一种基于 MODIS 影像可业务化运行的巢湖水体叶绿素 a 估算算法*

陶 慇^{1,2}, 段洪涛^{1**}, 齐 琳^{1,2}, 张玉超¹, 马荣华¹

(1: 中国科学院南京地理与湖泊研究所湖泊与环境国家重点实验室, 南京 210008)

(2: 中国科学院大学, 北京 100049)

摘 要: 现有水色卫星主要是针对大洋清洁水体设计, 内陆浑浊水体多数波段经常饱和; 而发展可以业务化运行的内陆水体叶绿素 a 算法, 为生产实践服务, 一直是水色遥感的重点和难点之一. 利用 2013 年巢湖星地同步数据 ($N=55$), 通过经验正交函数 (empirical orthogonal function, EOF) 分析方法, 选用 MODIS 唯一不饱和的 4 个波段 (469, 555, 645, 859 nm) 数据进行分解, 然后回归建模; 并使用第三方独立的巢湖实测数据 ($N=40$) 进行验证 ($R^2=0.63$, $URMSE=85.46\%$). 结果表明: 该算法用于 MODIS 影像上, 空间分布合理, 季节差异明显, 且在高悬浮物水体、不同气溶胶条件下均有很好的抗扰动性. 实践证明 EOF 算法可以应用于业务化运行的内陆水体叶绿素 a 浓度估算, 并对其他水色参数反演具有一定的借鉴意义.

关键词: 业务化运行; 叶绿素 a; MODIS; 经验正交函数; 内陆水体; 巢湖

An operational algorithm to estimate chlorophyll-a concentrations in Lake Chaohu from MODIS imagery

TAO Min^{1,2}, DUAN Hongtao¹, QI Lin^{1,2}, ZHANG Yuchao¹ & MA Ronghua¹

(1: State Key Laboratory of Lake Science and Environment, Nanjing Institute of Geography and Limnology, Chinese Academy of Sciences, Nanjing 210008, P. R. China)

(2: University of Chinese Academy of Sciences, Beijing 100049, P. R. China)

Abstract: Ocean color satellite is mainly designed for oceanic waters, but most of their bands are easily saturated in highly turbid waters. It's a great challenge to monitor water quality using satellites, because there are no operational algorithm to estimate chlorophyll-a concentration (Chl. a) in inland waters. In this paper, a novel approach based on the Empirical Orthogonal Function (EOF) analysis was developed using the spectral variance from the Rayleigh-corrected reflectance data (R_{rc}) at 469, 555, 645 and 859 nm, and Chl. a was estimated by considering the regression relationships between the spectral variance and 55 concurrent field measurements. The validation was then performed using independent data ($N=40$) from other laboratory, with $R^2=0.63$, $URMSE=85.46\%$. Application of the algorithm to MODIS images showed that spatial distribution patterns and seasonal changes are reasonable. Besides, this algorithm is immune to perturbations from thick aerosols and different sediments. Hence, we suggest that EOF algorithm could be applied to estimate Chl. a concentrations in inland waters.

Keywords: Operational algorithm; chlorophyll-a; MODIS; empirical orthogonal function; inland water; Lake Chaohu

巢湖位于安徽省, 是我国第 5 大淡水湖, 是长江中下游重要的淡水资源和生态湿地^[1], 在当地人民生活和经济发展中发挥着重要的作用. 近年来, 随着巢湖流域经济的快速发展, 大量点源、面源污染物进入水体, 巢湖水质污染严重; 另外, 巢湖闸、裕溪闸等的修建使巢湖逐渐成为半封闭水域^[2], 水体自净功能逐渐丧失, 巢湖富营养化情况日益严峻. 为了保证巢湖水质安全和流域经济可持续发展, 亟需加强水质状况的实时监测. 卫星遥感具有大范围、周期性和快捷性等特点, 已成为湖泊水质实时监测的重要手段. 叶绿素 a (Chl. a)

* 国家高技术研究发展计划“863”项目 (2014AA06A509) 和国家自然科学基金重点项目 (41431176) 联合资助.
2014-11-10 收稿; 2015-04-10 收修改稿. 陶慇 (1991~), 女, 硕士研究生, E-mail: tmalbee@126.com.

