

叶红云,熊黑钢,包青岭,等. 基于灰度关联-极限学习机的土壤有机质含量高光谱多尺度反演[J]. 江苏农业科学,2019,47(24):289-295.
doi:10.15889/j.issn.1002-1302.2019.24.064

基于灰度关联-极限学习机的土壤有机质含量高光谱多尺度反演

叶红云¹,熊黑钢²,包青岭¹,王 宁¹,马利芳¹

(1. 新疆大学资源与环境科学学院/教育部绿洲生态重点实验室,新疆乌鲁木齐 830046; 2. 北京联合大学应用文理学院城市系,北京 100083)

摘要:为了提高干旱区土壤有机质含量的预测精度,以新疆阜康人类不同干扰程度下的土壤有机质含量为研究对象,先利用小波变换对其原始光谱进行 1~8 层的离散小波分解,然后通过相关性分析确定最大分解层数,并进行 5 种常规数学变换以筛选敏感波段,最后引入灰色关联分析(GRA)筛选出具有较好预测能力的波段,结合极限学习机算法(ELM)分别建立不同分解层的土壤有机质含量高光谱预测模型。作为对照,同时采用传统相关分析方法选择的敏感波段构建偏最小二乘模型。研究表明:(1)小波变换获得的各层重构光谱中,随着分解层数增加,土壤光谱反射率与有机质含量的相关性降低,在 L6 之后二者相关性变化幅度不大且敏感波段数下降较多,即重构光谱对原始光谱信息的反映能力减弱,因此 L6 为本研究的最大分解层;(2)通过对比各分解层有机质含量模型,I 区(轻度干扰区)、II 区(中度干扰区)、III 区(重度干扰区)均是 L3、L4 建模效果最好;(3)无论是哪一个区,灰度关联-极限学习机所建模型均为最优,I、II、III 区验证集 R^2 和 RMSE(均方根误差)分别为 0.825、0.787、0.729 和 0.980、1.387、1.982,RPD(相对分析误差)分别达到 2.382、2.204、2.139,均优于传统的建模方法,且能对土壤有机质含量进行很好的预测。特别在有机质含量较低的情况下,灰度关联-极限学习机方法优势更加明显,表明该方法适用于干旱区有机质含量的预测。

关键词:高光谱;有机质;小波变换;灰色关联度;极限学习机;人类活动

中图分类号: TP79;S127;S153.6⁺21 **文献标志码:** A **文章编号:** 1002-1302(2019)24-0289-07

土壤有机质(SOM)主要由各种动植物的残体、微生物体及其分解和合成的各种有机物质组成,虽然其含量仅占土壤总量的很小一部分,但在土壤肥力上的作用却十分重要。不同土壤类型和区域由于气候以及人类干扰程度的不同,使得有机质分解程度不同,其光谱反射特性也存在差异。近年来,利用高光谱对土壤有机质进行估测研究因其便利性而被普遍使用^[1-3]。但土壤高光谱曲线是土壤属性的综合体现,存在很多与有机质无关的噪声,且信息相关性强、信息冗余等,因此对于高光谱噪声的去除成为近年来研究的热点^[4-6]。目前光谱去噪方法有很多,如 Savitzky-Golay 平滑去噪、移动平均法、傅里叶变换(FFT)滤波等,但这些方法能检测的信噪比信息有限。小波变换近几十年来在高光谱遥感领域以其在信号去噪和数据压缩方面独特的优势为地物光谱信息提取提供了新途径^[7-9]。但目前的研究大都以人类干扰的土壤为对象,且均是有机质含量较高的地区,未充分考虑到人类干扰程度不同且有机质含量较低时光谱的估测情况。灰钙土是新疆北部典型的土壤类型之一,其有机质质量分数大多低于 2%。研究表明,有机质含量较低会减弱其光谱反射率的吸收特性^[10-11],导致在估测有机质含量时存在光谱响应较弱、模型

预测精度较低等问题,因此有必要探究进一步提高预测精度的方法。

灰度关联法和分析各因素之间关联程度的方法,在光谱分析中可以很好地识别敏感波段,有利于优化模型精度^[12]。极限学习机(ELM)算法可随机产生输入层与隐含层间的连接权值及隐含层神经元的阈值,且在训练过程中无需调整,只需设置隐含层神经元的数量,便可获得唯一的最优解,与传统的 BP 神经网络算法相比,ELM 方法学习速度快、泛化性能好^[13]。虽然目前这 2 种方法应用较多,但将二者结合应用于干旱区小波变换最佳光谱层的确定中还未见报道。

本研究从人类干扰程度出发,结合植被特征、土地利用方式等指标,将研究区划分为轻度、中度和重度干扰区 3 种干扰类型,采用离散小波技术从土壤光谱数据中分别提取有机质信息,并利用相关分析和灰色关联分析法挑选有机质含量的敏感波段,结合极限学习机构建估测土壤有机质含量的模型,以期提高估测精度,为精准农业的发展提供借鉴和科学依据。

1 材料与方法

1.1 研究区概况

研究区位于新疆北部的阜康市(87°44'~88°46'E, 43°29'~45°45'N)。地势南高北低,平均海拔为 452 m。气候为典型的温带大陆性气候,夏季高温,冬季严寒,光能资源丰富,热量资源丰富,降水稀少且空间分布不均,年降水量为 163 mm,年蒸发潜力为 2 000 mm 左右^[14]。

