第43卷，第10期
2023年10月

光 谱 学 与 光 谱 分 析
Spectroscopy and Spectral Analysis

Vol. 43，No. 10，pp3075-3080
October，2023

主成分分析和连续投影融合的海洋沉积物粒度分类研究

贾宗潮[1]，王子鉴[1]，李雪莹[1,2*]，邱慧敏[1]，侯广利[1]，范萍萍[1*]

1. 齐鲁工业大学(山东省科学院)，山东省科学院海洋仪器仪表研究所，山东 青岛 266061
2. 中国石油大学(华东)计算机科学与技术学院，山东 青岛 266590

摘 要 海洋沉积物的粒度研究有助于了解人类活动对自然环境的影响。将主成分分析(PCA)和连续投影算法(SPA)融合能够综合利用两种光谱特征提取方法的优势，获得比单一特征提取方法更丰富的特征波长，实现无关特征和干扰信息的剔除，最大限度减少特征信息的丢失，有利于沉积物粒度的分析。以青岛市东大洋村潮间带表层32份沉积物为例，将海洋沉积物划分为0.3～0.2、0.2～0.1、0.1～0.075和<0.075 mm四个不同粒径的沉积物样品，分别测定不同粒径的32份沉积物的可见-近红外反射光谱，共计128条光谱。将128条光谱数据分别以2∶1，1∶1和1∶2的比例划分建模集和检验集进行分析；采用主成分分析和连续投影融合算法(FOPAS)提取不同粒径沉积物的特征光谱，利用支持向量机算法建立粒径分类模型。结果显示，对2∶1、1∶1、1∶2比例的数据集，融合算法检验集正确率分别为83.33%、82.81%、75.29%，仅在2∶1比例下正确率低于连续投影算法检验集的正确率90.47%，其余正确率相对于单一特征提取算法均有显著的提高，表明使用融合算法提取特征光谱建立的分类模型在训练集样本量少，粒径清晰的条件下，其分类模型相较于单独使用两个特征提取算法的模型更具有优势。采用基于主成分分析和连续投影融合算法的海洋沉积物粒度分类模型，能够提高海洋沉积物粒度分类结果的正确率，建立正确率更高的粒度分类模型，对快速粒度分类提供了解决方法。

关键词 海洋沉积物；粒度分类；主成分分析；连续投影算法；融合算法

中图分类号：O657.3 **文献标识码**：A **DOI**：10.3964/j.issn.1000-0593(2023)10-3075-06

引 言

海洋沉积物是指经过漫长而复杂的海洋沉积作用形成的海底沉积物，记录了古气候变化，海陆变迁，化学循环等过程的详细信息[1-2]。粒度分析作为沉积学与沉淀学的重要研究方法，在地球地质、海洋地质和海洋环境保护的研究中有着非常重要的应用价值。沉积物的粒度是一个非常重要的物理参数，反映了沉积物的运动过程和沉积物的结构特征，在沉积环境研究中有非常重要的意义[3-5]。海洋沉积物的在漫长的沉积过程中，海水中的有机质、碳酸盐等在不同的沉积物上的分布呈现的特征也是不均匀的，其中污染物等其他有害物质也会在沉积物上不断叠加从而引起沉积物自身粒度的变化，海洋沉积物粒度的研究有助于了解人类活动对自然环境的影响，从而为海洋环境保护提供理论指导[6]。因此近海沉积物粒径分析对海洋环境保护和生态修复有着重要的意义。

传统粒度测量技术一般包括直接测量、筛析法、双目显微镜、沉降法和电子显微镜法，其中常用的是沉降法和筛析法[7]。沉降法是基于颗粒在悬浮体系中以恒定速度沉降来测定分类的，测试时间较长，操作繁琐。筛析法是直径不同孔径的筛子将沉积物过筛，分出不同的粒级，目前筛析法对于小于0.045 mm的粒子不具备测量能力[5]。随着分析技术的发展，出现了多种非接触式的测量技术。如激光粒度分析法，该方法具有精度高，分析速度快等优点，但粒子取样要求较高，且对大体积质子测量误差较大[8]。图像法和超声谱分析法也是目前较为常用的粒度分析方法，图像法能够表达每个颗粒的大小及粒形信息，但存在着数据处理复杂等问题[9]，超声谱分析法可以取得比图像法偏差更小的测量结果，但对缓冲块介质的要求较严格[10]。

