sensors

MDPI

Article

Identifying the Vertical Stratification of Sediment Samples by Visible and Near-Infrared Spectroscopy

Pingping Fan [1,2], Zongchao Jia [3], Huimin Qiu [1], Hongru Wang [1] and Yang Gao [1,2,*]

[1] Institute of Oceanographic Instrumentation, Qilu University of Technology (Shandong Academy of Sciences), Qingdao 266061, China; fanpp_sdioi@126.com (P.F.); qiuhm@qlu.edu.cn (H.Q.); wanghr@qlu.edu.cn (H.W.)
[2] Laoshan Laboratory, Qingdao 266237, China
[3] College of Engineering, Ocean University of China, Qingdao 266404, China; zongchao_j@126.com
* Correspondence: gy@sdioi.com

Abstract: Vertical stratification in marine sediment profiles indicates physical and chemical sedimentary processes and, thus, is the first step in sedimentary research and in studying their relationship with global climate change. Traditional technologies for studying vertical stratification have low efficiency; thus, new technologies are highly needed. Recently, visible and near-infrared spectroscopy (VNIR) has been explored to rapidly determine sediment parameters, such as clay content, particle size, total carbon (TC), total nitrogen (TN), and so on. Here, we explored vertical stratification in a sediment column in the South China Sea using VNIR. The sediment column was 160 cm and divided into 160 samples by 1 cm intervals. All samples were classified into three layers by depth, that is, 0–50 cm (the upper layer), 50–100 cm (the middle layer), and 100–160 cm (the bottom layer). Concentrations of TC and TN in each sample were measured by Elementa Vario EL III. Visible and near-infrared reflectance spectra of each sample were collected by Agilent Cary 5000. A global model and several classification models for vertical stratification in sediments were established by a Support Vector Machine (SVM) after the characteristic spectra were identified using Competitive Adaptive Reweighted Sampling. In the classification models, K-means clustering and Density Peak Clustering (DPC) were employed as the unsupervised clustering algorithms. The results showed that the stratification was successful by VNIR, especially when using the combination of unsupervised clustering and machine learning algorithms. The correct classification rate (CCR) was much higher in the classification models than in the global model. And the classification models had a higher CCR using K-means combined with SVM (94.8%) and using DPC combined with SVM (96.0%). The higher CCR might be derived from the chemical classification. Indeed, similar results were also found in the chemical stratification. This study provided a theoretical basis for the rapid and synchronous measurement of chemical and physical parameters in sediment profiles by VNIR.

Keywords: sediment; spectroscopy; South China Sea; vertical profile; unsupervised clustering

check for updates

Citation: Fan, P.; Jia, Z.; Qiu, H.; Wang, H.; Gao, Y. Identifying the Vertical Stratification of Sediment Samples by Visible and Near-Infrared Spectroscopy. *Sensors* **2024**, *24*, 6610. https://doi.org/10.3390/s24206610

Academic Editor: Barry K. Lavine

Received: 9 September 2024
Revised: 10 October 2024
Accepted: 11 October 2024
Published: 14 October 2024

Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

Sediments are one of the three most important components in marine ecosystems, recording information about the ocean's past and future and regulating the ocean's health. Marine sediments undergo a long and complex deposition process, forming different layers vertically [1–3]. These layers have distinctive physical and chemical characteristics, which are related to different carbon cycling processes and global climate change [4,5]. Therefore, vertical stratification in marine sediment profiles is the first step in sedimentary research and in studying their relationships with global climate change [6–8].

Traditionally, stratifying vertical profiles in marine sediments is usually carried out in a lab and is laborious, time-consuming, and costly [9,10]. Therefore, new technologies are urgently needed for rapidly determining vertical stratification [9,10]. Visible and near-infrared reflectance spectroscopy (VNIR) is a green technology with fast speeds, high

sensitivity, and easy operation, being widely used in many fields [7,8]. In sediments or soils, many achievements have been made in spectral analysis of parameters by VNIR, e.g., clay content, particle size, total carbon, organic carbon, and total nitrogen [6,11–14].

Some studies have explored the use of VNIR on soil taxonomic classification or soil chronosequence classification by machine learning techniques. Using a Support Vector Machine (SVM), Zheng et al. (2019) studied soil chronosequence classification and achieved an average classification accuracy of 93.1% [6]. Jiang et al. (2021) built soil horizons using a deep learning model based on the U-net network architecture, reaching an average classification accuracy of 83% [15]. Zhang et al. (2021) developed a random forest model to identify soil horizons using a radial basis kernel SVM and achieved an accuracy of more than 70% [16]. These studies demonstrated the efficacy of VNIR in stratifying soil profiles [7,8].

Few studies have reported vertical stratification by VNIR in sediments, especially in marine sediments. Since sediments are similar to soils in terms of deposition processes and chemistry, studies on soil taxonomic/chronosequence classification could provide a sufficient reference for sediment profile stratification. Here, we explored the vertical stratification of sediments in the South China Sea by VNIR using different strategies and algorithms.

