Eco-Environmental Knowledge Web

环境工程学报
Chinese Journal of Environmental Engineering

第 16 卷 第 9 期 2022 年 9 月
Vol. 16, No.9 Sep. 2022

http://www.cjee.ac.cn E-mail: cjee@rcees.ac.cn (010) 62941074

文章栏目：环境监测技术
DOI 10.12030/j.cjee.202204166 中图分类号 X87 文献标识码 A

谢恩弘, 吴骏恩, 杨昆. 基于 Sentinel-2 影像的洱海叶绿素 a 质量浓度反演[J]. 环境工程学报，2022, 16(9): 3058-3069. [XIE Enhong, WU Junen, YANG Kun. Mass concentration inversion for chlorophyll a in Erhai lake based on Sentinel-2[J]. Chinese Journal of Environmental Engineering, 2022, 16(9): 3058-3069.]

基于 Sentinel-2 影像的洱海叶绿素 a 质量浓度反演

谢恩弘[1,2], 吴骏恩[1,2], 杨昆[1,2,✉]

1. 云南师范大学地理学部，昆明 650500；2. 西部资源环境地理信息技术教育部工程研究中心，昆明 650500

摘　要　为动态监测洱海水体富营养污染物，利用遥感技术对反映水体富营养化的核心参数——叶绿素 a 质量浓度进行反演，建立适合当地当季的反演模型，对水体叶绿素 a 质量浓度进行宏观监测；通过洱海的秋季 Sentinel-2 影像和实测叶绿素 a 质量浓度数据，使用参数相关分析方法选取反演波段，建立 BP 神经网络模型和多元线性回归模型，随机选择 7 个样本点对 2 种模型进行交叉验证后，对洱海叶绿素 a 质量浓度进行反演。结果表明：Sentinel-2 数据与叶绿素 a 质量浓度具有显著的相关关系 (Pearson 积矩相关系数的绝对值大于 0.7, $P < 0.001$)，且分别在单波段、单波段比值和双波段比值中相关系数最大的波段及波段组合为 B6、B7/B6 和 (B6+B8)/(B7+B8a)；隐含层神经元节点数为 4 的 3 层 BP 神经网络模型的均方根误差最小，决定系数最大，分别为 0.002 8 和 0.925；2019 年 10 月 12 日、11 月 9 日，洱海叶绿素 a 质量浓度在空间上均呈北部高于南部的分布状态；BP 神经网络模型的平均绝对误差百分比为 21.36%、均方根误差为 0.002 8、决定系数为 0.925，多元线性回归模型的平均绝对误差百分比为 27.90%、均方根误差为 0.004 5、决定系数为 0.788。总体而言，BP 神经网络模型的叶绿素 a 质量浓度反演精度高于多元线性回归模型。本研究成果可为相关部门对洱海水质进行动态监测以及制定洱海水质保护措施提供参考。

关键词　叶绿素反演；Sentinel-2；洱海；BP 神经网络

湖泊是淡水资源的重要载体，对区域生态环境的维护发挥着重要的作用[1]。随着中国经济的快速发展以及城镇面积的大幅扩张，国内生产与生活用水需求日益增长，由此造成的湖泊环境与生态破坏现象不胜枚举，湖泊水华暴发、水体缺氧等问题频频发生，并由此引发了一系列经济与社会问题[2]。洱海亦出现了水体富营养化的现象[1]，水生生态系统逐步退化、湖泊水体富营养化进程加快等问题日益突出[3]，同时针对洱海的相关研究[4]发现，其水质呈现变坏的趋势，暗藏蓝藻水华暴发的隐患。

蓝藻水华通常是指在富营养化水体中出现蓝藻大量繁殖的现象，主要表现为水体表面覆盖着一层蓝绿色并伴有恶臭气味的浮沫，水体藻细胞浓度一般都达到并超过 1.5×10^7 个·L^{-1}，叶绿素 a 质量浓度大于 10 mg·m^{-3}[5]。叶绿素 a(chlorophyll a, Chl-a) 是浮游植物 (藻类) 中最常见的色素，其质量浓度是评价水体富营养化程度的核心参数[6]。传统的湖泊营养状态评价方法主要依赖于对水体的实地取样分析，然而该方法易受局部天气与环境影响，取样与测试过程也比较耗时耗力、成本较高，较难实现对湖泊富营养状态在精细时空尺度上的监测[7]。与之相比，遥感 (remote sensing,

