



利用随机森林算法预测中国东部海域表层沉积物有机碳含量分布

张 潇, 李 超, 窦衍光, 王凤梅, 褚明欣, 密蓓蓓, 梅 西, 杨守业

Predicting the distribution of organic carbon content in surface sediments of the eastern China seas using random forest algorithm

ZHANG Xiao, LI Chao, DOU Yanguang, WANG Fengmei, CHU Mingxin, MI Beibei, MEI Xi, and YANG Shouye

在线阅读 View online: <https://doi.org/10.16562/j.cnki.0256-1492.2025012401>

您可能感兴趣的其他文章

Articles you may be interested in

基于机器学习的大洋玄武岩构造环境判别研究

Tectonic discrimination of oceanic basalt by machine learning

海洋地质与第四纪地质. 2024, 44(4): 190–199

南海天然气水合物智能识别方法与应用

Intelligent identification and application of gas hydrate in South China Sea

海洋地质与第四纪地质. 2024, 44(6): 25–33

南海南沙海区沉积有机质分布特征及其指示意义

Spatial distribution of organic matter in surface sediments from the Nansha sea area of the South China Sea and its implications for marine productivity and monsoon

海洋地质与第四纪地质. 2023, 43(2): 45–54

辽河口浅海湿地表层沉积物中正构烷烃的组成、分布及来源

Composition, distribution and origination of *n*-alkanes in surface sediments from shallow sea wetlands in the Liaohe River estuary

海洋地质与第四纪地质. 2025, 45(1): 18–28

水库调控下黄河口沉积有机碳的分布、来源与输运特征

Distribution, source, and transport of particulate organic carbon in the Yellow River estuary as affected by the water–sediment regulation

海洋地质与第四纪地质. 2024, 44(2): 120–130

南黄海表层沉积物稀土元素分布特征及其物源指示意义

Distribution of rare earth elements in surface sediments of the South Yellow Sea and its implication to sediment provenances

海洋地质与第四纪地质. 2022, 42(6): 93–103



关注微信公众号, 获得更多资讯信息

张潇, 李超, 窦衍光, 等. 利用随机森林算法预测中国东部海域表层沉积物有机碳含量分布 [J]. 海洋地质与第四纪地质, 2025, 45(4): 187-197.
ZHANG Xiao, LI Chao, DOU Yanguang, et al. Predicting the distribution of organic carbon content in surface sediments of the eastern China seas using random forest algorithm[J]. Marine Geology & Quaternary Geology, 2025, 45(4): 187-197.

利用随机森林算法预测中国东部海域表层沉积物有机碳含量分布

张潇¹, 李超¹, 窦衍光², 王凤梅², 褚明欣¹, 密蓓蓓², 梅西², 杨守业¹

1. 同济大学海洋地质全国重点实验室, 上海 200092

2. 中国地质调查局青岛海洋地质研究所, 青岛 266237

摘要:厘清中国边缘海沉积有机碳的分布特征和控制因素有助于建立东亚边缘海有机碳循环模型及其“源-汇”格局。当前中国东部边缘海有机碳分布图的绘制,主要是通过数学插值对采样点之间进行填充。该方法一方面极大地受限于采样站位的位置和数量,另一方面通过数学插值填图也忽视了样品与海水理化性质、海底地形和洋流等环境因素的差异,将复杂的地质问题简单化。机器学习方法能够从高维和复杂数据中提取关键信息,构建环境属性特征和预测变量的映射关系。本文借助机器学习方法中常用的随机森林算法,通过对 405 个海洋沉积物有机碳数据与 50 个环境属性特征映射关系的学习,预测了中国东部边缘海表层沉积物的有机碳含量。相比根据同样数量样品由克里金插值计算绘制的有机碳分布图,随机森林算法对沉积物有机碳含量预测结果的平均绝对误差、均方根误差、最大残差等误差评价指标均更小,十折交叉检验的 R^2 达到 0.6,表现出较高的拟合精度。尤其对于采样密度较低或因采样困难存在样品空缺的海区,随机森林算法能更准确的预测表层沉积物有机碳含量,体现出更符合实际情况的预测潜力和外推性优势。本文所建立的随机森林算法对于未来其他海洋沉积物地球化学指标的预测也同样具有借鉴作用,对于中国东部边缘海的资源调查和环境保护具有重要的现实意义。

关键词:空间插值;机器学习;随机森林;表层沉积物;有机碳含量

中图分类号:P736

文献标识码:A

DOI: 10.16562/j.cnki.0256-1492.2025012401

Predicting the distribution of organic carbon content in surface sediments of the eastern China seas using random forest algorithm

ZHANG Xiao¹, LI Chao¹, DOU Yanguang², WANG Fengmei², CHU Mingxin¹, MI Beibei², MEI Xi², YANG Shouye¹

1. State Key Laboratory of Marine Geology, Tongji University, Shanghai 200092, China

2. Qingdao Institute of Marine Geology, China Geological Survey, Qingdao 266237, China

Abstract: Clarifying the distribution characteristics and controlling factors of sedimentary organic carbon in China marginal seas is crucial for establishing an organic carbon cycle model for the East Asian marginal seas and its “source-to-sink” pattern. Currently, the distribution map of organic carbon in the Eastern China marginal seas is constructed mainly based on mathematical interpolation of existing data. However, this method is significantly limited by the location and quantity of sampling stations, and in addition, the mathematical interpolation mapping neglects the differences between the samples and environmental factors such as seawater physicochemical properties, seabed topography, and ocean currents, thus oversimplifying the complex geological issues. Machine learning methods can extract key information from high-dimensional and complex data and establish mapping relationships between geological property features and predictive variables. In this study, the commonly used Random Forest (RF) algorithm in machine learning was employed to predict the organic carbon content in the surface sediments of the Eastern China marginal seas by learning the mapping relationship among 405 marine sediment organic carbon data and 50 geological property features. Compared to the organic carbon distribution map generated by the Kriging interpolation calculations based on the same number of samples, the RF algorithm showed smaller errors of evaluation indicators, including mean absolute error, root mean square

资助项目:上海市科委国际合作项目“利用人工智能算法预测洋底 Th 同位素分布及全球海洋沉积通量”(24510714500);国家自然科学基金项目“根据河流沉积物($^{234}\text{U}/^{238}\text{U}$)计算流域尺度的平均侵蚀速率”(42273057)

作者简介:张潇(2000—),女,硕士研究生,海洋科学专业, E-mail: 2231657@tongji.edu.cn

通讯作者:李超(1983—),男,副教授,主要从事海洋沉积地球化学相关研究, E-mail: cli@tongji.edu.cn

收稿日期:2025-01-24; 改回日期:2025-02-23. 张现荣编辑

error, and maximum residual error. The ten-fold cross-validation R^2 reached 0.60, indicating high fitting accuracy. Notably, for regions with low sampling density or missing data due to sampling difficulties, the RF algorithm demonstrated a superior predictive accuracy for surface sediment organic carbon content, reflecting its potential for more realistic predictions and extrapolation advantages. The RF model established in this study provided valuable insights for predicting other geochemical indicators of marine sediments in the future and holds significant practical implications for resource investigation and environmental protection in the Eastern China marginal seas.

