

基于谐波分析算法的干旱区绿洲土壤光谱特性研究

张子鹏^{1,2,3}, 丁建丽^{1,2,3*}, 王敬哲^{1,2,3}

¹新疆大学资源与环境科学学院, 新疆 乌鲁木齐 830046;

²新疆大学绿洲生态教育部重点实验室, 新疆 乌鲁木齐 830046;

³新疆大学智慧城市与环境建模自治区普通高校重点实验室, 新疆 乌鲁木齐 830046

摘要 土壤有机质(SOM)含量是评价土壤肥力的重要指标。以新疆渭-库绿洲为研究区,对室内获取的 SOM 含量及反射光谱数据进行 Savitzky-Golay (S-G)平滑和一阶微分(FD)预处理。在此基础上,为减小敏感波段遴选对建模精度的影响,引入谐波分析(HA)算法对全波段光谱数据进行谐波分解。基于主成分分析(PCA)降维后的 7 个主分量对 SOM 含量进行基于反向传播(BP)神经网络、遗传算法(GA)-BP 神经网络和多元线性回归(MLR)方法的定量估算,并对估算精度进行比较。结果表明:HA 预处理后的光谱数据与 SOM 含量的相关性相较于 FD 数据有了明显提高;非线性模型 BP 神经网络的估算精度明显高于线性模型 MLR;在非线性模型中,GA-BP 模型的估算精度最高,其决定系数为 0.92,预测集的均方根误差为 3.92×10^{-3} ,相对分析误差为 1.93。验证了 HA 算法深度挖掘光谱数据的有效性,经过 GA 优化的 BP 神经网络模型可以提高 SOM 含量的估算精度,为土壤属性的光谱定量估算提供借鉴。

关键词 遥感; 高光谱; 有机质; 谐波分析; 主成分分析; 遗传算法; 反向传播神经网络

中图分类号 S127

文献标识码 A

doi: 10.3788/AOS201939.0228003

Spectral Characteristics of Oasis Soil in Arid Area Based on Harmonic Analysis Algorithm

Zhang Zipeng^{1,2,3}, Ding Jianli^{1,2,3*}, Wang Jingzhe^{1,2,3}

¹ College of Resources & Environmental Science, Xinjiang University, Urumqi, Xinjiang 830046, China;

² Oasis Ecology Key Laboratory of Ministry of Education, Xinjiang University, Urumqi, Xinjiang 830046, China;

³ Key Laboratory of Smart City and Environment Modelling of Higher Education Institute, Xinjiang University, Urumqi, Xinjiang 830046, China

Abstract The soil organic matter (SOM) content is an important index for evaluating soil fertility. Weigan-Kuqa region in Xinjiang is selected as the study area, based on the laboratory-derived SOM content and reflectance data, the pretreatment of Savitzky-Golay (S-G) smoothing and first order derivative (FD) are carried out. In order to further reduce the influence of sensitive band selection on modeling accuracy, we introduce the harmonic analysis (HA) algorithm to conduct the harmonic decomposition of all wavelengths. Seven principal components are obtained using dimensional reduction treatment of the principal component analysis (PCA). Subsequently, the SOM contents of soil samples are quantified by means of three methods: back propagation (BP) neural network, Genetic Algorithm (GA)-BP, and multiple linear regression (MLR). The accuracy of these methods is compared here. The results show that the correlation coefficient between SOM content and HA pretreated spectral data is improved effectively compared with those of FD data. The estimate accuracy of the non-linear model, BP neural network, is better than that of the linear model, MLR. In terms of non-linear models, the estimate accuracy of the GA-BP model is the best, with the optimal determining coefficient of 0.92, root mean square error of prediction set of 3.92×10^{-3} , and the relative analysis error of 1.93. This study validates the effectiveness of the HA algorithm for the depth mining of spectral data, and the BP neural network model optimized by GA can improve estimate accuracy of SOM content, which can further provide scientific reference for the quantitative estimation of multiple soil properties.

Key words remote sensing; hyperspectral; organic matter; harmonic analysis; principal component analysis;

收稿日期: 2018-08-13; 修回日期: 2018-09-20; 录用日期: 2018-10-08

基金项目: 国家自然科学基金项目(41771470, U1303381, 41661046)、自治区重点实验室专项基金项目(2016D03001)

* E-mail: watarid@xju.edu.cn

genetic algorithm; back propagation neural network

OCIS codes 280.4788; 300.6170; 200.4260

1 引 言

土壤有机质(SOM)是土壤肥力的重要组成部分,它对于促进土壤结构的形成,改善土壤理化性质,提高土壤保肥能力具有重要作用,是评估土壤肥力与生产力的重要指标^[1]。传统的 SOM 含量测定多基于化学分析方法,费时费力,无法满足现代精细农业发展的需要。与传统的化学分析方法相比,遥感技术是一种很有前途的土壤属性定量评估方法,该技术具有低成本、高效益的优点,可以很好地描述和评估不同尺度的表层土壤的各种特征^[2]。因此,光谱分析技术可以作为一种替代方法来估算 SOM 含量。

SOM 在可见-近红外波段具有不同的光谱反射和吸收特征,这为其定量化分析奠定了基础^[3-7]。近年来,国内外众多学者在利用土壤光谱估算 SOM 含量方面进行了大量研究,但基于不同建模方法得到的 SOM 含量的估算精度具有较大差异^[8-11]。其中,反向传播(BP)神经网络因具有操作简单、稳健性高等优点而被广泛应用于 SOM 的建模研究中,但 BP 神经网络的本质属于梯度下降算法,存在局部最优的问题,因此不同的参数设定可能会导致网络陷入局部最优解^[12]。针对 BP 神经网络的本质特点以及初始权值、阈值的随机性,采用遗传算法(GA)优化上述两个参量,并比较 BP 神经网络与 GA-BP 神经网络对 SOM 含量估算的拟合效果。

土壤是由多种物质组成的混合物,其反射光谱常包括背景噪声、基线漂移和倾斜等干扰信息,将反射光谱直接用于 SOM 含量的估算得到的效果并不好^[13]。Bioucas-Dias 等^[14]的研究表明,地物反射光谱以光的反射强度为特征信号,属于电磁波谱的一部分,而任何复杂的波都由许多不同频率、振幅和相位的正弦分量组成。谐波分析(HA)算法是研究非线性畸变和提取信号的振幅、相位等分量的较为成熟的方法,其应用范围也比较广泛^[15-16]。采用 HA 算法处理土壤光谱,获得与 SOM 含量相关性较高的谐波特征分量,并将其作为建模因子,可以构建 SOM 含量的估算模型。

