

基于随机森林的内陆湖泊水体有色可溶性有机物 (CDOM) 浓度遥感估算*

吴志明¹, 李建超¹, 王睿¹, 时蕾¹, 苗松¹, 吕恒^{1,2**}, 李云梅^{1,2}

(1: 南京师范大学虚拟地理环境教育部重点实验室, 南京 210023)

(2: 江苏省地理信息资源开发与利用协同创新中心, 南京 210023)

摘要: 水体中的有色可溶性有机物 (CDOM) 是湖泊生态系统中氮、磷等有机营养物质的重要来源, 利用卫星遥感数据反演内陆水体中 CDOM 浓度一直是个挑战。因此本文基于滇池 2009 年 9 月、2017 年 4 月以及太湖 2016 年 7 月的现场原位观测和室内实验, 在分析水体固有光学特性的基础上, 引入机器学习算法, 建立了基于哨兵-3A OLCI 传感器的我国内陆湖泊水体 CDOM 浓度随机森林反演模型。利用独立的验证数据集对所构建的随机森林模型及常用的波段比值模型、一阶微分模型、半分析模型、BP 神经网络模型等的反演精度进行评价。结果表明: 随机森林模型的均方根误差为 0.14 m^{-1} , 平均相对误差为 21%, 与反演效果相对较好的 BP 神经网络模型相比, 均方根误差降低了 50%, 平均相对误差降低了 38%, 反演精度得到了显著的提高。根据随机森林算法的特征重要性参数提供的各自变量影响力结果, 发现 B11 (709 nm) 和 B6 (560 nm) 波段贡献率最大, 是反演 CDOM 的敏感波段。最后将随机森林模型应用到滇池 2017 年 4 月 12 日、太湖 2017 年 5 月 18 日的哨兵-3A OLCI 影像上, 得到滇池、太湖水体 CDOM 浓度分布图。滇池 CDOM 浓度的分布特征大致符合东北、西南高, 中西部低的趋势, 且河口处的 CDOM 浓度高于湖泊水体, 表明径流的输入给滇池水体带来了大量的 CDOM。太湖 CDOM 浓度的分布特征大致符合西部高, 湖心区和东部低的趋势。太湖西部以及北部梅梁湾受入湖河流影响较大, CDOM 浓度较高, 太湖开敞区远离河口处, 受外源河流的影响逐渐减小, 且由于湖水的不断稀释, CDOM 浓度不断降低。太湖东部水生植物很多, 湖水较为清澈, CDOM 浓度较低。

关键词: 滇池; 太湖; 有色可溶性有机物 (CDOM); 随机森林; 哨兵-3A OLCI

Estimation of CDOM concentration in inland lake based on random forest using Sentinel-3A OLCI

WU Zhiming¹, LI Jianchao¹, WANG Rui¹, SHI Lei¹, MIAO Song¹, LV Heng^{1,2**} & LI Yunmei^{1,2}

(1: *Key Laboratory of Virtual Geographic Environment, Ministry of Education, Nanjing Normal University, Nanjing 210023, P.R. China*)

(2: *Jiangsu Center for Collaborative Innovation in Geographical Information Resource Development and Application, Nanjing 210023, P.R. China*)

Abstract: The chromophoric dissolved organic matter (CDOM) in the water is an important source of organic nutrients such as nitrogen and phosphorus in the lake ecosystem. It is a challenge to use the satellite data to inverse the CDOM concentration in inland water. Based on the *in situ* observations in Lake Dianchi (September 2009 and April 2017) and Lake Taihu (July 2016), and experiments in the laboratory and on the basis of analyzing the inherent optical properties of water bodies, this paper introduces a machine learning algorithm, the random forest inversion model of CDOM concentration in inland lake water, was established based on Sentinel-3A OLCI sensor. The inversion accuracy of the random forest model and other commonly used model such as band ratio model, the first order differential model, the semi-analysis model and the BP neural network model are evaluated by using an inde-

* 国家重点研发计划 (2017YFB0503902)、国家自然科学基金项目 (41471282)、国家水体污染控制与治理科技重大专项 (2017ZX07302003) 和民用航天技术预先研究项目联合资助。2017-08-09 收稿; 2017-10-25 收修修改稿。吴志明 (1991 ~), 男, 硕士研究生; E-mail: wuzhimingkaka@163.com。

** 通信作者; E-mail: Heng.lyu@njjnu.edu.cn。

pendent verification data set. The results show that the mean square error of the random forest model is 0.14 m^{-1} , with a mean relative error 21%. Compared with the BP neural network model with relatively good inversion effect, the root mean square error is reduced by 50% and the mean relative error is reduced by 38%. The accuracy of inversion has been significantly improved. According to the characteristic parameters of the random forest algorithm, the influence of the respective variables is provided, it is found that the contribution rate of B11 (709 nm) and B6 (560 nm) is the largest, which is the sensitive band of CDOM. Finally, the random forest model was applied to the Sentinel-3A OLCI image in Lake Dianchi on April 12, 2017 and Lake Taihu on May 18, 2017, the distribution of CDOM concentration in the two lakes were obtained. The distribution of CDOM concentration in Lake Dianchi show that the northeast and the southwest region is higher than that in the central and western region. Meanwhile, the CDOM concentration in estuarine is higher than that in the lake water, which indicates that the input of runoff brings a lot of CDOM to Lake Dianchi. The distribution of CDOM concentration in Lake Taihu show a trend that higher in the west region while lower in the middle and east region. The western part of Lake Taihu and the Meiliang Bay in the northern part are greatly affected by the rivers inflow, resulted in a high CDOM concentration, although such influence is gradually reduced for the open area far from the estuary. The CDOM concentration in the eastern Lake Taihu is low due to the abundant aquatic plants and clear water.

Keywords: Lake Dianchi; Lake Taihu; chromophoric dissolved organic matter (CDOM); random forest; Sentinel-3A OLCI

有色可溶性有机物 (CDOM), 又称黄色物质, 主要由腐质酸和富里酸等物质组成, 普遍存在于海洋和湖泊水体, 是水体中最大的溶解性有机碳贮库^[1]. CDOM 的来源主要有两种类型: 内源型 CDOM 主要来源于土壤和水生植物的降解, 外源型 CDOM 主要来源于陆地地表^[2]. 常用 355、375 和 440 nm 等波长处的吸收系数表示其浓度^[3-4], 吸收系数越大, 对应的 CDOM 浓度就越高.

CDOM 能够强烈吸收短波辐射, 限制中波紫外线 (UV-B, 波长为 280~315 nm) 辐射在水中的传播, 从而保护水体的生态系统^[5]. CDOM 还有固碳作用, 水体中 CDOM 浓度的增加会导致地表以及大气层中的碳含量减少^[6]. 此外, 水体中的 CDOM 也是湖泊生态系统中氮、磷等有机营养物质的重要来源^[7]. 因此, 监测水体中 CDOM 浓度变化状况是认识水体中碳循环的基础, 也是了解湖泊生态系统中营养物质循环的重要途径^[8]. 正因为 CDOM 具有诸多重要作用, 所以对其进行的遥感反演研究具有重要意义.

二类水体中光谱的主要影响因素有 CDOM、悬浮物和浮游植物, 通常称为水色三要素. 叶绿素和悬浮物目前已经有大量的反演模型, 而 CDOM 的反演研究大多应用于海洋水色遥感中^[9-11]. 内陆浅水湖泊通常是浮游植物及悬浮物吸收占主导, CDOM 受到其他的光学活性物质的影响较大, 光谱信号微弱, 因此反演精度较低. 由于 CDOM 的成分比较复杂, 很难直接从水体中提取出来, 所以目前大部分研究都从其荧光和吸收特性方面展开^[12].

