doi:10.6041/j.issn.1000-1298.2020.01.017

基于 FOD 和 SVMDA - RF 的土壤有机质含量高光谱预测

张智韬[1,2]　劳聪聪[1,2]　王海峰[1,2]　ARNON Karnieli[3]　陈俊英[1,2]　李　宇[1,2]

(1. 西北农林科技大学水利与建筑工程学院, 陕西杨凌 712100;
2. 西北农林科技大学旱区农业水土工程教育部重点实验室, 陕西杨凌 712100;
3. 本古里安大学 Blaustein 沙漠研究所, 思德博克 84990)

摘要: 为探讨分数阶微分(FOD)联合支持向量机分类-随机森林模型改善高光谱监测荒漠土壤有机质含量(SOM)的效果,对以色列 Sde Boker 荒漠地区采集的砂质土(SS)和黏壤土(CLS)样品进行理化分析和室内光谱测定,依据光谱的平均反射率建立支持向量机分类模型(SVMAD),并对不同土质高光谱原始反射率分别经 0 ~ 2 阶(间隔 0.2)的分数阶微分处理,构建归一化光谱指数(NDI),分析 NDI 和 SOM 之间的二维相关性,并筛选敏感的 NDI 指数,以此建立不同 FOD 的随机森林(RF)模型,并以不同土质中的最佳模型进行组合,构建新的 SVMDA - RF 模型。结果表明:基于光谱平均反射率的 SVMDA 对土壤质地的分类正确率可达 100%;分数阶微分耦合光谱指数具有放大波长间与 SOM 有关隐含信息的能力,经 FOD 提升敏感指数的数量在 0.6 阶时达到峰值,但黏壤土的敏感指数数量远大于沙质土;由不同 FOD 敏感指数建立的 RF 模型中,砂质土在 1.2 阶的模型最佳(R_C^2 = 0.962, R_P^2 = 0.920, RMSEP 为 0.435 g/kg, RPD 为 3.658),黏壤土在 0.6 阶的模型最佳(R_C^2 = 0.942, R_P^2 = 0.944, RMSEP 为 0.554 g/kg, RPD 为 4.316);经最佳模型组合后的 SVMDA - RF 模型,砂质土和黏壤土的模型精度都有所提高,其中 R_C^2 = 0.980, R_P^2 = 0.979, RMSEP 为 0.481 g/kg, RPD 为 7.004。研究成果可为快速评估荒漠土壤有机质含量提供依据。

关键词: 荒漠土壤; 有机质含量; 高光谱; 支持向量机分类; 分数阶微分; 随机森林

中图分类号: S127; S153.6　　**文献标识码**: A　　**文章编号**: 1000-1298(2020)01-0156-12　　**OSID**:

Estimation of Desert Soil Organic Matter through Hyperspectra Based on Fractional - Order Derivatives and SVMDA - RF

ZHANG Zhitao[1,2]　LAO Congcong[1,2]　WANG Haifeng[1,2]　ARNON Karnieli[3]　CHEN Junying[1,2]　LI Yu[1,2]

(1. *College of Water Resources and Architectural Engineering, Northwest A&F University, Yangling, Shaanxi* 712100, *China*
2. *Key Laboratory of Agricultural Soil and Water Engineering, Ministry of Education, Northwest A&F University, Yangling, Shaanxi* 712100, *China*
3. *Jocob Blaustein Institute for Desert Research, Ben - Gurion University of the Negev, Seder Boker* 84990, *Israel*)

Abstract: Aiming to explore the effect of fractional - order derivatives (FOD) combined with support vector machine discriminant analysis - random forest model (SVMDA - RF) on hyperspectral monitoring of desert soil organic matter content (SOM). The desert soil samples collected in the Sde Boker area of Israel were analyzed. These soil samples were through pretreatment, physical and chemical analysis, soil classification (divided into sandy soil (SS) and clay loam soil (CLS)), indoor spectral acquisition and spectral resampling (interval 10 nm). In order to avoid the influence of soil quality on the inversion model, the support vector machines discriminant analysis (SVMAD) was established based on the average reflectance of the spectrum. The spectral reflectance was processed by 0 ~ 2 order (interval 0.2) FOD. Then NDI was constructed by using the spectral data that through fractional order derivatives processing and the two-dimensional correlation between SOM and NDI was analyzed. In order to obtain all different FOD enhanced-NDI, the highest coefficient of determination (R^2) of 0-order NDI was used as the threshold (sand soil $R^2 > 0.901$, clay loam soil $R^2 > 0.763$). By using the different FOD enhanced-NDI to establish random forest (RF) models. All models based on different soils were compared and analyzed, the best models of different soils were combined to establish the SVMDA - RF model. The

收稿日期: 2019 - 06 - 01　修回日期: 2019 - 07 - 08

基金项目: 国家重点研发计划项目(2017YFC0403302、2016YFD0200700)和杨凌示范区科技计划项目(2018GY - 03)

作者简介: 张智韬(1976—),男,副教授,博士,主要从事遥感技术在节水灌溉及水资源中的应用研究,E-mail: zhitaozhang@126.com

results showed that SVMDA based on spectral average reflectance, the classification rate of soil texture could reach 100%. Fractional order derivatives coupling normalized spectral index, which had ability to amplify SOM-related implicit information between bands, but it had different effects on different soils. For the paper, clay loam soil was superior to sandy soil, and two soils of the FOD-enhanced sensitive index peaked at 0.6-order, but the number of sensitive index of clay loam soil was much larger than that of sandy soil. In the sandy soil RF models, the model based on 1.2-order NDI was the best($R_C^2=0.962$, $R_P^2=0.920$, RMSEP was 0.435 g/kg, and RPD was 3.658). In the loam RF models, the model based on 0.6-order NDI was the best ($R_C^2=0.942$, $R_P^2=0.944$, RMSEP was 0.554 g/kg, and RPD was 4.316). Combining the optimal models of the two soils to get the high-precision SVMDA-RF model, $R_P^2=0.979$, RMSEP was 0.481 g/kg, and RPD was 7.004. The model could provide effective support for quickly assessing the desert soil types and fertility.

Key words: desert soil; soil organic matter; hyperspectral; support vector machines discriminant analysis; fractional－order derivative; random forest

0　引言

土壤有机质含量(Soil organic matter,SOM)是评价土壤肥力的最主要指标,也是了解土壤退化状态的主要指标,监测 SOM 是掌握土壤肥力和土壤退化情况的重要手段。高光谱遥感可以快速、无损监测土壤有机质含量,为荒漠土壤整治提供依据。然而土壤高光谱具有非特异性,波长相邻的反射率互相重叠,且具有较强的相关性,当土壤肥力较低(SOM 小于 2%)时,土壤质地等特性会对光谱产生较大影响,在一定程度上掩盖了 SOM 对光谱的作用,使得高光谱反演荒漠土壤的 SOM 具有一定困难[1]。

国内外学者对利用高光谱技术反演 SOM 已进行了大量研究[2-8]。但是上述研究均是以 SOM 很高或较高的地区为研究区域,其成果无法很好应用于 SOM 较低的干旱荒漠地区。侯艳军等[1]利用高光谱反演新疆准噶尔盆地东部的荒漠土壤有机质,发现基于一阶微分光谱的 PLSR 最优,其 R^2 可达 0.78。王海峰等[9]通过对以色列 Sde Boker 地区采集的荒漠化土壤进行研究得出,利用灰度关联分析结合标准正态变换光谱建立的岭回归模型可以较好预测 SOM,R^2 可达 0.866。这说明高光谱遥感监测荒漠土壤 SOM 具有一定的可行性,但其所研究的模型均为线性模型,而线性模型无法较好解决非线性问题。因此,研究非线性组合模型反演荒漠土壤有机质含量十分必要。

此外,分数阶微分[10](Fractional－order derivative,FOD)与一阶、二阶微分相比,能以较小微分间隔确保光谱曲线形状缓慢变化,可以检测到光谱信号的更多特征,而敏感光谱指数是与对应 SOM 经二维相关分析筛选确定,既考虑了波长间的关系,也考虑 SOM 与波长之间的关系,可减少无关波长的影响。根据 HONG 等[11]研究,将两者耦合可进一步提升效果,但现有研究[11-12]一般取相关系数最高的指数建模,未完全利用经分数阶微分优化的光谱指数。

本文以在以色列荒漠地区采集的砂质土与黏壤土为对象,在室内进行理化分析与光谱采集,利用平均光谱反射率建立支持向量机分类(Support vector machines discriminant analysis,SVMAD)模型,进行分数阶微分后建立归一化光谱指数(Normalized difference index,NDI),再对 NDI 与 SOM 进行二维相关性分析,并以 0 阶 NDI 的最高决定系数作为阈值筛选敏感指数,结合随机森林模型(Random forest,RF),构建 SOM 的 SVMAD－RF 模型,研究其对荒漠土壤的监测精度。