** 通信作者; E-mail: htduan@niglas.ac.cn.

是所有藻类的共同色素,是表征湖泊水质变化程度的重要指标。因此,利用卫星遥感定量估测 Chl. a 浓度,可以很好地评估湖泊水质,具有重要的科学价值和实际研究意义。

对于内陆水体 Chl. a 浓度的估算方法,以水体辐射传输理论和流体力学为基础的生物光学模型^[3],可以很好地解释水体中生物光学机理问题。但由于内陆湖泊等Ⅱ类水体物质来源复杂,水体间光学特性差异较大,并且卫星数据波段数量很有限,导致这种方法很难直接应用于卫星上,因此无法实现业务化运行。而针对 Chl. a 吸收或者反射峰,提取相关峰值信息的形状指数模型(MPH、FLH 算法),已证实具有很大的潜力并且能够很好地应用于卫星上。如 Hu 等基于 SeaWiFS 和 MODIS 三波段差值提出的 OCI 算法,改变 NASA 近 40 年来一直延续使用的算法(OC1-4),2014 年被 NASA 采用作为所有水色卫星传感器(MODIS、VIIRS、MERIS、SeaWiFS、CZCS、OCTS)的标准算法^[4]。但卫星遥感数据中特征波段的缺失导致这类算法无法使用,这些算法的应用在很大程度上受限于大气校正的效果。除了以上的半经验和半分析算法,近年来一些人工智能模型也应用到 Chl. a 浓度反演上,包括神经网络模型^[5]、支持向量机算法^[6]、经验正交函数(empirical orthogonal function, EOF)分解算法等,这类人工智能类算法只分析数值变化的特点,对波段并无特殊要求;并且基于实测光谱 EOF 分解的研究成果表明,该方法在反演沿海 Chl. a 浓度的应用效果良好^[7]。

目前常用于水色遥感的卫星传感器主要包括 SeaWiFS、MERIS、MODIS 等^[8]。考虑到巢湖等长江中下游湖泊多为浅水湖泊,平均水深只有 2.69 m,由于受风浪等影响容易引起沉积物再悬浮,一天之内水质(特别是 Chl. a 浓度)变化较大,对于卫星数据时相要求较高;而 MODIS 一天之内过境 2 次,数据获取较为容易,便于业务化监测湖泊水质,是在轨传感器中相对理想的数据源^[9]。需要说明的是,MODIS Terra/Aqua 传感器海洋波段主要针对大洋清洁水体设计,比较敏感,在内陆浑浊水体中多数波段经常饱和^[4],大气校正(主要是气溶胶校正)困难,导致无法获得足够的波段及其遥感反射比数据。事实上,目前还没有完全适用于内陆浑浊水体的传感器。

本文以巢湖为研究区,尝试利用 MODIS 只经过瑞利校正的反射率数据(rayleigh corrected reflectance, R_{re}),不依赖于多数算法中常用的 Chl. a 特征峰波段,而是用纯粹的 EOF 分解技术,揭示 R_{re} 数据的变化模式,构建一种能够业务化运行的巢湖水体 Chl. a 浓度反演算法。由于瑞利校正后得到的巢湖地区 R_{re} 波段较少,只有 4 个波段(469、555、645、859 nm);因此,针对没有去除气溶胶影响的 R_{re} 数据,利用对波段数量要求不高的 EOF 分解方法进行 Chl. a 浓度业务化反演,也许是目前比较有效的一种解决方法。

1 数据与方法

1.1 研究区概况

巢湖(31°25'28"~31°43'28"N, 117°16'54"~117°51'46"E)地处安徽省中部,位于长江流域下游,是我国五大淡水湖之一(图 1)。巢湖水域面积约 769.55 km²,平均水深 2.69 m,多年平均水位 8.37 m,蓄水量 20.7 × 10⁸ m³^[1]。受当地经济快速发展的影响,巢湖水域富营养化问题严重,蓝藻水华频繁暴发。

1.2 实测数据采集与处理

1.2.1 水样采集 2013 年 5—10 月期间,在巢湖分别进行 5 次野外实验,具体日期和样点分布如图 1 所示。每个样点采集表层水样用于化学分析,同时测量透明度和光谱,并利用 GPS 记录采样位置和采样时间等辅助数据。

