

DOI: 10.13758/j.cnki.tr.2023.04.022

张清文, 吴风华, 宋敬茹, 等. 基于光谱变换的滨海湿地土壤全氮含量建模预测. 土壤, 2023, 55(4): 880–886.

基于光谱变换的滨海湿地土壤全氮含量建模预测^①

张清文¹, 吴风华¹, 宋敬茹¹, 汪金花¹, 张永彬¹, 刘明月^{1,2,3,4}, 李孟倩¹, 李春景⁵, 郝玉峰⁶, 满卫东^{1,2,3,4*}

(1 华北理工大学矿业工程学院, 河北唐山 063210; 2 唐山市资源与环境遥感重点实验室, 河北唐山 063210; 3 河北省矿区生态修复产业技术研究院, 河北唐山 063210; 4 矿产资源绿色开发与生态修复协同创新中心, 河北唐山 063210; 5 延边大学地理与海洋科学学院, 吉林延吉 133002; 6 河北省地质矿产勘查开发局第八地质大队, 河北秦皇岛 066001)

摘要: 基于 133 个滨海湿地土样的全氮(TN)含量和光谱反射率(R)及其对数($\lg R$)、对数的一阶微分($(\lg R)'$)、倒数($1/R$)、倒数的一阶微分($(1/R)'$)、一阶微分(R')、平方根(\sqrt{R})、一阶微分的倒数($1/(R)'$)变换, 采用偏最小二乘回归(PLSR)、随机森林回归(RFR)和支持向量机回归(SVR)3 种算法分别建立土壤 TN 含量估测模型。结果表明: ①土壤 TN 含量与光谱变换形式相关性由高到低为: $(1/R)' > R' > (\lg R)' > 1/R > \lg R > 1/(R)' > \sqrt{R} > R$, 经光谱变换, 土壤 TN 含量与变换光谱的相关性均高于 R , 其中与 $(1/R)'$ 的 Pearson 相关系数最大为 0.746。②PLSR 和 SVR 基于 R' 、 $(1/R)'$ 、 $(\lg R)'$ 和 $1/(R)'$ 变换构建的模型、RFR 方法构建的所有模型 R^2 均大于 0.732, 均可用于滨海湿地土壤 TN 含量的估算。③基于 $1/(R)'$ 建立的 SVR 模型预测精度最高, 其 R^2 为 0.987, RMSE 为 0.057 g/kg, MAE 为 0.050 g/kg, 是预测滨海湿地土壤 TN 含量的最优模型, 可为准确获取滨海湿地土壤 TN 含量提供稳定方法。

关键词: 光谱变换; 土壤全氮含量; 偏最小二乘回归; 随机森林回归; 支持向量机回归

中图分类号: S151.9 **文献标志码:** A

Estimating of Soil Total Nitrogen Content in Coastal Wetland Based on Spectral Transformation

ZHANG Qingwen¹, WU Fenghua¹, SONG Jingru¹, WANG Jinhua¹, ZHANG Yongbin¹, LIU Mingyue^{1,2,3,4}, LI Mengqian¹, LI Chunjing⁵, HAO Yufeng⁶, MAN Weidong^{1,2,3,4*}

(1 College of Mining Engineering, North China University of Science and Technology, Tangshan, Hebei 063210, China; 2 Tangshan Key Laboratory of Resources and Environmental Remote Sensing, Tangshan, Hebei 063210, China; 3 Hebei Industrial Technology Institute of Mine Ecological Remediation, Tangshan, Hebei 063210, China; 4 Collaborative Innovation Center of Green Development and Ecological Restoration of Mineral Resources, Tangshan, Hebei 063210, China; 5 College of Geography and Ocean Sciences, Yanbian University, Yanji, Jilin 133002, China; 6 The 8th Geological Brigade of Hebei Bureau of Geology and Mineral Resource Exploration, Qinhuangdao, Hebei 066001, China)

Abstract: Based on total nitrogen (TN) contents, spectral reflectance (R) of and their logarithm ($\lg R$), logarithm first derivative ($(\lg R)'$), reciprocal ($1/R$), reciprocal first derivative ($(1/R)'$), first derivative (R'), square root (\sqrt{R}) and reciprocal first derivative ($1/(R)'$) transformations of 133 coastal wetland soil samples, the predicating models of soil TN contents were established by partial least squares regression (PLSR), random forest regression (RFR) and support vector regression (SVR). The results showed that: Correlations between soil TN contents and spectral forms from high to low were: $(1/R)' > R' > (\lg R)' > 1/R > \lg R > 1/(R)' > \sqrt{R} > R$. Correlations between soil TN contents and spectral transformations were higher than those of R , and Pearson correlation coefficient of $(1/R)'$ was highest (0.746). R^2 of all models established by PLSR and SVR based on R' , $(1/R)'$, $(\lg R)'$ and $1/(R)'$ transformations and RFR method were greater than 0.732, indicating their applicable for soil TN content estimation, and SVR model based on $1/(R)'$ had the highest accuracy, with R^2 of 0.987, RMSE of 0.057 g/kg and MAE of 0.050 g/kg, which was the optimal model for accurately predicting TN content in coastal wetland soil.