遥感技术可以实现迅速、大范围地获取水体中 CDOM 浓度. 目前很多遥感反演 CDOM 的方法建立在 CDOM 浓度与遥感数据之间的线性相关分析上. 531 nm/551 nm^[13]、670 nm/443 nm^[14] 和 490 nm/590 nm^[15] 常作为 CDOM 反演的最佳波段组合, 用来反演不同水体的 CDOM 浓度, 模型的平均相对误差均低于 30%, 反演结果令人满意. 许多学者也采用了一阶微分方法反演内陆湖泊水体水质参数浓度^[16-17], 如冯龙庆^[16] 利用一阶微分方法建立了太湖水体 CDOM 浓度估算模型, 平均相对误差为 15.2%, 反演精度较高. 经验模型算法简单, 便于操作和测试, 但是经验模型缺乏理论依据, 统计关系常常不稳定, 有一定的局限性, 重复利用性差, 难于对比和推广^[18-19]. 且上述模型各个遥感因子间存在多重共线性问题, 而非参数模型能克服这种缺陷^[20]. 近年来, 陆续有学者将机器学习算法应用于遥感估算研究^[21-24], 如孙德勇等^[21] 利用神经网络模型估算太湖水体 CDOM 浓度, 平均相对误差远远低于 30%, 对反演湖泊水体 CDOM 浓度具有一定的应用价值. 神经网络模型虽然能处理复杂的非线性问题, 但是也存在收敛速度慢, 稳健性不足和过拟合等问题. 且神经网络模型的普适性一直是一个没有解决的难题, 能否用于其他湖泊还有待验证.

此外, 利用光学遥感数据反演水体要素时, 完全由解析模型直接反演得到的只有部分水体固有光学量, 大部分水体参量的反演模型一般都会加入经验关系, 在解析方法的推导过程中引入一些经验公式, 得到的反演模型就称为半解析模型. 在半解析模型中, 各个参数都有明确的物理意义, 水体要素反演结果更为可靠^[25]. 但是这种算法需要对水体光学组分进行分离, 并对各自的固有光学性质进行准确测量, 难度较大, 而且内陆水体成分复杂, 多种水色组分共同影响水体光谱特性, 并影响反演精度.

因此本文引入了随机森林算法. 随机森林指的是利用多棵树对样本进行训练并预测的一种分类器,该分类器最早由 Leo Breiman 等提出^[26]. 随机森林的优势是不容易产生过拟合,并且抗噪声能力比较强,对数据集的适应能力也比较强,被广泛应用于各个领域^[27-28]. 但随机森林算法是否能用于内陆光学特征复杂水体水质参数遥感反演还需要进一步研究.

哨兵-3A OCLI 传感器搭载的海陆色度仪(Ocean and Land Colour Instrument)传感器以 300 m 的瞬时视场对地球的陆地和海洋进行观测,利用 21 个谱段提供高信噪比数据,从而使 MERIS 15 个谱段所产生的数据产品具有延续性,且增加的 6 个谱段将为人们实现改进型水色要素反演提供有效途径^[29]. 由此可见哨兵-3A OCLI 传感器以其中等尺度的空间分辨率、高时间分辨率是内陆湖泊水体水质遥感监测的不可或缺的数据源,但是有关利用哨兵-3A OCLI 传感器监测内陆湖泊水环境的报道还不多.

因此本研究以滇池和太湖为实验区,在研究采样水体的反射光谱数据和固有光学量数据的基础上,构建并检验了针对滇池和太湖水体 CDOM 浓度的随机森林模型,并比较了随机森林模型与波段比值模型、半分析模型等其他常用模型的反演效果. 最终将随机森林模型应用到滇池和太湖的哨兵-3A OLCI 影像上,得到滇池和太湖水体 CDOM 浓度分布图,并对其空间分布各自进行了分析和探讨.

1 研究方法

1.1 研究区

本文以滇池和太湖为研究区,这 2 个湖泊在夏季有大量的蓝藻出现,都是我国重点监测和治理的富营养化湖泊. 2 个湖泊由于其地理位置、形成原因、气候条件、周边入湖河流和经济发展等因素的不同,水体的光学特性也表现出较大的差异性.

滇池(24.40°~25.02°N, 102.36°~102.47°E)是高原湖泊,我国第 6 大淡水湖,位于云南省昆明市南端. 滇池分内海和外海两部分,北部为内海,不属于本研究区范围;南部为外海. 当水面海拔高度为 1887.4 m 时,流域面积为 2920 km²,湖面面积为 309 km²,最大水深为 11.2 m,平均水深为 5.3 m. 滇池平均水温为 19℃,季节性变化不大,这对于藻类的生长和繁殖有利. 且由于滇池处在昆明市下游,城市的污水都排入滇池,造成滇池的水质日益恶化.

太湖(30.9°~31.6°N, 119.9°~120.6°E)位于江苏省和浙江省的交界处,长江三角洲的南部,其流域面积为 36900 km²,平均水深为 1.9 m,最大水深一般只有 3 m,水位较为稳定,是中国五大淡水湖之一,湖泊面积为 2338.1 km². 经济的快速发展伴随着工农业生产和人类活动排放的污染物增多,入湖污染负荷超重,湖泊的富营养化程度加重,蓝藻水华的频繁暴发严重影响了周围居民的生活饮水以及太湖流域经济的发展. 此外,由于东太湖是水草区,水质相对较好,因此不属于本研究区范围.

1.2 地面实验数据

本研究采用的数据来自滇池和太湖的 3 次野外实验,分别是 2016 年 7 月 22—23 日在太湖,2009 年 9 月 19—20 日以及 2017 年 4 月 13—16 日在滇池进行. 采集了各样点的光谱数据、位置坐标、光谱数据测量时的风速风向等辅助数据以及湖体表层水样,水样放入冷藏箱带回实验室,测定相关参数. 主要测量的水质参数有叶绿素 a 浓度(Chl.a)、悬浮物浓度(SS)、藻蓝蛋白浓度(PC)、CDOM 吸收系数($a_{\text{CDOM}}(\lambda)$)、总磷(TP)和总氮(TN)浓度等等.

在剔除了部分由于仪器操作和数据记录不全等人为因素的异常点之后,选取了数据质量较好采样点的表观光学量、固有光学量和水质参数浓度等数据用于研究分析. 本研究共有 100 个样本数据,其中,2016 年 7 月 22—23 日在太湖的采样数据有 53 个,2009 年 9 月 19—20 日在滇池的采样数据有 17 个,2017 年 4 月 13—16 日在滇池的采样数据有 30 个. 随机选取了 70 个样点作为训练样本,30 个样点用于精度评价. 图 1 为太湖、滇池水体实测采样点及研究区位置示意图.

1.3 光谱测定

光谱数据测量使用的仪器是美国 ASD 公司生产的 ASD FieldSpecPro 便携式光谱辐射计,其波段范围为 350~1050 nm. 光谱测量采用水表面以上测量法^[30]. 观测天顶角和方位角分别为 40°和 135°,优点是可以更好地避免太阳直射反射,且与剖面观测的固有差异较小. 每个对象采集 10 条光谱数据,剔除异常光谱,其余

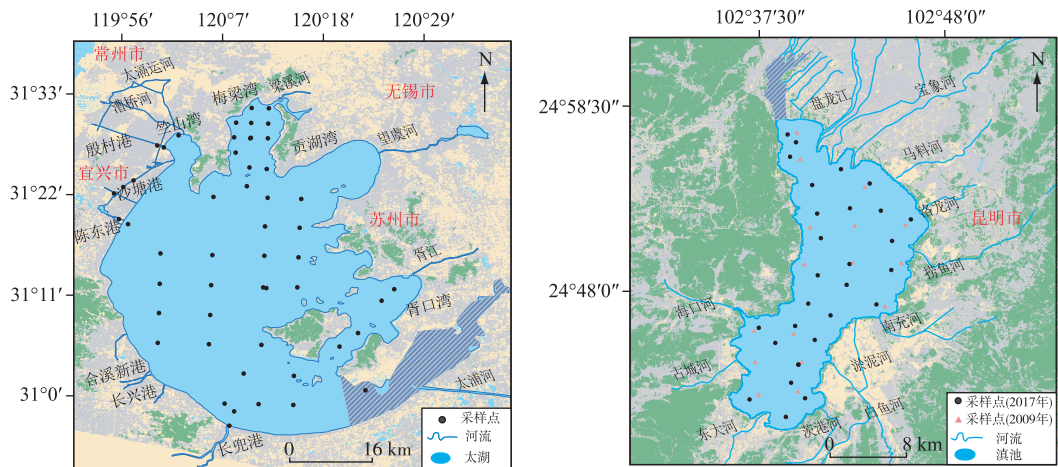


图 1 太湖、滇池水体实测采样点分布及研究区位置

Fig.1 Spatial distribution of sampling sites and location of the study area in Lake Taihu and Lake Dianchi

光谱数据作均值处理.