收稿日期：2022-04-25；**录用日期**：2022-07-03

基金项目：国家自然科学基金资助项目 (42071381)

第一作者：谢恩弘 (1994—)，男，硕士研究生，ynenhong.xie@qq.com；**✉通信作者**：杨昆 (1963—)，男，博士，教授，kmdcynu@163.com

RS) 技术具有覆盖范围广、获取资料快、周期短等优点，可弥补传统水质采样的诸多不足，因此，已被广泛应用于湖泊水环境和水生态等方面的监测。目前，在利用遥感技术的水质参数反演方面，反演方法经历了分析法、经验法、半经验法、机器学习和综合法[8]，已建立起具有较高精度和一定普适性的水质参数反演模型，可以用于宏观的水质评价[9-10]。并且形成了遥感反演叶绿素 a 质量浓度的多种算法，但不同的算法也存在一定的局限性[11]，且不同的算法在不同的传感器之间的适应性也存在差异[12]。

已有许多针对不同地区、不同季节、不同水质参数、不同的反演方法和算法、不同的卫星遥感数据源的水质反演方面的研究。潘鑫等[13] 利用高分六号卫星影像，采用 3 种模型对太湖进行叶绿素 a 质量浓度反演，得出了适合高分六号卫星影像太湖叶绿素 a 质量浓度反演的模型。郑震[14] 基于 OLI 影像，建立了叶绿素 a 质量浓度反演的数学回归模型，分析了东张水库叶绿素 a 质量浓度的时空分布特点。陈命男[15] 利用 Landsat 8 数据，建立了淀山湖的叶绿素 a 反演的回归模型。但雨生等[16] 基于 Sentinel-2 数据建立了可靠的 BP 神经网络模型，用以监测平寨水库水质。马丰魁等[17] 以密云水库为研究对象，采用 BP 神经网络算法反演 4 个水质参数，并且得到了较为可信的研究结果。徐鹏飞等[18] 建立了神经网络模型，对千岛湖清洁水体的叶绿素 a 质量浓度进行反演，并利用该模型对千岛湖的叶绿素 a 质量浓度进行时空特征分析。

由此可见，已有许多利用遥感数据反演水质参数的研究，分析方法亦较为成熟，这些研究为不同地区湖泊的水质监测提供了可靠的参考依据。但内陆水体光学特征具有较强的区域性和季节性[8]，而且针对叶绿素 a 反演的各种算法仍受到季节和地理位置等的限制[11]，致使各地区建立的水质参数反演模型不具有普适性。为此，针对不同地区、不同季节及不同的传感器[11]，仍需要根据实际情况有针对性地建立适合当地的相关模型，为水污染防治提供合理的数据支撑[19]。近年来，利用遥感技术监测洱海水质情况的研究主要包括蓝藻水华的空间分布特征[20]、土地利用变化与水质的关系[21]、干季水质的时空变化[22] 等。毕顺等[23] 利用 OLCI 数据，采用了三波段模型对洱海 2017 年 4 月 19 日叶绿素 a 质量浓度的分布进行了估算。但该研究的三波段模型中第三波段的选取需要满足一系列的假设条件，且三波段模型主要适用于中高浓度叶绿素水体，不适用于高度浑浊水体[11]；还有研究[24] 表明，OLCI 数据虽具有较高辐射分辨率，但其空间分辨率 (为 300 m) 较低，在中小型内陆水体的监测上能力有限。由于洱海属中型湖泊，因此，选用空间分辨率较高的多光谱遥感影像能较为准确地获取水质采样点的反射率数据，这也是提高模型水质反演精度较为主要的因素之一[25]。

鉴于以上所述，本研究选取空间分辨率较高、也是近些年最流行的多光谱遥感数据之一的 Sentinel-2 数据，以较少利用 Sentinel-2 数据反演叶绿素 a 质量浓度的洱海作为研究区域，建立 2 种叶绿素 a 质量浓度反演模型，反演洱海叶绿素 a 质量浓度的空间分布，旨在利用不同数据源和方法探索适用于洱海流域的叶绿素 a 质量浓度反演模型，为相关部门的水质监测和水污染防治提供参考。

