

杭鑫,曹云,杭蓉蓉,等,2021.基于随机森林算法与高分观测的太湖叶绿素 a 浓度估算模型[J].气象,47(12):1525-1536. Hang X, Cao Y, Hang R R, et al, 2021. Estimation model of chlorophyll-a concentration in Taihu Lake based on random forest algorithm and Gaofen Observations[J]. Meteor Mon, 47(12): 1525-1536 (in Chinese).

基于随机森林算法与高分观测的 太湖叶绿素 a 浓度估算模型*

杭鑫¹ 曹云² 杭蓉蓉³ 徐萌¹ 霍焱⁴ 孙良宵¹ 朱士华¹

1 江苏省气候中心,南京 210019

2 中国人民解放军 61175 部队,南京 210049

3 上海翔弘信息技术有限公司,上海 202172

4 江苏省常州市气象局,常州 213000

提 要: 基于 2018 年和 2019 年有效观测的高分 1 号(GF-1)卫星影像和湖面原位观测的叶绿素 a 浓度数据,利用随机森林算法定量评价特征变量重要性的功能,选择有效特征波段组合,建立了基于原位自动监测数据的太湖叶绿素 a 浓度的遥感反演模型。结果表明:绿光波段(0.52~0.59 μm)和红光波段(0.63~0.69 μm)是遥感估算叶绿素 a 浓度的关键波段,与其他波段组合可以定量估算叶绿素 a 浓度;分季节构建太湖叶绿素 a 浓度估算模型效果更好,春、夏、秋、冬各季模型的决定系数 R^2 分别达 0.84、0.85、0.96 和 0.82;太湖叶绿素 a 浓度夏季最高,秋、春季次之,冬季最低;春、秋和夏季叶绿素 a 浓度空间变化较明显,而冬季变化不明显,叶绿素 a 浓度高值区主要集中在西部沿岸区、竺山湖、梅梁湖和部分湖心区。研究表明:随机森林模型可以客观确定遥感反演叶绿素 a 浓度的有效波段,实现大面积内陆水体叶绿素 a 浓度的高精度估算。

关键词: 叶绿素 a 浓度,随机森林模型,高分 1 号,原位监测数据,太湖

中图分类号: P412, Q938

文献标志码: A

DOI: 10.7519/j.issn.1000-0526.2021.12.008

Estimation Model of Chlorophyll-a Concentration in Taihu Lake Based on Random Forest Algorithm and Gaofen Observations

HANG Xin¹ CAO Yun² HANG Rongrong³ XU Meng¹

HUO Yan⁴ SUN Liangxiao¹ ZHU Shihua¹

1 Jiangsu Climate Center, Nanjing 210019

2 61175 Troops of the Chinese People's Liberation Army, Nanjing 210049

3 Xianghong Information Technology Co., Ltd., Shanghai 202172

4 Changzhou Meteorological Office of Jiangsu Province, Changzhou 213000

Abstract: Based on the GF-1 satellite images effectively observed in 2018 and 2019 and the chlorophyll-a concentration data in-situ observed on the lake surface, a random forest machine learning algorithm is used to quantitatively evaluate the importance measures and contribution rate of the band reflectance and select effective feature band combinations. Then a remote sensing inversion model of chlorophyll-a concentration in Taihu Lake based on in-situ automatic monitoring data is established in this paper. The results show

* 国家重点研发计划(2018YFC1506500)、风云卫星应用先行计划(FY-APP-2021.0403)和江苏省气象局重点科研项目(KZ202003)共同资助

2020 年 7 月 16 日收稿; 2021 年 7 月 20 日收修定稿

第一作者:杭鑫,主要从事卫星遥感与生态气象研究与业务服务. E-mail:570702005@qq.com

通讯作者:徐萌,主要从事生态遥感与天气、气候方面研究与业务服务. E-mail:554941393@qq.com

that the green light band (0.52–0.59 μm) and the red light band (0.63–0.69 μm) are the key bands, which can be combined with other bands to estimate chlorophyll-a concentration. It is better to construct the estimation model of chlorophyll-a concentration in Taihu Lake by seasons, and the determination coefficients R^2 of the spring, summer, autumn, and winter models are 0.84, 0.85, 0.96, and 0.82, respectively. The concentration of chlorophyll-a in Taihu Lake is highest in summer, followed by autumn and spring, and lowest in winter. The spatial changes of chlorophyll-a concentration in spring, autumn and summer are more obvious, while that in winter is not obvious. The areas with high chlorophyll-a concentration are mainly concentrated in the western coastal area, Zhushan Lake, Meiliang Lake and some lake core areas. Studies have shown that the random forest model can objectively determine the effective bands for chlorophyll-a concentration inversion, and achieve high-precision estimation of chlorophyll-a concentration in large inland water bodies.

Key words: chlorophyll-a concentration, random forest model, GF-1, in-situ monitoring data, Taihu Lake

引 言

叶绿素 a 是浮游植物或藻类植物中最丰富的色素,准确定量估算叶绿素 a 浓度对于客观评价水体富营养化程度、实施水环境治理和生态保护具有重要意义(朱广伟等,2018;尹艳娥等,2014)。太湖是一个典型的内陆 II 类水体,水质状况存在明显的空间异质性(水利部太湖流域管理局,2019)。传统的叶绿素 a 浓度测量方法主要是人工采样监测,在实验室采用分光光度法进行分析,最快也要近一天时间才能获取结果,无法满足政府防控需求。2008 年以来采用了浮标站自动监测水质参数,这是一种基于荧光法的半定量分析方法,可在半小时内获得结果,其精密性和时效性都较高(刘苑等,2010)。但传统方法缺乏时空连续性,观测站点有限,观测数据可能有缺失,难以准确描述复杂水体的叶绿素 a 浓度与光谱特征之间的关系(王桥等,2008)。卫星遥感技术具有监测范围广、时效快和连续动态等优势(张鹏等,2018),越来越多地应用于大面积水体叶绿素 a 浓度估算(He et al, 2020;赵少华等,2019)。遥感反演叶绿素 a 浓度的方法大体可分为经验统计方法(Zhang et al, 2014)、半经验半分析方法和分析模型(张明慧等,2018)。经验方法主要通过建立水体反射率与实测叶绿素 a 浓度之间的定量关系来估算叶绿素 a 浓度,半经验半分析方法则是在经验方程的基础上结合辐射传输模型,有一定的物理基础。由于叶绿素 a 光学特征复杂和图像的大气校正不精确,光谱特征与组分浓度之间的关系较为复杂,经验、半经验模型较难解决复杂的非线性问题,反演精度可能较差(张玉超等,2009b),而分析模型虽具有

较高的反演精度和较好的通用性(韩留生等,2014),但机理复杂,建模难度大,实用性较差(潘应阳等,2017)。总体而言,精确遥感反演浑浊水体叶绿素 a 浓度仍是当前较为困难的任务(Nazeer and Nichol, 2016)。

随着人工智能技术的发展,机器学习算法开始在水质参数遥感反演中得到应用。由于机器学习方法不依赖于固定的模型框架,而是通过不断地“学习”模型校正过程中反馈的误差,来完善自变量与因变量之间的复杂关系,因而是解决非线性回归问题的有效方法(Lary et al, 2016;孙全德等,2019)。已有研究证实 BP 神经网络(BP neural network)模型在遥感反演太湖叶绿素 a 浓度时是可行的(朱云芳等,2017),但神经网络模型需要确定网络结构,参数较多,且随着训练样本的增加,所构建的最优网络可能是局部最优,会出现“过学习”的现象。支持向量机(support vector machine, SVM)模型在反演叶绿素 a 浓度时也获取了较高的精度(Kong et al, 2017),且与 BP 模型相比, SVM 模型的反演精度可能更高,稳定性、鲁棒性和空间泛化能力也可能更好,但对于大规模训练样本的 SVM 模型可能会耗费大量的机器内存和运行时间,导致学习效率降低(张玉超等,2009a)。Breiman 提出的随机森林模型是一种组合分类智能算法,具有极高的准确率、极强的数据挖掘能力及分析复杂相互作用分类特征的能力等很多优点,并且可以给出变量重要性估计,甚至被誉为当前最好的机器学习算法之一(Breiman, 2001;李文娟等,2018;刘扬和王维国,2020;王可心等,2021)。近年来已有一些学者尝试利用随机森林机器学习算法开展近岸水色遥感监测,但在浑浊水体定量估算叶绿素 a 浓度的研究仍相对较少,其中

