第 42 卷 第 18 期/2022 年 9 月/光学学报 研究论文

光学学报

基于现场多波段激发荧光的浮游植物多种色素含量XGBoost反演

王琳淇[1]，王胜强[1,2*]，孙德勇[1]，李俊生[2]，朱元励[3]，许永久[4]，张海龙[1]

[1]南京信息工程大学海洋科学学院，江苏 南京 210044；

[2]中国科学院空天信息创新研究院遥感科学国家重点实验室，北京 100101；

[3]自然资源部第二海洋研究所，浙江 杭州 310012；

[4]浙江海洋大学水产学院，浙江 舟山 316022

摘要 针对浮游植物的总叶绿素a和7种诊断色素(叶绿素b、岩藻黄素、多甲藻素、19-己酰基氧化盐藻黄素、19-丁酰基氧化盐藻黄素、别藻黄素和玉米黄素)，基于现场多波段激发荧光光谱数据，通过构建激发荧光光谱特征表征量，利用极限梯度提升(XGBoost)机器学习算法，建立了浮游植物色素浓度的反演模型。验证结果表明，反演模型具有良好的估算精度，其中总叶绿素a的反演模型精度最高(决定系数为0.87，平均绝对相对百分比误差为28.1%，均方根误差为1.168 mg·m^{-3})。将建立的色素反演模型应用于东海典型断面处，成功获取了色素浓度的垂向分布特征。

关键词 光谱学；激发荧光光谱；浮游植物色素浓度；反演模型；XGBoost机器学习算法

中图分类号 P76 **文献标志码** A **DOI**：10.3788/AOS202242.1830002

XGBoost-Based Inversion of Phytoplankton Pigment Concentrations from Field Measured Fluorescence Excitation Spectra

Wang Linqi[1], Wang Shengqiang[1,2*], Sun Deyong[1], Li Junsheng[2], Zhu Yuanli[3], Xu Yongjiu[4], Zhang Hailong[1]

[1]*School of Marine Sciences, Nanjing University of Information Science and Technology, Nanjing* 210044, *Jiangsu, China*;

[2]*State Key Laboratory of Remote Sensing Science, Aerospace Information Research Institute, Chinese Academy of Sciences, Beijing* 100101, *China*;

[3]*Second Institute of Oceanography, MNR, Hangzhou* 310012, *Zhejiang, China*;

[4]*School of Fishery, Zhejiang Ocean University, Zhoushan* 316022, *Zhejiang, China*

Abstract In this study, inversion models of phytoplankton pigment concentrations are built for the total chlorophyll a and seven diagnostic pigments (i. e., chlorophyll b, fucoxanthin, peridinin, 19′-hexanoyloxyfucoxanthin, 19′-butanoyloxyfucoxanthin, alloxanthin, and zeaxanthin). Specifically, given the field measured data of fluorescence excitation spectra, the feature representations of fluorescence excitation spectra are constructed, and the machine learning algorithm eXtreme Gradient Boosting (XGBoost) is employed to build these models. The validation indicates that the inversion models have good estimation accuracy, among which the inversion model of the total chlorophyll a has the highest accuracy (with the determination coefficient of 0.87, the mean absolute percentage error of 28.1%, and the root mean square error of 1.168 mg·m^{-3}). In addition, these pigment inversion models are applied to typical sections of the East China Sea, and vertical distribution features of pigment concentrations are obtained.

Key words spectroscopy; fluorescence excitation spectra; phytoplankton pigment concentration; inversion models; XGBoost machine learning algorithm

收稿日期：2022-01-18；**修回日期**：2022-02-20；**录用日期**：2022-02-28

基金项目：国家自然科学基金(42176181，42176179，42106176)、遥感科学国家重点实验室开放基金(OFSLRSS202103)、江苏省基础研究计划(自然科学基金)(BK20211289，BK20210667)、浙江省基础公益研究计划(LGF21D060001)

通信作者：*shengqiang. wang@nuist. edu. cn

1 引 言

浮游植物广泛分布于海洋中，是全球初级生产力的重要贡献者，在物质循环和能量流动中发挥着重要作用[1-2]。浮游植物细胞内包含的色素种类多样，不同种类浮游植物细胞中的色素种类和含量不尽相同，如硅藻中富含岩藻黄素（Fucox）、甲藻中富含多甲藻黄素（Perid）和隐藻中含有其他藻类所不具备的别藻黄素（Allox）[3]。浮游植物的色素构成及其浓度可用于指示浮游植物群落的组成和变化，为浮游植物群落结构的确定与判断提供依据[4-5]。此外，浮游植物色素浓度还可用于指示浮游植物的生理状态[4,6]。因此，准确监测浮游植物的色素浓度具有很重要的生理生态意义[7-8]。

针对浮游植物的色素浓度信息，实验室中多采用高效液相色谱法（HPLC）进行测量[4,9-11]。HPLC具有分离性能高和检测灵敏度高等优点，但该方法需要通过在站点采集样本，再带入实验室中进行分析，过程繁琐，且测试分析费用高。更为重要的是，受空间采样站点数量限制，HPLC分析法所得的信息往往在空间上存在明显的断点，尤其是水下剖面的测量，通常只利用有限的几个水层来代表整个剖面的特征，进而不能客观地反映剖面的真实信息。

荧光的激发是浮游植物典型的生理光学特征之一[12]。不同的色素具有不同的吸收光谱，如叶绿素a（Chla）在440 nm和675 nm波长处有着明显的吸收峰、墨角藻黄素的吸收峰在480 nm左右和藻胆色素在480～650 nm波长范围内呈现出不同的吸收峰[13]。不同的色素具有不同的光吸收特性，进而不同种类的浮游植物具有不同的激发荧光光谱，这种生理光学特性为利用激发荧光光谱反演浮游植物色素浓度提供了有力的理论基础[14]。

早在1966年，Lorenzen[15]成功地利用浮游植物的激发荧光特性，使用单波段船载荧光计监测了浮游植物的叶绿素含量。在此之后，荧光技术就发展成为了浮游植物色素浓度现场监测的重要手段之一。最初的荧光计是通过单波段光源进行激发获取荧光的，进而可实现对浮游植物叶绿素浓度的监测[16-17]。随着荧光技术的发展，目前已经研制出了多种多波段激发荧光光谱仪，如德国BBE公司生产的FluoroProbe和日本JFE公司生产的Multi-Exciter[18-19]。这些多波段激发荧光光谱仪均可在现场进行水平空间（走航式观测）和水下剖面的快速、连续原位观测，相比传统实验室方法具有高效、低成本等优势，为浮游植物色素浓度的快速连续监测提供了技术契机。

然而，现有的现场多波段激发荧光光谱仪不具备对多种浮游植物色素浓度进行反演的能力，阻碍了其在浮游植物色素浓度监测上的应用。因此，亟需开发一种利用激发荧光光谱反演浮游植物色素浓度的技术。激发荧光光谱与浮游植物色素浓度之间具有复杂的内在关系。近年来，极限梯度提升（XGBoost）算法作为一种新兴的机器学习优化算法，能解决复杂的非线性关系[20-21]，这为利用现场多波段激发荧光光谱反演Chla、Fucox、Perid和Allox等多种浮游植物色素浓度提供了思路和方法。