具体计算步骤如下:

1) 在避开太阳直射反射、忽略或避开水面泡沫的情况下,光谱仪测量的水体数据为:

$$L_{sw} = L_w + r \cdot L_{sky} \quad (1)$$

式中, L_w 为离水辐亮度; L_{sky} 为天空漫散射光, 不带有任何水体信息, 必须去掉; r 为气-水界面对天空光的反射率, 根据唐军武等^[30]的经验, 在上述观测几何条件下, 平静水面可取 $r = 0.022$, 在 5 m/s 风速的情况下, r 可取 0.025 , 由于实际测量时风速不超过 5 m/s , 在本研究中 r 取值为 0.022 .

2) 水表面入射总辐照度 $E_d(0^+)$ 可由测量标准板的反射 L_p 计算而得, 即:

$$L_p = \rho_p \cdot E_d(0^+) / \pi \quad (2)$$

式中, ρ_p 为标准板的反射率, 通常采用 $10\% \leq \rho_p \leq 30\%$ 的标准板.

3) 遥感反射率 (R_{rs}) 定义为:

$$R_{rs} = L_w / E_d(0^+) \quad (3)$$

1.4 CDOM 光谱吸收测定

先利用 Whatman GF/F 滤膜过滤一定体积的水样, 留取滤液. 再使用 $0.22 \mu\text{m}$ 的 millipore 滤膜过滤已经过滤过的水样, 即可得 CDOM 水样, 将适量 CDOM 水样倒入 1:4 的比色皿中, 利用岛津 UV2550 分光光度计进行测量, 所测得的值为 CDOM 吸光度. 然后可通过公式 (4) 计算各波长的吸收系数^[31]并通过公式 (5) 作散射校正:

$$a_{CDOM}(\lambda)' = 2.303 D(\lambda) / r \quad (4)$$

$$a_{CDOM}(\lambda) = a_{CDOM}(\lambda)' - a_{CDOM}(700)' \cdot \lambda / 700 \quad (5)$$

式中, $D(\lambda)$ 为吸光度; r 为光程路径; $a_{CDOM}(\lambda)'$ 为波长 λ 的未校正吸收系数; $a_{CDOM}(\lambda)$ 为波长 λ 的吸收系数 (m^{-1}). 本文利用 440 nm 处的 CDOM 吸收系数表示其浓度, 野外实验 100 个采样点的 CDOM 吸收系数曲线如图 2 所示. 太湖水体 $a_{CDOM}(440)$ 的变化范围为 $0.02 \sim 2.02 \text{ m}^{-1}$, 均值为 $0.49 \pm 0.31 \text{ m}^{-1}$; 滇池水体 $a_{CDOM}(440)$ 的变化范围为 $0.30 \sim 1.81 \text{ m}^{-1}$, 均值为 $0.68 \pm 0.45 \text{ m}^{-1}$.

1.5 哨兵-3A OLCI 影像数据获取及预处理

哨兵-3A OLCI 传感器是欧空局第三颗环境监测卫星, 于 2016 年 2 月 16 号发射并成功到达预定轨道 ($802 \text{ km} \times 807 \text{ km} \times 98.6^\circ$ 轨道). 哨兵-3A 卫星上将搭载海陆色度仪 (OLCI) 和海陆表面温度辐射计 (SLSTR) 两台光学仪器. 这两台仪器都将延续 ERS 卫星和 Envisat 卫星等的有效业务观测功能, 为了支持协同产品的

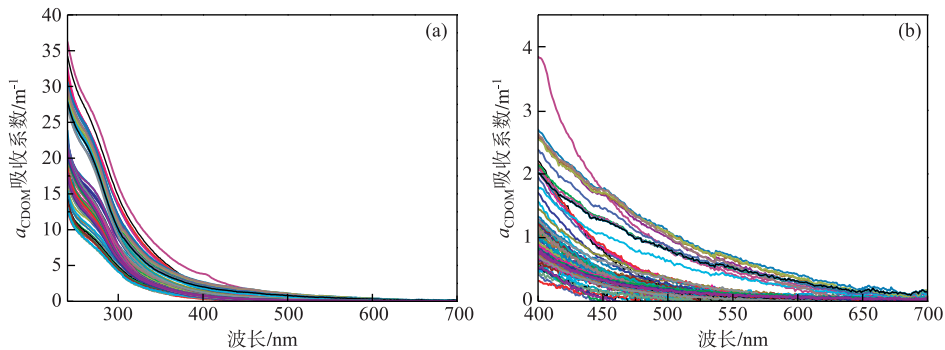


图2 100个野外采样点的CDOM吸收系数曲线((a):240~700 nm,(b):400~700 nm;
包括2016年7月太湖53个采样点,2017年4月滇池30个采样点,2009年9月滇池17个采样点)

Fig.2 The curves of CDOM absorption coefficient for 100 field sampling points((a): is 240~700 nm,
(b): is 400~700 nm; including 53 sampling points in Lake Taihu on Jul. 2016 and 30 sampling points
in Lake Dianchi on Apr., 2017 and 17 sampling points in Lake Dianchi on Sep., 2009)

研制工作,这两台光学仪器将被设定成对地球进行准同步观测.它们所取的轨道和扫描宽度将可以使卫星在任何海洋和陆地位置上的重访时间分别小于3.8 d(考虑太阳耀斑,但不考虑云)和1.4 d(太阳耀斑和云均不考虑).

在陆地上,哨兵-3A OLCI 传感器可应用于监测火灾、土地利用变化、检测植被健康以及测量河流和湖泊的高度等;在海洋上,可以测量海水温度、海色、海平面高度和海冰厚度等信息,监测海洋污染和气候变化等.哨兵-3A OLCI 传感器波段设置见哨兵-3A 用户手册(https://earth.esa.int/documents/247904/685236/Sentinel-3_User_Handbook).

对获取的2017年4月12日滇池和2017年5月18日太湖的哨兵-3A OLCI 影像,使用SeaDAS 7.3 软件进行辐射定标等预处理.

1.6 精度评价方法

本研究采用SPSS 软件进行相关统计分析,引入 Pearson 相关系数(R)、均方根误差($RMSE$)和平均相对误差(MRE)3个指标作为统计量来检验模拟值和实测值是否一致.其中, R 用来衡量两个数据集是否在同一条线上,用来衡量定距变量间的线性关系.相关系数的绝对值越接近于1,相关度越强. $RMSE$ 和 MRE 的计算公式分别为:

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (X_{Est,i} - X_{Obs,i})^2}{n}} \quad (6)$$

$$MRE = \frac{1}{n} \sum_{i=1}^n \left| \frac{X_{Est,i} - X_{Obs,i}}{X_{Obs,i}} \right| \cdot 100\% \quad (7)$$

式中, $X_{Est,i}$ 和 $X_{Obs,i}$ 分别为第*i*波长处的模拟值和实测值, n 为400~900 nm 波长总数.

2 CDOM 反演方法的构建

2.1 随机森林算法

随机森林顾名思义,是用随机的方式建立一个森林,森林里面有很多的决策树组成,随机森林的每一棵决策树之间是没有关联的.在得到森林之后,当有一个新的输入样本进入时,就让森林中的每一棵决策树分别进行一下判断,给出一个预测值,最后对所有的决策树进行求平均得到最终预测值.大量的理论和实践表明,随机森林算法具有很高的预测准确性,而且对异常值和噪声都具备很好的容忍度.由于随机森林选取样本的方法是有放回地随机抽样,因此,建立回归树时有一部分样本不会被选中,因而其作为检验样本出现,

这样就起到了样本内部交叉验证的作用,可以减少过拟合现象的出现^[32].