1　研究区域简介和数据获取

1.1　研究区域

洱海位于大理市，湖泊面积约 256 km^2，是云南省第二大高原淡水湖泊，也是大理市和周边乡村居民的生产生活用水供给源地。其水量补给主要为大气降水和入湖径流，周边主要入湖河流有 29 条。洱海区域属高原季风气候类型，四季温和、平均温度较小，日照差大，光照充足，干湿季节分明，雨季季节分配不均[1]。作为受人类活动干扰严重的中型湖泊，2010—2019 年期间洱海综合营养状态指数为 38.8 ~ 43.1，属中营养水平；2014—2019 年，整体水质类别为Ⅱ ~ Ⅲ类，水质

呈现变坏趋势，污染物主要来源于降水产生的地表径流所携带的禽畜养殖、农村生活和农田污染[3-4]。

1.2 研究数据来源

1) 叶绿素 a 质量浓度数据。叶绿素 a 质量浓度实测数据来源于云南省水环境监测中心大理州分中心，水质监测点共计 15 个 (图 1)，采用 2019 年 10 月 9 日、2019 年 11 月 8 日 2 期监测数据，共计 30 条，即每月监测 1 次，每次采集 15 个水质样品，水质采样在当天完成。

2) Sentinel-2 卫星数据及预处理。此次研究采用的是 Sentinel-2 数据，卫星数据从欧洲空间局 (European Space Agency, ESA) 网站下载，数据信息见表 1。Sentinel-2 是欧洲空间局哥白尼计划下的一个地球观测任务，该计划是由 2 颗相同的卫星 Sentinel-2A 与 Sentinel-2B 组成的卫星群，单颗卫星重返周期为 10 d，Sentinel-2 的卫星 2 颗互补，重返周期为 5 d。Sentinel-2 的每颗卫星都搭载相同的多光谱影像仪 (multispectral instrument, MSI)。该影像仪可拍摄涵盖可见光、近红外与短波红外的 13 个波段影像。MSI 的拍摄方式是推扫式，影像幅宽达到 290 km。通过 Sentinel-2 获取的各波段信息如表 2 所示。

在光学数据中，Sentinel-2 卫星在红边范围含有 3 个波段的数据，为快捷反演大区域叶面

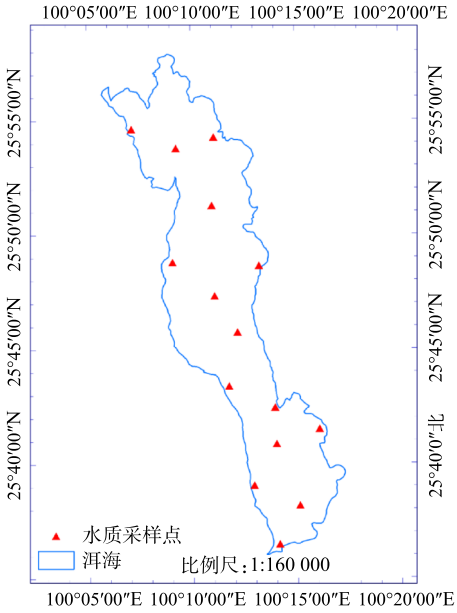


图 1 水质采样点分布图

Fig. 1 Distribution map of water quality sampling

表 1 Sentinel-2 数据信息

Table 1 Sentinel-2 data information

序号	日期	卫星	产品级别
1	2019年11月9日	Sentinel-2A	Level-2A
2	2019年10月12日	Sentinel-2B	Level-2A

表 2 Sentinel-2 传感器波段信息

Table 2 Spectral bands for the Sentinel-2 sensors

波段号	波段	Sentinel-2A		Sentinel-2B		空间分辨率/m
		中心波长/nm	波宽/nm	中心波长/nm	波宽/nm	
B1	沿海气溶胶	442.7	21	442.3	21	60
B2	蓝	492.4	66	492.1	66	10
B3	绿	559.8	36	559.0	36	10
B4	红	664.6	31	665.0	31	10
B5	植被红边	704.1	15	703.8	16	20
B6	植被红边	740.5	15	739.1	15	20
B7	植被红边	782.8	20	779.7	20	20
B8	近红外	832.8	106	833.0	106	10
B8a	窄波近红外	864.7	21	864.0	22	20
B9	水蒸气	945.1	20	943.2	21	60
B10	短波红外-卷云	1 373.5	31	1 376.9	30	60
B11	短波红外	1 613.7	91	1 610.4	94	20
B12	短波红外	2 202.4	175	2 185.7	185	20

