

赵海阳, 沈芳, 孙雪融, 等. 中国东部海域浮游植物类群遥感反演研究[J]. 海洋学报, 2022, 44(4): 153–168, doi:10.12284/hyxb2022062
Zhao Haiyang, Shen Fang, Sun Xuerong, et al. Remote sensing retrieval of phytoplankton group in the eastern China seas[J]. Haiyang Xuebao, 2022, 44(4): 153–168, doi:10.12284/hyxb2022062

中国东部海域浮游植物类群遥感反演研究

赵海阳¹, 沈芳^{1*}, 孙雪融¹, 魏小島²

(1. 华东师范大学 河口海岸学国家重点实验室, 上海 200241; 2. 华东师范大学 崇明生态研究院, 上海 202162)

摘要: 浮游植物类群遥感反演能够为全面认识浮游植物在海洋生态系统中的作用提供重要的数据资料。但由于复杂的水体光学特性, 近海浮游植物类群遥感反演存在着巨大挑战。本研究以复杂光学二类水体——中国东部海域为研究区, 通过使用 3 种建模方法, 即波段组合法、基于奇异值分解的多元线性回归法、基于奇异值分解的 XGBoost 回归法, 利用遥感反射率数据反演浮游植物类群。经原位实测数据集验证, 基于奇异值分解的 XGBoost 回归法构建的 8 类浮游植物叶绿素 *a* 浓度反演模型的精度最高, 其中硅藻、甲藻的叶绿素 *a* 浓度反演模型在验证集上的决定系数均大于 0.7。相比之下, 3 种建模方法估算得到的绿藻、蓝藻和金藻的叶绿素 *a* 浓度精度较低 (验证结果的决定系数小于 0.45)。同时, 研究评估了 OLCI 影像的 3 种大气校正方法 (C2RCC、POLYMER、MUMM) 在中国东部海域的适用性。结果显示, 相对于其他两种大气校正算法, C2RCC 在各波段有较好的表现 (均方根误差小于 0.0048 sr^{-1})。将 3 种浮游植物类群反演模型应用到大气校正后的 OLCI 影像, 验证结果显示, 利用基于奇异值分解的多元线性回归法建立的硅藻叶绿素 *a* 浓度模型有较好的反演精度 (决定系数为 0.56)。

关键词: 中国东部海域; 浮游植物类群; 遥感反演; OLCI; 大气校正

中图分类号: TP79

文献标志码: A

文章编号: 0253-4193(2022)04-0153-16

1 引言

浮游植物在生物地球化学循环中扮演着十分重要的角色。研究表明, 尽管浮游植物的生物量只有陆地植物生物量的 1%, 但贡献了全球大约 50% 的初级生产力^[1-2]。由于形态和生理特征的差异, 不同浮游植物在生物地球化学过程和海洋生态系统中的作用有所差别^[3]。如硅藻贡献了海洋初级生产力的 40%, 同时也是全球硅循环的重要驱动因素^[4]; 甲藻在生长代谢过程中产生的二甲基硫, 能够形成硫酸盐气溶胶, 进而维持全球辐射平衡^[5]。

因此, 为了更加全面地认识浮游植物在全球生态系统中所起的作用, 需要对浮游植物的生物量、空间

分布、类群和种群组成等信息进行有效观测。其中, 遥感技术的发展, 有效弥补了传统调查方式较难进行大范围、连续观测的不足, 同时使得对地观测变得更加经济高效。叶绿素 *a* (Chlorophyll *a*, Chl *a*) 浓度常被用来指示浮游植物生物量^[6], 其卫星反演一直以来是水色遥感领域的重要研究内容^[7]。但近年来随着人们对海洋生态系统、生物地球化学循环、气候变化等研究的逐步深入, 大量研究显示, Chl *a* 浓度已无法满足人们对浮游植物类群进行深入了解的迫切需求^[8-10], 使得发展浮游植物类群遥感反演方法变得十分必要。

目前, 浮游植物类群遥感反演方法主要分为 3 种: 基于浮游植物丰度法、基于生态环境因子法、基于光学辐射法^[9, 11]。不同浮游植物因大小、形状、外部和

收稿日期: 2021-08-17; 修订日期: 2021-11-20。

基金项目: 国家自然科学基金 (42076187, 41771378)。

作者简介: 赵海阳 (1997—), 男, 河南省许昌市人, 主要从事海洋水色遥感研究。E-mail: 51193904013@stu.ecnu.edu.cn

* 通信作者: 沈芳, 女, 教授, 主要从事海洋水色遥感研究。E-mail: fshen@sklec.ecnu.edu.cn

内部结构以及色素成分上的差异会改变光学辐射信号,同时卫星影像经过大气校正,首先得到水体的遥感反射率(Remote Sensing Reflectance, R_{rs})。因此,利用遥感反射率反演浮游植物类群是最为直接的方法^[11]。然而,对于水体光学性质复杂的中国东部海域,使用遥感反射率反演浮游植物类群的研究仍相对较少^[12-13]。

已有研究表明,基于波段组合的 Chl *a* 浓度反演模型简单高效,被广泛使用^[14-15]。因此,可以通过寻找与浮游植物类群 Chl *a* 浓度相关性最高的波段组合,建立反演模型。此外,基于遥感反射率数据以及其他表观光学变量的奇异值分解法,在大洋一类水体估算浮游植物类群 Chl *a* 浓度时取得了较好的结果^[16-17]。上述的研究方法主要是基于相对简单的线性回归模型,而机器学习算法,特别是极端梯度提升算法(eXtreme Gradient Boosting, XGBoost),因其高效的可扩展性和灵活性以及快速准确的特点,近年来在水色遥感领域中得到了不断的发展和应用^[18-19]。基于大量实测数据,本研究使用了3种建模方法构建浮游植物类群反演模型,即基于波段组合法、基于奇异值分解的多元线性回归法、基于奇异值分解的 XGBoost 回归法,并对比验证了不同模型的估算精度。

Sentinel-3/OLCI(Ocean and Land Colour Instrument)传感器作为新一代的水色传感器,其在波段设置方面继承了 Envisat/MERIS(Medium Resolution Imaging Spectrometer)传感器的参数,并针对水色遥感探测进行了优化^[20]。为了获得较高精度的遥感反射率产品,本研究利用实测的遥感反射率数据对 OLCI 3种大气校正算法(C2RCC(Case 2 Regional Coast Color)、POLYMER(POLYnomial-based algorithm applied to MERIS)、MUMM(Management Unit of Mathematical Models))进行了评估。

综上,本研究以中国东部海域为研究区,构建并对比了3种浮游植物类群遥感反演模型,验证了针对 OLCI 传感器的3种大气校正算法,并评估了大气校正对浮游植物类群反演模型的影响。将精度较高的硅藻 Chl *a* 浓度反演模型(基于奇异值分解的多元线性回归法)应用至使用 C2RCC 大气校正算法计算得到的 OLCI 遥感反射率数据,估算中国东部海域硅藻 Chl *a* 浓度。

2 数据与方法

2.1 研究区概况

本研究的实测原位数据来自于中国东部海域(图1)。中国东部海域包括渤海、黄海和东海,总面积约为 $1 \times 10^6 \text{ km}^2$ ^[21]。中国东部海域也是我国重要的碳汇,年

固碳量达 $222 \times 10^6 \text{ t}$ ^[22]。受到黑潮、台湾暖流、浙闽沿岸流等多种流系^[23]以及河流输入泥沙和营养盐的影响,该区域的浮游植物类群存在较大的时空变化^[24]。

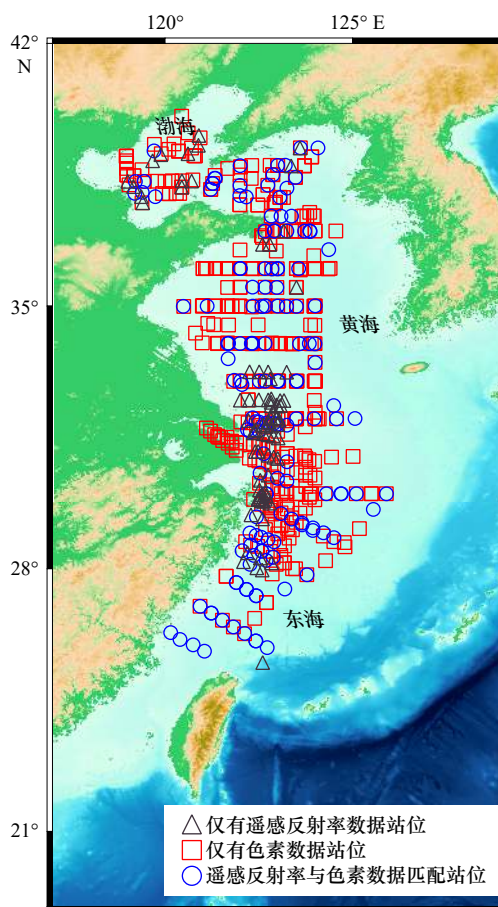


图1 研究区位置以及采样站点
Fig. 1 Location of the study area and sampling stations

2.2 实测遥感反射率数据

本研究共采集获取了2015–2020年8个航次的228条遥感反射率数据(图1)。该数据由 Sea-Bird Scientific 公司的海面高光谱测量仪(HyperSAS)测定,波段范围为349~856 nm,分辨率为1 nm。数据测定和处理参考文献[25]。根据 OLCI 的波谱响应函数,利用式(1)将实测的遥感反射率卷积计算为 OLCI 等效波长测得的遥感反射率^[26]。卷积完成后得到16个等效波段,即400 nm、412.5 nm、442.5 nm、490 nm、510 nm、560 nm、620 nm、665 nm、673.75 nm、681.25 nm、708.75 nm、753.75 nm、761.25 nm、764.375 nm、767.5 nm、778.75 nm。由于受到水汽吸收作用的影响,761.25 nm、764.375 nm、767.5 nm 波段不参与大气校正,因此,之后的研究只使用了经过大气校正的13个波段。

$$R_{rs} = \frac{\int_{\lambda_1}^{\lambda_2} S(\lambda_i) R_{rs}(\lambda_i) d\lambda}{\int_{\lambda_1}^{\lambda_2} S(\lambda_i) d\lambda}, \quad (1)$$

式中, R_s 为传感器测量的等效遥感反射率; λ_1 、 λ_2 为波段范围内的下限和上限; $S(\lambda_i)$ 为波长 λ_i 处传感器的光谱响应函数; $R_s(\lambda_i)$ 为 λ_i 处实测的遥感反射率。