1　材料与方法

1.1　研究区概况

研究区为 Sde Boker 等地区(34°26′～34°47′E,30°17′～30°52′N),该区域位于以色列南部,海拔 640 m,属于干旱半干旱气候,年降水量介于 100～200 mm 之间,年均蒸发量达 2 500 mm。尤其以色列南部全年降水稀少,占以色列一半以上的面积,被荒漠所覆盖。故而此地的荒漠土壤占大部分且有机质含量低(0.5～18 g/kg)[13],主要的土壤类型为黏壤土(CLS)与砂质土(SS)。图 1 为取样分布示意图,其中图 1e 所示样本为砂质土,图 1b、1c、1d 所示样本均为黏壤土。

1.2　试验材料及土壤样本有机质含量的测定

通过十字法在研究区域土层(0～20 cm)采集 134 份土样。土样经过风干、磨细、经孔径 2 mm 筛处理后,将每个土样分成 2 份,分别用于光谱测试和理化性质分析。由激光粒度仪测得土样粒度分布,根据国际土壤颗粒分类标准,将所取样本划分为两类,其中砂质土样本 78 个、黏壤土样本 56 个。土样 SOM 通过重铬酸钾外加热法测定。

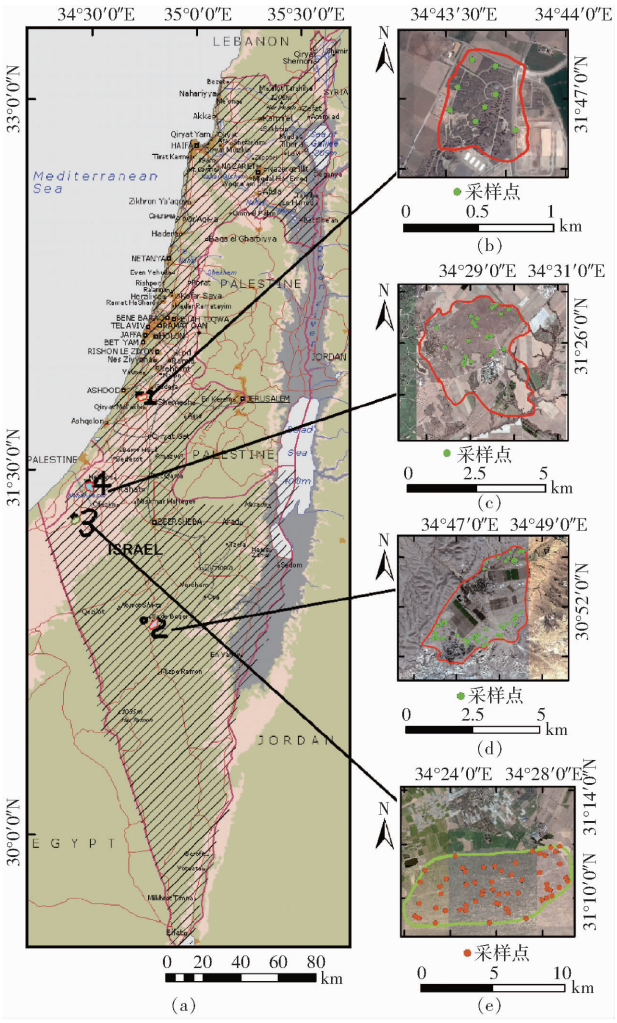

图 1 取样点分布示意图

Fig. 1 Sketch of study area and soil sampling locations

1.3 光谱测定

在照明控制的暗室,利用 ASD FieldSpec 3 型光谱仪测量土壤光谱反射率。将制备的 134 份土样装入直径为 10 cm、深度为 2 cm 的黑色容器中,土样填满容器后将表面刮平。光谱仪性能参数、光谱测试的几何参数、光谱数据的采集方式等参照文献[14]。

土样光谱通过 Savitzky - Golay 算法(多项式阶数为 2,点数为 10)进行平滑,并将噪声较大的波段(350 ~ 399 nm、2 401 ~ 2 500 nm)剔除。为了降低冗余度,按 10 nm 为间隔重采样,所得 400 ~ 2 400 nm 的光谱数据用于后续分析。去噪后的砂质土和黏壤土样品光谱曲线如图 2(图中灰色区域表示光谱的标准偏差,下同)所示。

1.3.1 平均反射率

为提高对土壤质地的分类正确率,尝试以平均反射率作为标准进行分类。全部波段(400 ~ 2 400 nm)的平均反射率计算公式为

$$\overline{R} = \frac{1}{n}\sum_{i=1}^{n} R_i \tag{1}$$

式中 R_i——波长 i 的土壤光谱反射率

利用 Matlab 从 134 个样本中随机选取 89 个样本作为分类建模集,其余 45 个样本作为验证集。由图 3 可知,建模集与验证集的砂质土与黏壤土均以 $\overline{R}=0.3$ 为界,砂质土集中在 $\overline{R}=0.3$ 上方,黏壤土则处于其下方,可为分类提供可靠依据。

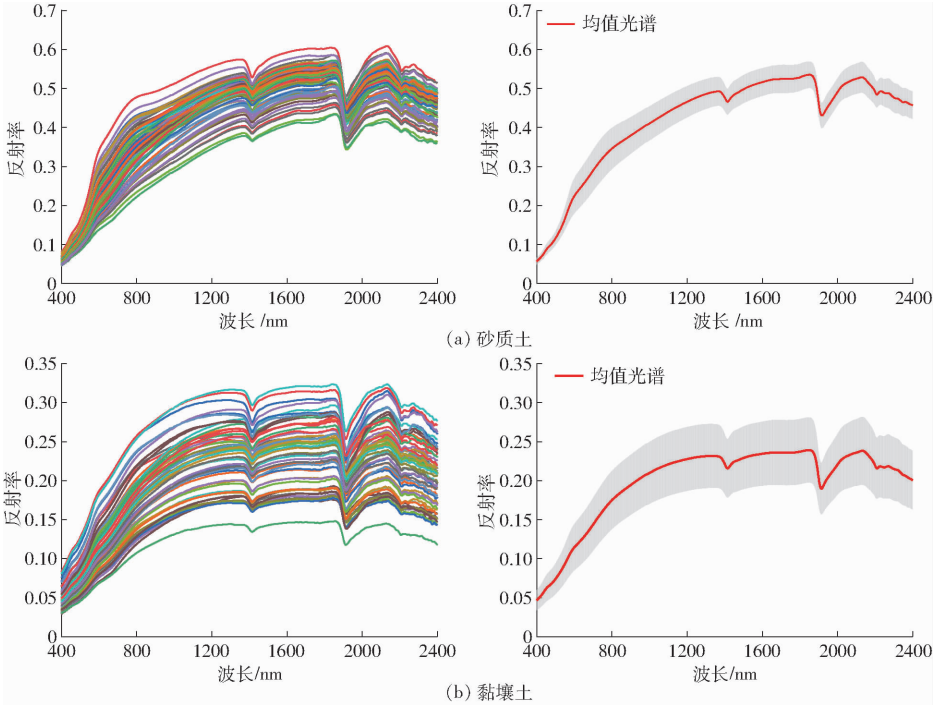

图 2 平滑后的土样反射率曲线

Fig. 2 Reflectance curves of soil samples after smoothing

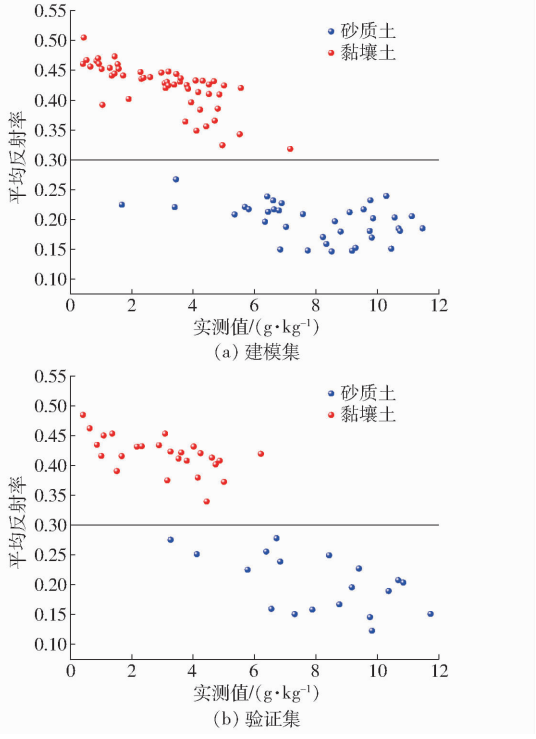

图 3　不同有机质含量土壤平均反射率散点图

Fig. 3　Scatter plots of average reflectance values of soils with different organic matter contents

