

# 光谱数据预处理对潮间带沉积物氮 LSSVM 模型的影响研究

吕美蓉<sup>1</sup>, 任国兴<sup>1,2</sup>, 李雪莹<sup>1</sup>, 范萍萍<sup>1</sup>, 刘杰<sup>1</sup>, 孙中梁<sup>1</sup>, 侯广利<sup>1</sup>, 刘岩<sup>1\*</sup>

1. 齐鲁工业大学(山东省科学院)海洋仪器仪表研究所, 山东省海洋监测仪器装备技术重点实验室, 国家海洋监测设备工程技术研究中心, 山东 青岛 266100
2. 中国海洋大学信息科学与工程学院, 山东 青岛 266100

**摘要** 光谱数据变换和光谱特征波长提取是二种重要的光谱预处理方法, 对消除环境等干扰具有重要的作用。以往文献主要对比研究不同的光谱数据变换方法, 光谱特征波长提取方法的对比研究以及二者的组合研究较少。为了获取适宜的光谱预处理方法, 提高潮间带沉积物氮的最小二乘支持向量机(LSSVM)模型精度, 研究了4种光谱变换方法与3种特征波长提取方法组合对沉积物氮 LSSVM 模型精度的影响, 以期实现潮间带沉积物氮的精确预测。研究结果表明, 多元散射校正(MSC)或标准正态变换(SVN)光谱变换方法提高了光谱与氮含量的相关性, 最高相关系数分别达到0.69和0.71; 并且提高了LSSVM模型的预测精度, 模型的预测 $R^2$ 和RPD分别为0.88, 0.87和2.78, 2.69。无信息变量消除(UVE)特征波长提取方法也提高了LSSVM模型的预测精度, 模型预测 $R^2$ 和RPD分别0.89和2.70。但是, UVE提取的特征波长并不都与氮含量具有高相关性。此外, 组合运用UVE特征波长提取方法和MSC或SVN光谱变换方法, 也提高了模型预测精度, 但并不优于单独运用UVE特征波长提取方法或单独运用MSC及SVN光谱变换方法。研究结果可为潮间带沉积物氮估算和光谱数据预处理提供技术参考。

**关键词** 光谱数据变换; 光谱特征波长提取; 光谱预处理; 潮间带沉积物

**中图分类号:** O657.3 **文献标识码:** A **DOI:** 10.3964/j.issn.1000-0593(2020)08-2409-06

## 引言

可见-近红外光谱法具有快速、便捷等优点, 已广泛应用于土壤/沉积物氮含量估算<sup>[1]</sup>。尽管采用光谱法预测氮含量的研究比较多, 但是以往主要针对土壤或湖泊沉积物, 未有潮间带沉积物氮含量的预测研究。潮间带是海陆交替的过渡地带之一, 人类活动造成大量氮元素被输送到潮间带地区, 并最终蓄积在沉积物中。在一定的条件下(温度、pH值、氧化还原电位等), 沉积物中蓄积的氮元素会通过一系列物理、化学和生物过程释放到水体中, 影响着潮间带及近岸水域环境<sup>[2]</sup>。因此定量化快速监测潮间带沉积物氮含量具有重要的意义。

由于土壤/沉积物成分复杂, 而且在光谱采集过程中往往受到仪器、环境等干扰, 因此光谱信息中包含了大量的噪音和冗余信息。采用原光谱数据建模, 不但计算复杂, 而且模型预测精度低。有必要探寻一种方法, 在保留有用光谱信

息的同时, 去除噪声, 剔除冗余信息, 为氮含量估算提供良好的数据基础。光谱变换、特征波长提取是两个重要的去除噪音和冗余信息的方法。光谱变换不仅可以去噪, 而且可以减少光谱信息的重叠度, 集聚与加强有用的光谱信息<sup>[3]</sup>。特征波长提取可以剔除冗余变量, 保留相关信息变量<sup>[4]</sup>。然而以往的研究主要对比研究不同的光谱数据变换方法, 光谱特征波长提取方法的对比研究以及二者的组合研究较少。有必要研究特征波长提取方法及其与光谱变换方法的组合, 优选最佳的光谱数据预处理方法, 以期实现最优的去噪、去冗余信息效果。

