

小波变换的潮滩沉积物含水量预测

李雪莹^{1,2}, 李宗民^{3*}, 陈光源⁴, 邱慧敏², 侯广利², 范萍萍^{2*}

1. 中国石油大学(华东)地球科学与技术学院, 山东 青岛 266580
2. 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
3. 中国石油大学(华东)计算机科学与技术学院, 山东 青岛 266580
4. 山东科技大学海洋科学与工程学院, 山东 青岛 266590

摘要 潮滩沉积物水分的分布在空间和时间上会有很大的变化, 含水量的变化会导致沉积物中生源要素含量的变化。因此, 实时、准确、快速的监测潮滩沉积物含水量, 对了解潮滩的各种特性, 掌握潮滩生源要素信息, 潮滩资源的开发有着重要意义。采集青岛市东大洋村潮间带的沉积物 115 份, 分别测定新鲜样品、风干 4 周、风干 8 周样品的可见近红外光谱和含水量。以 db10 小波基和 sym6 小波基对原始光谱进行小波变换, 采用偏最小二乘回归建立潮滩沉积物含水量模型。通过 10 阶小波变换获取原始光谱的低频信息 A_n 和 高频信息 D_n ($n=1, 2, \dots, 10$), 通过原始光谱 S 分别与高频信息 D_n 做差值, 得到 $S-D_n$, 对 A_n , D_n 和 $S-D_n$ 建立潮滩沉积物含水量模型, 并对模型结果进行分析。原始光谱建立模型的 R_p^2 为 0.841, RMSEP 为 2.767, RPD 值为 2.481。通过对 db10 小波基变换后的低频和高频信息分析, 无用信息主要集中在 D_3 和 D_4 , 去除 D_3 和 D_4 建立的含水量模型, 相比于原始光谱模型精度有明显提高, R_p^2 为 0.878, RMSEP 为 2.501, RPD 值为 2.749; 通过 sym6 小波基变换后进行分析, 无用信息主要集中在 D_5 和 D_9 , 去除 D_5 和 D_9 建立含水量模型与原始光谱模型相比, 精度也有一定提高, R_p^2 为 0.87, RMSEP 为 2.475, RPD 值为 2.768。因此通过小波变换对原始光谱划分低频信息和高频信息进行分析, 能够有效找到潮滩沉积物含水量的干扰信息, 实现特征信息提取, 从而建立准确度更高的潮滩沉积物含水量模型, 为潮滩沉积物含水量实时、动态监测提供理论基础。

关键词 潮滩沉积物; 小波变换; 含水量; 可见-近红外光谱

中图分类号: O657.3 **文献标识码:** A **DOI:** 10.3964/j.issn.1000-0593(2022)04-1156-06

引言

潮滩是海洋与陆地之间的交互地带, 是内外陆以及海动力共同作用的产物^[1]。潮滩沉积物包含自然灾害、生源要素预测的关键信息^[2]。地表土壤含水量是地表与大气进行水和能量交换的必要途径^[3], 受潮汐和波浪作用蒸发和降水的复杂水力学控制。潮滩沉积物水分的分布在空间和时间上会有很大的变化^[4], 含水量的变化会导致沉积物中生源要素含量的变化。因此, 实时、准确、快速的监测潮滩沉积物含水量, 对了解潮滩的各种特性, 掌握潮滩生源要素信息, 潮滩资源的开发有着重要意义。

传统的土壤含水量获取采用烘干法、时域反射仪法

等^[5], 沉积物含水量也可采用该方法。传统方法虽然精准, 但是成本高, 且费时费力, 同时也无法快速、实时获取含水量。光谱技术具有快速、无损等优点, 已广泛应用于多个领域。水分对沉积物表面短波辐射的反射有很大的影响, 包括光谱的可见-近红外和短波红外区域^[6]。因此光谱能够动态、实时、快速的监测潮滩沉积物含水量, 有望成为潮滩沉积物含水量监测的主流技术。目前基于光谱的土壤和沉积物含水量的测定已有一定的进展。Ma 等^[7]基于近红外光谱对农田土壤含水量进行检测和分析, 找到黑土水分的特征吸收谷, 利用吸收谷的光谱特征参数预测黑土的含水量。Shin 等^[8]分析了韩国海岸沉积物的矿物学特征及其与含水量相关的光谱特征, 提出了基于传统 SAM 模块的星载热发射和反射辐射计短波红外遥感方法进行含水量制图的方法。Shin 等^[9]于两