因此，本文针对浮游植物8种典型的色素浓度，通过构建多种激发荧光光谱指数形式，利用XGBoost机器学习方法，建立浮游植物色素浓度反演模型，最终实现对浮游植物色素浓度的快速监测。

2 数据与方法

2.1 数据来源

本文的现场实测数据来自2011年7月和2013年7月的东海调查航次，以及2012年7月的对马海峡调查航次，具体的航次调查信息见文献[22]。现场实测数据主要包括浮游植物色素浓度和激发荧光光谱，具体的调查测试方法如下所述。在调查过程中，共获得了色素浓度和激发荧光匹配的样本141组。此外，在东海调查过程中，在纬向断面（纬度为32.8°N，经度为124.5°E～127.5°E）上开展了激发荧光光谱剖面测量调查（垂向测量分辨率为0.2 m左右，共有7个调查站点，每个站点间隔50 km），用于反演获取东海典型断面上浮游植物色素浓度的垂向分布。

2.1.1 浮游植物色素浓度

在现场调查中，利用安装在温盐深仪（CTD）上的采水瓶采集水样。在暗光环境、低压（<0.01 MPa）下进行过滤，滤膜采用Whatman GF/F玻璃纤维滤膜（直径为47 mm、孔径为0.7 μm）。在过滤完成后，将滤膜用锡纸包裹好，并立刻放入液氮中进行保存。在返回实验室后，根据van Heukelem等[23]的测量标准，在实验室中使用HPLC法测量浮游植物的色素浓度。本文研究所涉及的浮游植物色素主要包含总叶绿素a（Tchla）和7种诊断色素，7种诊断色素包括Fucox、Perid、19-丁酰基氧化盐藻黄素（19Butfu）、19-己酰基氧化盐藻黄素（19Hexfu）、Allox、玉米黄素（Zeax）和叶绿素b（Chlb），具体的英文名称和英文缩写见表1。这7种诊断色素是较为常见的诊断色素，常被用于浮游植物群落结构的诊断分析中[1,24-25]。

表1 本文涉及的浮游植物色素的英文名称（缩写和全称）

Table 1 English names (symbols and full names) of phytoplankton pigments involved in this paper

Symbol	Pigment
Tchla	Total chlorophyll a
Chlb	Chlorophyll b
Fucox	Fucoxanthin
Perid	Peridinin
19Hexfu	19′-Hexanoyloxyfucoxanthin
19Butfu	19′-Butanoyloxyfucoxanthin
Allox	Alloxanthin
Zeax	Zeaxanthin

2.1.2 激发荧光光谱

本研究所使用的激发荧光光谱是由日本JFE公司生产的Multi-Exciter采集得到的，该仪器的详细介绍见官网（https://www.jfe-advantech.co.jp/eng/products/ocean-tahachou.html）。Multi-Exciter具有9个波段（375、400、420、435、470、505、525、570、590 nm）的LED激发光源，其主要工作原理为每个波长的光源依次发射脉冲光，对浮游植物进行荧光激发，在685 nm左右利用硅光电传感器接收记录光合系统II发射的荧光，从而获取浮游植物的激发荧光光谱[19]。在航次调查期间，将Multi-Exciter仪器固定在架子上，利用水文绞车缓慢下放仪器，进行剖面观测。首先，将仪器缓慢下放至水下3～5 m处静置5 min左右，使仪器的温度与周围环境温度保持一致。然后，将仪器缓慢提升至海表处，开始缓慢下放采集数据。

2.2 反演模型建立

本研究使用XGBoost机器学习算法构建基于激发荧光光谱的浮游植物色素浓度反演模型。考虑到荧光光谱与浮游植物色素之间复杂的内在关系，XGBoost算法作为一种新兴的机器学习算法，能够适应复杂的非线性关系，进而被广泛应用于解决复杂的回归问题。

为充分利用浮游植物色素的激发荧光光谱特征，本研究构建了7种激发荧光光谱的指数形式，具体如表2所示，其中$F(\cdot)$为激发荧光光谱，λ为波长，λ_1为波段1的波长，λ_2为波段2的波长。针对每种光谱指数形式，考虑了所有可能的波段组合，将X_1、X_2、X_3、X_5和X_6形式分别生成的72种光谱指数，X_4形式生成的9种光谱指数和X_7形式生成的36种光谱指数作为XGBoost的输入。然后，将141组实测数据按照8∶2的比例划分为训练数据集与验证数据集，利用训练数据集构建色素浓度反演模型，并利用验证数据集对模型精度进行评价。最后，通过对比分析7种光谱指数形式的反演效果，将反演误差最小的光谱指数形式作为最优反演模型。对于每种色素，分别采用上述模型构建方法，建立其最优的浓度反演模型。需要说明的是，在141组实测数据中，对于部分色素有少数站点的质量浓度为0，在模型建立和验证过程中，去掉了色素质量浓度为0的站点，这使得每种色素浓度在模型建立和验证过程中的数据量稍有不同。

2.3 模型精度评价

模型精度评价采用均方根误差（RMSE）、平均绝对百分比误差（MAPE）和决定系数R^2 3个指标，对每一种色素的7种激发荧光光谱指数形式所建立的模型进行精度评价。精度评价的表达式分别为

$$E_{\mathrm{RMS}}=\sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(y-y_{\mathrm{predict}}\right)^2},\tag{1}$$

$$E_{\mathrm{MAP}}=\frac{1}{N}\sum_{i=1}^{N}\left|\frac{y-y_{\mathrm{predict}}}{y}\right|\times 100\%,\tag{2}$$

表2 7种激发荧光光谱指标指数形式

Table 2 Exponential forms of seven excitation fluorescence spectral indicators

Spectral indicator	Exponential form
X_1	$\frac{\lg[F(\lambda_1)]+\lg[F(\lambda_2)]}{\lg[F(\lambda_1)]/\lg[F(\lambda_2)]}$
X_2	$\frac{\lg[F(\lambda_1)]-\lg[F(\lambda_2)]}{\lg[F(\lambda_1)]/\lg[F(\lambda_2)]}$
X_3	$\frac{\lg[F(\lambda_1)]-\lg[F(\lambda_2)]}{\lg[F(\lambda_1)]+\lg[F(\lambda_2)]}$
X_4	$\lg[F(\lambda)]$
X_5	$\frac{\lg[F(\lambda_1)]}{\lg[F(\lambda_2)]}$
X_6	$\lg\frac{F(\lambda_1)}{F(\lambda_2)}$
X_7	$\lg[F(\lambda_1)+F(\lambda_2)]$

$$R^2=1-\frac{\sum_i\left(y_{\mathrm{predict}}-y\right)^2}{\sum_i\left(\bar{y}-y\right)^2},\tag{3}$$

式中：$i=1,\cdots,N$，其中N为样本数量；y为HPLC测量的色素浓度值；$\bar{y}$为HPLC测量的色素浓度平均值；y_{predict}为色素浓度的反演值。