支持向量机(support vector machine, SVM)是一种无监督式学习算法, 学习泛化能力很强。其中, 最小二乘支持向量机(least square support vector machine, LSSVM)是一种改进的支持向量机模型, 采用最小二乘线性系统作为损失函数, 代替传统SVM采用的二次规划方法, 进一步简化了计算复杂性, 并提高了预测结果的准确性<sup>[5]</sup>。本研究在潮间带沉积物光谱采集的基础上, 组合运用4种光谱变换方法(多

收稿日期: 2019-07-08, 修订日期: 2019-11-12

基金项目: 山东省自然科学基金项目(ZR2019PD011, ZR2018LD007), 山东省科学院青年基金项目(2019QN0035, 2018QN0037)资助

作者简介: 吕美蓉, 女, 1983年生, 齐鲁工业大学(山东省科学院)海洋仪器仪表研究所助理研究员 e-mail: 444868063@qq.com

\* 通讯联系人 e-mail: sqdqliuyan@126.com

元散射校正、标准正态变换、一阶微分、二阶微分)和 3 种特征波长提取方法(连续投影算法、无信息变量消除法、遗传算法)以去除光谱噪音,挖掘沉积物氮敏感波长,突出相关光谱特征。通过比较沉积物氮的 LSSVM 模型精度,评价光谱预处理组合方法的效果,以期为潮间带沉积物氮的光谱速测提供研究基础和参考。

## 1 实验部分

### 1.1 沉积物样品采集

于青岛崂山区会场附近海洋潮间带采集了 197 份沉积物样品,采样深度为 0~10 cm。每个采样点之间至少间隔 10 m。采集的潮间带沉积物主要由黏土和粉砂组成。将采集的样品自然风干,研磨、过 60 目筛,用于光谱采集和化学分析。

### 1.2 分析测定方法

沉积物氮含量采用过高氯酸-硫酸消化法测定<sup>[6]</sup>,以此作为光谱建模的标准值。光谱反射率测定采用海洋光学 QE65000 光谱仪,光谱采样间隔为 1 nm,积分时间 600 ms,谱区范围 200~1 100 nm。取 3~5 g 沉积物样品放在自制样品盒中,轻轻刮平,采用 45° 视场角光纤探头进行光谱采集。每个土壤样品采集 5 次光谱反射率,取平均值。为减少噪声影响,剔除信噪比较低的边缘波段,保留 230~970 nm 的光谱数据。

### 1.3 数据处理

#### 1.3.1 光谱预处理

采用多元散射校正(multiplicative scatter correction, MSC)、标准正态变换(standardized normal variation, SVN)、一阶微分和二阶微分进行沉积物光谱数据变换。采用连续投影算法(successive projections algorithm, SPA)、无信息变量消除法(uninformative variable elimination, UVE)和遗传算法(genetic algorithm, GA)进行特征波长提取。

#### 1.3.2 模型建立

采用 Kennard-Stone(KS)方法对 197 个沉积物样品进行分类。采用最小二乘支持向量机法(least squares support vector machine, LSSVM)进行建模。

#### 1.3.3 模型检验

模型检验采用决定系数( $R^2$ )、误差均方根(RMSEC)和剩余估计偏差(RPD)为评价参数。当  $R^2$  越大, RMSEC 越小,表示模型的精度越好。此外,当  $RPD < 1.0$  时,表明模型预测能力很差,模型不可靠;当  $1.0 < RPD < 1.4$  时,表明模型预测能力较差;当  $1.4 < RPD < 1.8$  时,表明模型预测能力较好,可以对样本进行估测;当  $1.8 < RPD < 2.0$  时,表明模型预测能力好,可对样本进行定量估测,当  $2.0 < RPD < 2.5$  时,表明模型具有很好的定量预测能力;当  $RPD > 2.5$  时,模型具有极好的预测能力<sup>[7]</sup>。