Key words: spatial interpolation; machine learning; random forest; surface sediment; organic carbon contents

边缘海沉积有机碳的地球化学行为和埋藏特征,是表生过程碳循环研究的关键环节之一。在世界级大河的影响下,边缘海具有高陆源输入和高生产力的特点,因而其有机质的埋藏效率也更高。虽然边缘海只占全球海洋面积的8%,但其沉积有机质埋藏量约占全球海洋有机质埋藏量的80%^[1],相当于全球海洋通过沉积作用长期固定的大气CO₂总量的20%左右^[2]。厘清中国边缘海沉积有机碳的分布特征和控制因素,对于建立东亚边缘海有机碳循环模型和其“源-汇”格局具有重要意义;同时,也为中国应对有关碳排放的环境政策制定和未来气候变化趋势的预测提供科学依据。

中国东部边缘海接收了大量来自长江、黄河等周边大型河流注入的有机碳、颗粒物和营养盐等陆源物质^[3-4]。在中-晚全新世当海平面达到高海面后,这些物质在东亚边缘海独特的地形和洋流体系作用下沉积埋藏形成了多个泥质沉积区^[4]。而不同泥质区的地理位置和沉积环境差异导致有机碳的埋藏机制也各不相同^[5-9],这一差异可能影响TOC含量的空间分布模式。例如,高沉积速率区主要集中在黄河口水下三角洲和长江口及其邻近海域,黄河的现代沉积速率显著低于渤海和东海^[10]。中国东部边缘海沉积物中的总有机碳含量在空间分布上与泥质区的分布相似,渤海湾西部和渤海中部泥质区的有机碳含量较高,而辽东浅滩和渤海海峡,粒径较粗且沉积速率较低,有机碳含量也低。东海高有机碳含量区主要分布在长江口泥质区和内陆架泥质区,这里具有较高的沉积速率。而外陆架的砂质沉积区具有较低的有机碳含量,主要是因为有机碳主要与黏土等细颗粒物质结合,在砂粒级沉积物中含量较低^[11-12]。

目前已有不少学者对中国东部边缘海沉积物中有机碳的分布开展了广泛的调查,并系统绘制了有机碳等地球化学指标的分布图^[5-6,13],但是图件绘制中样品点之间大多采用数学插值(如克里金插值)进行预测和填充。另一方面,由于海洋沉积物采集成本较高,且涉及争议的水域很难获得样品,因此海洋沉积物样品的空间分辨率不高,甚至存在

一些样品空白区域。

在海洋学相关参数预测方面,相比于统计学和传统数值模拟模型,机器学习因其无需条件假设,凭借数据的时空依赖性,利用多元线性回归模型进行数值预测^[14],在处理非线性数据方面具有更大的优势。此外,机器学习对于处理大范围的时空数据也有着传统方法所不具备的优势。目前地质学研究常使用的是机器学习中的监督式学习算法,如随机森林(RF, Random Forest)、支持向量机(SVM, Support Vector Machines)、人工神经网络(ANN, Artificial Neural Networks)和K近邻(KNN, K-Nearest Neighbor)。这4种监督式学习算法与地质学研究融合的趋势已初见成效,在矿物识别^[15-16]、岩石分类^[17-19]、地质填图^[20]、找矿勘探^[21-22]等领域均得到较好的效果。在众多机器学习算法中,随机森林算法基于集成学习的思想对传统的决策树方法进行优化,其本质上是一个整合了多个弱分类器的强分类器^[23-24]。由于该方法能有效提升决策树的泛化性能,很好地处理具有高维特征的输入样本,并可评估各个特征在分类和参数预测等问题上的重要性,且不需要进行大量的参数调试工作,具有较高的准确率和迁移学习能力。

本文利用机器学习算法,利用现有的中国东部海域表层沉积物有机碳数据,结合样品所处海域的沉积物粒度、水深、地形、溶解氧、盐度、悬浮颗粒有机碳(POC)、沉积速率等多种可能影响有机碳分布的各种环境属性作为“特征”,构建特征和有机碳含量之间的映射关系,从而预测未知海区表层沉积物中有机碳的含量。在此基础上,通过与已发表相应海区表层沉积物有机碳分布图进行比对,分析随机森林算法与数学插值算法预测结果的差异,探讨随机森林算法在沉积物有机碳含量预测上的优势。尝试用更少的数据量获得与实际更相符的预测结果,绘制较高精度的中国东部海区表层沉积物有机碳分布特征图,极大减少人力、物力以及资源消耗,还为未来基于该模型对不同元素和对象的分布进行预测提供了研究思路,具有重要的科学意义和广泛的应用潜力。

1 材料与方法

1.1 研究区概况

中国东部海域包括渤海、黄海、东海及周边海区。其中, 渤海是中国的半封闭内海, 由辽东湾、渤海湾、莱州湾、中央浅海盆地和渤海海峡 5 个海区组成, 仅由东面的渤海海峡与黄海相通。渤海周边入海河流有黄河、海河、滦河和辽河等。黄海为西太平洋边缘岛弧与亚洲大陆之间的边缘海, 位于中朝准地台和扬子准地台两大构造单元之间, 黄海地形比较复杂, 沉积物受到潮汐、黄海暖流和沿岸流等影响, 沉积环境多变。东海是西太平洋典型的开放型边缘海, 地形平坦开阔, 接纳了长江和黄河等入海河流携带的大量陆源碎屑物质。末次冰盛期, 大部分陆架暴露成陆; 冰后期海侵期间, 在太平洋潮流的强烈作用下, 陆架沉积物被冲刷改造, 形成广泛分布的潮流沙脊和浅滩^[25-26]。

1.2 样品来源与资料收集

本文系统收集整理了国内科研院所近年来在中国东部海域采集并测试的 405 个表层沉积物有机碳含量数据(图 1)。样品用去离子水洗盐后, 置于烘箱内 60 °C 烘干, 准确称重后锡纸包样, 使用有机元素分析仪进行有机碳测试。不同样品的有机碳测试精度都在 <0.05% 标准偏差范围内。本文还收集了样品点位置处的相关海洋、水文、生物、化学和地质属性作为特征, 供机器学习算法开展后期学习和预测。用于机器学习的相关特征主要来自 Graw 等^[27] 提供的 49 个相关环境属性。各个特征均以 5×5 弧分的分辨率进行网格化处理, 涵盖了渤海的全部, 黄海和东海的大部分海域。所收集的特征包括海拔高度、沉积物厚度、海底孔隙度、含氧量、温度等(表 1)。此外, 本文还根据已发布的中国东部海域沉积物粒度属性分布图^[28], 通过数字化提取了图中沉积物粒度和岩性分类的相关资料, 作为一个重要的地质属性特征和前文中提到的 49 个特征一起共计 50 个“特征”参与机器学习和预测。

1.3 随机森林算法

随机森林算法(Random Forest)是由 Breiman^[24]提出的一种以决策树为基础构建 Bagging 框架的集成算法, 利用多个决策树组合进行回归分析。其中, 每个决策树逐次分裂形成节点, 每个节点既是