3 结果与分析

3.1 浮游植物色素浓度及激发荧光光谱特征

通过HPLC法测量得到的浮游植物色素浓度数据呈现出明显的变化特征，如表3所示。对8种浮游植物色素浓度进行分析，发现其均呈现对数正态分布，如图1所示，其中横轴C_1～C_8分别代表Perid、19Butfu、Fucox、19Hexfu、Allox、Zeax、Chlb和Tchla的质量浓度。在8种色素样本中，色素值变化范围最大的为Tchla，变化范围最小的色素为19Butfu。Tchla的质量浓度为0.078～6.730 mg·m^{-3}，均值为1.354 mg·m^{-3}。Allox的平均浓度最低，为0.032 mg·m^{-3}。8种色素按平均质量浓度从多到少排序为Tchla、Fucox、Chlb、19Hexfu、Zeax、Perid、19Butfu、Allox。

由于浮游植物色素浓度变化很大，故其激发荧光光谱也呈现出了显著的变化，结果如图2所示。可以看出：大部分样本的激发荧光光谱在470 nm波长处有光谱峰，在435 nm波长处有所下降；部分样本在570 nm波长处出现了第二个光谱峰；在470 nm处的激发荧光光谱呈现出较大的变化，其变化范围为0.140～24.082；在443 nm和570 nm处的激发荧光光谱分别在0.147～17.225和0.131～3.580之间显著变化。

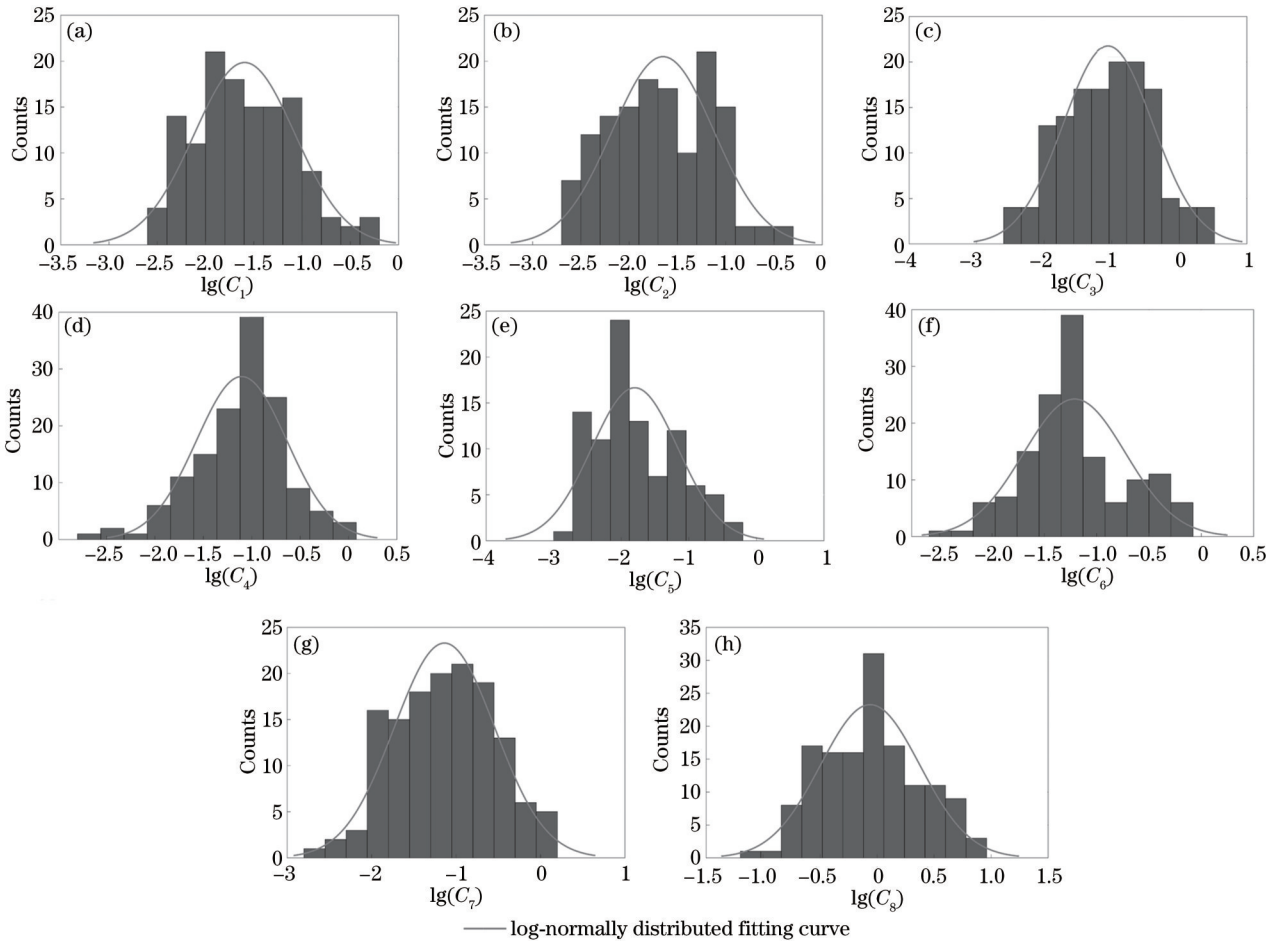

图 1　8种色素浓度的分布直方图。(a) Perid;(b) 19Butfu;(c) Fucox;(d) 19Hexfu;(e) Allox;(f) Zeax;(g) Chlb;(h) Tchla

Fig. 1　Distribution histograms of eight pigment concentrations. (a) Perid; (b) 19Butfu; (c) Fucox; (d) 19Hexfu; (e) Allox; (f) Zeax; (g) Chlb; (h) Tchla

表 3　HPLC 测得的色素浓度分布范围

Table 3　Statistics of pigment concentration measured by HPLC

unit: mg·m^{-3}

Pigment	Minimum value	Maximum value	Average value
Tchla	0.078	6.730	1.354
Fucox	0	2.000	0.221
Perid	0	0.616	0.048
19Hexfu	0	1.038	0.125
19Butfu	0	0.363	0.040
Allox	0	0.525	0.032
Chlb	0	1.325	0.163
Zeax	0.003	0.813	0.120

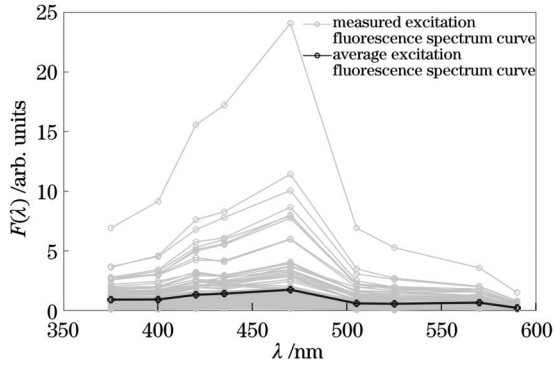

图 2　现场实测的激发荧光光谱曲线

Fig. 2　Excitation fluorescence spectrum curves measured in field

3.2　基于 XGBoost 机器学习算法的色素浓度反演模型

本文针对每一种色素，构建了 7 种激发荧光光谱指数形式，利用 XGBoost 机器学习算法分别建立了 7 种激发荧光光谱色素浓度的反演模型。针对每一种色素，通过对比分析 7 种光谱指数形式的反演效果，确定反演误差最小的光谱指数形式作为最优反演模型。基于 XGBoost 机器学习算法确定的 8 种色素浓度反演的最优激发荧光光谱指数形式及其估算效果如表 4 所示，大部分色素浓度反演的最优激发荧光光谱指数形式为 X_6。