2.3 实测浮游植物色素浓度数据

不同的浮游植物含有特定的特征色素, 在原位测量中, 常通过测定特征色素来确定浮游植物类群^[27]。本研究采用来自 2015–2020 年 9 个航次共 511 个站位(图 1) 的海水表层特征色素浓度数据(航次及采样信息详见文献 [28–29])。在野外采集特征色素时, 将水样过滤富集到孔径为 0.7 μm , 直径为 25 mm 的 Whatman GF/F 玻璃纤维滤膜上, 然后放入 -40°C 的冰箱中冷冻避光保存。返回实验室后通过高效液相色谱(High Performance Liquid Chromatography, HPLC)测定色素浓度。

CHEMTAX 是根据浮游植物特征色素浓度估算不同浮游植物类群 Chl *a* 浓度的程序^[30]。程序通过最速下降法不断迭代, 调整特征色素比率矩阵, 定量确定浮游植物类群组成。本研究参考 Sun 等^[31] 在中国东部海域给出的初始特征色素比率矩阵, 使用实测的多甲藻黄素、19'-丁酰氧基-岩藻黄素、岩藻黄素、19'-己酰氧基-岩藻黄素、新黄素、青绿藻黄素、紫黄素、别藻黄素、叶黄素(Lutein)、玉米黄素、叶绿素 *b*、叶绿素 *a* 的浓度, 计算得到青绿藻(Prasinophytes)、甲藻(Dinoflagellates)、隐藻(Cryptophytes)、绿藻(Chlorophytes)、蓝藻(Cyanobacteria)、硅藻(Diatoms)、金藻(Chrysophytes)、定鞭藻(Prymnesiophytes) 的 Chl *a* 浓度。

2.4 卫星影像数据

OLCI 传感器是欧洲空间局(The European Space Agency, ESA) Sentinel-3A(2016 年 2 月 16 日发射)和 Sentinel-3B(2018 年 4 月 25 日发射)卫星上的有效载荷之一。OLCI 的波段范围为 400~1 020 nm, 包括 21 个波段。两颗卫星在同一轨道运行, 于当地太阳时 10 点过境。OLCI 传感器的 L1B(传感器接收到的天顶辐亮度)产品包括了两种空间分辨率的产品, 分别是 300 m 分辨率(Full Resolution, FR)和 1.2 km 分辨率(Reduce Resolution, RR)^[32]。本研究使用的是 300 m 分辨率的 L1B 数据(<https://scihub.copernicus.eu/dhus>)。

2.5 大气校正算法

C2RCC 大气校正算法通过大量的实测数据, 利用水体光学传输模型和大气传输模型, 模拟了不同情况下的天顶辐亮度和离水反射数据集, 再利用模拟的数据集构建神经网络模型, 完成天顶辐亮度和离水反射之间的计算。在构建离水反射数据集时, 使用了从

世界各地光学复杂水体采集的光学参数, 因此能够适用于光学二类水体^[33–34]。

POLYMER 大气校正算法利用从蓝光到近红外光的整个光谱范围, 将大气和水体的信号解耦, 完成大气校正。该方法依赖于两个基本模型, 一个是关于大气的估计模型, 该大气模型使用全光谱建立多项式拟合大气贡献, 使得最终的模型可以适用于有太阳耀光污染的区域; 另一个是离水反射模型, 该模型使用 Chl *a* 浓度和后向散射系数模拟离水反射。其中后向散射系数的引入使得模型能够适用于光学二类水体^[35–36]。

MUMM 大气校正算法是对标准 SeaWiFS 大气校正算法的扩展。SeaWiFS 标准大气校正算法假设水体在近红外波段的离水反射为 0, 卫星近红外波段接收到的信号都是由气溶胶反射贡献, 通过将气溶胶贡献从近红外波段推算至短波波段, 完成大气校正^[37]。然而, 这一假设多适用于水色组分相对简单的一类水体。针对浑浊水体, Ruddick 等^[38] 将近红外波段离水反射为 0 的假设替换为在兴趣区内 765 nm 和 865 nm 波段离水反射比和气溶胶散射比为定值的假设, 从而推算气溶胶散射, 完成大气校正。

在使用实测数据进行大气校正结果验证时, 本研究选择在采样站位前后 3 h 过境的卫星影像上, 以采样点为中心, 空间上 3×3 窗口平均值, 作为该站位卫星探测的遥感反射率。

2.6 浮游植物类群遥感反演模型

2.6.1 波段组合法

波段组合法(Band Combination, BC)是建立遥感反射率波段组合值与不同浮游植物类群 Chl *a* 浓度之间经验关系的方法。利用相关系数, 选择与各浮游植物类群 Chl *a* 浓度相关性最高的遥感反射率波段组合(表 1)。在找到各浮游植物类群最优波段组合后, 分别使用一次函数、二次函数、三次函数、幂函数和指数函数将波段组合值与浮游植物 Chl *a* 浓度进行拟合, 确定最优的建模方式。

2.6.2 基于奇异值分解的多元线性回归法

在构建基于奇异值分解的多元线性回归法(SVD (Singular Value Decomposition)+MLR (Multiple Linear Regression))模型时, 首先对建模数据集的遥感反射率进行标准化, 将标准化后的遥感反射率矩阵($Rrs_{\text{standardized}}$, 矩阵大小为 $M \times N$, M 为光谱数量, N 为波段数量)进行奇异值分解, 如下所示,

$$Rrs_{\text{standardized}} = UAV^T, \quad (2)$$

式中, 正交矩阵 U ($M \times M$) 的列向量为左奇异向量, 代

表 1 本研究采用的波段组合形式

Table 1 The band combinations used in this study

序号	波段(B)组合形式	序号	波段(B)组合形式
BC ₁	$B_i + B_j$	BC ₆	$\frac{B_i - B_j}{B_k}$
BC ₂	$B_i - B_j$	BC ₇	$\frac{B_i}{B_j + B_k}$
BC ₃	$B_i \times B_j$	BC ₈	$\frac{B_i}{B_j - B_k}$
BC ₄	B_i / B_j	BC ₉	$B_i \times \left(\frac{1}{B_j} + \frac{1}{B_k} \right)$
BC ₅	$\frac{B_i + B_j}{B_k}$	BC ₁₀	$B_i \times \left(\frac{1}{B_j} - \frac{1}{B_k} \right)$

表建模数据的主成分; $\Lambda(M \times N)$ 为矩阵 $Rrs_{standardized}$ 的奇异值矩阵, 是一个对角阵, 奇异值由大到小在矩阵对角线排列; 正交矩阵 $V(N \times N)$ 的列向量为右奇异向量。计算过程中, 当第 $n+1$ 个主成分代表的方差小于第一个主成分代表方差的 0.000 1 倍时, 该主成分被当作是无意义的而被舍弃^[16]。最终, 将前 n 个主成分与浮游植物 Chl a 浓度的对数 ($\lg C_p$) 进行多元线性回归, 公式为

$$\lg C_p = a + b_1 u_1 + b_2 u_2 + \dots + b_n u_n, \quad (3)$$

式中, a 和 b_1, b_2, \dots, b_n 为方程的回归系数; C_p 为浮游植物 Chl a 浓度; u_1, u_2, \dots, u_n 为前 n 个主成分。

当将该方法应用至遥感影像进行浮游植物类群 Chl a 浓度估算时, 首先将影像的遥感反射率矩阵进行标准化 ($Rrs_{standardized}^{application}$), 随后将数据投影到建模数据奇

异值分解后的特征空间上得到 $U^{application}$ 矩阵 (式 (4)), 通过使用 $U^{application}$ 矩阵的列向量和建模时拟合的回归系数来估算各浮游植物类群 Chl a 浓度的对数 ($\lg C_p^{application}$)。该方法流程如图 2 所示。

$$U^{application} = Rrs_{standardized}^{application} V A^{-1}. \quad (4)$$

2.6.3 基于奇异值分解的 XGBoost 回归法

XGBoost 是一种基于梯度提升树的机器学习方法^[39]。其将多个弱学习器结合成为一个强的学习器, 通过不断地向集成模型中添加预测模型, 且每一个新添加的预测模型都对前一个模型的预测残差进行拟合, 以完成最后的模型构建。基于奇异值分解的 XGBoost (记为 SVD+XGBoost) 的浮游植物类群 Chl a 浓度反演算法的流程与 SVD+MLR 法一致, 仅将多元线性回归模型替换为 XGBoost 回归模型 (图 2)。

2.7 精度评价指标

为评估建立的各类群浮游植物 Chl a 浓度反演模型的精度, 以及 3 种大气校正算法的精度, 研究使用了决定系数 (R^2)、均方根误差 (Root Mean Square Error, RMSE)、平均绝对误差 (Mean Absolute Error, MAE)、中值绝对百分比误差 (Median Absolute Percentage Error, MAPE) 作为评价指标。各指标公式为

$$R^2 = 1 - \frac{\sum_{i=1}^N (E_i - M_i)^2}{\sum_{i=1}^N (E_i - \bar{M})^2}, \quad (5)$$

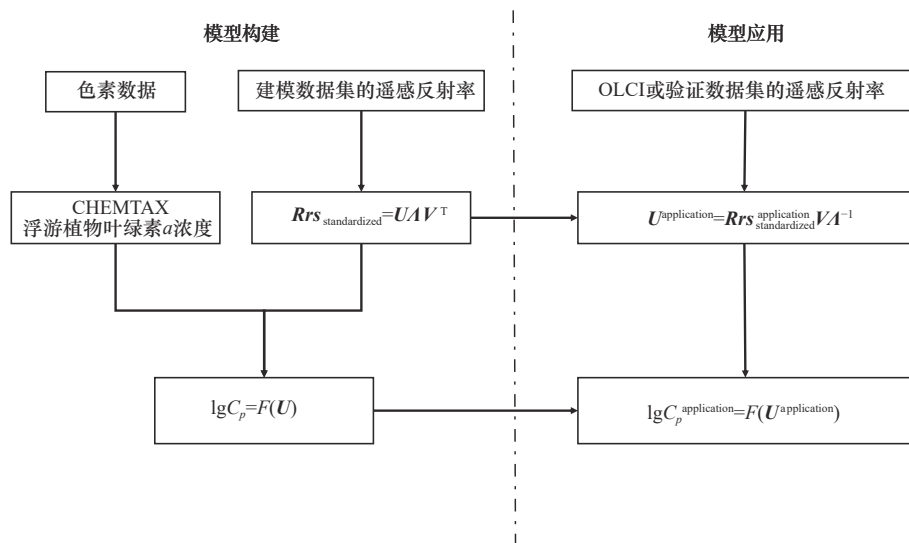


图 2 基于奇异值分解的浮游植物类群 Chl a 浓度反演算法流程

Fig. 2 Flow chart of Chl a concentration inversion algorithm of phytoplankton groups based on singular value decomposition