本试验将研究区分为 I、II、III 3 个区。I 区为轻度干扰区,因距离人类居住地较远,并未开发利用,该区人类活动

收稿日期:2018-09-10

基金项目:国家自然科学基金(编号:41671198)。

作者简介:叶红云(1995—),女,河南驻马店人,硕士,主要从事干旱区资源与环境研究。E-mail:2359837706@qq.com。

通信作者:熊黑钢,博士,教授,博士生导师,主要从事资源环境研究。E-mail:heigang@buu.edu。

很少,偶有人类进入,基本保持了其原有风貌;Ⅱ区为中度干扰区,处于新疆生产建设兵团 102 团附近,大部分为废弃耕地,人类目前对其干扰较小;Ⅲ区为重度干扰区,位于研究区的西南部,人类对其开发利用程度较大,主要开发利用方式为林地(包括人工梭梭林地、榆树林地、育苗地),尚处于开发利用的初期阶段(表 1)。

参照余作岳提出的人类干扰分类体系^[15],研究中的Ⅰ区长期以来的主导干扰因子只是人类偶尔进入该区,并未对该区原有植被类型造成干扰,即几乎不引起土壤中组分的变化,属于无效干扰;而Ⅱ区、Ⅲ区经过人类开垦、植树等干扰活动后,使得植被类型发生改变,正常演替序列被打破,土壤中的盐分、有机质等组分也随之发生变化,即表 1 中的 3 种干扰改变了生态系统的正常动态发展,属于有效干扰。因此可以用轻度干扰区的各项指标作为对照来比较分析中度干扰、重度干扰区土壤有机质含量的变化等信息。

表 1 研究区 3 类典型干扰区基本情况

类型	干扰强度	植被类型	土壤类型	植被盖度(%)
Ⅰ	轻度	原生植被	灰钙土	>30
Ⅱ	中度	原生植被、经济作物	灰钙土	15~30
Ⅲ	重度	经济作物、人工林	灰钙土	<15

1.2 实地调查与采样

本研究于 2017 年 10 月进行了土壤样本数据采集。在Ⅰ区由南向北布设 5 条间距 500~700 m 的东西向采样线,每条采样线布设 6 个采样点;Ⅱ区按同样方向布设 5 条间距 400~600 m 的采样线,每条采样线上选择 5~7 个具有代表性的采样点;Ⅲ区面积较小,因此布设 6 条间距 200~400 m

的采样线,每条采样线布设 5 个采样点。每个区均为 30 个采样点,由此构成由 90 个采样点组成的空间网格,并分别对其进行 GPS(全球定位系统)定位。采样点具体分布见图 1。本次野外数据采集主要包括以下 2 个部分:

1.2.1 野外土壤高光谱数据的采集 光谱测量采用美国 ASD 公司生产的便携式 FieldSpec® 3Hi-Res 光谱仪,波段为 350~2 500 nm,350~1 000 nm 光谱采样间隔为 1.4 nm,光谱分辨率为 3 nm;1 001~2 500 nm 范围内光谱采样间隔为 1.1 nm,光谱分辨率为 8 nm。由于本试验光谱测量采用野外实测方法,为了减少太阳高度对光谱采集造成的不利影响,试验均选在当地时间 11:00—15:00、晴朗少云、无风的天气进行。每次采集光谱前对光谱仪进行白板校正以去除暗电流的影响。采用 25°视场角探头,且距采样地面 15 cm 处垂直角度,在每个采样点按梅花桩对表层土壤原始光谱进行采集,每个位置重复测量 10 次,得到的 50 条光谱曲线的平均值即为该采样点的实测光谱值,共测定 90 个样点的采样光谱曲线。野外测量时观察每个样点的实测光谱值,将出现异常的光谱曲线删除并重新测定,以便获得更加准确的实测光谱反射率。

1.2.2 土壤有机质含量的采集 土样的采集与高光谱数据的测定同时、同地进行,选择地势平坦、能代表样点周围区域特征的地点作为采样单元,每个采样点分别从周边 1 m 范围内不同的地点挖取 3 个 0~10 cm 的土坑分别采样,并混合均匀放入密封袋中,封口标记编号,用手持 GPS 定位后记录经纬度。将采集的样品带回实验室经自然风干、去除杂质后,研磨过 1 mm 筛,送至中国科学院新疆生态与地理研究所,由专业人士采用重铬酸钾容量-外加加热法测定^[16]有机质含量。