2. Materials and Methods

2.1. Sampling and Preparation

A sediment column was collected from the Northern Slope of the South China Sea (18°53.03′ N, 114°47.29′ E) in September 2020 using an SDIOI-SC100 gravity column sampler (Figure 1). This is one of the most important tropical ecosystems in China, including coral reefs, mangroves, seagrass beds, etc. It is rich in resources for the biological carbon pump [17]. The sampler was constructed from stainless steel and had a weight of 100 kg. The length and inside diameter of the inner tube of the sampler were 180 cm and 12 cm, respectively.



Figure 1. Sampling site in the South China Sea.

The sediment column was 160 cm long (Figure 2). Here, we explored rapid classification for sediment vertical profiles by spectroscopy, so we used a random physical segmentation. The column was segmented at 1 cm intervals. All samples were classified into three layers, that is, 0–50 cm (the upper layer), 50–100 cm (the middle layer), and 100–160 cm (the bottom layer). Then, all samples were freeze-dried, ground, and sieved. Each sample was divided into two parts: one for spectrum collection and the other for chemical analysis.

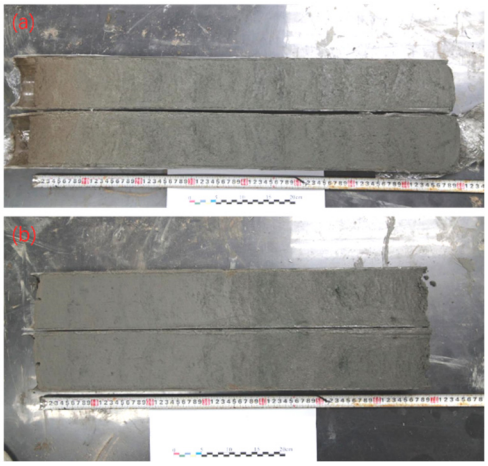


Figure 2. Sediment column. (**a**) 0–80 cm. (**b**) 81–160 cm.

2.2. Data Collection

The visible and near-infrared reflectance spectra of these samples were collected using an Agilent Cary 5000 (Specifications see Table S1). Cary 5000 was equipped with a special diffuse reflectance module designed for powdered samples, known as DRA-2500 (Agilent). The wavelength interval was set at 1 nm, and the scanning speed was set to 1200 nm/min. For each sample, five replicates of spectral data were collected. Consequently, the average spectra were utilized for each sample in the subsequent analyses. The concentrations of total carbon (TC) and total nitrogen (TN) in each sample were determined using a Vario EL III Elemental Analyzer at the Institute of Botany, Chinese Academy of Sciences. A total of 160 sample sets were collected. These 160 sample sets were analyzed, as shown in Figure 3.

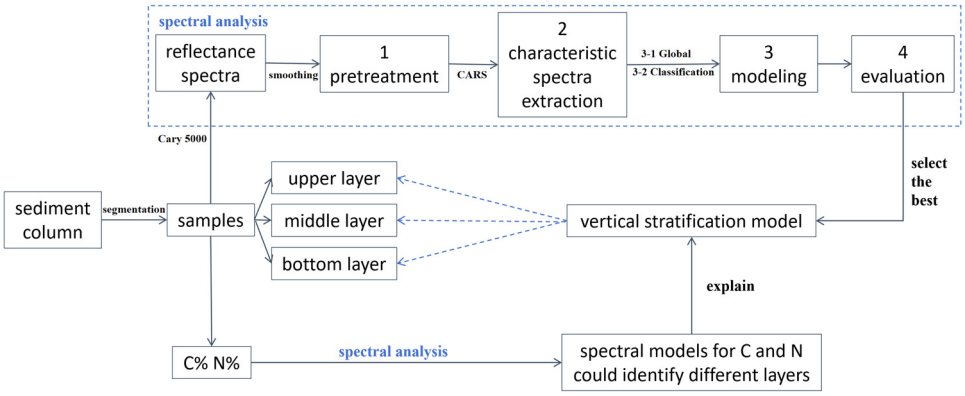


Figure 3. A flow diagram of data analysis in this study.

2.3. Spectral Analysis

All reflectance spectra were pretreated using Savitzky–Golay (SG) filtering, with a window size of 10 wavelengths and a polynomial order of 2. Then, a global model and classification models were established to identify the stratification of each sample.

2.3.1. Global Model

All samples were divided into a calibration set and a validation set, with a ratio of 2:1 by Kennard-Stone (K-S) [18]. Then, Competitive Adaptive Reweighted Sampling (CARS) was employed to identify the characteristic spectra [19]. Finally, a Support Vector Machine (SVM) and Partial Least Squares Regression (PLSR) were used to establish the spectral stratification model [20].

CARS was first proposed by Li et al. (2009) [19]. Specifically, N iterative Monte Carlo sampling was used to generate N wavelength subsets first. Secondly, the regression

coefficient of each wavelength feature was calculated within each subset of the PLSR model. The regression coefficient of the wavelength represents the importance score of the feature. The regression coefficient for a wavelength serves as an indicator of the feature's importance. Thirdly, the results were ranked in descending order based on the absolute value of these regression coefficients. Then, the root mean square error of cross-validation (RMSECV) for each wavelength subset was determined using an exponential decreasing function. After N Monte Carlo sampling was completed, the wavelength subset with the smallest RMSECV was selected as the characteristic wavelength.

