

吴芳芳, 傅智一, 胡林舒, 等. 基于随机森林方法反演墨西哥湾海表盐度[J]. 海洋学报, 2021, 43(9): 126–136, doi:10.12284/hyxb2021146  
Wu Fangfang, Fu Zhiyi, Hu Linshu, et al. Retrieval of sea surface salinity in the Gulf of Mexico based on random forest method[J]. Haiyang Xuebao, 2021, 43(9): 126–136, doi:10.12284/hyxb2021146

# 基于随机森林方法反演墨西哥湾海表盐度

吴芳芳<sup>1,2</sup>, 傅智一<sup>1,2</sup>, 胡林舒<sup>1,2</sup>, 张丰<sup>1,2\*</sup>, 杜震洪<sup>1,2</sup>, 刘仁义<sup>1,2</sup>

(1. 浙江大学 浙江省资源与环境信息系统重点实验室, 浙江 杭州 310028; 2. 浙江大学 地球科学学院, 浙江 杭州 310027)

**摘要:** 盐度是表征物理和生物地球化学过程的重要参数之一, 光学遥感可满足较高分辨率的监测需要并避免射频频干扰问题, 为沿海水域的海表盐度研究提供可行的途径。本文基于 MODIS-Aqua 的 412 nm、443 nm、488 nm、555 nm 和 667 nm 波段的遥感反射率 (Rrs412、Rrs443、Rrs488、Rrs555、Rrs667)、海表温度以及实测的海表盐度数据构建随机森林模型, 基于模型结果分析墨西哥湾海表盐度时空异质性及海表盐度与影响因子 (海表温度和遥感反射率) 之间的相关关系。研究结果表明: (1) 随机森林模型能较准确地反演墨西哥湾海表盐度, 其均方根误差为 0.335, 决定系数为 0.931; (2) 湾区海表盐度空间分布呈近岸-河口低、离岸高, 环状向内增值的态势, 其变化受河流流量、风力以及环流的影响; (3) 海表温度与海表盐度存在较强的相关性, 海表温度对海表盐度的反演影响显著; (4) 海表温度、遥感反射率与海表盐度的相关性呈现空间异质性。

**关键词:** 海表盐度; 随机森林; 时空异质性; 墨西哥湾

中图分类号: P731.12

文献标志码: A

文章编号: 0253-4193(2021)09-0126-11

## 1 引言

盐度是海水的基本物理参数之一, 可用于追踪陆源物质的路径, 研究陆源河流冲淡水在近海海水混合变化过程、藻类暴发过程以及缺氧区的生态环境等<sup>[1-5]</sup>。传统盐度数据的获取方式包括现场测量获取以及利用浮标获取, 难以实现海表盐度的时空连续观测。卫星遥感凭借其大范围、连续性的时空优势可及时有效地监测海表盐度, 对于海洋的相关研究、海洋资源的管理与决策具有重要意义。

微波遥感在复杂天气条件下具有一定的优势, 常被用于大面积开阔海域的海表盐度反演研究<sup>[6]</sup>, 但其时空分辨率低且易受射频频干扰, 不适用于沿海水域。在沿海水域及近岸-河口地区, 海表盐度的监测与研究通常利用光学遥感。与现有的盐度卫星相比, 光学

卫星具有较高的分辨率与较短的重返周期并可避免射频频干扰的问题。Vandermeulen 等<sup>[7]</sup>利用 SNPP-VIIRS 卫星 410 nm、443 nm、486 nm、551 nm 和 671 nm 波段的遥感反射率 (Remote Sensing Reflectance, Rrs) (Rrs410、Rrs443、Rrs486、Rrs551、Rrs671) 反演了墨西哥湾河口区域的海表盐度。Liu 等<sup>[8]</sup>基于静止轨道卫星 (GOCI) 的遥感反射率 (Rrs412、Rrs443、Rrs490) 反演了渤海的海表盐度。这些研究通常将有色溶解有机物 (Chromophoric Dissolved Organic Matter, CDOM) 的吸收系数作为海表盐度反演的中介参数, 来建立遥感反射率与海表盐度之间的关系, 但仅靠单一的遥感反射率数据并不能得到较好的盐度反演结果。近年来, 也有学者利用遥感反射率数据和海表温度数据反演海表盐度, 并证明海表温度是反演海表盐度的重要参数<sup>[9-10]</sup>。

收稿日期: 2021-02-03; 修订日期: 2021-06-16。

基金项目: 国家重点研发计划重点专项(2018YFB0505000); 国家自然科学基金(41671391, 41922043, 41871287)。

作者简介: 吴芳芳(1994—), 女, 贵州省长顺县人, 从事海洋 GIS 研究。E-mail: wyl19952021@126.com

\* 通信作者: 张丰(1977—), 女, 副教授, 从事海洋 GIS 研究。E-mail: zfcarnation@zju.edu.cn

现有的盐度遥感反演方法多采用数理统计模型, 如比值模型<sup>[11]</sup>、多元线性回归模型<sup>[12]</sup>, 利用遥感反射率数据实现对海表盐度的反演。但基于数理统计模型的方法存在模型反演精度低、应用范围有限的缺点, 无法适用于盐度呈复杂非线性特征或遥感信息间存在时空自相关性<sup>[13]</sup>等的复杂场景。基于机器学习的方法能有效分析出数据内部的关联性, 提高预测精度。Geiger等<sup>[9]</sup>利用MODIS-Aqua的归一化离水辐亮度、海表温度数据以及位置信息建立了反演大西洋中部海表盐度的神经网络模型。Chen等<sup>[14]</sup>基于MODIS和SeaWiFS卫星的遥感反射率数据和海表温度数据, 构建了基于多层感知神经网络(Multilayer Perceptron Neural Network, MPNN)的海表盐度反演模型。虽然基于神经网络模型可以有效地提高盐度的反演精度, 但模型的构建方法复杂, 内部运行机理未知, 难以对变量进行合理地解释。随机森林算法是基于决策树的一种集成学习算法<sup>[15]</sup>, 它可以高效地划分多颗决策树。树中的各节点可用于反映不同区域、不同环境下的盐度状况, 可应用于复杂环境下的盐度反演研究。墨西哥湾受入海河流的淡水输送、环流以及风力的共同影响, 是一个具有复杂水动力环境的典型半封闭海域。另外, 该区域内实测数据丰富且易获取, 因此, 选择墨西哥湾作为研究区域, 构建适用于复杂环境的海表盐度反演模型。

本文选取MODIS-Aqua卫星的遥感反射率数据(Rrs412、Rrs443、Rrs488、Rrs555、Rrs667)和海表温度数据, 协同实测盐度数据, 利用随机森林方法构建墨西哥湾海表盐度反演模型, 反演2018年墨西哥湾的海表盐度。基于海表盐度的反演结果, 研究湾区内海表盐度的时空变化特征, 探究不同空间区域内各影响因子与海表盐度的相关性, 对其空间异质性进行分析, 为复杂环境下的海表盐度反演模型提供模型变量的选取思路, 为其他海域的海表盐度反演研究提供参考。