1.3.2　分数阶微分

FOD 是数学理论的分支，其将经典的整数阶微分拓展至任意阶微分。而 Grünwald－Letnikov 是一种离散形式的定义，便于进行数值计算，且运算效率较高[10]，故本文采用此表达式，具体为

$$\mathrm{d}^q f(x) = \lim_{h\to\infty}\frac{1}{h^q}\sum_{m=0}^{\frac{t-a}{h}}(-1)^m\frac{\Gamma(q+1)}{m!\ \Gamma(q-m+1)}f(x-mh) \tag{2}$$

式中　d——微分函数　　q——微分阶数

h——步长　　t、a——微分的上限和下限

其中 Gamma(Γ)函数[15]为

$$\Gamma(\beta) = \int_0^\infty \mathrm{e}^{-t}t^{\beta-1}\mathrm{d}t = (\beta-1)! \tag{3}$$

式中　β——任意变量(本文指微分阶数)

设 $f(\lambda)$ 为一维光谱，$[a,t]$ 为波长区间，$\lambda\in[a,t]$，将波长区间按单位 h 进行等分，令 $h=1$，$n=\frac{t-a}{h}=t-a$，则由式(2)推导出分数阶微分的差值表达式为

$$\frac{\mathrm{d}^\nu f(\lambda)}{\mathrm{d}\lambda^\nu}\approx f(\lambda)+(-\nu)f(\lambda-1)+\frac{(-\nu)(-\nu+1)}{2}f(\lambda-2)+\cdots+\frac{\Gamma(-\nu+1)}{n!\ \Gamma(-\nu+n+1)}f(\lambda-n) \tag{4}$$

其中 ν 为 1 或 2 时，式(4)与一、二阶微分的方程相同；而当 $\nu=0$ 时，表示未对光谱数据进行处理。本文利用式(4)分别计算两种土壤的 0～2 阶微分(间隔 0.2 阶)。

1.3.3　光谱指数

基于经分数阶微分处理后的光谱数据，构建归一化指数(Normalized difference index，NDI)，其计算式为

$$NDI = \frac{A_{\lambda1}-A_{\lambda2}}{A_{\lambda1}+A_{\lambda2}} \tag{5}$$

式中　$A_{\lambda1}$、$A_{\lambda2}$——400～2 400 nm 中任意两波长，且 $A_{\lambda1}\neq A_{\lambda2}$

分析不同分数阶微分下的 NDI 与 SOM 的相关性，设立阈值，确定估算 SOM 的敏感指数。

1.4　支持向量机分类和随机森林模型

SVMDA 具有强大的小样本学习能力以及在高维空间中良好的扩展性能，因而广泛用于分类[16]。RF 是 BREIMAN[17]提出的一个组合分类器算法，其主要思想是用自助法重采样技术，基于原始样本集，生成若干个自助样本集，并将每个自助样本集作为每棵分类树的所有训练数据。该模型是多个分类树的集合，可以估计独立变量之间复杂的非线性关系并且响应变量。有研究表明，RF 预测性能较强，特别在高光谱技术估算土壤相关特征含量方面[18]。因而本文采用此法来预测 SOM。

1.5　建模集和验证集的划分

SVMDA 建模，以随机抽样法取 89 个样本作为建模集，剩余 45 个样本作为验证集。RF 建模采用 Kennard－Stone(K－S)算法划分建模集、验证集，结果如图 4 所示，图中 C_V 为变异系数，Mean 为平均值，SD 为标准差。

1.6　模型建立与验证

以土壤平均光谱反射率作为自变量，土壤质地类型作为因变量，采用非线性的 SVMDA，建立土壤质地判别模型，以 leave one out 法进行交叉验证。再对不同质地的土壤 SOM 分别建模，通过 0 阶 NDI 的最高决定系数作为阈值获取敏感指数，以所获敏感指数作为自变量，SOM 作为因变量，建立不同分数阶的非线性 RF 模型。

SVMDA 以分类准确率进行评价，分类准确率越大，表明分类越精准；RF 模型精度通过决定系数 R^2、均方根误差(RMSE)、相对分析误差(RPD)等参数评价，建模集决定系数 R_C^2、验证集决定系数 R_P^2 趋向于 1，则模型稳定性越好；建模集均方根误差(RMSEC)及验证集均方根误差(RMSEP)越小表明模型精度越高。当 RPD 小于 1 时，模型不可用；当

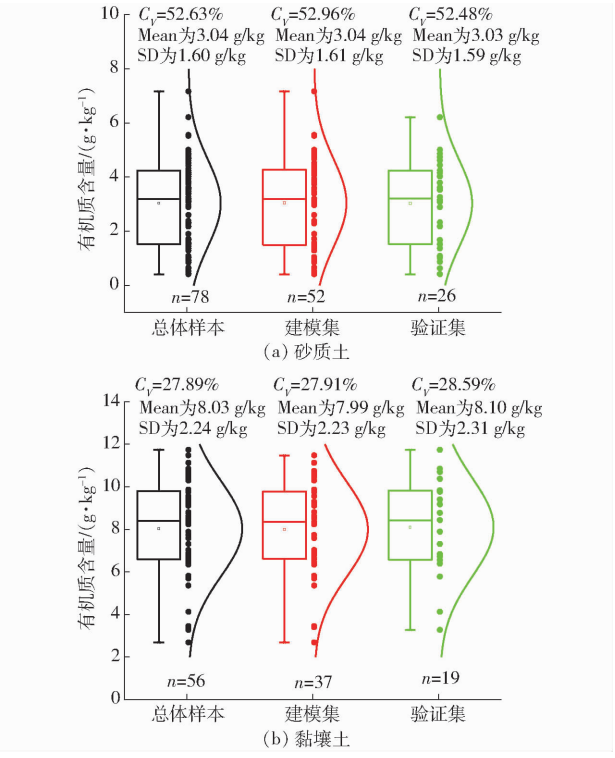

图 4 土壤有机质含量描述性统计特征箱线图

Fig. 4 Box plots of density and descriptive statistics of SOM

RPD 大于等于 1、小于 1.4 时,模型不佳,只能对样本粗略估计;当 RPD 大于等于 1.4、小于 1.8 时表明模型具有一定预测能力,可以对样本进行评估;当 RPD 大于等于 1.8、小于 2 时模型具有较好的预测能力;当 RPD 大于等于 2、小于 2.5 时表示模型具有很好的预测能力,可以定量预测;当 RPD 大于等于 2.5 时表示模型优秀,具有极好的预测能力[19-20]。

1.7 数据分析软件

Savitzky-Golay 算法、平均反射率、分数阶微分、光谱指数计算和筛选、K-S 算法及 SVMDA 均在 Matlab R2018a 平台进行。图表的制作在 OriginPro 2017C 中实现。RF 通过 R3.5.2 软件中的 Random Forest 软件包实现和优化,具体参数设置有分类树数目(ntree)、训练子集数目(mtry)。本文中 ntree 为 500,mtry 设置为 1/3 总变量数。

2 结果与分析

2.1 曲线特征分析

2.1.1 土壤样本高光谱曲线特征

图 5 为 2 种土壤不同 SOM 的光谱反射率变化曲线,可以看出,在 400~1 800 nm 内,砂质土反射率曲线斜率大,变化快,黏壤土反射率曲线斜率较小,变化慢;土样虽已风干,但残留水分对光谱的作用依然显著,在 1 400、1 900、2 200 nm 附近存在水分吸收谷[21];在 SOM 基本相同时,黏壤土的反射率明显小于砂质土;在全波段范围内,就同一种土壤而言,随着 SOM 的增加,光谱反射率逐渐降低。