## 2 结果与讨论

### 2.1 光谱数据变换

图 1 显示了潮间带沉积物氮含量与光谱反射率的正相关性。从图中可以看出,随着波长增加,原光谱反射率与氮含量的相关系数呈现先增加后降低的趋势,其中 350~620 nm 波段的相关系数较高,位于 0.40~0.55 之间,最敏感的波长出现在 503 nm 处,相关性为 0.55。经过一阶微分和二阶微分光谱变换后,整个波段的相关系数起伏非常大,连续性差,且与原光谱相比,大多数波长的相关系数降低了。经过多元散射校正(MSC)和标准正态变换(SVN)光谱变换后,相关性从 230~325 nm 波段范围呈现上升趋势,在 330~425 nm 波段范围呈现急剧下降又急剧上升的趋势,之后从 425~554 nm 波段范围内继续升高,但增加幅度较为平缓,到 554 nm 处相关性曲线再次呈现下降—上升—再下降的趋势。相对于原光谱,除 355 和 740 nm 左右,其他波段的相关系数都增加了。这暗示着, MSC 或 SVN 可以降低内外部环境因子的干扰,将原本湮没在原光谱数据中的一些微弱的沉积物氮特征光谱信息挖掘出来。

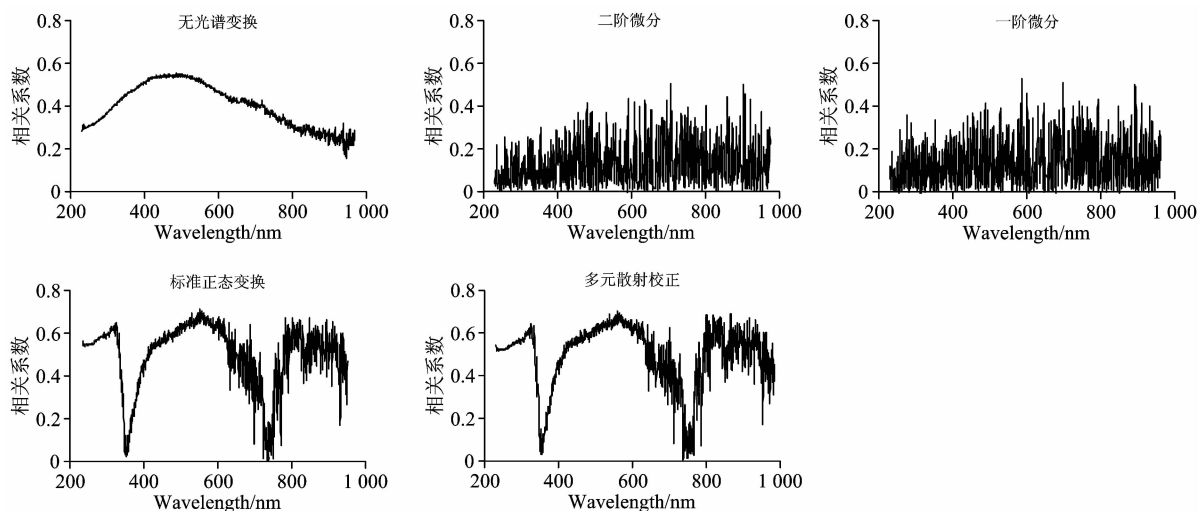


图 1 潮间带沉积物氮含量与光谱反射率的正相关性

Fig. 1 Positive correlation between nitrogen content and spectral reflectivity of sediment in intertidal zone