叶绿素 a 浓度实测数据也大多采用人工定点采样实验室分析而来,自动监测数据的应用较为少见,而水质参数自动监测应是今后的必然趋势,因此,有必要充分利用好水质自动观测数据,提高叶绿素 a 浓度定量估算精度。

本文利用 2018 年和 2019 年有效观测的 GF-1 影像和同步自动观测的叶绿素 a 浓度数据,采用随机森林机器学习算法,选择有效特征波段组合,建立太湖叶绿素 a 浓度的遥感估算模型,了解太湖水体叶绿素 a 浓度时空分布特征,试图为蓝藻水华发生发展预测预警提供重要参数,为蓝藻水华防控提供科学依据。

1 数据和方法

1.1 研究区域

选取太湖作为研究区,其地理范围在 $30^{\circ}55' \sim 31^{\circ}30' \text{N}$ 、 $119^{\circ}55' \sim 120^{\circ}40' \text{E}$;湖面面积约为 $2\,338 \text{ km}^2$,平均水深约为 2 m 。

1.2 遥感数据

选取 GF-1 卫星搭载的 Wide-Field-of-View (WFOV) 传感器观测得到的卫星影像作为遥感数据源,所有数据均来自于高分辨率对地观测系统江苏数据与应用中心,时间范围为 2018 年 1 月至 2019 年 5 月。选择有蓝藻水华发生、晴朗或少云天气时质量较高的影像,最后共获得 18 天 27 景影像,日期分别为,2018 年:1 月 12 日,2 月 5 日,2 月 6 日,2 月 13 日,2 月 23 日,4 月 8 日,4 月 28 日,5 月 15 日,5 月 23 日,6 月 25 日,7 月 19 日,7 月 20 日,10 月 6 日,10 月 27 日,12 月 18 日;2019 年:1 月 17 日,1 月 24 日,5 月 3 日。

由于原始影像为 1 级(相对辐射校正产品),为准确获得水体表面反射率,需对数据进行正射校正、辐射定标、大气校正和影像镶嵌等预处理。

1.3 叶绿素 a 浓度实测数据

叶绿素 a 浓度数据取自江苏省生态环境厅布设在太湖的 19 个水质浮标站自动监测数据,站点分布如图 1。由于卫星影像过境时间为每日 11:30(北京时,下同)左右,因此实测数据选取与卫星观测相同日期每日 11:30 的瞬时观测值。

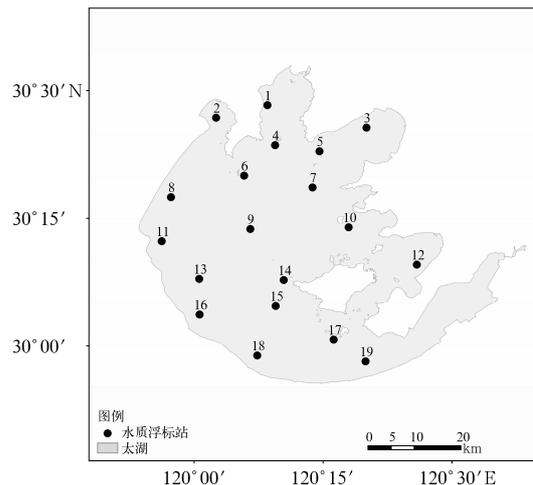


图 1 水质浮标站点分布

Fig. 1 Distribution of water quality buoy sites

1.4 随机森林基本原理

随机森林算法基本思路是:(1)首先确定原始总样本集 D 和变量个数 M ;(2)基于原始训练样本集 D ,采用重采样技术从中抽取 N 个决策树数目 N_{tree} (number of trees)与 D 中样本数量相同的子训练样本集 $D_1, D_2, D_3, \dots, D_n$,分别建立 N_{tree} 个回归树模型,未抽取的部分组成袋外数据(OOB)作为测试样本;(3)确定树节点预选变量个数 M_{try} (number of variable per level)的数值, M_{try} 代表在决策树节点做决定时所用变量数,一般 M_{try} 须小于等于原始数据变量个数 M ;(4)针对每个训练集生长一棵分类回归树,按照节点不纯度最小原则在树的每个节点处,依据法则在 M_{try} 个特征变量中选择高分类能力的特征进行分支生长,并且每棵树都不做任何裁剪,任其最大限度的生长;(5)重复步骤(4) n 次,得到由 n 棵决策树组成的随机森林;(6)随机森林的最终回归结果为每棵树结果的平均值,预测精度则以每棵回归树的平均 OOB 误差来确定。

随机森林模型精度取决于 N_{tree} 和 M_{try} 。 N_{tree} 决定了随机森林的总体规模, M_{try} 决定了单棵决策树的生长状况,两者分别从宏观和微观层面决定随机森林的精度。在回归模型中, M_{try} 值通常为变量数的三分之一, N_{tree} 值根据模型误差随决策树数目变化情况来决定。

随机森林模型支持量化比较各自变量之间对于模型的重要程度,在建模时,可以从大量特征变量中选取对最终结果影响较大的数目较少的特征变

量,删除一些和任务无关或者冗余的特征变量,简化的特征数据集也常常会得到更精确的模型。变量重要性度量的主要评价指标为精度平均减少值 IncMSE (increasing the mean square error)和节点不纯度减少值 IncNodePurity (increasing the node impurity),值越大,表明该变量越重要,反之则相对不重要。

2 结果与分析

2.1 样本处理与潜在变量分析

从所有 18 天 GF-1 卫星过境时刻的实测叶绿素 a 浓度数据中,剔除缺测、异常值及受云影响的

152 个数据,最终得到 190 个数据组成了实测叶绿素 a 浓度样本数据集。再根据不同季节分组,分别得到春季(3—5 月)57 个,夏季(6—8 月)30 个,秋季(9—11 月)20 个和冬季(12—2 月)83 个实测叶绿素 a 浓度的样本子集。从各样本子集中随机选取四分之三的数据作为训练样本集,剩下四分之一数据作为测试样本集。

基于随机森林模型的叶绿素 a 浓度遥感估算模型是以 GF-1 WFV 影像为主要数据源。采用 GF-1 WFV 4 个波段的反射率进行水质参数反演有其合理性,但单一波段往往不能很好地反映影响因子与水质参数之间的关系(朱云芳等,2017)。参考方馨蕊等(2019),将 GF-1 WFV 4 个波段及不同组合的 39 个变量因子作为潜在变量进行筛选(表 1)。

表 1 参与随机森林建模的 39 个潜在变量因子
Table 1 39 latent variable factors involved in random forest modeling

变量名	获取公式	个数
单一波段类 B_i	B_1, B_2, B_3, B_4	4
增强型植被指数类 EVI	$EVI = 2.5 \times \frac{B_1 - B_3}{B_1 + 6B_3 - 7.5B_1 + 1}$	1
归一化植被指数类 $NDVI_{(i,j)}$	$NDVI_{(1,2)} = (B_1 - B_2)/(B_1 + B_2), NDVI_{(1,3)} = (B_1 - B_3)/(B_1 + B_3),$ $NDVI_{(1,4)} = (B_1 - B_4)/(B_1 + B_4), NDVI_{(2,3)} = (B_2 - B_3)/(B_2 + B_3),$ $NDVI_{(2,4)} = (B_2 - B_4)/(B_2 + B_4), NDVI_{(3,4)} = (B_3 - B_4)/(B_3 + B_4)$	6
差值植被指数类 $DVI_{(i,j)}$	$DVI_{(1,2)} = B_1 - B_2, DVI_{(1,3)} = B_1 - B_3, DVI_{(1,4)} = B_1 - B_4,$ $DVI_{(2,3)} = B_2 - B_3, DVI_{(2,4)} = B_2 - B_4, DVI_{(3,4)} = B_3 - B_4$	6
比值植被指数类 $RVI_{(i,j)}$	$RVI_{(1,2)} = B_1/B_2, RVI_{(1,3)} = B_1/B_3, RVI_{(1,4)} = B_1/B_4,$ $RVI_{(2,3)} = B_2/B_3, RVI_{(2,4)} = B_2/B_4, RVI_{(3,4)} = B_3/B_4$	6
其他类 VI	$VI_{(4,1,2,3)} = B_4/(B_1 + B_2 + B_3), VI_{(1,2,3,4)} = B_1/(B_2 + B_3 + B_4),$ $VI_{(2,1,3,4)} = B_2/(B_1 + B_3 + B_4), VI_{(3,1,2,4)} = B_3/(B_1 + B_2 + B_4),$ $VI_{(1,2,3)} = B_1/(B_2 + B_3), VI_{(1,2,4)} = B_1/(B_2 + B_4), VI_{(1,3,4)} = B_1/(B_3 + B_4),$ $VI_{(2,1,3)} = B_2/(B_1 + B_3), VI_{(2,1,4)} = B_2/(B_1 + B_4), VI_{(2,3,4)} = B_2/(B_3 + B_4),$ $VI_{(3,1,2)} = B_3/(B_1 + B_2), VI_{(3,2,4)} = B_3/(B_2 + B_4), VI_{(3,1,4)} = B_3/(B_1 + B_4),$ $VI_{(4,2,3)} = B_4/(B_2 + B_3), VI_{(4,1,3)} = B_4/(B_1 + B_3), VI_{(4,1,2)} = B_4/(B_1 + B_2)$	16

注: B_1 为蓝波段(0.45~0.52 μm), B_2 为绿波段(0.52~0.59 μm), B_3 为红波段(0.63~0.69 μm), B_4 为近红外波段(0.77~0.89 μm)。

Note: B_1, B_2, B_3, B_4 refer to the blue wave band (0.45~0.52 μm), the green wave band (0.52~0.59 μm), the red wave band (0.63~0.69 μm), and the near-infrared wave band (0.77~0.89 μm), respectively.