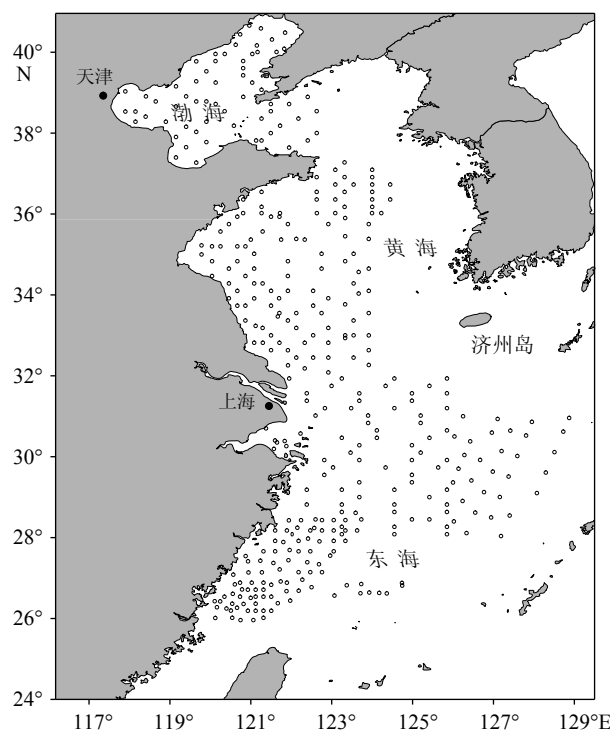


图 1 沉积物有机碳样品分布

Fig.1 Distribution of organic carbon in sediment samples

单个特征组成的一个预测器, 决定着预测走向, 同时在节点上做出决策并进一步分裂, 从而引入两个新的预测器, 形成两个新节点, 最终分裂停止留下叶节点。通过构建大量的决策树形成随机森林, 随机森林的预测结果为各个决策树结果的均值。相比于其他机器学习方法, 随机森林具有原理简单、易于使用即参数较少的特点。在 scikit-learn 的随机森林库中, 该算法的参数主要包括 bagging 框架参数中的决策树棵数(n_estimators)、决策树参数中的最大特征数(max_features)。决策树棵数太少容易欠拟合, 影响模型的预测效果, 增多决策树数量能提升模型预测准确率但同时会增加计算量, 而且当其增大到一定的数量后, 模型性能提升会很小, 模型泛化效果降低。

2 特征选择

预测模型训练和预测本身使用的特征或预测变量并不总是与观测数据相关, 也就是说本文选择的 50 个环境属性“特征”可能和表层沉积物有机质含量密切相关, 也可能是无关的。总体来说这些特征并不是越多越好, 有些无关特征可能会干扰模型的预测性能, 因此本文采用单变量特征选择方法对输入预测器网格执行特征选择^[29]。首先使用单一网格的特征利用随机森林算法进行预测, 通过决定

表 1 用于执行初始预测的 50 个环境属性特征

Table 1 The 50 environmental attribute features used for initial prediction

	预测特征	特征名称
1	GL_DIST_TO_COAST_KM_ETOPO.5m.ggg	离海岸距离
2	GL_ELEVATION_M_ASLSRTM15+V2.5m.ggg	海拔
3	GL_LATITUDE_DD.5m.ggg	纬度
4	GL_LONGITUDE.5m.ggg	经度
5	GL_RIVERMOUTH_CO2_TGCRYR-1_ORNL.5m.ggg	河口CO ₂
6	GL_RIVERMOUTH_DOC_TGCRYR-1_ORNL.5m.ggg	河口DOC
7	GL_RIVERMOUTH_HCO3_TGCRYR-1_ORNL.5m.ggg	河口HCO ₃ ⁻
8	GL_RIVERMOUTH_POC_TGCRYR-1_ORNL.5m.ggg	河口POC
9	GL_RIVERMOUTH_TSS_TGYR-1_ORNL.5m.ggg	河口TSS
10	GL_TOT_SED_THICK_M_CRUST1_NOAA.5m.ggg	沉积物厚度（据CRUST1模型）
11	GL_TOT_SED_THICK_M_GLOBSED_Straume.5m.ggg	沉积物厚度（据GLOBSED模型）
12	SF_AVG_SEA_DENSITY_KGM3_DECADAL_MEAN_woa13x.5m.ggg	平均海水密度（十年平均）
13	SF_AVG_SEA_SOUNDSPPEED_MS_DECADAL_MEAN_woa13x.5m.ggg	平均海水声波速度（十年平均）
14	SF_CALSIL_FRAC_DUTKIEWICZ.5m.ggg	硅酸钙盐含量
15	SF_CLAYFRACTION_FRAC_NGDC.5m.ggg	黏土含量
16	SF_CURRENT_EAST_MS_2012_12_HYCOMx.5m.ggg	东向洋流的速度
17	SF_CURRENT_NORTH_MS_2012_12_HYCOMx.5m.ggg	北向洋流的速度
18	SF_GRAINSIZE_D16_MM_NGDC.5m.ggg	D16粒径
19	SF_GRAINSIZE_D50_MM_NGDC.5m.ggg	D50粒径
20	SF_GRAINSIZE_D84_MM_NGDC.5m.ggg	D84粒径
21	SF_SEA_CONDUCTIVITY_SM_DECADAL_MEAN_woa13v2x.5m.ggg	海水电导率（十年平均）
22	SF_SEA_NITRATE_MCML_DECADAL_MEAN_woa13v2x.5m.ggg	海水硝酸盐浓度
23	SF_SEA_OXYGEN_MLL_DECADAL_MEAN_woa13v2x.5m.ggg	海水溶解氧浓度
24	SF_SEA_OXYGEN_PCTSAT_DECADAL_MEAN_woa13v2x.5m.ggg	海水溶解氧饱和度
25	SF_SEA_PHOSPHATE_MCML_DECADAL_MEAN_woa13v2x.5m.ggg	海水磷酸盐浓度
26	SF_SEA_SALINITY_PSU_DECADAL_MEAN_woa13v2x.5m.ggg	海水盐度
27	SF_SEA_SEA_OXYGEN_UTILIZATION_MOLM3_DECADAL_MEAN_woa13v2x.5m.ggg	海水氧利用率
28	SF_SEA_SILICATE_MCML_DECADAL_MEAN_woa13v2x.5m.ggg	海水硅酸盐浓度
29	SF_SEA_TEMPERATURE_C_DECADAL_MEAN_woa13v2x.5m.ggg	海水温度
30	SF_TERBIO_FRAC_DUTKIEWICZ.5m.ggg	陆源生物硅
31	SF_TOC_PDW.5m.ggg	总有机碳
32	SL_GEOID_M_ABOVE_WGS84_NGA_egm2008.5m.ggg	大地水准面
33	SS_BIOMASS_BACTERIA_LOG10_MGCM2_WEI2010x.5m.ggg	细菌生物量
34	SS_BIOMASS_FISH_LOG10_MGCM2_WEI2010x.5m.ggg	鱼类生物量
35	SS_BIOMASS_INVERTEBRATE_LOG10_MGCM2_WEI2010x.5m.ggg	无脊椎动物生物量
36	SS_BIOMASS_MACROFAUNA_LOG10_MGCM2_WEI2010x.5m.ggg	大型动物生物量

续表 1

预测特征		特征名称
37	SS_BIOMASS_MEGAFAUNA_LOG10_MGCM2_WEI2010x.5m.ggg	巨型动物生物量
38	SS_BIOMASS_MEIOFAUNA_LOG10_MGCM2_WEI2010x.5m.ggg	较小型底栖生物生物量
39	SS_BIOMASS_TOTAL_LOG10_MGCM2_WEI2010x.5m.ggg	总生物量
40	SS_CHLOROPHYLL_LOG_MG_M3_MODIS_Aqua_MISSION_MEANx.5m.ggg	叶绿素
41	SS_DENSITY_KGM-3_SACD_Aquarius_MISSION_MEANx.5m.ggg	平均海水密度
42	SS_PHOTO_AVAIL_RAD_EINSTEIN_M2_DAY_SNPP_VIIRS_MISSION_MEANx.5m.ggg	可利用光照
43	SS_PHYTO_ABSORPTION_443NM_M-1_SNPP_VIIRS_MISSION_MEANx.5m.ggg	光吸收量
44	SS_PIC_LOG_MOL_M3-1_MODIS_Aqua_MISSION_MEANx.5m.ggg	PIC
45	SS_POC_LOG_MOL_M3-1_MODIS_Aqua_MISSION_MEANx.5m.ggg	POC
46	SS_WAVE_DIRECTION_DEG_2012_12_WAVEWATCH3x.5m.ggg	波向
47	SS_WAVE_HEIGHT_M_2012_12_WAVEWATCH3x.5m.ggg	波高
48	SS_WAVE_PERIOD_S_2012_12_WAVEWATCH3x.5m.ggg	波周期
49	SS_WINDSPEED_MS-1_SACD_Aquarius_MISSION_MEANx.5m.ggg	风速
50	Sediment type	沉积物类型