An SVM is the most commonly used classification method [20]. It maps the vector into a higher-dimensional space through the kernel function to construct an optimal classification hyperplane. It identifies two parallel hyperplanes that are maximally distant and aligned with the classification hyperplane. Then, a hyperplane is defined by the equation $f(x) = \omega x + b = 0$, where ω is the normal vector of the classification plane, b is the bias of the classification plane, and $f(x) = \omega x + b$ is the classification function. The greater the distance between parallel hyperplanes is, the higher the classification accuracy of the classifier is [21]. An SVM follows the principle of structural risk minimization, which can significantly reduce the occurrence of locally optimal solutions and nonlinear overfitting when dealing with a small number of samples in other machine learning algorithms [8].

PLSR is a widely used multivariate statistical analysis technique in scientific research. Compared to traditional multivariate linear regression and principal component regression, PLSR has the advantage of both effectively compressing spectral data and fully analyzing spectral data. Consequently, spectral models could be established with enhanced stability and superior prediction capabilities [22].

In this study, we selected the Gaussian radial basis function as the kernel function in SVM models. It projects the data into a higher-dimensional space, thereby effectively solving the problem of linear inseparability in the original data space. All data analyses were conducted in Matlab R2021a.

2.3.2. Classification Model

Unsupervised clustering was used to divide all spectra into two subsets. These subsets were defined by K-means clustering and the Density Peak Clustering (DPC). In each subset, a spectral vertical stratification model was constructed as a global model.

K-means clustering is a distance-based clustering algorithm. Its basic idea is to find a partition scheme for K clusters interactively so the loss function corresponding to the clustering can be minimized. Here, the loss function was defined as the sum of error squares for the distance between each sample and the central point of the cluster:

$$J(c, \mu) = \sum_{i=1}^{M} ||x_i - \mu_{c_i}||^2 \quad (1)$$

where x_i is the sample i, c_i is the cluster x_i belongs to, μ_{c_i} is the central point corresponding to the cluster, and M is the total number of samples.

DPC is an unsupervised clustering algorithm based on distance and density. The selection and number of clustering centers are determined by the local density ρ and the minimum distance δ. The local density is calculated by the Gaussian kernel method [23]. The execution steps of the DPC algorithm are as follows:

(1) Calculate the distance matrix from sample set data;
(2) Determine the neighborhood truncation distance d_c;
(3) Calculate the local density of each point ρ_i, j;

$$\rho_i = \sum_{i \neq j} \exp\left[-\left(\frac{d_{ij}}{d_c}\right)^2\right] \quad (2)$$

$$\delta_i = \min_{j:\rho_j > \rho_i} (d_{ij}) \quad (3)$$

(1) Calculate the offset distance of each point δ_i;
(2) Estimate the cluster center points;
(3) Classify the non-clustered central data points.

2.3.3. Model Evaluation

In the stratification models, the Correct Classification Rate (CCR) was used to evaluate the classification results:

$$CCR = \frac{TP}{N} \times 100\% \quad (4)$$

where TP represents the correct number, and N represents the total number. The higher the CCR is, the higher the accuracy of the classification model is [24].

Meanwhile, the determination coefficient (R^2), root mean square error (RMSE), and relative predictive deviation (RPD) were also used to evaluate the spectral models.

$$R^2 = 1 - \frac{\sum_{i-1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \overline{y})^2} \quad (5)$$

$$RMSE = \sqrt{\frac{\sum_{i-1}^{n}(y_i - \hat{y}_i)^2}{n}} \quad (6)$$

$$RPD = \frac{\sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(y_i - \overline{y})^2}}{\sqrt{\frac{1}{n-1}\sum_{i-1}^{n}(y_i - \hat{y}_i)^2}} \quad (7)$$

where y_i represents the true value, $\hat{y}_i$ represents the modeled value, $\overline{y}$ represents the mean of the true values of all samples, and n represents the number of samples. A better model had a lower RMSE, higher R^2, and larger RPD [25].

3. Results

We explored the use of visible and near-infrared spectroscopy in the vertical stratification of marine sediments. By employing machine learning algorithms, we established excellent spectral models to realize vertical stratification in sediment profiles in the South China Sea.

3.1. Physical Stratification of Sediment Profiles

All samples were classified into three layers, that is, the upper layer (0–50 cm), the middle layer (50–100 cm), and the bottom layer (100–160 cm). We first studied the spectral characteristics in the three layers and then established the stratification models.

3.1.1. Spectral Characteristics in the Physical Profile

The reflectance rates of all samples were 10–50% (Figure 4a). The reflectance rates increased with wavelength at the range of 10–40% in the visible band (350–750 nm) while being concentrated at 40–50% in the near-infrared band (750–2500 nm).

The three layers had different reflectance spectra (Figure 4b). The bottom layer (100–160 cm) had the largest reflectance rates compared to the other layers. The upper layer (0–50 cm) had the smallest reflectance rates but had no significant difference compared to the middle layer (50–100 cm).

Principal component analysis (PCA) also showed the differences among these layers (Figure 5). The first two principal components could explain the difference in reflectance among the three layers, in which PC1 explained 82.3% of the difference and PC2 explained 12.0% of the difference. These results supported the evidence on the classification of the three layers.