Key words: Spectral transformation; Soil total nitrogen content; Partial least squares regression; Random forest regression; Support vector regression

①基金项目: 国家自然科学基金项目(41901375, 42101393)和河北省自然科学基金项目(D2022209005)资助。

* 通讯作者(manwd@ncst.edu.cn)

作者简介: 张清文(1998—), 男, 山东临沂人, 硕士研究生, 主要从事生态环境遥感研究。E-mail: zhangqingwen@stu.ncst.edu.cn

氮元素是土壤中最大的限制性营养元素,在动植物生长、微生物呼吸中有着不可或缺的作用^[1]。土壤全氮(TN)含量的高低直接影响滨海湿地生态系统的初级生产力,且对生态系统碳循环等其他生物化学循环有重要影响^[2]。滨海湿地是陆地生态系统和海洋生态系统的缓冲区,在净化水体、涵养水源和保持生物多样性等方面有重要作用^[3],同时也是全球气候变化和人类活动共同作用下的生态敏感区,生态脆弱性高。人类生产生活排放已经造成滨海湿地氮循环失衡^[4],所以准确预测土壤 TN 含量对生态系统氮循环有重要意义。

传统化学分析方法测定土壤 TN 含量精度较高,但劳动强度大,人力财力消耗多,且产生大量环境污染^[5]。高光谱技术的发展为实现快速、简便、无污染估测土壤各种理化性质提供了新的方法。很多学者通过对原始光谱进行变换、计算光谱指数、筛选敏感波段等方法确定自变量,再基于数学模型对土壤各种理化性质进行估测,如土壤含水量^[6]、重金属含量^[7]、有机质含量^[8]、有机碳含量^[9]等,均达到了较好的预测效果。高光谱所含土壤信息丰富,合理选取建模波段对土壤 TN 含量的准确预测极其重要。刘凡等^[10]利用 677 nm 和 1 414 nm 波段反射率建立了土壤 TN 含量模型,但 2 个波段难以概括全部土壤信息,所建模型有一定的局限性。李焱等^[11]选取 Pearson 相关系数最大的 10 个波段与土壤 TN 含量进行建模分析,建模效果得到一定程度的改进。于雷等^[12]比较了全波段建模与敏感波段建模,发现虽然基于全波段建模精度要优于敏感波段,但敏感波段建立的模型复杂度远远低于全波段建模。为了包含绝大部分光谱信息,同时尽量降低模型复杂度,筛选敏感波段是必要的。

目前,用于估测土壤理化性质的线性模型有多元线性回归、多元逐步回归和偏最小二乘回归(Partial Least Square Regression, PLSR)等^[13-14]。其中,多元线性回归适用于自变量少,且自变量之间相关性较小的情况,对于土壤光谱,波段数较多且波段之间相关性强,所以多元线性回归有一定的局限性^[15];多元逐步回归可以筛除干扰变量,选择合适的解释变量回归,但是依然不能解决波段间的共线性问题;而 PLSR 通过提取主成分的方法可以解决这一问题,且可以实现对全部波段信息的概括,建模效果优于多元线性回归和多元逐步回归^[16]。除此之外,机器学习算法中如决策树回归、随机森林回归(Random Forest Regression, RFR)和支持向量机回归(Support Vector Regression, SVR)等具有极强的非线性拟合能力,被

广泛应用于土壤性质的估算^[17]。决策树回归复杂度低且适用于高维数据,是构建非线性模型的常用方法,但单棵决策树回归容易过拟合,泛化能力差。RFR 由多棵决策树集成,改善了单棵决策树的缺点,更加稳定,是高光谱估测土壤理化性质的理想方法^[18]。除机器学习外,深度学习也被广泛应用,有学者比较了 RFR、SVR 和神经网络建模,结果表明 RFR 优于神经网络和 SVR^[19]。

滨海湿地土壤受潮汐影响大,海水携带的营养物质多,理化性质更新快,湿地土壤 TN 含量与内地差异较大。由于野外采样受潮汐影响,采样可达性差,所以基于原始光谱及其多种光谱变换形式的滨海湿地土壤 TN 含量建模研究较少。基于此,本研究通过 PLSR、RFR 和 SVR 三种建模方法结合不同光谱变换构建土壤 TN 含量的预测模型,分析模型效果优劣,探求可用于滨海湿地土壤 TN 含量的光谱预测方法,以期为估测滨海湿地土壤 TN 储量提供方法依据,同时为揭示滨海湿地生态系统氮循环提供帮助。