基于 XGBoost 机器学习算法的模型训练效果如图 3 和表 4 所示。可以看出，所有色素反演模型的训练效果均比较理想，实测值与估测值基本分布于 1∶1 线

表 4 8种色素浓度反演的最优激发荧光光谱指数形式及其估算效果

Table 4 Optimal indictor forms of fluorescence excitation spectra and performances inverted by eight pigment concentrations

Pigment	Optimal indictor of fluorescence excitation spectrum	Training dataset			Validation dataset		
		R^2	RMSE / (mg·m^{-3})	MAPE /%	R^2	RMSE / (mg·m^{-3})	MAPE /%
Perid	X_3	0.84	0.036	39.2	0.77	0.110	49.9
19Butfu	X_6	0.94	0.025	24.1	0.67	0.024	50.6
Fucox	X_6	0.96	0.083	25.7	0.87	0.382	46.9
19Hexfu	X_6	0.78	0.103	39.2	0.68	0.125	35.8
Allox	X_5	0.96	0.030	26.5	0.86	0.037	38.2
Zeax	X_6	0.85	0.064	34.5	0.86	0.135	47.2
Chlb	X_5	0.80	0.171	41.1	0.59	0.241	64.2
Tchla	X_6	0.98	0.210	7.5	0.87	1.168	28.1

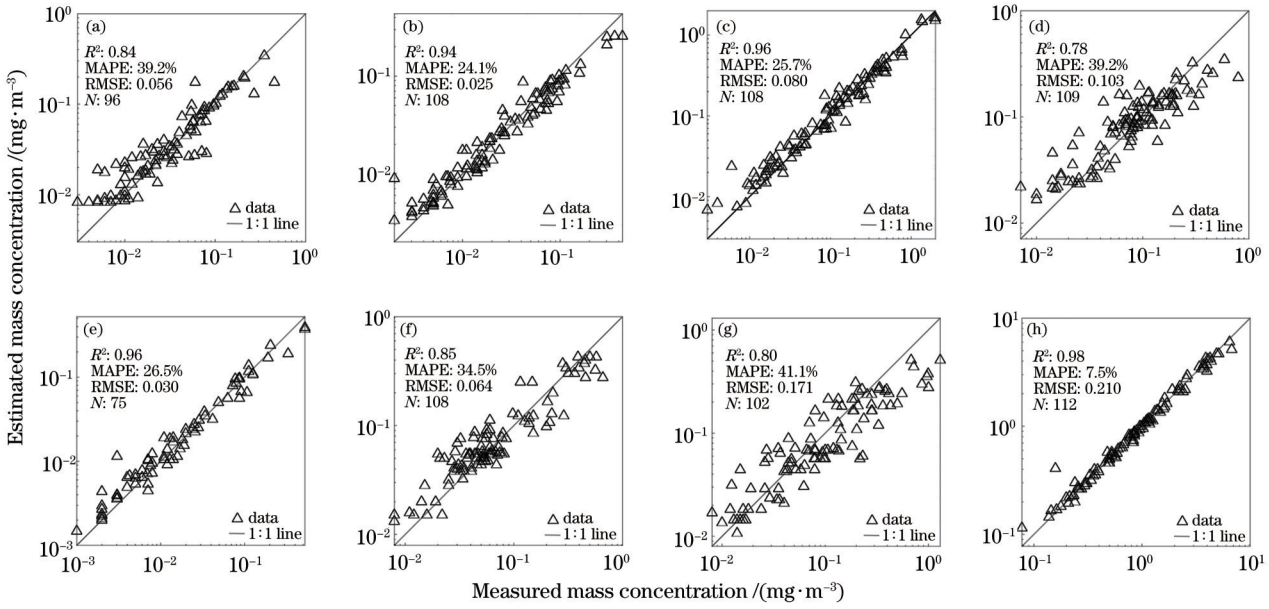

图 3 基于XGBoost机器学习算法的色素浓度反演模型训练效果图。(a) Perid;(b) 19Butfu;(c) Fucox;(d) 19Hexfu;(e) Allox;(f) Zeax;(g) Chlb;(h) Tchla

Fig. 3 Training performances of pigment concentration inversion models based on XGBoost machine learning algorithm. (a) Perid; (b) 19Butfu; (c) Fucox; (d) 19Hexfu; (e) Allox; (f) Zeax; (g) Chlb; (h) Tchla

附近,R^2均达到0.75以上,MAPE基本在40%以下,RMSE基本小于0.2 mg·m^{-3}。

利用验证集数据,对8种色素浓度反演模型进行检验,结果如图4与表4所示。可以看出,基于XGBoost机器学习算法的模型验证效果也相对理想。由模型验证结果可知:R^2>0.80的模型有4个,分别为Tchla、Fucox、Allox与Zeax的浓度反演模型;MAPE小于40%的模型有3个,分别为19Hexfu、Allox与Tchla的浓度反演模型。在8种色素反演模型中,Tchla的反演模型验证效果最好,R^2高达0.87,MAPE低至28.1%,RMSE为1.168 mg·m^{-3},而Fucox、Allox与Zeax的反演模型验证效果同样理想,R^2均超过0.86,MAPE最低可至38.2%,RMSE最低可达0.037 mg·m^{-3}。

结合表4、图3和图4可以发现,基于XGBoost机器学习算法建立的8种色素浓度估算模型均具有较好的估算精度,反演模型估算精度由高到低依次为Tchla、Fucox、Zeax、Allox、Perid、19Hexfu、19Butfu、Chlb。

3.3 典型断面上的浮游植物色素浓度剖面分布

将基于XGBoost机器学习算法建立的激发荧光光谱反演浮游植物色素浓度反演模型应用于东海典型断面(32.8°N,124.5°E~127.5°E)处,成功获取了浮游植物色素浓度的垂向分布,如图5所示。可以看出:所有色素均出现了水下次表层最大层,大部分色素的次表层最大层出现在深度约30 m处,只有Zeax的次表层最大层出现在大于40 m处;Tchla的表层质量浓度在1.00 mg·m^{-3}左右,而次表层的色素质量浓度范围为2.00~3.00 mg·m^{-3};19Hexfu与Tchla的次表层最大层色素浓度的垂向变化梯度较小,19Hexfu的表层