2.2 关键波段变量因子确定

首先,确定随机森林回归模型中最重要的两个输入参数 M_{try} 和 N_{tree} 。 M_{try} 取潜在变量总数 39 的三分之一,即 $M_{\text{try}} = 13$,而 N_{tree} 值根据误差分析来确定。由于参数 M_{try} 值固定不变,那么参数 N_{tree} 值越大,则误差 Error 越小或趋于稳定,代表模型精度越

高。为考察不同季节叶绿素 a 浓度反演的情况,分别建立了全部样本模型(MOD_{All})、春季模型(MOD_{Spr})、夏季模型(MOD_{Sum})、秋季模型(MOD_{Aut})和冬季模型(MOD_{Win})共 5 个模型。5 个模型在不同的 N_{tree} 和 M_{try} 组合下所对应的分类精度如图 2 所示。由图看出,MOD_{All} 和 MOD_{Spr} 的误差在 N_{tree} 为 600 时趋于稳定,MOD_{Sum} 的误差在 N_{tree}

为 500 时趋于稳定,MOD_{Aut} 的误差在 N_{tree} 为 200 时趋于稳定,MOD_{Win} 的误差在 N_{tree} 为 400 时趋于稳定。综上所述,我们确定在 MOD_{All} 和 MOD_{Spr} 中, $N_{tree}=600$;在 MOD_{Sum} 中, $N_{tree}=500$;在 MOD_{Aut} 中, $N_{tree}=200$;在 MOD_{Win} 中, $N_{tree}=400$ 。

然后,分别根据单一指标 IncMSE 和 IncNodePurity 筛选波段因子变量。随机森林回归模型通过评估每个变量对总体模型预测精度提高的能力,对这些变量进行排序,评估各变量在模型中的相对重要性。利用上述方法确定的参数 M_{try} 和 N_{tree} 分别进行建模和模型优化,MOD_{All}、MOD_{Spr}、MOD_{Sum}、MOD_{Aut} 和 MOD_{Win} 模型各单独训练 5000 次,从中各选择精度相对较高(相关系数 $R>0.85$) 的 20 个模型,对应每个变量都可以得到 20 个 IncMSE 和 IncNodePurity 值。分别计算每个变量的 IncMSE 和 IncNodePurity 指标的平均值,然后根据这两个指标

平均值的大小分别进行排序。将指标平均值排名相对靠前且重要性度量曲线出现较明显拐点前的特征变量认为是相对重要的变量,由此得到全部样本模型 MOD_{All}、春季模型 MOD_{Spr}、夏季模型 MOD_{Sum}、秋季模型 MOD_{Aut} 和冬季模型 MOD_{Win} 共 5 个模型的单指标重要变量筛选结果:

(1) 基于 IncMSE 指标的筛选结果

- MOD_{All}: $DVI_{(2,3)}, VI_{(2,1,3)}, RVI_{(2,3)}, NDVI_{(2,3)}, VI_{(2,1,3,4)}, VI_{(2,3,4)}, DVI_{(2,4)}, VI_{(2,1,4)}, B_1, VI_{(3,1,2)}$;
- MOD_{Spr}: $VI_{(2,1,4)}, DVI_{(2,3)}, VI_{(2,1,3)}, VI_{(2,1,3,4)}, DVI_{(1,2)}$;
- MOD_{Sum}: $VI_{(3,1,2)}, RVI_{(1,3)}, DVI_{(2,3)}, RVI_{(2,3)}, DVI_{(1,3)}$;
- MOD_{Aut}: $VI_{(2,3,4)}, RVI_{(1,2)}, VI_{(1,2,3,4)}, VI_{(1,2,3)}, VI_{(4,1,2)}, NDVI_{(2,4)}, VI_{(1,2,4)}, VI_{(1,3,4)}, DVI_{(1,2)}$;
- MOD_{Win}: $RVI_{(2,3)}, RVI_{(2,4)}, VI_{(2,1,3,4)}, VI_{(2,3,4)}$ 。