积指数、叶绿素质量浓度等生物物理量指标提供了可能[26]。本次研究直接获取 Sentinel-2 卫星的 L2A 级数据。L2A 级数据在 L1C 级数据的基础上，利用 ESA 提供的 Sen2cor 模型进行处理生成。从 2018 年 3 月开始，ESA 逐渐向全球用户提供 L2A 级产品，并于 2018 年 12 月覆盖到全球。L2A 级数据在生成过程中，会对 L1C 级数据进行大气校正、云雪检测、地形校正、场景分类等一系列处理[27]，可直接用于下游产品。

为便于使用 ENVI 5.3 提取各采样点反射率，利用 SNAP 软件将 2 期影像都按照同样的方法进行重采样，将其重采样为 10 m，之后，波段由 13 减至 12 个 (减少了短波红外-卷云 B10 波段)。分别提取水质采样点 2 期影像的水体反射率。

2　叶绿素 a 质量浓度反演模型的建立与验证

2.1　反演波段的选择

在水质参数的反演工作中，获取与水质参数关系密切的敏感波段，将其作为模型的输入因子，这样建立的模型具有更高的预测精度[28]。本研究通过 Pearson 相关分析来获取与叶绿素 a 质量浓度关系密切的敏感波段，Pearson 相关系数是一种线性相关系数，通过 Pearson 相关分析可以得出不同波段或者波段组合与叶绿素 a 质量浓度的相关性强弱，可进一步剔除弱相关的、可能干扰反演模型建立的波段信息。由于水体的反射特性主要位于可见光和近红外波段，且有研究[29]表明，可见光和近红外的反射率可以成功地用以反演水体的叶绿素 a 质量浓度，因此，首先选择 4 个可见光波段和 4 个红边波段进行相关性分析，然后利用置信度为 99.9%($P < 0.001$) 的波段进行波段组合，进一步分析各波段组合的反射率与叶绿素 a 质量浓度的相关性。

目前，常用的波段组合方式有单波段比值、双波段比值、波段差值、三波段和四波段模型。对于 Sentinel-2 数据来说，双波段比值对叶绿素质量浓度更加敏感[30]，因此，双波段比值是本研究首选的波段组合方式之一，另外还选择了单波段比值。根据但雨生等[16]基于 Sentinel-2 卫星数据与平寨水库叶绿素 a 质量浓度的相关分析研究结果，叶绿素 a 质量浓度与组合波段 B5/B4、(1/B4-1/B5)×B6、(1/B4-1/B5)×B7 和 (1/B4-1/B5)×B8 的 Pearson 相关系数最大。另外，叶绿素质量浓度与归一化植被指数 (normalized difference vegetation index, NDVI) 呈线性关系[31]，所以本研究选择具有代表性的波段比值、三波段模型以及 NDVI 筛选与叶绿素 a 质量浓度显著相关的敏感波段。利用 15 个采样点提取的 2 期共 30 条水体反射率数据，与实测叶绿素 a 质量浓度数据进行 Pearson 相关分析，分别在单波段、单波段比值和双波段比值中选取相关系数最大的叶绿素 a 反演波段。如表 3 所示，

表 3　各波段/波段组合反射率与叶绿素 a 质量浓度的相关性

Table 3　Correlation between reflectance of each band/band combination and chlorophyll-a mass concentration

波段/波段组合	Pearson相关系数	P	样本数	波段/波段组合	Pearson相关系数	P	样本数
B2	-0.148	0.435	30	B8/B6	0.297	0.110	30
B3	-0.099	0.602	30	B8/B7	0.470	0.009	30
B4	0.233	0.216	30	B7/B6	-0.713	<0.001	30
B5	0.186	0.324	30	(1/B4-1/B5)×B6	-0.128	0.501	30
B6	0.811	<0.001	30	(1/B4-1/B5)×B7	-0.086	0.653	30
B7	0.724	<0.001	30	(1/B4-1/B5)×B8	-0.098	0.606	30
B8	0.788	<0.001	30	(1/B4-1/B5)×B8a	-0.070	0.713	30
B8a	0.694	<0.001	30	(B6+B7)/(B8+B8a)	-0.382	0.037	30
B5/B4	-0.009	0.963	30	(B6+B8)/(B7+B8a)	0.821	<0.001	30
B8a/B6	0.081	0.669	30	(B8-B6)/(B8+B6)	0.363	0.048	30
B8a/B7	0.264	0.158	30	(B8-B7)/(B8+B7)	0.477	0.008	30
B8a/B8	-0.621	<0.001	30	(B8-B4)/(B8+B4)	0.489	0.006	30