注：表中加粗的变量表示经过特征选择方法筛选出的38个重要环境特征。

系数 R^2 来指示预测效果, 然后对预测效果进行排序 (图 2)。这里决定系数 R^2 是一种用于衡量回归模型预测效果的统计指标, 表示模型解释目标变量总变异的程度, R^2 越接近于 1, 表示模型的预测越接近真实值。此外, 在特征选择过程中引入均匀随机噪声的预测器网格, 假设 R^2 在噪声网格线之下的特征对于模型没有任何提升。筛选掉这部分特征后, 本文在原有 50 个环境属性特征中筛选获取出了 41 个 R^2 高于随机噪声的特征。

为了进一步明确需要多少个环境属性特征才能进行有效的预测, 本文对误差小于均匀随机噪声的 41 个特征进行了集成排序 (Ensemble ranking)。以增量的形式逐步增加预测中使用的预测特征数量, 从 R^2 最高的特征开始, 迭代增加特征数量, 并计算出相关特征的 R^2 (图 2)。由图 2 可知, 在特征的数目达到 8 个的时候, 模型的 R^2 增速放缓, 已经开始稳定; 在特征数目达到 19 个的时候, R^2 首次达到最高值 0.6。之后从 20 个预测变量的集合开始, 集合排名显示预测误差趋于平稳, 在第 38 个特征组成的集合处 R^2 再次达到最高值 0.6。虽然特征数目为 19 个的时候同样达到 R^2 的最高值, 但是机器学习模型在样本少、特征维度低的时候建立的联系过于简单, 无约束收敛到局部最小值的优化过程会导致模型泛化能力有限, 因此本文选择 38 个特征来执行沉积物有机碳预测。

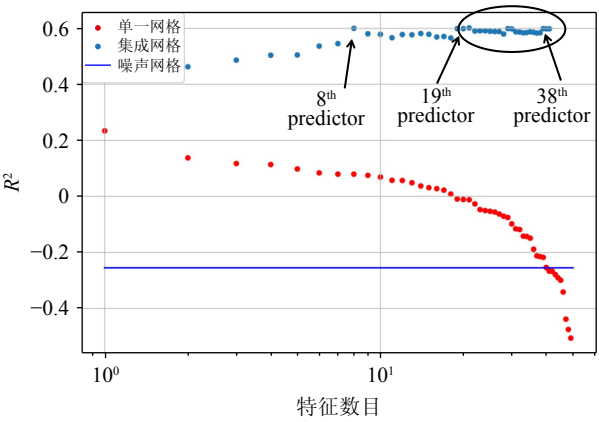


图 2 单一网格特征的 R^2 排名 (红点) 与集成网格的 R^2 排名 (蓝点)
蓝色直线代表随机噪声网格的 R^2 。

Fig.2 The R^2 ranking of single-grid features (red dots) and that of integrated grid (blue dots)
The blue line represents the R^2 of the random noise grid.

3 交叉验证

为了评估模型的性能, 本文对于机器学习训练的结果和克里金插值的结果分别进行了十折交叉验证。在十折交叉验证中, 数据集被均匀划分为 10 个子集, 每次使用其中 9 个子集进行训练, 剩余的 1 个子集作为验证集进行评估。该过程重复进

行 10 次, 每个子集都轮流作为验证集。最终, 所有 10 次评估结果的平均值作为模型的性能指标。十折交叉验证能够有效降低过拟合风险, 确保每个数据点都参与训练和验证, 最大程度地利用数据集。此外, 通过多次独立的评估使得性能评估更加稳定和可靠, 减少了数据划分带来的偶然性和偏差, 因此被广泛应用于各种机器学习模型的评估和选择中。

4 结果

4.1 机器学习计算结果与传统插值结果比较

本文机器学习算法的十折交叉验证结果如图 3 所示, 将所有样品点的有机碳预测值和样品有机碳实际测试值进行对比, 利用最小二乘法对数据点拟合, 机器学习算法的决定系数 R^2 为 0.6。该 R^2 结果在机器学习领域中处于较高的水平。

克里金插值方法得到的结果与随机森林算法

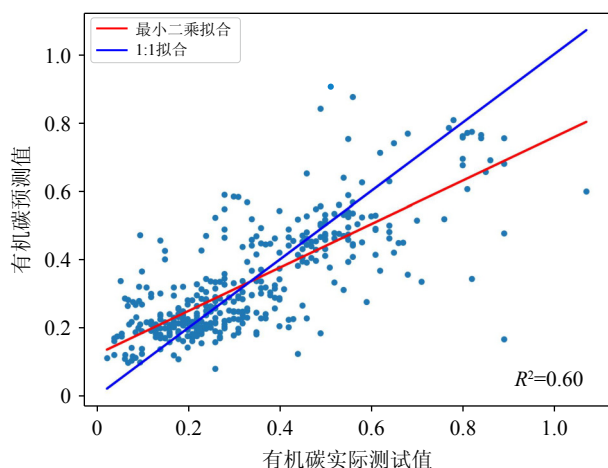


图 3 模型的交叉验证结果

Fig.3 The Cross-validation results of the model

预测的结果如图 4 所示。根据决定系数 R^2 , 绝对中位差 (Median Absolute Deviation, MAD)、平均绝对误差 (Mean Absolute Error, MAE)、均方根误差 (Root Mean Square Error, RMSE)、最大残差 (Max Error) 等误差评价指标分析, 本文对两种方法的预测结果与观测值之间的误差进行了比较。结果发现, 随机森林算法对有机碳含量的预测误差均小于传统克里金插值方法, 误差指数 (MAD、MAE、RMSE 和 Max Error) 分别下降了 11.7%、17.5%、19.9%、28.5%; 同时随机森林算法结果具有更高的决定系数 R^2 。

本文绘制了中国东部海域表层沉积物有机碳含量分布的预测结果 (图 5a、b), 具体包括基于随机森林算法预测的有机碳分布图 (图 5a) 和基于克里金插值方法计算的有机碳分布图 (图 5b)。通过对比图 5a 和 5b 可以看出基于两种不同算法绘制的有机碳含量分布特征总体上一致, 含量在 0~2.12% 之间波动, 平均值约为 $0.47\% \pm 0.26\%$ 。有机碳高含量区主要分布在渤海泥质区、北黄海中部、黄海中部泥质区、黄海东南部泥质区、济州岛西南泥质区、闽江与瓯江三角洲及邻近海域、台湾岛东北和西南部以及冲绳海槽中北部等区域。有机碳含量低值区主要分布在辽东湾、渤海海峡、黄海北部、长江口东北部、南黄海西南部以及东海中外陆架等区域, 其中以东海中外陆架砂质区和东海南部的砂质区有机碳含量最低。

4.2 基于两种不同算法绘制的有机碳分布图差异

虽然克里金插值算法与随机森林算法在采样点所覆盖的区域表层沉积物有机碳含量空间分布趋势与前人研究^[6-8]基本一致, 但在采样点未覆盖的部分区域存在明显的不同, 主要分布在图 5 北黄海中东部 (A 区)、济州岛周边海域 (B 区) 及冲绳海槽