收稿日期：2022-04-30，修订日期：2022-09-08

基金项目：国家自然科学基金项目(32171578)，山东省自然科学基金项目(ZR2021QF028，ZR2021MD093，ZR2021MD103)，齐鲁工业大学科教产融合试点工程基础研究类项目(2022PYI008)资助

作者简介：贾宗潮，1995年生，齐鲁工业大学(山东省科学院)海洋仪器仪表研究所硕士研究生 e-mail：zongchao_j@126.com

*通讯作者 e-mail：fanpp_sdioi@126.com；412973984@qq.com

光谱分析具有检测速度快、灵敏度高、无损伤检测等优点，在化学成分分析、质量检测等领域应用广泛[11]。可见-近红外吸收/反射光谱中富含样品 O—H、N—H、C—H 等有机官能团的种类和数量信息，在土壤和海洋沉积物 C，N 等有机质含量快速测定方面取得了非常多的成果。不同粒径海洋沉积物有机官能团的种类和数量信息有一定的不同，故其吸收/反射光谱信息也有一定的不同，利用可见-近红外光谱特性对海洋沉积物粒径进行分类具有很好的研究前景。

以青岛市东大洋村潮间带表层沉积物为例，将海洋沉积物划分为 0.3～0.2、0.2～0.1、0.1～0.075 和＜0.075 mm 四个不同粒径的沉积物样品，分别测定不同粒径下沉积物的可见-近红外反射光谱。采用主成分分析和连续投影算法融合的特征光谱提取方法（fusion of principal component analysis and successive projection algorithm，FOPAS）提取不同粒径沉积物的特征光谱。该方法能够获得比单一特征提取方法更丰富的特征波长，综合两种特征提取方法的优势，既实现无关特征和干扰信息的剔除，又能够最大限度减少特征信息的丢失，提高模型的正确率和稳定性。将主成分分析和连续投影融合算法分类结果和单一特征提取方法的分类结果进行比对分析，寻找最优海洋沉积物粒度特征信息，从而建立正确率更高的分类模型，实现对沉积物粒度的快速分类。

1 实验部分

1.1 试验材料

采样地点位于青岛市东大洋村潮间带，于 2019 年 8 月借助竹筏采样，共采集 32 份沉积物样品。把采集到的沉积物样品放在实验室风干，破碎，全部通过 0.3 mm 筛，低温烘干后研磨，把研磨过后的样品进行筛分，分别过 0.2、0.1 和 0.075 mm 三层筛子，对应得到 0.3～0.2、0.2～0.1、0.1～0.075 和＜0.075 mm 不同粒径的沉积物样品，共计 32 份样品，用于后续可见-近红外光谱的检测。

1.2 光谱数据采集

使用海洋光学 QE65000 光谱仪搭配 DH-2000-BAL 型光源采集沉积物样品的反射光谱，光谱采样间隔为 1 nm，积分时间 600 ms，谱区范围 200～1 100 nm，通过 Y 型光纤（QR400-7-UV-VIS）连接光谱仪和光源，Y 型光纤探头由支架固定，将样品放置于自制的样品盒中，探测样品的反射光谱。每个样品测定 5 次光谱反射率，取 5 次光谱反射率的平均值作为该样品的反射光谱。