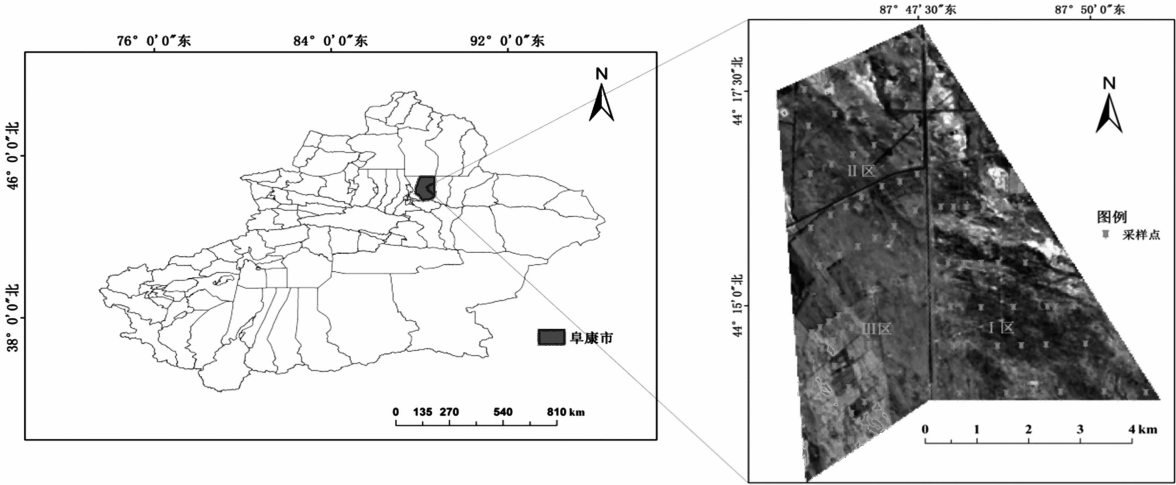


图1 研究区采样示意

1.3 高光谱数据处理

因为光谱仪器波谱两端的边缘波段存在较大的仪器噪声,首先剔除 350~399 nm 和 2 451~2 500 nm 波段。然后运用 OriginPro 9.0 对光谱反射率进行 Savitaky-Golay(9 点)平滑去噪处理。通过观察研究区 90 个采样点的土壤光谱曲线,发现野外实测光谱在 1 350~1 450 nm 和 1 830~1 950 nm 附近出现异常,这主要受到水分吸收带和大气的影

响,因此将其剔除。

1.4 离散小波变换

离散小波变换是基于傅里叶变换发展起来的数据分析方

法,可以将原始信号按照不同的尺度、频率进行多级分解,得到不同分解尺度下的小波系数。通过小波系数可以实现原始光谱的信息重构。借鉴前人研究结果^[17-18],本研究对原始光谱进行 1~8 层小波分解并重构各层光谱进行分析。

1.5 灰色关联度分析

灰色关联分析是通过灰色关联度来揭示 2 个因素的贴近程度,曲线几何形状越接近,灰色关联度越大。一般而言,若关联度 ≥ 0.8 ,则 2 个因素贴近程度很好;若关联度介于 0.6 与 0.8 之间,则贴近程度较好;若关联度小于 0.5 时,则认为基本不相关^[19-20]。

1.6 极限学习机(ELM)

ELM 是发展于单隐含层前馈神经网络的新型神经网络算法^[21]。在算法执行过程中随机设定输入层与隐含层之间的权值和阈值,无需反复调整迭代网络的输入权值及隐元的偏置,并且产生唯一的最优解,避免陷入局部最优解的情况,因此具有学习速度快和强泛化性的优点^[22]。

1.7 模型的建立与检验

模型的精度检验主要通过判定系数(coefficient of determination,简称 R^2)、均方根误差(root mean square error of calibration,简称 RMSE)和相对分析误差(residual prediction deviation,简称 RPD)来检验实测值和估测值的拟合效果。 R^2 越大, RMSE 越小,说明模型精度越高^[23]。Viscarra 等对模型 RPD 做了如下分类, $RPD < 1.0$, 预测能力极差; $1.0 < RPD < 1.4$, 预测能力较差; $1.4 < RPD < 1.8$, 预测能力一般; $1.8 < RPD <$

2.0 , 预测能力较好; $2.0 < RPD < 2.5$, 说明预测能力极好^[24]。

2 结果与分析

2.1 土壤有机质含量统计特征

选用 SPXY(光谱-理化值共生距离)算法^[25]分别计算出 I、II、III 区各个样品有机质含量之间的欧式距离,按照 3:2 的比例划分为建模集和检验集。研究区土壤有机质含量总体较低,且随着干扰强度的减少,经营方式逐渐回归自然状态,有机质含量逐渐增加,变异系数逐渐减少(表 2)。这是因为 I 区原始植被状态保存得完好,主要分布有梭梭、柽柳、猪毛菜、盐爪爪等,植被覆盖度较高;而 III 区受人类活动干扰的影响,土地利用方式多样,主要有农林、枸杞育苗地以及其他农业土地利用类型,使得空间异质性与其他 2 个区相比加强,变异系数也增加。

表 2 土壤有机质含量的描述统计量

类型	样本集	样本数 (个)	含量(g/kg)				变异系数
			最小值	最大值	均值	标准差	
I	全集	30	6.507	14.707	10.257	2.004	0.195
	建模集	18	7.952	14.170	10.344	1.753	0.169
	验证集	12	6.507	14.707	10.127	2.409	0.238
II	全集	30	6.182	15.221	8.939	2.284	0.255
	建模集	18	6.786	14.797	8.658	1.884	0.218
	验证集	12	6.182	15.221	9.360	2.818	0.301
III	全集	30	3.619	13.107	7.977	2.297	0.288
	建模集	18	3.619	12.133	7.483	1.935	0.259
	验证集	12	4.695	13.107	8.717	2.670	0.306