从表 1 可以看出, 光谱变换对模型的建模精度影响不大。无论光谱变换与否, 模型的建模集  $R^2$  都很高 ( $>0.99$ ), RMSEC 都很低 ( $<0.02$ )。但是, 光谱变换对模型的预测精度影响较大。基于原光谱构建的模型检验集  $R^2$ , RMSEC 和 RPD 分别为 0.83, 0.13 和 2.37。经过一阶微分和二阶微分光谱变换后, 模型的检验集  $R^2$  极大的降低了, 分别为 0.62 和 0.51, 且 RPD 都小于 1.4, 这可能与一阶微分和二阶微分降低了光谱反射率与氮含量的相关性有关。但是, MSC 和 SVN 均提高了模型的检验集  $R^2$ , 降低了 RMSEC, 且 RPD 都高于 2.5, 这暗示着 MSC 和 SVN 均可以强化沉积物氮的相关信息, 提高模型预测精度。

表 1 光谱变换和特征波长提取对潮间带沉积物氮 LSSVM 模型精度的影响

Table 1 Effect of spectral transformation and characteristic wavelength extraction on the accuracy of LSSVM model of nitrogen in intertidal sediments

光谱变换	特征波长提取	建模集		检验集		
		$R^2$	RMSEC	$R^2$	RMSEC	RPD
无	无	1	0.02	0.83	0.13	2.37
一阶微分	无	1	0.00	0.62	0.14	1.37
二阶微分	无	0.99	0.03	0.51	0.15	1.22
MSC	无	1	0.01	0.88	0.12	2.78
SVN	无	1	0.02	0.87	0.12	2.69
无	SPA	0.80	0.13	0.77	0.16	1.91
一阶微分	SPA	0.87	0.11	0.47	0.23	1.33
二阶微分	SPA	0.81	0.14	0.54	0.20	1.47
MSC	SPA	0.79	0.14	0.77	0.14	2.08
SVN	SPA	0.81	0.13	0.79	0.15	2.11
无	UVE	0.94	0.07	0.89	0.12	2.70
一阶微分	UVE	1	0.00	0.79	0.17	1.93
二阶微分	UVE	1	0.01	0.78	0.17	2.04
MSC	UVE	0.96	0.06	0.86	0.11	2.72
SVN	UVE	0.93	0.08	0.87	0.13	2.53
无	GA	0.77	0.15	0.39	0.24	1.28
一阶微分	GA	0.88	0.13	0.45	0.23	1.32
二阶微分	GA	0.06	0.29	0.22	0.29	1.07
MSC	GA	0.69	0.16	0.55	0.23	1.42
SNV	GA	0.88	0.11	0.72	0.15	1.88

## 2.2 特征波长提取

识别有用的特征波长, 剔除无关的特征波长, 突出潮间带沉积物氮的光谱信息非常重要。但是, 不恰当的波长提取, 反而会造成有用光谱信息损失, 不利于模型精度的提高。表 1 显示了 3 种特征波长提取方法对 LSSVM 模型精度的影响。结果表明, 连续投影算法 (SPA)、无信息变量消除法 (UVE) 和遗传算法 (GA) 三种特征波长提取方法都降低了模型的建模精度, 这可能是因为特征波长提取减少了建模变量。但是, 这三种特征波长提取方法对模型预测精度的影响并不一致。SPA 稍微降低模型的预测精度。而 UVE 增加了模型的预测精度, 检验集  $R^2$ , RMSEC 和 RPD 分别为 0.89, 0.12, 2.70, 表明 UVE 能更加充分、准确地挖掘光谱信息,

使模型更加稳定, 预测误差更小。从图 2 看, SPA 提取了 24 个特征波长, 主要集中在 830~970 nm 波段间, 而 UVE 提取了 72 个波长, 且分布范围广。此外, 遗传算法极大的降低了模型的预测精度, 检验集  $R^2$ , RMSEC 和 RPD 仅为 0.39, 0.24 和 1.28, 即基于 GA 建立的模型预测能力非常差, 分析认为 GA 在运行过程中容易陷入了局部极小化。从表 2 中可以看出 GA 提取的特征波长约为 900 nm, 且提取的特征波长仅有 17 个。