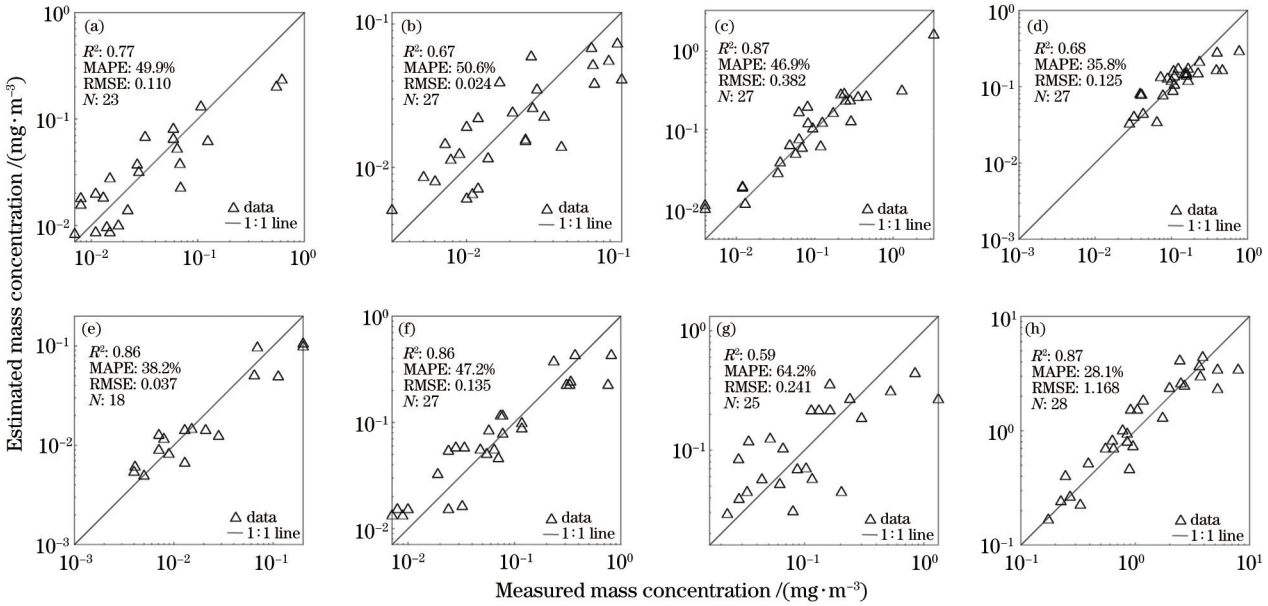

图4 基于XGBoost机器学习算法的色素浓度反演模型验证效果图。(a) Perid;(b) 19Butfu;(c) Fucox;(d) 19Hexfu;(e) Allox;(f) Zeax;(g) Chlb;(h) Tchla

Fig. 4 Validation performances of pigment concentration inversion models based on XGBoost machine learning algorithm. (a) Perid; (b) 19Butfu; (c) Fucox; (d) 19Hexfu; (e) Allox; (f) Zeax; (g) Chlb; (h)Tchla

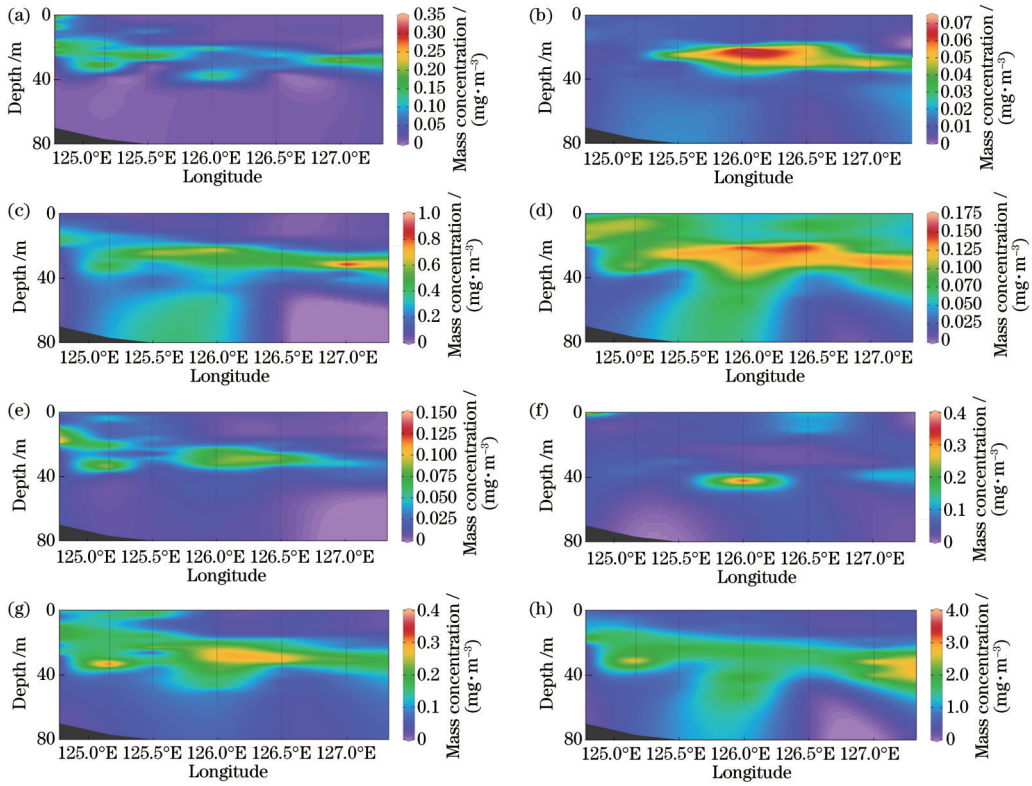

图5 基于激发荧光光谱估算的32.8°N断面上的8种色素浓度剖面分布。(a) Perid;(b) 19Butfu;(c) Fucox;(d) 19Hexfu;(e) Allox;(f) Zeax;(g) Chlb;(h) Tchla

Fig. 5 Profile distributions of eight pigment concentrations in 32.8°N section estimated from fluorescence excitation spectra. (a) Perid; (b) 19Butfu; (c) Fucox; (d) 19Hexfu; (e) Allox; (f) Zeax; (g) Chlb; (h) Tchla

质量浓度在0.06 mg·m^{-3}左右,次表层最大层质量浓度在0.15 mg·m^{-3}左右;19Butfu与Zeax的次表层最大层色素浓度的垂向变化梯度较大,19Butfu的次表层最大层质量浓度在0.06 mg·m^{-3}左右,Zeax的次表层最大层质量浓度在0.35 mg·m^{-3}左右。从水平方向来看,除19Butfu与Zeax外,其余色素浓度均在西侧站点呈现出高值,靠近西侧站点的Tchla的质量浓度在3.00 mg·m^{-3}左右,其原因可能是西侧站点的海水受

到长江冲淡水的影响，营养盐含量较高，进而促进了浮游植物的生长[26-28]。

需要指出的是，考虑到目前现场多波段激发荧光光谱仪不具备对多种浮游植物色素浓度进行反演的能力，本文主要聚焦在反演模型的建立上，关于模型应用只开展了以上的示例研究。在下一步工作中，将会把建立的反演模型应用到多个航次的观测资料中，开展浮游植物色素浓度分布特征与环境影响因素的相关研究。