不同粒径 32 份沉积物的反射光谱共 128 条。由于光谱前段和后段受噪声影响，因此去掉前段和后段光谱，取 226～975 nm 波段光谱，如图 1 所示。

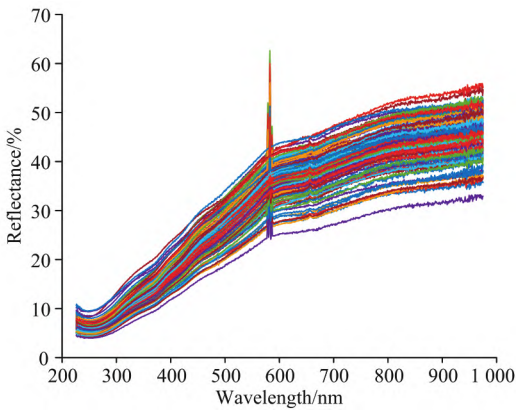

图 1 不同粒径沉积物反射光谱图

Fig. 1 Reflectance spectra of sediments with different particle sizes

对不同粒径的沉积物样品测定碳氮含量（北京植物所，碳氮分析仪）。TN、TC 的含量的实测值统计列表如表 1 所示。不同粒径下的沉积物 CN 含量值有所差异。

表 1 C、N 含量平均值统计（Mean±SD）

Table 1 Mean values of C，N contents（Mean±SD）

粒径/mm	TN/(g·kg^{-1})	TC/(g·kg^{-1})
0.3～0.2	0.265±0.015	2.920±0.135
0.2～0.1	0.366±0.014	2.652±0.108
0.1～0.075	0.350 ±0.010	2.067±0.067
＜0.075	0.360±0.010	2.383±0.062

1.3 主成分分析和连续投影融合算法

主成分分析和连续投影融合算法是利用两种特征提取方法，对原始光谱数据进行降维和特征提取，可以用更少的变量去代替更多的原始变量，降低模型的复杂度，使模型更加高效和稳定。

主成分分析（principal component analysis，PCA），是一种经典的特征提取法，旨在降低数据集复杂性的同时能够最大限度的减少信息的丢失。它通过正交变换的方式可以将一组变量的观察值转换成一组不存在相关性的变量，转换完成后获得的变量被称为主成分[12]。在光谱分析中，PCA 通过正交变换将光谱数据中具有相关性的数据变量转换成不相关的光谱变量即是主成分，从而达到了降低光谱数据的复杂性的同时也能够获取不同粒径沉积物的特征光谱，最大限度的减少了光谱信息的丢失。

连续投影算法（successive projections algorithm，SPA）是一种使矢量空间共线最小化的前向变量选择算法，能够很好的消除波长数值间共线性的影响，优选出能够反映样本关键信息的有效特征波段从而降低模型的复杂度，提高模型的稳定性和准确性。它是一种前向循环筛选方法，即从一个波长作为起点，每次循环合并一个新的波长，直到达到指定数目的波长为止[13]。通过连续投影算法对光谱数据波长进行筛选，得到共线性最小的波长，即得到了最能反映关键特征的波长，降低了数据的复杂性，减少了无关信息的干扰，提高了分类模型的准确率和稳定性。

主成分分析和连续投影融合算法（fusion of principal component analysis and successive projection algorithm，FOPAS）是将两种算法得到的特征光谱融合作为建模光谱值，具体如式(1)所示

$$F = F_1 + F_2 \tag{1}$$

式(1)中：F 为融合后得到的特征光谱，F_1 为主成分分析法

得到的特征光谱，F_2 为连续投影算法得到的特征波长。

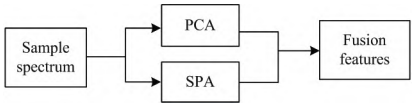

图 2　主成分分析和连续投影融合算法流程图

Fig. 2　Flow chart of principal component analysis and continuous projection fusion algorithm

主成分分析(PCA)和连续投影算法(SPA)作为多变量分析中常用的降维和特征提取方法，在消除光谱变量间的多重共线性，降低模型复杂度等研究方面取得了很好的效果。主成分分析(PCA)不仅能够浓缩光谱数据，同时还具有不相关的性质，最大限度的避免信息重叠带来的虚假性。连续投影算法(SPA)通过向量投影能够选出冗余度低，共线性小又能反映光谱关键特征的有效波段。融合算法能够有效地融合两种算法的优点，将两种特征提取方法获取的特征光谱融合，获得比单一特征提取方法更丰富的特征波长，从而既达到无关特征和干扰信息的剔除，又能够最大限度的减少特征信息的丢失，既提高了建模速度又降低了模型的复杂度，进而提高了模型的正确率和稳定性。

1.4　分类算法

支持向量机(support vector machine，SVM)作为最常用的分类方法，它的核心思想是通过核函数将向量映射到更高维的空间中，构造一个最优分类超平面。寻找两个距离最大且平行于分类超平面的平行超平面。通过构造一个超平面 $f(x)=\omega x+b=0$，其中 ω 为分类平面的法向量，b 为分类平面的偏移量，则构造的分类函数为 $f(x)=\omega x+b$。平行超平面之间的距离越大，分类器的分类准确率越高。SVM 算法遵循结构风险最小化原则，能有效的解决其他机器学习算法中小样本、非线性的情况下过拟合以及陷入局部最优解等问题。