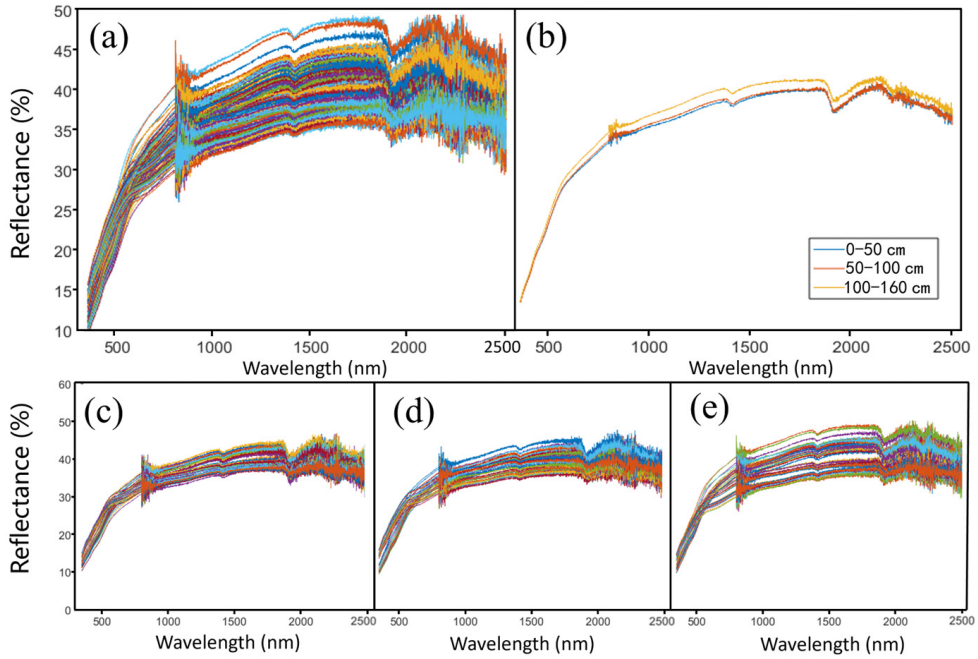


Figure 4. Reflectance spectra of sediments in all samples and different layers. (**a**) All samples. (**b**) The average spectrum of each layer. (**c**) The upper layer. (**d**) The middle layer. (**e**) The bottom layer.

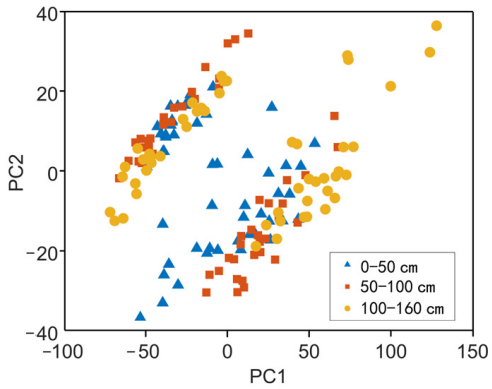


Figure 5. Principle component analysis of reflectance spectra in sediments of different layers.

3.1.2. Physical Stratification Models

Classification models could greatly improve the performance of physical stratification, especially after the extraction of characteristic spectra. The Correct Classification Rate (CCR) was higher in the classification model than in the global model (Tables 1 and 2). The CCR was 70.6% in the global model and 76.9–88.3% in the classification models without any characteristic spectrum extraction algorithms (Table 1). After CARS was used, the CCR increased to 76.5% in the global model and >92.6 in the classification models (Table 2).

Table 1. Correct classification rates (CCRs) in the global model.

Characteristic Spectra Selection Method	Correct Classification Rates (CCRs)	
	Calibration Set	Validation Set
None	100%	70.6%
Competitive Adaptive Reweighted Sampling	100%	76.5%

Table 2. Correct classification rates (CCRs) in the classification models.

Clustering Method	Subset	Characteristic Spectra Selection Method	Correct Classification Rates (CCRs)	
			Calibration Set	Validation Set
K-means	1	None	100%	77.8%
		Competitive Adaptive Reweighted Sampling	100%	92.6%
	2	None	100%	78.3%
		Competitive Adaptive Reweighted Sampling	97.9%	95.7%
Density Peak Clustering	1	None	100%	76.9%
		Competitive Adaptive Reweighted Sampling	100%	96.2%
	2	None	100%	83.3%
		Competitive Adaptive Reweighted Sampling	100%	95.8%

The use of unsupervised clustering could produce a priming effect of CARS on the CCR. In the classification models, CARS played a more important role (Table 2). After using K-means combined with CARS, the CCR increased from 78% to more than 92.6%. And after using DPC combined with CARS, the CCR increased to more than 95.8%. Compared to K-means, the use of DPC combined with CARS could generate a larger CCR.

3.2. Physical Stratification Is Correlated with Chemistry in Sediment Profiles

The formation of physical layers in marine sediments is highly related to marine chemistry, especially organic matter. Therefore, we studied the chemical characteristics in the sediment profile, including total carbon (TC) and total nitrogen (TN).