F 表示本研究中使用的回归模型, 即为多元线性回归模型或 XGBoost 回归模型

F represents the regression model used in this study, such as the multiple linear regression model or XGBoost regression model

$$\text{RMSE} = \sqrt{\frac{\sum_{i=1}^N (E_i - M_i)^2}{N}}, \quad (6)$$

$$\text{MAE} = \frac{1}{N} \sum_{i=1}^N |E_i - M_i|, \quad (7)$$

$$\text{MAPE} = \text{median} \left(\frac{|E_i - M_i|}{M_i} \times 100 \right), \quad (8)$$

式中, E 代表模型预测值; M 代表实测值; N 代表实测值与预测值匹配到的数量。

3 结果与分析

3.1 浮游植物类群 Chl a 浓度统计

利用实测的色素浓度数据和 CHEMTAX 软件计算得到的 8 类浮游植物的 Chl a 浓度的统计结果如表 2 所示。为避免 Chl a 浓度较低所带来的不确定性^[16], 研究将浮游植物类群 Chl a 浓度的阈值设为 0.001 mg/m^3 , 即当某一类群 Chl a 浓度小于该阈值时, 样本被剔除。由于 8 类浮游植物并不是在每个站位都出现, 因此, 各浮游植物类群 Chl a 浓度的样本数量并不相同, 其中观测到蓝藻的站位个数最多(465 个), 观测到绿藻的站位个数最少(216 个)。不同浮游植物类群的平均 Chl a 浓度也有较大差别, 其中硅藻的 Chl a 浓度平均值在 8 类浮游植物中最高 (1.13 mg/m^3), 甲藻次之 (0.29 mg/m^3), 定鞭藻的 Chl a 浓度平均值 (0.08 mg/m^3) 最低。在实测数据中, 8 个站位的硅藻 Chl a 浓度大于 10 mg/m^3 , 而甲藻 Chl a 浓度大于 10 mg/m^3 的站位仅有 1 个。青绿藻、隐藻、金藻、定鞭藻的 Chl a 浓度相对较低, 仅有较少的站位大于 1 mg/m^3 , 分别为 4 个、5 个、7 个、3 个。虽然绿藻 Chl a 浓度大于 1 mg/m^3

表 2 8 类浮游植物叶绿素 a 浓度统计特征

Table 2 Statistical characteristics of Chl a concentration of eight phytoplankton groups

浮游植物	站位数 (N)	平均值/ ($\text{mg}\cdot\text{m}^{-3}$)	标准差/ ($\text{mg}\cdot\text{m}^{-3}$)	最大值/ ($\text{mg}\cdot\text{m}^{-3}$)
青绿藻	450	0.15	0.27	4.45
甲藻	380	0.29	0.89	10.91
隐藻	410	0.17	0.22	1.31
绿藻	216	0.22	0.77	7.45
蓝藻	465	0.24	0.68	9.85
硅藻	446	1.13	2.10	15.23
金藻	360	0.11	0.23	2.51
定鞭藻	458	0.08	0.14	1.74

的站位有 10 个, 但大多数站位的浓度低于 0.3 mg/m^3 。

将各浮游植物类群 Chl a 浓度与对应站位的实测遥感反射率进行匹配, 8 类浮游植物匹配到的点对数量为: 青绿藻 119 对、甲藻 90 对、隐藻 107 对、绿藻 73 对、蓝藻 128 对、硅藻 116 对、金藻 92 对、定鞭藻 128 对。按照 4 : 1 的比例将上述匹配数据随机分成建模数据集和验证数据集。

3.2 不同大气校正算法验证与比较

将影像大气校正后得到的遥感反射率与匹配的实测值进行比较, 对 3 种大气校正方法进行验证, 结果见表 3。由于有些大气校正算法会产生无效值, 因此, 不同大气校正算法得到的匹配点数量不同。经 C2RCC 校正后的影像具有最多的匹配点(48 个), 经 POLYMER 校正后的影像具有 46 个匹配点, 而经 MUMM 校正后的影像只包含 27 个匹配点。在 400 nm、412.5 nm、442.5 nm、490 nm 波段, C2RCC 大气校正效果优于另外两种算法, MUMM 效果最差。在其余波段, POLYMER 相比于另外两种大气校正算法效果最差 ($\text{RMSE} \geq 0.0032 \text{ sr}^{-1}$)。如图 3 所示, 3 种大气校正方法在红光和近红外波段, 即 665 nm、673.75 nm、681.25 nm、708.75 nm、753.75 nm、778.75 nm 波段处易低估遥感反射率。经 MUMM 算法校正后的遥感反射率在 400 nm、412.5 nm、445.5 nm 易出现高估现象, 且与实测值之间差别较大 ($\text{MAPE} \geq 140\%$)。除少数点外, POLYMER 大气校正后的各波段遥感反射率均比对应的实测值低, 且在 708.75 nm、753.75 nm、778.75 nm 波段处, POLYMER 大气校正后的数据存在负值。

受陆源和人类活动的影响, 中国东部沿海上空存在多种气溶胶类型, 其中城市型气溶胶和混合型气溶胶是该区域主要的气溶胶类型^[40]。C2RCC 和 MUMM 大气校正算法中的气溶胶模型在建立过程中, 很少考虑到中国东部沿海气溶胶的光学特性^[41-42], 因而造成对气溶胶信号贡献的错误估计^[43-44]。此外, MUMM 大气校正算法遵循兴趣区内 765 nm 和 865 nm 波段离水反射比为定值的假设, 而已有研究表明, 在浑浊水域两个波段的离水反射存在明显的变化^[45]。POLYMER 大气校正中的离水反射模型仅考虑 Chl a 浓度和后向散射系数, 对于浑浊水域的其他水色组分, 模型假设有色可溶有机物随 Chl a 浓度变化, 且未考虑非藻类颗粒物的吸收, 而该假设并不适用于水色组分复杂的中国东部海域。

3.3 不同浮游植物类群反演模型验证与比较

利用建模数据集, 即由 CHEMTAX 计算得到的 8 类浮游植物 Chl a 浓度及其对应的实测遥感反射率,

表 3 3 种大气校正算法的精度评价

Table 3 Accuracy evaluation of three atmospheric correction algorithms

波长/nm	RMSE/(10 ⁻³ sr ⁻¹)			MAE/(10 ⁻³ sr ⁻¹)			MAPE/%		
	C2RCC	POLYMER	MUMM	C2RCC	POLYMER	MUMM	C2RCC	POLYMER	MUMM
400	4.13	4.41	10.79	3.21	3.67	9.83	41	44	213
412.5	3.99	4.36	7.68	3.01	3.53	6.69	36	41	141
442.5	4.10	4.64	6.81	3.21	3.68	6.02	36	39	140
490	4.79	4.97	4.90	3.74	3.82	4.43	40	37	89
510	4.72	5.28	4.21	3.65	4.12	3.74	40	41	70
560	4.34	5.31	3.53	3.31	4.03	3.04	33	43	56
620	3.22	4.21	3.05	2.24	3.21	2.44	49	66	81
665	3.21	3.75	2.71	2.21	2.82	2.05	61	72	84
673.75	3.23	3.64	2.62	2.29	2.72	1.98	63	71	80
681.25	3.21	3.64	2.60	2.29	2.75	1.91	64	73	83
708.75	3.12	4.45	3.21	2.21	3.62	1.93	70	116	69
753.75	3.21	3.24	2.38	2.30	2.40	1.49	88	94	81
778.75	3.04	3.23	2.40	2.18	2.40	1.44	86	99	90

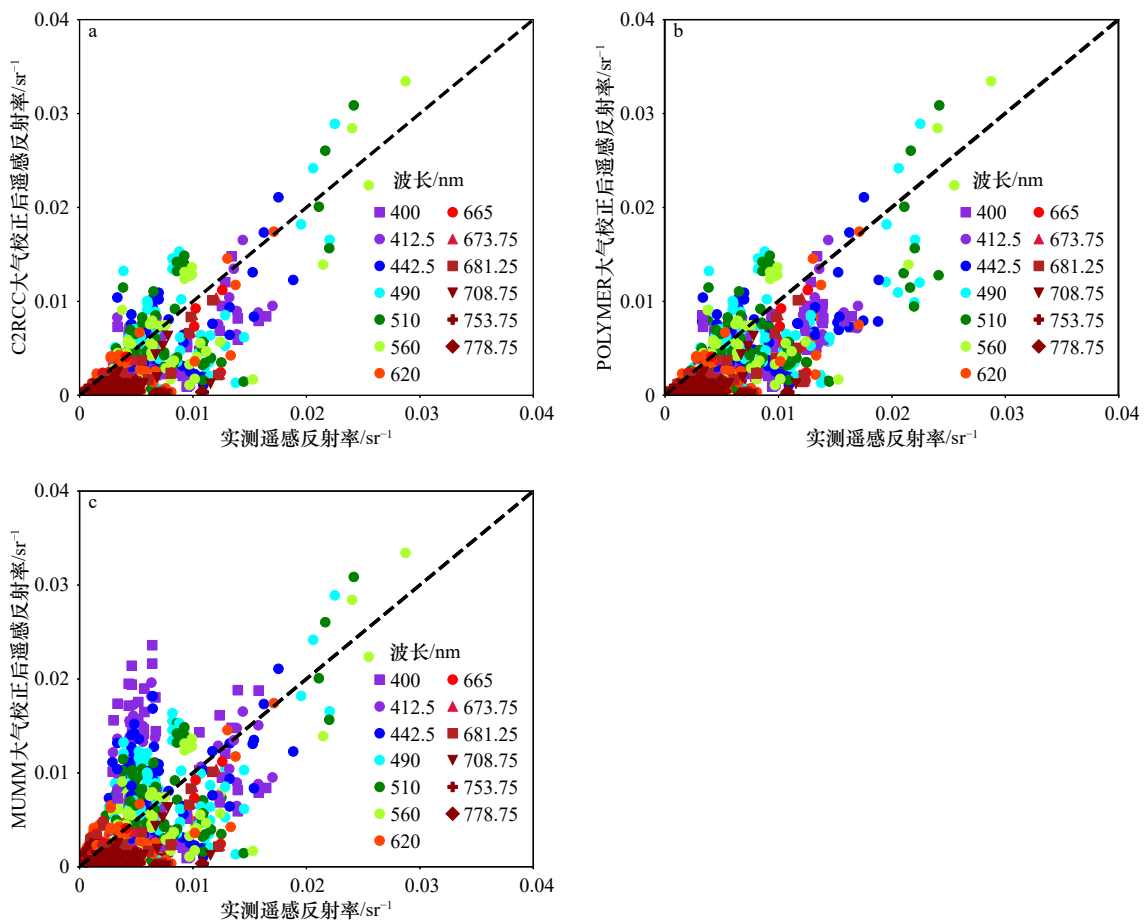


图 3 3 种大气校正方法获得的遥感反射率与实测遥感反射率验证结果

Fig. 3 Validation of remote sensing reflectance obtained by three atmospheric correction methods with *in-situ* remote sensing reflectance