1.2.2 水质参数分析 水体经 1.2 μm 的 Whatman GF/F 玻璃纤维膜过滤后避光保存,使用液氮反复冻融 3 次后,加入 90% 丙酮溶液避光提取 12~15 h,离心后采用分光光度法测定 Chl. a 浓度^[10]。悬浮物采用称重法测定。用 0.7 μm 的 Whatman GF/F 滤膜过滤水样后,烘箱 105℃ 条件下烘 4~6 h 后称重并减去膜重,得到总悬浮物(SPM)浓度^[11-13]。

1.3 MODIS 卫星数据收集与处理

MODIS 卫星数据处理主要分为 3 步:①从 NASA 网站上选取与实测叶绿素同步的 0 级数据,然后进行辐射定标得到 LIB 数据;②利用 SeaDAS 软件去除气体吸收(臭氧吸收)和瑞利散射(分子效应)的影响,从而得到瑞利校正后的反射率数据;③根据实测采样点的经纬度位置信息,提取实测采样点对应的 R_{re} 数据,每个 R_{re} 数据都通过 3 × 3 窗口检验,以减少杂散光的影响并保证空间同质性^[14]。

图 1 巢湖水体采样点位置分布

Fig.1 Spatial distribution of sampling sites in Lake Chaohu

由于巢湖水体较为浑浊,MODIS 水色波段通常处于饱和状态;469 ~ 2130 nm 间的 7 个陆地波段,由于动态范围较宽、敏感性低因此很少饱和^[15]. 对前 6 个波段的 R_{re} 数值进行盒图分析(图 2),可以看到第 5 波段(1240 nm)易受噪声影响,离群值太大不能用于建模;第 6 波段(1640 nm)的 R_{re} 值普遍偏小,基本接近 0,对 EOF 分解结果意义不大. 因此,本文选取 469、555、645 和 859 nm 前 4 个波段用于 EOF 建模.

图 2 MODIS- R_{re} 6 个波段的盒图分析

Fig.2 Boxplot analysis of 6 bands using MODIS R_{re}

4 个波段的 R_{re} 光谱如图 3a 所示. 通过对光谱特征的分析发现,不同 Chl. a 浓度下 R_{re} 光谱变化的特征很类似,但振幅完全不同. 值得注意的是,有 2 个实测点 859 nm 波段的 R_{re} 反射率明显高于其他点,且在 645 nm 波段出现吸收峰,这 2 个点位代表水华发生时 R_{re} 光谱的变化情况,这样的点位于建模可以提高模型的适用性.

1.4 模型训练点的筛选

部分训练点由于受到影像质量和大气条件的影响,光谱形状异常,这些训练点会影响模型的精度,因此需要对模型训练点进行筛选. 通过把所有训练点位,都标记在对应天数的 RGB 原图上,删除靠近影像边缘,易受水陆混合效应影响的样点;另外,根据当天的影像质量,需要删除天气、影像明显损坏等有问题,以及光

谱特征比较异常的样点. 筛选后用于建模的样点为 55 个,有效样点水质参数的统计结果(表 1)表明,用于建模的数据分布范围很宽,Chl. a 最小浓度为 8.19 μg/L,最大浓度可以达到 727.16 μg/L,SPM 浓度的变化范围也比较大,基于该数据集构建的模型适用性更广.

表 1 EOF 建模的有效样点的水质参数统计
Tab. 1 Statistics of water quality parameters of EOF training data

采样批次	采样点	Chl. a/(μg/L)			SPM/(mg/L) *		
		均值	最小值	最大值	均值	最小值	最大值
201305	15	48.69 ± 57.72	8.19	176.66	39.67 ± 15.86	20	85
201306	6	96.19 ± 83.22	15.16	254.13	63.83 ± 26.38	27	102
201307	18	39.41 ± 17.02	13.27	69.06	85.64 ± 44.30	38	182
201309	13	92.73 ± 93.76	20.11	370.94	39.23 ± 10.45	20	57
201310	3	488.12 ± 229.64	178.21	727.16	149.33 ± 76.01	42	208

* 缺少 7 个 5 月 SPM 数据.