## 2 研究区概况

墨西哥湾为北美洲大陆东南沿海水域, 湾内主要受北部河流输入的影响, 密西西比河、阿查法拉亚河等众多河流为墨西哥湾输送了大量淡水、泥沙、营养物质和有机物<sup>[16]</sup>。位于墨西哥湾东南部的环流系统是墨西哥湾区环流的主要驱动力, 该环流将加勒比海的高温、高盐水向北输送到墨西哥湾, 再从佛罗里达海峡输出至北大西洋。另外, 环流也将北部的淡水沿着北部大陆架和西佛罗里达大陆架向东和向南输

送<sup>[17-18]</sup>。整体上, 墨西哥湾由于受到入海河流淡水输送、环流和风力等因素的影响, 其物理和生物化学过程活跃, 海水盐度也呈现空间异质性(图1, 参考Otis等<sup>[17]</sup>的研究)。

## 3 数据与方法

### 3.1 数据的收集与处理

#### 3.1.1 实测盐度数据

实测盐度数据来自2018年美国国家海洋和大气管理局(National Oceanic and Atmospheric Administration, NOAA)在墨西哥湾海域开展的15个共享航次, 该数据从海洋碳数据系统(Ocean Carbon Data System, OCADS)(<https://www.nodc.noaa.gov/ocads>)中下载。盐度观测仪器为美国Sea-Bird公司的温盐深仪(CTD), 测量的方式为走航测量。根据世界海洋环流实验(World Ocean Circulation Experiment, WOCE)指南, 数据均被标上质量控制标识, 共有3个标识值:“2”表示质量良好;“3”表示可疑数据;“4”表示质量较差。为保证数据的准确性, 本研究保留具有质量良好标识“2”的实测盐度数据。数据来源的航次信息如表1所示, 空间分布情况如图2所示。

#### 3.1.2 卫星数据

卫星数据来自搭载于Aqua卫星上的中分辨率成像光谱辐射计(MODIS), 该数据具有较高的光谱分辨率和中等空间分辨率, 可满足研究的需求。本文选取MODIS的遥感反射率数据(Rrs412、Rrs443、Rrs488、Rrs555、Rrs667)以及海表温度(Sea Surface Temperature, SST)数据用于反演海表盐度, 具体的统计信息如表2所示。数据来源于美国航天局戈达德航天飞行中心(Goddard Space Flight Center, GSFC), 是空间分辨率约1 km的标准L2级单轨产品。按照美国航天局质量控制标准, 剔除含有云、杂散光等低质量的卫星数据, 剔除质量等级大于1的L2级低质量海表温度影像。基于3×3像元窗口, 获取与实测盐度采样时间间隔为±6 h内的L2级遥感反射率和海表温度数据。利用Arcmap软件的Extract Values to Points工具, 将卫星数据和实测盐度数据进行匹配, 以便进行盐度反演。本文数据经过预处理后, 共获得7 963组匹配数据, 用于海表盐度反演模型的开发与验证。

### 3.2 方法

随机森林是一种用于分类和回归的集成学习方法, 其将Bagging方法、随机子空间方法与决策树方法相结合<sup>[19-21]</sup>, 为解决单棵决策树存在过拟合的瓶颈问题, 集成多颗决策树来提高预测性能。算法融入自

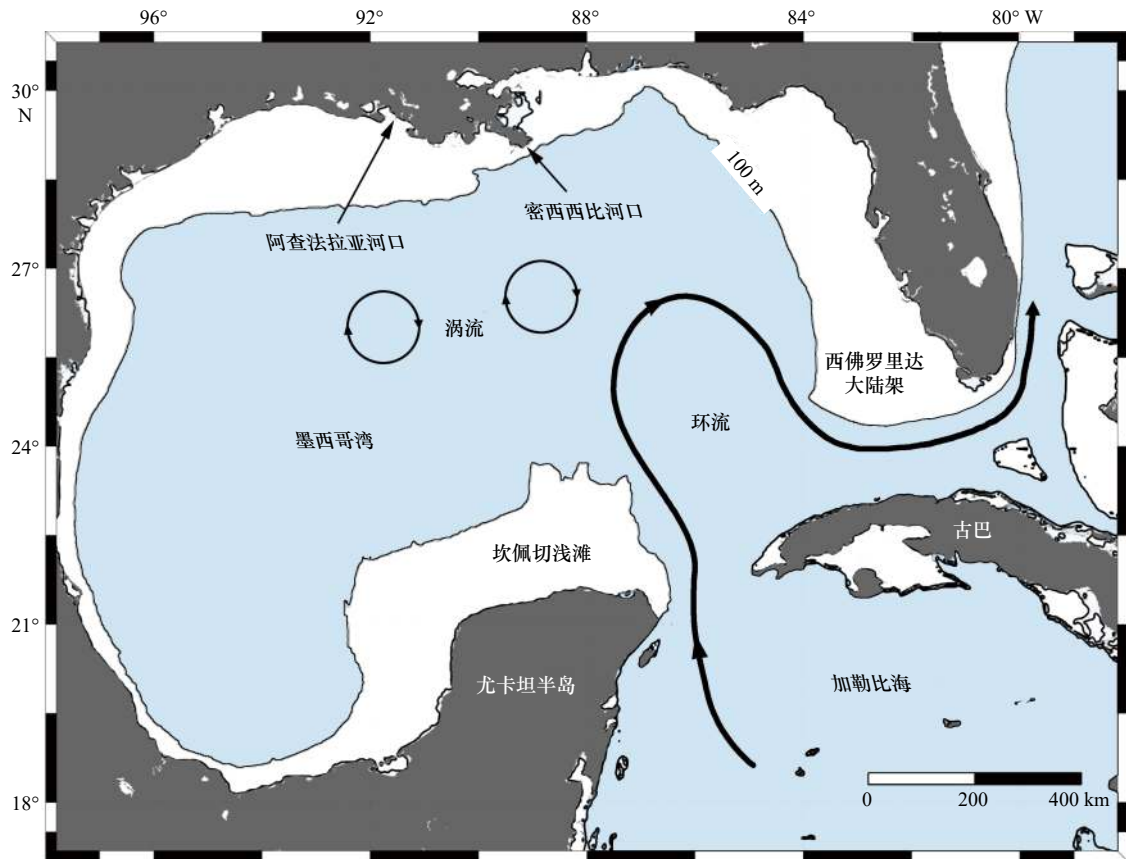


图 1 研究区概况

Fig. 1 Overview of the study area

黑色曲线代表 100 m 等深线; 水深大于 100 m 区域用浅蓝色表示; 黑色带箭头曲线代表环流; 圆环代表环流扩张时引起的涡流  
 The black curve is for the 100 m isobath; areas with water depths greater than 100 m are shaded in light blue; the curve with black arrow is for circulation; the ring is for the eddy current caused by circulation expansion