1 材料与方法

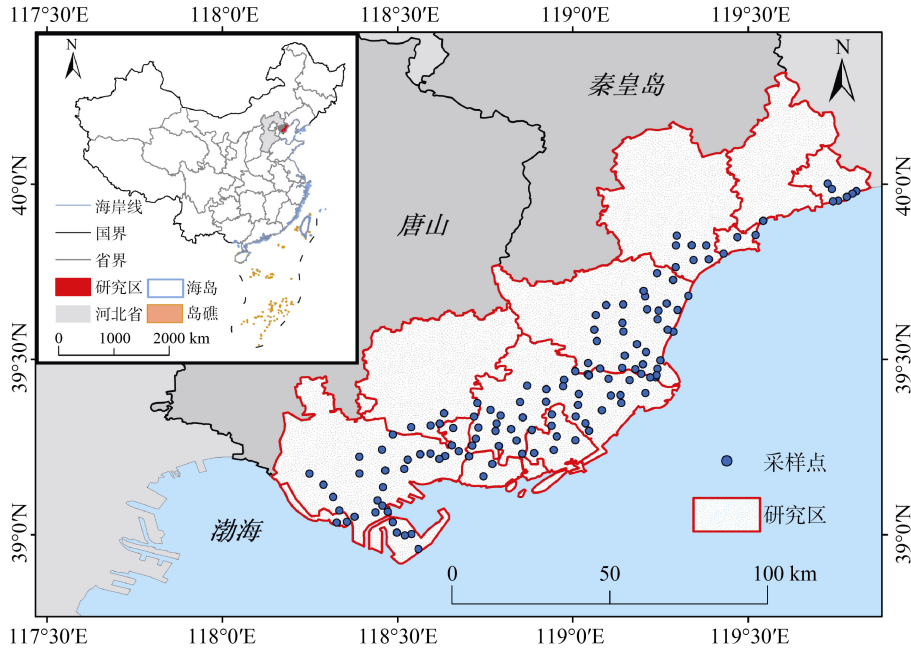
1.1 土样的采集与制备

采用环刀法采集 133 个唐山市和秦皇岛市滨海湿地表层土样(采样深度为 0~20 cm),采样点分布如图 1 所示。为减少外界因素对样品 TN 含量的影响,对所采集的土样进行室内自然风干,剔除异物,碾磨,过 100 目筛,然后分为 2 份,分别用于凯氏蒸馏法测定土样 TN 含量^[20]和高光谱数据采集。

土壤 TN 含量统计特征如表 1 所示。土壤 TN 含量最大值为 3.124 g/kg,采集于芦苇湿地;TN 含量最小值为 0.024 g/kg,采集于海边砂土;TN 平均含量为 0.733 g/kg,标准差为 0.501 g/kg,变异系数为 68.5%。133 个土样中包含 84 个粉土和 49 个砂土,不同质地的土样土壤 TN 含量平均值有显著差异,粉土 TN 含量平均值为 0.881 g/kg,砂土 TN 含量平均值为 0.478 g/kg。砂粒含量较高的土样 TN 含量低,粉粒含量高的土样 TN 含量高^[21]。

1.2 土样光谱测定与预处理

采用 ASD 公司生产的 FieldSpec4 获取土样的高光谱反射率(R),测定范围为 350~2 500 nm,光谱采样间隔在 350~1 000 nm 波段范围为 1.4 nm,1 000~2 500 nm 波段范围为 2 nm,重采样间隔为 1 nm。每个样品重复测定 10 条光谱曲线,并使用光谱仪自带软件 ViewSpec Pro 取平均值作为该样品原始光谱信息。在光谱采集的过程中会不可避免地受到测试环



(该图基于国家测绘地理信息局标准地图服务网站下载的审图号为 GS(2020)4619 的标准地图制作, 底图无修改)

图 1 研究区位置和采样点分布图

Fig. 1 Location of study area and soil sampling sites

表 1 土壤质地及其 TN 含量描述性统计特征
Table 1 Descriptive statistics of soil texture and TN content

质地类型	样点数	最小值 (g/kg)	最大值 (g/kg)	均值 (g/kg)	标准差 (g/kg)	变异系数 (%)
全部	133	0.024	3.124	0.733	0.501	68.5
粉土	84	0.314	2.255	0.881	0.333	37.8
砂土	49	0.024	3.124	0.478	0.625	76.5

境、仪器自身的影响,导致光谱曲线噪声较大,因此本研究中,对平均后的光谱反射率曲线进行 S-G 平滑处理得到 R ,再对 R 做 7 种光谱变换:对数($\lg R$)、对数的一阶微分($(\lg R)'$)、倒数($1/R$)、倒数的一阶微分($(1/R)'$)、一阶微分(R')、平方根(\sqrt{R})、一阶微分的倒数($1/(R)'$)。数据的对数、倒数、平方根处理采用 R 软件的“stats”包完成,数据的一阶微分处理采用 Origin Pro 2021 完成。

本研究对土壤 TN 含量与光谱反射率及其各种变换形式进行相关性分析, Pearson 相关系数范围为 $-1.0 \sim 1.0$, 绝对值越大,其相关性越高,波段越敏感;选取相关系数大于 0.4 的波段,进行 $P < 0.01$ 水平上的显著性检验,通过显著性检验的波段确定为敏感波段,用于土壤 TN 含量高光谱模型的估算。