随机森林算法步骤如下:

- 1) 运用自助法 (Bootstrap) 从 N 个原始样本中有放回地抽取 n 个样本构建 n 棵决策树, 未被抽到的样本为袋外数据集.
- 2) 如果有 P 个输入变量, 每棵树的节点都将随机选择 k ($k < P$) 个特定的变量, 然后运用这 k 个变量来确定最佳的分裂点.
- 3) 每棵决策树都最大可能地进行生长而不进行剪枝, 直到满足分割终止条件.
- 4) 通过对所有的决策树进行求平均来预测新的数据.

由于随机森林算法在回归时是使用多个决策树的预测结果求平均值作为最终模拟值, 因此不会产生一个具体的回归方程. 正是利用其选择自变量和样本时的随机性, 在寻找自变量与因变量关系的同时, 也注重了不同样本和自变量的差异, 使得回归的结果既可以考虑到每个样本和自变量的影响, 又不会过分趋近于个别样.

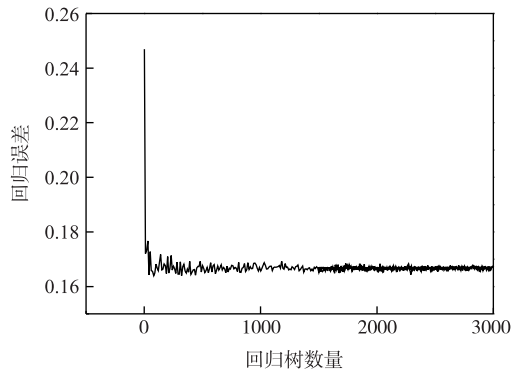


图3 随机森林模型的回归树数量
与回归误差关系图

Fig.3 The relationship between regression error
and regression tree number
in random forest model

本研究采用 python 自带的 scikit-learn 机器学习库实现随机森林算法. 其中, 70 个样点作为训练样本用于参数寻优和模型的构建, 30 个样点作为验证样本用于模型检验. 首先将实测遥感反射率通过光谱响应函数模拟到哨兵-3A OCLI 传感器的 B3 (442 nm) ~ B11 (709 nm) 波段, 以这 9 个波段的遥感反射率作为输入自变量. 其次, 确定回归过程中涉及的几个重要参数: 首先, 调试决策树的数量 ($n_estimators$), 也就是自助法 (Bootstrap) 的重抽样次数. $n_estimators$ 参数的确定主要是设置不同决策树的数量, 从而得到不同模型的误差, 然后观察误差的变化趋势. 当 $n_estimators$ 达到一定的数量之后, 随着 $n_estimators$ 数量的增加, 模型的误差也趋于稳定, 我们就选取该值作为最终选取的 $n_estimators$ 值. 由图 3 可知, 回归树数量达到 1000 后回归误差趋于稳定, 因此在不影响计算效率的情况下, 我们以 1000 作为回归树数量. 其次调试最大特征数量 ($max_features$)、最大深度 (max_depth) 和分

裂性能函数 (criterion) 等几个关键参数. $max_features$ 是考虑内部节点分支的时候, 属性的选择个数, 这里设置为 '自动'. criterion 是用来测试分裂性能的函数, 设置为 '均方误差'. max_depth 默认不输入, 这样就能满足训练过程中每棵决策树都最大可能地进行生长而不进行剪枝, 直到满足分割终止条件.

2.2 其他常规算法的参数率定

2.2.1 波段比值方法 CDOM 反演的经验模型目前有很多, 其中以波段比值法最为常见^[13-15]. 531 nm/551 nm^[13]、670 nm/443 nm^[14] 和 490 nm/590 nm^[15] 常作为 CDOM 反演的最佳波段组合, 用来反演不同水体的 CDOM 浓度, 模型的平均相对误差均低于 30%, 反演结果令人满意. 本研究对 B4、B5、B7、B8、B9 和 B10 等波段进行组合研究, 利用不同的波段组合与 $a_{CDOM}(440)$ 进行回归分析, 最终发现 B10/B9 在取以十为底的对数值时反演 CDOM 的效果最好.

2.2.2 一阶微分方法 许多学者也采用了一阶微分方法反演内陆湖泊水体水质参数^[16-17]. 本研究首先对原始遥感数据进行一阶微分处理, 然后分别求各个波段遥感反射率微分值与 $a_{CDOM}(440)$ 的相关系数. $a_{CDOM}(440)$ 与各波段遥感反射率微分值的相关系数呈正负交叉的趋势, 其中在 666 nm 处呈最大负相关 ($R = -0.52763$), 因此本研究以 666 nm 处的遥感反射率微分值为自变量, 以 $a_{CDOM}(440)$ 为因变量建立模型 (图 4).

2.2.3 神经网络算法 本研究建立了 3 层神经网络模型, 首先将实测遥感反射率通过光谱响应函数模拟到哨

兵-3A OCLI 传感器的 B3(442 nm)~B11(709 nm)波段,以这 9 个波段的遥感反射率作为输入的 9 个神经元,以 $a_{\text{CDOM}}(440)$ 作为输出层唯一的神经元. 分别设置了 4~14 个隐含层节点,并对各节点数的神经网络模型进行训练,用验证样点进行验证,并利用模型实测值与模拟值的 R 、 $RMSE$ 和 MRE ,对含不同隐含层节点数的神经网络模型的测试结果进行比较,从而选出最佳节点数^[33]. 最终发现节点数为 8 时效果最好. 具体 BP 算法参数如表 1 所示.

2.2.4 QAA(V5)-CDOM 算法 QAA-CDOM 算法是 Zhu 等^[34]在 2013 年提出来的,Ogashawara 等^[35]针对内陆水体对 QAA-CDOM 算法进行了重新参数化. 该算法以 $R_{\text{rs}}(\lambda)$ 作为输入,计算过程包括 8 个步骤,其中前 6 步取自 QAA-V5 算法,得到水体总吸收 $a(440)$ 和悬浮物后向散射 $b_{\text{bp}}(560)$;后两步借鉴了 QAA-E 算法中悬浮物吸收系数 a_{p} 与悬浮物后向散射系数 b_{bp} 的统计关系,并对其系数进行了调整,计算得到了 $a_{\text{CDOM}}(440)$. 结合哨兵-3A 卫星的波段设置,本研究采用 560 nm 作为参考波长,并对相关波段进行了改进,如 640 nm 替换为 665 nm 等.

3 结果与讨论

3.1 CDOM 光学特性以及敏感波段分析

CDOM 是水体重要的光衰减因子,其浓度和光学特性显著地改变了水体水色和水下光场强度的分布^[36]. CDOM 的光谱特征:在紫外和蓝光范围具有强烈的吸收特性,在黄色波段吸收最小,使其呈黄色,故常称这类复杂的混合物为“黄色物质”. 此外,在近紫外、可见光范围(250~700 nm),CDOM 吸收系数随波长基本上呈指数衰减^[37]. CDOM 是水体固有光学特性的重要组成部分,是影响水色要素反演的重要因素. 水体总吸收等于水体中各组分吸收的线性叠加^[38].

随机森林算法中的特征重要性参数可以给出各自变量影响力大小. B11(709 nm)和 B6(560 nm)波段占的比重最大(图 5). 张运林等^[39-40]对太湖夏季 CDOM 的吸收特性研究表明,夏季水华暴发时 CDOM 吸收系数与叶绿素 a 浓度的空间分布较为一致,CDOM 吸收系数与叶绿素 a 浓度存在正相关,浮游植物新陈代谢可能是水体中 CDOM 的重要来源. 因此本研究分析了 CDOM 与各水质参数之间的相关性,其中,CDOM 吸收系数与叶绿素 a 浓度存在显著性正线性相关,相关系数达到 0.46,且 CDOM 与悬浮物等其他水质参数之间的相关性并不明显,表明在夏季水体中浮游藻类浓度较高,CDOM 除了来自河流外,浮游藻类死亡降解的产物也是 CDOM 的重要来源. 同时还分析了叶绿素 a 浓度与 400~900 nm 波段遥感反射率的相关关系,叶绿素 a 浓度与 709 nm 附近遥感反射率相关性最好(图 6). 因此本研究发现 709 nm 是 CDOM 反演的重要敏感波段,可能是因为 CDOM 主要来源于浮游植物的新陈代谢.