2 EOF 算法构建、验证与应用

2.1 EOF 方法概述

EOF 方法是一种数学统计方法,可用于分析变量的结构特征,这种数学分析的方法最初是引入气象和气候学的研究中,近年来,有关海岸带水体反射率光谱与水体组分含量的研究中也逐渐采用该方法^[16]. EOF 的结果是对变量数据的所有变化模式进行分析之后,得出一组正交不相关的变量集. 第一模态代表最大部分的贡献方差,后几个模态逐渐减少,贡献方差表征对变量变化模式的解释程度,因此选择前几个模态,就可以在不损失变量信息的情况下进行重新归类.

对 R_{rc} 数据进行 EOF 分析,即对 R_{rc} 数据进行重新归类,对 Chl. a 总体变化贡献显著的波段数据进行重新组合. 另外,回归分析的一个重要前提要求变量不能线性相关,而 EOF 分解的模态彼此正交不相关,便于进行 Chl. a 反演的回归建模.

2.2 基于 EOF 分解方法模型构建

2.2.1 光谱归一化预处理 对 R_{rc} 数据进行 EOF 分析表明,第一模态含有 89.52% 的贡献方差(图 3c),除第一波段外,它的形状与 R_{rc} 平均形状类似,这表明它所描述的是光谱振幅的变化. 对 R_{rc} 进行归一化之后(公式(1))得到 NR_{rc} (normalized reflectance raleigh corrected, NR_{rc}) 光谱^[17],归一化过程大大减弱不同浓度的光谱分散程度,光谱变化的模式更容易被识别(图 3b). 累积方差结果(图 3c 和 3d)表明,归一化对各模态累积方差的改变并不大,归一化后前 3 个模态累计贡献方差达 100%,而归一化前前 3 个模态累计贡献方差只有 99.47%.

$$NR_{rc}(\lambda) = \frac{R_{rc}(\lambda)}{\int_{469}^{859} R_{rc}(\lambda) d\lambda}$$

(1)

光谱归一化并不影响对光谱变量的解释程度,然而归一化后各个模态的形态却发生重大变化. 归一化后第一模态的变化形态与归一化前第二模态的变化形态类似,但贡献方差(图 3d)的比例更大(58.24%),归一化前(图 3c)第二模态的贡献方差只有 7.03%,这种变化表明对 R_{rc} 数据进行归一化处理非常关键,它更加真实地捕捉光谱形状的变化模式,从而可以更精确地研究光谱变化所驱动的 Chl. a 浓度变化.

2.2.2 EOF 建模理论基础 对光谱预处理后,利用 EOF 分解 NR_{rc} 数据计算出每个模态的得分值,不同 Chl. a 浓度下的得分值呈现一定的特征. 将 Chl. a 浓度分为 4 个梯度——高浓度:大于 100 μg/L;中浓度:50 ~ 100 μg/L;正常:20 ~ 50 μg/L;低浓度:小于 20 μg/L,对 NR_{rc} 光谱分别进行渲染,显然不同 Chl. a 浓度下的 NR_{rc} 是杂乱无序的(图 4a),归一化过程并没有把不同 Chl. a 浓度区分出来. 若以 4 个模态为横坐标,以

图3 建模数据点: R_{rc} 光谱特征(a)和 NR_{rc} 光谱特征(b);
 R_{rc} (c)和 NR_{rc} (d) EOF分解后的模态特征

Fig. 3 MODIS R_{rc} (a) and NR_{rc} (b) spectra (469, 555, 645 and 859 nm) corresponding to the 55 *in situ* measurements; Loadings, variance explained represented for modes 1–4 of EOF analysis of R_{rc} (c) and NR_{rc} (d)

图4 不同 Chl. a 浓度下的 NR_{rc} (a)和 EOF 分解得分值(b)

Fig. 4 NR_{rc} spectra (a) and scores after EOF analysis (b) under different Chl. a concentrations

每个模态的得分值为纵坐标按不同梯度 Chl. a 浓度进行渲染,结果表明经过 EOF 分解后,各波段重新线性组合的结果使得不同 Chl. a 浓度条件下的得分谱具有一定特征(图 4b),即 EOF 分解提取出了波段贡献的

数值特征. 正是因为存在这样的特征,才能以每个模式的得分值为因变量作 Chl. a 浓度的回归分析,这是进行 EOF 回归建模的基础.