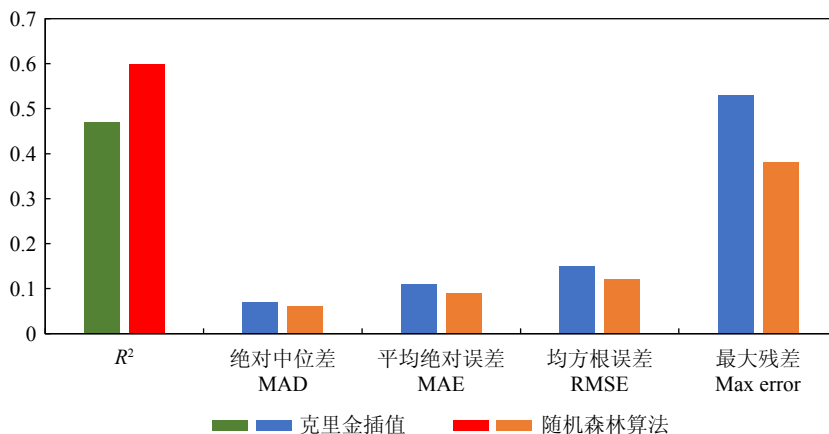


图 4 机器学习算法与克里金插值预测精度比较

Fig.4 Comparison in prediction accuracy between machine-learning algorithms and kriging interpolation

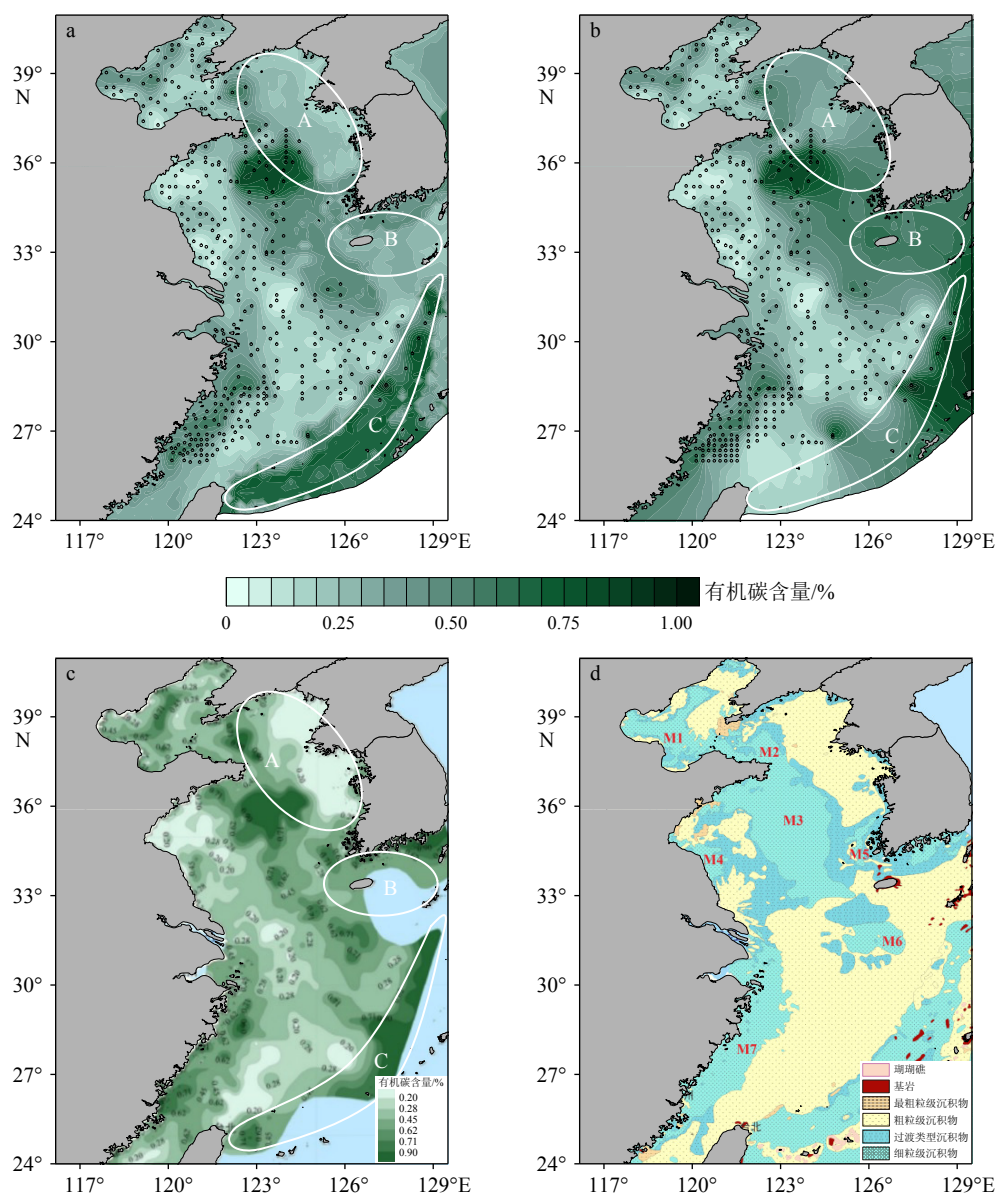


图 5 中国东部海域表层沉积物有机碳含量空间分布

a: 随机森林算法预测有机碳分布图, b: 克里金空间插值算法预测有机碳分布图, c: 实测有机碳含量分布图^[6], d: 沉积物类型分布图^[28]。

Fig.5 Spatial distribution of organic carbon content in surface sediments of the eastern China seas

a: Organic carbon distribution predicted by the random forest algorithm, b: organic carbon distribution predicted by the Kriging spatial interpolation algorithm, c: measured organic carbon content distribution^[6], d: sediment type distribution^[28].

(C 区)等区域。

对于 A 区海域, 图 5a 显示该区域的有机碳含量为 0.2%~0.3%(平均值 0.23%), 特别是靠近朝鲜半岛西侧的区域有机碳含量最低。从分布趋势看, 由朝鲜半岛西侧海岸向北黄海内部的沉积物有机碳含量递增, 区域内的梯度变化明显, 平行海岸呈条带状分布。图 5b 中 A 区域的有机碳含量为 0.25%~0.5%(平均值 0.38%), 比图 5a 中 A 区平均值略高, 在 A 区中心附近最低, 由 A 区域中心向四周逐渐扩散增加。

对于 B 区海域, 图 5a 显示该区域的有机碳含量为 0.2%~0.5%(平均值 0.35%), 整体有机碳含量从北部韩国近岸向南逐渐降低。图 5b 中 B 区有机碳含量比图 5a 中 B 区有机碳含量高, 约为 0.25%~0.7%(平均值 0.5%)。从分布趋势看, 该区域的有机碳空间分布差异不大, 靠近济州岛西北部相对有机碳含量最高。

对于 C 区海域, 图 5a 显示该区域有机碳含量约为 0.5%~0.9%(平均值 0.72%), 有机碳高值分布与冲绳海槽走向一致, 呈新月状分布, 与西侧陆架区

有机碳含量差距较明显。图 5b 显示 C 区域的有机碳含量分布较为离散, 范围为 0.1%~1.0% (平均值 0.45%), 变化范围较大。靠近台湾岛一侧, 有机碳含量较低。冲绳海槽北部有机碳在个别点处含量较高, 向周围呈辐射状降低。