1.3 数据分析方法

1.3.1 偏最小二乘回归(PLSR) PLSR 兼顾了多元线性回归、主成分分析和自变量因变量相关性分析 3 种分析方法的优点,不仅能解决自变量相关性较强

的问题,还可以通过提取主成分的方式降低自变量维数,在样本数远小于自变量数的情况下建模。本研究通过判断留一法(leave One Out, LOO)交叉验证的预测均方根误差(Root Mean Square Error of Prediction, RMSEP)确定主因子个数。理想的 RMSEP 值随主因子个数的增加先递减,达到最低点后随主因子个数的增加出现微小上升或波动。取 RMSEP 最小时对应的主因子个数建立 PLSR 模型可有效避免过拟合问题。本研究选择与土壤 TN 含量相关性强的成分,不考虑全部的自变量,因此建立的模型有更好的鲁棒性。采用 R 软件中“pls”包完成 PLSR 建模。

1.3.2 随机森林回归(RFR) RFR 是多个决策树集成的学习器,可以高精度地处理大数据集,具有数据适应性强、不易受极大值或极小值的影响即稳定性高等优点,所以模型稳定性比 PLSR 高。RFR 通过对多个决策树的集成融合,输出结果为多棵决策树的平均值。构建 RFR 模型:

$$\bar{h}(x) = \frac{1}{T} \sum_{i=1}^T \{h(x, \theta_i)\} \tag{1}$$

式中: $\bar{h}(x)$ 为模型预测结果; $\{h(x, \theta_i)\}$ 为基于 x 和 θ_i 的输出; x 为建模波段; θ_i 是独立同分布随机向量; T 为回归决策树的个数^[22]。本研究 RFR 建模通过 R 软件中“randomForest”包实现,在建模过程中需要定义两个主要的参数:每棵树的输入变量数量(mtry)和树的数量(ntree),利用“caret”包的网格搜索法确

定最优参数 $mtry$ 和 $ntree$ ^[23]，采用预测误差最小的参数组合进行最终建模。

1.3.3 支持向量机回归(SVR) SVR 是一种基于统计学习理论的机器学习技术。SVR 使用核函数将数据投影到高维空间，并进行回归。在 SVR 建模中，选取径向核函数，调整两个参数 γ 和 $cost$ 。其中， $cost$ 为惩罚系数， $cost$ 过大，容易出现过拟合， $cost$ 过小，容易拟合不充分； γ 决定数据映射到高维空间后的分布， γ 越大，支持向量越少， γ 值越小，支持向量越多。本研究利用“caret”包的网格搜索法确定最优参数 γ 和 $cost$ ，采用预测误差最小的参数组合进行最终建模。使用 R 软件“e1071”包建立 SVR 模型。

1.3.4 模型精度验证 建立土壤 TN 预测模型，采用 LOO 交叉验证。模型的评价指标采用预测值与实测值的决定系数(R^2)、均方根误差(Root Mean Squared Error, RMSE)和平均绝对误差(Mean Absolute Error, MAE)。其中，RMSE 用来衡量观测值同真值之间的偏差；MAE 是绝对误差的平均值，能更好地反映预测值误差的实际情况。模型精度评价标准如表 2 所示， R^2 越接近 1，拟合效果越好，RMSE、MAE 值越小，说明模型估算精度越高。计算公式为：

$$R^2 = \frac{\sum_{i=1}^n (\hat{y}_i - \bar{y})^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad (2)$$

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (\hat{y}_i - y_i)^2} \quad (3)$$

$$MAE = \frac{1}{n} \sum_{i=1}^n |\hat{y}_i - y_i| \quad (4)$$

式中： \hat{y}_i 为第 i 个样本的预测值； y_i 第 i 个样本的实测值； n 为样本容量； \bar{y} 为所有样本实测值的平均值。

表 2 模型整体精度评价标准^[24]

Table 2 Evaluation standard of model overall accuracy

R^2	模型预测能力描述
<0.5	模型精度差，不具备预测能力
0.5~0.66	模型精度一般，具备粗略预测能力
0.67~0.89	模型精度好，具备预测能力
>0.89	模型精度极好，具备准确估测能力

2 结果与分析

2.1 土壤光谱特征和全氮含量的相关性

由图 2 可以看出，研究区土壤样本的原始光谱反

射率(R)曲线形态基本一致，各曲线间差异较小，呈曲折上升的弧形，在 1 410、1 920、2 210 nm 出现 3 个明显的水分吸收谷。其中，在可见光波段(350 ~ 780 nm)范围内， R 迅速上升；除吸收谷外，在近红外波段 780 ~ 1 900 nm 范围内缓慢上升，各样本光谱间差异较大，至 2 100 nm 处 R 达到最大值；在近红外波段 2 150 ~ 2 500 nm 范围内， R 平缓下降。结合野外采样点，砂土样品的 R 曲线在 2 300 ~ 2 400 nm 处出现明显吸收谷，此区间是硅酸盐矿物中的水分子羟基(-OH)伸缩振动的合频谱带，由于砂土中硅酸盐矿物含量较大^[25]，所以在图 2 中砂土样品的反射率曲线在 2 300 ~ 2 400 nm 出现明显吸收谷，与粉土样品曲线出现交叉现象。