收稿日期: 2021-03-16, 修订日期: 2021-05-06

基金项目: 国家自然科学基金项目(U2006209, 32171578)和山东省自然科学基金项目(ZR2021QF028, ZR2021MD093, ZR2021MD103)资助

作者简介: 李雪莹, 女, 1990 年生, 中国石油大学(华东)地球科学与技术学院博士研究生 e-mail: 412973984@qq.com

*通讯作者 e-mail: lizongmin@upc.edu.cn; fanpp_sdioi@126.com

年后研究短波红外反射率与海滩含水量之间的关系，发现最佳光谱带，并根据选定的波带建立的回归模型的 R^2 大于 0.74，RMSE 小于 0.05。

本工作基于可见-近红外光谱，以青岛市东大洋村潮间带的沉积物为例，采用小波变换找到能够表征潮滩沉积物含水量的特征信息，旨在消除具有复杂成分组成的潮滩沉积物的干扰和噪声信息，建立准确度更高的潮滩沉积物含水量模型，实现含水量反演，为潮滩沉积物含水量实时、动态监测提供科技支撑。

1 实验部分

1.1 试验材料

采集青岛市东大洋村潮间带的沉积物 155 份。将采集的一部分样品立即测定可见-近红外反射光谱，同时测定其含水量。含水量测定采用烘干法，110 °C 烘干，烘干前后的重量差值与新鲜样品百分比即为水分含量。另一部分放在实验室通风处风干，分别风干 4 周和 8 周后，采集可见-近红外反射光谱，测定含水量，不同条件下潮滩沉积物样品的含水量见表 1。将新鲜样品、风干 4 周、风干 8 周共计 465 个样品进行后续分析。

表 1 潮滩沉积物水分含量 (%) 统计表
Table 1 The statistical results of moisture content (%) of tidal flat sediments

样品种类	样品个数	最大值	最小值	平均值	标准差
新鲜样品	155	63.57	43.42	54.77	2.87
风干 4 周	155	55.41	31.76	47.82	3.61
风干 8 周	155	47.32	18.81	38.23	5.09
所有样品	465	63.57	18.81	46.94	7.86

1.2 光谱数据采集

采用海洋光学 QE65000 光谱仪搭配 DH-2000-BAL 型光源采集沉积物样品的反射光谱，采样间隔 1 nm，积分时间 600 ms，谱区范围 200~1 000 nm。通过 QR400-7-UV-VIS 型号的 Y 型光纤采集样品的反射光谱，每个样品测定 5 次，取平均光谱作为该样品的反射光谱。删去样品的前段和后段受噪声影响的光谱，保留 220~970 nm 光谱。

1.3 小波变换

在光谱分析中，小波变换可用于数据压缩、平滑、滤除噪声等，常用的小波基有 Haar 小波、dbN 小波、Symlets 小波。小波变换是通过选择合适的小波基、小波分解尺度，对信号进行小波变换，得到低频系数和低频系数，然后进行小波逆变换，得到重构的光谱数据^[10]。离散小波变换是对离散时间信号进行时频分析的一种数学方法，其主要思想是多分辨率分析过程^[11]。假设为小波母函数，将其进行平移和伸缩得到如式(1)函数

$$\Psi_{a,\tau}(t) = a^{-\frac{1}{2}} \Psi\left(\frac{t-\tau}{a}\right), a > 0, \tau \in R \quad (1)$$