综合沉积物有机碳在北黄海中东部(A)、济州岛附近海域(B)和冲绳海槽区域(C)3个海区的分布特点, 通过机器学习算法计算的有机碳含量在各区域分布普遍表现为相对均匀, 局部区域的梯度变化幅度较小(图 5a)。而在通过克里金插值算法计算的有机碳含量分布中, 上述3个海域沉积物有机碳含量相对较高, 且在区域内变化大, 尤其是在冲绳海槽区域梯度变化显著(图 5b)。

5 讨论

5.1 利用随机森林算法预测有机碳分布的优势

通过对比随机森林算法和克里金插值方法的各项误差评价指标(图 4)可以明显看出, 随机森林算法在预测精度方面具有更好的表现。随机森林算法获得的 R^2 高于克里金插值方法得到的 R^2 , 表明随机森林模型能够更准确地模拟有机碳含量在中国东部边缘海的分布规律。高的 R^2 值反映了模型对真实值的拟合程度较高, 说明随机森林算法能够更好地总结和自主学习现有沉积物数据的规律, 找到环境属性特征与沉积物有机碳含量的内在联系, 从而对样品空白海域进行有效的预测。在绝对中位差 MAD、平均绝对误差 MAE 和均方根误差 RMSE 等误差判断指标上, 随机森林算法的误差值均显著低于克里金插值方法, 表明随机森林模型的预测值与真实值之间的偏差更小, 预测结果更加精确和稳定。尤其是均方根误差(RMSE)的显著下降, 显示了随机森林算法在应对复杂的有机碳含量分布时具有更好的鲁棒性和适应性。此外, 在最大残差这一极端误差指标上, 随机森林算法相比克里金插值方法降低了 28.5%。这表明随机森林算法能够有效减少预测中出现的大偏差值, 从而显著提升预测结果的可靠性。通过以上各种分析表明, 随机森林算法作为一种非参数机器学习模型, 在预测中国东部海域表层沉积物有机碳含量的过程中比传统的克里金插值法展示出了明显的优势。

随机森林算法和克里金插值方法在样品点所覆盖的区域对表层沉积物有机碳的预测结果基本一致, 但在采样点未覆盖区域存在明显的不同。例

如朝鲜半岛西侧-北黄海海域(A区)、济州岛周边海域(B区)等区域, 机器学习预测的有机碳含量普遍小于克里金插值方法的预测结果。而在冲绳海槽海域(C区), 机器学习算法预测的有机碳含量又高于克里金插值的预测。石学法等^[6]利用采自渤海、黄海和东海海域 5796 个站位的沉积物有机碳含量, 系统编制了中国东部海域表层沉积物有机碳分布图(图 5c)。该图件数据量大, 样品分布范围广、精度高, 是目前已发表的分辨率最高的中国东部海域沉积物有机碳分布图。以该图为“标准”, 本文将随机森林算法和克里金插值算法两种预测结果与石学法等^[6]的有机碳分布图进行对照。

对于 A 区, 前人的研究(图 5c)表明, 该区域的有机碳含量较低, 平均值在 0.2% 左右, 整体分布较均一。该特征与随机森林算法的预测结果(图 5a)更接近, 而克里金插值的预测结果(图 5b)普遍高于石学法等^[6]的插值结果, 且分布特征也与石学法等^[6]的插值分布有较大区别。对于 B 区, 石学法等^[6]插值计算的有机碳含量分布图(图 5c)存在一部分空白区域。但总的来说, 越靠近北部的朝鲜半岛, 有机质含量越高, 并且明显高于周边海域。经过对比发现随机森林计算的有机碳预测(图 5a)的分布规律更接近石学法等^[6]的插值结果。而克里金插值方法预测的有机碳含量(图 5b)在 B 区域未体现出明显的由近岸向远海的变化规律。对于 C 区海域, 石学法等^[6]实测的有机碳含量分布图(图 5c)显示, 该区域的有机碳含量整体偏高, 随机森林预测(图 5a)该区域有机碳含量显著高于陆架地区, 更接近冲绳海槽有机碳含量高达 0.9% 以上的事实。随机森林预测结果显示有机碳的分布更广, 且与冲绳海槽长条状弧形地势一致。而对于克里金插值的结果(图 5b), 在冲绳海槽区域表现出较大的空间异质性, 明显与石学法等^[6]研究结果不符。

通过以上与石学法等^[6]中国东部边缘海有机碳分布图的比较, 发现根据随机森林算法计算得到的有机碳预测结果较克里金空间插值方法的预测结果更好, 误差更小, 预测精度也更高。中国东部边缘海表层沉积物有机碳的分布受到周边河流沉积物输入、水深、地形、洋流等因素的共同影响, 不同区域的沉积物有机碳空间分布上可能存在较大的差异。传统数学插值算法, 如克里金插值, 并没有考虑实际海洋沉积环境的影响, 而是通过数值差值计算对空白区域进行填图和预测。相反, 随机森林算法所建立的模型是根据有机碳的真实分布数据来决定的, 在充分考虑沉积物自身特点和周边环境

变量(如粒径、地形、水动力等)的基础上构建有机碳分布与环境特征的定量映射关系,显著提升了预测的科学性,有效避免有机碳含量在预测过程中存在的“高值低估,低值高估”问题。克里金插值法基于空间自相关特性,适用于样品数据点较密集的区域,但在空间异质性高或样品点稀疏的区域,预测结果容易偏离实际分布。随机森林算法通过综合环境参数克服了这一局限性,因此对局部海区分布细节有明显的改善,空间分辨率更高,能清晰呈现其变化分布趋势的同时,细节信息更加突出。此外,对于采样无法或难以覆盖的区域,如朝鲜半岛西侧或冲绳海槽附近,随机森林算法的预测则体现出了巨大的优势,除较为快捷高效,还具有更加准确的预测能力和优秀的外推性能。

5.2 影响有机碳分布的环境参数特征分析

本文在环境属性特征选择过程中依据单一特征排序选择了前 41 个特征进行预测,降维后特征向量的维度为 38。特征选择的主要依据都是基于随机森林算法筛选的,结果显示,在参与随机森林计算的特征中,“沉积物类型”的相关性占比最高(0.40),其次是“海水温度”(0.07)和“POC”(0.07)。沉积物有机碳的含量受沉积物粒度的控制已经成为共识,即有机质更容易赋存在细颗粒的黏土组分中。本文的预测结果显示有机碳含量的高值区分布趋势与细粒沉积区分布基本一致,即有机质在泥质区含量相对较高,而在渤海东部与东北部、北黄海中东部、苏北浅滩和东海中外陆架等砂质区普遍较低。对比石学法等^[28]绘制的中国东部边缘海沉积物类型分布(图 5d),基于随机森林算法所预测的有机碳高值区(图 5a)与渤海泥质区(M1)、山东半岛东部沿岸泥质区(M2)、黄海中部泥质区(M3)、黄海东南部泥质区(M5)、济州岛西南泥质区(M6)和浙闽沿岸泥质区(M7)基本一一对应。

除沉积物粒度特征外,沉积物采样点处表层水体中多年平均颗粒有机碳含量也与沉积物有机质含量有一定关系。中国东部边缘海受周边河流影响显著,由河流携带的陆源有机质在河口及邻近海域大量沉积,如长江口、闽江口和黄河口地区^[30],这些区域的颗粒有机碳含量高,导致表层沉积物中的有机碳含量也相应较高。另一方面,由海洋初级生产力产生的有机质在上升流和营养盐丰富的区域富集,如冲绳海槽和济州岛附近,也会造成沉积物中有机质的富集^[31]。