由表 3 可以看出， R 与土壤 TN 含量有一定的相关性，但相关性较低，两者之间的相关系数在 584 nm 处达到最大 0.547。经 7 种变换处理后，变换光谱参数与土壤 TN 含量相关系数均有不同程度的提高，基于简单数学变换的光谱参数 $1/R$ 、 \sqrt{R} 和 $\lg R$ 与土壤 TN 含量的最大相关系数仅提升 0.013。基于微分变换的光谱参数 R' 、 $(1/R)'$ 和 $(\lg R)'$ 与土壤 TN 含量相关系数提高较为明显，最大相关系数均达到 0.7 以上，表明经过微分变换可以消除背景噪声的干扰，分解混合重叠峰，提高光谱分辨率和灵敏度，凸显土壤光谱特征。光谱预处理能够较好地消除土壤质地、温度、湿度等背景干扰，从而提高光谱反射率与土壤 TN 含量的相关性^[26]。光谱变换后，除 $(\lg R)'$ 建模波段的个数有所减少外，其余明显增加， R' 、 $(1/R)'$ 、 $\lg R$ 、 $(1/R)'$ 分别增加 119、277、161、121 个。不同变换形式光谱参数敏感波段集中分布区间不同，进一步表明经过不同形式光谱变换有利于挖掘不同土壤信息。

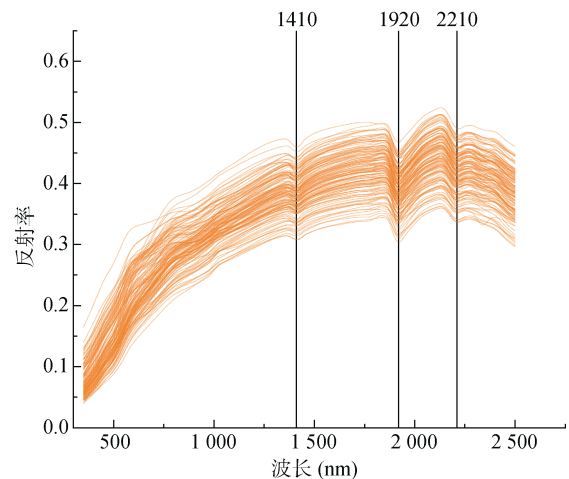


图 2 供试土壤样本原始光谱反射率

Fig. 2 Original spectral reflectance of soil samples

表 3 不同变换形式下土壤 TN 含量与光谱波段的相关性
Table 3 Correlations between soil TN contents and spectral bands under different spectrat transformations

光谱变换形式	波段数	敏感波段区间(nm)	最大相关系数	最大相关系数波长(nm)
R	415	350 ~ 764	-0.547**	584
$1/R$	443	350 ~ 792	0.560**	593
R'	534	350 ~ 792、815 ~ 1 155、1 279 ~ 1 883、1 925 ~ 2 499	0.707**	1 130
\sqrt{R}	426	350 ~ 775	0.555**	584
$1/(R)'$	692	391 ~ 1120	-0.555**	901
$\lg R$	576	350 ~ 872、2 448 ~ 2 500	-0.559**	584
$(1/R)'$	536	352 ~ 1 130、1 862 ~ 2 495、1 670 ~ 1 700、2 027 ~ 2 474	-0.746**	718
$(\lg R)'$	388	445 ~ 456、596 ~ 1 135、2 302 ~ 2 495	-0.703**	718

注：*、** 分别表示在 $P < 0.05$ 和 $P < 0.01$ 水平显著相关。

2.2 模型对比分析

在利用 PLSR 建模过程中, 经不同光谱变换形式所建立的模型精度: $(1/R)' > (\lg R)' > R' > 1/(R)' > 1/R > \sqrt{R} > \lg R > R$, 以微分变换光谱建立的模型 R^2 均高于未经微分变换光谱的 R^2 , 而 RMSE、MAE 则低于未经微分处理的对应值(表 4), 由此说明, 经微分变换光谱参数建立的模型精度要优于未经微分变换的光谱参数。经微分变换光谱参数建立的模型中, $(1/R)'$ 变换建立的模型性能最好, R^2 为 0.775, MAE 为

0.351 g/kg, RMSE 为 0.472 g/kg, 相比基于 R 建立的模型, R^2 提升了 38.9%, RMSE 和 MAE 分别降低了 28.7% 和 25.3%, 可以用于预测土壤 TN 含量。基于 $1/R$ 、 \sqrt{R} 、 $\lg R$ 和 R 建立的预测模型 R^2 为 0.558 ~ 0.598, 在 0.5 ~ 0.66, 可粗略预测滨海湿地土壤 TN 含量(表 2)。在对 R 进行光谱变换后再建模均提升了模型精度, 说明对 R 进行合适的变换能较好地消除外界因素对光谱的影响, 以增强光谱对土壤 TN 的敏感性。

表 4 各预测模型的验证与对比
Table 4 Validation and comparison of different prediction models

光谱变换形式	PLSR			RFR			SVR		
	RMSE (g/kg)	R^2	MAE (g/kg)	RMSE (g/kg)	R^2	MAE (g/kg)	RMSE (g/kg)	R^2	MAE (g/kg)
$1/R$	0.635	0.598	0.455	0.193	0.852	0.134	0.376	0.438	0.212
R	0.662	0.558	0.470	0.199	0.843	0.138	0.387	0.404	0.224
$\lg R$	0.643	0.583	0.466	0.204	0.835	0.140	0.347	0.521	0.183
\sqrt{R}	0.588	0.588	0.461	0.197	0.845	0.139	0.383	0.418	0.219
R'	0.501	0.748	0.370	0.146	0.916	0.096	0.057	0.987	0.052
$1/(R)'$	0.516	0.732	0.380	0.147	0.915	0.095	0.057	0.987	0.050
$(\lg R)'$	0.488	0.760	0.357	0.155	0.905	0.104	0.120	0.943	0.070
$(1/R)'$	0.472	0.775	0.351	0.151	0.910	0.103	0.059	0.987	0.050