本课题组从光谱数据的预处理入手,对土壤的原始光谱进行 Savitzky-Golay(S-G)平滑和一阶微分(FD)处理,并在此基础上,引入 HA 算法,对全波

段数据进行多次 HA。根据 HA 的结果确定各谐波分量的最优分解次数,基于最优 HA 次数选择谐波特征分量,以此作为建模因子,实现 SOM 含量的估算。最后对各估算模型的精度进行比较,研究结果表明:非线性 BP 神经网络模型的估算能力明显优于线性模型;经 GA 优化的 BP 神经网络的估算能力最好。

2 实验数据与方法流程

2.1 研究区介绍及土壤样品的制备

研究区位于为渭-库绿洲,地处新疆南部,塔里木盆地中北部,包括新和县、沙雅县和库车县 3 个行政县域。该地区属于典型的温带大陆性气候,具有冬冷夏热、昼夜温差大、气候干燥、降水少、蒸发强等特点^[17]。同时,该地区属于绿洲灌溉农业区,土壤类型以潮土、灰漠土、荒漠土、棕漠土等为主。由于土地利用方式及土壤类型的不同,SOM 含量的空间分异性较为明显^[18]。研究区采样图如图 1 所示。

根据研究区的特点,共设置 101 个采样点,以梅花采样法采集 5 点的样本,混合后作为土壤样品,采样深度为 0~20 cm,使用手持全球定位系统(GPS)记录样点的位置,以便用于检验。采集后的样品带回实验室后进行自然风干处理,粗略去除沙砾及植物残体后进行碾磨,并过 2 mm 孔筛。将土壤样品分为两份,一份用于室内采集光谱数据,另一份基于重铬酸钾-外加热法^[19]测定 SOM 含量。为了保证建模集和验证集数据都尽可能地涵盖 SOM 含量的全部范围,使用 Kennard-Stone(K-S)算法对样本集合进行划分^[20],101 个土壤样本被分成两部分,其中的 71 个样本用于建模,剩余的 30 个样本用于验证建模的结果。SOM 含量的统计特征如表 1 所示。

由表 1 可知,研究区土壤样本的 SOM 含量(质量分数)为 $2.32 \times 10^{-3} \sim 18.06 \times 10^{-3}$,平均含量为 10.94×10^{-3} ,变异系数为 31.51%,属于中等变异程度。建模集和验证集的 SOM 含量均值分别为 10.91×10^{-3} 和 11.04×10^{-3} ,变异系数分别为 31.61% 和 31.79%。数据集的划分结果表明,两个集合的统计结果相似,确保了划分的样本具有代表性。

2.2 光谱数据的采集

测定土壤光谱的光谱仪型号为美国 ASD 公司的 Field Spec 3 型,其采样间隔为 1.4 nm 和 2 nm (350~1000 nm 与 1000~2500 nm 区间),重采样间

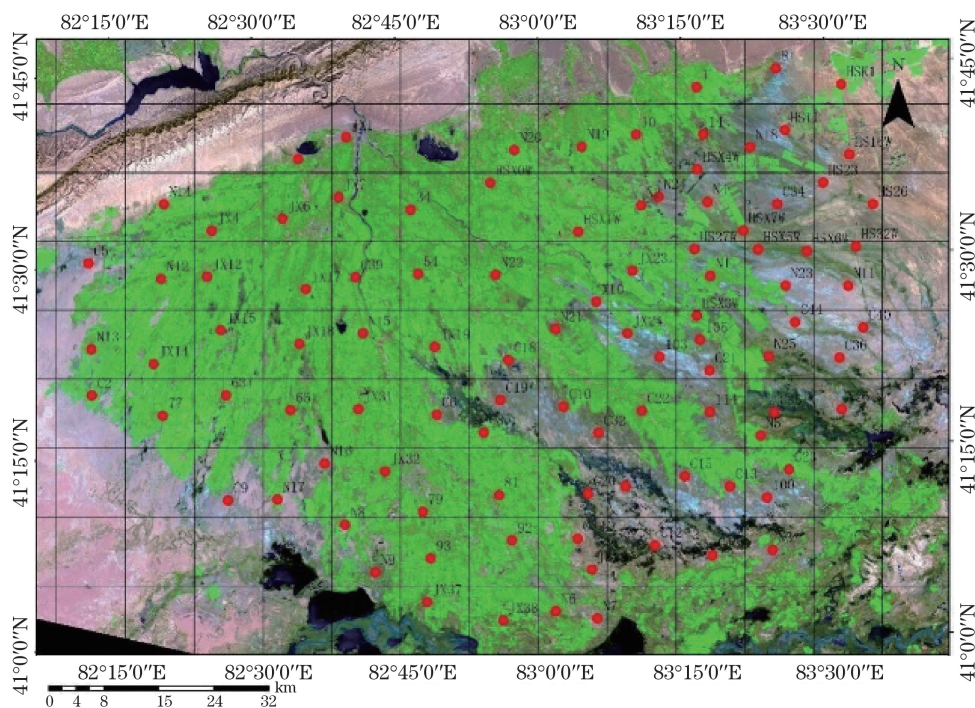


图 1 研究区采样图

Fig. 1 Research area sampling chart

表 1 SOM 含量的统计特征

Table 1 Statistical characteristics of SOM content

Type of samples	Number of sample	SOM content (mass fraction) / 10 ⁻³			Standard deviation / 10 ⁻³	Coefficient of variation / %
		Minimum	Maximum	Mean		
Whole set	101	2.32	18.06	10.94	3.45	31.51
Calibration set	71	3.68	18.06	10.91	3.45	31.61
Validation set	30	2.32	16.45	11.04	3.51	31.79

隔为 1 nm,输出波段数为 2151 个^[21]。土壤光谱的测定在暗室中进行,选用 50 W 的卤素灯,探头的视场角为 5°,探头距离土壤表面 30 cm,每次测定前均进行白板定标,每个土壤样本重复测定 10 次,取算术平均值作为该样品的反射率。

2.3 光谱数据的预处理

位于两端处的光谱数据(350~400 nm 和 2401~2500 nm)通常存在高频噪声,需要剔除。对 101 个土壤样本在 401~2400 nm 波段的反射率光谱进行二阶 9 点 S-G 平滑^[22]处理。光谱微分是增加信噪比的主要技术手段,其中 FD 可以去除不同的背景噪声,修复基线漂移,能反映被测地物的本质特征。同时,经 FD 处理后的光谱的分辨率更高,且光谱轮廓更加清晰^[23]。基于此,利用 FD 对平滑后的光谱数据进行后处理,并将其应用于后续的研究中。