表 1 实测海表盐度数据来源航次信息

Table 1 The source and voyage information of field sea surface salinity

航次名称	船名	时间范围	观测点个数
EQ17	M/V <i>Celebrity Equinox</i>	2018年1月1-6日	2 179
AS17	M/V <i>Allure of the Seas</i>	2018年1月4-7日	1 198
GU1801_Leg1	R/V <i>Gordon Gunter</i>	2018年1月14-22日	4 178
GU1801_Leg2	R/V <i>Gordon Gunter</i>	2018年1月26日至2月9日	7 421
GU1801_Leg3	R/V <i>Gordon Gunter</i>	2018年2月12-27日	5 428
GU1801_Leg4	R/V <i>Gordon Gunter</i>	2018年3月1-16日	7 941
GU1802	R/V <i>Gordon Gunter</i>	2018年6月24日至7月9日	7 609
GU1803-transit	R/V <i>Gordon Gunter</i>	2018年7月11-14日	1 340
GU1803_Leg1	R/V <i>Gordon Gunter</i>	2018年7月20日至8月3日	7 196
GU1803_Leg2	R/V <i>Gordon Gunter</i>	2018年8月6-19日	4 727
GU1804	R/V <i>Gordon Gunter</i>	2018年8月23-31日	4 445
GU1805_Leg1	R/V <i>Gordon Gunter</i>	2018年9月2-9日	3 563
GU1805_Leg2	R/V <i>Gordon Gunter</i>	2018年9月11-30日	9 656
EQ18	M/V <i>Celebrity Equinox</i>	2018年6月1日至12月22日	872
GU1806	R/V <i>Gordon Gunter</i>	2018年11月10日至1月4日	10 127
观测点总数			77 883
用于模型的开发与验证的点数			7 963

助聚集 (bootstrap aggregating) 方法生成子集, 即从原样本集  $D$  中通过 bootstrap 有放回地随机抽取  $M$  个与原样本集同样大小的训练样本集  $D_m (m=1,2,3,\dots,n)$ , 据此构建多颗决策树。在对决策树的每个节点进行分裂时, 引入随机子空间方法, 从全部  $K$  个特征中均匀随机地抽取一个特征子集, 从该子集中选择一个最优分裂特征<sup>[15]</sup>。最后采用多颗决策树并行的方式, 求

取多棵决策树结果的均值作为最终预测结果, 可简要表示为

$$\hat{f} = \frac{1}{B} \sum_{b=1}^B f_b(x), \quad (1)$$

式中,  $\hat{f}$  表示最终预测结果;  $B$  为决策树数量;  $f_b$  表示回归函数;  $x$  表示训练样本值。

随机森林的优点在于利用 bootstrap 缓解高方差的问题, 弱化决策树间的相关性; 利用随机子空间方法选取特征的随机性优势, 增强模型的泛化能力; 平均多颗决策树预测值提高预测精度, 实现数据的准确分类和预测。随机森林方法简单高效, 只需调整森林中树的数量 (ntree) 和每个结点的特征数 (mtry) 就可以快速有效地生成合理的模型。与其他机器学习方法相比, 随机森林抗过拟合能力强, 异常值和噪声的容忍度高<sup>[15]</sup>, 在参数优化、变量排序以及后续变量分析解释等方面存在明显优势<sup>[22]</sup>。因此, 本文选择随机森林算法整合多元参数来反演海表盐度, 模型构建流程如图 3 所示。

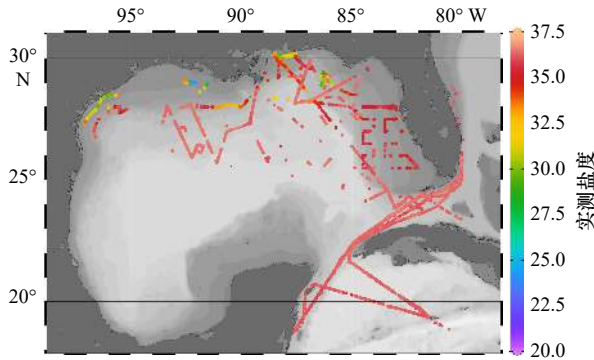


图 2 实测海表盐度数据分布  
Fig. 2 Spatial distribution of field sea surface salinity

表 2 实测数据与卫星数据统计信息

Table 2 Statistics of measured data and satellite data

变量	实测盐度	海表温度/°C	Rrs412/sr <sup>-1</sup>	Rrs443/sr <sup>-1</sup>	Rrs488/sr <sup>-1</sup>	Rrs555/sr <sup>-1</sup>	Rrs667/sr <sup>-1</sup>
最大值	36.587	32.525	0.030 18	0.036 44	0.045 42	0.031 41	0.010 12
最小值	22.328	18.270	-0.001 74	-0.000 42	0.000 85	-0.000 11	-0.000 65
中位数	36.147	26.475	0.008 00	0.007 00	0.005 41	0.001 46	0.000 15
平均值	35.554	26.514	0.007 51	0.006 71	0.005 35	0.001 68	0.000 21

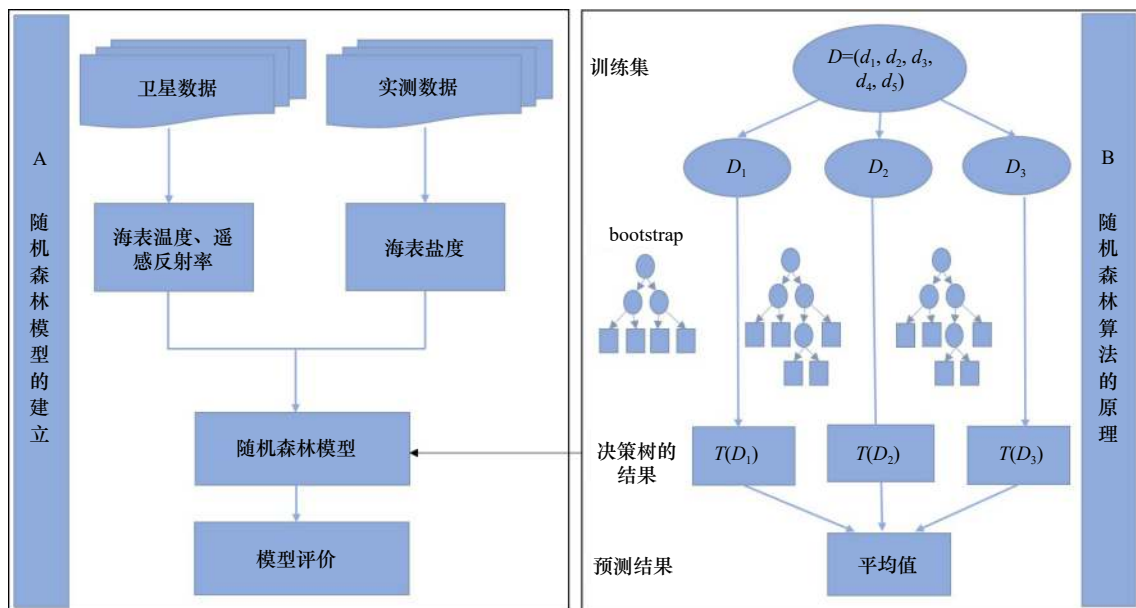


图 3 基于随机森林算法的海表盐度反演模型构建流程图

Fig. 3 Flow chart of sea surface salinity retrieval model based on random forest algorithm



本文利用 R 软件中的 randomForest 包来实现海表盐度反演模型的构建,将遥感反射率以及海表温度作为模型的输入,海表盐度作为模型的输出,通过多次迭代训练建立各影响因子与海表盐度之间的映射关系。模型结果使用均方根误差(Root Mean Square Error, RMSE)、决定系数( $R^2$ )、平均误差(Mean Bias, MB)、平均绝对误差(Mean Absolute Error, MAE)进行评估。