4 讨 论

本文基于XGBoost机器学习算法建立了激发荧光光谱的浮游植物色素浓度的反演模型，通过R^2、MAPE和RMSE评价指标进行精度验证，结果表明所建立的8个浮游植物色素浓度反演模型均具有良好的估算精度。需要指出的是，最小二乘回归方法是目前比较常见且应用相对广泛的建模方法，其最大的优势是模型简单易操作。为此，本文也采用了最小二乘回归方法构建了色素浓度反演模型。考虑到色素浓度数据呈现出如图1所示的对数正态分布特征，采用指数形式建立反演模型，可表示为

$$C_{\mathrm{Pig}}=10^{c_0+c_1X+c_2X^2+c_3X^3}, \tag{4}$$

式中：C_{Pig}为浮游植物的色素浓度；X为相关性最高的波段组合对应的激发荧光光谱指数；c_0、c_1、c_2和c_3为模型系数，通过最小二乘回归拟合确定。

基于最小二乘回归方法，同样选择了表2中的7种激发荧光光谱指数形式建立浮游植物色素浓度反演模型。在模型训练与验证时选择与XGBoost机器学习算法相同的数据。首先，针对每种光谱指数形式，本研究同样考虑了所有可能的波段组合，X_1、X_2、X_3、X_5与X_6形式分别生成了72种光谱指数，X_4形式生成了9种光谱指数，X_7形式生成了36种光谱指数。然后，计算每种光谱指数与色素浓度的相关系数，将相关系数最高的情况作为最佳波段组合，并用于色素浓度反演模型的构建。接着，利用式(4)进行模型训练，并利用验证集数据对模型进行精度评价。最后，通过对比分析7种光谱指数形式的反演效果，确定反演误差最小的光谱指数形式作为最优的反演模型。对于每种色素浓度，分别采用上述模型构建方法，建立其最优的反演模型。基于最小二乘回归方法获取的色素浓度估算的最优激发荧光光谱指数形式、最佳波段及其估算效果如表5所示，大部分色素的最优激发荧光光谱指数形式为X_6。

表5 基于最小二乘回归方法的8种色素浓度反演的最优激发荧光光谱指数形式、最佳波段及其估算效果

Table 5 Optimal indictor forms of fluorescence excitation spectra, best band combinations and performances inverted by eight pigment concentration based on least square regression method

Pigment	Optimal indictor of fluorescence excitation spectrum	Best band combination /nm	Training dataset			Validation dataset		
			R^2	RMSE /(mg·m^{-3})	MAPE /%	R^2	RMSE /(mg·m^{-3})	MAPE /%
Perid	X_1	$\lambda_1=570$, $\lambda_2=505$	0.54	0.063	72.1	0.54	0.151	58.7
19Butfu	X_6	$\lambda_1=505$, $\lambda_2=590$	0.55	0.053	95.4	0.56	0.021	69.7
Fucox	X_1	$\lambda_1=375$, $\lambda_2=400$	0.74	0.162	94.5	0.87	0.770	68.6
19Hexfu	X_6	$\lambda_1=375$, $\lambda_2=435$	0.43	0.142	62.6	0.07	0.172	57.3
Allox	X_6	$\lambda_1=435$, $\lambda_2=505$	0.51	0.091	126.9	0.36	0.057	120.2
Zeax	X_6	$\lambda_1=435$, $\lambda_2=505$	0.42	0.119	80.0	0.46	0.186	124.5
Chlb	X_6	$\lambda_1=505$, $\lambda_2=590$	0.58	0.211	71.0	0.38	0.268	92.4
Tchla	X_7	$\lambda_1=420$, $\lambda_2=505$	0.76	0.786	43.4	0.75	0.893	45.1

基于最小二乘回归的模型训练与验证效果如图6和图7所示。从图6可以看出，所有色素估算模型的训练效果均不太理想，实测值与估测值偏离1∶1线较远，$R^2<0.6$，MAPE基本高于60%，RMSE基本大于0.150 mg·m^{-3}，最高可达0.786 mg·m^{-3}。从图7可以看出，模型验证效果也不太理想，绝大多数色素的R^2不超过0.6，MAPE基本高于65%，RMSE基本大于0.200 mg·m^{-3}，最高可达0.893 mg·m^{-3}。

基于最小二乘回归方法建立的色素浓度反演模型的精度均低于基于XGBoost机器学习算法所建立的反演模型，且二者的R^2与MAPE差距较大。对于Allox，两种方法所建立的色素浓度反演模型训练精度的MAPE的差值最大，高达100.4%。对于19Hexfu，两种方法所建立的色素浓度反演模型验证精度的R^2的差值最大，高达0.55。其他色素浓度反演模型的精度对比结果如图8所示。

以上结果表明，利用XGBoost机器学习算法建立的色素浓度反演模型精度均高于最小二乘回归方法所建立的反演模型。虽然基于XGBoost机器学习算法的模型相对复杂，但相较于应用广泛的最小二乘回归模型，基于XGBoost机器学习算法所建立的色素反演模型精度更高。因此，本研究推荐使用XGBoost机器学习算法构建的浮游植物色素反演模型。然而，需要指出的是XGBoost机器学习算法作为一种机器学习方法，利用其所建立的模型对训练数据资料可能具有一定的依赖性，而本文所使用的数据资料来自东海和对马海峡，这可能会导致本文建立的浮游植物多种色素浓度反演模型局限于这两个海域。因此，在利用现

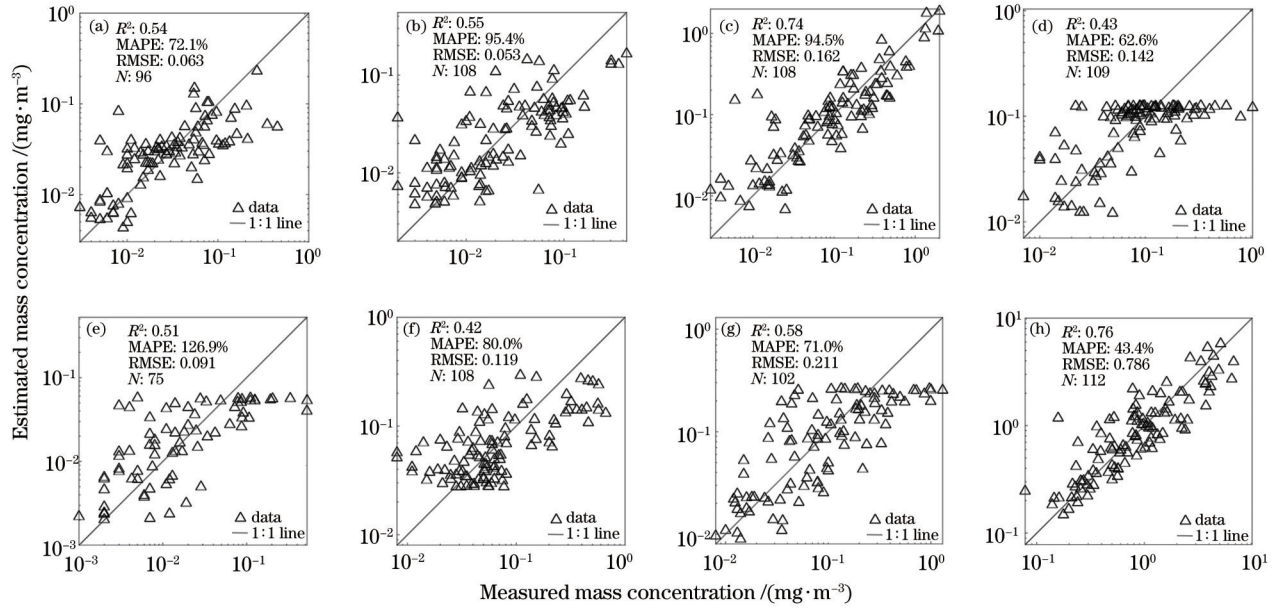

图 6　基于最小二乘回归方法的色素浓度反演模型训练效果图。(a) Perid;(b) 19Butfu;(c) Fucox;(d) 19Hexfu;(e) Allox;(f) Zeax;(g) Chlb;(h) Tchla