式(1)中， a 为尺度参数， τ 为平移参数。在离散小波变换中，

$a = 2^j, \tau = 2^j k, j, k \in Z$ ，式(1)转换为如式(2)

$$\Psi_{j,k}(t) = 2^{-\frac{j}{2}} \Psi(2^{-j}t - k), j, k \in Z \quad (2)$$

经 M 阶离散小波变换后，有限长度离散信号 $S(t)$ 可表达为

$$S(t) = A_M + \sum_{j=1}^M D_j(t) \quad (3)$$

式(3)中， $A_M = \sum_k A_{M,k} \varphi_{j,k}$ ， $D_j(t) = \sum_k D_{j,k} \Psi_{j,k}(t)$ ， $A_{M,k}$ 和 $D_{j,k}$ 分别为原始信号的低频和高频系数。

小波基分别选用 db10 小波和 Sym6 小波。通过小波变换获取原始光谱信号的高频和低频信息，分析每个高频和低频以及 $S - D_j$ 信息，找到并去除干扰和噪声信息，实现潮滩沉积物含水量光谱特征信息的提取。

1.4 建模方法及评价标准

建模方法采用偏最小二乘回归 (partial least squares regression, PLSR)，在已知可见-近红外光谱数据和含水量数据的基础上，通过 PLSR 建立潮滩沉积物含水量模型，实现对未知沉积物样品含水量的预测。

样本分类采用 SPXY 算法以 1 : 1 比例划分建模集和预测集，它是在 Kennard-Stone 算法基础上发展的，计算如式(4)~式(6)

$$d_x(p, q) = \sqrt{\sum_{j=1}^N [x_p(j) - x_q(j)]^2} \quad p, q \in [1, N] \quad (4)$$

$$d_y(p, q) = \sqrt{\sum_{j=1}^N (y_p - y_q)^2} = |y_p - y_q| \quad p, q \in [1, N] \quad (5)$$

$$d_{xy}(p, q) = \frac{d_x(p, q)}{\max_{p, q \in [1, N]} d_x(p, q)} + \frac{d_y(p, q)}{\max_{p, q \in [1, N]} d_y(p, q)} \quad (6)$$

其中 x 为光谱数据， y 为含水量数据。

评价标准采用建模集和预测集绝对系数 R_c^2 和 R_p^2 ，校正均方根误差 RMSEC，预测均方根误差 RMSEP 和相对分析误差 RPD 值，当 R_c^2 ， R_p^2 和 RPD 值越大，RMSEC 和 RMSEP 越小，则建立的潮滩沉积物含水量模型精度越高。

2 结果与讨论

2.1 基于原始光谱潮滩沉积物含水量预测结果

潮滩沉积物原始光谱以 SPXY 算法划分建模集(233 个样本)和预测集(232 个样本)，通过 PLSR 算法建立含水量模型。图 1 为实测和预测含水量拟合图。潮滩沉积物含水量模型的 R_c^2 为 0.878， R_p^2 为 0.841，RMSEC 和 RMSEP 分别为 3.043 和 2.767，RPD 值为 2.481，模型精度较好。

2.2 基于小波变换 (db10 小波基) 潮滩沉积物含水量预测结果

将潮滩沉积物原始光谱以 db10 为小波基进行离散小波变换，分别得到 10 阶低频信息和高频信息，图 2 分别为所有样本平均光谱的低频 A_n 和高频信息 $D_n (n=1, 2, \dots, 10)$ 。

由图 2，低频和高频信息均是随 n 增加而趋于平滑。分别以原始光谱的低频信息 A_n 和高频信息 $D_n (n=1, 2, \dots, 10)$ 建立含水量模型，结果见表 2。