在利用 RFR 建模过程中, 不同光谱变换形式建模效果: $R' > 1/(R)' > (1/R)' > (\lg R)' > 1/R > \sqrt{R} > R > \lg R$ (表 4)。基于 R' 建立的预测模型效果最佳, R^2 为 0.916, RMSE 为 0.146 g/kg, MAE 为 0.096 g/kg。RFR 模型与 PLSR 模型相比, 基于不同光谱变换形式建立的 RFR 模型 R^2 变化均较小, R^2 集中在 0.835 ~ 0.916, 说明光谱变换对 RFR 模型影响较小。RFR 模型 MAE 和 RMSE 小于 PLSR 模型, MAE 在 0.095 ~ 0.140 g/kg, RMSE 在 0.146 ~ 0.204 g/kg。其中, 经微分变换光谱参数建立的模型 $R^2 \geq 0.905$, RMSE ≤ 0.155 g/kg, 模型精度极好, 均可精确估算土壤 TN 含量; 未经微分变

换光谱参数建立的模型 $R^2 \leq 0.852$, RMSE ≥ 0.193 g/kg, 模型精度好, 具备预测能力。

在利用 SVR 建模过程中, 不同光谱变换形式建模效果: $1/(R)' > (1/R)' > R' > (\lg R)' > \lg R > 1/R > \sqrt{R} > R$ (表 4)。建模效果出现明显差异, 即 $\lg R$ 、 $1/R$ 、 \sqrt{R} 和 R 未经微分变换光谱参数建立的模型 RMSE 在 0.347 ~ 0.387 g/kg, R^2 在 0.404 ~ 0.521, MAE 在 0.183 ~ 0.224 g/kg, 模型精度差, 不能估测滨海湿地土壤 TN 含量; 而 R' 、 $(1/R)'$ 、 $1/(R)'$ 和 $(\lg R)'$ 经微分变换光谱参数建立的模型 RMSE 在 0.057 ~ 0.120 g/kg, R^2 在 0.943 ~ 0.987, MAE 在 0.050 ~ 0.070 g/kg, 模型

精度极好,可精确估算土壤 TN 含量,其中,基于 $1/(R)'$ 建立的模型精度最高, R^2 达到 0.987, 是准确预测滨海湿地土壤 TN 含量的最理想模型。RFR 和 SVR 建模中最优模型要高于 PLSR 建模中最优模型,证明波段反射率与土壤 TN 含量呈非线性关系,非线性模型具有更强的解释能力^[27-28]。土壤光谱是水分、盐分、pH 和有机质等各种理化性质综合作用的结果^[29], 土壤 TN 含量也是多种因子共同作用的结果,为进一步探究土壤 TN 含量,应从多角度、多维度分析各种因子的综合作用。

对 R 进行 7 种光谱变换之后,建模 R^2 、RMSE 和 MAE 得到了改善,但不同变换形式所建模型精度差异较大。 R 和 R' 、 $1/R$ 和 $(1/R)'$ 、 $\lg R$ 和 $(\lg R)'$ 相比,一阶微分变换参数建模效果显著优于未经过一阶微分变换参数。 R 经一阶微分变换,反射率大小与土壤 TN 含量之间的关系转化为反射率斜率与土壤 TN 含量之间

的关系,消除了反射率大小中的噪声,所以一阶微分变换参数建模的效果普遍较好。 R 和 $1/R$ 、 R' 和 $1/R'$ 在 3 种建模方法中表现相近,说明经简单的倒数运算对光谱信息的挖掘能力较弱。Brunet 等^[30]、徐永明等^[31] 也比较了基于微分变换与其他光谱变换所建立的预测模型精度,结果证明,一阶微分变换所建模型精度更高。因此,对 R 进行合适的光谱变换,对于提高土壤 TN 含量预测模型的精度和稳定性具有关键作用。

本研究选取敏感波段为自变量,以土壤 TN 含量为因变量进行 PLSR、RFR 和 SVR,建立 R 及其 7 种光谱变换形式对土壤 TN 含量的预测模型。图 3 显示了 3 种回归方法各自的最优模型的土壤 TN 含量预测值与实测值的散点图,可以明显看出,SVR 模型预测的土壤 TN 含量与实测值的拟合曲线比 PLSR 和 RFR 模型更接近 1:1 线,表明实测值与预测值较为接近,模型预测精度更高。