而在 560 nm 附近,虽然 CDOM 和悬浮颗粒物的吸收都小于蓝紫光波段,但悬浮颗粒物的吸收减小得更快,而 CDOM 吸收光谱相对平缓,使得此处 CDOM 占较高吸收比例,CDOM 光谱信号较强,与遥感反射率相关性较大. 因此,560 nm 附近也是反演 CDOM 的敏感波段.

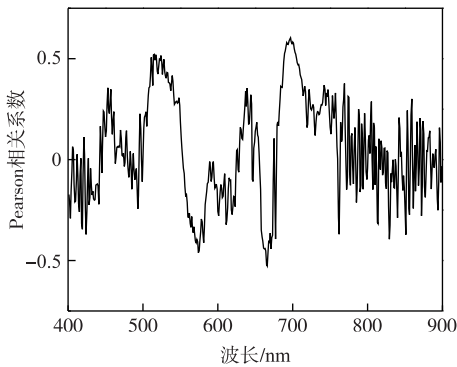


图 4 各波段遥感反射率一阶微分值与 440 nm 处 CDOM 吸收系数的 Pearson 相关系数
Fig.4 Pearson correlation coefficient of first order derivative value of remote sensing reflectance at each band and the CDOM absorption coefficient at 440 nm

表 1 BP 算法参数设置
Tab.1 Settings of the BP algorithm

参数	取值
输入层到隐层激励函数	双曲正切 S 型函数
隐层到输出层激励函数	线性输出函数
学习函数	梯度下降权重函数
训练函数	Levenberg-Marquardt 算法
隐层结点数	8
迭代次数	10000
学习目标	1e-7
学习率	0.2

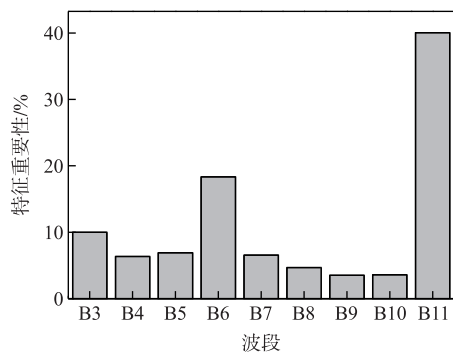


图5 随机森林算法的特征重要性参数给出的各波段重要性大小
Fig.5 The importance of each band is given by the feature importance parameter of the random forest algorithm

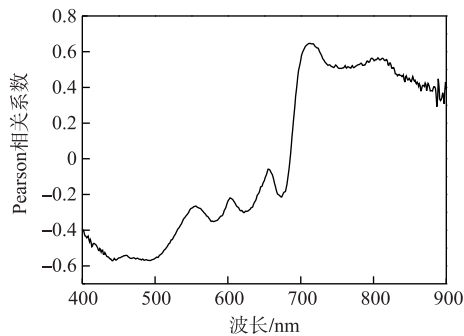


图6 100个采样点叶绿素a浓度与400~900 nm波段遥感反射率的相关关系
Fig.6 Correlation between Chl.a concentration and remote sensing reflectance at 400~900 nm for 100 field sampling points

表2 随机森林模型以及波段比值等其他常用模型反演精度评价

Tab.2 The accuracy verification results of the Random Forest model and other commonly used model such as the Band Ratio model

反演模型	RMSE/m ⁻¹	MRE/%
随机森林模型	0.14	21
波段比值模型	0.63	69
一阶微分模型	0.23	36
神经网络模型	0.24	40
半分析模型	0.49	68

3.2 反演精度评价

利用验证数据集对本文构建的随机森林模型,以及常用的波段比值模型、一阶微分模型、神经网络模型和半分析模型进行精度评价. 表2是利用验证数据对所建立的模型进行精度验证的结果. 图7是各模型反演值与实测值的散点图. 从RMSE和MRE来看,波段比值模型反演精度较差,散点图显示所有样点都分布在1:1线以下;一阶微分模型和神经网络模型的大部分样点的实测值和预估值都能够较好地吻合,较均匀地分布在1:1线附近,但是神经网络模型有4个点明显偏离1:1线,使模型精度较差,可见BP神经网络模型虽然能处理复杂的非线性问题,但是也存在收敛速度慢、稳健性不足和过拟合等问题;

半分析模型反演值与实测值相比偏低,所有样点都分布在1:1线以下.

随机森林回归模型的样点基本分布在1:1线附近,没有偏离极度异常的点,且当CDOM浓度小于约0.6 m⁻¹时产生高估现象,CDOM浓度大于约0.8 m⁻¹时产生低估现象(图7). 但从RMSE和MRE的提高水平来看,随机森林回归模型的优越性还是很明显的.

随机森林的优势是不容易产生过拟合,并且抗噪声能力比较强,对数据集的适应能力也比较强:既能处理连续型数据,也能处理离散型数据,而且数据集无需规范化,运行速度很快,并且在构建随机森林的过程中可以生成一个泛化误差的内部无偏估计,现有的算法中随机森林的精确度是无可比拟的.

随机森林算法能够取得比神经网络算法更好的反演结果,得益于随机森林算法中2个随机性的引入,使得随机森林模型的抗噪声能力比较强,也不容易陷入过拟合. 而BP神经网络在训练过程中可能学习了过多的样本细节,学习出的模型已不能反映样本内含的规律. 综上所述,随机森林回归算法可作为滇池和太湖水体CDOM浓度遥感估算建模的首选方法.

4 太湖、滇池 CDOM 空间格局分析

4.1 大气校正精度评价

对于一类水体,标准大气校正算法通过基于近红外波段离水反射率为0的假设来实现,取得了较好的效果. 但是对于内陆混浊水体,由于水体悬浮物以及藻类的存在,使得近红外波段往往具有较强的后向散射,从而导致标准大气校正算法高估了大气辐射信号,无法获取准确的离水反射信号. 近年来,国内外学者构建了

许多针对于内陆水体的大气校正算法试图获取更为准确的离水反射率信号. 在本研究中大气校正采用的是 SeaDAS 7.3 的 MUMM 算法, $R_{rs}(779)$ 与 $R_{rs}(865)$ 的比值是一个关键参数, 设定为 1.9.

滇池影像的拍摄时间是 2017 年 4 月 12 日, 实际采样是从 2017 年 4 月 13 日开始, 卫星过境时间与地面监测同步时差相超过了 24 h. 对比 4 月 12 日和 13 日的气象条件可以发现, 这两天的天气都极为晴朗, 大气状况极为接近, 因此 4 月 12 日获取的滇池哨兵-3A OLCI 影像以及 4 月 13 日采集的 5 个受云、太阳耀斑影响较小的地面匹配点可以用来验证大气校正效果.

利用上述大气校正方法对哨兵-3A OLCI 影像进行大气校正, 得到哨兵-3A OLCI 遥感反射率影像. 基于本文的研究内容, 我们得到 B3(442 nm)~B11(709 nm) 波段处大气校正的结果, 图 9 为 5 个同步点大气校正后得到的哨兵-3A OLCI 影像不同波段遥感反射率与实测遥感反射率的对比结果.

经过大气校正后的哨兵-3A OLCI 影像的遥感反射率相比于实测遥感反射率都存在着一定程度系统性的低估现象(图 8). 绿光、红光波段与近红外波段大气校正精度较高(表 3). 除 442 nm 外, 其他绿光、红光波段与近红外波段大气校正的平均相对误差都小于 25%. 其中, 665 nm 大气校正精度最好, 其平均相对误差仅为 16.714%. 影像的大气校正结果较为理想.