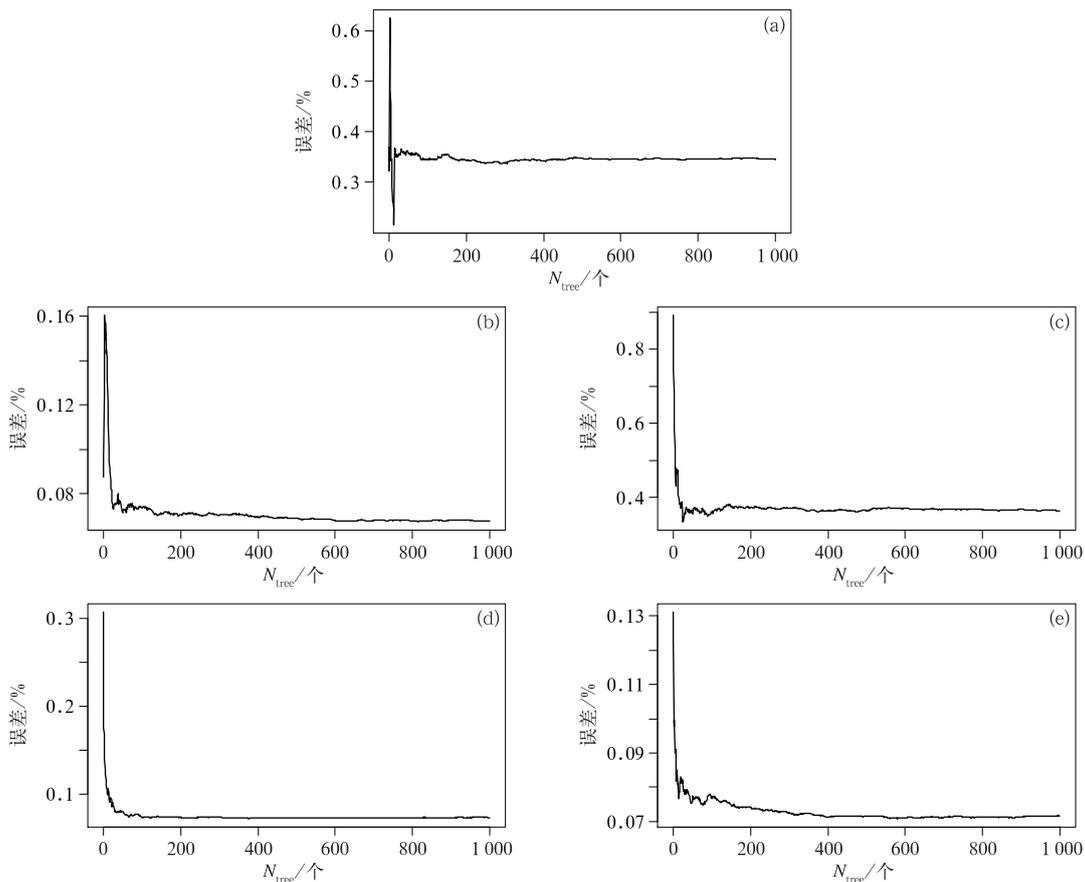


图 2 决策树数目 N_{tree} 与模型误差的关系

(a) MOD_{All}, (b) MOD_{Spr}, (c) MOD_{Sum}, (d) MOD_{Aut}, (e) MOD_{Win}

Fig. 2 The relationship between decision tree N_{tree} and model error

(a) MOD_{All}, (b) MOD_{Spr}, (c) MOD_{Sum}, (d) MOD_{Aut}, (e) MOD_{Win}

(2) 基于 IncNodePurity 指标的筛选结果

$$\begin{aligned} \text{MOD}_{\text{All}}: & VI_{(2,1,3,4)}, VI_{(2,3,4)}, VI_{(2,1,3)}, B_4, VI_{(2,1,4)}, \\ & DVI_{(2,3)}; \\ \text{MOD}_{\text{Spr}}: & VI_{(4,1,3)}, NDVI_{(1,4)}, VI_{(4,1,2,3)}, EVI; \\ \text{MOD}_{\text{Sum}}: & RVI_{(1,3)}, VI_{(3,1,2)}, VI_{(2,3,4)}, VI_{(1,3,4)}; \\ \text{MOD}_{\text{Aut}}: & VI_{(2,1,3)}, RVI_{(2,3)}, DVI_{(1,2)}, VI_{(4,1,2,3)}, \\ & VI_{(4,1,2)}, DVI_{(2,3)}, VI_{(4,2,3)}, B_4, NDVI_{(2,3)}, EVI; \\ \text{MOD}_{\text{Win}}: & RVI_{(2,3)}, VI_{(2,3,4)}, VI_{(3,1,2)}, RVI_{(2,4)}. \end{aligned}$$

从上述两类筛选结果中,发现同时包含 B_2 和 B_3 波段变量因子出现次数较多,在 IncMSE 指标的筛选结果中出现 19 次,占比为 58%,在 IncNodePurity 指标的筛选结果中出现 16 次,占比为 57%;而包含 B_2 或 B_3 波段变量因子出现次数更多,在 IncMSE 指标筛选结果中共出现 32 次,占比为 97%,仅 1 个变量因子不包含 B_2 或 B_3 波段,在 IncNodePurity 指标筛选结果中累计出现 25 次,占比为 89%,仅 3 个变量因子不包含 B_2 或 B_3 波段,说明 B_2 和 B_3 波段在所有变量因子中占主导地位。

最后,根据结合 IncMSE 和 IncNodePurity 的综合因子筛选关键波段因子。已有文献在筛选特征

$$RIEI = \frac{\left(\frac{IncMSE_i - IncMSE_{\min}}{IncMSE_{\max} - IncMSE_{\min}} + \frac{IncNodePurity_i - IncNodePurity_{\min}}{IncNodePurity_{\max} - IncNodePurity_{\min}} \right)}{2} \quad (1)$$

式中: $IncMSE_i$ 为第 i 个 IncMSE 值, $IncMSE_{\min}$ 为 20 个 IncMSE 中的最小值, $IncMSE_{\max}$ 为 20 个 IncMSE 中的最大值; $IncNodePurity_i$ 为第 i 个 IncNodePurity 值, $IncNodePurity_{\min}$ 为 20 个 IncNodePurity 中的最小值, $IncNodePurity_{\max}$ 为 20 个 IncNodePurity 中的最大值。

将包含全部样本和春、夏、秋和冬季共 5 个模型的 RIEI 值绘制成变量重要性评估曲线(图 3)。

将变量重要性排名相对靠前且曲线出现较明显拐点前的特征变量认为是相对重要的变量,分别筛选出 5 个模型的重要特征变量:

$$\begin{aligned} \text{MOD}_{\text{All}}: & VI_{(2,1,3,4)}, VI_{(2,3,4)}, VI_{(2,1,3)}, B_4, VI_{(2,1,4)}, \\ & DVI_{(2,3)}, NDVI_{(2,3)}; \\ \text{MOD}_{\text{Spr}}: & DVI_{(1,2)}, VI_{(2,1,4)}, VI_{(2,1,3,4)}, RVI_{(1,2)}, \\ & VI_{(2,1,3)}, DVI_{(2,3)}, VI_{(1,2,3)}; \\ \text{MOD}_{\text{Sum}}: & RVI_{(1,3)}, VI_{(3,1,2)}, RVI_{(2,3)}, VI_{(2,3,4)}, \\ & VI_{(1,3,4)}, DVI_{(2,3)}, DVI_{(1,3)}, VI_{(1,2,3,4)}, VI_{(2,1,3,4)}, \\ & VI_{(1,2,3)}, VI_{(2,1,3)}; \\ \text{MOD}_{\text{Aut}}: & VI_{(4,1,2)}, DVI_{(1,2)}, VI_{(2,3,4)}, VI_{(4,1,2,3)}, \end{aligned}$$

变量因子时,大多选择 IncMSE 和 IncNodePurity 中的一种指标(罗晓春等,2019)。但从以上两种指标的筛选结果中可以发现,同一种模型不同指标筛选出的变量因子并不完全相同。相对应于这两个指标,所有模型经过筛选得到两组不同的变量因子。对于 MOD_{All} ,单用指标 IncMSE 筛选得到 10 个变量,用指标 IncNodePurity 得到 6 个变量,两组变量中出现 5 个相同因子。相类似的, MOD_{Win} 有 3 个相同因子, MOD_{Sum} 中有 2 个相同因子, MOD_{Aut} 中仅有 1 个相同因子。我们注意到模型 MOD_{Spr} 的两组筛选结果中甚至没有出现相同的变量因子。表明仅用 IncMSE 或 IncNodePurity 一个指标并不能完全反映变量的重要性,存在一定的局限性。为此,考虑综合 IncMSE 和 IncNodePurity 两种指标构建一组新的变量相对重要性评价指标(relative importance evaluation index, RIEI),具体计算方法是:对于 MOD_{All} 、 MOD_{Spr} 、 MOD_{Sum} 、 MOD_{Aut} 和 MOD_{Win} 这 5 个模型中的每个模型,首先分别将以上 20 个模型的 IncMSE 值和 IncNodePurity 值进行归一化处理,然后再求平均值,得到每一个变量的 RIEI 值:

$$\begin{aligned} & VI_{(1,2,3,4)}; \\ \text{MOD}_{\text{Win}}: & RVI_{(2,3)}, VI_{(2,3,4)}, RVI_{(2,4)}, VI_{(2,1,3,4)}, \\ & DVI_{(2,3)}, VI_{(3,1,2)}. \end{aligned}$$

在以上模型的变量中,同时包含 B_2 和 B_3 波段的变量因子出现 25 次,占比为 69%,明显高于 IncMSE 指标的 58%和 IncNodePurity 指标的 57%,表明使用综合指标筛选的结果明显优于单一指标。同时,所有的变量因子都包含了 B_2 或 B_3 波段,再次证明绿光波段 B_2 ($0.52 \sim 0.59 \mu\text{m}$) 和红光波段 B_3 ($0.63 \sim 0.69 \mu\text{m}$) 是遥感反演叶绿素 a 浓度的关键波段因子,对准确估算浑浊水体叶绿素 a 浓度具有重要的意义。

2.3 模型建立

根据上述 5 个模型筛选的重要特征变量,分别重新构建随机森林模型,其中参数 M_{try} 为特征变量个数的三分之一,分别取 1、2、3 和 4 四个数值, N_{tree} 则根据前面误差分析结果分别选取 400、500 和 600 三个数值。对应每组参数组合 $(M_{\text{try}}, N_{\text{tree}})$ 重复建