为了全面了解沉积物有机质分布的影响因素,

在随机森林计算中共有 38 个特征参与了模型的筛选。而一些意想不到的特征,如海水温度、可利用光照、海水电导率等特征对有机碳含量分布的贡献程度也很高,这表明它们与沉积物有机碳的分布也存在某种联系。例如温度和有机质含量的相关性占比是 0.07。温度通过影响有机质的分解速率和保存效率间接控制有机碳的分布,在温暖的浅水区域,有机质分解速率较快,导致有机碳保存效率较低^[32]。沉积物中有机碳含量相对较低,这种趋势在长江口东北部和东海南部的浅水砂质区域尤为显著^[11,33]。在冷水区域,较低的温度减缓了有机质的分解速率,有利于有机碳的保存。例如,北黄海中部和冲绳海槽等区域,较低的水温对有机碳的保存具有一定的促进作用。此外,温度还与海洋环流密切相关,不同温度的水团对沉积物输运和堆积过程也有显著影响,从而间接影响有机碳的空间分布。

除以上单独预测中排名较高的几个特征外,其他一些特征对有机质分布也起到一定的贡献^[34-35]。盐度的变化直接影响水体密度和分层,控制有机碳的垂直输送和分布^[36-37]。通过调控浮游植物、微生物等生物群落的生长和代谢,改变有机碳的生成、降解及转移模式^[38-39]。电导率在河口和沿海区域的变化影响咸淡水交汇处的有机碳混凝和沉降过程。此外,高盐度环境下的静电效应可能加速颗粒有机碳的凝聚和沉降,同时溶解有机碳的稳定性也随之变化。由此可见,机器学习算法可以发现特征之间一些不可预见的相关性,而这些关系可能在一般认识和传统研究中被忽视,这也正充分体现了机器学习在数据挖掘和规律总结中的优势。

5.3 应用随机森林算法预测中国东部边缘海有机碳特征分布的应用前景

本文初步建立了一个应用随机森林算法预测中国东部海域表层沉积物有机碳含量分布的基础模型。通过对 405 个沉积物样品有机碳含量的学习,随机森林算法预测的有机碳分布图比根据同样数量样品由克里金插值计算绘制的有机碳分布图更加准确,甚至十分接近石学法等^[6]根据 5 796 个样品由克里金插值计算绘制的有机碳分布图。而对于样品空白区域的填充,随机森林算法则展示了远超克里金插值算法的巨大优势。

未来如果需要进一步提升预测精度,需要将更多更丰富的地质特征添加到原始数据集,并重新运行机器学习算法完成优化。该过程不仅简单高效,同时也为模型的动态改进和可持续应用提供了极

大的灵活性。在模型优化过程中,添加更多的样本数据(有机碳含量)是关键,特别是在无样品区域较广或数据稀缺的地区,如地质特性复杂或当前观测不足的海域,通过引入更多、更精确的环境属性特征将有助于更全面地表征影响有机碳分布的海洋水文、环境和地质因素,从而进一步提升预测结果的科学性和准确性,弥补这些调查区域的数据空缺。

此外,目前所使用的模型输入特征数据(如温度、盐度、粒度等)大多为长期平均值(如年均值或多年均值),这与表层沉积物有机碳采样的时间不完全对应。地质特征的长期平均值可能会掩盖短时间内指标变化对有机碳保存的实际影响,特别是在季节性差异显著的区域,可能会引入一定的不确定性。未来的研究应重点关注数据稀疏区域的采样工作,以及优化和扩展现有环境属性特征的数量与精度。通过结合高分辨率、高时效性的地质特征信息和更加全面的样品数据,基于时间序列分析的方法,构建时间动态模型,将沉积物有机碳分布与时间维度上的特征变化相结合,可以在区域或全球尺度上进一步提升模型的预测能力,为海洋资源调查和海洋环境保护提供更可靠的科学支撑。

6 结论

(1)通过对405个沉积物样品有机碳含量的学习,本文根据随机森林算法预测并绘制了中国东部边缘海表层沉积物的有机碳分布图。与根据同样数量样品由克里金插值计算绘制的有机碳分布图相比,随机森林算法对沉积物有机碳含量预测结果的平均绝对误差、均方根误差、最大残差等评价指标均更小,十折交叉检验的 R^2 达到0.60,表现出更高的拟合精度和更小的误差。

(2)通过与石学法等^[6]绘制的中国东部边缘海沉积物有机质含量比较,随机森林算法所预测的表层沉积物有机碳空间分布特征效果非常理想,尤其对于一些样品欠缺或空白的海区,随机森林算法的预测将体现出更大的潜力和优势。随机森林预测是由数据驱动的,能够基于少数的现场测量数据直接进行较为精确的估计。本文应用的随机森林算法对于未来其他海洋沉积物地球化学指标的预测也有重要的借鉴意义;对于中国东部边缘海的资源调查和环境保护具有重要的现实意义。

致谢: 感谢同济大学海洋与地球科学学院蔡佳宁、黄柯华和邓家榕对本文撰写提供的帮助。

参考文献 (References)

- [1] Mayer L M. Sedimentary organic matter preservation: an assessment and speculative synthesis-a comment[J]. *Marine Chemistry*, 1995, 49(2-3): 123-126.
- [2] Battle M, Bender M L, Tans P P, et al. Global carbon sinks and their variability inferred from atmospheric O₂ and $\delta^{13}\text{C}$ [J]. *Science*, 2000, 287(5462): 2467-2470.
- [3] Wright L D, Wiseman W J, Bornhold B D, et al. Marine dispersal and deposition of Yellow River silts by gravity-driven underflows[J]. *Nature*, 1988, 332(6165): 629-632.
- [4] Yang Z S, Liu J P. A unique Yellow River-derived distal subaqueous delta in the Yellow Sea[J]. *Marine Geology*, 2007, 240(1-4): 169-176.
- [5] 石学法, 胡利民, 乔淑卿, 等. 中国东部陆架海沉积有机碳研究进展: 来源、输运与埋藏 [J]. 海洋科学进展, 2016, 34(3): 313-327. [SHI Xuefa, HU Limin, QIAO Shuqing, et al. Progress in research of sedimentary organic carbon in the East China Sea: sources, dispersal and sequestration[J]. *Advances in Marine Science*, 2016, 34(3): 313-327.]
- [6] 石学法, 吴斌, 乔淑卿, 等. 中国东部近海沉积有机碳的分布、埋藏及碳汇效应 [J]. 中国科学: 地球科学, 2024, 54(10): 3113-3133. [SHI Xuefa, WU Bin, QIAO Shuqing, et al. Distribution, burial fluxes and carbon sink effect of sedimentary organic carbon in the eastern China seas[J]. *Science China: Earth Sciences*, 2024, 67(10): 3062-3082.]
- [7] 赵美训, 丁杨, 于蒙. 中国边缘海沉积有机质来源及其碳汇意义 [J]. 中国海洋大学学报, 2017, 47(9): 70-76. [ZHAO Meixun, DING Yang, YU Meng. Sources of sedimentary organic matter in China marginal sea surface sediments and implications of carbon sink[J]. *Periodical of Ocean University of China*, 2017, 47(9): 70-76.]
- [8] 焦念志, 梁彦韬, 张永雨, 等. 中国海及邻近区域碳库与通量综合分析 [J]. 中国科学: 地球科学, 2018, 48(11): 1393-1421. [JIAO Nianzhi, LIANG Yantao, ZHANG Yongyu, et al. Carbon pools and fluxes in the China Seas and adjacent oceans[J]. *Science China Earth Sciences*, 2018, 61(11): 1535-1563.]
- [9] Bao R, McIntyre C, Zhao M X, et al. Widespread dispersal and aging of organic carbon in shallow marginal seas[J]. *Geology*, 2016, 44(10): 791-794.
- [10] 李军, 胡邦琦, 窦衍光, 等. 中国东部海域泥质沉积区现代沉积速率及其物源控制效应初探 [J]. 地质论评, 2012, 58(4): 745-756. [LI Jun, HU Bangqi, DOU Yanguang, et al. Modern sedimentation rate, budget and supply of the muddy deposits in the East China Seas[J]. *Geological Review*, 2012, 58(4): 745-756.]
- [11] Deng B, Zhang J, Wu Y. Recent sediment accumulation and carbon burial in the East China Sea[J]. *Global Biogeochemical Cycles*, 2006, 20(3): GB3014.
- [12] Wu Y, Eglinton T, Yang L Y, et al. Spatial variability in the abundance, composition, and age of organic matter in surficial sediments of the East China Sea[J]. *Journal of Geophysical Research: Biogeosciences*, 2013, 118(4): 1495-1507.
- [13] Qiao S Q, Shi X F, Wang G Q, et al. Sediment accumulation and budget in the Bohai Sea, Yellow Sea and East China Sea[J]. *Marine Geology*, 2017, 390: 270-281.