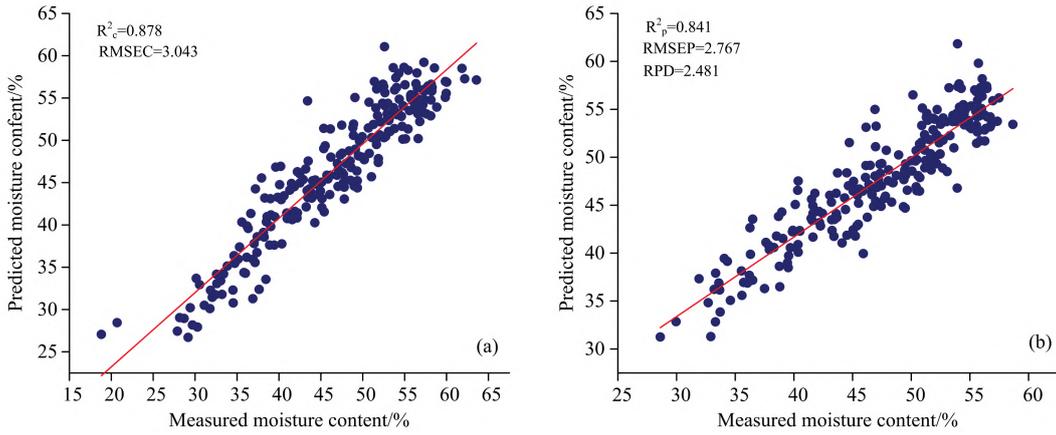


图 1 潮滩沉积物原始光谱含水量拟合图

(a): 建模集; (b): 预测集

Fig 1 Scatter plot of the measured against predicted moisture content using original spectra of tidal flat sediments