2.2 不同程度人类干扰下原始光谱反射率的小波变换分析

应用小波分解与重构去噪方法需要对各种光谱信号进行分解以选择分解层次,从而在重构时达到去噪与保留细微光谱特征的平衡点。在 MATLAB R2017b 中选取常用的 db4 函数作为小波母函数,并对原始高光谱数据进行 8 层离散小波分解,然后将分解后的每层小波系数分别进行光谱重构(图 2)。随着分解层数的增加,吸收谷由较为“尖锐”逐渐变得“圆润”。与 I0 相比, I1、I2 的光谱反射率曲线变化不大,而 I3~I5 则凸显了光谱曲线的局部细节,使得波峰、波谷更加清晰,到 I6、I7 光谱反射率趋于平滑, I8 则几乎看不到吸收谷。表明适当的分解尺度可以减少背景和噪音的影响,但分解尺度过高可能会导致某些有效光谱信息的丢失。I 区、II 区、III 区的光谱反射率曲线形态大致相同,土壤高光谱反射率曲线呈现 I 区 < II 区 < III 区,即随着人类干扰程度的增加,土壤高光谱反射率增强。

2.3 不同程度人类干扰下 SOM 与重构光谱的相关性分析

以 0.05 显著性水平为阈值,通过对比不同分解层数与土壤有机质含量的相关性来选择适宜的分解层(表 3)。随着分解层的增加, I 区、II 区、III 区通过 0.05 显著性水平检验的相关系数逐渐减小,在 I6 附近变化程度不大,且通过检验的敏感波段数均在 I6 处达到最高值,分别为 405、198、167 个,而 I7、I8 的敏感波段数量则呈现快速下降的状态。表明分解层数的增加,在一定程度上使得小波系数输出原始光谱信息的能力逐渐减弱。因此,本研究选取相关性较强且通过的敏感波段数量较多的前 6 层特征光谱用于土壤有机质含量模型的