## 2.3 光谱数据变换与特征波长提取组合

从表 1 中可以看出, 在无特征波长提取的情况下或在 GA 特征波长提取的情况下, MSC 和 SVN 光谱变换都增加了模型的预测精度。但是, 在 SPA 或 UVE 特征波长提取的情况下, MSC 和 SVN 光谱变换对模型的预测精度影响不大。此外, 在无特征波长提取以及 SPA 特征波长提取的情况下, 一阶微分和二阶微分都极大的降低了模型的预测精度。而在 UVE 特征波长提取的情况下, 一阶微分和二阶微分对模型预测精度的影响幅度有所减弱。在 GA 特征波长提取的情况下, 一阶微分对模型预测精度影响不大, 二阶微分极大的降低了模型预测精度。

在无光谱变换、一阶微分和二阶微分光谱变换的情况下, UVE 特征波长提取都提高了模型的预测精度; 而在 MSC 和 SNV 光谱变换的情况下, UVE 特征波长提取对模型的预测精度影响不大。在无光谱变换、MSC、SNV 和一阶微分光谱变换的情况下, SPA 都降低了模型的预测精度; 但在二阶微分光谱变换的情况下, SPA 对模型预测精度影响不大。在无光谱变换、MSC、二阶微分光谱变换的情况下, GA 极大的降低了模型预测精度; 而在一阶微分、SNV 光谱变换的情况下, GA 对模型预测精度的影响幅度有所减轻。

尽管在无光谱变换或在 MSC, SNV 和一阶微分光谱变换的情况下, SPA 对模型精度的影响趋势基本一致。然而, 经过光谱变换后, SPA 提取的特征波长点发生变化 (图 2)。未经过光谱变换时, SPA 提取的特征波主要集中在 850~970 nm 之间; 经过 MSC 光谱变换后, SPA 提取的特征波长主要在 750~800 和 940~960 nm; 经过 SVN 光谱变换后, SPA 提取的特征波长主要在 650~800 和 940~970 nm。有意思的是, 无论是无光谱变换还是经过光谱变换, SPA 提取的特征波长区域与图 1 中高相关性区域并不对应, 这暗示着合适的建模波长并不只是高相关性波长。经过一阶微分和二阶微分光谱变换后, SPA 提取的特征波长在全波段都有分布, 与未经过光谱变换提取的特征波长区别明显。另外, 经 MSC 光谱变换后, UVE 提取的特征波长数量大幅度降低了; 经 SVN 光谱变换后, UVE 提取的特征波长较为聚集, 主要集中在 320 和 710 nm 附近, 但特征波长点的这些变化并没有影响模型精度。此外, 经一阶微分和二阶微分光谱变换后, UVE 提取的特征波长数量极大的增加了, 这可能是模型精度增加的一个重要原因。最后, 无论光谱变换与否, GA 提取的特征波长数量均较少, 且主要分布于 900 nm 左右。尤其是经过二阶微分光谱变换后, GA 提取的特征波长的数量仅为 1 个, 这可能是模型精度降低的一个关键原因。

相对于无光谱变换和无特征波长提取预处理, 仅 MSC

光谱变换, 或仅 SVN 光谱变换, 或仅 UVE 特征波长提取, 或 MSC 与 UVE 结合, 或 SVN 与 UVE 结合都能提高模型的预测精度。但是, 相对于单独运用 UVE 特征波长提取方