The concentrations of TC and TN in the profile are shown in Figure 6. The distribution of TC and TN is complicated and unclear. There were no distinctive layers according to the distribution of TC or TN, respectively.

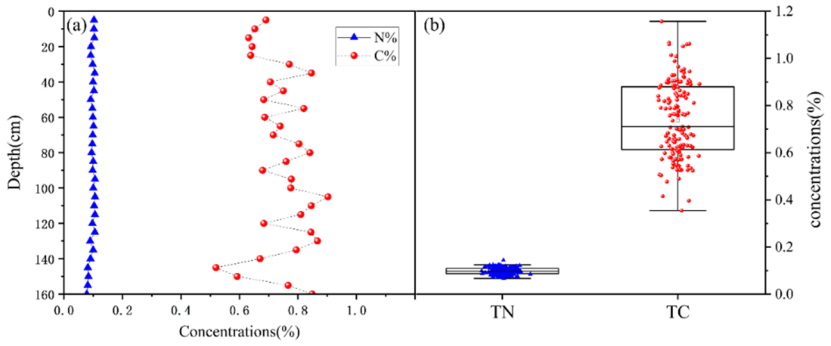


Figure 6. Chemical distribution of total carbon and total nitrogen. (**a**) Concentrations of total carbon and total nitrogen in the sediment profile, (**b**) boxplot of total carbon and total nitrogen.

However, considering both TC and TN simultaneously, the difference among the three layers was significant. As shown in Figure 7, the upper layer had the smallest TC concentrations and the largest TN concentrations. The middle layer had the largest TC concentrations and the largest TN concentrations. The bottom layer had the largest TC concentrations and the smallest TN concentrations. Although the TC concentrations were not different among the layers, the TN concentrations were significantly different (Figure 7). These results indicated that the three layers had different chemistry.

Indeed, TC and TN could be accurately determined by spectroscopy. In our study, spectral models for TC and TN performed better for each layer (Figure 8). For the TC models, the determination coefficient (R^2) was as high as 0.96, and the relative prediction deviation (RPD) was as high as 4.7 (Figure 8c). The TN models performed as well as the TC models (Figure 8).

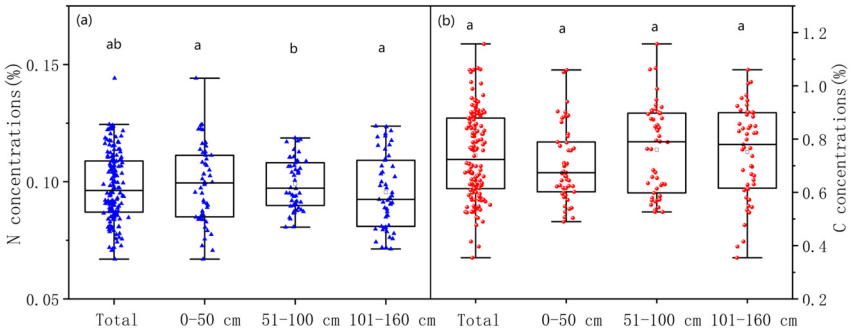


Figure 7. Boxplot of concentrations of total nitrogen (**a**) and total carbon (**b**) in all samples and different layers. In each subfigures, the characters "a", "b", and "ab" are the results of ANOVA for average concentrations of total carbon and total nitrogen.