2.4 HA 原理

HA 又称调和函数,其通过正弦函数和余弦函

数来拟合数字信号或时间序列的周期函数,将原始数据从时域变换到频率域,并根据相应的谐波分解式提取信号的振幅、相位等信息^[24]。进行 HA 时,将光谱曲线的波段数 L 视为函数的周期 L 。HA 利用谐波余项($S_0/2$)、振幅(A_t 、 B_t 、 C_t)和相位(φ_t)等谐波分量的叠加来表示每个波段的光谱反射率。若每个样本由 L 个波段组成,则该样本的光谱可以表示为 $S_{(r)} = \{S_1, S_2, \dots, S_r, \dots, S_L\}$,每个波段的光谱值表示为 S_r , r 为波段序列号($r = 1, 2, \dots, L$), L 为总的波段数, t 次谐波分解为

$$S_{(r)} = \frac{S_0}{2} + \sum_{t=1}^{\infty} \left[A_t \cos\left(\frac{2\pi tr}{L}\right) + B_t \sin\left(\frac{2\pi tr}{L}\right) \right] = \frac{S_0}{2} + \sum_{t=1}^{\infty} C_t \sin\left(\frac{2\pi tr}{L} + \varphi\right), \quad (1)$$

t 次 HA 的特征分量为

$$\frac{S_0}{2} = \frac{1}{L} \sum_{r=1}^L S_r, \quad (2)$$

$$A_t = \frac{2}{L} \sum_{r=1}^L S_r \cos\left(\frac{2\pi tr}{L}\right), \quad (3)$$

$$B_t = \frac{2}{L} \sum_{r=1}^L S_r \sin\left(\frac{2\pi tr}{L}\right), \quad (4)$$

$$C_t = \sqrt{(A_t^2 + B_t^2)}, \quad (5)$$

$$\varphi_t = \arctan\left(-\frac{A_t}{B_t}\right), \quad (6)$$

式中： $S_0/2$ 为谐波余项； t 为 HA 次数， $t=1,2,3,\dots$ ； $C_t \sin(\pi tr/L + \varphi_t)$ 为第 t 次谐波分量，当 $t=1$ 时表示基波分量； A_t 、 B_t 、 C_t 均为描述第 t 次 HA 的振幅信息， A_t 为余弦振幅； B_t 为正弦振幅； C_t 为谐波分量振幅； φ_t 为谐波分量的相位信息。各谐波分量的物理意义如下：

1) $S_0/2$ 为地物反射光谱的平均反射强度，但其只描述地物反射电磁波的平均水平，并不反映光谱

曲线的细节特征；2) C_t 表示不同波段的反射率基于反射率均值的变化范围， C_t 越大，表明其对光谱形变的贡献率越大，且在一定程度上能描述其对待测物体的响应程度；3) φ_t 表示波峰或波谷在横坐标上的位置，对于光谱数据来说，相位是描述光谱波形成变化的标度，它代表了 C_t 在光谱上的位置信息。不同地物光谱的 C_t 不同，其对应的相位值也不一样。相位是地物反射电磁波的特有性质，承载着地物光谱反射峰值的位置信息。

2.5 SOM 含量的估算流程

使用主成分分析(PCA)后的主分量作为建模因子，采用线性模型中的多元线性回归(MLR)和非线性模型中的 BP 神经网络作为估算模型，并引入 GA 对 BP 神经网络进行参量优化。SOM 含量的估算流程如图 2 所示。

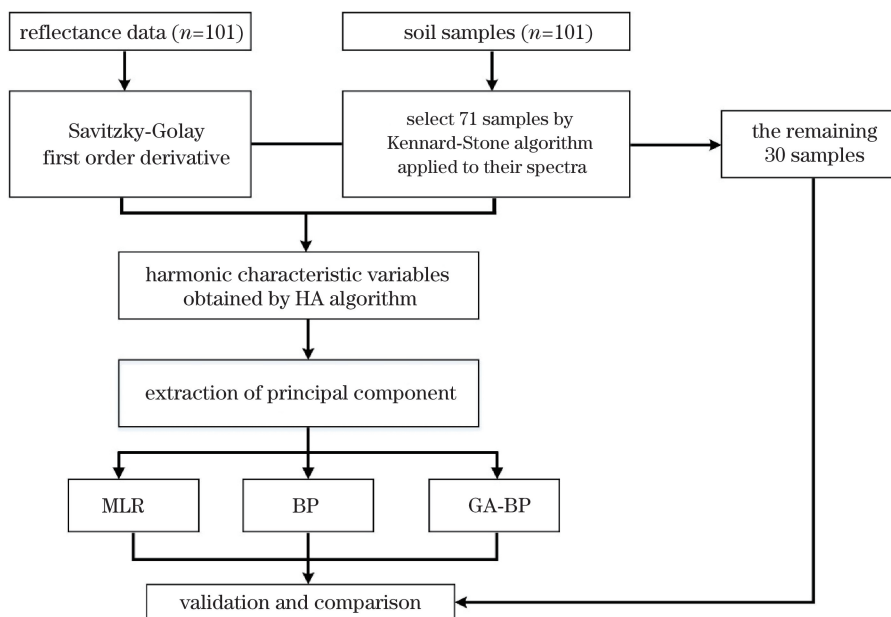


图 2 SOM 含量估算的流程图

Fig. 2 Flow chart of SOM content estimation

2.6 模型的精度检验

分别采用 BP 神经网络和 MLR 模型构建 SOM 含量的高光谱估测模型，并从拟合程度、估算能力、准确性 3 个方面对模型进行验证。模型的拟合度用决定系数 R^2 评价， R^2 越接近 1，说明模型的拟合程度越高。模型的估算能力用验证集均方根误差(RMSEP)来评价，RMSEP 越小，说明模型的估算能力越好。估算模型的准确性用相对分析误差(RPD)来评价，当 $RPD < 1.0$ 时，表明模型的估算能力非常差，不推荐使用；当 $1.0 \leq RPD < 1.4$ 时，表明模型的量化能力较弱，仅能区分样本含量的高低；当 $1.4 \leq RPD < 1.8$ 时，该模型可以提供较为

合理的估算结果；当 $1.8 \leq RPD < 2.0$ 时，说明其具有较好的量化性能，可用来进行样本含量的定量估算^[1]。

3 结果分析与讨论

3.1 光谱的预处理

由图 3 可知：原始光谱与 SOM 含量的相关性曲线较为平滑，并且呈负相关关系，最大相关系数为 0.26，通过 0.01 显著性水平检验($P^{**} = \pm 0.255$)的波段仅有 62 个；经过 FD 处理的光谱相关性明显提升，其中通过 0.01 显著性水平检验的波段共 416 个，最大正相关系数为 0.29，最大负相关系数为 -0.30，