(a): Modeling set; (b): Prediction set

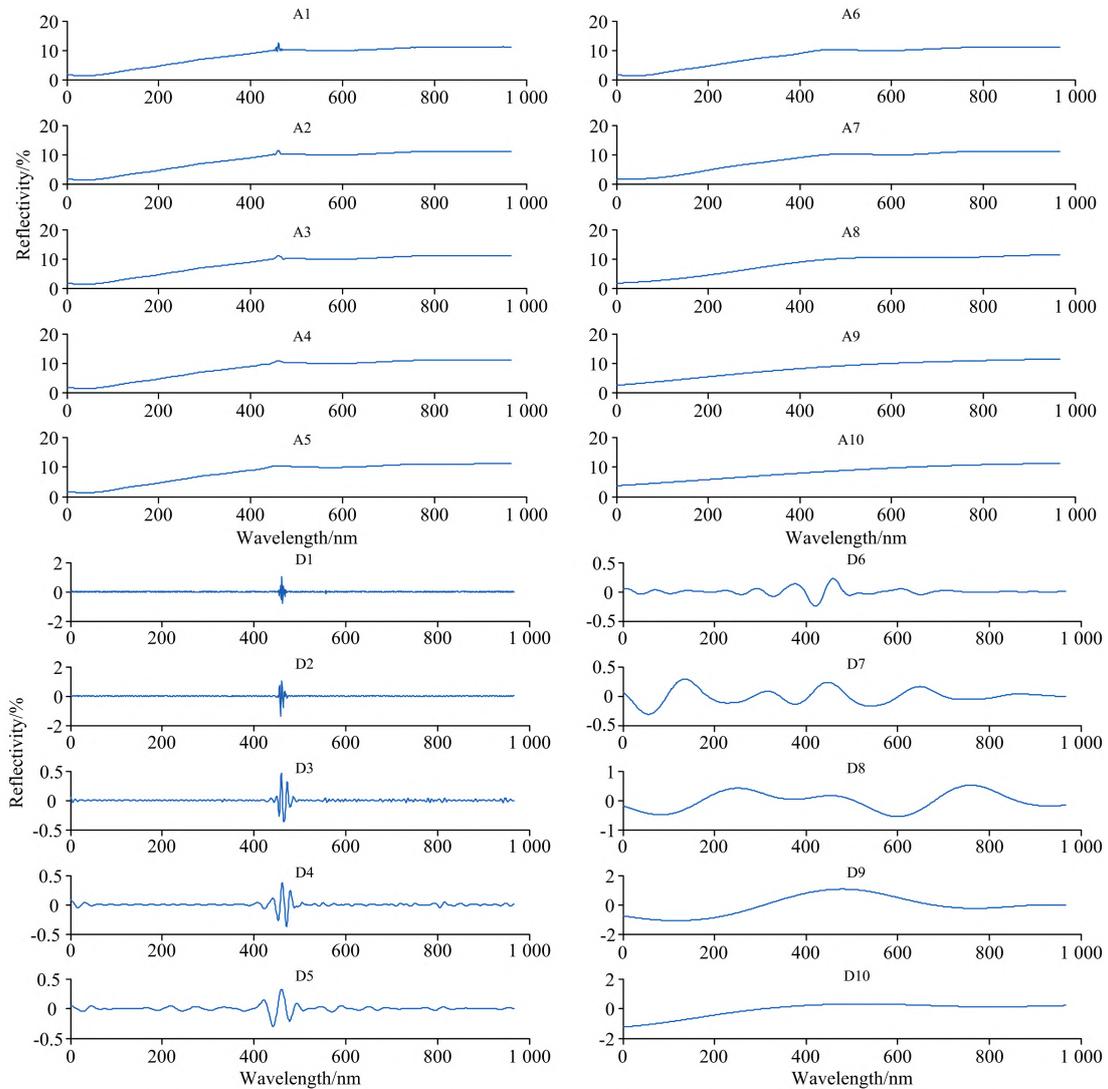


图 2 潮滩沉积物离散小波变换 (db10 小波基) 低频信息和高频信息光谱图

Fig 2 Low and high frequency information spectrograms of tidal flat sediment based on discrete wavelet transform (db10 wavelet basis)

表 2 db10 小波基潮滩沉积物低频和高频信息建模结果
Table 2 Modeling results of low and high frequency informations of sediment based on db10 wavelet

n	A_n			D_n		
	R_p^2	RMSEP	RPD	R_p^2	RMSEP	RPD
1	0.830	2.882	2.388	0.629	4.540	1.532
2	0.818	2.960	2.321	0.572	4.917	1.408
3	0.820	2.952	2.335	0.572	4.949	1.405
4	0.837	2.836	2.454	0.294	6.155	1.132
5	0.827	2.939	2.376	0.467	5.421	1.313
6	0.817	3.039	2.298	0.572	4.645	1.495
7	0.779	3.340	2.083	0.702	3.986	1.792
8	0.783	3.291	2.098	0.728	3.715	1.908
9	0.749	3.563	1.949	0.728	3.731	1.889
10	0.809	3.118	2.267	0.623	4.181	1.616

由表 2，低频信息 A_n 中， A_1 的预测结果最好， R_p^2 为 0.837，RPD 值为 2.454，RMSEP 为 2.837，但仍不如原始光谱预测结果。由表 3，高频信息 D_n 结果均较差，其中 D_1 模型结果最差，而 A_1 光谱信息中不包含 D_1 ，且模型结果最好，可初步判定， D_1 为干扰信息。为进一步确定 D_1 是否为干扰信息，分别将原始光谱 S 与高频信息 D_n 的差值 $S-D_n$ ($n=1, 2, \dots, 10$) 作为光谱输入，得到模型结果见表 3。

由表 3， $S-D_2, S-D_3, S-D_4, S-D_5$ 和 $S-D_6$ 光谱信息模型预测结果均优于原始光谱模型结果， R_p^2 在 0.845 以上，RMSEP 在 2.72 以下，RPD 值在 2.5 以上。其中 $S-D_3$ 和 $S-D_4$ 模型结果基本一致，且精度最高， R_p^2 分别为 0.928 和 0.927，RMSEC 分别为 2.348 和 2.365， R_p^2 分别为 0.870 和 0.873，RMSEP 分别为 2.5 和 2.495，RPD 值分别为 2.742 和 2.748。由于 $S-D_3$ 和 $S-D_4$ 结果明显高于其他 $S-D$ 结果，可判断 D_3 和 D_4 含有的无用信息较多，因此将原始光谱 S

与 D_3, D_4 做差值，去除 D_3 和 D_4 高频信息，得到 $S-D_3-D_4$ ，结果如表 3 所示。相比于 $S-D_3$ 和 $S-D_4$ ， $S-D_3-D_4$ 的 R_p^2 略低，为 0.924；RMSEC 略高，为 2.397；预测集中的 R_p^2 和 RPD 值略高，分别为 0.878 和 2.749；RMSEP 略高，为 2.501。 $S-D_3, S-D_4$ 和 $S-D_3-D_4$ 三个光谱信息模型整体结果相差不大。