采用 3 种方法(2.6 节所述)构建浮游植物类群遥感反演模型,并利用验证数据集进行模型精度验证。由于

本研究中 Chl *a* 浓度的数值跨越多个量级,因此在回归建模以及之后的精度评价中,全部采用 Chl *a* 浓度

对数(以 10 为底)形式。

采用波段组合法进行回归建模时,对于各浮游植物类群选择的最优波段组合及对应的回归方程见表 4。硅藻 Chl *a* 浓度对数与对应的遥感反射率波段组合值的相关性为 0.76,使用波段组合法建立的反演模型相比于其他类群浮游植物在验证集上的精度较高, R^2 为 0.55(表 5)。青绿藻、甲藻和隐藻的

Chl *a* 浓度对数与其对应的遥感反射率波段组合值的相关性大于 0.5,基于波段组合法的反演模型在实测数据组成的验证集上的 R^2 分别为: 0.29、0.31、0.31。对于绿藻、蓝藻、金藻和定鞭藻,根据对应的波段组合建立的模型反演精度较差($R^2 < 0.2$),可能与 Chl *a* 浓度的对数和其相关性最高的波段组合值相关性较低有关(相关系数小于 0.42)。

表 4 8类浮游植物叶绿素 *a* 浓度反演模型使用的波段组合形式及回归方程

Table 4 Band combinations form and regression equations used in Chl *a* concentration inversion model of eight phytoplankton groups

浮游植物	波段组合	与Chl <i>a</i> 浓度对数的相关系数	回归方程
青绿藻	$X = \frac{R_{rs}(442.5) + R_{rs}(620)}{R_{rs}(560)}$	0.53	$\lg C_p = -0.74X + 0.09$
甲藻	$X = \frac{R_{rs}(442.5)}{R_{rs}(510)} - \frac{R_{rs}(442.5)}{R_{rs}(560)}$	0.54	$\lg C_p = -7.83X^3 - 0.21X^2 + 3.97X - 1.05$
隐藻	$X = \frac{R_{rs}(442.5) + R_{rs}(490)}{R_{rs}(510)}$	0.61	$\lg C_p = -2.10X + 2.87$
绿藻	$X = \frac{R_{rs}(560)}{R_{rs}(442.5) - R_{rs}(620)}$	0.37	$\lg C_p = 2.2 \times 10^{-4}X^2 - 3.8 \times 10^{-2}X - 1.25$
蓝藻	$X = \frac{R_{rs}(412.5)}{R_{rs}(442.5) - R_{rs}(620)}$	0.40	$\lg C_p = -0.006X - 0.951$
硅藻	$X = \frac{R_{rs}(490) + R_{rs}(620)}{R_{rs}(560)}$	0.76	$\lg C_p = -1.93X + 2.75$
金藻	$X = R_{rs}(665) - R_{rs}(673.75)$	0.34	$\lg C_p = -477.853.92X^2 - 1.569.03X - 1.46$
定鞭藻	$X = \frac{R_{rs}(490) - R_{rs}(510)}{R_{rs}(560)}$	0.42	$\lg C_p = -2.143X - 1.285$

表 5 列出了 3 种方法建立的遥感反演模型在估算 8 类浮游植物 Chl *a* 浓度的精度评价结果。对于各浮游植物类群,使用 SVD+XGBoost 方法建立的模型精度均为最优,其中甲藻、硅藻 Chl *a* 浓度反演模型的 R^2 大于 0.7,青绿藻、隐藻、定鞭藻的 Chl *a* 浓度反演模型的 R^2 大于 0.47;相比于波段组合法和 SVD+MLR 建立的绿藻、蓝藻和金藻 Chl *a* 浓度遥感反演模型,使用 SVD+XGBoost 建立的反演模型的精度有显著提高。说明 SVD+XGBoost 回归模型具有一定的优势。

对比 3 种建模方法、8 类浮游植物,硅藻的估算精度均较好($R^2 \geq 0.55$),可能是因为相对于其他浮游植物,其 Chl *a* 浓度较高,光学信号更容易捕捉。相比之下,绿藻的估算精度在 3 种建模方法建立的模型中均表现较差($R^2 \leq 0.32$),特别是基于 SVD+MLR 法的绿藻 Chl *a* 浓度反演模型,验证结果 R^2 小于 0。波段组合法和 SVD+MLR 法建立的绿藻、蓝藻和金藻 Chl *a* 浓度反演模型的验证结果 $R^2 \leq 0.2$,说明波段组合法和 SVD+MLR 法可能不适合绿藻、蓝藻、金藻 Chl *a* 浓度反演。

由图 4 至图 6 的散点图可以看出,硅藻和青绿藻

的验证结果在实测 Chl *a* 浓度低于 0.01 mg/m³时均存在一个高估的样本点,且基于 3 种建模方法的反演模型对该样本点的预测值与实测值差别均较大,这可能是由于本研究所采集的青绿藻和硅藻的 Chl *a* 浓度小于 0.01 mg/m³的样本点较少,使得模型在构建过程中无法很好地覆盖到 Chl *a* 浓度小于 0.01 mg/m³的情况。

3.4 基于卫星影像的浮游植物类群遥感反演模型验证

由 3.2 节大气校正算法验证结果可知,C2RCC 大气校正方法在精度上具有一定的优势,本研究选择经 C2RCC 大气校正后的影像进行进一步研究。3.3 节结果显示,利用 3 种建模方法建立的反演模型在估算甲藻、隐藻、硅藻 Chl *a* 浓度时精度较高,而其他类群浮游植物 Chl *a* 浓度的反演模型仍有待进一步提高,因此,本研究仅选取甲藻、隐藻、硅藻 Chl *a* 浓度反演模型进行卫星应用验证。选择完全独立于建模数据集的实测浮游植物类群 Chl *a* 浓度,在经大气校正后的影像上取对应站位的遥感反射率(时空匹配算法仍按照在时间上选择卫星过境前后 3 h,空间上选择 3×3 的窗口平均值),利用获取到的遥感反射率和建立的反演模型估算 3 类浮游植物 Chl *a* 浓度。

表 5 8 类浮游植物叶绿素 *a* 浓度遥感反演模型的精度评价
Table 5 Accuracy of inversion model for Chl *a* concentration of eight phytoplankton groups

浮游植物类群	建模方法	R^2	MAE	RMSE	MAPE/%
青绿藻	BC	0.29	0.33	0.44	29
	SVD+MLR	0.38	0.33	0.42	33
	SVD+XGBoost	0.47	0.30	0.38	26
甲藻	BC	0.31	0.63	0.76	40
	SVD+MLR	0.42	0.62	0.69	52
	SVD+XGBoost	0.77	0.34	0.44	24
隐藻	BC	0.31	0.48	0.60	41
	SVD+MLR	0.43	0.42	0.55	25
	SVD+XGBoost	0.49	0.38	0.52	29
绿藻	BC	0.13	0.68	0.80	46
	SVD+MLR	-0.06	0.77	0.89	56
	SVD+XGBoost	0.32	0.56	0.71	26
蓝藻	BC	0.15	0.35	0.52	24
	SVD+MLR	0.06	0.39	0.55	29
	SVD+XGBoost	0.44	0.31	0.43	21
硅藻	BC	0.55	0.41	0.57	71
	SVD+MLR	0.64	0.39	0.52	63
	SVD+XGBoost	0.73	0.32	0.45	48
金藻	BC	-0.03	0.54	0.71	27
	SVD+MLR	0.06	0.52	0.68	27
	SVD+XGBoost	0.36	0.46	0.56	25
定鞭藻	BC	0.14	0.37	0.46	24
	SVD+MLR	0.37	0.32	0.40	20
	SVD+XGBoost	0.51	0.26	0.35	13

由图 7 至图 9 可以看出, 当将 3 种建模方法建立的反演模型应用到影像上时, 各浮游植物类群 Chl *a* 浓度的反演精度都有所下降。特别是甲藻与隐藻, 基于 3 种方法建立的甲藻 Chl *a* 浓度反演模型精度的决定系数均小于 0。基于波段组合法建立的甲藻 Chl *a* 浓度反演模型, 虽然有部分预测值和实测值处于 1 : 1 线上, 但是部分样本点存在严重高估的现象。对于隐藻, 3 种建模方法得到的反演值大部分都低于实测值。对于硅藻, 基于 SVD+MLR 法建立的模型的反演精度最高 ($R^2=0.56$, MAPE=45%), 使用波段组合法建立的模型反演精度也相对较好 ($R^2=0.44$, MAPE=58%)。

实测验证数据集上的结果显示, 基于多个波段进

行奇异值分解后建立的反演模型都有较高的反演精度, 但是将模型应用到卫星影像上时, 精度下降明显, 特别是基于 SVD+XGBoost 法建立的反演模型。这可能是由于卫星大气校正后得到的遥感反射率和真实值存在差距, 大气校正的准确性对使用多个波段进行奇异值分解后建立的反演模型的估算精度的影响较大。