B6、B7/B6 和 (B6+B8)/(B7+B8a) 反射率与叶绿素 a 质量浓度呈现 0.001 水平的显著性，且相关系数最大。因此，最终选择 B6、B7/B6 和 (B6+B8)/(B7+B8a) 作为叶绿素 a 质量浓度的反演波段。

需要说明的是，在进行 Pearson 相关分析之前，需要对各波段、波段组合提取的采样点反射率数据以及实测叶绿素 a 质量浓度数据进行 Shapiro-Wilk 检验。在 25 组数据中 (波段或波段组合的采样点反射率数据 24 组 (表 3)；实测叶绿素 a 质量浓度数据 1 组)，6 组数据符合正态分布，其余数据基本符合正态分布。

2.2 BP 神经网络模型的构建

BP (back propagation) 神经网络是一种按误差逆向传播算法训练的多层前馈网络，能学习和存储大量的“输入-输出”模式的映射关系，而无需事前揭示描述这种映射关系的数学方程[32]。BUCKTON 等[33] 和 SCHILLER 等[34] 利用 MERIS 数据和 2 层 BP 神经网络模型反演了叶绿素、悬浮物、黄色物质 3 个水质参数，得出神经网络模型完全可以用来反演Ⅰ类水质和Ⅱ类水质参数，且反演精度远高于经验模型的结论。BP 神经网络具有自适应、自学习、自组织和非线性映射的功能，非常适用于模拟各种错综复杂的非线性关系。已有研究[16] 证明，具有 1 个隐含层的 3 层网络可以逼近任意非线性函数。本研究所采用的即是具有 1 个输入层、1 个隐含层和 1 个输出层的 3 层 BP 神经网络结构。其中，隐含层的神经元节点数可根据 3 种经验公式[32] 选择，本研究由经验公式 (式 (1)) 进行确定。

$$M = \sqrt{m+n}+a \tag{1}$$

式中：M 为隐含层神经元个数；m 和 n 分别为输入层和输出层神经元个数；a 为 1 ~ 10 的整数。

根据输入输出层神经元个数 3 和 1，由式 (1) 计算得到 M 的取值为 3 ~ 12 的整数。分别从单波段、单波段比值、双波段比值中选取与叶绿素 a 质量浓度显著相关且相关系数最大的波段和波段组合 B6、B7/B6 和 (B6+B8)/(B7+B8a) 作为神经网络的输入数据 (波段组合使用软件 PIE-Basic 6.3 完成)，与之相对应的实测叶绿素 a 质量浓度作为输出数据。30 组数据随机分为训练数据 23 组和检验数据 7 组，采用 feedforwardnet 函数建立神经网络。trainglm 作为训练函数，双曲正切函数 sigmoid为传递函数，线性函数 purelin 作为输出层函数。训练次数设置为 1 000 次，学习速率为 0.01，训练目标为 1×10^{-6}。设置参数后，分别选择不同的神经元节点数对神经网络进行反复训练，最终获得决定系数 R^2 最大、均方根误差 (root mean square error, RMSE) 最小的神经网络结构，此时隐层神经元节点数为 4 (表 4)。

表 4 隐含层不同神经元个数的 R^2 和 RMSE

Table 4 R^2 and RMSE of different numbers of neurons in the hidden layer

神经元个数	决定系数	均方根误差
3	0.770	0.004 5
4	0.925	0.002 8
5	0.564	0.005 1
6	0.335	0.004 7
7	0.587	0.006 1
8	0.361	0.004 9
9	0.791	0.007 4
10	0.851	0.004 5
11	0.661	0.007 1
12	0.418	0.007 9