2.2.3 模型构建过程 模型构建可以简单分为 3 个过程,具体如下:

1) 对 R_{re} 数据进行归一化,得到 NR_{re} 数据. 对归一化光谱利用 MATLAB™ 的 princomp 函数进行 EOF 分解. EOF 分解的输出包括 EOF 每个模式的得分向量,每个得分向量都是原 4 个波段的线性组合,即波段的重新归类. 除此之外,还包括每个波段的载荷值,即原波段进行线性组合的系数,还有每个模式的贡献方差,用于说明描述原波段变化模式的解释程度.

2) 对训练点位进行筛选,确定训练点位集,并根据所得到的 EOF 各个模式的得分值与实测 Chl. a 浓度作多元逐步回归分析. 使用 MATLAB™ 的 regress 函数,以每个点位 4 个模式的得分值 T_1 、 T_2 、 T_3 、 T_4 为自变量,对应实测 Chl. a 浓度为因变量,可以得到回归系数($\beta_0 \sim \beta_4$):

$$\beta_0 + \beta_1 T_1 + \beta_2 T_2 + \beta_3 T_3 + \beta_4 T_4 = \text{实测 Chl. a}$$

(2)

3) 将 EOF 算法应用于 MODIS 影像上,对 R_{re} 影像上所有有效像素进行归一化处理,经 EOF 分解得到每个像元的 EOF 得分. $\beta_0 \sim \beta_4$ 是构建模型时训练点位的回归系数, $T_1 \sim T_4$ 是影像上每个像素的 EOF 得分. 因此,反演所得到的 Chl. a 浓度为:

$$\text{反演 Chl. a} = \beta_0 + \beta_1 T_1 + \beta_2 T_2 + \beta_3 T_3 + \beta_4 T_4$$

(3)

基于 55 个训练点位模拟出来的算法结果如图 5a 所示,表明 EOF 模拟 Chl. a 浓度和实测 Chl. a 浓度之间具有显著相关关系($R^2 = 0.94$),无偏差 $RMSE(URMSE)^{[18]}$ 为 65.27%. 实测 Chl. a 浓度 20 $\mu\text{g/L}$ 以下的点位相较于其他浓度下的点位回归情况,分散度会更高一些. 鉴于 EOF 模型只使用 MODIS 的 4 个非特征峰波段的前提下,该算法的性能是可以接受的.

2.3 模型验证与精度评价

利用第三方提供的 5 个采样点 8 次野外巡测(即 2012 年 3、5、10、12 月和 2013 年 4、5、10、11 月 8 次野外调查)独立 Chl. a 浓度数据集($N = 40$)进行算法验证. 验证数据覆盖湖区的不同位置,并且在没有参与构建模型的 3、4、11 月均有数据,这样的数据集更有利于检验算法的应用效果. 验证结果(图 5b)表明,数据的相关性较好($R^2 = 0.63$),误差范围也可以接受($URMSE = 85.46\%$);在 1:1 回归线附近分散程度略高的验证点,可能是由于野外采集及室内提取 Chl. a 浓度过程中存在一定偏差,同时尺度效应所带来的单点误差,都会对验证结果带来一定的影响.

2.4 巢湖 Chl. a 浓度时空变化

野外实测的 Chl. a 浓度受到时间和空间的限制,因此,验证结果可能无法完全检验算法的性能. 为了进

图 5 EOF 建模的模型误差(a);独立数据集的 EOF 模型验证结果(b)

Fig. 5 Relationships between *in situ* measured Chl. a concentrations and estimated Chl. a concentrations from MODIS(a); Validation using independent field measurements from the 6 fixed stations (b)

一步验证 EOF 方法反演 Chl. a 浓度性能,也是为了检验模型业务化运行的能力,将该算法用于 MODIS 卫星影像上来检查空间分布和季节差异是否合理.

本文获取了 2013 年 130 景的 MODIS 影像,通过 EOF 方法进行批量反演,对得到的 MODIS Chl. a 浓度分布图进行逐一检查,对照真彩色 RGB,以确定相同位置 Chl. a 浓度的空间格局是否一致,并分析 Chl. a 浓度空间分布特征的季节平均变化趋势(图 6b~e). Chl. a 浓度较高的地区主要位于巢湖西半湖地区,东半湖的浓度略小一些(图 6a),这与夏季平均 Chl. a 浓度(图 6c)的变化规律一致,而 RGB 图像上中部地区出现小片区域的藻华,在反演的结果图上也展现出,这个现象也进一步证明该算法是合理可信的.