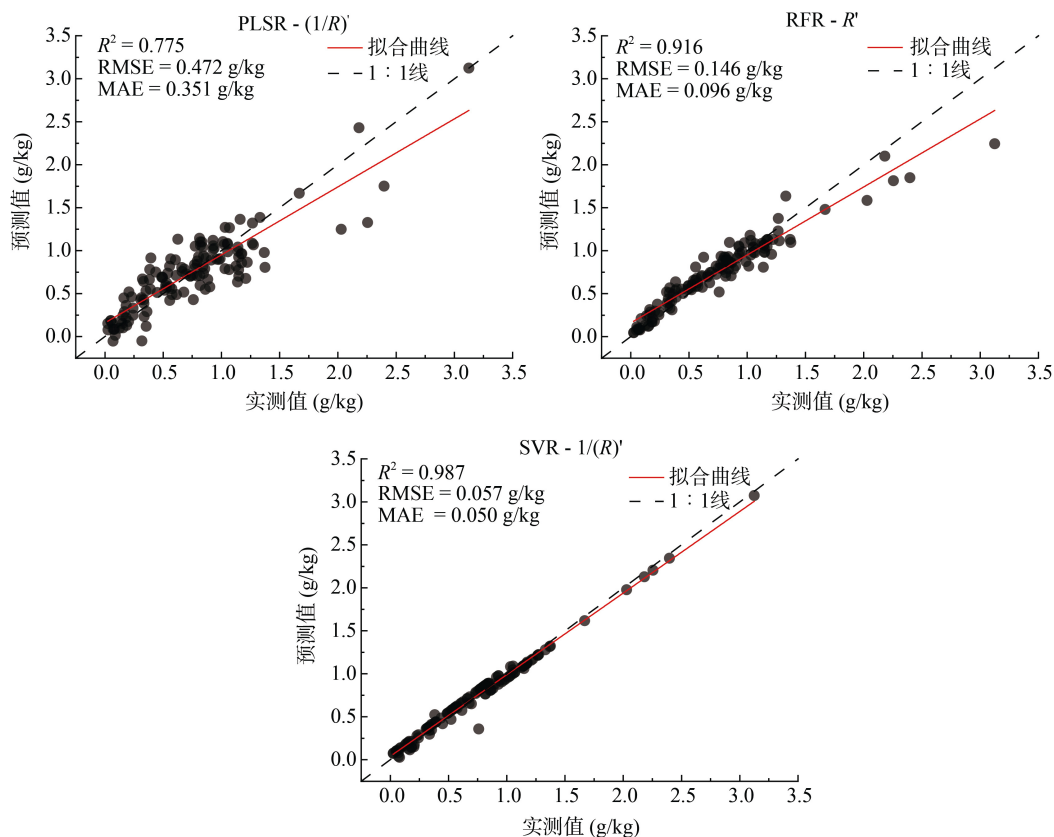


图 3 PLSR、RFR、SVR 建模中最优模型土壤 TN 含量实测值与预测值比较

Fig. 3 Comparison between measured and predicted TN contents in optimal soil TN models established by PLSR, RFR and SVR

3 结论

1) 滨海湿地土壤原始光谱反射率与 TN 含量有较好的相关性,可以用于对土壤 TN 含量进行估测,而对原始光谱反射率曲线进行变换,可进一步提高光

谱反射率与土壤 TN 含量的相关性,其中提高最明显的变换形式为 $(1/R)'$ 。微分变换可以有效消除噪音,提高光谱反射率对土壤全氮含量的敏感度,故利用土壤光谱反射率预测土壤 TN 含量是可行的。

2) 根据光谱反射率及其变换形式建立的 PLSR、

RFR 和 SVR 模型, 每种算法里均有较好的预测模型可以粗略预测土壤 TN 含量。在 PLSR 模型中, 基于 $(1/R)$ 建立的模型预测精度最高, 可以粗略预测土壤 TN 含量; 在 RFR 模型中, 基于 R' 建立的模型可精确预测土壤 TN 含量; 在 SVR 模型中, 基于 $1/(R)$ 建立的模型可准确预测土壤 TN 含量。在建立的模型中, 经光谱变换后建立的模型预测精度比基于 R 建立的模型精度有所提升, 故对 R 变换后建模是科学的。

3) 基于 $1/(R)$ 建立的 SVR 模型 R^2 为 0.987, RMSE 为 0.057 g/kg, MAE 为 0.050 g/kg, 是预测滨海湿地土壤 TN 含量的最优模型。

参考文献:

- [1] 王莉雯, 卫亚星. 植被氮素浓度高光谱遥感反演研究进展[J]. 光谱学与光谱分析, 2013, 33(10): 2823–2827.
- [2] 宋长春, 宋艳宇, 王宪伟, 等. 气候变化下湿地生态系统碳、氮循环研究进展[J]. 湿地科学, 2018, 16(3): 424–431.
- [3] Kirwan M L, Megonigal J P. Tidal wetland stability in the face of human impacts and sea-level rise[J]. Nature, 2013, 504(7478): 53–60.
- [4] Socolow R. Fitting on the earth: Challenges of carbon and nitrogen cycle to preserve the habitability of the planet[J]. Engineering, 2016, 2(1): 21–22.
- [5] 王红, 刘高焕, 宫鹏. 利用 Cokriging 提高估算土壤盐离子浓度分布的精度——以黄河三角洲为例[J]. 地理学报, 2005, 60(3): 511–518.
- [6] 刘焕军, 王翔, 张小康, 等. 松嫩平原主要土壤类型含水量高光谱预测模型[J]. 土壤通报, 2018, 49(1): 38–44.
- [7] 郭云开, 张思爱, 王建军, 等. 特征变量选择结合 SVM 的耕地土壤 Hg 含量高光谱反演[J]. 测绘工程, 2022, 31(1): 17–23.
- [8] 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.
- [9] 陈秋宇, 杨仁敏, 朱长明. 基于 VIS-NIR 光谱的互花米草入侵湿地土壤有机碳预测研究[J]. 土壤学报, 2021, 58(3): 694–703.
- [10] 刘凡, 马玲, 杨光, 等. 灰漠土土壤全氮含量的高光谱特征分析及估测[J]. 新疆农业科学, 2017, 54(1): 140–147.
- [11] 李焱, 王让会, 管延龙, 等. 基于高光谱反射特性的土壤全氮含量预测分析[J]. 遥感技术与应用, 2017, 32(1): 173–179.
- [12] 于雷, 洪永胜, 耿雷, 等. 基于偏最小二乘回归的土壤有机质含量高光谱估算[J]. 农业工程学报, 2015, 31(14): 103–109.
- [13] 涂宇龙, 邹滨, 姜晓璐, 等. 矿区土壤 Cu 含量高光谱反演建模[J]. 光谱学与光谱分析, 2018, 38(2): 575–581.
- [14] 尼加提·卡斯木, 师庆东, 郭玉川, 等. 基于 EM38 和 WorldView-2 影像的土壤盐渍化建模研究[J]. 土壤, 2019, 51(3): 594–601.
- [15] 钱佳, 郭云开, 蒋明, 等. 不同类型土壤 Cu 含量高光谱联合反演建模[J]. 测绘科学, 2020, 45(8): 138–144.
- [16] 陈红艳, 赵庚星, 李玉环, 等. 消除水分因素影响的野外原状土壤盐分高光谱建模估测[J]. 农业工程学报, 2018, 34(12): 119–125.
- [17] 张东辉, 赵英俊, 秦凯. 一种新的光谱参量预测黑土养分含量模型[J]. 光谱学与光谱分析, 2018, 38(9): 2932–2936.
- [18] 王腾军, 方珂, 杨耘, 等. 随机森林回归模型用于土壤重金属含量多光谱遥感反演[J]. 测绘通报, 2021(11): 92–95.
- [19] Cui L J, Dou Z G, Liu Z J, et al. Hyperspectral inversion of *Phragmites communis* carbon, nitrogen, and phosphorus stoichiometry using three models[J]. Remote Sensing, 2020, 12(12): 1998.
- [20] 鲍士旦. 土壤农化分析[M]. 3 版. 北京: 中国农业出版社, 2000.
- [21] 林俊杰, 杨振宇, 刘丹, 等. 干湿交替下三峡支流消落带沉积物粒径组成及氮分布特征[J]. 土壤学报, 2016, 53(3): 602–611.
- [22] 孙兴亮, 郝晓华, 王建, 等. 基于光谱-环境随机森林回归模型的 MODIS 积雪面积比例反演研究[J]. 冰川冻土, 2022, 44(1): 147–158.
- [23] Song J R, Gao J H, Zhang Y B, et al. Estimation of soil organic carbon content in coastal wetlands with measured VIS-NIR spectroscopy using optimized support vector machines and random forests[J]. Remote Sensing, 2022, 14(17): 4372.
- [24] Viscarra Rossel R A, McGlynn R N, McBratney A B. Determining the composition of mineral-organic mixes using UV-vis-NIR diffuse reflectance spectroscopy[J]. Geoderma, 2006, 137(1/2): 70–82.
- [25] 王惠敏. 基于光谱吸收特征的土壤有机质与重金属含量估算研究[D]. 徐州: 中国矿业大学, 2019.
- [26] 殷哲, 雷廷武, 陈展鹏, 等. 近红外传感器测量不同种类土壤含水率的适应性研究[J]. 农业机械学报, 2014, 45(3): 148–151, 190.
- [27] 殷彩云, 白子金, 罗德芳, 等. 基于高光谱数据的土壤全氮含量估测模型对比研究[J]. 中国土壤与肥料, 2022(1): 9–15.
- [28] 王海江, 刘凡, Yungler J A, 等. 不同粒径处理的土壤全氮含量高光谱特征拟合模型[J]. 农业机械学报, 2019, 50(2): 195–204.
- [29] 徐彬彬. 土壤剖面的反射光谱研究[J]. 土壤, 2000, 32(6): 281–287.
- [30] Brunet D, Barthès B G, Chotte J L, et al. Determination of carbon and nitrogen contents in Alfisols, Oxisols and Ultisols from Africa and Brazil using NIRS analysis: Effects of sample grinding and set heterogeneity[J]. Geoderma, 2007, 139(1/2): 106–117.
- [31] 徐永明, 蔺启忠, 黄秀华, 等. 利用可见光/近红外反射光谱估算土壤总氮含量的实验研究[J]. 地理与地理信息科学, 2005, 21(1): 19–22.