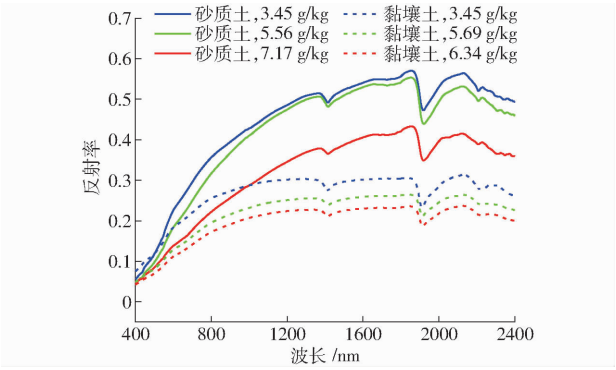

图 5 不同土壤有机质含量光谱反射率变化曲线

Fig. 5 Reflectance spectra curves of soils with different SOM

2.1.2 分数阶微分光谱曲线特征

利用式(4)计算 2 种土壤 0~2 阶分数阶微分(间隔 0.2 阶),经预处理后的光谱如图 6、7 所示。图 6a~6f、7a~7f 显示,当阶数从 0 增加到 1 时,在 580 nm 处的负峰有所增加,此峰主要受 SOM 的影响,水分子振动作用逐渐凸显,在 1 400、1 900 nm 处的吸收谷更加明显[22]。由图 6d~6f、7d~7f 可知,可见光区域的光谱吸收特征不断增强,特别在 430、550 nm 处有 2 个正峰,而在 480 nm 处有 1 个负峰。由图 6e~6k、7e~7k 可知,在 0.8 阶以后,与水分有关的 2 200 nm 峰值较快增加,此处峰值也受 SOM 的影响,因为 SOM 越大,土壤吸附水分的能力就越强[23]。分数阶微分光谱与原始光谱相比,可以改善光谱曲线分辨率,深度挖掘与 SOM 有关的潜在信息。

2.2 建模与分析

2.2.1 土壤质地分类准确率分析

SVMDA 对土壤质地的分类判别效果中,建模集和验证集准确率都达到 100%。这也表示土壤质地对土壤光谱反射率有显著的影响。而平均反射率可以较好地表示砂质土和黏壤土对光谱响应的差异,用此模型可以精确地区分砂质土和黏壤土,为后续针对不同土壤质地建立 SOM 模型,并进行组合,提供了有效支撑。

2.2.2 SOM 与不同阶数 NDI 的二维相关性分析

由图 8 可知,砂质土决定系数高于黏壤土,这可能与土壤质地有较大关系,黏壤土的反射率较小,导致光谱中所含的 SOM 信息无法凸显。

分数阶微分对于反射率较小的黏壤土作用更加明显,将决定系数从 0.763 提高到 0.820。而对于反射率较大的砂质土作用较小,只将决定系数从 0.901 提高到 0.917。在 NDI 指数中相关性较好的波段主要集中在可见光和短波近红外区域(400~1 000 nm)(图 8),这与 HONG 等[23]的研究结果一

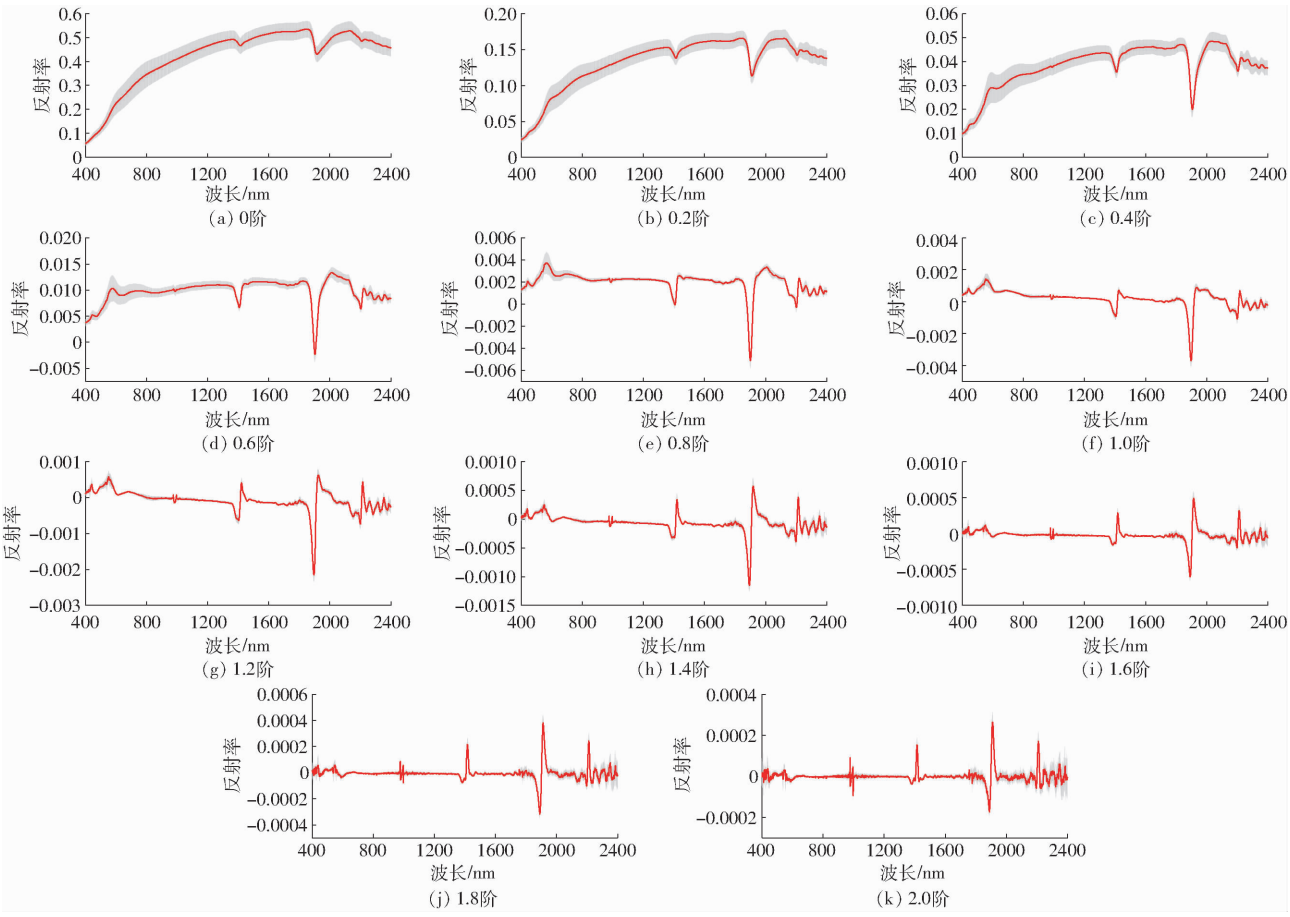

图 6　砂质土光谱反射率分数阶微分曲线

Fig. 6　Spectral reflectance fractional differential curves of sandy soil

致。而砂质土、黏壤土的 NDI 指数与 SOM 的决定系数分别在 0.8、0.6 阶时，达到最大值，但在 1.4 阶以后，分数阶微分对 NDI 的效果降低，这与 WANG 等[12]的研究结果基本一致。提取两种土壤的每个分数阶最佳组合波长，并计算 NDI 与 SOM 的决定系数，结果见表 1。砂质土与黏壤土 NDI 的最佳组合波长都不相同，但砂质土的组合波长都稍大于黏壤土。在 0.4～1.2 阶微分时，黏壤土最佳组合波长都集中在可见光区域。而砂质土的所有最佳组合波长中，均含有至少 1 个红外波段。这表明黏壤土 SOM 对可见光区域更敏感，而砂质土 SOM 对红外区域更敏感。

2.2.3　敏感光谱指数筛选

为获得通过分数阶微分所提升的敏感指数 NDI，以 0 阶最佳 NDI 所对应的决定系数作为阈值，筛选富含 SOM 信息的敏感指数。由表 1 可知，砂质土的筛选阈值为 0.901，组合波长为 R2210、R2170；黏壤土的筛选阈值为 0.763，组合波长为 R860、R850。

筛选结果如图 9 所示，两种土壤均在 0.6 阶达到峰值，且黏壤土提升效果优于砂质土，差异的造成与土壤质地有较大关系。这说明分数阶微分耦合光谱指数均能够有效放大高光谱与 SOM 有关的潜在信息，尤其当阶数为 0.6 时，可以获得最多高相关性的自变量，以此建模可进一步提高模型性能，且此法对提取被其他土壤属性遮蔽的 SOM 信息十分有效，且对遮蔽作用越大（决定系数小）的土壤，其提升效果越明显。

2.2.4　敏感光谱指数建模

基于不同阶数的敏感 NDI 所建 RF 模型参数见表 2，表中 N 为模型自变量数目。为与敏感指数 RF 模型对比，建立不同土质的全波段 RF 模型（Full spectrum，FS），其结果见表 2。敏感 NDI 所建模型均具有较好的效果，均远高于全波段模型。总体而言，利用敏感指数建模，可以一定程度上提高预测效果。RPD 随阶数先增加后减小，但不同质地的土壤达到峰值的阶数不同，砂质土 1.2 阶模型（SS－1.2－NDI）的 RPD 最大，可达 3.658；黏壤土 0.6 阶模型（CLS－0.6－NDI）RPD 最大，可达 4.316。此差异可能与土壤质地有关。而基于全波段的不同土质 FS 模型中，砂质土模型较优，RPD 为 1.579，与敏感指数所建模型有较大差距。