图 6 EOF 算法应用于巢湖影像结果(a);2013 年巢湖 Chl. a 浓度季节分布(b~e)

Fig. 6 The EOF algorithm applied to MODIS images (a) and the seasonal mean Chl. a distribution in 2013 (b-e)

3 讨论

3.1 训练点位选择

EOF 属于统计学方法,它要求抽样必须代表总体才有效,因此训练数据集必须反映总体的特征. 本文使用的光谱数据集是人为抽样的,影响到参数估计,无法代表总体参数的点位不能参与参数估计. 如果更改数据集,回归系数将产生很大的变化,本文的模拟结果也证实了这一结论:从原始数据集随机抽取 N 个数据 ($17 < N < 55$) 进行回归分析,抽取数据集的过程重复 100 次用于减小随机误差,模拟后得到 $\beta_0 \sim \beta_4$ 的平均值和标准差的统计结果如表 2 所示,各个系数的变化范围均比较大,数据离散程度高,系数不确定性较强. 利用 SPSS 软件对 100 次模拟的参数 $\beta_0 \sim \beta_4$ 进行单样本 t 检验,除常数项 β_0 外,其它参数的变化差异显著

($Sig < 0.05$), 而 $\beta_1 \sim \beta_4$ 与得分值直接相乘, 它的变化直接影响 Chl. a 浓度的反演结果. 因此, 在算法建立的过程中, 数据集的挑选十分重要, EOF 算法对训练样本的质量要求较高. 综上可知, 训练光谱点位、挑选合适的数据集是模型成功的关键所在.

表 2 更换数据集系数的统计结果*
Tab. 2 Statistics of the regression coefficient

系数	β_0	β_1	β_2	β_3	β_4
平均值	83.615	314381.185	-192407.299	171477.181	0
标准差	14.154	114284.267	146545.105	149622.578	0
Sig 值	0.265	0	0	0	0

* 系统默认小于 0.001 时记为 0.

光谱形状异常的离群值固然会影响到 EOF 得分的提取, 但是还有一些光谱形状异常的点位并非离群值, 它们处于 EOF 得分的聚集区, 但是回归之后的结果与实测值的差值较大, 因此影响回归方程的建立. 对于这样的点位, 仅依靠前 2 个 EOF 得分的形状无法进行剔除, 因此对点位进行标注的步骤非常有必要, 这样方便找到异常目标点位进行删除. 由于 4 个变量的 EOF 属于高维空间分解, 所以无法直观展示异常光谱点位对 EOF 分解的影响. 根据前 3 个模态的累计贡献方差图 (图 7a), 前 2 个模态的 EOF 累计贡献方差达到 85%, 可以基本保证波段信息没有损失, 因此选择前两个主成分得分, 二维散点图方便说明问题. 将异常光谱 (图 7b) 的异常点位进行删除, 所有点位的得分值均发生重大变化, 这样的离群点位不能参与参数回归, 因此, 这类异常光谱通常需要删除.

图 7 EOF 各模态累计贡献方差(a)和不同光谱下的 EOF 形状(b)

Fig. 7 Variance explained for modes 1-4 of EOF analysis (a) and EOF shape of different spectral(b)

3.2 训练点位阈值的确定

在不断筛选点位的过程中, 面临着使用 EOF 方法时训练样本的数目多少才会合适的问题. 因此本文使用统计学中的刀切法 (Jackknife) 对训练集进行重采样来确定样本数目的阈值. 刀切法是再抽样的一种统计方法^[19], 每次从训练集中删除一些样本, 剩余的样本成为刀切集, 由一系列这样的刀切集计算统计量的估计值. 刀切法要求样本集需要事先给定, 恰好符合我们的训练点位集已经确定的情况. 对训练集的所有 55 个点位进行 EOF 得分值的计算, 随机抽取 N 个点的得分值作为刀切集, 与对应的实测 Chl. a 浓度进行多元逐步回归分析, 得到的回归系数用于拟合余下 $(55 - N)$ 个点位的得分值, 与实测 Chl. a 浓度进行验证, 并计算验证结果的 R^2 以及 $RMSE$. 选取点位的过程重复 1000 次用于减小随机误差, 并提高运算效率.