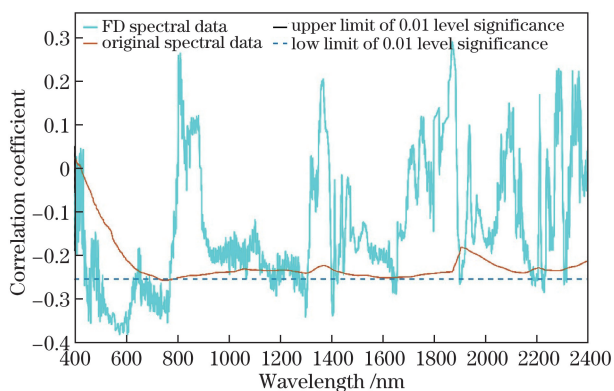


图 3 原始光谱、FD 光谱与 SOM 含量的相关性曲线

Fig. 3 Correlation curves of original spectrum, FD spectrum and SOM content

其在可见光波段(401~760 nm)的变化较大,并呈明显的负相关。研究表明,FD 的土壤光谱与 SOM 含量的相关性较原始光谱明显增强,提高了土壤光谱对 SOM 的敏感度。

3.2 谐波分析

以 500~4000 次为 HA 区间,500 次为 HA 间隔,对 FD 处理后的土壤光谱利用(1)~(6)式进行 HA,获得谐波特征分量 A_i 、 B_i 、 C_i 、 φ_i ,并计算其与 SOM 含量的皮尔森相关系数。图 4 为在不同的 HA 次数下, A_i 、 B_i 、 C_i 、 φ_i 与 SOM 含量的相关系数图。表 2~5 为不同 HA 次数下各谐波分量与 SOM 含量的相关系数统计特征。

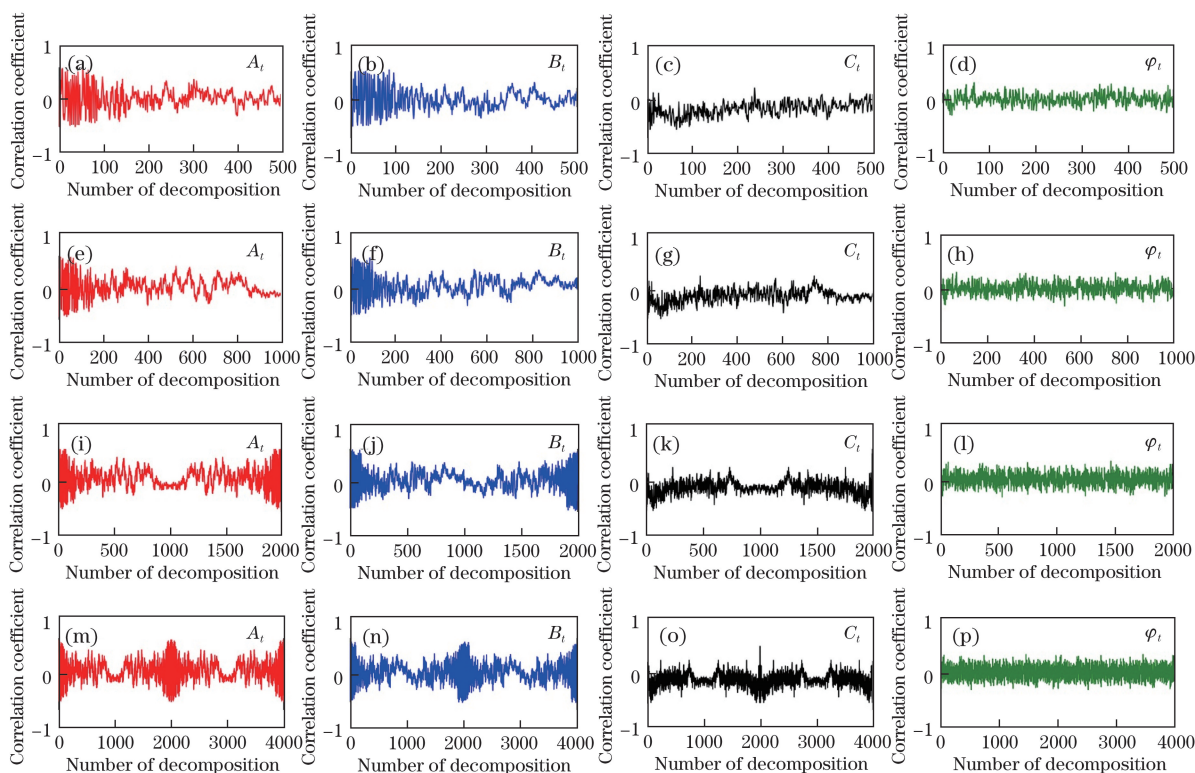


图 4 不同 HA 次数下谐波分量与 SOM 含量的相关系数图。(a)(e)(i)(m) A_i 与 SOM 的相关系数;

(b)(f)(j)(n) B_i 与 SOM 的相关系数;(c)(g)(k)(o) C_i 与 SOM 的相关系数;(d)(h)(l)(p) φ_i 与 SOM 的相关系数

Fig. 4 Correlation coefficient diagrams of harmonic component and SOM content at different HA times. (a)(e)(i)(m)

Correlation coefficient of A_i and SOM content; (b)(f)(j)(n) correlation coefficient of B_i and SOM content; (c)(g)

(k)(o) correlation coefficient of C_i and SOM content; (d)(h)(l)(p) correlation coefficient of φ_i and SOM content

由图 4(a)~(d)可知: A_i 、 B_i 与 SOM 含量的相关系数在 HA 前 100 次较高; C_i 、 φ_i 与 SOM 含量的相关性比 A_i 、 B_i 与 SOM 含量的相关性低,且没有明显的规律变化。

由图 4(e)~(h)可知:当 HA 次数为 1000 时, A_i 、 B_i 、 C_i 与 SOM 含量的相关系数具有减小的趋势,且变化逐渐趋于稳定, φ_i 与 SOM 含量的相关系

数变化没有规律。

由图 4(i)~(l)结合表 2~5 可知:当 HA 次数为 2000 时, A_i 、 C_i 与 SOM 含量的相关系数都呈轴对称,最大正相关系数和最大负相关系数均出现在 HA 的前后 10 次内; B_i 、 φ_i 与 SOM 的相关系数分别呈中心对称,最大正相关系数和最大负相关系数的数值相等,且对应的 HA 次数之和为 2000。

表 2 不同 HA 次数下 A_i 与 SOM 含量的相关系数统计特征

Table 2 Statistical characteristics of correlation coefficients between A_i and SOM content at different HA times