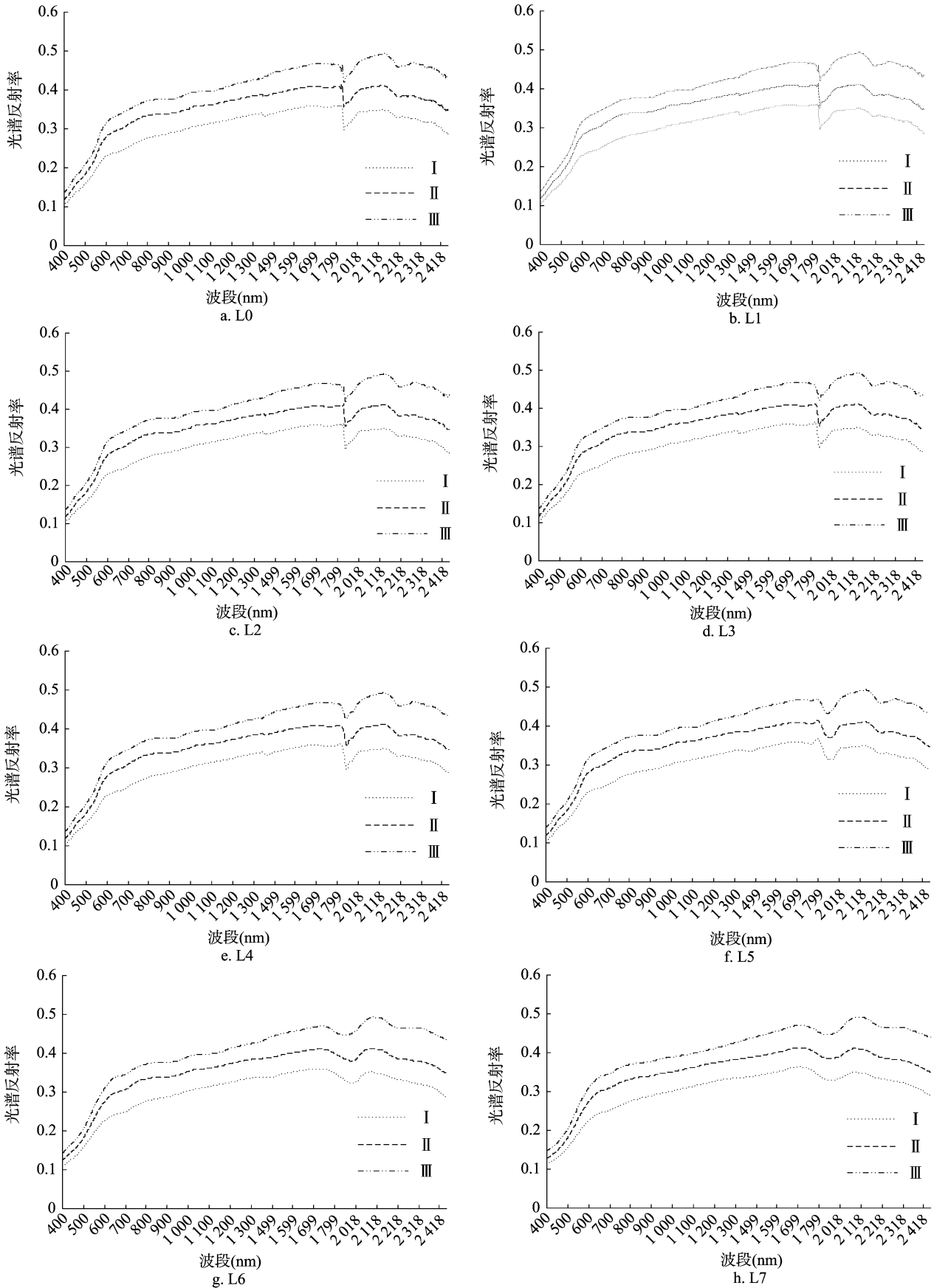
反演研究。

2.4 各层敏感波段的选取

基于以上分析,将通过 0.05 水平检验的原始光谱(I0)和小波分解后(I1~I6)的特征光谱分别进行 $1/R$ 、 $\lg R$ 、 R' 、 $(1/R)'$ 、 $(\lg R)'$ 等 5 种数学变换,以获得各分解层的敏感波段。由于高光谱数据量较大,且通过小波变换后信息含量更大,使得选择敏感波段难度加大,因此本研究以通过 0.05 显著性水平以及相关性最大的原则选择每层的敏感波段(表 4)。总体来看, I 区敏感波段主要分布在可见光区(752~875 nm)和近红外区(1 820 nm), II 区相关性最大的波段主要集中在可见光区(400~600 nm),而 III 区主要集中在近红外区(1 200~1 300 nm);从不同数学变换来看,无论是哪个区,微分变换后整体上相关系数均有所增加;对于不同分解层来说, I 区、II 区、III 区相关系数较高的波段均主要集中在 I3 和 I4 层,以 I 区的 $1/R$ 为例, I4 层相关系数最高,为 0.498 3,比 I0 处的相关系数提高了 0.05 左右,说明微分处理可以将可见光区微弱的有效信息放大,有利于敏感波段的选择。

2.5 特征光谱变换与 SOM 灰色关联度分析

单纯依赖简单的相关性分析确定敏感波段并不能完全确定有机质含量的有效信息,具有一定的局限性。为了进一步衡量不同分解层所选的敏感波段与有机质的关联程度,更好地筛选出表征有机质的敏感光谱指标因素,本研究对其进行了灰色关联分析(表 5)。总体来看, I 区、II 区和 III 区各层特征光谱及其不同数学变换与有机质含量的灰色关联度均高于小波变换前(I0)所对应的数据,表明小波变换在去除噪声的



L0 为未经小波变换前的原始光谱反射率曲线

图2 不同程度人类干扰下小波分解重构光谱

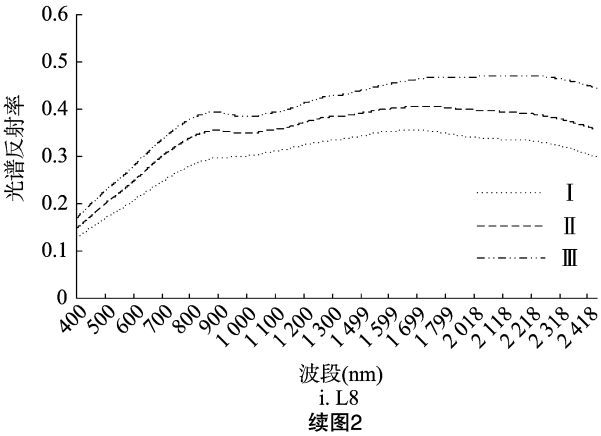


表 3 不同程度人类干扰下 SOM 与各层敏感光谱的相关性分析

分解层	敏感波段数(个)			波段(nm)			最大相关系数		
	I	II	III	I	II	III	I	II	III
L1	340	177	151	1 824	522	1 254	-0.432	-0.418	0.369
L2	343	179	150	1 824	450	1 254	-0.427	-0.392	0.368
L3	344	179	149	1 822	450	1 243	-0.414	-0.392	0.366
L4	350	181	135	1 820	451	1 245	-0.388	-0.375	0.367
L5	354	185	114	775	516	1 247	-0.381	-0.374	0.365
L6	405	198	167	758	488	1 236	-0.382	-0.372	0.363
L7	333	171	109	757	458	1 250	-0.38	-0.370	0.363
L8	302	164	82	862	427	1 301	-0.378	-0.369	0.362