## 4 结果与分析

### 4.1 随机森林模型的适用性评估

将数据集按照 8:2 的比例随机分为训练集和测试集,利用本文建立的随机森林模型进行海表盐度反演,图 4 为训练集和测试集的实测盐度与模型反演盐度的线性回归密度散点图。在训练集中, $R^2$  为 0.986, RMSE 为 0.203, MAE 为 0.073, MB 为 -0.001。在测试集中, $R^2$  为 0.931, RMSE 为 0.335, MAE 为 0.154, MB 为 -0.002。从结果可知,实测盐度与模型反演的盐度之间相关性高、误差小,模型的拟合度高,验证了随机森林模型反演海表盐度的可行性。

为进一步评估模型在不同地理区域的性能,探究随机森林在不同环境下的适用性,本文选取墨西哥湾典型的河口区域(图 5d 中的 A 区域)以及环流影响的区域(图 5d 中的 B 区域)作为子研究区,得到各子区域的实测盐度与模型反演盐度的线性回归结果,见图 5a 和图 5b。如图所示,模型在河口区域与环流区域都得到了较好的拟合结果, RMSE 均小于 0.3,  $R^2$  均大于 0.8, 图 5c 表明, 92.4% 的误差在 -0.2~0.2 范围内。模型在河口区域取得更好结果的原因可能是河口区域接收了北部大量含有 CDOM 或其他陆源成分

的河流淡水。淡水在红光波段具有反射特性,同时,陆源成分在蓝光波段和绿光波段分别具有强吸收以及强反射的特性,而模型的输入参数(遥感反射率)能有效地捕捉到这些由光学特性表征的海水-淡水混合信息,因此该模型能很好地反演河口的盐度。环流区域由于羽流的扩散<sup>[14]</sup>,其海水-淡水混合信息能被模型很好地捕捉,故在该区域也得到了理想的拟合结果。

为验证随机森林模型的可行性,本文将随机森林模型与多元线性回归、人工神经网络、支持向量机模型的反演结果进行比较,用 RMSE、 $R^2$ 、MAE、MB 评价指标衡量模型的反演效果和稳健性。由图 6 结果可知,随机森林模型的  $R^2$  最高, RMSE 最小。综合相关性和误差,随机森林模型相比其他方法而言具有更强的泛化能力以及更佳的预测能力。

### 4.2 基于随机森林模型的盐度时空分布特征分析

探究盐度的空间分布特征可揭示淡水以及陆源物质的输送范围,对于研究与管理海洋环境具有重要意义。空间分辨率为约 1 km 的 MODIS 卫星数据能较好地展现海表盐度在空间上的差异性,捕捉盐度变化的细微特征。随机森林模型不同的树节点对应不同区域、不同环境下的盐度状况,可较为准确地反演出湾区内盐度的分布模式。基于此,将模型应用于 MODIS 卫星数据,得到墨西哥湾海域海表盐度反演分布结果,并对其时空分布特征进行分析。

由 2018 年墨西哥湾海表盐度的反演结果(图 7)可知,墨西哥湾海表盐度的空间分布呈近岸-河口低、离岸高,环状向内增值的态势。受密西西比河、阿查法拉亚河等河流输入的影响,盐度在近岸-河口较低。湾区北部呈现明显的低盐度水和高盐度水混合的现象,在风、环流以及涡流的影响下<sup>[23]</sup>,河流入海

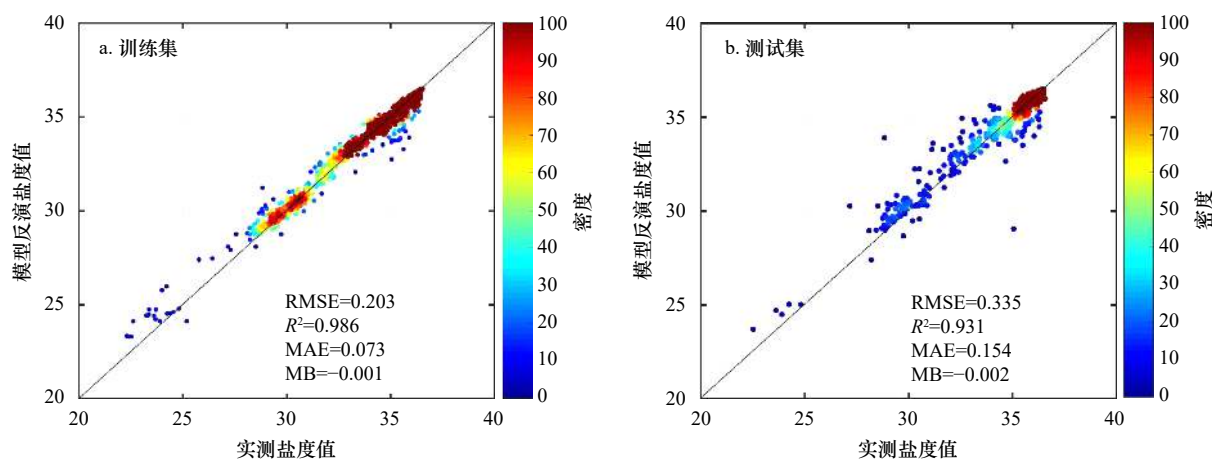


图 4 随机森林模型性能对比

Fig. 4 Performance comparison of random forest model

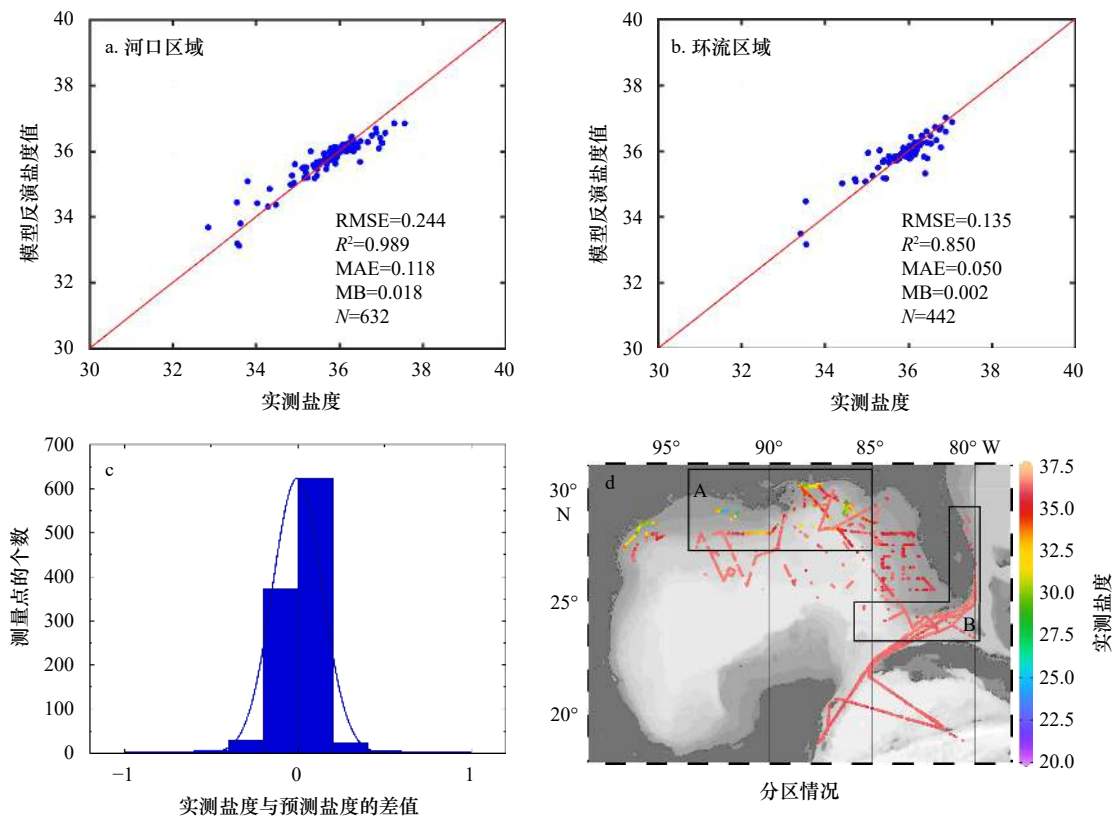


图5 随机森林模型在河口区域与环流区域的性能对比

Fig. 5 Random forest model performance comparison in estuary region and circulation region