Harmonic decomposition times	Maximum positive correlation coefficient	Harmonic characteristic component	Maximum negative correlation coefficient	Harmonic characteristic component
500	0.57	7	-0.54	1
1000	0.57	7	-0.54	1
1500	0.57	7	-0.54	1
2000	0.57	7,1993	-0.54	1,1999
2500	0.57	7,1993,2007	-0.54	1,1999,2001
3000	0.57	7,1993,2007	-0.54	1,1999,2001
3500	0.57	7,1993,2007	-0.54	1,1999,2001
4000	0.57	7,1993,2007,3993	-0.54	1,1999,2001,3999

表 3 不同 HA 次数下 B_i 与 SOM 含量的相关系数统计特征

Table 3 Statistical characteristics of correlation coefficient between B_i and SOM content at different HA times

Harmonic decomposition times	Maximum positive correlation coefficient	Harmonic characteristic component	Maximum negative correlation coefficient	Harmonic characteristic component
500	0.54	21	-0.51	4
1000	0.54	21	-0.51	4
1500	0.54	21	-0.51	4
2000	0.54	21	-0.51	1979
2500	0.54	21,2021	-0.51	1979
3000	0.54	21,2021	-0.51	1979
3500	0.54	21,2021	-0.51	1979
4000	0.54	21,2021	-0.51	1979,3979

表 4 不同 HA 次数下 C_i 与 SOM 含量的相关系数统计特征

Table 4 Statistical characteristics of correlation coefficient between C_i and SOM content at different HA times

Harmonic decomposition times	Maximum positive correlation coefficient	Harmonic characteristic component	Maximum negative correlation coefficient	Harmonic characteristic component
500	0.47	1	-0.56	6
1000	0.47	1	-0.56	6
1500	0.47	1	-0.56	6
2000	0.47	1,1999	-0.56	6,1994
2500	0.47	1,1999,2001	-0.56	6,1994,2006
3000	0.47	1,1999,2001	-0.56	6,1994,2006
3500	0.47	1,1999,2001	-0.56	6,1994,2006
4000	0.47	1,1999,2001,3999	-0.56	6,1994,2006,3994

由图 4(m)~(p)结合表 2~5 可知:当 HA 次数为 4000 时,所有谐波分量与 SOM 含量的相关系数都呈周期对称,最大正/负相关系数增加至 24 个,相比于 HA 为 2000 次时增加了 1 倍。

根据 HA 为 2000 次时的对称性和 HA 为 4000 次时的周期对称性,将 HA 的整周期确定为 2000

次,与光谱曲线的周期(波段数) L 相同。在确保谐波分量完整性的基础上,为了减少数据冗余,确定 A_i 和 C_i 的最优谐波分解次数为 1000 次($L/2$ 次), B_i 和 φ_i 的最优谐波分解次数为 2000 次,即 L 次,并按照相关性大小选取了相关系数大于 0.5 的 21 个谐波分量用于模型构建。

表 5 不同 HA 次数下 φ_i 与 SOM 含量的相关系数统计特征

Table 5 Statistical characteristics of correlation coefficient between φ_i and SOM content at different HA times

Harmonic decomposition times	Maximum positive correlation coefficient	Harmonic characteristic component	Maximum negative correlation coefficient	Harmonic characteristic component
500	0.30	68	-0.30	20
1000	0.30	68	-0.32	532
1500	0.30	68	-0.32	532
2000	0.32	1468	-0.32	532
2500	0.32	1468	-0.32	532
3000	0.32	1468	-0.32	532,2532
3500	0.32	1468,3468	-0.32	532,2532
4000	0.32	1468,3468	-0.32	532,2532

3.3 模型的构建

为了提高 BP 神经网络的运算速率和减小运算误差,对选出的 21 个谐波特征分量进行 PCA,并提取主成分特征值。由表 6 可知前 7 个主成分的累计方差贡献率可达 99.13%,很好地涵盖了谐波分量的特征信息^[7]。基于 PCA 的结果,选择前 7 个主分量作为 BP 神经网络的输入层。

BP 神经网络的本质是梯度下降算法,该算法本身存在收敛速度慢和易陷入局部极小值的缺陷,同时,初始权值和阈值的随机性对网络学习结果的影响较大^[25]。因此,引入 GA 对 BP 神经网络连接层的权值和阈值进行优化(GA 算法与神经网络算法相结合可以避免陷入局部最优解),以达到更好的估算结果^[26]。结合实际情况对 BP 神经网络的参数进行多次训练,最终确立网络拓扑结构为 7-5-1。其

表 6 PCA 特征值和方差贡献率

Table 6 Characteristic value of PCA and variance contribution rate

Principle component	Characteristic value	Variance contribution /%	Accumulative contribution /%
PCA1	16.33	77.80	77.80
PCA2	1.45	6.91	84.71
PCA3	1.09	5.19	89.90
PCA4	0.91	4.34	94.24
PCA5	0.52	2.58	96.82
PCA6	0.26	1.43	98.25
PCA7	0.16	0.88	99.13
...
PCA21	-3.13×10^{-16}	-1.49×10^{-15}	100

中,网络迭代运算次数为 2000 次,学习率为 0.01,学习误差为 0.001,如图 5 所示。

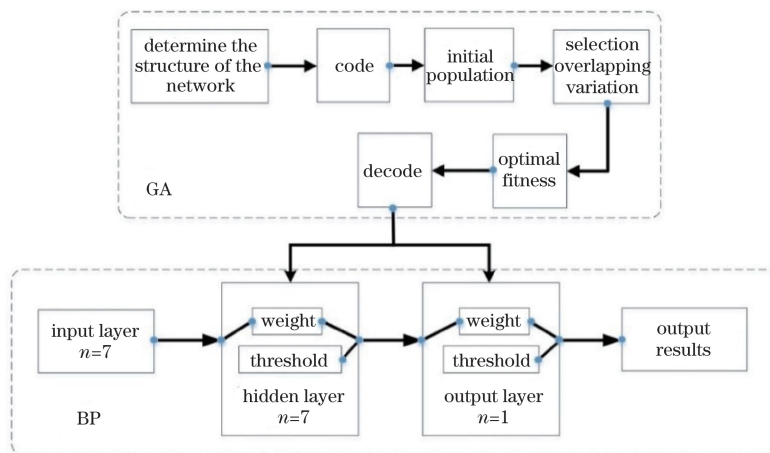


图 5 GA-BP 神经网络的技术路线

Fig. 5 Technology roadmap of GA-BP neural network

根据 BP 神经网络的拓扑结构,计算得出需要优化的权值和阈值共 46 个,基于对算法精度、可行性和复杂度的考虑,经过反复测试确定 GA 算法的

染色体数目为 50,进化代数为 50,选择概率为 0.9,选择方法为轮盘赌法,交叉概率为 0.9,变异概率为 0.1。选择误差平方和的倒数作为适应度函数值,

即最优个体的适应度函数值越小,它与期望值之间的误差就越小,通过反复的迭代运算寻找 BP 神经网络连接层的最优权值和阈值。从图 6 中可以得出初始的适应度函数值为 23.2,经过多次迭代运算后适应度函数值逐渐减小,迭代第 25 次时停止进化,适应度函数值降为 7.30,确定该网络为 GA-BP 的网络。

