): 202.

Estimation of Soil Organic Matter in Coastal Wetlands by SVM and BP Based on Hyperspectral Remote Sensing

ZHANG Sen¹, LU Xia^{1*}, NIE Ge-ge², LI Yu-rong¹, SHAO Ya-ting¹, TIAN Yan-qin¹, FAN Li-qiang¹, ZHANG Yu-juan¹

1. School of Geomatics and Marine Information, Jiangsu Ocean University, Lianyungang 222005, China

2. School of Resources and Environment, Henan University of Economics and Law, Zhengzhou 450046, China

Abstract In recent years, although the nutrient content in the soil can be quickly obtained with the emergence of hyperspectral technology, but different soil types have great differences in the accuracy of estimation. The soil type of coastal wetland is greatly affected by the marine environment, and its hyperspectral reflectance and inland soil type will be different. This will reduce the precision in the same estimation model when inverting the nutrient content of coastal wetland soil types. With the development of marine resources and the ecological restoration of coastal wetlands in recent years, it is urgent to explore a suitable estimation model to quickly and accurately obtain nutrient content in soil. This study aimed to verify the use of visible-near infrared hyperspectral reflectivity to construct a nonlinear model so as to invert the feasibility of organic matter (SOM) in coastal wetland soils. The topsoil in the third core area of Dafeng Elk National Nature Reserve in Yancheng, Jiangsu province was taken as the investigated object. The sensitive bands corresponding to Soil Organic Matter (SOM) content were retrieved based on correlation coefficient after 5 point S-G filtering and four differential transformations of R' , $(1/R)'$, $(1/R)''$, $(\lg R)'$ by spectral reflectance of soil samples. The estimation models of SOM by Support Vector Machine (SVM) and BP neural network were determined, and the prediction accuracy of the model was verified by using the decision coefficient R^2 and the root mean square error RMSE. The research results indicated that the effective bands can be identified by S-G filtering, differential transformation and correlation coefficient method based on the original spectra of soil samples. The characteristic bands of SOM based on transformations $(1/R)'$ were 498~501, 1 180~1 182, 1 946, 1 947, 2 323~2 326 nm. Estimation accuracy of SVM was better than that of BP neural network for SOM in Yancheng coastal wetland. The estimation model of SOM by SVM based on $(1/R)'$ spectra had the highest precision, with the determination coefficients (R^2) and root mean square error (RMSE) of 0.93 and 0.23. Therefore, it is suitable to use hyperspectral remote sensing to quickly estimate the nutrient contents of topsoil in coastal wetland.

Keywords Coastal wetland; S-G filtering; Support vector machine; BP neural network

(Received Dec. 8, 2018; accepted Apr. 22, 2019)

* Corresponding author