口的低盐度水向南扩散和沿着大陆岸线向东西两侧扩散。由于河流的到达范围有限, 离岸区域的盐度较高且相对均质, 盐度值大部分为33~36, 以上分析结果表明, 入海河流的淡水输送及扩散直接影响湾区内盐度的分布状况。

2018年墨西哥湾逐月的海表盐度分布(图8)显示, 湾区盐度空间形态分布特征相似, 均呈现近岸—河口低、离岸高的态势。离岸区域的盐度值大部分为33~36。

受河流流量的影响, 湾区北部盐度较低且变化最为显著。春季, 湾区北部的盐度受密西西比河以及阿查法亚河等河流流量影响, 出现明显的盐度低值区。夏季, 在西南风和环流的作用下<sup>[24]</sup>, 低盐度羽流向东和向南扩散, 并沿环流东缘流向佛罗里达海峡。在河流流量以及风力的作用下, 低盐度水的扩散效应持续至9月, 此后, 由于河流流量的减小<sup>[25]</sup>, 秋季盐度的变化较小。河流流量在冬季持续减小, 并在1月中旬左右达到年度最低值<sup>[26]</sup>, 河口盐度低值区持续收缩。盐度空间分布成因分析表明, 盐度的变化与河流流量存在季节性响应, 在风力的影响下, 低盐度羽流于夏季出现明显的扩散现象。另外, 环流带来的加勒比海的海水也会影响湾区内的盐度。

#### 4.3 影响因子与海表盐度的相关性分析

输入变量的重要性可用于表示不同输入变量对模型的影响程度, 本文利用 InMSE 方法和 IncNodePurity 方法得到输入变量排序, 以此来增强随机森林模型的可解释性, 进而揭示海表温度和遥感反射率对于海表盐度反演的重要性。InMSE 与 IncNodePurity 同为 randomForest 包自带的评估变量重要性的方法。InMSE 表示逐一加入各变量后均方误差的增加, IncNodePurity 指节点在分裂前后的残差平方和的增加, 两种方法均通过量化误差来衡量输入变量的重要程度。海表温度以及遥感反射率数据(紫光波段(Rrs412)、蓝绿光波段(Rrs443、Rrs488、Rrs555)、红光波段(Rrs667))的重要性排序结果如图9所示, 两种不同重要性排序方法得出的结果都说明海表温度对海表盐度反演精度的影响最大, Rrs488 的对海表盐度反演精度的影响最小, 而在不同的方法下, Rrs412、Rrs443、Rrs555、Rrs667 的重要性不完全一致。

海表温度与海表盐度的响应程度很高主要是因为入海河流输入的淡水和海水之间会产生温差, 通过温差可捕捉河流的入海信息。而遥感反射率与盐度两者在本质上并无直接关系, 需要引入 CDOM 作为中介物进行分析。海表盐度与 CDOM 的吸收系数之

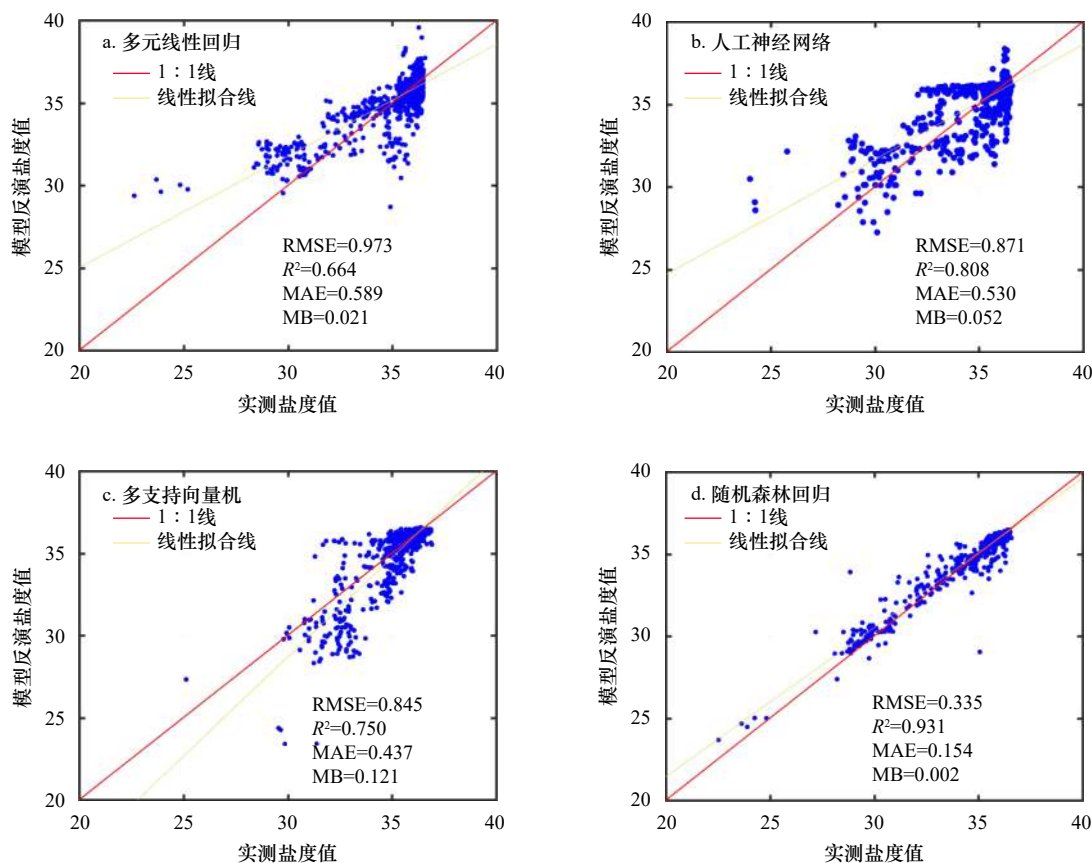


图 6 4 种模型反演的海表盐度验证

Fig. 6 Validation of sea surface salinity retrieved by four models

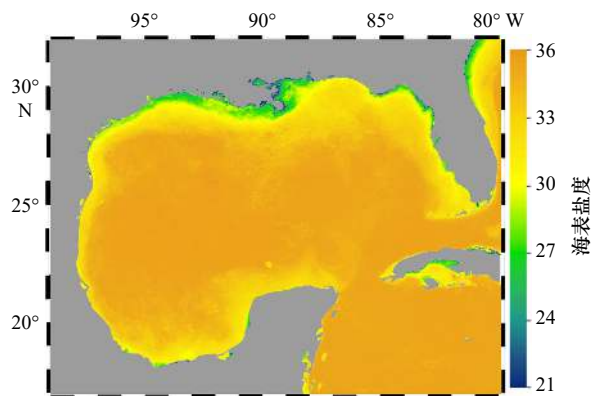


图 7 随机森林模型生成的 2018 年墨西哥湾年平均海表盐度反演图

Fig. 7 Annual mean sea surface salinity generated by the random forest model in the Gulf of Mexico in 2018