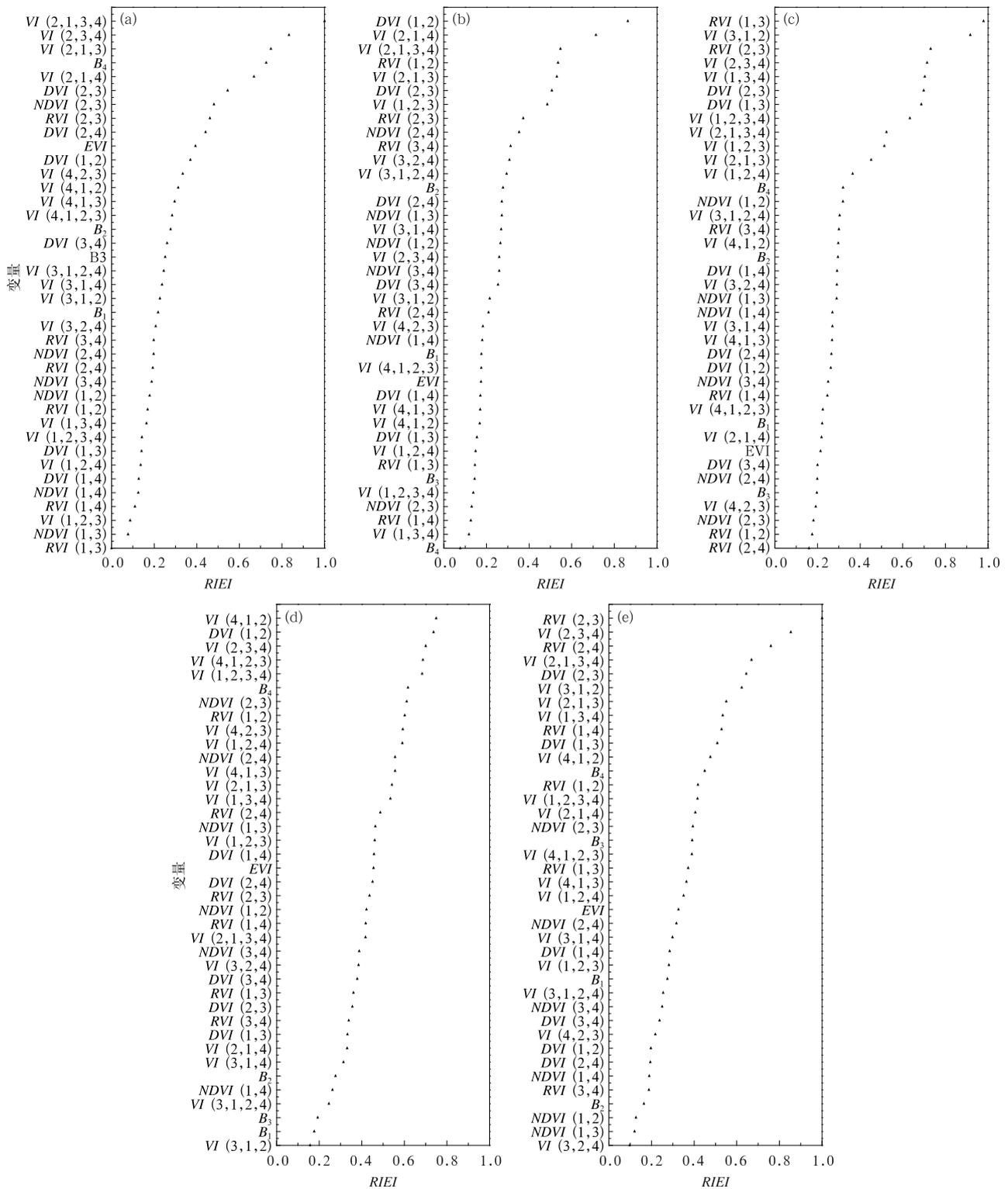


图 3 基于 RIEI 值的 5 个模型的变量重要性评估曲线

(a) MOD_{All}, (b) MOD_{Spr}, (c) MOD_{Sum}, (d) MOD_{Aut}, (e) MOD_{Win}

Fig. 3 Variable importance evaluation curve of five models based on RIEI value

(a) MOD_{All}, (b) MOD_{Spr}, (c) MOD_{Sum}, (d) MOD_{Aut}, (e) MOD_{Win}

模 5000 次, 从中选出各模型精度最高的参数组合 (表 2)。由表 2 可知, 在 MOD_{All}、MOD_{Spr}、MOD_{Sum}、

MOD_{Aut} 和 MOD_{Win} 共 5 个模型中, 秋季模型 MOD_{Aut} 精度 (R) 最高达 0.99, 对应的参数组合为 (2, 400),

包含全部样本的模型 MOD_{All} 和冬季模型 MOD_{Win} 精度 (R) 最小, 均为 0.84, 对应的参数组合分别为 (3,

表 2 各模型精度

Table 2 Each model accuracy

模型序号	MOD_{All}	MOD_{Spr}	MOD_{Sum}	MOD_{Aut}	MOD_{Win}
M_{try}	3	2	3	2	2
N_{tree}	400	500	500	400	600
精度 (R)	0.84	0.94	0.94	0.99	0.84

2.4 模型验证

进一步验证模型的反演精度, 将上述建立的 5 个随机森林模型 MOD_{All} 、 MOD_{Spr} 、 MOD_{Sum} 、 MOD_{Aut} 和 MOD_{Win} 反演的叶绿素 a 浓度值, 分别与实测叶绿素 a 浓度值进行比较, 各模型叶绿素 a 浓度估算值和实测值拟合关系见图 4。由图可知, 各模型估算值与实测值之间均呈现较高的相关性, 均方根误差

($RMSE$) 均较低。其中 MOD_{Aut} 模型的拟合精度最高, 决定系数 (R^2) 为 0.96, $RMSE$ 为 $2.1 \text{ mg} \cdot \text{m}^{-3}$, MOD_{Sum} 模型次之, R^2 为 0.85, $RMSE$ 为 $2.0 \text{ mg} \cdot \text{m}^{-3}$, MOD_{All} 模型的拟合精度最低, R^2 为 0.77, $RMSE$ 为 $2.2 \text{ mg} \cdot \text{m}^{-3}$ 。这一结果表明用所有样本构建的模型的拟合效果不如分季节构建的模型, 说明分季节模型估算的叶绿素 a 浓度值更加接近实测值, 其中秋季模型拟合效果又明显好于其他 3 个季节模型。

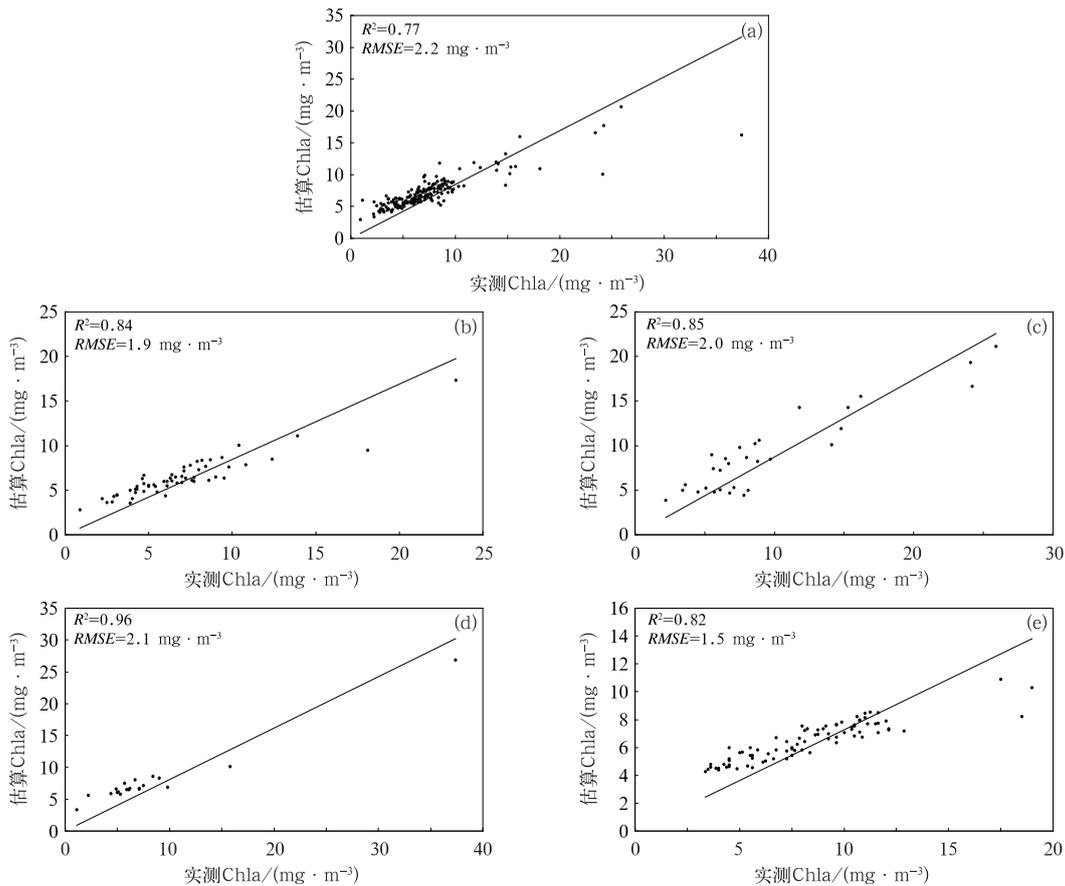


图 4 各模型叶绿素 a 浓度估算值和实测值之间的散点关系

(a) MOD_{All} , (b) MOD_{Spr} , (c) MOD_{Sum} , (d) MOD_{Aut} , (e) MOD_{Win}

Fig. 4 Scatter plots between estimated and measured chlorophyll a concentration of each model