法或单独运用 MSC 及 SVN 光谱变换方法, 组合运用 UVE 特征波长提取方法和 MSC 或 SVN 光谱变换方法并没有提高模型的预测精度。

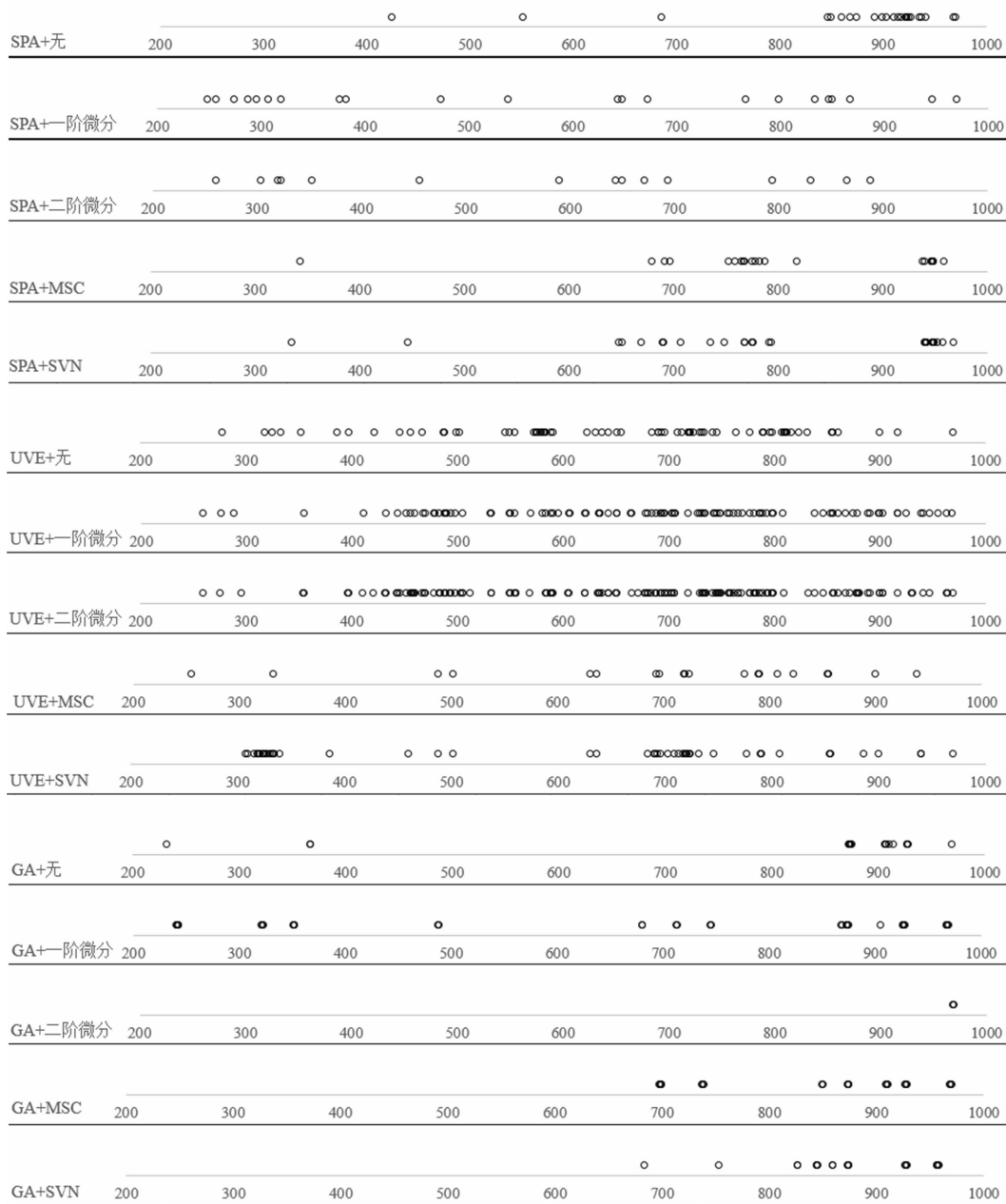


图 2 提取的潮间带沉积物氮特征波长点

Fig. 2 Extracted characteristic wavelength points of intertidal sediment nitrogen