灰色代表陆地  
Land are shaded in grey

间存在显著的负相关关系<sup>[27-28]</sup>, 而 CDOM 的吸收系数可通过遥感反射率反演得出<sup>[29]</sup>, 因此, 遥感反射率能间接反演盐度信息。CDOM 的吸收系数在可见光波段内呈指数衰减的特性, 通常表示为遥感反射率的线性函数, 例如, 可被表示为蓝绿光波段的比值<sup>[30]</sup>或紫

光、蓝绿光的线性组合<sup>[31]</sup>。特别是在河口和近岸海域, CDOM 的浓度较高, 其对光的强吸收从紫光波段延伸到蓝绿光波段<sup>[30]</sup>。同时, 红光波段对盐度的反演也很重要<sup>[10]</sup>。综上所述, 反演 CDOM 的算法大多依赖于紫光波段、蓝绿光波段以及红光波段的遥感反射率数据, 故以上波段对海表盐度的反演起到了重要作用。

从 InMSE 方法排序结果可知, 在本文中海表温度对海表盐度反演精度的影响最大, 其次为绿光波段的遥感反射率 (Rrs555) 以及红光波段的遥感反射率 (Rrs667), 最后为紫光波段和蓝光波段的遥感反射率 (Rrs412、Rrs443、Rrs488)。从 IncNodePurity 方法的排序结果可知, 海表温度仍然对海表盐度反演精度的影响最大, 其次为蓝光波段、紫光波段、红光波段的遥感反射率 (Rrs443、Rrs412、Rrs667), 最后为蓝绿光波段的遥感反射率 (Rrs555、Rrs488)。InMSE 方法与 IncNodePurity 方法定义变量重要性标准的概念不同, 所以变量的排名也不尽相同。

#### 4.4 影响因子相关性的区域差异分析

区域差异是地理学的传统研究视角, 为探索海表盐度影响因子的区域差异, 本文对河口区域、环流区



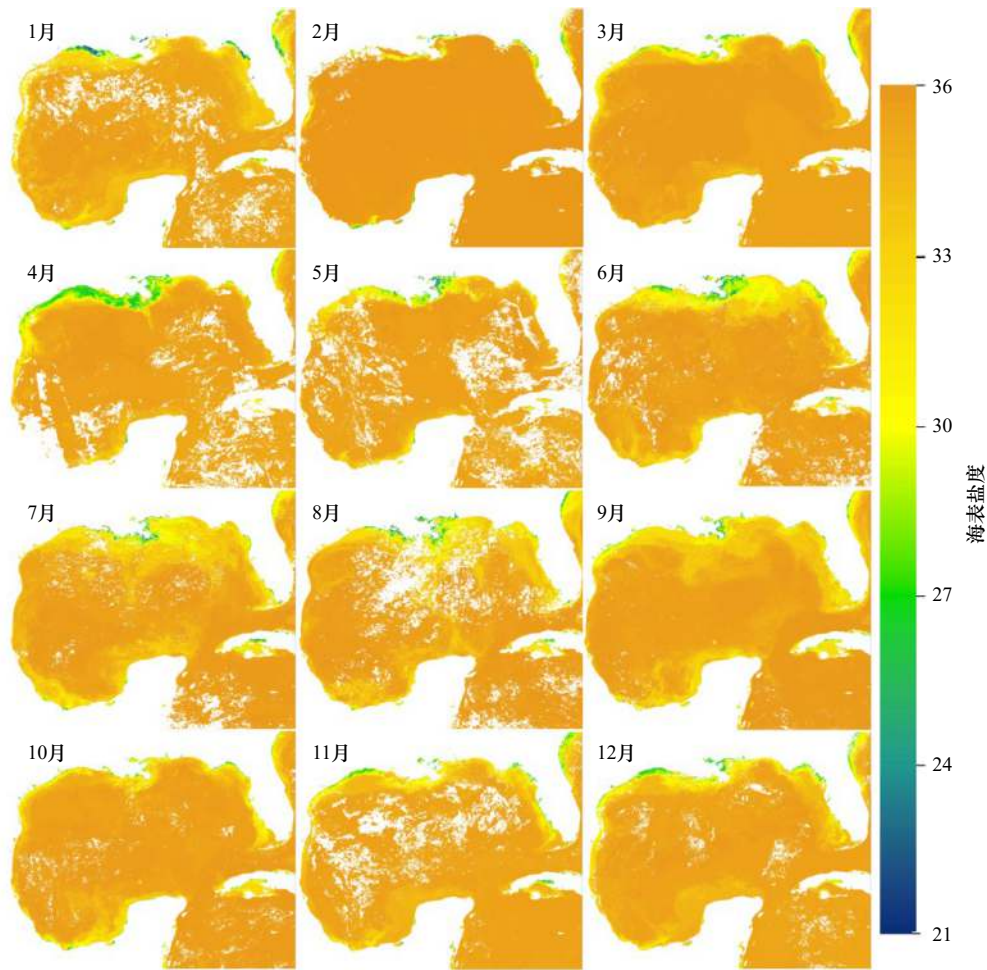


图 8 随机森林模型生成的 2018 年墨西哥湾月平均海表盐度反演图

Fig. 8 Monthly mean sea surface salinity generated by the random forest model in the Gulf of Mexico in 2018

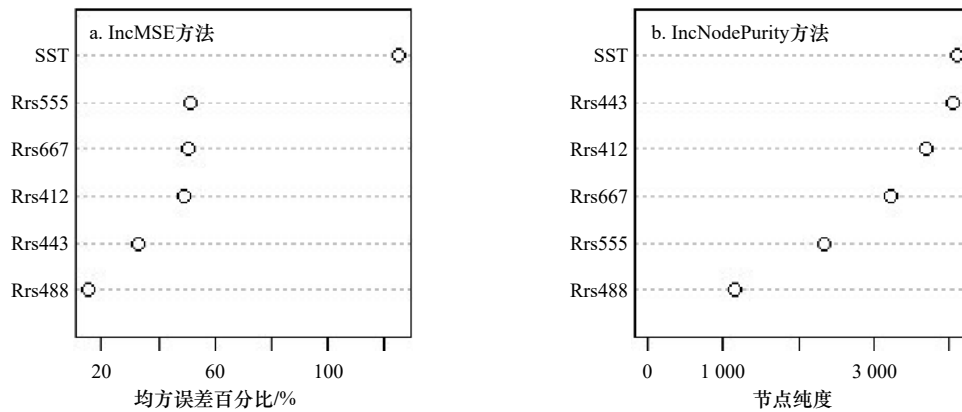


图 9 影响因子重要性排序

Fig. 9 Importance ranking of influence factors

域以及整个墨西哥湾区域影响因子与海表盐度间的相关性进行研究,各影响因子与海表盐度的相关性如图 10 所示。研究表明,在河口区域和环流区域,海表温度与海表盐度的相关性最大,而对于整个墨西哥湾海域, Rrs667 与海表盐度的相关性最大。