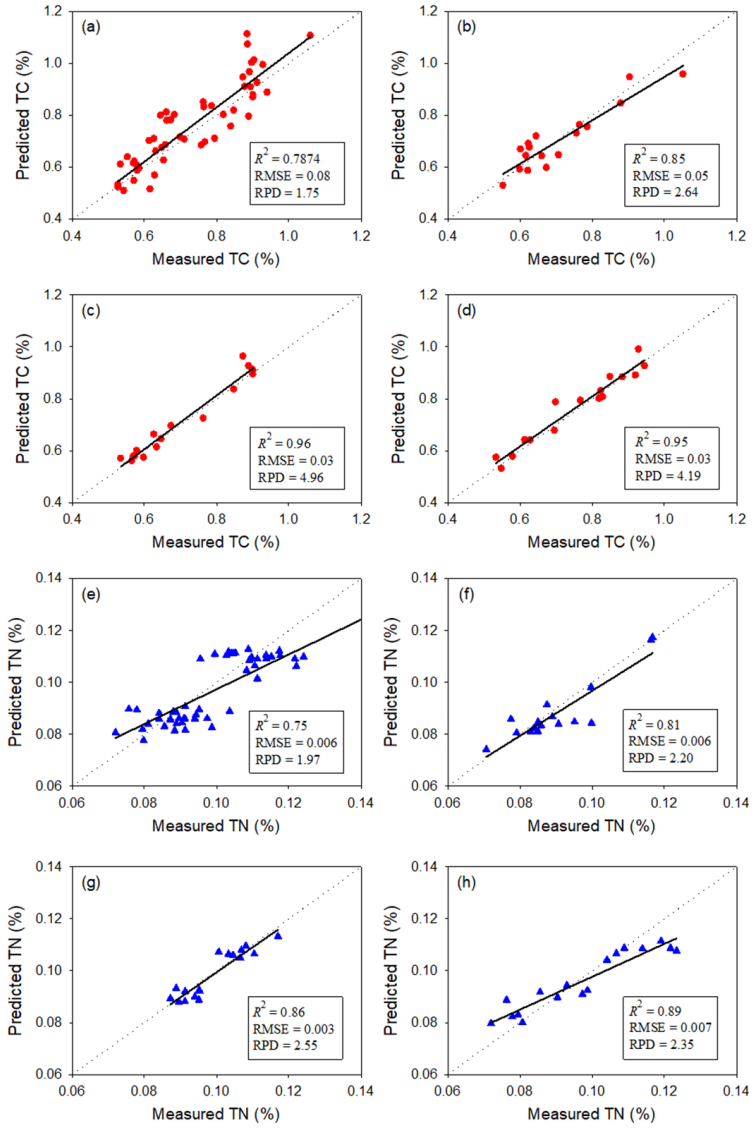


Figure 8. Results of total carbon and total nitrogen by global model and classification models. (**a**) TC in all samples, (**b**) TC in the upper layer, (**c**) TC in the middle layer, (**d**) TC in the bottom layer, (**e**) TN in all samples, (**f**) TN in the upper layer, (**g**) TN in the middle layer, (**h**) TC in the bottom layer. In subfigures, the red circles and blue triangles were TC data and TN data, respectively; in the dotted lines (1:1 lines) the predicted values equals to the measured values; the solid lines were the fitted curves whose parameters were listed in the left box.

In addition, we further cross-validated the classification spectral models across all layers. For example, we used the TC model of the upper layer to predict the samples in the middle layer and bottom layer. The results showed that a spectral model in a specific layer was not well suited for samples in other layers (Table 3). These results also support the results of physical stratification.

Table 3. Validation results of TC and TN in sediments of all three layers from the spectral model established using the samples in a specific layer.

Spectral Model	Sample	TC			TN		
		R^2	RMSE	RPD	R^2	RMSE	RPD
Upper layer	Upper layer	0.85	0.05	2.64	0.81	0.006	2.20
	Middle layer	0.26	0.17	0.96	0.63	0.008	1.22
	Bottom layer	0.15	0.21	0.81	0.59	0.013	1.28
Middle layer	Upper layer	0.22	0.16	0.85	0.39	0.014	1.24
	Middle layer	0.96	0.03	4.66	0.86	0.003	2.55
	Bottom layer	0.35	0.15	1.13	0.66	0.010	1.23
Bottom layer	Upper layer	0.35	0.15	1.13	0.33	0.014	1.23
	Middle layer	0.43	0.13	1.31	0.38	0.008	1.19
	Bottom layer	0.95	0.03	4.19	0.89	0.007	2.35

4. Discussion

Our study showed that the marine sediment vertical profile exhibited a complicated but regular structure and morphology. Visible and near-infrared spectroscopy (VNIR) is a green technology with fast speeds, high sensitivity, and easy operation. And VNIR has been proven to be an efficient technology in many fields. There are some studies on soil taxonomic classification or soil chronosequence that have used VNIR [6–8,15,16], but application on sediment vertical stratification is rare. Our study showed that the physical layer could be classified by spectroscopy, mainly because these layers were highly related to chemical properties, such as TC and TN. Our study provided a new application for visible and near-infrared spectroscopy (VNIR) on marine sediments. In the future, more parameters should be developed by VNIR synchronously, which could make marine sediments easily "seen" in the sea.

In our study, stratification models performed better than the past studies on soil profiles. In our study, the R^2 of spectral models could reach as high as 0.96, and the correct classification rate (CCR) achieved up to 96%. In past studies on soil profiles, the highest CCR was 93.1% [6], which is smaller than that in our study and may be because effective algorithms were used in our study.

Obviously, algorithms play an important role in establishing excellent spectral models. Twenty years ago, Confalonieri et al. (2001) and Smith et al. (2001) compared the influence of different algorithms on spectral models [26,27]. Mouazen et al. (2010) found that machine learning algorithms had the advantage of establishing spectral models for soil nutrients [28]. Jia et al. (2014) proposed that the predominant step in establishing spectral models for soil nutrients was extracting the characteristic spectra [29]. In our study, CARS was used as the characteristic spectra extraction algorithm and showed a great effect on spectral models. And the unsupervised clustering of DCP performed better than K-means in classification models. These studies revealed the importance of algorithms in establishing spectral models.

This study showed that the combination of unsupervised clustering and CARS effectively improved the performance of vertical stratification. Because these algorithms played different roles in spectral analysis, their combination could nonlinearly amplify their different roles. Therefore, CARS showed a priming effect on vertical stratification after using unsupervised clustering in our study.

The results also highlighted the effectiveness of combining unsupervised clustering with machine learning to construct stratification models for sediments or soils. This was also suggested in other studies [6,30]. For example, Jiang et al. (2021) built soil horizons using a deep learning model based on the U-net network architecture, reaching an average classification accuracy of 83% [15].

Our study explored a new application of VNIR on marine sediments. Here, we used random physical segmentation to explore the rapid classification of sediment vertical profiles. In the future, we will increase the measurement accuracy of both carbon and nitrogen to improve the rapid determination of other related parameters by VINR, for example, how to scientifically segment a sediment profile and calculate the sedimentation rates using VNIR.