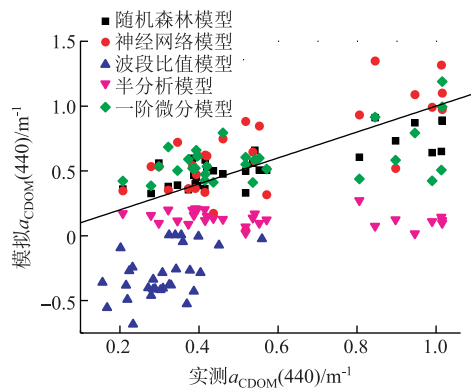


图 7 随机森林模型与波段比值等其他常用模型实测值与模拟值关系的散点图

Fig.7 The relationship between measured and simulated values in Random Forest model and other commonly used models such as the Band Ratio model

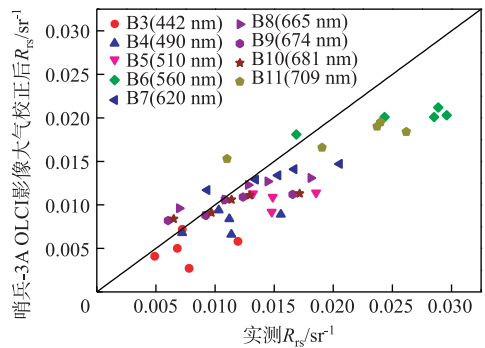


图 8 5 个星地匹配样点哨兵-3A OLCI 影像大气校正后的 R_{rs} 和实测 R_{rs} 关系的散点图

Fig.8 The relationship between the R_{rs} of the 5 satellite-ground matchups via atmosphere correction from the Sentinel-3A OLCI data and the measured R_{rs}

表 3 哨兵-3A OLCI 影像各波段大气校正的 RMSE 和 MRE

Tab.3 Root mean square error and mean relative error of the atmospheric correction for the bands from the Sentinel-3A image

	B3 (442 nm)	B4 (490 nm)	B5 (510 nm)	B6 (560 nm)	B7 (620 nm)	B8 (665 nm)	B9 (674 nm)	B10 (681 nm)	B11 (709 nm)
RMSE/ sr^{-1}	0.0037	0.0039	0.0045	0.0069	0.0032	0.0027	0.0027	0.0029	0.0050
MRE/%	31.940	24.907	24.415	22.435	17.166	16.714	17.327	18.119	24.007

4.2 空间格局分析

本研究将随机森林模型应用到 2017 年 4 月 12 日滇池、2017 年 5 月 18 日太湖的哨兵-3A OLCI 影像上, 得到滇池、太湖 CDOM 浓度分布图(图 9). 滇池 CDOM 浓度的空间格局与程庆霖等^[41-42]的研究保持一致, 太湖 CDOM 浓度的空间格局与冯龙庆等^[16,43-44]的研究保持一致.

利用 5 个星地匹配点对影像反演精度进行评价, 结果表明: 5 个准同步匹配点的 CDOM 反演值与实测值相比, 存在一定程度的高估现象, 均方根误差为 0.13 m^{-1} , 平均相对误差为 14%. 大气校正利用高光谱影像准确

地进行 CDOM 反演的前提条件,但是由于内陆水体的影响因素过多,大气状况比较复杂,大气校正精度有待于提高,这在一定程度上限制了影像的应用,也是影响本研究利用哨兵-3A OLCI 影像反演 CDOM 精度的一个不可忽视的原因。

2017 年 4 月 12 日滇池 CDOM 浓度分布特征大致符合东北、西南高,中西部低的趋势,且位于河口处的 CDOM 浓度都高于湖泊水体中的浓度,表明径流的输入给滇池水体带来了大量的 CDOM(图 9a)。滇池环湖有 28 条主要入湖河流,主要分布在内海以及外海的东北部,污染物主要来自主城区的城镇生活污水以及重点工业企业的废水排放^[45],河口区的地表径流和河流输入会带来大量外源性溶解性有机质,滇池外海的东北部 CDOM 浓度较高,而北部靠近内海地区 CDOM 浓度较低(图 9a),这可能是由于受到了风浪影响,滇池外海区域风浪较大,风浪可能会导致 CDOM 浓度降低。此外,外海东部及南部入湖河流相对较少,污染物主要来源于农业面源、城市生活污染源及少量的工业企业废水排放,大部分河流水质相对较好,从 CDOM 浓度分布图上也可以看出,滇池外海的东南部 CDOM 浓度大多要比东北部低。不过,在滇池外海的西南部以及东南部分地区,也就是古城河、茨巷河以及白鱼河的入湖口附近有一片高值区存在,这可能是由于受到河流污染的影响,茨巷河污染最严重,径流输入会带来大量的可溶性有机质。滇池外海的中西部大片区域,很少受到入湖河流的影响,受风浪影响最大。在风浪的作用下,近岸和远岸的海水混合比较厉害,使其高浓度区域变小,中等区域增大,整体趋于均质和稳定,在 CDOM 浓度分布图上也可以明显看出这一趋势。

2017 年 5 月 18 日太湖 CDOM 浓度的分布特征大致符合西部高,湖心区和东部低的趋势,且位于河口处的 CDOM 浓度都高于湖泊水体中的浓度(图 9b)。冯龙庆^[16]认为太湖春季湖泊内浮游植物还没有大量生长,浮游植物腐烂产生的有机物并非太湖 CDOM 的主要来源,水体中的 CDOM 应主要来自于河流输入的陆源有机物。太湖西部入湖河流河口处 CDOM 浓度明显大于开敞区,这是由于入湖河流周边分布着大量的村庄、乡镇,又受到工、农业较发达的无锡、常州市的影响,居民生活用水和工业废水排入河流,导致其 CDOM 浓度较高,而北部梅梁湾紧靠无锡市区,也是太湖湖水的主要入口之一,悬浮物质含量高,水质处于富营养化水平,CDOM 浓度较其他区域也偏高,这反映了外源是太湖 CDOM 的重要来源。太湖开敞区远离河口处,受外源河流的影响逐渐减小,且由于湖水的不断稀释,CDOM 浓度不断降低。太湖西南沿岸长兜港附近,虽然也有入湖河流的影响,但该处入湖河流污染物较少,在春夏之交,盛行东南风,污染物能够得到很快的稀释,因此其 CDOM 浓度并没有比湖心区和太湖东部的 CDOM 浓度有明显的偏大。而太湖东部是太湖湖水的主要出水通道,水生植物很多,水体比较清澈,CDOM 浓度较低。

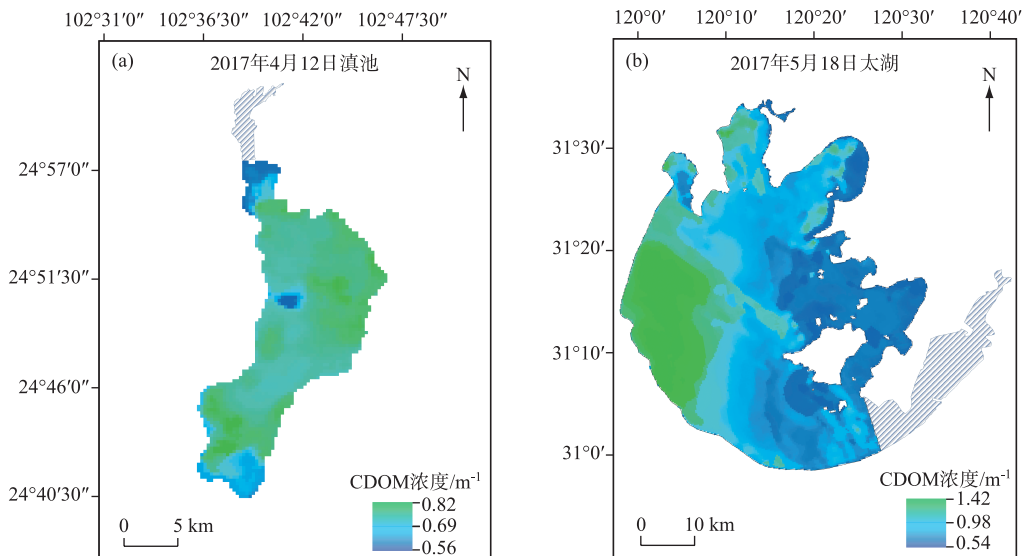


图9 滇池(a)和太湖(b)CDOM 浓度分布

Fig.9 Distribution of CDOM concentration in Lake Dianchi(a) and Lake Taihu(b)