Fig. 6　Training performances of pigment concentration inversion models based on least square regression method. (a) Perid; (b) 19Butfu; (c) Fucox; (d) 19Hexfu; (e) Allox; (f) Zeax; (g) Chlb; (h) Tchla

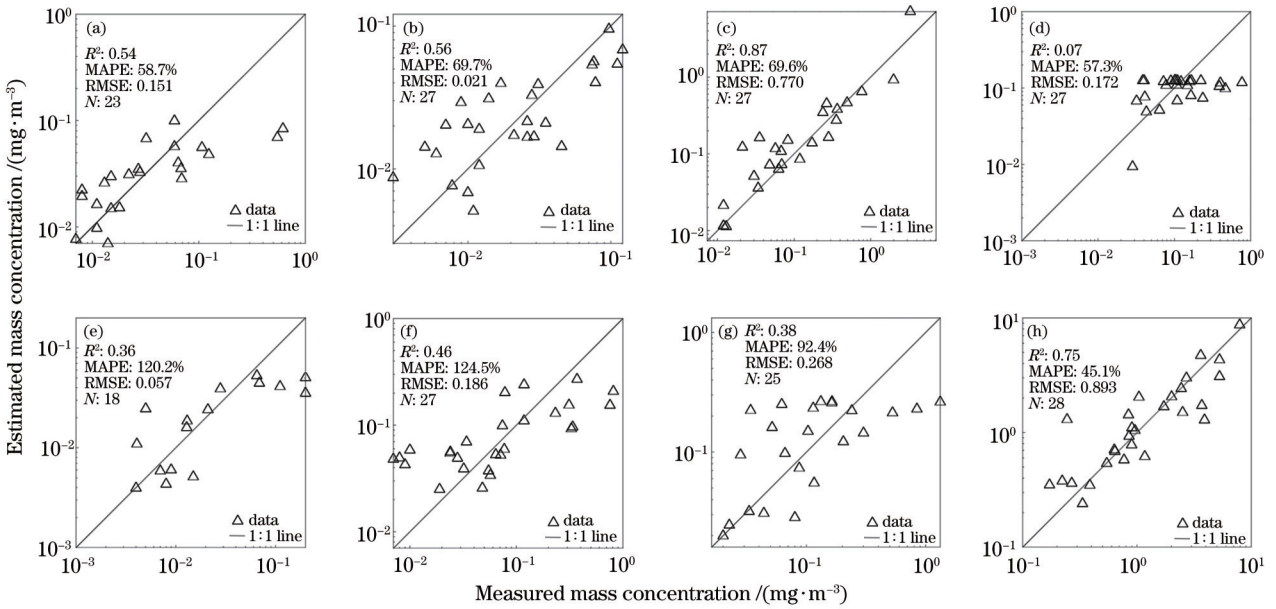

图 7　基于最小二乘回归方法的色素浓度反演模型验证效果图。(a) Perid;(b) 19Butfu;(c) Fucox;(d) 19Hexfu;(e) Allox;(f) Zeax;(g) Chlb;(h) Tchla

Fig. 7　Validation performances of concentration inversion models based on least square regression method. (a) Perid; (b) 19Butfu; (c) Fucox; (d) 19Hexfu; (e) Allox; (f) Zeax; (g) Chlb; (h) Tchla

场多波段激发荧光光谱仪监测浮游植物色素浓度时，虽然本文建议应使用 XGBoost 机器学习算法构建的反演模型，但是这些模型在其他海域中的适用性还需要进行进一步检验。

5　结　　论

针对 Perid、19Butfu、Fucox、19Hexfu、Allox、Zeax、Chlb 与 Tchla 这 8 种浮游植物色素，基于 XGBoost 机器学习算法，建立了基于激发荧光光谱的反演模型，模型精度良好，其中 Tchla 反演模型效果最好（R^2=0.87，MAPE 为 28.1%，RMSE 为 1.168 mg·m^{-3}）。与应用广泛的最小二乘回归建模方法相比，利用 XGBoost 机器学习算法构建的 8 种色素浓度反演模型均呈现出了更高的精度。将构建的色素浓度反演模型，应用至东海典型断面（32.8°N，124.5°E～127.5°E）处的实测激发荧光数据，成功获得了浮游植物色素浓度的垂向分布特征。

本研究在利用 XGBoost 机器学习算法构建基于

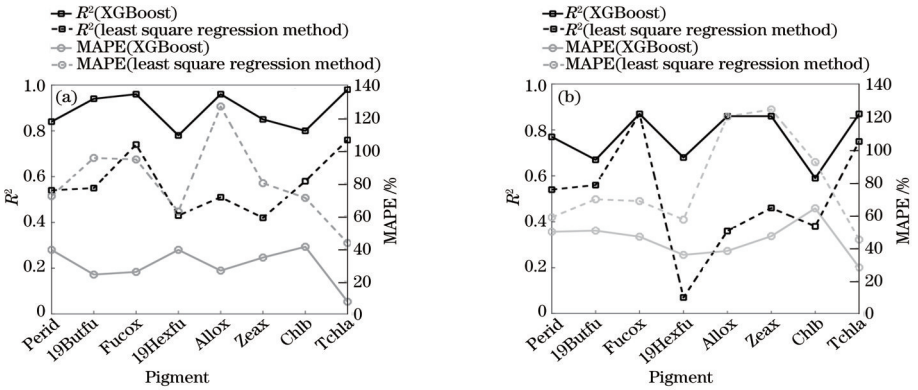

图 8 基于XGBoost机器学习算法与最小二乘回归方法的色素浓度反演模型精度对比图。(a)模型训练;(b)模型验证

Fig. 8 Comparison of accuracies of pigment concentration inversion models based on XGBoost machine learning algorithm and least square regression method. (a) Model training; (b) model validation

激发荧光光谱的浮游植物色素浓度反演模型时,所使用的数据资料主要来自东海和对马海峡海域,因此所建立的反演模型可能主要适用于这两个海域,反演模型在其他海域的适用性还有待进一步检验。

参 考 文 献

[1] 李正浩. 近海浮游植物诊断色素遥感反演及时空分异特征研究[D]. 南京: 南京信息工程大学, 2021: 1-7.
Li Z H. Remote sensing retrieval and temporal and spatial differentiation of phytoplankton diagnostic pigments in offshore waters[D]. Nanjing: Nanjing University of Information Science & Technology, 2021: 1-7.