将 SS－1.2－NDI 与 CLS－0.6－NDI 组合，建立 SVMDA－RF 模型，并计算其模型评价参数，结果

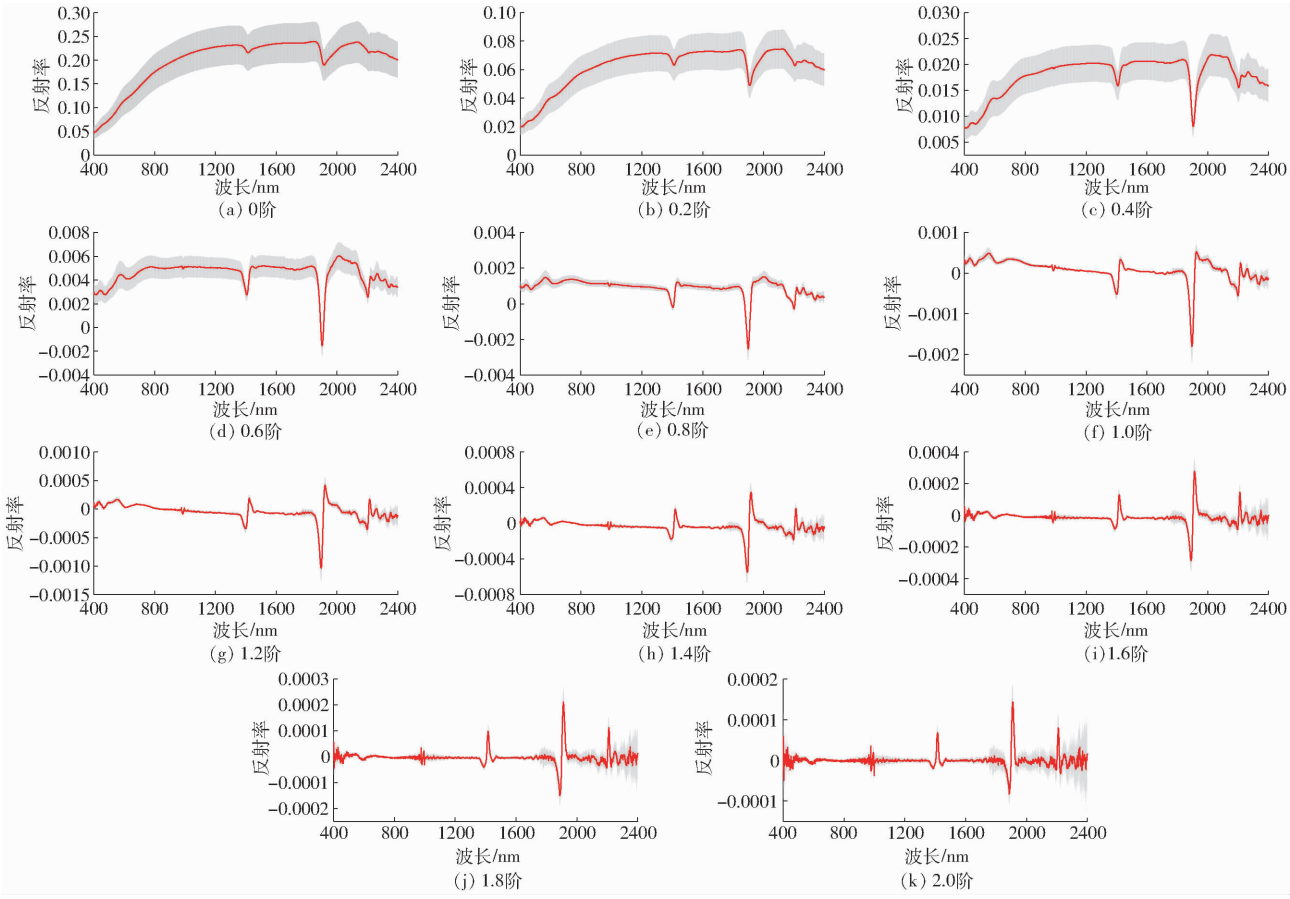

图 7 黏壤土光谱反射率分数阶微分曲线

Fig. 7 Spectral reflectance fractional differential curves of clay loam soil

见表 3。为便于比较,建立将 SS－FS、CLS－FS 组合的 SVMDA－RF 模型,以及不区分土质(Unclassified samples,US)的全波段 RF 模型。从表 3 可以看出,SVMDA－RF(SS－1.2－NDI、CLS－0.6－NDI)效果最好,SVMDA－RF(SS－FS、CLS－FS)次之,RF(US)效果最差。且 SVMDA－RF(SS－1.2－NDI、CLS－0.6－NDI)其 RPD 可达 7.004,能精确预测 SOM。而 RF(US)的 RPD 只有 2.228,只具有一定的预测能力。这表明区分土质分别建模并二次组合可以提高建模效果,结合敏感指数,能使模型具有较高精度,实现精确预测 SOM。

SVMDA－RF(SS－1.2－NDI、CLS－0.6－NDI)的建立不仅可以快速区分土壤质地类型,还能够针对不同质地的土壤精准预测土壤 SOM。但如果土壤种类过多时,SVMDA 区分土壤质地的正确率可能会下降,这将影响 SVMDA－RF 模型的稳定及其精度,因此需要寻找不同土质对光谱影响的特点,以此作为分类依据,为后续的 RF 模型精度提供保障。

高光谱数据具有高维度和数据冗余的特点,导致难以从中提取到与 SOM 相关性高的敏感数据。而且在 SOM 低于 20 g/kg 时,土壤中的其他组成物质(水分、盐分、矿物等)和土壤质地的光谱反射将占主导,其中土壤含水率对光谱的作用较强。本文已将土样经过风干处理、光谱重采样和设置阈值剔除低相关性的指数,一定程度上消除了水分对光谱特征的影响。如果预测变量和响应变量之间存在非线性关系,则 RF 这类非线性模型通常会有较好的拟合效果,产生优异的估计精度。如 WANG 等[24]所建土壤盐分的 RF 与偏最小二乘模型,RF 优于线性的偏最小二乘模型,最佳模型的验证效果较好,其 RPD 可达 2.78。

图 10 中,RF(US)模型的预测值与实测值在 1∶1线上偏差较大,点的离散程度高。但利用 SVMDAD 对土质分类,再建模其精度得到明显提升。由图 10c、10f 可知,SVMDA－RF(SS－1.2－NDI、CLS－0.6－NDI)的实测值与预测值的散点拟合线与 1∶1直线基本重合,达到极高的精度,可以为高光谱遥感监测土壤荒漠程度提供较为可靠的依据。由图 10a、10b、10d、10e 可知,黏壤土比砂质土的点更为分散,这与黏壤土的吸附特性优于砂质土有关。虽然经长时间的风干,但黏壤土可能还会吸附一些水分等物质。这些物质对于光谱的影响,更加掩盖低含量 SOM 对于光谱的作用。故而有些研究者[1]会将 1 400、1 900、2 200 nm 附近存在的 3 个水分吸收谷