5 结论

1)利用太湖2016年7月、滇池2009年9月、2017年4月的实测光谱数据以及水质参数数据,建立了滇池、太湖水体CDOM反演随机森林模型,并对随机森林模型和其他常用模型的适用性以及反演精度进行对比分析,最终发现随机森林模型效果最好,并将该模型应用到了2017年4月12日滇池以及2017年5月18日太湖的哨兵-3A OLCI影像上,得到滇池、太湖水体CDOM浓度分布图。

2)以往的研究中多采用经验波段比值方法来反演CDOM浓度,本研究采用多波段参数与随机森林回归算法组合来进行CDOM浓度反演,显著提高了反演精度。随机森林算法利用其选择样本和自变量时的随机性,在寻找自变量与因变量间关系的同时,也注重了不同样本和自变量的差异,使回归结果既可以考虑到每一个样本和自变量的影响,又不会过分趋近于个别样本,这对于进行CDOM浓度这种变化复杂、总体趋势不明显的变量的反演具有较好的效果。

3)随机森林模型适合计算大规模数据,随着选取解释变量个数增加,树与树之间的相关性会增强,同时泛化误差和OOB(Out-of-bag)误差也会增大。在未来的研究中应该尝试进一步增加样本量,采用装袋法(Bagging)构建随机森林,随机森林模型的预测精度可能会更高。

致谢:感谢参加滇池、太湖实验的所有工作人员在数据获取过程中所做的贡献!

6 参考文献

- [1] Green SA, Blough NV. Optical absorption and fluorescence properties of chromophoric dissolved organic matter in natural waters. *Limnology and Oceanography*, 1994, **39**(8): 1903-1916.
- [2] Kirk JTO ed. Light and photosynthesis in aquatic ecosystems. Cambridge: Cambridge University Press, 1983.
- [3] Stedmon CA, Markager S, Kaas H. Optical properties and signatures of chromophoric dissolved organic matter(CDOM) in Danish coastal waters. *Estuarine, Coastal and Shelf Science*, 2000, **51**(2): 267-278.
- [4] Castillo CED, Coble PG. Seasonal variability of the colored dissolved organic matter during the 1994-95 NE and SW Monsoons in the Arabian Sea. *Deep Sea Research Part II: Topical Studies in Oceanography*, 2000, **47**(7/8): 1563-1579.
- [5] Zhu WJ. Preliminary study on optical characteristics and remote sensing inversion of colored dissolved organic matter (CDOM) in Changjiang Estuary and adjacent sea area[Dissertation]. Shanghai: East China Normal University, 2010. [朱伟健. 长江口及邻近海域有色溶解有机物(CDOM)的光学特性和遥感反演的初步研究[学位论文]. 上海: 华东师范大学, 2010.]
- [6] Hedges FE, Farrington J. Comments from the editors on the Suzuki statement. *Marine Chemistry*, 1993, **41**: 289-290.
- [7] Chen J, Wang BJ, Sun JH *et al.* Spatial distribution of CDOM concentration in Taihu Lake based on Landsat/TM image. *Spectroscopy and Spectral Analysis*, 2011, **31**(1): 34-38. [陈军, 王保军, 孙记红等. 基于 Landsat/TM 影像提取太湖 CDOM 浓度空间分布. 光谱学与光谱分析, 2011, **31**(1): 34-38.]
- [8] Hanson CE, Waite AM, Thompson PA *et al.* Phytoplankton community structure and nitrogen nutrition in Leeuwin Current and coastal waters of the Gascoyne region of Western Australia. *Deep-Sea Research II*, 2007, **54**(8/9/10): 902-924.
- [9] Kowalczyk P, Olszewski J, Darecki M *et al.* Empirical relationships between coloured dissolved organic matter(CDOM) absorption and apparent optical properties in Baltic Sea waters. *International Journal of Remote Sensing*, 2005, **26**(2): 345-370.
- [10] Kutser T, Pierson DC, Kallio KY *et al.* Mapping lake CDOM by satellite remote sensing. *Remote Sensing of Environment*, 2005, **94**(4): 535-540.
- [11] Kutser T, Pierson DC, Tranvik L *et al.* Using satellite remote sensing to estimate the colored dissolved organic matter absorption coefficient in lakes. *Ecosystems*, 2005, **8**(6): 709-720.
- [12] Feng LQ, Shi ZQ, Pan JJ *et al.* Absorption fluorescence characteristics and remote sensing algorithm of colored soluble organic matter in Taihu Lake in winter. *J Lake Sci*, 2011, **23**(3): 348-356. DOI:10.18307/2011.0306. [冯龙庆, 时志强, 潘剑君等. 太湖冬季有色可溶性有机物吸收荧光特性及遥感算法. 湖泊科学, 2011, **23**(3): 348-356.]
- [13] Zhou YM, Li JS, Shen Q *et al.* Inversion of colored organic matter in Guanting Reservoir based on surface spectral data. *Spectroscopy and Spectral Analysis*, 2015, **35**(4): 1015-1019. [周亚明, 李俊生, 申茜等. 基于水面光谱数据的官厅水库有