(a) MOD_{All} , (b) MOD_{Spr} , (c) MOD_{Sum} , (d) MOD_{Aut} , (e) MOD_{Win}

3 讨 论

有效波段的选择是高精度估算叶绿素 a 浓度的关键(姜广甲等,2013)。利用随机森林模型识别、量化特征变量重要性的功能,客观筛选出相对重要的变量因子,本文确定了绿光波段($0.52\sim 0.59\ \mu\text{m}$)和红光波段($0.63\sim 0.69\ \mu\text{m}$)及其组合为定量反演太湖叶绿素 a 浓度的关键波段。尽管目前对于浑浊 II 类水体叶绿素 a 的光谱特征和敏感波段的研究还较少(潘应阳等,2017),但已有一些学者针对不同内陆水体的实测光谱数据进行了分析,得到了类似的结果。有统计表明水体叶绿素 a 浓度与 $0.54\ \mu\text{m}$ 和 $0.701\ \mu\text{m}$ 反射峰的相关系数接近 1,相近波段的特征光谱能较好反演叶绿素 a 浓度(吴传庆等,2009);水中藻类最显著的光谱特征是 $0.56\ \mu\text{m}$ 附近的反射峰,该峰值的存在与否通常被认为是判断水体是否含有藻类的依据,而 $0.682\ \mu\text{m}$ 波段与叶绿素 a 浓度相关性最好(杨婷等,2011);此外,叶绿素 a 浓度的变化也会影响浮游植物吸收峰的数值和位置,从而影响到最佳波段的选择(潘应阳等,2017)。这些大多是针对特定水体特定区域、选用较少的测量数据的研究结果可能并不具有普适性(冯驰等,2015)。对于较为浑浊的内陆水体,由于存在浮游植物、悬浮物、溶解有机物等许多影响叶绿素 a 吸收的物质,各组分之间彼此混合、交互作用,水体的光谱特征更加复杂,实际测量的叶绿素 a 反射吸收峰也会有明显不同(罗建美等,2017)。因此,利用随机森林模型确定叶绿素 a 光谱特征的关键波段,可以避免特定水域、特定叶绿素 a 浓度测量的局限性,对于

定量遥感反演大面积浑浊 II 类水体叶绿素 a 浓度不失为一次有益的尝试。

太湖叶绿素 a 浓度具有明显的时空分布特征(乐成峰等,2008)。本文利用春、夏、秋、冬四个季节的随机森林模型,分析了太湖叶绿素 a 浓度的时空分布特点。图 5 为 2018 年全湖叶绿素 a 浓度均值随时间的变化情况,可见夏季平均叶绿素 a 浓度最高,冬季最低,二者平均浓度分别为 $9.6\ \text{mg}\cdot\text{m}^{-3}$ 和 $7.1\ \text{mg}\cdot\text{m}^{-3}$,秋季由于受到夏季高浓度的影响,叶绿素 a 浓度高于春季,分别为 $8.6\ \text{mg}\cdot\text{m}^{-3}$ 和 $7.7\ \text{mg}\cdot\text{m}^{-3}$,这与乐成峰(2008)研究较为一致。以一景影像代表各季节的空间分布情况(图 6),可以看到冬季叶绿素 a 浓度较低,空间变化不明显,这与冬季温度降低有关(贾春燕,2008);春季叶绿素 a 浓度开始出现较明显的空间变化,西北部湖区,特别是梅梁湖和竺山湖附近叶绿素 a 浓度较高,向湖心区逐渐减小,这主要是由于存在众多的入湖河流和高密度城市排污口,造成水体富营养化严重(刘聚涛等,2011);而夏、秋季叶绿素 a 浓度空间变化最为显著,西部沿岸区、竺山湖、梅梁湖和部分湖心区叶绿素 a 浓度明显偏高,除了与富营养化程度有关外,还与夏、秋季盛行东南风引起的湖流有关(秦伯强等,2004)。东太湖部分水域叶绿素 a 浓度始终呈现相对较高的水平,则可能是受该区域丰富的水生植物影响。考虑到太湖各区域均存在不同程度的水生植物,且随着季节交替,水生植物面积变化幅度较大(Zhao et al, 2013),因此,本研究参考了相关文献(杨婷等,2011;朱云芳等,2017),没有将水生植物剔除,可能会对叶绿素 a 浓度反演产生干扰,后续将会进一步研究水生植物的影响。另外,本文仅采用了

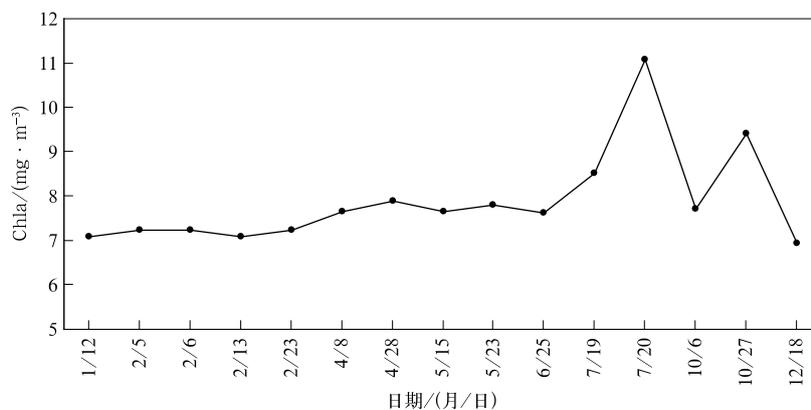


图 5 2018 年太湖叶绿素 a 浓度时间变化

Fig. 5 Temporal variation of chlorophyll a concentration in Taihu Lake in 2018

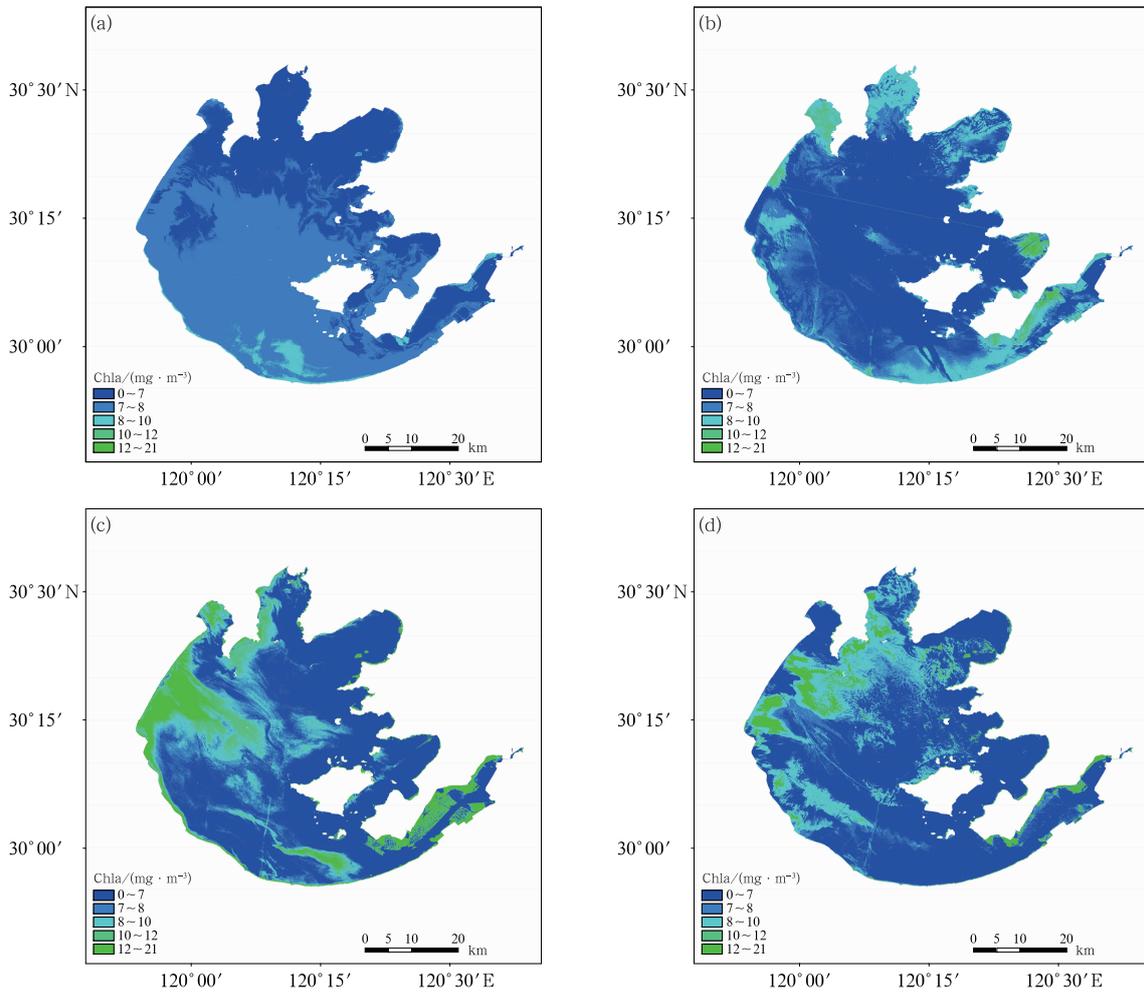


图 6 各季节太湖叶绿素 a 浓度空间分布

(a) 冬季, (b) 春季, (c) 夏季, (d) 秋季

Fig. 6 Spatial distribution of chlorophyll-a concentration in Taihu Lake in each seasons