表 4 SOM 与各层特征光谱的不同数学变换的最大相关性及其波段所处位置

类型	分解层	R		1/R		lgR		R'		(1/R)'		(lgR)'	
		波段(nm)	r	波段(nm)	r	波段(nm)	r	波段(nm)	r	波段(nm)	r	波段(nm)	r
I	L0	1 825	-0.433 6	1 824	0.441 2	1 824	-0.433 7	864	0.632 3	864	-0.638 3	864	0.645 8
	L1	1 824	-0.431 8	1 824	0.439 9	1 824	-0.438 1	864	0.577 0	807	-0.583 7	864	0.594 3
	L2	1 824	-0.426 5	1 824	0.434 7	1 824	-0.432 9	863	0.614 1	862	-0.608 5	862	0.633 1
	L3	1 822	-0.433 8	1 821	0.453 7	1 821	-0.441 0	865	0.692 0	835	-0.643 5	865	0.643 7
	L4	1 820	-0.407 6	765	0.498 3	766	-0.449 3	845	0.681 9	840	-0.643 8	1 036	0.650 0
	L5	775	-0.381 1	755	0.427 4	761	-0.392 3	855	0.596 7	855	-0.546 0	855	0.542 0
	L6	758	-0.381 8	752	0.398 8	755	-0.393 2	2 079	-0.577 6	875	-0.508 9	875	0.506 6
II	L0	450	-0.418 3	523	0.376 9	523	-0.403 2	408	-0.566 2	473	-0.596 6	415	-0.424 8
	L1	522	-0.418 3	523	0.376 8	523	-0.403 1	427	-0.558 7	557	-0.570 4	427	-0.428 1
	L2	450	-0.418 3	522	0.376 7	522	-0.403 0	405	-0.494 6	559	-0.546 4	404	-0.422 4
	L3	450	-0.422 4	524	0.379 6	524	-0.405 4	401	0.506 3	559	-0.627 4	400	0.431 4
	L4	450	-0.428 5	524	0.379 6	524	-0.402 9	434	-0.571 4	456	-0.528 3	403	-0.498 6
	L5	516	-0.419 0	516	0.377 4	516	-0.403 7	412	-0.469 1	471	-0.561 8	400	-0.493 8
	L6	488	-0.420 8	488	0.379 4	489	-0.405 1	406	-0.465 9	578	-0.518 2	584	0.474 1
III	L0	1 254	0.367 3	1 244	-0.269 4	1 254	0.318 0	1 248	-0.563 6	1 240	0.5318	1 248	-0.525 9
	L1	1 254	0.367 1	1 254	-0.269 3	1 254	0.318 0	1 252	0.569 7	1 240	0.564 5	1 248	-0.515 3
	L2	1 254	0.366 8	1 244	-0.269 3	1 244	0.317 8	1 205	0.580 0	1 227	0.570 5	1 207	-0.518 7
	L3	1 243	0.367 7	1 244	-0.269 1	1 244	0.318 5	1 205	0.581 4	1 227	0.587 0	1 203	0.517 9
	L4	1 245	0.366 7	1 245	-0.269 7	1 245	0.318 8	1 206	0.582 3	1 233	0.587 5	1 203	0.527 7
	L5	1 247	0.365 8	1 315	-0.267 9	1 248	0.316 6	1 191	0.545 8	1 247	0.573 3	1 191	0.513 8
	L6	1 236	0.366 6	1 277	-0.264 7	1 237	0.317 4	1 191	0.542 3	1 238	0.436 5	1 202	0.420 9

注：“r”代表波段对应的相关系数。

同时也最大程度地保留了光谱中的有效信息,使得小波变换的各层重构光谱与有机质含量的关联程度增加。从同一分解层来看,经过微分处理后小波变换重构特征光谱与有机质含量之间的灰色关联度总体上均高于对应的未经微分处理的数

学变换,如 I 区 L0,其关联度排序依次是 $(1/R)' > R' > (lgR)' > 1/R > lgR > R$,进一步说明在小波去噪的基础上,通过微分数学变换可以显著增强光谱曲线上局部位置对有机质含量变化的响应差异。

表 5 不同程度人类干扰下各层特征光谱不同数学变换的灰色关联度

类型	分解层	灰色关联度					
		<i>R</i>	1/ <i>R</i>	lg <i>R</i>	<i>R</i> '	(1/ <i>R</i>)'	(lg <i>R</i>)'
Ⅰ	L0	0.642 8	0.685 1	0.672 6	0.707 6	0.749 3	0.691 6
	L1	0.679 9	0.694 3	0.682 4	0.727 5	0.715 8	0.704 0
	L2	0.670 5	0.694 0	0.684 3	0.709 3	0.698 9	0.722 0
	L3	0.681 3	0.689 1	0.680 9	0.736 5	0.703 7	0.691 1
	L4	0.646 4	0.687 3	0.681 6	0.735 5	0.697 9	0.672 8
	L5	0.665 2	0.687 8	0.691 2	0.729 6	0.704 1	0.695 0
	L6	0.677 4	0.703 2	0.691 0	0.707 6	0.777 2	0.733 1
Ⅱ	L0	0.652 0	0.669 0	0.673 3	0.688 1	0.682 3	0.663 9
	L1	0.675 7	0.669 1	0.682 6	0.692 6	0.708 5	0.702 2
	L2	0.682 0	0.669 2	0.673 4	0.680 1	0.703 6	0.707 4
	L3	0.682 0	0.669 0	0.693 3	0.662 4	0.706 1	0.702 1
	L4	0.681 9	0.649 0	0.676 4	0.678 7	0.664 2	0.694 4
	L5	0.676 5	0.670 1	0.691 5	0.673 1	0.712 7	0.694 7
	L6	0.680 3	0.674 0	0.713 8	0.699 3	0.739 4	0.691 5
Ⅲ	L0	0.561 3	0.616 8	0.601 6	0.613 7	0.644 3	0.658 1
	L1	0.632 1	0.632 4	0.620 2	0.642 8	0.673 1	0.662 9
	L2	0.624 8	0.616 2	0.619 7	0.608 2	0.672 7	0.672 5
	L3	0.627 9	0.627 7	0.605 5	0.673 0	0.634 3	0.669 3
	L4	0.600 3	0.616 9	0.634 8	0.684 1	0.707 4	0.651 5
	L5	0.612 6	0.636 6	0.619 5	0.616 7	0.705 5	0.669 2
	L6	0.630 3	0.637 2	0.640 3	0.632 1	0.713 7	0.697 6