### 3 结 论

复杂的沉积物成分以及测试环境都会对目标光谱产生影响, 而光谱变换可消除部分影响。周清等<sup>[8]</sup>认为导数变换可以去掉部分线性或接近线性的背景值干扰, 降低低频噪声对光谱的影响, 突显细微信息差异。杨红飞等<sup>[9]</sup>认为光谱数据

归一化比值处理可增强光谱与检测目标之间的相关性, 突出光谱敏感波段, 提高模型预测精度和稳定性。研究发现, MSC 或 SVN 可改变光谱反射率与氮含量的相关性, 释放被掩盖的有用光谱信息, 提升模型预测精度。而一阶微分和二阶微分则降低了光谱反射率与氮含量的相关性。对光谱数据进行微分变换时, 光谱反射率变化小, 则微分值接近于零, 光谱反射率变化较大, 则微分值较大。因此, 微分有利于限

制低频背景光谱的影响, 但会放大高频噪声的干扰, 这可能是本研究微分降低光谱反射率与氮含量相关性的重要原因。

本研究发现, 采用全部的波长点建模, 比选择部分波长点建模更能够提高模型的建模精度。但是, 并不是采用的波长点数量越多, 越有利于模型的预测精度。UVE 可以有效地挖掘出有用的光谱信息, 剔除无关信息, 提高模型预测精度。类似的, 梁梦醒等<sup>[10]</sup>也表示, 采用 UVE 优选敏感波长后, 模型的准确性更高。有意思的是, UVE 提取的特征波长与氮含量高相关性的波长并不是对应的, 这可能是因为, 合适的特征波长并非仅包括高相关性波长, 无吸收或吸收较弱的波长也可以作为特征波长, 用来矫正散射等影响。本研究还发现, 并非所有的特征波长提取方法都能够增加模型的预测精度。遗传算法降低了潮间带沉积物氮的模型预测精度, 推测可能是因为遗传算法导致变量过少而影响了模型的预测能力。

## References

- [ 1 ] REN Hong-yan, SHI Xue-zheng, ZHUANG Da-fang, et al(任红艳, 史学正, 庄大方, 等). Remote Sensing Technology and Application (遥感技术与应用), 2012, 27: 372.
- [ 2 ] NIE Jia-qin, WANG Dong-qi, CHEN Jie, et al(聂家琴, 王东启, 陈杰, 等). Environmental Science(环境科学), 2018, 39: 4219.
- [ 3 ] SHEN Zhang-quan, YE Ling-bin, SHAN Ying-jie(沈掌泉, 叶领宾, 单英杰). Acta Pedologica Sinica(土壤学报), 2014, 51: 1011.
- [ 4 ] ZHANG Rui, LI Zhao-fu, PAN Jian-jun(张锐, 李兆富, 潘剑君). Transactions of the Chinese Society of Agricultural Engineering(农业工程学报), 2017, 33: 175.
- [ 5 ] FAN Chang-sheng, LIU Ze-zhao, GUO Qiang, et al(范昌胜, 刘泽照, 郭强, 等). Journal of Taiyuan University of Science and Technology(太原科技大学学报), 2019, 40: 159.
- [ 6 ] SUN Biao, YANG Zhi-yan, ZHAO Sheng-nan, et al(孙标, 杨志岩, 赵胜男, 等). Soil and Fertilizer Sciences in China(中国土壤与肥料), 2019, (2): 194.
- [ 7 ] CUI Xia, SONG Qing-jie, ZHANG Yao-yao, et al(崔霞, 宋清洁, 张瑶瑶, 等). Acta Prataculturae Sinica(草业学报), 2017, 10: 20.
- [ 8 ] ZHOU Qing, ZHOU Bin, ZHANG Yang-zhu, et al(周清, 周斌, 张杨珠, 等). Acta Pedologica Sinica(土壤学报), 2004, 41(6): 905.
- [ 9 ] YANG Hong-fei, ZHENG Li-ming, HAO Zhong-yao, et al(杨红飞, 郑黎明, 郝中要, 等). Journal of Anhui Agricultural University(安徽农业大学学报), 2018, 45(1): 101.
- [ 10 ] LIANG Meng-xing, DAI Jun, HAN Bang-xing, et al(梁梦醒, 戴军, 韩邦兴, 等). Journal of West Anhui University(皖西学院学报), 2017, 33: 1.