(a) winter, (b) spring, (c) summer, (d) autumn

2018 年 1 月至 2019 年 5 月的资料,且由于高分卫星观测条件限制,数据时序不连续,因此需要进一步搜集更多数据,增加样本数,以期构建更为精确的模型。

由于人工采样实验室分析法和水质浮标站自动监测法在原理、方法和步骤等方面不同,加之太湖湖体各处的藻种、水质存在较大差异,人工采样实验室分析与自动监测的叶绿素 a 浓度之间的差异较大。据江苏省环境监测中心利用全湖人工观测数据与原位观测数据对比,表明人工采样分析与自动监测数据之间的相关性较差,整体而言,人工采样分析得到的结果约为自动监测的 2.4 倍。因此,本文利用自动监测数据来估算的叶绿素 a 浓度,比目前大多数采用人工采样实验室分析数据的估算结果要明显偏

小(曹红业等,2016;冯驰等,2015;宋挺等,2017),但可能更客观地反映了太湖湖体叶绿素 a 浓度及其分布状况。水质参数的自动监测分析应当是今后的必然趋势,事实上,近年来江苏省太湖水污染防治工作中使用的叶绿素 a 浓度数据就主要来自于自动监测结果,因此有必要在后续的研究中搞清楚自动观测数据和人工采样数据之间的关系。

4 结 论

(1)用随机森林模型可以确定绿光波段和红光波段为遥感反演太湖叶绿素 a 浓度的关键波段,但仍需结合其他波段,这样可以避免特定水域、特定叶绿素 a 浓度测量的局限性,结果更客观。

(2)分季节构建的太湖叶绿素 a 浓度随机森林估算模型能够得到更加接近实测值的结果,春、夏、秋、冬各季模型的决定系数 R^2 分别为 0.84、0.85、0.96 和 0.82, $RMSE$ 分别为 1.9、2.0、2.1 和 $1.5 \text{ mg} \cdot \text{m}^{-3}$ 。

(3)太湖叶绿素 a 浓度呈明显时空变化特征,其中夏季叶绿素 a 浓度最高,秋、春季次之、冬季最低;春、秋和夏季全湖叶绿素 a 浓度空间变化较明显,冬季空间变化不明显,叶绿素 a 浓度高值区主要集中在西部沿岸区、竺山湖、梅梁湖和部分湖心区。

(4)模型使用实测数据来源于水质浮标站自动监测,跟人工采样实验室分析法获得结果相比明显偏低,这是因为两种监测方式在原理、方法和步骤等方面有所不同,而且太湖湖体不同区域的藻种、水质也存在较大差异。但水质自动化监测是未来的发展方向,充分利用好自动观测数据具有重要的现实意义。

参考文献

- 曹红业,龚涛,袁成忠,等,2016. 基于 RBF 模型的太湖北部叶绿素 a 浓度定量遥感反演[J]. 环境工程学报, 10(11): 6499-6504. Cao H Y, Gong T, Yuan C Z, et al, 2016. Quantitative retrieval of chlorophyll-a concentration in northern part of Lake Taihu based on RBF model[J]. Chin J Environ Eng, 10(11): 6499-6504 (in Chinese).
- 方馨蕊,温兆飞,陈吉龙,等,2019. 随机森林回归模型的悬浮泥沙浓度遥感估算[J]. 遥感学报, 23(4): 756-772. Fang X R, Wen Z F, Chen J L, et al, 2019. Remote sensing estimation of suspended sediment concentration based on random forest regression model [J]. J Remote Sens, 23(4): 756-772(in Chinese).
- 冯驰,金琦,王艳楠,等,2015. 基于 GOCI 影像和水体光学分类的内陆湖泊叶绿素 a 浓度遥感估算[J]. 环境科学, 36(5): 1557-1564. Feng C, Jin Q, Wang Y N, et al, 2015. Remote sensing estimation of chlorophyll-a concentration in inland lakes based on GOCI image and optical classification of water body[J]. Environ Sci, 36(5): 1557-1564(in Chinese).
- 韩留生,陈水森,李丹,等,2014. 近岸二类水体生物光学模型参数优化[J]. 热带地理, 34(3): 351-358. Han L S, Chen S S, Li D, et al, 2014. Optimization of bio-optical algorithm parameters in coastal CASE-II waters [J]. Trop Geogr, 34(3): 351-358 (in Chinese).
- 贾春燕,2008. 基于遥感技术的湖泊叶绿素 a 动态监测及改善空间制图详度的研究[D]. 杭州:浙江大学. Jia C Y, 2008. Dynamic monitoring of chlorophyll a for lakes and improving spatial mapping precision[D]. Hangzhou: Zhejiang University (in Chinese).
- 姜广甲,周琳,马荣华,等,2013. 浑浊 II 类水体叶绿素 a 浓度遥感反演(II): MERIS 遥感数据的应用[J]. 红外与毫米波学报, 32(4): 372-378. Jiang G J, Zhou L, Ma R H, et al, 2013. Remote sensing retrieval for chlorophyll-a concentration in turbid case II waters (II): application on MERIS image[J]. J Infrared Millim Waves, 32(4): 372-378(in Chinese).
- 乐成峰,李云梅,孙德勇,等,2008. 太湖叶绿素 a 浓度时空分异及其定量反演[J]. 环境科学, 29(3): 619-626. Le C F, Li Y M, Sun D Y, et al, 2008. Spatio-temporal distribution of chlorophyll a concentration and its estimation in Taihu Lake[J]. Environ Sci, 29(3): 619-626(in Chinese).
- 李文娟,赵放,酆敏杰,等,2018. 基于数值预报和随机森林算法的强对流天气分类预报技术[J]. 气象, 44(12): 1555-1564. Li W J, Zhao F, Li M J, et al, 2018. Forecasting and classification of severe convective weather based on numerical forecast and random forest algorithm[J]. Meteor Mon, 44(12): 1555-1564 (in Chinese).
- 刘聚涛,杨永生,高俊峰,等,2011. 太湖蓝藻水华分级及其时空变化[J]. 长江流域资源与环境, 20(2): 156-160. Liu J T, Yang Y S, Gao J F, et al, 2011. Characteristics of cyanobacteria bloom grading and its temporal and spatial variation in Taihu Lake[J]. Resour Environ Yangtze Basin, 20(2): 156-160(in Chinese).
- 刘扬,王维国,2020. 基于随机森林的暴雨灾害人口损失预估模型及应用[J]. 气象, 46(3): 393-402. Liu Y, Wang W G, 2020. Assessing model of casualty loss in rainstorms based on random forest and its application[J]. Meteor Mon, 46(3): 393-402 (in Chinese).
- 刘苑,陈宇炜,邓建明,2010. YSI(多参数水质检测仪)测定叶绿素 a 浓度的准确性及误差探讨[J]. 湖泊科学, 22(6): 965-968. Liu Y, Chen Y W, Deng J M, 2010. Discussion on accuracy and errors for phytoplankton chlorophyll-a concentration analysis using YSI (multi-parameter water analyzer)[J]. J Lake Sci, 22(6): 965-968(in Chinese).
- 罗建美,霍永伟,韩晓庆,2017. 基于 HJ 卫星的近岸 II 类水体叶绿素 a 浓度定量遥感反演研究—以滦河口北部海域为例[J]. 海洋学报, 39(4): 117-129. Luo J M, Huo Y W, Han X Q, 2017. Inversion of chlorophyll a concentration in offshore II waters using HJ satellite data—example in the north of the Luanhe Delta[J]. Acta Oceanol Sin, 39(4): 117-129(in Chinese).
- 罗晓春,杭鑫,曹云,等,2019. 太湖富营养化条件下影响蓝藻水华的主导气象因子[J]. 湖泊科学, 31(5): 1248-1258. Luo X C, Hang X, Cao Y, et al, 2019. Dominant meteorological factors affecting cyanobacterial blooms under eutrophication in Lake Taihu[J]. J Lake Sci, 31(5): 1248-1258(in Chinese).
- 潘应阳,国巧真,孙金华,2017. 水体叶绿素 a 浓度遥感反演方法研究进展[J]. 测绘科学, 42(1): 43-48. Pan Y Y, Guo Q Z, Sun J H, 2017. Advances in remote sensing inversion method of chlorophyll a concentration[J]. Sci Surv Mapp, 42(1): 43-48 (in Chinese).
- 秦伯强,胡维平,陈伟民,等,2004. 太湖水环境演化过程与机理[M]. 北京:科学出版社:136-137. Qin B Q, Hu W P, Chen W M, et al, 2004. Process and Mechanism of Environmental Changes of the Taihu Lake[M]. Beijing: Science Press: 136-137 (in Chinese).