在本研究中，SVM 算法选取高斯径向基函数作为核函数建立定性分析判别模型，从而将数据映射到高位空间，有效的解决了原始空间中线性不可分的问题。支持向量机的建模与预测在 Matlab R2016a 环境下完成。

1.5　模型评价标准

模型评价标准采用正确率 Accuracy、均方根误差(root mean square error，RMSE)、决定系数 R^2。

$$\text{Accuracy}=\frac{TP}{N}\times 100\% \tag{2}$$

$$\text{RMSE}=\sqrt{\frac{\sum_i (y_i-\hat{y}_i)^2}{N}} \tag{3}$$

$$R^2=1-\frac{\sum_i (y_i-\hat{y}_i)^2}{\sum_i (y_i-\bar{y})^2} \tag{4}$$

式(2)—式(4)中，TP 为正确个数，N 为总个数，y_i 为真实值，$\hat{y}_i$ 为预测值，$\bar{y}$ 为所有样本真实值的均值。通常情况下，均方根误差越小，正确率和决定系数越大，分类器准确率越高。

2　结果与讨论

2.1　全光谱模型海洋沉积物粒度分类

首先对沉积物全光谱数据先进行归一化处理，然后采用 SVM 方法建立模型，建模集和检验集的比例分别以 2∶1(建模集样品 86 个，检验集样品 42 个)，1∶1(建模集样品 64 个，检验集样品 64 个)和 1∶2(建模集样品 42 个，检验集样品 86 个)进行分析，分类顺序按照 0.3～0.2 mm 32 个、0.2～0.1 mm 32 个、0.1～0.075 mm 32 个、<0.075 mm 32 个等距离选取，全光谱模型分类结果如表 2 所示。

表 2　全光谱模型分类结果

Table 2　Full-spectrum model classification results

建模集∶检验集	建模集			检验集		
	正确率	rmse	R^2	正确率	rmse	R^2
2∶1	93.02%(80/86)	0.32	0.92	73.81%(31/42)	0.58	0.74
1∶1	95.31%(61/64)	0.22	0.96	70.31%(45/64)	0.70	0.65
1∶2	93.02%(40/42)	0.46	0.84	55.29%(47/86)	0.67	0.66

由表 2，在全光谱模型的下，建模集和检验集的比例分别以 2∶1、1∶1、1∶2 时，建模集的分类正确率分别为 93.02%、95.31%、93.02%，建模集的最小均方根误差为 0.22，决定系数最高的为 0.96，检验集分类正确率分别为 73.81%、70.31%、55.29%，当比例为 1∶2 时检验集的分类正确率最低，同样在比例为 1∶1 时均方根误差最大 0.70，决定系数最小为 0.65。

采用全光谱模型分类，在建模集和检验集为 1∶1 比例下，分别统计四个粒径的分类的正确率，见表 3。

表 3　全光谱模型各粒径建模集和检验集以 1∶1 比例的分类结果

Table 3　Classification results of modeling set and test set in a ratio of 1∶1 using entire-spectrum data for each particle size

序号	粒径/mm	建模集正确率	检验集正确率
1	0.3～0.2	93.75%(15/16)	62.50%(10/16)
2	0.2～0.1	100.0%(16/16)	68.75%(11/16)
3	0.1～0.075	93.75%(15/16)	68.75%(11/16)
4	<0.075	93.75%(15/16)	81.25%(13/16)

由表 3，建模集和检验集为 1∶1 比例下，0.1～0.075 粒径建模集的分类正确率达到了 100%，其余 3 个粒径的建模集的分类正确率都为 93.75%，错误样本都为 1 个。检验集分类正确率最高的是 <0.075 mm 粒径，分类正确率为 81.25%，0.2～0.1 mm 粒径和 0.1～0.075 mm 粒径分类正确率相同为 68.75%，分类正确率最低的是 0.3～0.2 mm 粒径，分类正确率为 62.50%，错误样本为 6 个。