注:灰色关联分析中取分辨系数 $\rho=0.5$;加粗字体为每层分解层灰色关联度最大处。

2.6 SOM 反演模型的优选

分别选取Ⅰ区、Ⅱ区和Ⅲ区的原始光谱和重构光谱及5种数学变换后与有机质含量相关系数最大的波段为自变量、土壤有机质含量为因变量,结合PLSR和ELM算法构建有机质含量的估测模型,通过对比得出灰色关联-极限学习机模型建模精度均优于对应的相关系数-PLSR模型,由于各层小波重构光谱所建立的模型太多,此处仅对各层灰色关联-ELM模型进行讨论。此外,由于小波变换把原始光谱分解成不同的离散光谱,每层光谱均表征了原始光谱中的特定信息,如果单一地选择某一层重构特征光谱建模,容易导致其他分解层中有效信息的丢失,使得所选变量不能完全反映有机质含量信息,对建模精度产生影响。因此,本研究再以各层相关系数最大的波段组合以及灰色关联度分析所选的波段组合,分别建立PLSR和ELM模型。仅列出利用相关系数最大波段组合所建的PLSR模型和灰色关联所建立的ELM模型,分别用L-R、L-GRD表示(表6)。

从各分解层的建模效果来看,以L3、L4所构建的模型最优。以Ⅰ区为例,前4层构建的验证集模型 R^2 呈不同程度的增加,到L4模型 R^2 最优(0.807),均方根误差最小(1.093 g/kg),且RPD大于2(2.109),表明第4层能够极好地预测有机质含量。此后,随着分解层数的增加,各模型的 R^2 降低,RMSE增大,RPD也随之减小。这与王延仓等的研究结果^[17,26]一致,即利用小波重构原始光谱时,随着分解层数的增加,使得原始光谱中噪声和信号信息更加明显,有利于有效光谱信息和噪声的分离。但分解尺度过大,会导致有效信息的丢失而使得模型误差较大。从总的建模效果来看,对于Ⅰ、Ⅱ、Ⅲ区利用相关分析-偏最小二乘法所建模型验证集决定系数 R^2 分别为0.568、0.517、0.544, RMSE分别为1.620、1.771、3.208, RPD均小于1.4,预测能力较差。而利

用灰色关联-极限学习机建模效果均最好。以验证集所建模型为例,与原始光谱所建模型相比,Ⅰ、Ⅱ、Ⅲ区 R^2 分别提高了9.8%、8.9%、13.6%, RMSE分别减少了18.3%、2.5%、7.3%, RPD分别增大了39.9%、22.2%、46.4%,这可能是因为单一地利用某一层的小波重构光谱,并不能完全将光谱中的细节体现出来,通过灰色关联度分析将不同层的特征光谱组合起来,可以很好地体现每一层特征光谱与有机质含量的关系,增加了模型的稳定性和模型精度。同时,灰色算法结合光谱和有机质信息,能够较好地挑选敏感波段,且ELM算法是在BP神经网络基础上的优化新型算法,具有学习速度快且泛化性能好的优点。

此外,Ⅰ区不同分解层下的土壤有机质建模精度均高于相应的Ⅱ区、Ⅲ区,这主要是因为Ⅰ区为轻度干扰区,基本保持原有状态,有机质分布较为均质,而Ⅱ区、Ⅲ区由于受到人类不同干扰程度的影响,不同地块的土地利用方式不同,使得有机质含量分布变异性较大,导致模型预测精度相对较低。

3 结论

对于人类不同干扰程度土壤的小波重构光谱来说,对其最大的影响因素主要是有机质含量的高低,整体上随着人类干扰程度的增加,小波分解的重构光谱反射率随之升高,排序为Ⅰ区<Ⅱ区<Ⅲ区。

在确定最佳分解尺度中,Ⅰ区、Ⅱ区、Ⅲ区土壤有机质和重构光谱通过0.05显著性水平检验的敏感波段在L6处变化幅度不大,且通过检验的敏感波段数达到最大值,分别达405、198、167个。因此L6为本研究的最大分解层。

对于不同层的重构光谱来说,小波分解使得重构光谱及各种数学变换与有机质含量之间的相关性和灰色关联度有不同程度的提高;对于同一分解层来说,经过微分变换后与有机

表 6 土壤有机质含量反演最优模型的建模集与验证集结果

类型	模型	建模集		验证集		
		R^2	RMSE	R^2	RMSE	RPD
I	L0	0.741	0.919	0.751	1.199	1.703
	L1	0.771	0.884	0.758	1.199	1.709
	L2	0.776	0.841	0.765	1.185	1.733
	L3	0.782	0.815	0.785	1.157	1.896
	L4	0.779	0.795	0.807	1.093	2.109
	L5	0.682	0.985	0.723	1.278	1.574
	L6	0.642	1.022	0.644	1.486	1.385
	L-R	0.568	1.163	0.583	1.620	1.161
	L- GRA	0.793	0.809	0.825	0.980	2.382
II	L0	0.745	0.964	0.723	1.422	1.804
	L1	0.731	0.959	0.714	1.586	1.551
	L2	0.703	1.033	0.681	1.526	1.671
	L3	0.794	0.842	0.762	1.398	1.816
	L4	0.764	0.872	0.753	1.380	1.765
	L5	0.711	0.995	0.704	1.473	1.694
	L6	0.638	1.131	0.662	1.615	1.509
	L-R	0.517	1.277	0.573	1.771	1.366
	L- GRA	0.796	0.828	0.787	1.387	2.204
III	L0	0.671	1.093	0.642	2.138	1.461
	L1	0.682	1.017	0.677	1.993	1.542
	L2	0.688	1.151	0.669	2.300	1.273
	L3	0.696	1.098	0.682	2.029	1.451
	L4	0.712	1.028	0.693	1.923	1.810
	L5	0.654	1.118	0.636	2.136	1.396
	L6	0.576	1.371	0.559	2.578	1.125
	L-R	0.544	1.274	0.470	3.208	0.997
	L- GRA	0.751	0.945	0.729	1.982	2.139