相比于同区域的遥感反射率,海表温度对海表盐

度的影响在河口区域和环流区域较大,这主要是由于水流混合导致的温度梯度造成的。河口区域低盐淡水和海水混合,环流将高温、高盐的加勒比海水输送至墨西哥湾,较大的温差导致盐度产生局部地变异。但对于整个墨西哥湾海域,海表温度对海表盐度的影响较小,因为大面积的离岸区域温度较为稳定,从而



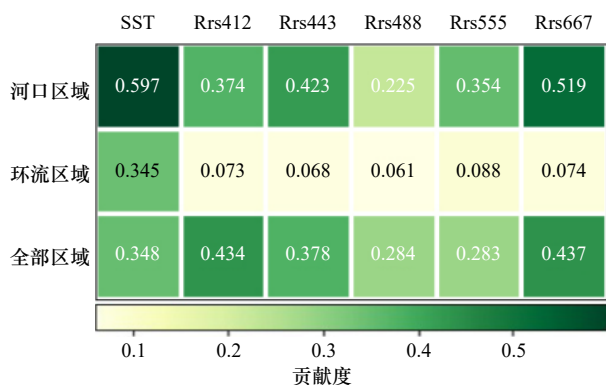


图 10 不同地理分区海表盐度影响因子的贡献度  
Fig. 10 Contribution of each factor to sea surface salinity in different geographical regions

弱化了温度的影响。针对不同的地理分区,海表温度在河口区域对海表盐度的影响最大,在环流区域的影响最小。该结果显示,在墨西哥湾海域,冲淡水对海表盐度的影响比环流更加明显,这主要是由于入海河流输入的淡水与湾区内海水的盐度差和温度差相较于加勒比海水与湾区内海水更大。

在富含沉积物的河口区域,各波段的遥感反射率对海表盐度的影响较大,其原因为遥感反射率能很好地捕捉到陆源物质。而在环流区域,同质海水使得遥感反射率对海表盐度的影响并不显著,从而强化了温度的影响。在全部海域,各波段的遥感反射率对海表盐度的影响也较大,主要是因为该区域不仅包含受冲淡水影响的河口区域、由暖流控制的环流区域,还包括了同样富含陆源物质的大陆架与受温度影响较小的离岸区域,因此,对于整个墨西哥湾,遥感反射率对海表盐度反演的影响也较为显著。

## 5 结论

本文将海表温度作为模型的输入变量之一,利

### 参考文献:

- [1] Devlin M J, Petus C, Da Silva, et al. Water quality and river plume monitoring in the Great Barrier Reef: an overview of methods based on ocean colour satellite data[J]. *Remote Sensing*, 2015, 7(10): 12909–12941.
- [2] Horner-Devine A R, Hetland R D, MacDonald D G. Mixing and transport in coastal river plumes[J]. *Annual Review of Fluid Mechanics*, 2015, 47: 569–594.
- [3] 白雨, 赵亮, 刘境舟. 生态因子在黄海绿潮生消过程中的作用[J]. *海洋学报*, 2019, 41(8): 97–105.  
Bai Yu, Zhao Liang, Liu Jingzhou. The role of ecological factors in the progress of the green tide in the Yellow Sea[J]. *Haiyang Xuebao*, 2019, 41(8): 97–105.
- [4] 韦钦胜, 王保栋, 陈建芳, 等. 长江口外缺氧区生消过程和机制的再认知[J]. *中国科学: 地球科学*, 2015, 58(4): 628–648.  
Wei Qinsheng, Wang Baodong, Chen Jianfang, et al. Recognition on the forming-vanishing process and underlying mechanisms of the hypoxia off the Yangtze River estuary[J]. *Science China: Earth Sciences*, 2015, 58(4): 628–648.
- [5] Cannizzaro J P, Carlson Jr P R, Yarbrow L A, et al. Optical variability along a river plume gradient: implications for management and remote sensing[J]. *Estuarine, Coastal and Shelf Science*, 2013, 131: 149–161.

用随机森林提供变量排序、变量解释的优势,证明了海表温度跟海表盐度之间具有较强的相关性,并验证了各影响因子与海表盐度的相关性存在空间异质性。基于 2018 年遥感与实测匹配的遥感反射率数据和海表温度数据构建的随机森林模型,对墨西哥湾的海表盐度进行反演,结合比较其他方法得出以下结论:

(1) 随机森林模型的反演精度高于多元线性回归、支持向量机以及人工神经网络模型,其 RMSE 为 0.335,  $R^2$  为 0.931, 更精确地反演了墨西哥湾海表盐度;(2) 墨西哥湾海表盐度时空分布特征结果表明,湾区内海表盐度的空间分布呈近岸-河口低、离岸高,环状向内增值的态势。海表盐度空间分布成因分析表明,湾区内海表盐度的变化与河流流量存在季节性响应,在风和环流的影响下,入海河流的淡水输送及扩散直接影响湾区内海表盐度的分布状况;(3) 对模型的影响因子进行排序,发现了海表温度是墨西哥湾海域反演海表盐度的重要因素;(4) 对不同的区域内海表温度和遥感反射率与海表盐度的相关性进行探析,验证了影响因子与海表盐度间的相关性存在空间异质性,其中,河口区域的海表温度和 Rrs667 与海表盐度的相关性最强。

本研究发现了海表温度与海表盐度存在较强的相关性,各影响因子与海表盐度存在空间异质性,可为沿海水域海表盐度反演模型提供变量选取的思路。建立的高精度海表盐度反演模型可适用于类似地理条件的区域,进一步的拓展应用还需进行相应海域及不同时相数据的分析和检验。

致谢: 感谢 NOAA 国家环境信息中心 (NCEI) (<https://www.ncei.noaa.gov/access/ocean-carbon-data-system/>) 提供的 2018 年墨西哥湾实测盐度数据。