2.2 基于主成分分析和连续投影算法融合的海洋沉积物粒度分类

在对沉积物光谱数据进行归一化处理后，采用主成分分析法对 128 个不同粒径的沉积物样品的光谱数据进行降维，选取贡献率大于 99%的前 4 个主成分。在对归一化后的光谱数据使用连续投影算法选出 11 个波长，分别是 226、228、229、286、581、583、584、685、942、944 和 950 nm。将主成分分析法获得的贡献率大于 99%，即前 4 个主成分与连续投影算法提取到的 11 个波长点融合使用，使用 SVM 进行建模，同样，建模集和检验集的比例分别以 2∶1，1∶1 和1∶2 进行分析，分类顺序与前文所述方法相同，将分类结果与单一特征提取算法对比。结果如表 4 所示。

表 4 两种算法单独使用和融合算法分类结果

Table 4 Classification results of the two algorithms alone and fused algorithms

建模集∶检验集	主成分分析						连续投影算法						融合算法					
	建模集			检验集			建模集			检验集			建模集			检验集		
	正确率	rmse	R^2	正确率	rmse	R^2	正确率	rmse	R^2	正确率	rmse	R^2	正确率	rmse	R^2	正确率	rmse	R^2
2∶1	82.56% (71/86)	0.53	0.79	42.86% (25/42)	0.76	0.55	98.84% (85/86)	0.11	0.99	90.47% (38/42)	0.31	0.93	98.84% (85/86)	0.11	0.99	83.33% (35/42)	0.60	0.75
1∶1	78.13% (50/64)	0.52	0.79	46.88% (30/64)	0.94	0.41	87.50% (56/64)	0.47	0.83	73.44% (47/64)	0.56	0.79	95.31% (61/64)	0.31	0.93	82.81% (53/64)	0.63	0.72
1∶2	76.71% (34/42)	0.61	0.73	44.71% (38/86)	0.83	0.49	90.69% (39/42)	0.40	0.88	72.94% (62/86)	0.61	0.72	93.02% (40/42)	0.57	0.78	75.29% (64/86)	0.69	0.68

由表 4 可得，建模集和检验集的比例分别以 2∶1、1∶1、1∶2 时，主成分分析方法的建模集的分类正确率分别为 82.56%、78.13%、76.71%，检验集分类正确率分别为 42.86%、46.88%、44.71%，3 种比例分类结果表现均衡，但检验集的均方根误差较建模集增大明显，决定系数减小明显，均未超过 0.6。连续投影算法的建模集的分类正确率分别为 98.84%、87.50%、90.69%，检验集分类正确率分别为 90.47%、73.44%、72.94%，当建模集与检验集的比例为 2∶1时，检验集分类结果最好，检验集的均方根误差比建模集略有增加，决定系数相差明显。融合算法建模集正确率均达到了 90%以上，高于其他方法，检验集分类正确率最低的是 1∶2 比例，为 75.29%，其他两个比例的检验集正确率分别为 83.33%、82.81%，平均分类正确率也均高于其他方法，但在检验集上均方根误差最低为 0.60，决定系数最高为 0.75，较建模集减小明显。

表 5 两种算法单独使用和融合算法的各粒径建模集和检验集以 1∶1 比例的分类结果

Table 5 Classification results of modeling set and test set in a 1∶1 ratio using each of PCA, SPA and PCA fused with SPA the two algorithms alone and the fusion algorithm

沉积物粒径/nm	主成分分析		连续投影算法		两种方法融合	
	建模集	检验集	建模集	检验集	建模集	检验集
0.3～0.2	81.25% (13/16)	43.75% (7/16)	75.00% (12/16)	43.75% (7/16)	93.75% (15/16)	93.75% (15/16)
0.2～0.1	81.25% (13/16)	50.00% (8/16)	100.0% (16/16)	81.25% (13/16)	93.75% (15/16)	81.25% (13/16)
0.1～0.075	75.00% (12/16)	25.00% (4/16)	87.50% (14/16)	75.00% (12/16)	93.75% (15/16)	75.00% (12/16)
＜0.075	75.00% (12/16)	68.75% (11/16)	87.50% (14/16)	93.75% (15/16)	100.0% (16/16)	81.25% (13/16)