表 3 db10 小波基潮滩沉积物 $S-D_n$ 光谱信息建模结果
Table 3 Modeling results of $S-D_n$ spectral information of tidal flat sediment based on db10 wavelet

尺度参数	建模集		预测集		
	R_c^2	RMSEC	R_p^2	RMSEP	RPD
$S-D_1$	0.871	3.127	0.829	2.882	2.388
$S-D_2$	0.918	2.489	0.853	2.658	2.586
$S-D_3$	0.928	2.348	0.870	2.500	2.742
$S-D_4$	0.927	2.365	0.873	2.495	2.748
$S-D_5$	0.887	2.927	0.850	2.699	2.541
$S-D_6$	0.879	3.035	0.845	2.720	2.521
$S-D_7$	0.881	3.005	0.838	2.799	2.456
$S-D_8$	0.846	3.392	0.834	2.896	2.405
$S-D_9$	0.924	2.408	0.841	2.774	2.466
$S-D_{10}$	0.866	3.192	0.835	2.804	2.459
$S-D_3-D_4$	0.924	2.397	0.878	2.501	2.749

2.3 基于小波变换 (sym6 小波基) 潮滩沉积物含水量预测结果

将潮滩沉积物原始光谱以 sym6 为小波基进行离散小波变换，同样得到 10 阶低频信息 A_n 和 10 阶高频信息 D_n ($n=1, 2, \dots, 10$)，分别以光谱的低频信息和高频信息建立 PLSR 含水量模型。为更直观看出 A_n 和 D_n 的含水量模型结果，将预测集 R_p^2 和 RPD 结果以折线图和柱状图表示，见图 3。

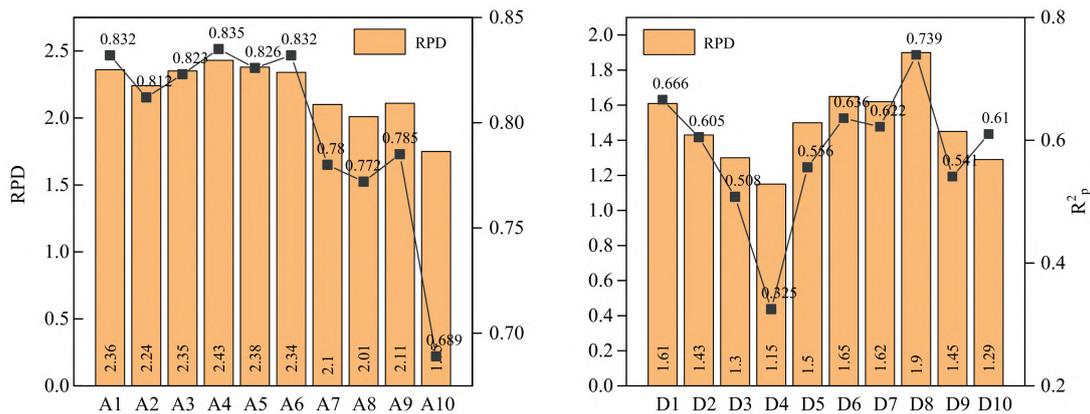


图 3 sym6 小波基潮滩沉积物高频和低频信息建模结果

Fig 3 Modeling results of low and high frequency informations of sediment based on sym6 wavelet

由图 3，低频信息中， A_1 预测结果最好，但仍差于原始光谱预测结果。高频信息 D_n ($n=1, 2, \dots, 10$) 结果均较差，其中 D_1 模型结果最差，与 db10 小波基结果一致。为判定 D_1 是否为主要干扰信息，分别将原始光谱 S 与高频信息 D_n (n

$=1, 2, \dots, 10$) 的差值作为光谱输入，得到模型结果见表 4。

由表 4， $S-D_4, S-D_5, S-D_6, S-D_8, S-D_9, S-D_{10}$ 光谱信息模型预测结果均优于原始光谱模型结果， R_p^2 在 0.845 以上，RMSEP 在 2.73 以下，RPD 值在 2.5 以上。对比图 4 结