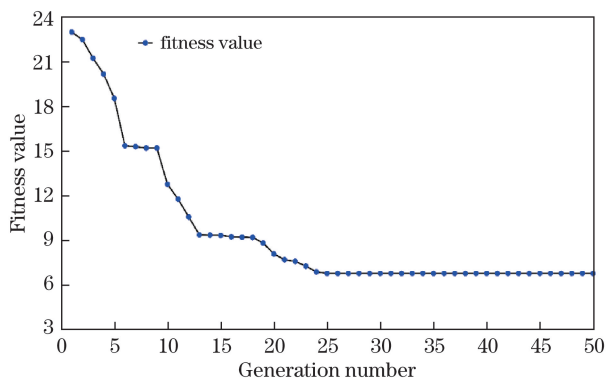


图 6 最优个体的适应度变化曲线

Fig. 6 Changing curve of optimal individual fitness

表 7 为不同估算模型的精度统计结果。由表 7 可知:GA-BP 模型的 R^2 最大,为 0.92;BP 次之,为 0.87;MLR 最小,为 0.77。GA-BP 模型的 RMSEP

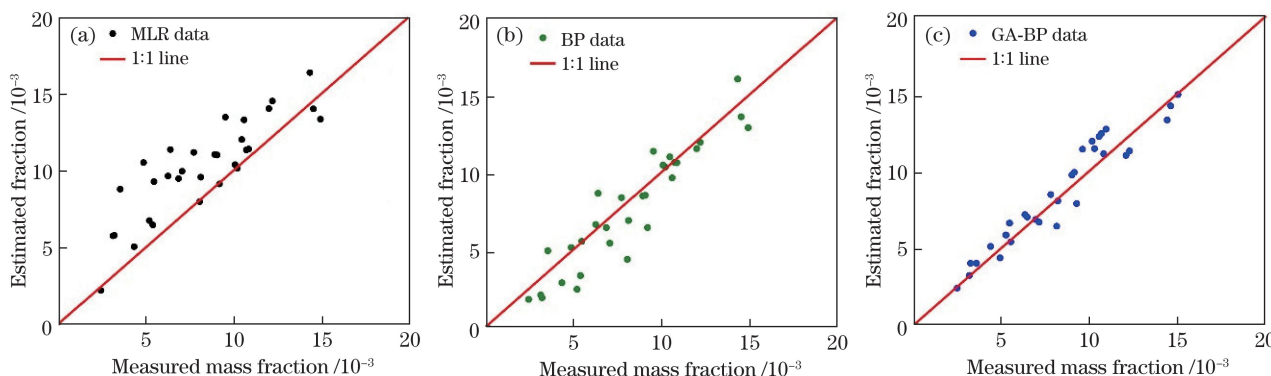


图 7 不同模型 SOM 含量的实测值与估算值。(a) MLR 模型; (b) BP 模型; (c) GA-BP 模型

Fig. 7 Measured and estimated SOM content for different models. (a) MLR model; (b) BP model; (c) GA-BP model

3.4 讨论

高光谱数据的预处理可以在不同程度上提高光谱数据对待测物的敏感程度,以进一步优化模型的量化效果^[20]。Shi 等^[27]对光谱数据进行多种数学变换后发现,SOM 的敏感波段集中在 600~800 nm 附近,并且光谱数据与 SOM 含量呈负相关关系。本研究所发现的敏感波段与其研究结果基本一致。姜雪芹等^[28]在黄土高光谱数据的处理中引入 HA 算法,选择相关系数大于 0.5 的 180 个特征波段进行 180 次谐波分解,并从相位、振幅等方面初步探讨了其与土壤水分含量的关系。

最小,为 3.92×10^{-3} ;BP 模型的 RMSEP 略高于 GA-BP 模型,为 4.07×10^{-3} ;MLR 的 RMSEP 最大,为 4.76×10^{-3} 。GA-BP 模型的 RPD 最大,为 1.93;BP 的 RPD 次之,为 1.85;MLR 的 RPD 最小,为 1.59。综上所述,3 种模型均可用于估算 SOM 含量,其中 GA-PB 模型的拟合程度、估算能力和准确性均优于另外两种模型。

表 7 不同估算模型的精度统计结果

Table 7 Accuracy statistical results of different estimate models

Validation model	R^2	RMSEP / 10^{-3}	RPD
MLR	0.77	4.76	1.59
BP	0.87	4.07	1.85
GA-BP	0.92	3.92	1.93

图 7 为不同模型 SOM 实测值与估算值的比较。由图 7 可知:MLR 模型的估算效果最差,其估算值普遍高于 1:1 线;BP 模型的估算效果略优于 MLR 模型,但估测值部分低于 1:1 线;GA-BP 模型的实测值和估算值基本分布于 1:1 线附近,模型的估算效果相较于前二者更为优异,说明 GA-BP 模型能够对 SOM 含量进行较好的估算。

本研究以 500~4000 次为谐波分解区间,以 500 次为谐波分解间隔对土壤的 FD 光谱进行 HA,并计算各次 HA 的特征分量与 SOM 含量的相关系数。研究发现: A_i 、 C_i 与 SOM 含量的相关系数在一个 HA 周期内分别具有对称性; B_i 、 φ_i 与 SOM 含量的相关系数在一个 HA 整周期内分别具有中心对称性。这与姜雪芹等^[28]的研究结果类似。在波段数的选择上,本研究选取了土壤光谱的全波段数据进行 HA,结合光谱信号的振幅、相位等特性深度挖掘了光谱的潜在信息,在提升光谱对 SOM 响应程度的同时减少了光谱信息的丢失。同时本研究发

现, A_i 、 B_i 与 SOM 含量的相关系数在 HA 的一个周期内的前后 100 次较高, 这表明土壤光谱的波形信息主要集中在 HA 前后 100 次的谐波特征分量中。 φ_i 代表了土壤光谱辐射峰值的位置信息, 反映了研究区土壤辐射电磁波的特有性质, 因此, φ_i 与 SOM 含量的相关系数在 HA 的一个周期内相对较小。