由表 5，在建模集和检验集为 1∶1 比例下，主成分分析 0.3～0.2 与 0.2～0.1 mm 粒径建模集的分类正确率都为 81.25%，0.1～0.075 与＜0.075 mm 粒径建模集分类正确率都是 75%。检验集分类正确率最高的是＜0.075 mm 粒径，分类正确率为 68.75%，分类正确率最低的是 0.1～0.075 mm 粒径，分类正确率为 25%。连续投影算法分类模型，除了 0.3～0.2 mm 粒径检验集正确率较低，仅为 43.75%，其他粒径的检验集分类正确率均大于等于 75%。融合算法的检验集分类正确率最低的是 0.1～0.075 mm 粒径，为 75%，除 0.1～0.075 mm 粒径外其他 3 个粒径的检验集分类正确率均高于 80%，分别为 93.75%、81.25%、81.25%。

全光谱模型分类，建模集分类正确率与检验集分类正确率相差明显，除＜0.075 mm 粒径，总体和其余粒径检验集分类效果均不理想。基于主成分分析的沉积物分类模型三种比例建模集与检验集分类结果都不好，各粒径的建模集和检验集分类正确率也低于全光谱模型分类。基于连续投影算法方法的分类，不仅在总体分类结果优于全光谱和基于主成分分析的，除 0.3～0.2 mm 粒径外其他 3 个粒径的分类正确率均有一定的提升。两种特征提取算法融合的分类方法，除低于在 2∶1 比例下连续投影算法检验集正确率，其余正确率

相对于单一特征提取算法均有显著的提高，另外由于在粒径＜0.075 mm时由于粒径过小，光谱特征较大粒径样品光谱变得不明显，所以建模集有过拟合的现象，造成了检验集结果变差。在均方根误差和决定系数方面，由于四个粒径的观测值0.3～0.2 mm用1，0.2～0.1 mm用2，0.1～0.075与＜0.075 mm分别用3和4代替，融合算法在0.1～0.075与＜0.075 mm粒径上的分类正确率的较低，这两种粒径的代替值为3和4，其模型预测值与真实值的误差相对于0.3～0.2与0.2～0.1 mm粒径的代替值1和2所占比较大，从而导致总体的均方根误差变大和决定系数变小。从以上结果分析表明使用融合算法的提取的特征光谱建立的分类模型在训练集样本量少，粒径更大的条件下，其分类模型相较于使用两个单独的特征提取算法的模型更具有优势，另外融合算法在大粒径样品的分类正确率上也有很大的提升。

在寻找最优沉积物粒径特征光谱中，通过使用两种不同的特征提取算法对光谱数据进行降维提取特征光谱，同时消除相关性高的波长数据对模型的干扰。采用两种特征提取算法融合的模型分类结果要好于单独使用两种特征提取方法和全光谱模型的结果。因此采用基于主成分分析和连续投影算法联用的海洋沉积物粒度分类模型，能够提高海洋沉积物粒度分类的正确率。后续将尝试其他预处理和特征提取算法处理小粒径样品光谱，以获得更好的分类结果。

3　结　论

以青岛市东大洋村潮间带表层沉积物为例，将海洋沉积物划分为0.3～0.2、0.2～0.1、0.1～0.075和＜0.075 mm四个不同粒径样品的沉积物样品，并分别测定不同粒径样品的可见-近红外反射光谱。由分类正确率可知采用两种特征提取算法融合的方法的分类模型，在训练集样本量少、粒径清晰的条件下，优于单独使用一种特征提取方法的分类模型和全光谱分类模型，其在总体分类正确率和各个粒径分类正确率上都有显著的提高。因此采用基于主成分分析和连续投影融合算法的海洋沉积物粒度分类模型，能够提高海洋沉积物粒度分类结果的正确率，建立正确率更高的粒度分类模型，对快速粒度分类提供了解决方法。

References

[1] Ling S D, Sinclair M, Levi C J, et al. Marine Pollution Bulletin, 2017, 121(1): 104.

[2] Giancarlo A Restreppo, Warren T Wood, Jordan H Graw, et al. Marine Geology, 2021, 440: 106577.

[3] WANG Jun-bo, ZHU Li-ping(王君波, 朱立平). Journal of Lake Sciences(湖泊科学), 2005, 1: 17.

[4] Li Xueying, Fan Pingping. IEEE Access, 2020, 8: 157151.