质含量之间的相关性和灰色关联度均优于微分变换前,表明小波变换和微分结合有利于挖掘光谱中的有效信息,提升其与有机质含量的相关性。

对 3 个区原始光谱分别进行不同尺度的分解后 (L1 ~ L6),总体上 L3、L4 的建模效果最好,与 L0 建模效果相比,验证集 R^2 分别提高了 0.06、0.04、0.05,表明 L3、L4 在去噪的同时又最大程度上保留了光谱的有效信息。

不论在哪个区,利用灰度关联与 ELM 结合均优于传统建模方法,且 I 区、II 区、III 区均以 L- GRA 所建模型的精度最高。其验证集 R^2 均大于 0.73,RPD 均大于 2,相比原始 L0 所建模型,RPD 分别提高了 0.679、0.400、0.678。预测能力由较差上升为极好,表明灰色关联-极限学习机方法有利于最大程度地筛选有机质敏感光谱信息,尤其适用于干旱区有机质含量较低情况下有机质含量的预测,且不受人类干扰程度的影响,是估算干旱区有机质含量的优选模型。

参考文献:

- [1] 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.
- [2] 石朴杰,王世东,张合兵,等. 基于高光谱的复垦农田土壤有机质含量估测[J]. 土壤,2018,50(3):558-565.
- [3] 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.

- [4] 南锋,朱洪芬,毕如田. 黄土高原煤矿区复垦农田土壤有机质含量的高光谱预测[J]. 中国农业科学,2016,49(11):2126-2135.
- [5] 黄威,高太长,刘磊,等. 一种基于改进累积方差百分比的红外高光谱数据降噪方法[J]. 光谱学与光谱分析,2016,36(11):3625-3629.
- [6] 王一丁,赵铭钦,刘鹏飞,等. 基于高光谱分析的植烟土壤有机质和全氮含量预测研究[J]. 中国烟草学报,2016,22(3):44-51.
- [7] 陈红艳,赵庚星,李希灿,等. 小波分析用于土壤速效钾含量高光谱估测研究[J]. 中国农业科学,2012,45(7):1425-1431.
- [8] 林鹏达,佟志军,张继权,等. 基于 CWT 的黑土有机质含量野外高光谱反演模型[J]. 水土保持研究,2018,25(2):46-52,57.
- [9] 梁福明,熊黑钢,王芳,等. 基于小波分析的土壤速效 K 含量高光谱反演[J]. 干旱区地理,2015,38(2):320-326.
- [10] 单海斌,蒋平安,颜安,等. 基于高光谱数据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.
- [11] 刘焕军,张柏,赵军,等. 黑土有机质含量高光谱模型研究[J]. 土壤学报,2007,44(1):27-32.
- [12] Li M L, Li X C, Tian Y, et al. Grey relation estimating pattern of soil organic matter with residual modification based on hyper-spectral data[J]. Journal of Grey System,2016,28(4):27-39.
- [13] 周鹏,杨玮,李民赞,等. 基于灰度关联-极限学习机的土壤全氮预测[J]. 农业机械学报,2017,48(增刊1):271-276.
- [14] 张林静,岳明,张远东,等. 新疆阜康绿洲荒漠过渡带植物群落物种多样性特征[J]. 地理科学,2003,23(3):329-334.
- [15] 余作岳. 热带亚热带退化生态系统植被恢复生态学研究[M]. 广州:广东科技出版社,1996.
- [16] 鲍士旦. 土壤农化分析[M]. 3 版. 北京:中国农业出版社,2000.
- [17] 王延仓,杨贵军,朱金山,等. 基于小波变换与偏最小二乘耦合模型估测北方潮土有机质含量[J]. 光谱学与光谱分析,2014,34(7):1922-1926.
- [18] 蔡亮红,丁建丽. 小波变换耦合 CARS 算法提高土壤水分含量高光谱反演精度[J]. 农业工程学报,2017,33(16):144-151.
- [19] 邓聚龙. 灰色系统基本方法[M]. 武汉:华中科技大学出版社,2001.
- [20] 曹明霞. 灰色关联分析模型及其应用的研究[D]. 南京:南京航空航天大学,2007.
- [21] Nahvi B, Habibi J, Mohammadi K, et al. Using self-adaptive evolutionary algorithm to improve the performance of an extreme learning machine for estimating soil temperature[J]. Computers and Electronics in Agriculture,2016,124:150-160.
- [22] 张森,孔盼,李雁华,等. 基于极限学习机的土壤硝态氮预测模型研究[J]. 农业机械学报,2016,47(6):93-99.
- [23] 崔霞,宋清洁,张瑶瑶,等. 基于高光谱数据的高寒草地土壤有机碳预测模型研究[J]. 草业学报,2017,26(10):20-29.
- [24] 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] 毛博慧,孙红,刘豪杰,等. 基于正交变换与 SPXY 样本划分的冬小麦叶绿素诊断[J]. 农业机械学报,2017,48(增刊1):160-165.
- [26] 陈至坤,张菡洁,王玉田,等. 基于小波变换的矿物油荧光光谱数据处理方法[J]. 激光杂志,2016,37(10):78-81.