此外, 本研究还尝试性地用谐波特征分量对 SOM 含量进行高光谱建模。 Wang 等^[20] 对盐渍土光谱进行 FD 处理, 得到的 RPD 为 1.14, RMSEP 为 8.09×10^{-3} , R^2 为 0.23。 本研究的结果表明, 采用 HA 的 FD 反射率建模, RPD 可达 1.93, RMSEP 较 Wang 等采用 FD 处理得到的 RMSEP 降低了 4.17×10^{-3} , R^2 则由 0.23 提升至 0.92, 说明利用谐波特征分量建模的精度和稳定性较 FD 数据有了大幅提升, 并达到了 SOM 含量高光谱估算的精度要求, 为高光谱数据定量估算 SOM 含量提供了新的视角, 证明了 HA 算法在高光谱建模中的可行性。

此外, 根据已有的研究可知: 当 SOM 含量(质量分数)大于 20×10^{-3} 时, 其成为土壤光谱特性的主要影响因素; 当 SOM 含量(质量分数)小于 20×10^{-3} 时, 土壤的其他组成物质对土壤光谱特性的影响有所增强, 这对于特征波段的选取和估算模型的建立都带来了巨大挑战^[29]。 周倩倩等^[30] 采用多种方法对光谱数据进行预处理, 其中 FD 光谱与 SOM 含量的相关性最为显著。 栾福明等^[31] 采用 MLR 模型和 BP 神经网络分别建立了 SOM 含量的估算模型, 结果发现, 利用 BP 神经网络建立的模型对 SOM 含量的估算效果较 MLR 模型更好, 估算值与实测值更接近。 本研究利用 GA 对 BP 神经网络连接层的权值和阈值进行多次迭代优化, 并建立估算模型, 结果发现: 非线性 BP 神经网络模型的估算能力要明显优于线性模型 MLR; 而经过 GA 优化的 BP 神经网络对 SOM 含量的估算能力最好, 其实测值和估测值基本分布于 1:1 线附近(如图 7 所示)。

本研究的土壤采样点集中在渭-库绿洲及其周边地区, 该地区存在土地荒漠化现象, 而且土壤的盐渍化程度十分严重。 土壤的酸碱度(pH)、含盐量等因素都会对干旱区 SOM 含量的估算精度产生影响。 因此, 本研究所建立的 SOM 含量估算模型对其他地区是否适用, 还有待于进一步深入研究。 因此, 后续研究在讨论其他光谱预处理对 HA 影响的同时, 需要进一步增加土壤采样点和扩大采样范围, 以提高模型的估算精度和普适性。

4 结 论

本研究以渭-库绿洲为研究区, 通过实验获取了 101 个土壤样本的 SOM 含量数据及其高光谱数据, 并在此基础上利用 HA 算法对高光谱数据进行处理, 选择相关系数大于 0.5 的 21 个谐波特征分量进行建模, 得出的结论如下:

1) 土壤的 FD 光谱具有较多的敏感波段, 为了避免特征波段遴选时的不确定性, 采用 HA 算法对土壤光谱进行时频变换和谐波特征分量的提取。 研究发现: A_i 、 B_i 与 SOM 含量的相关系数在 HA 前后 100 次较高; HA 1000 次时, A_i 、 B_i 、 C_i 与 SOM 含量的相关系数具有减小的趋势, 且变化趋于稳定。 φ_i 与 SOM 含量的相关系数相比于振幅的变化较弱, 且没有明显的变化规律。 本研究确定 HA 2000 次 (L) 为谐波分量与 SOM 含量相关系数变化的一个整周期。 一个周期的 HA 得到的谐波特征分量与 SOM 含量的相关系数具有对称性: A_i 、 C_i 与 SOM 的相关系数分别呈轴对称; B_i 、 φ_i 与 SOM 含量的相关系数分别呈中心对称。 HA 1000 次 ($L/2$) 为 A_i 和 C_i 的最优谐波分解次数, HA 2000 次 (L) 为 B_i 和 φ_i 的最优 HA 次数。 研究结果表明, 谐波特征分量与 SOM 含量的相关性相比于 FD 有了明显提高, 并呈现出更多的隐含信息, 说明 HA 算法提高了土壤光谱对 SOM 含量的响应力。

2) 对比分析了 MLR 模型、BP 神经网络模型和 GA-BP 模型对 SOM 含量的估算能力, 得出非线性 BP 神经网络模型的估算能力要明显优于线性模型 MLR 的结论。 而且, 在非线性模型中, GA-BP 的估算能力最好, 该模型的 RMSEP = 3.92×10^{-3} , $R^2 = 0.92$, RPD = 1.93, 与其他学者通过不同方法估算的 SOM 含量相比, 该估算结果明显更好。 该结果对研究土壤的信息提取和建模估测具有一定的参考价值, 为估测土壤的其他属性特征提供了新思路。

参 考 文 献

- [1] Rossel R A V, 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.
- [2] Ghamisi P, Yokoya N, Li J, et al. Advances in hyperspectral image and signal processing: a comprehensive overview of the state of the art[J]. Proceedings of the IEEE, 2017, 5(4): 37-78.
- [3] Minasny B, McBratney A B, Bellon-Maurel V, et al.