由于光谱变换能够提取微弱的光谱变化, 突出有用的光谱信息, 更有利于特征波长的选择, 因此推测光谱变换和特征波长提取组合将提高 LSSVM 模型精度; 部分光谱变换与特征波长提取组合比单独运用特征波长提取更有利于模型预测精度, 比如在无光谱变换的情况下, GA 极大的降低了模型预测精度; 而在 SNV 光谱变换的情况下, GA 对模型预测精度的影响幅度有所减轻。研究结果中, UVE 是效果最优的特征波长提取方法, MSC 和 SVN 是效果最优的光谱变换方法, 但是, 相对于单独运用 UVE 特征波长提取方法或单独运用 MSC 以及 SVN 光谱变换方法, 组合运用 UVE 特征波长提取方法和 MSC 或 SVN 光谱变换方法并没有提高模型的预测精度, 这与预想并不一致。推测可能是组合预处理反而造成了光谱数据的过度处理, 并不能够增加信噪比和模型精度。

# The Effect of Spectral Pretreatment on the LSSVM Model of Nitrogen in Intertidal Sediments

LÜ Mei-rong<sup>1</sup>, REN Guo-xing<sup>1, 2</sup>, LI Xue-ying<sup>1</sup>, FAN Ping-ping<sup>1</sup>, LIU Jie<sup>1</sup>, SUN Zhong-liang<sup>1</sup>, HOU Guang-li<sup>1</sup>, LIU Yan<sup>1\*</sup>

1. Institute of Oceanographic Instrumentation, Qilu University of Technology (Shandong Academy of Sciences), Shandong Provincial Key Laboratory of Marine Monitoring Instrument Equipment Technology, National Engineering and Technological Research Center of Marine Monitoring Equipment, Qingdao 266100, China

2. School of Information Science and Engineering, Ocean University of China, Qingdao 266100, China

**Abstract** Spectral data transformation and feature wavelength extraction are two important spectral pretreatment methods, which play an important role in eliminating environmental interference. Previous literature mainly compared different spectral data transformation methods and there was less study on the spectral feature wavelength extraction methods and the combination of the two methods. In order to obtain suitable spectral pretreatment method and improve the accuracy of LSSVM model of sediment nitrogen in the intertidal zone, the effect of 4 spectral transformation methods combined with 3 characteristic wavelength extraction methods on the accuracy of LSSVM model of sediment nitrogen is studied for accurate prediction of sediment nitrogen in the intertidal zone. The results showed that the spectral transformation methods of multivariate scattering correction (MSC) or normal distribution (SVN) increased the correlation between spectra and nitrogen content and the highest correlation reached 0.69 and 0.71 respectively. MSC and SVN improved the prediction accuracy of LSSVM model, and the prediction  $R^2$  and RPD are 0.88, 0.87 and 2.78, 2.69, respectively. The feature wavelength extraction method of uninformative variable elimination (UVE) also improved the prediction accuracy of LSSVM model, model test  $R^2$  and RPD were 0.89 and 2.70, respectively. However, not all of the characteristic wavelengths extracted by UVE have a high correlation with nitrogen content. In addition, the combination of UVE and MSC or SVN also improved the prediction accuracy of the model, but it is not better than UVE alone or MSC or SVN alone. The results of this paper can provide a technical reference for nitrogen estimation and spectral data preprocessing of intertidal sediments.

**Keywords** Spectral data transformation; Spectral feature wavelength extraction; Spectral pretreatment; Intertidalite sediment

(Received Jul. 8, 2019; accepted Nov. 12, 2019)

\* Corresponding author