- 色可溶性有机物反演. 光谱学与光谱分析, 2015, 35(4): 1015-1019.]
- [14] Chen CQ, Pan ZL, Shi P. Application of seawater spectrum and its application in remote sensing inversion of yellow matter. *Journal of Tropical Oceanography*, 2003, 22(5): 33-39. [陈楚群, 潘志林, 施平. 海水光谱模拟及其在黄色物质遥感反演中的应用. 热带海洋学报, 2003, 22(5): 33-39.]
- [15] Kowalczyk P, Olszewski J, Darecki M *et al.* Empirical relationships between colored dissolved organic matter (CDOM) absorption and apparent optical properties in Baltic Sea. *International Journal of Remote Sensing*, 2005, 26(2): 345-370.
- [16] Feng LQ. Study on estimation model of phycocyanin and CDOM concentration in Taihu Lake based on hyperspectral remote sensing [Dissertation]. Nanjing: Nanjing Agricultural University, 2011. [冯龙庆. 基于高光谱遥感的太湖水体藻蓝素和 CDOM 浓度估算模型研究 [学位论文]. 南京: 南京农业大学, 2011.]
- [17] Cheng CM, Wei YC, Zhang J *et al.* First-order differential estimation model of chlorophyll a concentration in turbid water based on spectral smoothing. *Geography and Geo-Information Science*, 2013, 29(3): 30-34. [程春梅, 韦玉春, 张静等. 基于光谱平滑的浑浊水体叶绿素 a 浓度一阶微分估算模型. 地理与地理信息科学, 2013, 29(3): 30-34.]
- [18] Morel A, Prieur L. Analysis of variations in ocean color. *Limnology and Oceanography*, 1977, 22(4): 709-722.
- [19] Sathyendranath S, Morel A. Light emerging from the sea—interpretation and uses in remote sensing//Remote sensing applications in marine science and technology. Netherlands: Springer, 1983: 323-357.
- [20] Sun XL, Shu QT, Ou GL *et al.* Study on biomass estimation of *Pinus kesiyavar langbianensis* plantation based on random forest regression. *Guangdong Agricultural Sciences*, 2015, 42(15): 148-153. [孙雪莲, 舒清态, 欧光龙等. 基于随机森林回归的不同龄组思茅松人工林生物量遥感估测. 广东农业科学, 2015, 42(15): 148-153.]
- [21] Sun DY, Li YM, Wang Q *et al.* Neural network model for estimating CDOM concentration in Taihu Lake by using hyperspectral data. *Journal of Wuhan University: Information Science Edition*, 2009, 34(7): 851-855. [孙德勇, 李云梅, 王桥等. 利用高光谱数据估算太湖水体 CDOM 浓度的神经网络模型. 武汉大学学报: 信息科学版, 2009, 34(7): 851-855.]
- [22] Xu C. Remote sensing inversion model of sea surface salinity based on machine learning. Beijing: China University of Geosciences, 2016. [徐朝. 基于机器学习的海面盐度遥感反演模型 [学位论文]. 北京: 中国地质大学, 2016.]
- [23] Wang XL, Zhou ZY, Yan JP *et al.* Application of GA-SVM to multi-spectral remote sensing inversion of Weihe River water quality. *Journal of Remote Sensing*, 2009, 13(4): 735-744. [汪西莉, 周兆永, 延军平等. 应用 GA-SVM 的渭河水质参数多光谱遥感反演. 遥感学报, 2009, 13(4): 735-744.]
- [24] Lv J. Study on hyperspectral inversion model of crop chlorophyll content based on machine learning and radiation transmission model [Dissertation]. Beijing: China University of Geosciences, 2012. [吕杰. 基于机器学习和辐射传输模型的农作物叶绿素含量高光谱反演模型 [学位论文]. 北京: 中国地质大学, 2012.]
- [25] Li SJ. Study on inversion algorithm of lake optical water quality parameters by using analytical method—Taking Chaohu Lake in Anhui Province as an example [Dissertation]. Beijing: Beijing University, 2003. [李素菊. 利用分析方法建立湖泊光学水质参数反演算法研究——以安徽巢湖为例 [学位论文]. 北京: 北京大学, 2003.]
- [26] Leo B. Random forests. *Machine Learning*, 2001, 45: 5-32.
- [27] Li XH. Application of stochastic forest model in classification and regression analysis. *Journal of Applied Insectology*, 2013, 50(4): 1190-1197. [李欣海. 随机森林模型在分类与回归分析中的应用. 应用昆虫学报, 2013, 50(4): 1190-1197.]
- [28] Wang LA, Ma C, Zhou XD *et al.* Remote sensing estimation of wheat leaf based on random forest regression algorithm. *Journal of Agricultural Mechanization*, 2015, 46(1): 259-265. [王丽爱, 马昌, 周旭东等. 基于随机森林回归算法的小麦叶片 SPAD 值遥感估算. 农业机械学报, 2015, 46(1): 259-265.]
- [29] Yue ZG. Brief introduction of European "Sentinel"-3 satellite and its optical payload. *Infrared*, 2011, (7): 43-46. [岳植干. 欧洲“哨兵”-3 号卫星及其光学有效载荷简介. 红外, 2011, (7): 43-46.]
- [30] Tang JW, Tian GL, Wang XY *et al.* Measurement and analysis of water spectrum I: Measurement of water surface. *Journal of Remote Sensing*, 2004, 8(1): 37-44. [唐军武, 田国良, 汪小勇等. 水体光谱测量与分析 I: 水面以上测量法. 遥感学报, 2004, 8(1): 37-44.]
- [31] Bricaud A, Morel A, Prieur L. Absorption by dissolved organic matter in the sea (yellow substance) in the UV and visible domain. *Limnology and Oceanography*, 1981, 26: 43-53.
- [32] Fang KN, Wu JB, Zhu JP *et al.* A review of random forest methods. *Journal of Statistics and Information*, 2011, 26(3): 32-38. [方匡南, 吴见彬, 朱建平. 随机森林方法研究综述. 统计与信息论坛, 2011, 26(3): 32-38.]
- [33] Shen HY, Wang ZX, Gao CY *et al.* Determination of the number of hidden layer elements in BP neural network. *Journal of*

- Tianjin University of Technology, 2008, **24**(5): 13-15. [沈花玉, 王兆霞, 高成耀等. BP 神经网络隐含层单元数的确定. 天津理工大学学报, 2008, **24**(5): 13-15.]
- [34] Zhu WN, Yu Q. Inversion of chromophoric dissolved organic matter from EO-1 Hyperion imagery for turbid estuarine and coastal waters. *IEEE Transactions on Geoscience and Remote Sensing*, 2013, **51**(6): 3286-3298.
- [35] Ogashawara I, Mishra DR, Nascimento RFF *et al.* Re-parameterization of a quasi-analytical algorithm for colored dissolved organic matter dominant inland waters. *International Journal of Applied Earth Observation and Geoinformation*, 2016, **53**: 128-145.
- [36] Zhang YL, Wu SC, Qin BQ *et al.* Photo-absorption of colored organic matter in Meiliang Bay, Taihu Lake. *China Environmental Science*, 2004, **24**(4): 405-409. [张运林, 吴生才, 秦伯强等. 太湖梅梁湾有色可溶性有机物对光的吸收. 中国环境科学, 2004, **24**(4): 405-409.]
- [37] Zhang YL, Qin BQ, Ma RH *et al.* Absorption and fluorescence properties of colored soluble organic compounds in typical grass and algal lakes of Taihu Lake. *Environmental Science*, 2005, **26**(2): 142-147. [张运林, 秦伯强, 马荣华等. 太湖典型草、藻型湖区有色可溶性有机物的吸收及荧光特性. 环境科学, 2005, **26**(2): 142-147.]
- [38] IOCCG. Remote sensing of ocean colour in coastal, and other optically-complex, waters[M]//Reports of the International Ocean-Colour Coordinating Group. 2000.
- [39] Zhang YL, Qin BQ. Characteristics and possible sources of CDOM in summer and winter in Meiliang Bay and Taihu Lake. *Advances in Water Science*, 2007, **18**(3): 415-423. [张运林, 秦伯强. 梅梁湾、大太湖夏季和冬季 CDOM 特征及可能来源分析. 水科学进展, 2007, **18**(3): 415-423.]
- [40] Zhang YL, Qin BQ, Yang LY. Absorption characteristics of suspended particulate matter and CDOM in Meiliang Bay of Taihu Lake. *Acta Ecologica Sinica*, 2006, **26**(12): 3969-3979. [张运林, 秦伯强, 杨龙元. 太湖梅梁湾水体悬浮颗粒物和 CDOM 的吸收特性. 生态学报, 2006, **26**(12): 3969-3979.]
- [41] Cheng QL, Zheng BH, Wang SR *et al.* Three-dimensional fluorescence spectra of colored soluble organic matter(CDOM) in Dianchi Lake. *Spectroscopy and Spectral Analysis*, 2014, **34**(3): 698-703. [程庆霖, 郑丙辉, 王圣瑞等. 滇池水体有色溶解性有机质(CDOM)三维荧光光谱特征. 光谱学与光谱分析, 2014, **34**(3): 698-703.]
- [42] Zhang H, Huang JZ, Li YM *et al.* Study on absorption characteristics of optical substances in Dianchi Lake. *Environmental Science*, 2011, **32**(2): 452-463. [张红, 黄家柱, 李云梅等. 滇池水体光学物质的吸收特性研究. 环境科学, 2011, **32**(2): 452-463.]
- [43] Zhang YL, Qin BQ, Yang LY *et al.* Spatial distribution and optical behavior of colored soluble organic matter in Meiliang Bay, Taihu Lake. *J Lake Sci*, 2006, **18**(4): 319-326. DOI: 10.18307/2006.0401. [张运林, 秦伯强, 杨龙元. 太湖梅梁湾有色可溶性有机物的空间分布及光学行为. 湖泊科学, 2006, **18**(4): 319-326.]
- [44] Liu ML. Study on absorption characteristics of colored soluble organic matter in Taihu Lake[Dissertation]. Nanjing: Nanjing Agricultural University, 2009. [刘明亮. 太湖有色可溶性有机物吸收特性研究[学位论文]. 南京: 南京农业大学, 2009.]
- [45] Li L, Wang HF, Wang SR *et al.* Spatiotemporal variation and morphological contribution of nitrogen into lake load in Dianchi Lake. *Environmental Science Research*, 2016, **29**(6): 829-836. [李乐, 王海芳, 王圣瑞等. 滇池河流氮入湖负荷时空变化及形态组成贡献. 环境科学研究, 2016, **29**(6): 829-836.]