- Removing the effect of soil moisture from NIR diffuse reflectance spectra for the prediction of soil organic carbon[J]. *Geoderma*, 2011, 167-168: 118-124.
- [4] Peón J, Fernández S, Recondo C, *et al.* Evaluation of the spectral characteristics of five hyperspectral and multispectral sensors for soil organic carbon estimation in burned areas[J]. *International Journal of Wildland Fire*, 2017, 26(3): 230-239.
- [5] Tahmasbian I, Xu Z H, Boyd S, *et al.* Laboratory-based hyperspectral image analysis for predicting soil carbon, nitrogen and their isotopic compositions[J]. *Geoderma*, 2018, 330: 254-263.
- [6] Lin L X, Xue F C, Wang Y J, *et al.* Photography measured-value magnification improves local correlation maximization-complementary superiority method of hyperspectral analysis of soil total nitrogen[J]. *Catena*, 2018, 165: 106-114.
- [7] Nocita M, Stevens A, Noon C, *et al.* Prediction of soil organic carbon for different levels of soil moisture using Vis-NIR spectroscopy[J]. *Geoderma*, 2013, 199: 37-42.
- [8] Stenberg B. Effects of soil sample pretreatments and standardized rewetting as interacted with sand classes on Vis-NIR predictions of clay and soil organic carbon[J]. *Geoderma*, 2010, 158(1/2): 15-22.
- [9] Gomez C, Lagacherie P, Coulouma G. Regional predictions of eight common soil properties and their spatial structures from hyperspectral Vis-NIR data[J]. *Geoderma*, 2012, 189-190: 176-185.
- [10] Yu L, Hong Y S, Zhou Y, *et al.* Inversion of soil organic matter content using hyperspectral data based on continuous wavelet transformation[J]. *Spectroscopy and Spectral Analysis*, 2016, 36(5): 1428-1433.
于雷, 洪永胜, 周勇, 等. 连续小波变换高光谱数据的土壤有机质含量反演模型构建[J]. *光谱学与光谱分析*, 2016, 36(5): 1428-1433.
- [11] Hong Y S, Yu L, Chen Y Y, *et al.* Prediction of soil organic matter by VIS-NIR spectroscopy using normalized soil moisture index as a proxy of soil moisture[J]. *Remote Sensing*, 2017, 10(2): 28.
- [12] Luo H L, Yang Y, Tong B, *et al.* Traffic sign recognition using a multi-task convolutional neural network[J]. *IEEE Transactions on Intelligent Transportation Systems*, 2018, 19(4): 1100-1111.
- [13] Peng X T, Shi T Z, Song A H, *et al.* Estimating soil organic carbon using VIS/NIR spectroscopy with SVMR and SPA methods[J]. *Remote Sensing*, 2014, 6(4): 2699-2717.
- [14] Bioucas-Dias J M, Plaza A, Camps-Valls G, *et al.* Hyperspectral remote sensing data analysis and future challenges[J]. *Proceedings of the IEEE*, 2013, 1(2): 6-36.
- [15] Xue Z H, Du P J, Su H J. Harmonic analysis for hyperspectral image classification integrated with PSO optimized SVM[J]. *Proceedings of the IEEE*, 2014, 7(6): 2131-2146.
- [16] Donoho D L, Maleki A, Rahman I U, *et al.* Reproducible research in computational harmonic analysis[J]. *Computing in Science & Engineering*, 2009, 11(1): 8-18.
- [17] Ding J L, Yu D L. Monitoring and evaluating spatial variability of soil salinity in dry and wet seasons in the Werigan-Kuqa Oasis, China, using remote sensing and electromagnetic induction instruments[J]. *Geoderma*, 2014, 235-236: 316-322.
- [18] Cao L, Ding J L, Yu H Y. Relationship between multi-scale landscape pattern and salinity in Weigan and Kuqa rivers delta oasis[J]. *Transactions of the Chinese Society of Agricultural Engineering*, 2016, 32(3): 101-110.
曹雷, 丁建丽, 于海洋. 渭-库绿洲多尺度景观格局与盐度关系[J]. *农业工程学报*, 2016, 32(3): 101-110.
- [19] Bao S D. *Soil agrochemical analysis*[M]. 3rd ed. Beijing: China Agriculture Press, 2000.
鲍士旦. *土壤农化分析*[M]. 3版. 北京: 中国农业出版社, 2000.
- [20] Wang J Z, Ding J L, 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.
- [21] Shi Z, Wang Q L, Peng J, *et al.* Development of a national VNIR soil-spectral library for soil classification and prediction of organic matter concentrations[J]. *Science China Earth Sciences*, 2014, 57(7): 1671-1680.
- [22] Savitzky A, Golay M J E. Smoothing and differentiation of data by simplified least squares procedures[J]. *Analytical Chemistry*, 1964, 36(8): 1627-1639.
- [23] Du J, Hu B L, Zhang Z F. Gastric carcinoma classification based on convolutional neural network and micro-hyperspectral imaging [J]. *Acta Optica Sinica*, 2018, 38(6): 0617001.
杜剑, 胡炳樑, 张周锋. 基于卷积神经网络与显微高光谱的胃癌组织分类方法研究[J]. *光学学报*, 2018, 38(6): 0617001.
- [24] Yang K M, Zhang T, Wang L B, *et al.* A new algorithm on hyperspectral image fusion based on the harmonic analysis[J]. *Journal of China University of Mining & Technology*, 2014, 43(3): 547-553.

- 杨可明, 张涛, 王立博, 等. 高光谱影像的谐波分析融合算法研究[J]. 中国矿业大学学报, 2014, 43(3): 547-553.
- [25] Ge X Y, Ding J L, Wang J Z, *et al.* Estimation of soil moisture content based on competitive adaptive reweighted sampling algorithm coupled with machine learning[J]. *Acta Optica Sinica*, 2018, 38(10): 1030001.
- 葛翔宇, 丁建丽, 王敬哲, 等. 基于竞争适应重加权采样算法耦合机器学习的土壤含水量估算[J]. 光学学报, 2018, 38(10): 1030001.
- [26] Wang W X, Tang R C, Li C, *et al.* A BP neural network model optimized by Mind Evolutionary Algorithm for predicting the ocean wave heights[J]. *Ocean Engineering*, 2018, 162: 98-107.
- [27] Shi Z, Ji W, Viscarra Rossel R A, *et al.* Prediction of soil organic matter using a spatially constrained local partial least squares regression and the Chinese vis-NIR spectral library[J]. *European Journal of Soil Science*, 2015, 66(4): 679-687.
- [28] Jiang X Q, Ye Q, Lin Y, *et al.* Inverting study on soil water content based on harmonic analysis and hyperspectral remote sensing[J]. *Acta Optica Sinica*, 2017, 37(10): 1028001.
- 姜雪芹, 叶勤, 林怡, 等. 基于谐波分析和高光谱遥感的土壤含水量反演研究[J]. 光学学报, 2017, 37(10): 1028001.
- [29] Liu H J, Zhang Y Z, Zhang X L, *et al.* Quantitative analysis of moisture effect on black soil reflectance[J]. *Pedosphere*, 2009, 19(4): 532-540.
- [30] Zhou Q Q, Ding J L, Tang M Y, *et al.* Inversion of soil organic matter content in oasis typical of arid area and its influencing factors[J]. *Acta Pedologica Sinica*, 2018, 55(2): 313-324.
- 周倩倩, 丁建丽, 唐梦迎, 等. 干旱区典型绿洲土壤有机质的反演及影响因素研究[J]. 土壤学报, 2018, 55(2): 313-324.
- [31] Luan F M, Zhang X L, Xiong H G, *et al.* Comparative analysis of soil organic matter content based on different inversion models[J]. *Spectroscopy and Spectral Analysis*, 2013, 33(1): 196-200.
- 栾福明, 张小雷, 熊黑钢, 等. 基于不同模型的土壤有机质含量高光谱反演比较分析[J]. 光谱学与光谱分析, 2013, 33(1): 196-200.