表 4 sym6 小波基潮滩沉积物 $S-D_n$ 建模结果
Table 4 Modeling results of $S-D_n$ spectral information
of tidal flat sediment based on sym6 wavelet

尺度参数	建模集		预测集		
	R_c^2	RMSEC	R_p^2	RMSEP	RPD
$S-D_1$	0.900	2.758	0.832	2.913	2.356
$S-D_2$	0.783	4.061	0.846	2.907	2.365
$S-D_3$	0.874	3.089	0.840	2.764	2.489
$S-D_4$	0.884	2.973	0.848	2.714	2.531
$S-D_5$	0.928	2.332	0.867	2.542	2.702
$S-D_6$	0.880	3.018	0.848	2.708	2.540
$S-D_7$	0.858	3.271	0.840	2.778	2.484
$S-D_8$	0.924	2.398	0.864	2.562	2.686
$S-D_9$	0.921	2.456	0.869	2.497	2.755
$S-D_{10}$	0.907	2.633	0.845	2.728	2.546
$S-D_5-D_9$	0.920	2.473	0.870	2.475	2.768

果判定的 D_4 为干扰信息, 虽然 $S-D_4$ 相比于原始光谱结果有所提高, 但是模型精度提高有限, 因此干扰信息有待于进一步讨论。 $S-D_9$ 模型精度最高, R_c^2 为 0.921, RMSEC 为 2.456, R_p^2 为 0.869, RMSEP 为 2.497, RPD 值为 2.755; 其次为 $S-D_5$, R_c^2 为 0.928, RMSEC 为 2.332, R_p^2 为 0.867, RMSEP 为 2.542, RPD 值为 2.702。 $S-D_9$ 和 $S-D_5$ 模型结果最好, 可判定 D_5 和 D_9 含有的干扰信息较多, 因此将原始光谱 S 与 D_5 、 D_9 做差值, 得到 $S-D_5-D_9$, 模型结果略有提高,

R_p^2 为 0.87, RMSEP 为 2.475, RPD 值为 2.768。

3 结 论

基于可见-近红外光谱技术, 采集青岛市东大洋村潮间带的沉积物, 分别以 db10 小波基和 sym6 小波基对原始光谱进行小波变换, 以偏最小二乘回归建立潮滩沉积物含水量模型。通过 10 阶小波变换获取原始光谱的低频信息 A_n 和 高频信息 D_n ($n=1, 2, \dots, 10$), 通过原始光谱 S 分别与高频信息 D_n 做差值, 得到 $S-D_n$, 对 A_n , D_n , $S-D_n$ 建立潮滩沉积物含水量模型, 并对模型结果进行分析。通过 db10 小波基变换后的低频和高频信息, 无用信息主要集中在 D_3 和 D_4 , 去除 D_3 和 D_4 建立的含水量模型, 相比于原始光谱模型精度有明显提高, R_p^2 为 0.878, RMSEP 为 2.501, RPD 值为 2.749; 通过 sym6 小波基变换后的低频和高频信息, 无用信息主要集中在 D_5 和 D_9 , 去除 D_5 和 D_9 建立含水量模型与原始光谱模型相比, 精度也有一定提高, R_p^2 为 0.87, RMSEP 为 2.475, RPD 值为 2.768。通过小波变换对原始光谱划分低频信息和 高频信息进行分析, 能够有效找到潮滩沉积物含水量的干扰信息, 实现提取特征信息, 从而建立准确度更高的潮滩沉积物含水量模型, 为潮滩沉积物含水量实时、动态监测提供理论基础。