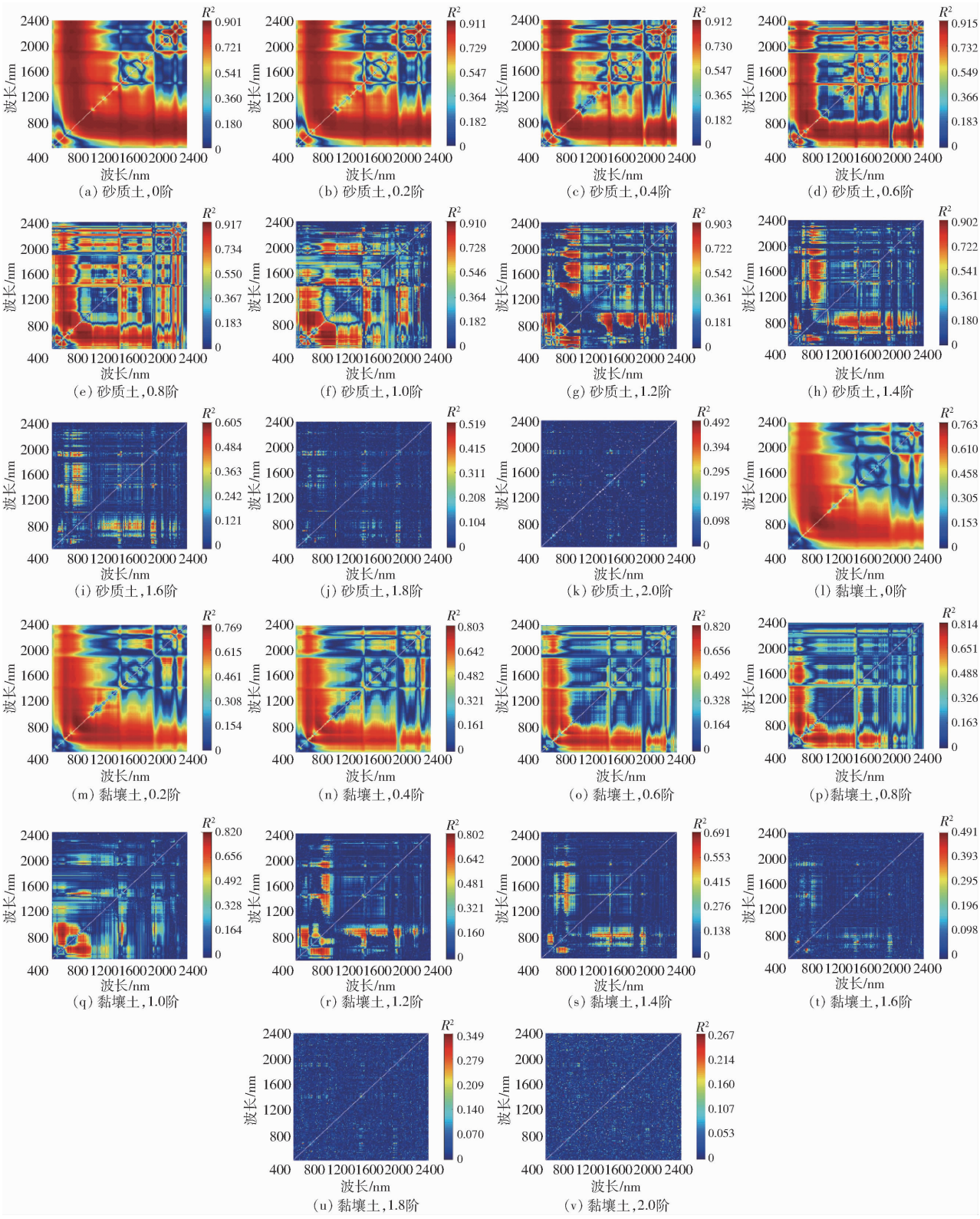

图 8　砂质土和黏壤土光谱反射率各阶微分的归一化光谱指数与有机质含量的决定系数

Fig. 8　Determination coefficients between SOM and normalized spectral index of sandy soil and clay loam soil reflectance treated by fractional order derivatives

作水汽吸收带的剔除处理，以减少对有机质反演的影响。但是其附近可能含有 SOM 信息的波段，利用这些波长可能会提升有机质反演的精度。如本文砂质土的 1.2 阶最佳 NDI 含有 1 420 nm 波长，然而利用其建模效果较好，RPD 可达 3.65。

3　讨论

高光谱数据具有高维度和数据冗余的特点，导致难以从中提取到与 SOM 相关性高的敏感数据。而且在 SOM 低于 20 g/kg 时，土壤中的其他组成物

表 1 砂质土与黏壤土在不同阶数下的最佳归一化指数及其决定系数

Tab. 1 Determination coefficients between SOM and optimal normalized spectral index for sandy soil and clay loam under different fractional order derivatives

阶数	砂质土		黏壤土	
	NDI 最佳组合波长	R^2	NDI 最佳组合波长	R^2
0	R2210、R2170	0.901	R860、R850	0.763
0.2	R2210、R2170	0.911	R800、R780	0.769
0.4	R2210、R2170	0.912	R640、R580	0.803
0.6	R820、R670	0.915	R640、R560	0.820
0.8	R800、R670	0.917	R640、R540	0.814
1.0	R850、R570	0.910	R670、R560	0.820
1.2	R1420、R680	0.903	R670、R560	0.802
1.4	R1420、R670	0.902	R670、R560	0.691
1.6	R1890、R560	0.605	R1410、R660	0.491
1.8	R1890、R1410	0.519	R660、R550	0.349
2.0	R1730、R1410	0.492	R2330、R1440	0.267

注:R2210 表示波长为 2 210 nm,以此类推。

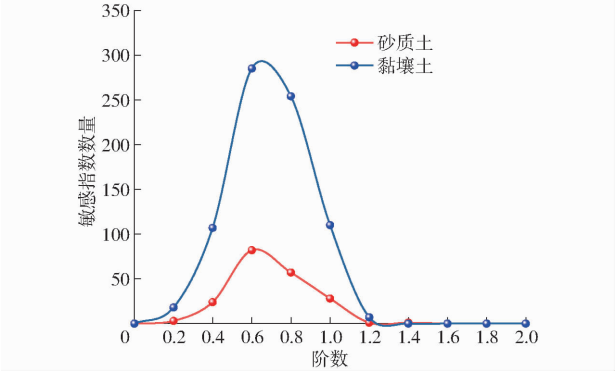

图 9 敏感指数数量随阶数的变化曲线

Fig. 9 Changing curves of number of sensitive indices under different fractional order derivatives

质(水分、盐分、矿物等)和土壤质地的光谱反射将占主导,其中土壤质地和黏粒含量是影响土壤光谱反射的重要因素之一[25]。故而为了减小土壤质地对定量反演 SOM 的影响,利用平均反射率建立 SVMDA 模型,其对所研究区域的土壤质地分类准确率

表 2 基于 RF 的 SOM 建模与预测结果

Tab. 2 Calibration and validation results of SOM based on RF

土壤类型	参数类型	阶数	N	R_C^2	RMSEC/(g·kg^{-1})	R_P^2	RMSEP/(g·kg^{-1})	RPD
砂质土	NDI	0.2	3	0.983	0.216	0.857	0.605	2.679
		0.4	24	0.987	0.198	0.874	0.590	2.791
		0.6	82	0.987	0.194	0.855	0.612	2.633
		0.8	57	0.984	0.208	0.869	0.615	2.692
		1.0	28	0.983	0.216	0.852	0.672	2.514
		1.2	1	0.962	0.317	0.920	0.435	3.658
		1.4	1	0.969	0.284	0.855	0.688	2.452
	FS		201	0.970	0.288	0.742	0.813	1.579
黏壤土	NDI	0.2	18	0.901	0.694	0.928	0.670	3.116
		0.4	107	0.936	0.580	0.942	0.572	3.971
		0.6	285	0.942	0.534	0.944	0.554	4.316
		0.8	254	0.944	0.533	0.941	0.568	4.055
		1.0	110	0.938	0.559	0.926	0.642	3.436
		1.2	7	0.918	0.643	0.925	0.736	2.683
	FS		201	0.861	0.842	0.708	1.402	1.005

表 3 分类组合模型与不分类模型的 SOM 建模与预测结果

Tab. 3 Calibration and validation results of SOM based on classification combination models and non-classification model

建模类型	N	R_C^2	RMSEC/(g·kg^{-1})	R_P^2	RMSEP/(g·kg^{-1})	RPD
SVMDA-RF(SS-1.2-NDI、CLS-0.6-NDI)	1、285	0.980	0.419	0.979	0.481	7.004
SVMDA-RF(SS-FS、CLS-FS)	201、201	0.961	0.597	0.903	1.092	3.086
RF(US)	201	0.843	1.187	0.832	1.512	2.228

可达 100%。虽然以往研究表明利用光谱技术可以较好分辨不同的土壤质地,但是所建模型自变量数目多,较为复杂。如宋海燕等[26]通过正交信号校正(OSC)处理砂土、壤土、黏土 3 类土壤光谱,并结合偏最小二乘(PLS)对土壤质地分类,其正确率最高达 93.3%,但无法很好区分壤土与砂土;WANG 等[25]利用土壤吸光度对基于垂直干旱指数(PDI)分组的样本,建立 PLS 分类模型,可以较好区分砂土与黏土,并预测黏粒含量。而本文利用平均反射率建立相对简单的分类模型,且分类效果更好。这表明平均反射率用于砂质土和黏壤土分类是可行的。