3.5 基于卫星影像的硅藻类群遥感反演应用

将在影像上表现较好, 基于 SVD+MLR 法构建的硅藻 Chl *a* 浓度反演模型应用到 2020 年 5 月和 2020 年 8 月过境中国东部海域所有 OLCI 影像上 (包括 Sentinel-3A 和 Sentinel-3B), 得到 2020 年 5 月和 2020 年 8 月的月平均硅藻 Chl *a* 浓度空间分布, 如图 10 所示。硅藻 Chl *a* 浓度的高值区域集中在近岸海域, 低值区域主要分布在黄海中部和东海外海以及长江口、杭州湾、苏北沿岸等悬浮泥沙含量较多的浑浊水域。在渤海南部, 2020 年 8 月的硅藻 Chl *a* 浓度存在一定的高值。据《2020 中国海洋灾害公报》^[46] 显示, 该区域在 2020 年 8 月份暴发过以硅藻浮游植物为主导的赤潮灾害事件。在苏北沿岸区域, 由于受到高浓度悬浮泥沙的影响, 导致离岸较近区域硅藻 Chl *a* 浓度较低, 相比于 2020 年 8 月, 2020 年 5 月的低值范围更大。在长江口外部海域, 由于长江径流带来的丰富营养盐, 该区域相较于东海其他区域硅藻 Chl *a* 浓度较高。由于海表温度的上升, 2020 年 5 月的硅藻 Chl *a* 浓度月均值要高于 2020 年 8 月。从时空分布来看, 上述结果与已有研究中针对中国东部海域 Chl *a* 浓度时空变化规律的研究结果相一致^[47-49]。

4 结论

本研究利用实测的遥感反射率和 8 类浮游植物 Chl *a* 浓度, 对比了 3 种建模方法建立的遥感反演模型的精度; 评估了 3 种大气校正方法在中国东部海域的适用性以及各类群浮游植物 Chl *a* 浓度反演模型的影响。主要研究结论如下:

(1) 相比于传统的回归模型构建方法, SVD+XGBoost 方法具有一定的优势。实测数据的验证结果显示, 基于 SVD+XGBoost 法的 8 类浮游植物 Chl *a* 浓度反演模型的精度高于其他两种建模方法。特别是对于硅藻和甲藻浮游植物, 基于 SVD+XGBoost 法的反演模型在验证集上具有较好的精度 ($R^2>0.70$), 而其他 6 类浮游植物 Chl *a* 浓度反演模型精度相对较低 ($R^2\leq 0.51$), 仍有待进一步研究。本研究中的浮游植物类群 Chl *a* 浓度反演算法是经验性的, 今后将考虑

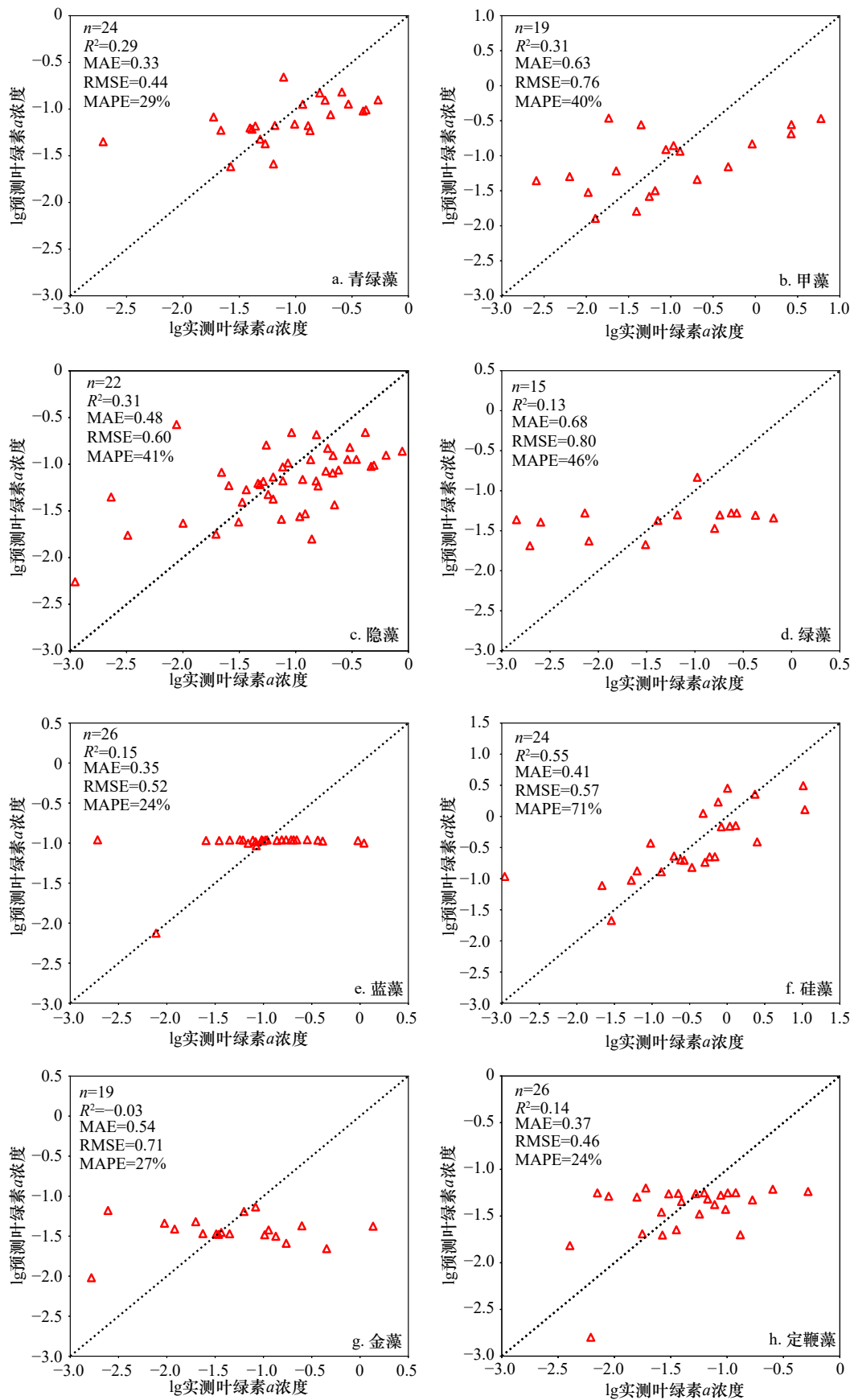


图4 基于波段组合法建立的8类浮游植物叶绿素a浓度反演模型的验证结果(叶绿素a浓度单位: mg/m³)

Fig. 4 Validation results of Chl *a* concentration inversion model of eight phytoplankton groups based on band combination method (the unit of Chl *a* concentration is mg/m³)

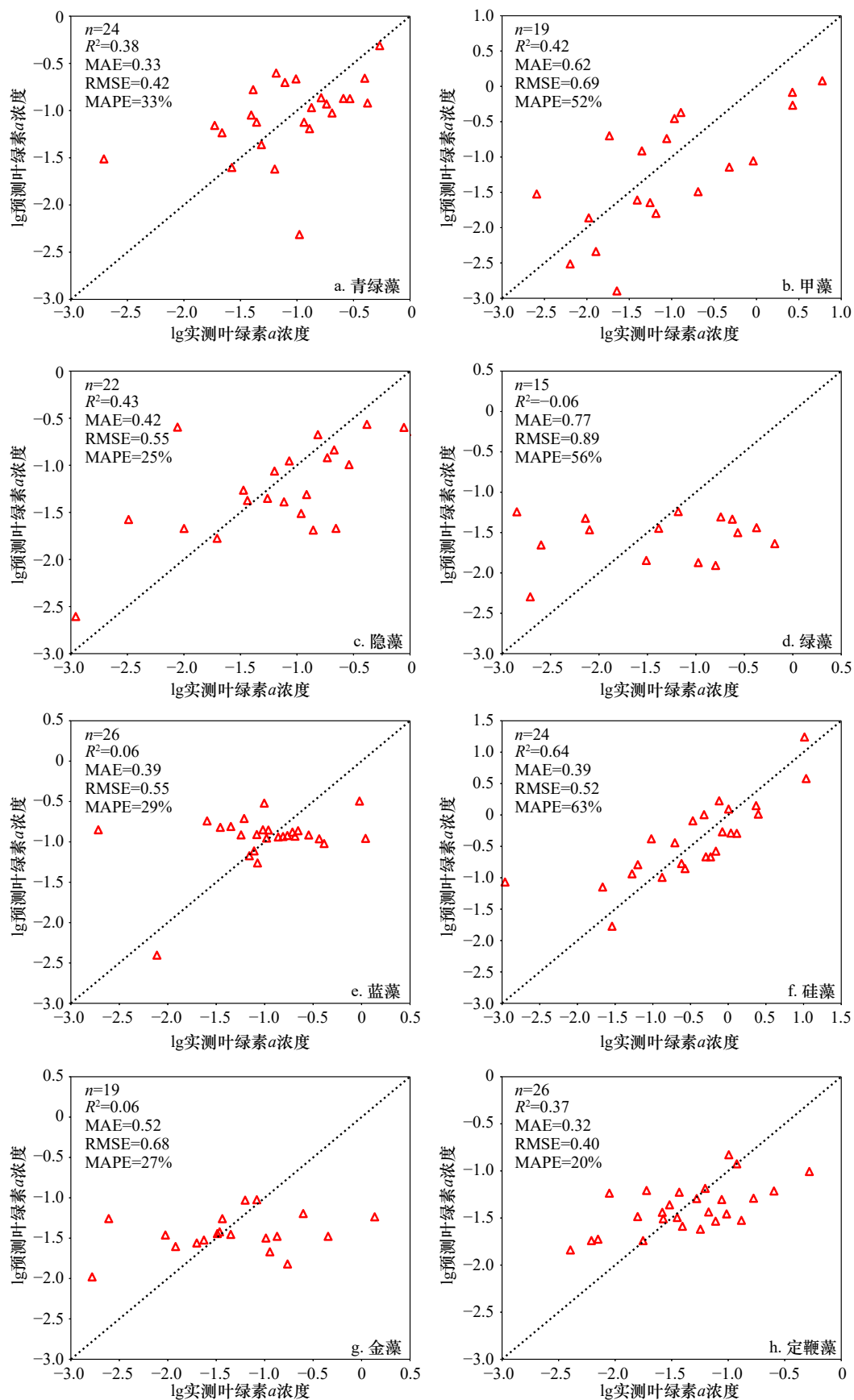
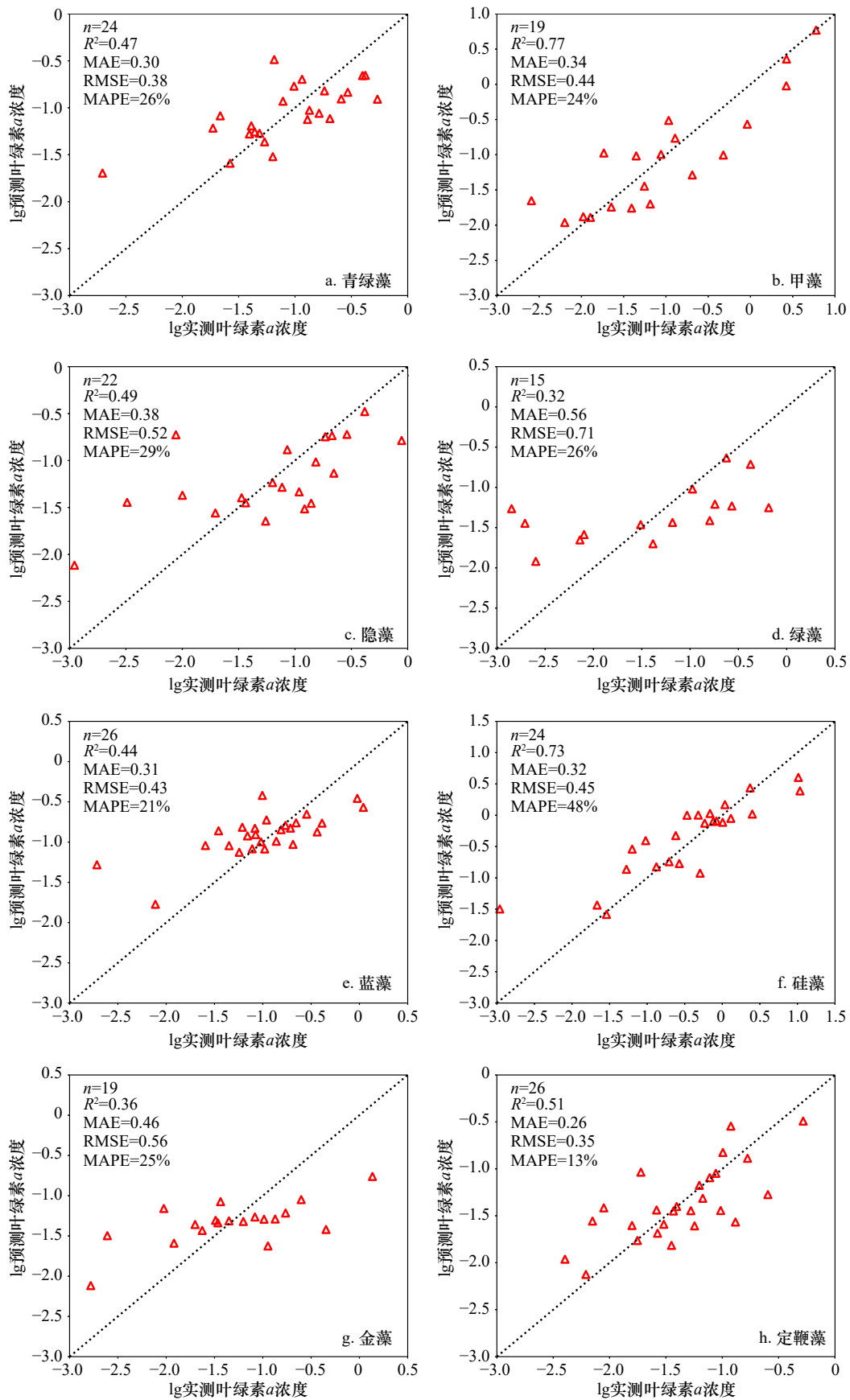