[2] 李正浩, 陈志钊, 王力彦, 等. 结合GOCI数据反演近海浮游植物叶绿素和类胡萝卜素浓度[J]. 光学学报, 2021, 41(2): 0201001.
Li Z H, Chen Z Z, Wang L Y, et al. Remote sensing inversion of concentration of phytoplankton chlorophyll and carotenoid from GOCI measurements in coastal waters[J]. Acta Optica Sinica, 2021, 41(2): 0201001.

[3] Barlow R, Kyewalyanga M, Sessions H, et al. Phytoplankton pigments, functional types, and absorption properties in the Delagoa and Natal Bights of the Agulhas ecosystem[J]. Estuarine, Coastal and Shelf Science, 2008, 80(2): 201-211.

[4] 殷高方, 赵南京, 胡丽, 等. 基于色素特征荧光光谱的浮游植物分类测量方法[J]. 光学学报, 2014, 34(9): 0930005.
Yin G F, Zhao N J, Hu L, et al. Classified measurement of phytoplankton based on characteristic fluorescence of photosynthetic pigments[J]. Acta Optica Sinica, 2014, 34(9): 0930005.

[5] 王桂芬, 张银雪, 徐文龙, 等. 基于高光谱吸收的南海浮游植物色素浓度估算[J]. 光学学报, 2021, 41(6): 0601002.
Wang G F, Zhang Y X, Xu W L, et al. Estimation of phytoplankton pigment concentration in the South China Sea from hyperspectral absorption data[J]. Acta Optica Sinica, 2021, 41(6): 0601002.

[6] 乔芮. 基于镜检和浮游植物色素分析的贝类食性研究[D]. 上海: 上海海洋大学, 2015: 2-7.
Qiao R. Microscope and phytoplankton pigments analysis of the shellfish feeding habits[D]. Shanghai: Shanghai Ocean University, 2015: 2-7.

[7] Quéré C L, Harrison S P, Colin Prentice I, et al. Ecosystem dynamics based on plankton functional types for global ocean biogeochemistry models[J]. Global Change Biology, 2005, 11(11): 2016-2040.

[8] Nair A, Sathyendranath S, Platt T, et al. Remote sensing of phytoplankton functional types[J]. Remote Sensing of Environment, 2008, 112(8): 3366-3375.

[9] Millie D F, Schofield O M, Kirkpatrick G J, et al. Using absorbance and fluorescence spectra to discriminate microalgae[J]. European Journal of Phycology, 2002, 37(3): 313-322.

[10] 唐晓静, 张前前, 类淑河, 等. 活体浮游植物同步荧光光谱特征分析研究[J]. 光谱学与光谱分析, 2007, 27(3): 556-559.
Tang X J, Zhang Q Q, Lei S H, et al. Research on characterization analysis of synchronous fluorescence spectra of living phytoplankton[J]. Spectroscopy and Spectral Analysis, 2007, 27(3): 556-559.

[11] 王志刚, 刘文清, 张玉钧, 等. 基于激发荧光光谱的浮游植物分类测量方法[J]. 中国环境科学, 2008, 28(4): 329-333.
Wang Z G, Liu W Q, Zhang Y J, et al. The phytoplankton classified measure based on excitation fluorescence spectra technique[J]. China Environmental Science, 2008, 28(4): 329-333.

[12] Falkowski P G, Raven J A. Aquatic photosynthesis[M]. Princeton: Princeton University Press, 2007.

[13] Bricaud A, Claustre H, Ras J, et al. Natural variability of phytoplanktonic absorption in oceanic waters: influence of the size structure of algal populations[J]. Journal of Geophysical Research: Oceans, 2004, 109(C11): C11010.

[14] Yentsch C S, Phinney D A. Spectral fluorescence: an ataxonomic tool for studying the structure of phytoplankton populations[J]. Journal of Plankton Research, 1985, 7(5): 617-632.

[15] Lorenzen C J. A method for the continuous measurement of *in vivo* chlorophyll concentration[J]. Deep Sea Research and Oceanographic Abstracts, 1966, 13(2):

223-227.

[16] Holm-Hansen O, Lorenzen C J, Holmes R W, et al. Fluorometric determination of chlorophyll[J]. ICES Journal of Marine Science, 1965, 30(1): 3-15.

[17] Kiefer D A. Fluorescence properties of natural phytoplankton populations[J]. Marine Biology, 1973, 22(3): 263-269.

[18] Beutler M, Wiltshire K H, Meyer B, et al. A fluorometric method for the differentiation of algal populations *in vivo* and *in situ*[J]. Photosynthesis Research, 2002, 72(1): 39-53.

[19] Yoshida M, Horiuchi T, Nagasawa Y. *In situ* multi-excitation chlorophyll fluorometer for phytoplankton measurements: technologies and applications beyond conventional fluorometers[C]//OCEANS'11 MTS/IEEE KONA, September 19-22, 2011, Waikoloa, HI, USA. New York: IEEE Press, 2011.

[20] Chen T Q, Guestrin C. XGBoost: a scalable tree boosting system[C]//Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, August 13-17, 2016, San Francisco California USA. New York: ACM Press, 2016: 785-794.

[21] 张亦然, 刘廷玺, 童新, 等. 基于XGBoost算法的草甸地上生物量的高光谱遥感反演[J]. 草业学报, 2021, 30(4): 1-12.
Zhang Y R, Liu T X, Tong X, et al. Hyperspectral remote sensing inversion of meadow aboveground biomass based on an XGBoost algorithm[J]. Acta Prataculturae Sinica, 2021, 30(4): 1-12.

[22] Wang S Q, Xiao C, Ishizaka J, et al. Statistical approach for the retrieval of phytoplankton community structures from *in situ* fluorescence measurements[J]. Optics Express, 2016, 24(21): 23635-23653.

[23] van Heukelem L, Thomas C S. Computer-assisted high-performance liquid chromatography method development with applications to the isolation and analysis of phytoplankton pigments[J]. Journal of Chromatography A, 2001, 910(1): 31-49.

[24] Kramer S J, Siegel D A. How can phytoplankton pigments be best used to characterize surface ocean phytoplankton groups for ocean color remote sensing algorithms? [J]. Journal of Geophysical Research: Oceans, 2019, 124(11): 7557-7574.

[25] 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.

[26] Gong G C, Chen Y L L, Liu K K. Chemical hydrography and chlorophyll a distribution in the East China Sea in summer: implications in nutrient dynamics [J]. Continental Shelf Research, 1996, 16(12): 1561-1590.

[27] Zhou M J, Shen Z L, Yu R C. Responses of a coastal phytoplankton community to increased nutrient input from the Changjiang (Yangtze) River[J]. Continental Shelf Research, 2008, 28(12): 1483-1489.

[28] Yamaguchi H, Kim H C, Son Y B, et al. Seasonal and summer interannual variations of SeaWiFS chlorophyll a in the Yellow Sea and East China Sea[J]. Progress in Oceanography, 2012, 105: 22-29.