- [6] 周博天. 海面盐度多源遥感协同反演方法研究[D]. 北京: 中国地质大学(北京), 2013.  
Zhou Botian. The study of sea surface salinity model based on multi-source remote sensing coordinated retrieval[D]. Beijing: China University of Geosciences (Beijing), 2013.
- [7] Vandermeulen R A, Arnone R, Ladner S, et al. Estimating sea surface salinity in coastal waters of the Gulf of Mexico using visible channels on SNPP-VIIRS[C]//Proceedings of SPIE 9111, Ocean Sensing and Monitoring VI. Baltimore, Maryland, United States: SPIE, 2014: 911109.
- [8] Liu Rongjie, Zhang Jie, Yao Haiyan, et al. Hourly changes in sea surface salinity in coastal waters recorded by Geostationary Ocean Color Imager[J]. *Estuarine, Coastal and Shelf Science*, 2017, 196: 227–236.
- [9] Geiger E F, Grossi M D, Trembanis A C, et al. Satellite-derived coastal ocean and estuarine salinity in the Mid-Atlantic[J]. *Continental Shelf Research*, 2013, 63(S1): S235–S242.
- [10] Kim D W, Park Y J, Jeong J Y, et al. Estimation of hourly sea surface salinity in the East China sea using geostationary ocean color imager measurements[J]. *Remote Sensing*, 2020, 12(5): 755.
- [11] Ahn Y H, Shanmugam P, Moon J E, et al. Satellite remote sensing of a low-salinity water plume in the East China Sea[J]. *Annales Geophysicae*, 2008, 26(7): 2019–2035.
- [12] Qing Song, Zhang Jie, Cui Tingwei, et al. Retrieval of sea surface salinity with MERIS and MODIS data in the Bohai Sea[J]. *Remote Sensing of Environment*, 2013, 136: 117–125.
- [13] 张文. 基于智能计算的水体悬浮物遥感反演研究——以长江中游为例[D]. 北京: 中国科学院研究生院, 2012.  
Zhang Wen. Remote sensing retrieval of suspended matters based on intelligent calculation—a case study of middle Yangtze River[D]. Beijing: University of Chinese Academy of Sciences, 2012.
- [14] Chen Shuangling, Hu Chuanmin. Estimating sea surface salinity in the northern Gulf of Mexico from satellite ocean color measurements[J]. *Remote Sensing of Environment*, 2017, 201: 115–132.
- [15] 王奕森, 夏树涛. 集成学习之随机森林算法综述[J]. *信息通信技术*, 2018, 12(1): 49–55.  
Wang Yisen, Xia Shutao. A survey of random forests algorithms[J]. *Information and Communications Technologies*, 2018, 12(1): 49–55.
- [16] Rabalais N N, Turner R E, Wiseman Jr W J. Gulf of Mexico Hypoxia, A. K. A. “The dead zone”[J]. *Annual Review of Ecology and Systematics*, 2002, 33(1): 235–263.
- [17] Otis D B, Le Hénaff M, Kourafalou V H, et al. Mississippi River and Campeche bank (Gulf of Mexico) episodes of cross-shelf export of coastal waters observed with satellites[J]. *Remote Sensing*, 2019, 11(6): 723.
- [18] Fu Zhiyi, Hu Linshu, Chen Zhende, et al. Estimating spatial and temporal variation in ocean surface  $pCO_2$  in the Gulf of Mexico using remote sensing and machine learning techniques[J]. *Science of the Total Environment*, 2020, 745: 140965.
- [19] Breiman L. Random forests[J]. *Machine Learning*, 2001, 45(1): 5–32.
- [20] Ho T K. The random subspace method for constructing decision forests[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1998, 20(8): 832–844.
- [21] Breiman L. Bagging predictors[J]. *Machine Learning*, 1996, 24(2): 123–140.
- [22] 方馨蕊, 温兆飞, 陈吉龙, 等. 随机森林回归模型的悬浮泥沙浓度遥感估算[J]. *遥感学报*, 2019, 23(4): 756–772.  
Fang Xinrui, Wen Zhaofei, Chen Jilong, et al. Remote sensing estimation of suspended sediment concentration based on Random Forest Regression Model[J]. *Journal of Remote Sensing*, 2019, 23(4): 756–772.
- [23] Vazquez J, Gierach M M, Leben R R, et al. Initial results on the variability of sea surface salinity from Aquarius/SAC-D in the Gulf of Mexico[C]//American Geophysical Union, Fall Meeting 2012. San Francisco, CA: AGU, 2012.
- [24] Wiseman W J, Rabalais N N, Turner R E, et al. Seasonal and interannual variability within the Louisiana coastal current: stratification and hypoxia[J]. *Journal of Marine Systems*, 1997, 12(1/4): 237–248.
- [25] Fournier S, Lee T, Gierach M M. Seasonal and interannual variations of sea surface salinity associated with the Mississippi River plume observed by SMOS and Aquarius[J]. *Remote Sensing of Environment*, 2016, 180: 431–439.
- [26] Fu Zhiyi, Wu Fangfang, Zhang Zhengliang, et al. Sea surface salinity estimation and spatial-temporal heterogeneity analysis in the gulf of Mexico[J]. *Remote Sensing*, 2021, 13(5): 881.
- [27] Binding C E, Bowers D G. Measuring the salinity of the Clyde Sea from remotely sensed ocean colour[J]. *Estuarine, Coastal and Shelf Science*, 2003, 57(4): 605–611.
- [28] Wouthuyzen S, Tarigan S, Kusmanto E, et al. Measuring sea surface salinity of the Jakarta bay using remotely sensed of ocean color data acquired by modis sensor[J]. *Marine Research in Indonesia*, 2011, 36(2): 51–70.
- [29] Darecki M, Stramski D. An evaluation of MODIS and SeaWiFS bio-optical algorithms in the Baltic Sea[J]. *Remote Sensing of Environment*, 2004, 89(3): 326–350.
- [30] 王林, 赵冬至, 杨建洪, 等. 大洋河口海域有色溶解性有机物的光学特性及遥感反演模型[J]. *海洋学报*, 2011, 33(1): 45–51.  
Wang Lin, Zhao Dongzhi, Yang Jianhong, et al. The optical properties and remote sensing retrieval model of chromophoric dissolved organic matter in the Dayang Estuary[J]. *Haiyang Xuebao*, 2011, 33(1): 45–51.
- [31] Johannessen S C, Miller W L, Cullen J J. Calculation of UV attenuation and colored dissolved organic matter absorption spectra from measurements of ocean color[J]. *Journal of Geophysical Research: Oceans*, 2003, 108(C9): 3301.

## Retrieval of sea surface salinity in the Gulf of Mexico based on random forest method

Wu Fangfang<sup>1,2</sup>, Fu Zhiyi<sup>1,2</sup>, Hu Linshu<sup>1,2</sup>, Zhang Feng<sup>1,2</sup>, Du Zhenhong<sup>1,2</sup>, Liu Renyi<sup>1,2</sup>

(1. Zhejiang Provincial Key Laboratory of Geographic Information Science, Zhejiang University, Hangzhou 310028, China; 2. School of Earth Sciences, Zhejiang University, Hangzhou 310027, China)

**Abstract:** Salinity is an important parameter to characterize physical and biogeo-chemical processes. Optical satellite images with high resolution can avoid radio frequency interference, and provide a feasible way to monitor sea surface salinity (SSS) in coastal regions. Using an extensive dataset of ship-based SSS and MODIS estimated remote sensing reflectance (Rrs) at 412 nm, 443 nm, 488 nm, 555 nm and 667 nm and sea surface temperature (SST) a random forest (RF) model has been utilized to retrieve the SSS. Based on the predicted SSS, we analyze the spatiotemporal heterogeneity of SSS in the Gulf of Mexico and contribution of each factor with correlations coefficient. The results show that: (1) the RF model can accurately estimate the SSS in the Gulf of Mexico (RMSE=0.335,  $R^2=0.931$ ); (2) the spatial distribution pattern of SSS shows a ring-shaped inward value increase, which is affected by river discharge, wind forcing and circulation; (3) there is a strong correlation between SSS and SST, and SST significantly impact in retrieving SSS; (4) the correlation between SST, Rrs and SSS appears spatial heterogeneity.

**Key words:** sea surface salinity; random forest; spatial-temporal heterogeneity; Gulf of Mexico