图 5 基于 SVD+MLR 建立的 8 类浮游植物叶绿素 *a* 浓度反演模型的验证结果(叶绿素 *a* 浓度单位: mg/m^3)

Fig. 5 Validation results of Chl *a* concentration inversion model of eight phytoplankton groups based on SVD+MLR method (the unit of Chl *a* concentration is mg/m^3)

图6 基于SVD+XGBoost建立的8类浮游植物叶绿素*a*浓度反演模型的验证结果(叶绿素*a*浓度单位: mg/m^3)Fig. 6 Validation results of Chl *a* concentration inversion model of eight phytoplankton groups based on SVD+XGBoost method (the unit of Chl *a* concentration is mg/m^3)

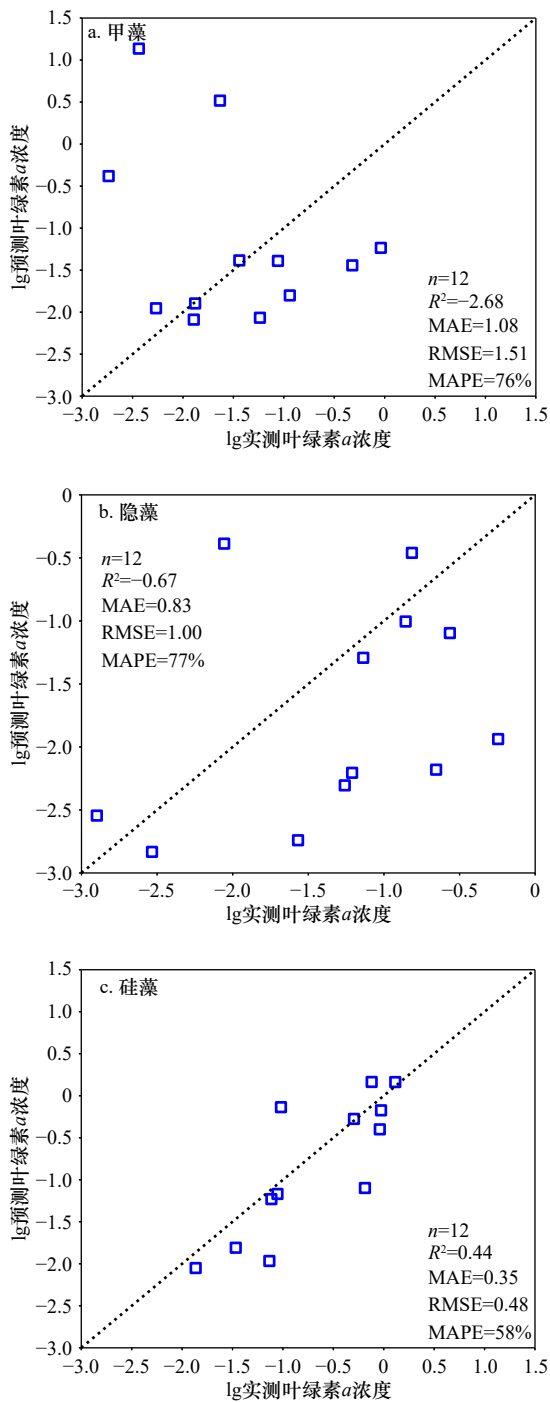


图7 基于波段组合法的浮游植物类群叶绿素 a 浓度卫星反演验证 (叶绿素 a 浓度单位: mg/m^3)
Fig. 7 Validation of satellite inversion results of phytoplankton group Chl a concentration based on band combination method (the unit of Chl a concentration is mg/m^3)

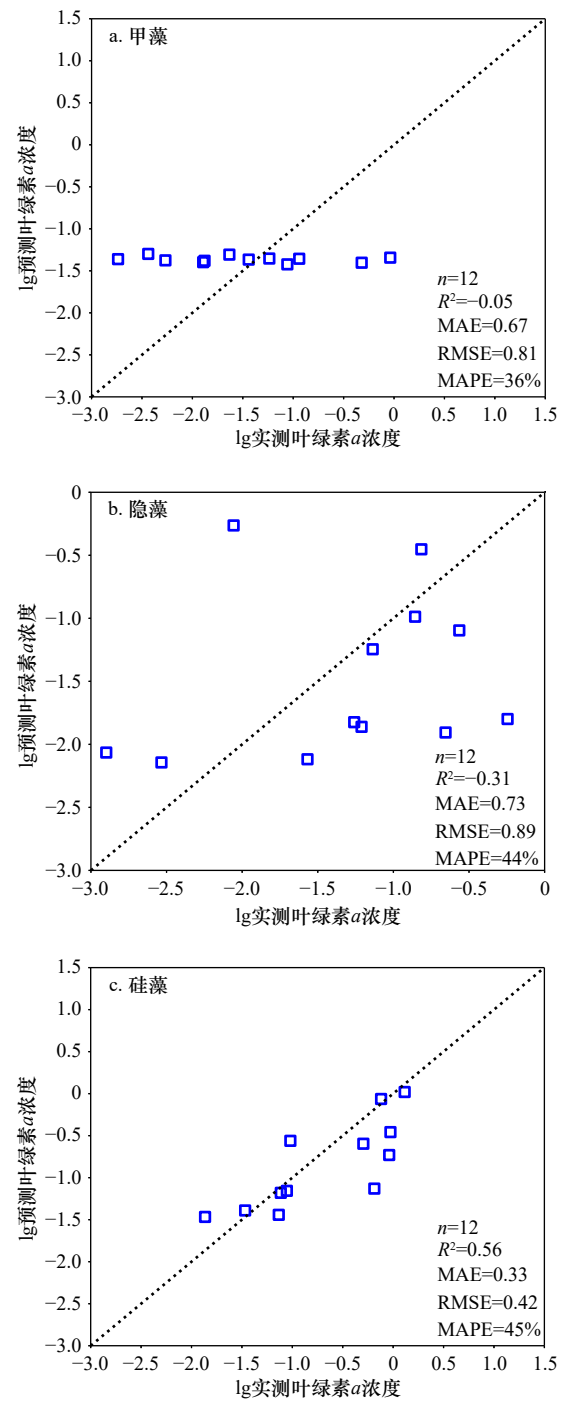


图8 基于SVD+MLR的浮游植物类群叶绿素 a 浓度卫星反演验证 (叶绿素 a 浓度单位: mg/m^3)
Fig. 8 Validation of satellite inversion results of phytoplankton group Chl a concentration based on SVD+MLR method (the unit of Chl a concentration is mg/m^3)

从光学机理出发, 深入探究不同浮游植物类群对遥感反射率贡献的差异。

(2) 在各波段, C2RCC 相对于其他两种大气校正算法 (MUMM、POLYMER) 有较好的表现 (RMSE 均小于 0.0048 sr^{-1}), 且能获得更多的有效值, 但与实

测值之间仍有差距, 需要继续发展适用于中国东部海域的高精度的大气校正算法。

(3) 受大气校正精度的影响, 将浮游植物类群 Chl a 浓度反演模型应用到 OLCI 影像上时, 模型估算精度有所下降, 特别是基于 SVD+XGBoost 建立的反演模