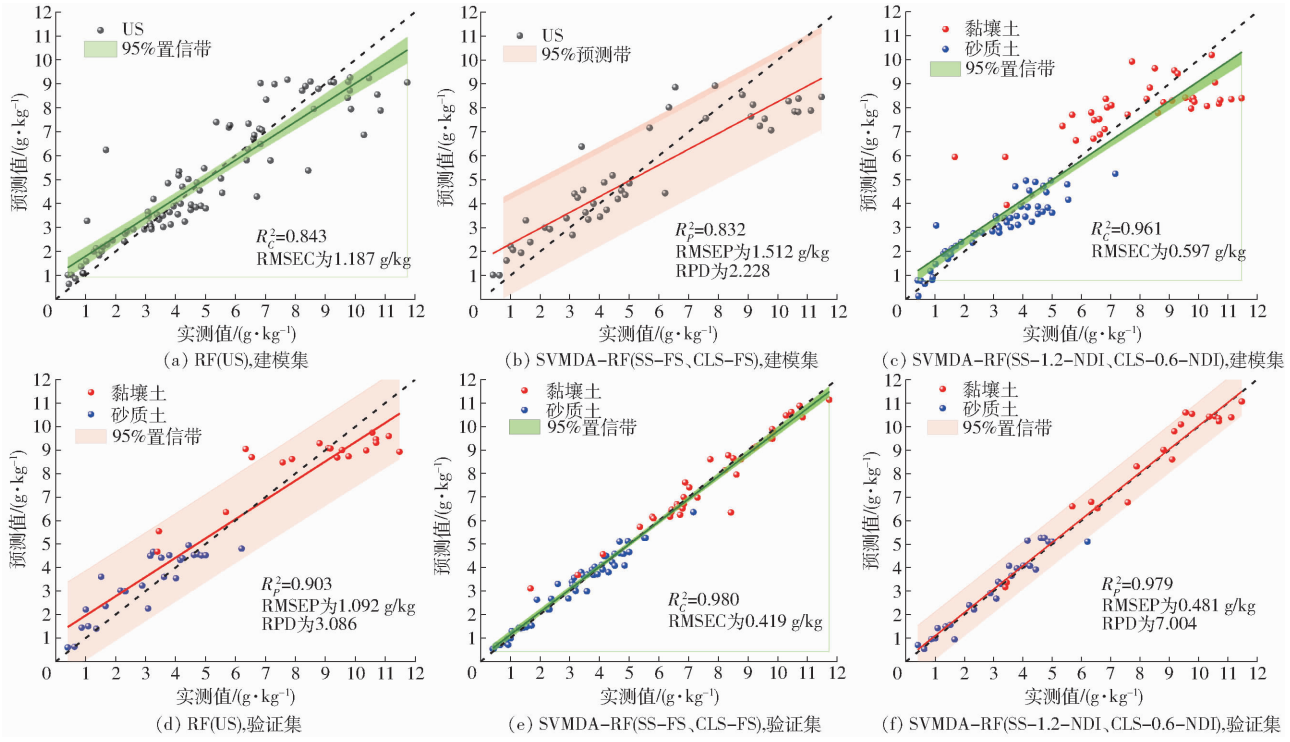

图 10　基于 RF 的 SOM 预测值与实测值的比较

Fig. 10　Comparison of measured and predicted SOM values based on RF

洪永胜等[27]分析基于 4 种不同光谱变换的归一化光谱指数与 SOM 的二维相关性，发现 R^2 最高达 0.55。郭燕等[28]分析 NDI、DI、RI 与 SOM 的二维相关性，得到最大相关系数均在 0.5 ~ 0.7 之间。HONG 等[23]分析基于 6 种不同光谱变换的光谱指数与 SOM 的二维相关性，得到最大相关系数为 0.81。虽然二维相关分析能从二维光谱空间中更精准地捕捉到与 SOM 有关的光谱特征，有效减小土壤中其他组成物质（包括水分）对预测 SOM 的影响，但上述研究均未结合分数阶微分，相关系数有待提高。而本文分析基于分数阶微分的光谱指数与 SOM 的相关性，得到决定系数最高可达 0.917。因此光谱数据先经过分数阶微分处理，可以扩大二维光谱的空间，为获取高相关性敏感指数提供了更大可能。且以 0 阶决定系数筛选指数可以获得更多特征信息，有利于提高建模的效果。

侯艳军等[1]所建最佳荒漠土壤 SOM 模型，为偏最小二乘回归模型，其 R_P^2 = 0.76。王海峰等[9]所建最佳荒漠土壤 SOM 模型，为岭回归模型，其 R_P^2 = 0.866，RPD 为 2.79。高志海等[29]所建最佳荒漠土壤 SOM 模型，为一元线性回归模型，其 R_P^2 达 0.882。上述学者的研究均使用线性模型预测荒漠土壤 SOM，虽然 PLSR、岭回归可以有效地解决多重共线性问题，但只是在一些特定的土壤属性和相应的光谱特征反射率之间模拟潜在的线性关系，得到相应的线性模型。况且土壤性质的分布大多是偏态分布而不是标准正态分布，如图 2 显示砂质土 SOM 分布是左偏、黏壤土 SOM 分布是右偏，在此情况下 PLSR 等线性回归模型拟合效果一般。本文利用 RF 可以估计独立变量之间复杂的非线性关系且响应变量的特性，对筛选后的敏感指数建立模型，发现 RF 模型在本文中的效果明显，尤其是黏壤土 RF 模型，其 R_P^2 达 0.944，RPD 达 4.316。因此，如果预测变量和响应变量之间存在非线性关系，则 RF 这类非线性模型通常会有较好的拟合效果，产生优异的估计精度。如 WANG 等[24]所建土壤盐分的 RF 与 PLSR 模型，RF 优于 PLSR，最佳模型的验证效果较好，其 RPD 可达 2.78。

高光谱遥感反演的难点在于反演变量不完全是主导遥感光谱信息的因子，特别是在反演变量含量较低时，仅可以为遥感光谱提供微弱信号。而本文利用 FOD 联合二维相关分析的方法可以得到高相关性的敏感指数。不过阈值的设立具有一定主观性，需要更严谨客观的方法进行筛选。今后的研究可以利用变量投影重要性分析法[30]、遗传算法[31]等来筛选敏感指数。所建立的 SVMDA－RF（SS－1.2－NDI、CLS－0.6－NDI）和 SVMDA－RF（SS－FS、CLS－FS）都具有极好的预测性能，但是模型的普适性还需要大量其他地区荒漠土壤 SOM 数据来验证。此外，本文只进行一种非线性模型（RF）的尝试，以后的研究中可以建立不同非线性模型进行组合，比较获得适应性更佳的模型。

4 结论

(1)基于平均反射率的 SVMDA 可以对土壤质地精确分类,将其与 RF 结合,可以提升荒漠土壤 SOM 的预测精度。

(2)0 阶 NDI 的最高决定系数作为阈值筛选敏感指数,可以获得经分数阶微分提升的 NDI,其数量达到峰值的阶数均为 0.6 阶,但峰值与土壤质地有关。

(3)利用分数阶微分结合 NDI 可获得高相关性的敏感指数,以此建模可以极大提升模型的整体效果。

参 考 文 献

[1] 侯艳军,塔西甫拉提·特依拜,买买提·沙吾提,等. 荒漠土壤有机质含量高光谱估算模型[J]. 农业工程学报, 2014, 30(16): 113 – 120.
HOU Yanjun, TASHPOLAT·Tiyip, MAMAT·Sawut, et al. Estimation model of desert soil organic matter content using hyperspectral data[J]. Transactions of the CSAE, 2014,30(16):113 – 120. (in Chinese)

[2] 叶勤,姜雪芹,李西灿,等. 基于高光谱数据的土壤有机质含量反演模型比较[J/OL]. 农业机械学报, 2017, 48(3): 164 – 172.
YE Qin, JIANG Xueqin, LI Xican, et al. Comparison on inversion model of soil organic matter content based on hyperspectral data[J/OL]. Transactions of the Chinese Society for Agricultural Machinery, 2017,48(3):164 – 172. http://www.j-csam.org/jcsam/ch/reader/view_abstract.aspx? file_no = 20170321&flag = 1&journal_id = jcsam. DOI:10.6041/j.issn.1000-1298.2017.03.021. (in Chinese)

[3] 尼加提·卡斯木,茹克亚·萨吾提,师庆东,等. 基于优化光谱指数的土壤有机质含量估算[J/OL]. 农业机械学报, 2018, 49(11): 155 – 163.
NIJAT Kasim, RUKEYA Sawut, SHI Qingdong, et al. Estimation of soil organic matter content based on optimized spectral index[J/OL]. Transactions of the Chinese Society for Agricultural Machinery, 2018,49(11):155 – 163. http://www.j-csam.org/jcsam/ch/reader/view_abstract.aspx? file_no = 20181118&flag = 1&journal_id = jcsam. DOI:10.6041/j.issn.1000-1298.2018.11.018. (in Chinese)

[4] 何东健,陈煦. 土壤有机质含量田间实时测定方法[J/OL]. 农业机械学报, 2015, 46(1): 127 – 132.
HE Dongjian, CHEN Xu. Real-time measurement of soil organic matter content in field[J/OL]. Transactions of the Chinese Society for Agricultural Machinery, 2015,46(1):127 – 132. http://www.j-csam.org/jcsam/ch/reader/view_abstract.aspx? file_no = 20150119&flag = 1&journal_id = jcsam. DOI:10.6041/j.issn.1000-1298.2015.01.019. (in Chinese)

[5] HUMMEL J W, SUDDDTH K A, HOLLINGER S E. Soil moisture and organic matter prediction of surface and subsurface soils using an NIR soil sensor[J]. Computers and Electronics in Agriculture, 2001, 32(2): 149 – 165.