结果表明不同训练点的个数条件下, $RMSE$ (图 8a) 存在一个急剧下降的过程, 7 个训练点 $RMSE = 0.32$ 到 17 个训练点条件下 $RMSE = 0.29$, $RMSE$ 迅速减小. 多于 17 个训练点时 $RMSE$ 的降低率明显变缓. 同样, R^2 (图 8b) 也呈现同样的规律, 从 7 个训练点 $R^2 = 0.70$, 到 17 个训练点条件下 $R^2 = 0.80$. 而当训练点大于

17 时, R^2 增加的速率大幅下降. 虽然之后随着训练点数量的增加, R^2 仍然呈现上升趋势, 但上升的幅度很小, 从 17 个训练点条件下 $R^2 = 0.80$, 到 37 个训练点下 $R^2 = 0.84$. 这个结果证明多于 17 个样本点对 EOF 反演 Chl. a 浓度的模型性能只有很小的改进, 这也提供了用于构建 EOF 模型样本数目的最低阈值, 对再使用此类算法进行参数反演具有一定的借鉴意义.

图 8 不同训练点条件下的 $RMSE$ (a) 和 R^2 (b)

Fig. 8 Number of training points versus Chl. a jackknife $RMSE$ (a) and training points versus jackknife R^2 (b)

3.3 高悬浮物水体 EOF 适用性

2013 年 3 月 15 日的影像显示水体整体呈现白色, 水体悬浮物含量较高(图 9a); 而通过 EOF 方法估算的 Chl. a 浓度(图 9b)并没有受到悬浮物的显著影响, 整体呈现出较为合理的空间分布; 特别是巢湖中部和南部, Chl. a 浓度整体偏低($< 30 \mu\text{g/L}$), 与巢湖春季(图 6b)特征符合, 而此处图像显示尤为浑浊, 证实 EOF 方法对于高悬浮物水体具有较好效果.

3.4 云、大气的影晌对算法应用的影像

EOF 算法只使用经过瑞利校正的 MODIS 数据, 避免大气校正带来的误差, 因此它对不同条件下的气溶胶浓度均表现出良好的应用效果. 图 10 展示算法在不同的气溶胶条件下的性能, 根据真彩色 RGB 图像的直观显示, 2 幅影像的气溶胶条件不同, b 图的气溶胶浓度显然要高于 a 图的情况^[20]. 而 EOF 反演的结果表明, 这 2 幅连续 2 天的影像在 Chl. a 的空间格局和浓度上大体趋于一致, 并没有因为气溶胶厚度的增大带来空间的异常, 因此也说明 EOF 算法对不同的气溶胶条件具有很好的抗扰动力.

对于有云条件下的影像, 如果不做掩膜将云覆盖的部分进行去除, 对整幅 R_{rs} 影像进行 EOF 分解后, 会导致无云部分的得分值偏高, 随之估算得到的 Chl. a 浓度也会普遍偏高或者局部出现异常. 因此, 需要根据 RGB 图像先勾出云层的掩膜, 再进行 Chl. a 浓度估算.

4 结论

EOF 分解很好地揭示了 R_{rs} 光谱的变化模式, 从而与 Chl. a 浓度的变化进行回归分析, 结果表明, 该模型可以利用很少的波段(4 个)有效估算巢湖水体 Chl. a 浓度, 空间分布合理, 季节差异明显, 符合巢湖实际情况. 该方法在高悬浮物水体、不同气溶胶条件下都具有很好的表现, 适用性较强. 另外, EOF 算法比较容易操作, 运行的时间效率较高, 通常一景影像在普通电脑上运行 1.8478 s 内就可完成, 满足业务化运行的需求. 反演结果精确、应用性能稳定, 因此, 该方法可以用于巢湖水体 Chl. a 浓度的业务化监测.

需要注意的是, 随着更多高性能水色卫星的发射, 特别是我国高分系列卫星的快速发展, 相信有更多更好的卫星数据可以选择, EOF 方法将有更多的波段可以使用, 水色参数反演的精度将进一步提高, 具有更广阔的应用潜力.

致谢: 感谢中国科学院南京地理与湖泊研究所“湖泊—流域科学数据共享平台”提供数据.

图 9 高悬浮物水体 EOF 算法应用情况

Fig. 9 Performance of the Chl. a retrieval algorithm under high suspended solids

图 10 不同气溶胶厚度条件下 (a、b) EOF 算法应用情况

Fig. 10 Performance of the Chl. a retrieval algorithm under different aerosol conditions

5 参考文献

[1] 王苏民, 窦鸿身. 中国湖泊志. 北京: 科学出版社, 1998: 230.