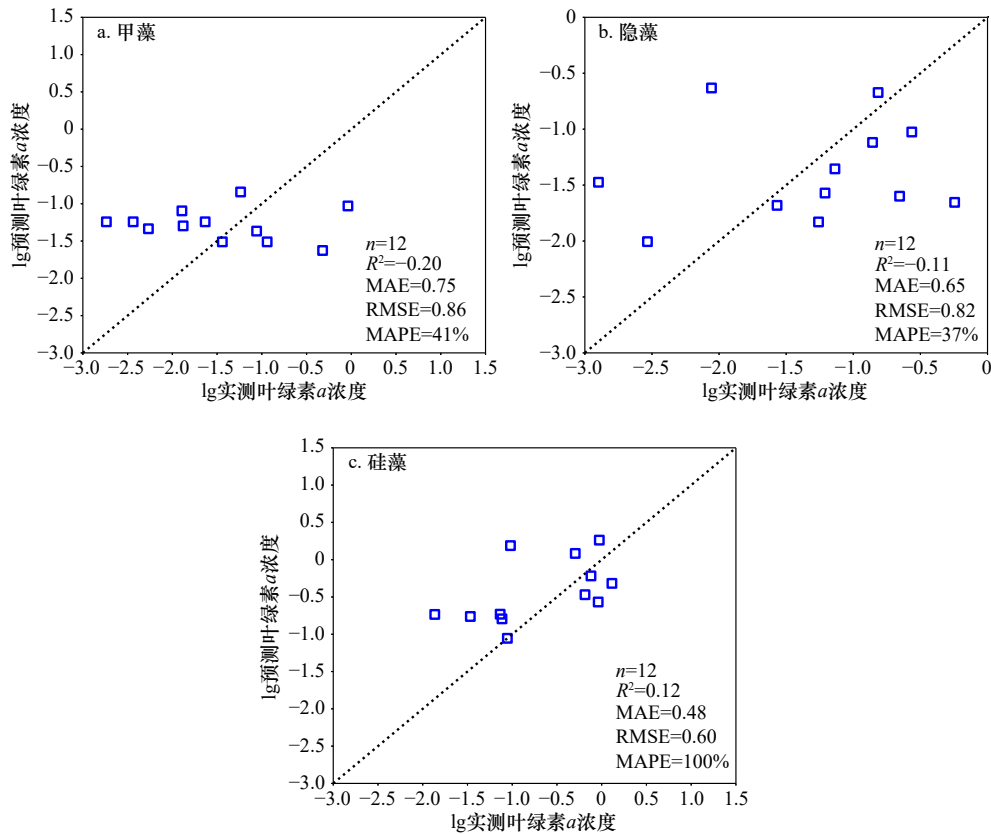


图9 基于SVD+XGBoost的浮游植物类群叶绿素a浓度卫星反演验证(叶绿素a浓度单位: mg/m^3)

Fig. 9 Validation of satellite inversion results of phytoplankton group Chl a concentration based on SVD+XGBoost method (the unit of Chl a concentration is mg/m^3)

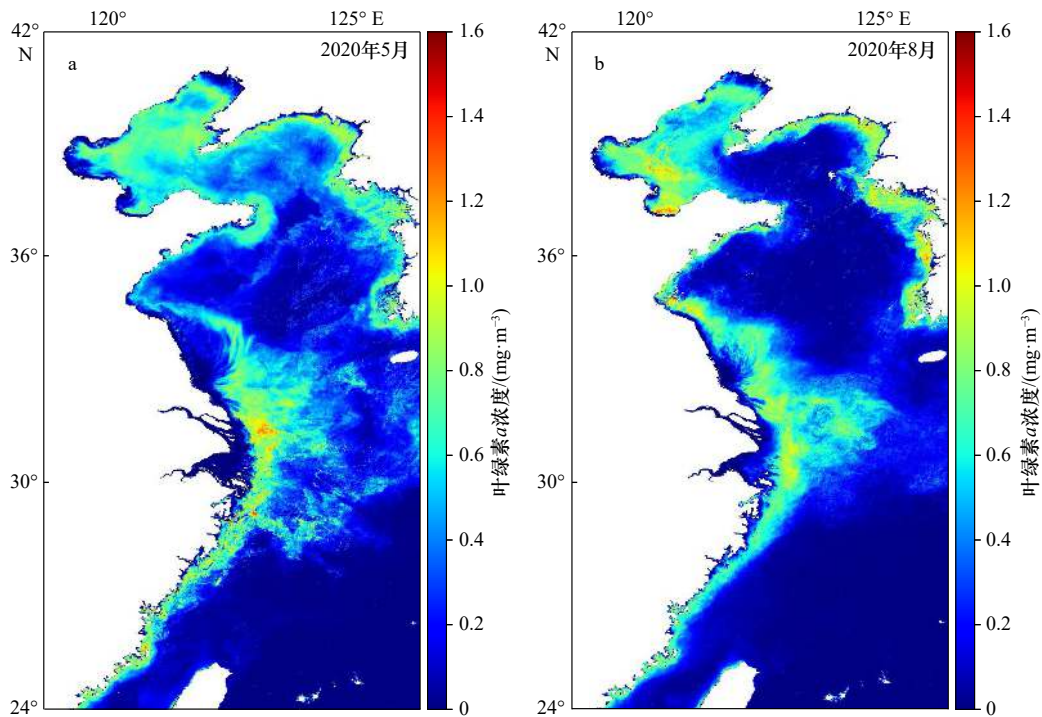


图10 2020年5月和2020年8月中国东部海域硅藻叶绿素a浓度空间分布

Fig. 10 The spatial distribution of diatom Chl a concentration in the eastern China seas in May 2020 and August 2020

型, 卫星反演验证结果 R^2 均小于 0.2。相比之下, 利用波段组合法、SVD+MLR 法建立的硅藻 Chl a 浓度

反演模型在应用到卫星影像上时仍保持一定的精度 (R^2 分别为 0.44、0.56), 具有一定的鲁棒性。

参考文献:

- [1] Field C B, Behrenfeld M J, Randerson J T, et al. Primary production of the biosphere: integrating terrestrial and oceanic components[J]. *Science*, 1998, 281(5374): 237–240.
- [2] Behrenfeld M J. Climate-mediated dance of the plankton[J]. *Nature Climate Change*, 2014, 4(10): 880–887.
- [3] Le Quéré C, Harrison S P, Prentice I C, et al. Ecosystem dynamics based on plankton functional types for global ocean biogeochemistry models[J]. *Global Change Biology*, 2005, 11(11): 2016–2040.
- [4] Tréguer P, Bowler C, Moriceau B, et al. Influence of diatom diversity on the ocean biological carbon pump[J]. *Nature Geoscience*, 2018, 11(1): 27–37.
- [5] Nair A, Sathyendranath S, Platt T, et al. Remote sensing of phytoplankton functional types[J]. *Remote Sensing of Environment*, 2008, 112(8): 3366–3375.
- [6] McClain C R. A decade of satellite ocean color observations[J]. *Annual Review of Marine Science*, 2009, 1(1): 19–42.
- [7] Tilstone G H, Pardo S, Dall’Olmo G, et al. Performance of ocean colour Chlorophyll *a* algorithms for Sentinel-3 OLCI, MODIS-Aqua and Suomi-VIIRS in open-ocean waters of the Atlantic[J]. *Remote Sensing of Environment*, 2021, 260: 112444.
- [8] Kostadinov T S, Cabré A, Vedantham H, et al. Inter-comparison of phytoplankton functional type phenology metrics derived from ocean color algorithms and Earth System Models[J]. *Remote Sensing of Environment*, 2017, 190: 162–177.
- [9] Bracher A, Bouman H A, Brewin R J W, et al. Obtaining phytoplankton diversity from ocean color: A scientific roadmap for future development[J]. *Frontiers in Marine Science*, 2017, 4: 55.
- [10] IOCCG. Phytoplankton functional types from space[R]. Dartmouth: International Ocean-Colour Coordinating Group, 2014.
- [11] Mouw C B, Hardman-Mountford N J, Alvain S, et al. A consumer’s guide to satellite remote sensing of multiple phytoplankton groups in the global ocean[J]. *Frontiers in Marine Science*, 2017, 4: 41.
- [12] Sun Xuerong, Shen Fang, Liu Dongyan, et al. In situ and satellite observations of phytoplankton size classes in the entire Continental Shelf Sea, China[J]. *Journal of Geophysical Research: Oceans*, 2018, 123(5): 3523–3544.
- [13] 李楠, 孙德勇, 环宇, 等. 黄渤海浮游植物种群比吸收光谱的确定及其应用[J]. *光学学报*, 2020, 40(6): 060100.
Li Nan, Sun Deyong, Huan Yu, et al. Determination and application of specific absorption spectra of phytoplankton species in Yellow Sea and Bohai Sea[J]. *Acta Optica Sinica*, 2020, 40(6): 060100.
- [14] O’Reilly J E, Maritorena S, Mitchell B G, et al. Ocean color chlorophyll algorithms for SeaWiFS[J]. *Journal of Geophysical Research: Oceans*, 1998, 103(C11): 24937–24953.
- [15] Hu Chuanmin, Lee Z, Franz B. Chlorophyll *a* algorithms for oligotrophic oceans: A novel approach based on three-band reflectance difference[J]. *Journal of Geophysical Research: Oceans*, 2012, 117(C1): C010011.
- [16] Xi Hongyan, Losa S N, Mangin A, et al. Global retrieval of phytoplankton functional types based on empirical orthogonal functions using CMEMS GlobColour merged products and further extension to OLCI data[J]. *Remote Sensing of Environment*, 2020, 240: 111704.
- [17] Bracher A, Xi Hongyan, Dinter T, et al. High resolution water column phytoplankton composition across the Atlantic Ocean from ship-towed vertical undulating radiometry[J]. *Frontiers in Marine Science*, 2020, 7: 235.
- [18] Cao Zhigang, Ma Ronghua, Duan Hongtao, et al. A machine learning approach to estimate chlorophyll-*a* from Landsat-8 measurements in inland lakes[J]. *Remote Sensing of Environment*, 2020, 248: 111974.
- [19] Liu Huizeng, Li Qingquan, Bai Yan, et al. Improving satellite retrieval of oceanic particulate organic carbon concentrations using machine learning methods[J]. *Remote Sensing of Environment*, 2021, 256: 112316.
- [20] Kyrilyuk D, Kratzer S. Evaluation of Sentinel-3A OLCI products derived using the case-2 regional coastcolour processor over the Baltic Sea[J]. *Sensors*, 2019, 19(16): 3609.
- [21] He X Q, Bai Yan, Pan Delu, et al. Satellite views of the seasonal and interannual variability of phytoplankton blooms in the eastern China seas over the past 14 yr (1998–2011)[J]. *Biogeosciences*, 2013, 10(7): 4721–4739.
- [22] 宋金明. 中国近海生态系统碳循环与生物固碳[J]. *中国水产科学*, 2011, 18(3): 703–711.
Song Jinming. Carbon cycling processes and carbon fixed by organisms in China marginal seas[J]. *Journal of Fishery Sciences of China*, 2011, 18(3): 703–711.
- [23] 陈建忠, 葛建忠, Richard B. 东海中部浮游生态系统季节变化的数值模拟[J]. *华东师范大学学报(自然科学版)*, 2019(6): 153–168.
Chen Jianzhong, Ge Jianzhong, Richard B. Numerical simulation of pelagic ecosystem’s seasonal variation in the central East China Sea[J]. *Journal of East China Normal University (Natural Science)*, 2019(6): 153–168.
- [24] 黄邦钦, 肖武鹏, 柳欣. 中国边缘海浮游植物群落时空格局与演变趋势[J]. *厦门大学学报(自然科学版)*, 2021, 60(2): 390–397.
Huang Bangqin, Xiao Wupeng, Liu Xin. Spatial-temporal distributions and successional patterns of phytoplankton communities in the Chinese marginal seas[J]. *Journal of Xiamen University (Natural Science)*, 2021, 60(2): 390–397.
- [25] Sokoletsky L G, Shen Fang. Optical closure for remote-sensing reflectance based on accurate radiative transfer approximations: the case of the Changjiang (Yangtze) River Estuary and its adjacent coastal area, China[J]. *International Journal of Remote Sensing*, 2014, 35(11/12): 4193–4224.
- [26] Zhang H K, Roy D P, Yan Lin, et al. Characterization of Sentinel-2A and Landsat-8 top of atmosphere, surface, and nadir BRDF adjust-