- [14] Liu W, Du P J, Zhao Z W, et al. An adaptive weighting algorithm for interpolating the soil potassium content[J]. *Scientific Reports*, 2016, 6(1): 23889.
- [15] 徐述腾, 周永章. 基于深度学习的镜下矿石矿物的智能识别实验研究[J]. *岩石学报*, 2018, 34(11): 3244-3252. [XU Shuteng, ZHOU Yongzhang. Artificial intelligence identification of ore minerals under microscope based on deep learning algorithm[J]. *Acta Petrologica Sinica*, 2018, 34(11): 3244-3252.]
- [16] 朱紫怡, 周飞, 王瑀, 等. 基于机器学习的锆石成因分类研究[J]. *地学前缘*, 2022, 29(5): 464-475. [ZHU Ziyi, ZHOU Fei, WANG Yu, et al. Machine learning-based approach for zircon classification and genesis determination[J]. *Earth Science Frontiers*, 2022, 29(5): 464-475.]
- [17] Zhao Y, Zhang Y G, Geng M, et al. Involvement of slab-derived fluid in the generation of cenozoic basalts in Northeast China inferred from machine learning[J]. *Geophysical Research Letters*, 2019, 46(10): 5234-5242.
- [18] Guo P, Yang T, Xu W L, et al. Machine learning reveals source compositions of intraplate basaltic rocks[J]. *Geochemistry, Geophysics, Geosystems*, 2021, 22(9): e2021GC009946.
- [19] Gion A M, Piccoli P M, Candela P A. Characterization of biotite and amphibole compositions in granites[J]. *Contributions to Mineralogy and Petrology*, 2022, 177(4): 43.
- [20] Kirkwood C, Cave M, Beamish D, et al. A machine learning approach to geochemical mapping[J]. *Journal of Geochemical Exploration*, 2016, 167: 49-61.
- [21] Rodriguez-Galiano V, Sanchez-Castillo M, Chica-Olmo M, et al. Machine learning predictive models for mineral prospectivity: An evaluation of neural networks, random forest, regression trees and support vector machines[J]. *Ore Geology Reviews*, 2015, 71: 804-818.
- [22] Sun T, Chen F, Zhong L X, et al. GIS-based mineral prospectivity mapping using machine learning methods: A case study from Tongling ore district, eastern China[J]. *Ore Geology Reviews*, 2019, 109: 26-49.
- [23] Breiman L. Bagging predictors[J]. *Machine Learning*, 1996, 24(2): 123-140.
- [24] Breiman L. Random forests[J]. *Machine Learning*, 2001, 45(1): 5-32.
- [25] Berné S, Vagner P, Guichard F O, et al. Pleistocene forced regressions and tidal sand ridges in the East China Sea[J]. *Marine Geology*, 2002, 188(3-4): 293-315.
- [26] Chough S K, Kim J W, Lee S H, et al. High-resolution acoustic characteristics of epicontinental sea deposits, central-eastern Yellow Sea[J]. *Marine Geology*, 2002, 188(3-4): 317-331.
- [27] Graw J H, Wood W T, Phrampus B J. Predicting global marine sediment density using the random forest regressor machine learning algorithm[J]. *Journal of Geophysical Research: Solid Earth*, 2021, 126(1): e2020JB020135.
- [28] 石学法, 刘焱光, 乔淑卿, 等. 渤海、黄海和东海沉积物类型图[M]. 北京: 科学出版社, 2021. [SHI Xuefa, LIU Yanguang, QIAO Shuqing, et al. Sediment Type Map of the Bohai Sea, Yellow Sea and East China Sea[M]. Beijing: Science Press, 2021.]
- [29] Lee T R, Phrampus B J, Obelcz J, et al. Global marine isochore estimates using machine learning[J]. *Geophysical Research Letters*, 2020, 47(18): e2020GL088726.
- [30] Zhao B, Yao P, Bianchi T S, et al. Contrasting controls of particulate organic carbon composition and age from riverine to coastal sediments of Eastern China Marginal Seas[J]. *Chemical Geology*, 2023, 624: 121429.
- [31] 王玥铭, 窦衍光, 李军, 等. 16ka 以来冲绳海槽中南部沉积物物源演化及其对古气候的响应[J]. *沉积学报*, 2018, 36(6): 1157-1168. [WANG Yueying, DOU Yanguang, LI Jun, et al. Sediment provenance change and its response to paleoclimate change in the middle Okinawa Trough since 16 ka[J]. *Acta Sedimentologica Sinica*, 2018, 36(6): 1157-1168.]
- [32] Malinverno A, Martinez E A. The effect of temperature on organic carbon degradation in marine sediments[J]. *Scientific Reports*, 2015, 5: 17861.
- [33] 高建华, 汪亚平, 潘少明, 等. 长江口外海域沉积物中有机物的来源及分布[J]. *地理学报*, 2007, 62(9): 981-991. [GAO Jianhua, WANG Yaping, PAN Shaoming, et al. Source and distribution of organic matter in seabed sediments of the Changjiang River estuary and its adjacent sea area[J]. *Acta Geographica Sinica*, 2007, 62(9): 981-991.]
- [34] Chen S Z, Lou S, Yang Z Y, et al. Distributions and influence factors of organic carbon in coastal area of the Yangtze River Estuary, China[J]. *Estuaries and Coasts*, 2024, 47(8): 2253-2266.
- [35] Zhang S S, Liang C, Xian W W. Spatial and temporal distributions of terrestrial and marine organic matter in the surface sediments of the Yangtze River Estuary[J]. *Continental Shelf Research*, 2020, 203: 104158.
- [36] Helber R W, Richman J G, Barron C N. The influence of temperature and salinity variability on the upper ocean density and mixed layer[J]. *Ocean Science Discussions*, 2010, 7(4): 1469-1495.
- [37] Wang Y J, Jiang C J, Cheng H Q, et al. Influence of the density stratification on the vertical distribution of suspended sediment in a partially mixed estuary[J]. *Marine Geology*, 2025, 481: 107461.
- [38] Masmoudi S, Tastard E, Guermazi W, et al. Salinity gradient and nutrients as major structuring factors of the phytoplankton communities in salt marshes[J]. *Aquatic Ecology*, 2015, 49(1): 1-19.
- [39] D'ors A, Bartolomé M C, Sánchez-Fortún S. Repercussions of salinity changes and osmotic stress in marine phytoplankton species[J]. *Estuarine, Coastal and Shelf Science*, 2016, 175: 169-175.