5. Conclusions

We explored vertical stratification in a sediment column sampled from the South China Sea by establishing a global model and classification models. Both the global model and classification models were established by a Support Vector Machine (SVM) and Partial Least Squares Regression (PLSR) after the characteristic spectra were identified using Competitive Adaptive Reweighted Sampling (CARS). In the classification models, K-means clustering and Density Peak Clustering (DPC) were employed as the unsupervised clustering algorithms. The results showed that physical stratification had a correct classification rate (CCR) of 76.5% in the global model. In the classification models, the CCR was 94.8% using K-means combined with CARS and 96.0% using DPC combined with CARS. Physical stratification could be influenced by chemical properties, such as total carbon and total nitrogen. These results suggested that vertical stratification in marine sediments could be rapidly identified by VNIR, especially when using the combination of unsupervised clustering and machine learning algorithms. This study provided an important technological basis for simulating marine sedimentation processes and establishing marine C cycling models.

Supplementary Materials: The following supporting information can be downloaded at: https://www.mdpi.com/article/10.3390/s24206610/s1, Table S1: The specifications of spectrometers Cary 5000.

Author Contributions: Conceptualization, P.F. and Y.G.; methodology, Z.J.; formal analysis, H.W.; investigation, H.Q. and Z.J.; writing—original draft preparation, P.F. and Z.J.; writing—review and editing, P.F. and Y.G.; funding acquisition, P.F., H.Q., and Y.G. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the National Natural Science Foundation of China (32171578, U2006209, 42307483, 42476188, and 41506110) and the Natural Science Foundation of Shandong Province, China (ZR2021MD093 and ZR2021MD103).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Data are contained within the article.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Dang, H.; Jiao, N. Perspectives on the microbial carbon pump with special reference to microbial respiration and ecosystem efficiency in large estuarine systems. *Biogeosciences* **2015**, *11*, 3887–3898. [CrossRef]
2. Luo, T.T.; Song, Y.C.; Zhu, Y.M.; Liu, W.G.; Liu, Y.; Li, Y.H.; Wu, Z.R. Triaxial experiments on the mechanical properties of hydrate-bearing marine sediments of South China Sea. *Mar. Petrol. Geol.* **2016**, *77*, 507–514. [CrossRef]
3. Qiu, H.; Fan, P.; Hou, G.; Li, X.; Wang, Y. Analysis and Model Comparison of Carbon and Nitrogen Concentrations in Sediments of the Yellow Sea and Bohai Sea by Visible-Near Infrared Spectroscopy. *Bull. Environ. Contam. Toxicol.* **2022**, *108*, 1124–1131. [CrossRef] [PubMed]
4. Du, J.; Guan, D.; Yao, Z.; Wang, Z.; Huo, C. Records of human-induced changes in sedimentation and carbon sequestration in Dalian Bay, north China. *Cont. Shelf Res.* **2019**, *178*, 51–58. [CrossRef]