- ted reflectance and NDVI differences[J]. *Remote Sensing of Environment*, 2018, 215: 482–494.
- [27] Catlett D, Siegel D A. Phytoplankton pigment communities can be modeled using unique relationships with spectral absorption signatures in a dynamic coastal environment[J]. *Journal of Geophysical Research: Oceans*, 2018, 123(1): 246–264.
- [28] Sun Xuerong, Shen Fang, Brewin Robert J W, et al. Twenty-year variations in satellite-derived chlorophyll *a* and phytoplankton size in the Bohai Sea and Yellow Sea[J]. *Journal of Geophysical Research: Oceans*, 2019, 124(12): 8887–8912.
- [29] 孙雪融. 东中国海浮游植物粒级结构定量遥感研究[D]. 上海: 华东师范大学, 2021
- Sun Xuerong. Remote sensing quantitative retrieval of phytoplankton size classes in eastern China seas[D]. Shanghai: East China Normal University, 2021.
- [30] Mackey M D, Mackey D J, Higgins H W, et al. CHEMTAX—A program for estimating class abundances from chemical markers: Application to HPLC measurements of phytoplankton[J]. *Marine Ecology Progress Series*, 1996, 144(1/3): 265–283.
- [31] Sun Deyong, Lai Wendian, Wang Shengqiang, et al. Synoptic relationships to estimate phytoplankton communities specific to sizes and species from satellite observations in coastal waters[J]. *Optics Express*, 2019, 27(16): A1156–A1172.
- [32] Donlon C, Berruti B, Buongiorno A, et al. The global monitoring for environment and security (GMES) Sentinel-3 mission[J]. *Remote Sensing of Environment*, 2012, 120: 37–57.
- [33] Brockmann C, Doerffer R, Peters M, et al. Evolution of the C2RCC neural network for sentinel 2 and 3 for the retrieval of ocean colour products in normal and extreme optically complex waters[C]// *Proceedings of the Living Planet Symposium*, [S.l.]: [s.n.], 2016.
- [34] Doerffer R., Schiller H. The MERIS Case 2 water algorithm[J]. *International Journal of Remote Sensing*, 2007, 28(3/4): 517–535.
- [35] Steinmetz F, Deschamps P Y, Ramon D. Atmospheric correction in presence of sun glint: application to MERIS[J]. *Optics Express*, 2011, 19(10): 9783–9800.
- [36] Steinmetz F, Ramon D. Sentinel-2 MSI and Sentinel-3 OLCI consistent ocean colour products using POLYMER[C]// *Proceedings of SPIE 10778 Remote Sensing of the Open and Coastal Ocean and Inland Waters*. Washington: SPIE, 2018: 107780E.
- [37] Gordon H R, Wang Menghua. Retrieval of water-leaving radiance and aerosol optical thickness over the oceans with SeaWiFS: a preliminary algorithm[J]. *Applied Optics*, 1994, 33(3): 443–452.
- [38] Ruddick K G, Ovidio F, Rijkeboer M. Atmospheric correction of SeaWiFS imagery for turbid coastal and inland waters[J]. *Applied Optics*, 2000, 39(6): 897–912.
- [39] Chen Tianqi, Guestrin C. XGBoost: A Scalable Tree Boosting System[C]// *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. California: Association for Computing Machinery, 2016: 785–794.
- [40] Tu Qiangang, Hao Zengzhou, Yan Yunwei, et al. Aerosol optical properties around the East China Seas based on AERONET measurements[J]. *Atmosphere*, 2021, 12(5): 642.
- [41] Zagolski F, Santer R, Aznay O. A new climatology for atmospheric correction based on the aerosol inherent optical properties[J]. *Journal of Geophysical Research: Atmospheres*, 2007, 112(D14): D14208.
- [42] Ahmad Z, Franz B A, McClain C R, et al. New aerosol models for the retrieval of aerosol optical thickness and normalized water-leaving radiances from the SeaWiFS and MODIS sensors over coastal regions and open oceans[J]. *Applied Optics*, 2010, 49(29): 5545–5560.
- [43] Wang Dian, Ma Ronghua, Xue Kun, et al. The assessment of Landsat-8 OLI atmospheric correction algorithms for inland waters[J]. *Remote Sensing*, 2019, 11(2): 169.
- [44] Shen Ming, Duan Hongtao, Cao Zhigang, et al. Sentinel-3 OLCI observations of water clarity in large lakes in eastern China: Implications for SDG 6.3. 2 evaluation[J]. *Remote Sensing of Environment*, 2020, 247: 111950.
- [45] Doron M, Bélanger S, Doxaran D, et al. Spectral variations in the near-infrared ocean reflectance[J]. *Remote Sensing of Environment*, 2011, 115(7): 1617–1631.
- [46] 自然资源部海洋预警监测司. 2020中国海洋灾害公报[R]. 北京: 自然资源部海洋预警监测司, 2021
- Department of Marine Early Warning and Monitoring Ministry of Natural Resources. Bulletin of China marine disaster, 2020[R]. Beijing: Department of Marine Early Warning and Monitoring, Ministry of Natural Resources, 2021.
- [47] 江杉, 王益澄, 马仁锋. 中国东海叶绿素浓度变化分析及其海水温度响应[J]. *测绘通报*, 2020(6): 39–44.
- Jiang Shan, Wang Yicheng, Ma Renfeng. Change conalysis of chlorophyll concentration in the East China Sea and its response to seawater temperature[J]. *Bulletin of Surveying and Mapping*, 2020(6): 39–44.
- [48] 陈楠生, 陈阳. 中国海洋浮游植物和赤潮物种的生物多样性研究进展(二): 东海[J]. *海洋与湖沼*, 2021, 52(2): 363–38.
- Chen Nansheng, Chen Yang. Advances in the study of biodiversity of phytoplankton and red tide species in China (II): the East China Sea[J]. *Oceanologia et Limnologia Sinica*, 2021, 52(2): 363–384.
- [49] 田洪阵, 刘沁萍, Goes J I, 等. 基于2002–2018年MODIS数据的黄海叶绿素 a 时空变化研究[J]. *海洋通报*, 2020, 39(1): 101–110.
- Tian Hongzhen, Liu Qiping, Goes J I, et al. Temporal and spatial changes in chlorophyll *a* concentrations in the Yellow Sea from 2002 to 2018 based on MODIS data[J]. *Marine Science Bulletin*, 2020, 39(1): 101–110.

Remote sensing retrieval of phytoplankton group in the eastern China seas

Zhao Haiyang¹, Shen Fang¹, Sun Xuerong¹, Wei Xiaodao²

(1. State Key Laboratory of Estuarine and Coastal Research, East China Normal University, Shanghai 200241, China; 2. Institute of Eco-Chongming, East China Normal University, Shanghai 202162, China)

Abstract: Remote sensing retrieval of phytoplankton group can provide important data for a comprehensive understanding of the role of phytoplankton in marine ecosystem. However, due to the complex optical characteristics, there are still great challenges in the remote sensing retrieval of phytoplankton group in offshore waters. In this study, the eastern China seas region, a complex optical class II water body, is taken as the research area. By using three modeling methods, namely band combination method, multiple linear regression method based on singular value decomposition (SVD+MLR) and XGBoost regression method based on singular value decomposition (SVD+XGBoost), the phytoplankton group is retrieved from remote sensing reflectance (R_{rs}) data. Verified by the *in-situ* measured data set, the chlorophyll *a* (Chl *a*) concentration retrieval model of eight phytoplankton groups by SVD+XGBoost has the highest accuracy, and the determination coefficient (R^2) of Chl *a* concentration inversion model of diatoms and dinoflagellates in the validation set is greater than 0.7. In contrast, the accuracy of Chl *a* concentration of chlorophytes, cyanobacteria and chrysophytes estimated by the three modeling methods is low (the R^2 of the validation results is less than 0.45). At the same time, the applicability of three atmospheric correction methods of OLCI images (C2RCC, POLYMER and MUMM) in the eastern China seas is evaluated. The results show that compared with the other two atmospheric correction algorithms, C2RCC has better performance in each band (root mean square error is less than 0.0048 sr^{-1}). Finally, the performance of the retrieval model on satellite images is verified by the *in-situ* data. The validation results show that the diatoms Chl *a* concentration model established by SVD+MLR has better accuracy (the R^2 is 0.56), while the Chl *a* concentration inversion models of other phytoplankton groups have poor results.

Key words: eastern China seas; phytoplankton group; remote sensing retrieval; OLCI; atmospheric correction