[6] RINNAN R, RINNAN Å. Application of near infrared reflectance (NIR) and fluorescence spectroscopy to analysis of microbiological and chemical properties of arctic soil[J]. Soil Biology and Biochemistry, 2007, 39(7): 1664 – 1673.

[7] 刘磊,沈润平,丁国香. 基于高光谱的土壤有机质含量估算研究[J]. 光谱学与光谱分析, 2011, 31(3): 762 – 766.
LIU Lei, SHEN Runping, DING Guoxiang. Studies on the estimation of soil organic matter content based on hyper-spectrum [J]. Spectroscopy and Spectral Analysis, 2011, 31(3): 762 – 766. (in Chinese)

[8] 谢文,赵小敏,郭熙,等. 基于 RBF 组合模型的山地红壤有机质含量光谱估测[J]. 林业科学, 2018, 54(6): 16 – 23.
XIE Wen, ZHAO Xiaomin, GUO Xi, et al. Spectrum based estimation of the content of soil organic matters in mountain red soil using RBF combination model[J]. Scientia Silvae Sinicae, 2018, 54(6): 16 – 23. (in Chinese)

[9] 王海峰,张智韬,ARNON Karnieli,等. 基于灰度关联–岭回归的荒漠土壤有机质含量高光谱估算[J]. 农业工程学报, 2018,34(14): 124 – 131.
WANG Haifeng, ZHANG Zhitao, ARNON Karnieli, et al. Hyperspectral estimation of desert soil organic matter content based on gray correlation – ridge regression model[J]. Transactions of the CSAE, 2018, 34(14): 124 – 131. (in Chinese)

[10] 张东. 分数阶微分在土壤盐渍化遥感监测中的应用研究[D]. 乌鲁木齐:新疆大学, 2017.

[11] HONG Y, LIU Y, CHEN Y, et al. Application of fractional – order derivative in the quantitative estimation of soil organic matter content through visible and near-infrared spectroscopy[J]. Geoderma, 2019, 337: 758 – 769.

[12] WANG X, ZHANG F, KUNG H, et al. New methods for improving the remote sensing estimation of soil organic matter content (SOMC) in the Ebinur Lake Wetland National Nature Reserve (ELWNNR) in northwest China[J]. Remote Sensing of Environment, 2018, 218: 104 – 118.

[13] PEN-MOURATOV S, MYBLAT T, SHAMIR I, et al. Soil biota in the arava valley of Negev desert, Israel[J]. Pedosphere, 2010, 20(3): 273 – 284.

[14] 洪永胜,于雷,耿雷,等. 应用 DS 算法消除室内几何测试条件对土壤高光谱数据波动性的影响[J]. 华中师范大学学报(自然科学版), 2016, 50(2): 303 – 308.
HONG Yongsheng, YU Lei, GENG Lei, et al. Using direct standardization algorithm to eliminate the effect of laboratory geometric parameters on soil hyperspectral data fluctuate characteristic[J]. Journal of Central China Normal University(Natural Sciences), 2016, 50(2): 303 – 308. (in Chinese)

[15] KARAAGAC B. New exact solutions for some fractional order differential equations via improved sub-equation method[J]. Discrete and Continuous, 2019, 12(3): 447－454.

[16] 黄玉萍,LU Renfu,戚超,等. 基于空间可分辨光谱的番茄成熟度判别方法研究[J]. 光谱学与光谱分析, 2018, 38(7): 2183－2188.
HUANG Yuping, LU Renfu, QI Chao, et al. Tomato maturity classification based on spatially resolved spectra [J]. Spectroscopy and Spectral Analysis, 2018, 38(7): 2183－2188. (in Chinese)

[17] BREIMAN L. Random forests[J]. Machine Learning, 2001, 45: 5－32.

[18] SHI T, HORVATH S. Unsupervised learning with random forest predictors[J]. Journal of Computational and Graphical Statistics, 2012, 15(1): 118－138.

[19] VISCARRA R R A, WALVOORT D J J, MCBRATENEY A B, et al. Visible, near infrared, mid infrared or combined diffuse reflectance spectroscopy for simultaneous assessment of various soil properties[J]. Geoderma, 2006, 131(1－2): 59－75.

[20] GHOLIZADEH A, BORUVKA L, SABERIOON M, et al. A memory-based learning approach as compared to other data mining algorithms for the prediction of soil texture using diffuse reflectance spectra[J]. Remote Sensing, 2016, 8(4): 341.

[21] DOTTO A C, DALMOLIN R S D, TEN C A, et al. A systematic study on the application of scatter-corrective and spectral-derivative preprocessing for multivariate prediction of soil organic carbon by Vis－NIR spectra[J]. Geoderma, 2018, 314: 262－274.

[22] MOHAMED E S, SALEH A M, BELAL A B, et al. Application of near-infrared reflectance for quantitative assessment of soil properties[J]. The Egyptian Journal of Remote Sensing and Space Science, 2018, 21(1): 1－14.

[23] HONG Y, CHEN S, ZHANG Y, et al. Rapid identification of soil organic matter level via visible and near-infrared spectroscopy: effects of two-dimensional correlation coefficient and extreme learning machine[J]. Science of the Total Environment, 2018, 644: 1232－1243.

[24] WANG J, DING J, ABULIMITI A, et al. Quantitative estimation of soil salinity by means of different modeling methods and visible-near infrared (Vis－NIR) spectroscopy, Ebinur Lake Wetland, Northwest China[J]. PeerJ, 2018, 6: e4703.

[25] WANG D, ZHANG G, ROSSITER D G, et al. The prediction of soil texture from visible-near-infrared spectra under varying moisture conditions[J]. Soil Science Society of America Journal, 2016, 80(2): 420.

[26] 宋海燕,秦刚,韩小平,等. 基于近红外光谱和正交信号-偏最小二乘法对土壤的分类[J]. 农业工程学报, 2012, 28(7):168－171.
SONG Haiyan, QIN Gang, HAN Xiaoping, et al. Soil classification based on near infrared reflectance spectroscopy and orthogonal signal correction-partial least square[J]. Transactions of the CSAE, 2012, 28(7): 168－171. (in Chinese)

[27] 洪永胜,朱亚星,苏学平,等. 高光谱技术联合归一化光谱指数估算土壤有机质含量[J]. 光谱学与光谱分析, 2017, 37(11):3537－3542.
HONG Yongsheng, ZHU Yaxing, SU Xueping, et al. Estimation of soil organic matter content using hyperspectral techniques combined with normalized difference spectral index [J]. Spectroscopy and Spectral Analysis, 2017, 37(11): 3537－3542. (in Chinese)

[28] 郭燕,纪文君,吴宏海,等. 基于野外 Vis－NIR 光谱的土壤有机质预测与制图[J]. 光谱学与光谱分析, 2013, 33(4): 1135－1140.
GUO Yan, JI Wenjun, WU Honghai, et al. Estimation and mapping of soil organic matter based on Vis－NIR reflence spectroscopy [J]. Spectroscopy and Spectral Analysis, 2013, 33(4): 1135－1140. (in Chinese)

[29] 高志海,白黎娜,王琦瑜,等. 荒漠化土地土壤有机质含量的实测光谱估测[J]. 林业科学, 2011, 47(6): 9－16.
GAO Zhihai, BAI Li'na, WANG Bengyu, et al. Estimation of soil organic matter content in desertified lands using measured soil spectral data[J]. Scientia Silvae Sinicae, 2011,47(6):9－16. (in Chinese)

[30] WANG H, CHEN Y, ZHANG Z, et al. Quantitatively estimating main soil water-soluble salt ions content based on visible-near infrared wavelength selected using GC, SR and VIP[J]. PeerJ, 2019, 7: e6310.

[31] HONG Y, CHEN Y, YU L, et al. Combining fractional order derivative and spectral variable selection for organic matter estimation of homogeneous soil samples by Vis－NIR spectroscopy[J]. Remote Sensing, 2018, 10(3): 479.