[2] 臧小平, 吴国平, 涂 敏. 长江流域湖泊水库水华防治对策. 人民长江, 2009, (9): 5-8.

[3] Hu W, Jørgensen SE, Zhang F. A vertical-compressed three-dimensional ecological model in Lake Taihu, China. *Ecological Modelling*, 2006, **190**(3/4): 367-398.

[4] Hu C, Lee Z, Franz B. Chlorophyll a algorithms for oligotrophic oceans: A novel approach based on three-band reflectance difference. *Journal of Geophysical Research*, 2012, **117**(C1). DOI 10.1029/2011 JC 007395.

[5] Schiller H, Doerffer R. Neural network for emulation of an inverse model operational derivation of Case II water properties from MERIS data. *International Journal of Remote Sensing*, 1999, **20**(9): 1735-1746.

[6] 刘朝相, 宫兆宁, 赵文吉. 基于 SVM 模型的枞水河 Chl. a 浓度的遥感反演. 遥感技术与应用, 2014, **29**(3): 419-427.

- [7] Craig SE, Jones CT, Li WKW *et al.* Deriving optical metrics of coastal phytoplankton biomass from ocean colour. *Remote Sensing of Environment*, 2012, **119**:72-83.
- [8] 姜广甲, 周琳, 马荣华等. 浑浊Ⅱ类水体 Chl. a 浓度遥感反演(Ⅱ): MERIS 遥感数据的应用. 红外与毫米波学报, 2013, **32**(4): 372-378.
- [9] 马荣华, 孔维娟, 段洪涛等. 基于 MODIS 影像估测太湖蓝藻暴发期藻蓝素含量. 中国环境科学, 2009, **29**(3): 254-260.
- [10] Ma R, Tang J, Dai J. Bio-optical model with optimal parameter suitable for Taihu Lake in water colour remote sensing. *International Journal of Remote Sensing*, 2006, **27**(19):4305-4328.
- [11] Cleveland JS, Weidemann AD. Quantifying absorption by aquatic particles: A multiply scattering correction for glass-fiber filters. *Limnology and Oceanography*, 1993, **38**(6):1321-1327.
- [12] Duan H, Ma R, Hu C. Evaluation of remote sensing algorithms for cyanobacterial pigment retrievals during spring bloom formation in several lakes of East China. *Remote Sensing of Environment*, 2012, **126**:126-135.
- [13] Duan H, Feng L, Ma R *et al.* Variability of particulate organic carbon in inland waters observed from MODIS Aqua imagery. *Environmental Research Letters*, 2014, **9**(8):084011.
- [14] Le C, Hu C, English D *et al.* Climate-driven chlorophyll-a changes in a turbid estuary: Observations from satellites and implications for management. *Remote Sensing of Environment*, 2013, **130**:11-24.
- [15] Hu C, Chen Z, Clayton TD *et al.* Assessment of estuarine water-quality indicators using MODIS medium-resolution bands: Initial results from Tampa Bay, FL. *Remote Sensing of Environment*, 2004, **93**(3):423-441.
- [16] Mueller JL. Ocean color spectra measured off the Oregon coast: characteristic vectors. *Applied Optics*, 1976, **15**(2): 394-402.
- [17] Barnes BB, Hu C, Cannizzaro JP *et al.* Estimation of diffuse attenuation of ultraviolet light in optically shallow Florida Keys waters from MODIS measurements. *Remote Sensing of Environment*, 2014, **140**:519-532.
- [18] Hooker SB, Lazin G, Zibordi G *et al.* An evaluation of above-and in-water methods for determining water-leaving radiances. *Journal of Atmospheric and Oceanic Technology*, 2002, **19**(4): 486-515.
- [19] 李宝慧. 浅谈复杂样本方差估计的一种方法——刀切法. 统计教育, 1998, (5):24.
- [20] Qi L, Hu C, Duan H *et al.* A novel MERIS algorithm to derive cyanobacterial phycocyanin pigment concentrations in a eutrophic lake: Theoretical basis and practical considerations. *Remote Sensing of Environment*, 2014, **154**: 298-317.