[5] TIAN Yu-xin(田宇昕). China CIO News(信息系统工程), 2019, 12: 147.

[6] Ishfaq Ahmad Mir, Maria Brenda Luzia Mascarenhas, Neloy Khare. Journal of Asian Earth Sciences, 2022, 227(2): 105102.

[7] Thorne P D, Hurther David. Continental Shelf Research, 2014, 73: 97.

[8] LIAN Jie(廉　洁). Information & Communications(信息通信), 2019, 5: 160.

[9] QU Pei-yu, JIANG Yu, SU Ming-xu, et al(曲佩玓, 蒋　瑜, 苏明旭, 等). Acta Metrologica Sinica(计量学报), 2021, 4: 469.

[10] JIA Nan, GU Jian-fei, SU Ming-xu(贾　楠, 顾建飞, 苏明旭). Acta Metrologica Sinica(计量学报), 2019, 3: 466.

[11] LI Xue-ying, LI Zong-min, HOU Guang-li, et al(李雪莹, 李宗民, 侯广利, 等). Spectroscopy and Spectral Analysis(光谱学与光谱分析), 2021, 41(9): 2898.

[12] ZHAO Shan, LI Yong-si(赵　珊, 李永思). Journal of Beijing University of Posts and Telecommunications(北京邮电大学学报), 2019, 2: 36.

[13] NIU Fang-peng, LI Xin-guo, MAMATTURSUN·Eziz, et al(牛芳鹏, 李新国, 麦麦提吐尔逊·艾则孜, 等). Journal of Zhejiang University: Agriculture and Life Sciences(浙江大学学报：农业与生命科学版), 2021, 5: 673.

Marine Sediment Particle Size Classification Based on the Fusion of Principal Component Analysis and Continuous Projection Algorithm

JIA Zong-chao[1], WANG Zi-jian[1], LI Xue-ying[1, 2*], QIU Hui-min[1], HOU Guang-li[1], FAN Ping-ping[1*]

1. Institute of Oceanographic Instrumentation, Qilu University of Technology (Shandong Academy of Sciences), Qingdao 266061, China
2. College of Computer Science and Technology, China University of Petroleum (East China), Qingdao 266590, China

Abstract The study on the granularity of marine sediments is helpful in understanding the impact of human activities on the natural marine environment. The fusion of principal component analysis and successive projection algorithm combines the advantages of both spectral feature extraction methods. It can obtain richer feature wavelengths than a single feature extraction method, achieve rejection of irrelevant features and interference information, minimize the loss of feature information, and facilitate the analysis of sediment grain size. In this paper, 32 sediments from the surface layer of the intertidal zone of East Dayang Village in Qingdao City were divided into four sediment samples with different grain sizes of 0.3～0.2, 0.2～0.1, 0.1～0.075 and <0.075 mm. The visible-NIR reflectance spectra of 32 sediments with different grain sizes were measured separately, with 128 spectra samples. The 128 spectral samples were divided into modeling set and test set in the 2∶1, 1∶1 and 1∶2 ratio for analysis. An algorithm fused with principal component analysis and successive projection algorithm was used to extract the characteristic spectra of different grain-size sediments, and the support vector machine algorithm was used to build a grain-size classification model. The results show that the fusion algorithm test set correct rates of 83.33%, 82.81%, and 75.29% at 2∶1, 1∶1 and 1∶2, respectively. All the correct rates were significantly improved relative to the single feature extraction algorithm, except for the lower than 90.47% correct rate for the test set of the continuous projection algorithm at the 2∶1 ratio, indicating that the classification models were built by using the extracted feature spectra of the fusion algorithm. The classification model using the fused algorithm with the extracted feature spectra has an advantage over the model using two separate feature extraction algorithms under the condition of a small training set and clear particle size. Adopting a classification model a marine sediment particle size based on principal component analysis and continuous projection fusion algorithm can improve the correct classification rate results of marine sediment particle size, establish a particle size classification model with a higher correct rate, and provide a solution for fast particle size classification.

Keywords Marine sediments; Particle size classification; Principal component analysis; Successive projection algorithm; Fusion algorithm

(Received Apr. 30, 2022; accepted Sep. 8, 2022)

* Corresponding authors