References

- [1] ZHANG Ying, ZHANG Dong, HU Ping-xiang(张 鹰, 张 东, 胡平香). Acta Oceanologica Sinica(海洋学报), 2008, 30(5): 29.
- [2] Klemas V, Finkl C, Kabbara N. Journal of Coastal Research, 2014, 30(4): 685.
- [3] Manage L P M, Greve M H, Knadel M, et al. Soil Science Society of America Journal, 2018, 82(6): 1333.
- [4] Nolet C, Poortinga A, Roosjen P, et al. Plos One, 2014, 9: e112151.
- [5] LI Xin-xing, LIANG Bu-wen, BAI Xue-bing, et al(李鑫星, 梁步稳, 白雪冰, 等). Spectroscopy and Spectral Analysis(光谱学与光谱分析), 2020, 40(12): 59.
- [6] Verpoorter C, Carrère, V, Combe J P. Journal of Geophysical Research Earth Surface, 2015, 119(10): 2098.
- [7] Ma Y, Fan X. Asia-Pacific Journal of Chemical Engineering, 2020, 15(1): e2507.
- [8] Shin H, Yu J, Kim J, et al. Remote Sensing Letters, 2015, 6(4-6): 488.
- [9] Shin H, Yu J, Jeong Y, et al. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 2017, 10(10): 4512.
- [10] Zhang F D, Liu J, Lin J, et al. Infrared Physics & Technology, 2019, 97: 224.
- [11] Song J, Li G, Yang X. Journal of the Science of Food and Agriculture, 2019, 99(11): 4898.

Prediction of Tidal Flat Sediment Moisture Content Based on Wavelet Transform

LI Xue-ying^{1,2}, LI Zong-min^{3*}, CHEN Guang-yuan⁴, QIU Hui-min², HOU Guang-li², FAN Ping-ping^{2*}

1. School of Geosciences, China University of Petroleum (Huadong), Qingdao 266580, China

2. Institute of Oceanographic Instrumentation, Qilu University of Technology (Shandong Academy of Sciences), Qingdao 266061, China

3. College of Computer Science and Technology, China University of Petroleum (Huadong), Qingdao 266580, China

4. College of Ocean Science and Engineering, Shandong University of Science and Technology, Qingdao 266590, China

Abstract The distribution of water in flat tidal sediments will change greatly in space and time, and the changes will lead to the changes of biogenic elements in sediments. Therefore, the tidal flat sediment water content data are monitored in real time, accurately and quickly, which is of great significance to understanding the tidal flat characteristics, grasp the information of tidal flat biogenic elements, and develop tidal flat resources. This paper collected 115 samples of intertidal sediments from Dongdayang village, Qingdao city. The visible near-infrared spectra and moisture content of fresh samples, air-dried for 4 weeks and 8 weeks were measured. The db10 and sym6 wavelet basis were used to transform the original spectrum, and partial least squares regression was used to establish the tidal flat sediment moisture content model. The low-frequency information A_n and high-frequency information D_n ($n=1, 2, \dots, 10$) of the original spectrum were obtained by 10 order wavelet transform. $S-D_n$ was calculated by the difference between the original spectrum S and D_n . The moisture content models were established using A_n , D_n and $S-D_n$, respectively, and the results were analyzed. The original spectrum model's R_p^2 , RMSEP and RPD were 0.841, 2.767 and 2.481. By analysing low-frequency and high-frequency information, after db10 wavelet basis transforms, the useless information was mainly concentrated in D_3 and D_4 , and the accuracy of the moisture content model established by removing D_3 and D_4 was significantly improved, R_p^2 was 0.878, RMSEP was 2.501, RPD was 2.749. Through the analysis of sym6 wavelet basis transform, the useless information was mainly concentrated in D_5 and D_9 , the R_p^2 , RMSEP and RPD by removing D_5 and D_9 were 0.87, 2.475 and 2.768. Therefore, by analyzing the low-frequency and high-frequency information using wavelet transform, the interference information of sediment moisture content can be effectively found, and the feature information can be extracted. The more accurate the tidal flat sediment moisture content model is established, it provides a theoretical basis for real-time and dynamic monitoring of tidal flat sediment moisture content.

Keywords Tidal flat sediment; Wavelet transform; Moisture content; Visible near infrared spectroscopy

(Received Mar. 16, 2021; accepted May 6, 2021)

* Corresponding authors