- 宋挺,周文麟,刘军志,等,2017.利用高光谱反演模型评估太湖水体叶绿素 a 浓度分布[J].环境科学学报,37(3):888-899. Song T, Zhou W L, Liu J Z, et al, 2017. Evaluation on distribution of chlorophyll-a content in surface water of Taihu Lake by hyperspectral inversion models[J]. Acta Sci Circum, 37(3): 888-899 (in Chinese).
- 水利部太湖流域管理局. (2019-12-05). 太湖健康状况报告(2018) [EB/OL]. http://www.tba.gov.cn/slbthlyglj/thjzkzkg/content/slth1_09f7d6b21629439f9891c7fd70ad49d8.html. Taihu Basin Administration of Ministry of Water Resources. (2019-12-05). Taihu health report (2018) [EB/OL]. http://www.tba.gov.cn/slbthlyglj/thjzkzkg/content/slth1_09f7d6b21629439f9891c7fd70ad49d8.html(in Chinese).
- 孙全德,焦瑞莉,夏江江,等,2019.基于机器学习的数值天气预报风速订正研究[J].气象,45(3):426-436. Sun Q D, Jiao R L, Xia J J, et al, 2019. Adjusting wind speed prediction of numerical weather forecast model based on machine learning methods[J]. Meteor Mon, 45(3): 426-436(in Chinese).
- 王可心,包云轩,朱承瑛,等,2021.随机森林回归法在冬季路面温度预报中的应用[J].气象,47(1):82-93. Wang K X, Bao Y X, Zhu C Y, et al, 2021. Forecasts of road surface temperature in winter based on random forests regression[J]. Meteor Mon, 47(1): 82-93(in Chinese).
- 王桥,张兵,韦玉春,等,2008.太湖水体环境遥感监测实验及其软件实现[M].北京:科学出版社. Wang Q, Zhang B, Wei Y C, et al, 2008. Remote Sensing Monitoring Experiment of Taihu Lake Water Environment and Its Software Implementation[M]. Beijing: Science Press(in Chinese).
- 吴传庆,杨志峰,王桥,等,2009.叶绿素 a 浓度的动态峰反演方法[J].湖泊科学,21(2):223-227. Wu C Q, Yang Z F, Wang Q, et al, 2009. A reverse method of chlorophyll-a based on dynamic apex[J]. J Lake Sci, 21(2): 223-227(in Chinese).
- 杨婷,张慧,王桥,等,2011.基于 HJ-1A 卫星超光谱数据的太湖叶绿素 a 浓度及悬浮物浓度反演[J].环境科学,32(11):3207-3214. Yang T, Zhang H, Wang Q, et al, 2011. Retrieving for chlorophyll-a concentration and suspended substance concentration based on HJ-1A HIS image[J]. Environ Sci, 32(11): 3207-3214 (in Chinese).
- 尹艳娥,沈新强,蒋玫,等,2014.长江口及邻近海域富营养化趋势分析及与环境因子关系[J].生态环境学报,23(4):622-629. Yin Y E, Shen X Q, Jiang M, et al, 2014. Analysis on the trend of eutrophication in the Changjiang (Yangtze River) Estuary and in the adjacent East China Sea and its relation to environmental factors[J]. Ecol Environ Sci, 23(4): 622-629(in Chinese).
- 张明慧,苏华,季博文,2018. MODIS 时序影像的福建近岸叶绿素 a 浓度反演[J].环境科学学报,38(12):4831-4839. Zhang M H, Su H, Ji B W, 2018. Retrieving nearshore chlorophyll-a concentration using MODIS time-series images in the Fujian Province (China)[J]. Acta Sci Circum, 38(12): 4831-4839(in Chinese).
- 张鹏,王春姣,陈林,等,2018.沙尘气溶胶卫星遥感现状与需要关注的若干问题[J].气象,44(6):725-736. Zhang P, Wang C J, Chen L, et al, 2018. Current status of satellite-based dust aerosol remote sensing and some issues to be concerned[J]. Meteor Mon, 44(6): 725-736(in Chinese).
- 张玉超,钱新,钱瑜,等,2009a.基于机器学习方法的太湖叶绿素 a 定量遥感研究[J].环境科学,30(5):1321-1328. Zhang Y C, Qian X, Qian Y, et al, 2009a. Quantitative retrieval of chlorophyll a concentration in Taihu Lake using machine learning methods [J]. Environ Sci, 30(5): 1321-1328(in Chinese).
- 张玉超,钱新,钱瑜,等,2009b.支持向量机在太湖叶绿素 a 非线性反演中的应用[J].中国环境科学,29(1):78-83. Zhang Y C, Qian X, Qian Y, et al, 2009b. Application of SVM on Chl-a concentration retrievals in Taihu Lake[J]. China Environ Sci, 29(1): 78-83(in Chinese).
- 赵少华,刘思含,刘芹芹,等,2019.中国城镇生态环境遥感监测现状及发展趋势[J].生态环境学报,28(6):1261-1271. Zhao S H, Liu S H, Liu Q Q, et al, 2019. Progress of urban ecological environment monitoring by remote sensing in China[J]. Ecol Environ Sci, 28(6): 1261-1271(in Chinese).
- 朱广伟,秦伯强,张运林,等,2018.2005-2017 年北部太湖水体叶绿素 a 和营养盐变化及影响因素[J].湖泊科学,30(2):279-295. Zhu G W, Qin B Q, Zhang Y L, et al, 2018. Variation and driving factors of nutrients and chlorophyll-a concentrations in northern region of Lake Taihu, China, 2005-2017[J]. J Lake Sci, 30(2): 279-295(in Chinese).
- 朱云芳,朱利,李家国,等,2017.基于 GF-1WV 影像和 BP 神经网络的太湖叶绿素 a 反演[J].环境科学学报,37(1):130-137. Zhu Y F, Zhu L, Li J G, et al, 2017. The study of inversion of chlorophyll a in Taihu based on GF-1 WV image and BP neural network[J]. Acta Sci Circum, 37(1): 130-137(in Chinese).
- Breiman L, 2001. Random forests[J]. Mach Learn, 45(1): 5-32.
- He J Y, Chen Y J, Wu J P, et al, 2020. Space-time chlorophyll-a retrieval in optically complex waters that accounts for remote sensing and modeling uncertainties and improves remote estimation accuracy[J]. Water Res, 171: 115403.
- Kong X Y, Sun Y Y, Su R G, et al, 2017. Real-time eutrophication status evaluation of coastal waters using support vector machine with grid search algorithm[J]. Mar Pollut Bull, 119(1): 307-319.
- Lary D J, Alavi A H, Gandomi A H, et al, 2016. Machine learning in geosciences and remote sensing[J]. Geosci Front, 7(1): 3-10.
- Nazeer M, Nichol J E, 2016. Development and application of a remote sensing-based Chlorophyll-a concentration prediction model for complex coastal waters of Hong Kong[J]. J Hydrol, 532: 80-89.
- Zhang F F, Zhang B, Li J S, et al, 2014. Validation of a synthetic chlorophyll index for remote estimates of chlorophyll-a in a turbid hypereutrophic lake[J]. Int J Remote Sens, 35(1): 289-305.
- Zhao D H, Lv M T, Jiang H, et al, 2013. Spatio-temporal variability of aquatic vegetation in Taihu Lake over the past 30 years[J]. PLoS One, 8(6): e66365.