5. Seki, A.; Tada, R.; Kurokawa, S.; Murayama, M. High-resolution quaternary record of marine carbon content in the hemipelagic sediments of the Japan Sea from bromine counts measured by XRF core scanner. *Prog. Earth Planet. Sci.* **2019**, *6*, 1. [CrossRef]
6. Zheng, G.H.; Ryu, D.; Jiao, C.; Xie, X.; Cui, X.; Shang, G. Visible and near-Infrared reflectance spectroscopy analysis of a coastal soil chronosequence. *Remote Sens.* **2019**, *11*, 2336. [CrossRef]
7. Linderholm, J.; Geladi, P.; Gorretta, N.; Bendoula, R.; Gobrecht, A. Near infrared and hyperspectral studies of archaeological stratigraphy and statistical considerations. *Geoarchaeology* **2019**, *34*, 311–321. [CrossRef]
8. Gholizadeh, A.; Rossel, R.A.V.; Mohammadmehdi, S.; Luboš, B.; Josef, K.; Lenka, P. National-scale spectroscopic assessment of soil organic carbon in forests of the Czech Republic. *Geoderma* **2021**, *385*, 114832. [CrossRef]
9. Song, T.; Song, T.; Pang, C.; Hou, B.; Xu, G.; Xue, J.; Sun, H.; Meng, F. A review of artificial intelligence in marine science. Front. *Earth Sci.* **2023**, *11*, 1090185.
10. Dennis, Y.C.L.; Giorgio, C.; Maroto-Valer, M.M. An overview of current status of carbon dioxide capture and storage technologies. *Renew. Sust. Energ. Rev.* **2014**, *39*, 426–443.
11. Liu, S.; Shen, H.; Chen, S.; Zhao, X.; Biswas, A.; Jia, X.; Shi, Z.; Fang, J. Estimating forest soil organic carbon content using vis-NIR spectroscopy: Implications for large-scale soil carbon spectroscopic assessment. *Geoderma* **2019**, *348*, 37–44. [CrossRef]
12. Ribeiro, S.G.; Adunias, S.T.; Marcio, R.R.O.; Mirian, C.G.C.; Isabel, C.S.A.; Luis, C.J.M.; Fernando, B.L. Soil organic carbon content prediction using soil-reflected spectra: A comparison of two regression methods. *Remote Sens.* **2021**, *13*, 4752. [CrossRef]
13. Semella, S.; Hutengs, C.; Seidel, M.; Ulrich, M.; Schneider, B.; Ortner, M.; Thiele-Bruhn, S.; Ludwig, B.; Vohland, M. Accuracy and reproducibility of laboratory diffuse reflectance measurements with portable VNIR and MIR spectrometers for predictive soil organic carbon modeling. *Sensors* **2022**, *22*, 2749. [CrossRef] [PubMed]
14. Xie, S.G.; Ding, F.J.; Chen, S.G.; Wang, X.; Li, Y.H.; Ma, K. Prediction of soil organic matter content based on characteristic band selection method. *Spectrochim. Acta A* **2022**, *273*, 120949. [CrossRef]
15. Jiang, Z.D.; Phillip, R.O.; Zhang, C.L.; Kristofor, R.B.; Weindorf, D.C.; Kabindra, A.; Sun, Z.X.; Sun, F.J.; Wang, Q.B. Towards a dynamic soil survey: Identifying and delineating soil horizons in-situ using deep learning. *Geoderma* **2021**, *401*, 115341. [CrossRef]
16. Zhang, Y.K.; Alfred, E.H.; Huang, J.Y. Spectral signatures of soil horizons and soil orders—An exploratory study of 270 soil profiles. *Geoderma* **2021**, *389*, 114961. [CrossRef]
17. Chen, J.F.; Wiesner, M.G.; Li, H.L.; Ran, L.H.; Sun, L. Biogeochemical fluxes and biological pump composition in the south china sea. *Bull. Miner. Petrol. Geoch.* **2015**, *34*, 911–919.
18. Kennard, R.W.; Stone, L.A. Computer aided design of experiments. *Technometrics* **1969**, *11*, 137–148. [CrossRef]
19. Li, H.D.; Liang, Y.Z.; Xu, Q.S.; Cao, D.S. Key wavelengths screening using competitive adaptive reweighted sampling method for multivariate calibration. *Anal. Chim. Acta* **2009**, *648*, 77–84. [CrossRef]
20. Vladimir, N.V. *Statistical Learning Theory*; Wiley-Interscience: New York, NY, USA, 1998.
21. Rossel, R.A.V.; Behrens, T. Using data mining to model and interpret soil diffuse reflectance spectra. *Geoderma* **2010**, *158*, 46–54. [CrossRef]
22. Cheng, H.; Shen, R.L.; Chen, Y.Y.; Wan, Q.J.; Shi, T.Z.; Wang, J.J.; Wan, Y.; Hong, Y.S.; Li, X.C. Estimating heavy metal concentrations in suburban soils with reflectance spectroscopy. *Geoderma* **2019**, *336*, 59–67. [CrossRef]
23. Rodriguez, A.; Laio, A. Machine learning. Clustering by fast search and find of density peaks. *Science* **2014**, *344*, 1492–1496. [CrossRef] [PubMed]
24. Saberioon, M.; Císař, P.; Labbé, L.; Souček, P.; Pelissier, P.; Kerneis, T. Comparative performance analysis of support vector machine, random forest, logistic regression and *k*-nearest neighbours in Rainbow Trout (*Oncorhynchus mykiss*) classification using image-based features. *Sensors* **2018**, *18*, 1027. [CrossRef] [PubMed]
25. Rossel, R.A.V.; McGlynn, R.N.; McBratney, A.B. Determining the composition of mineral-organic mixes using UV-vis-NIR diffuse reflectance spectroscopy. *Geoderma* **2006**, *137*, 70–82. [CrossRef]
26. Confalonieri, M.; Fornasier, F.; Ursino, A.; Boccardi, F.; Pintus, B.; Odoardi, M. The potential of near infrared reflectance spectroscopy as a tool for the chemical characterisation of agricultural soils. *J. Near Infrared Spec.* **2001**, *9*, 123–131. [CrossRef]
27. Smith, G.D.; Jervis, A.; Lennartsson, M.; Bourne, W.F. Laboratory methods of estimating potentially mineralizable nitrogen in organic potting mixes. II. Development of near infrared reflectance spectroscopy method. *Commun. Soil Sci. Plant.* **2001**, *32*, 2769–2781. [CrossRef]
28. Mouazen, A.M.; Kuang, B.; Baerdemaeker, J.D.; Ramon, H. Comparison among principal component, partial least squares and back propagation neural network analyses for accuracy of measurement of selected soil properties with visible and near infrared spectroscopy. *Geoderma* **2010**, *158*, 23–31. [CrossRef]
29. Jia, S.Y.; Yang, X.L.; Zhang, J.M.; Li, G. Quantitative analysis of soil nitrogen, organic carbon, available phosphorous, and available potassium using near-infrared spectroscopy combined with variable selection. *Soil Sci.* **2014**, *179*, 211–219. [CrossRef]
30. Fajardo, M.; McBratney, A.; Whelan, B. Fuzzy clustering of Vis-NIR spectra for the objective recognition of soil morphological horizons in soil profiles. *Geoderma* **2016**, *263*, 244–253. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.