2.3 模型精度的检验

根据构建 BP 神经网络时随机产生的训练样本 23 组，也就是利用与 BP 神经网络相同的特征波段 B6、B7/B6 和 (B6+B8)/(B7+B8a) 及其数据，建立多元线性回归模型 (式 (2))。

$$C = 0.031\,0x_1 + 0.001\,6x_2 + 0.033\,4x_3 - 0.030\,2 \tag{2}$$

式中：C 为 Chl-a 质量浓度，$mg\cdot L^{-1}$；x_1 代表 B6 波段；x_2 代表单波段比值 B7/B6；x_3 代表双波段比值 (B6+B8)/(B7+B8a)。

采用决定系数 R^2、均方根误差 (RMSE)、平均绝对误差百分比 (mean absolute percentage error, MAPE)3 个指标评价模型效果。MAPE 的计算方法见式 (3)，7 条检验数据对 2 种模型的检验结果如表 5 所示。

$$\delta = \frac{1}{m}\sum_{i=1}^{m}\left|\frac{y_i - \widehat{y_i}}{y_i}\right| \times 100\% \tag{3}$$

式中：δ 为所有检验样本的相对误差绝对值的平均值；m 为检验样本总数；y_i 为第 i 个检验样本的叶绿素 a 质量浓度的实测值；$\widehat{y_i}$为第 i 个检验样本的叶绿素 a 质量浓度的估测值。

表 5　检验数据对 2 种模型误差检验结果

Table 5　Error testing of two models using test data

样本编号	实测值/(mg·L^{-1})	线性回归模型估测值	BP神经网络估测值	线性回归模型相对误差/%	BP神经网络相对误差/%
1	0.014 0	0.007 0	0.013 1	49.82	6.36
2	0.006 5	0.008 3	0.011 3	27.90	74.01
3	0.015 6	0.008 6	0.013 8	44.72	11.66
4	0.017 0	0.012 0	0.014 7	29.33	13.73
5	0.021 8	0.019 0	0.017 6	12.69	19.48
6	0.011 0	0.009 2	0.012 6	16.63	14.81
7	0.015 4	0.013 2	0.013 9	14.20	9.49

注：线性回归模型的平均绝对误差百分比(MAPE)为27.90%，均方根误差(RMSE)为0.004 5，决定系数(R^2)为0.788；BP神经网络的平均绝对误差百分比(MAPE)为21.36%，均方根误差(RMSE)为0.002 8，决定系数(R^2)为0.925。

3　结果与讨论

3.1　建模数据的时间匹配

从本研究的叶绿素 a 质量浓度数据的采集时间和 Sentinel-2 卫星影像的时间来看，2 种数据没有在时间上做到非常好的匹配。10 月 9 日，叶绿素 a 质量浓度实测数据与 10 月 12 日 Sentinel-2 卫星数据间隔 3 d；11 月 8 日，叶绿素 a 质量浓度实测数据与 11 月 9 日 Sentinel-2 卫星数据间隔 1 d。根据来源于美国国家海洋和大气管理局国家环境信息中心 (NOAA National Centers for Environmental Information) 的气象站点 (该气象站点位于洱海湖体西南部，大理市气象局东南部) 数据，研究区 10 月 9 日—10 月 10 日无降水，10 月 11 日降水量为 0.508 mm，10 月 12 日降水量为 0.254 mm，10 月 9 日—10 月 12 日 4 d 内最大平均风速为 2.16 m·s^{-1}，日平均温度变化最大为 1.2 ℃；11 月 8 日—11 月 9 日无降水量，最大平均风速为 1.59 m·s^{-1}，日平均温度升高 2.3 ℃。此外，来源于美国国家航空航天局 (National Aeronautics and Space Administration, NASA) 提供的大气再分析资料 MERRA-2 数据显示，10 月 9 日—10 月 12 日，大理市地区日照时数的最大值和最小值分别为 5.92 h 和 4.85 h，变化值最大为 1.07 h；11 月 8 日—11 月 9 日，日照时数减少 0.10 h。因此，时间上的差异对数据产生的影响很小。此外，鉴于本研究存在的数据在时间匹配上的误差，且叶绿素 a 质量浓度变化受风速、降水、流速、营养元素等因子影响[35]，可建立包含风速、降水、流速、营养元素等因子的模型，以减少因为数据不能在时间上完全匹配所产生的叶绿素 a 质量浓度的估算误差，这可以在一定程度上避免因卫星数据时间分辨率以及卫星数据云量过多造成的与实测数据的时间匹配问题。

3.2　相关分析结果

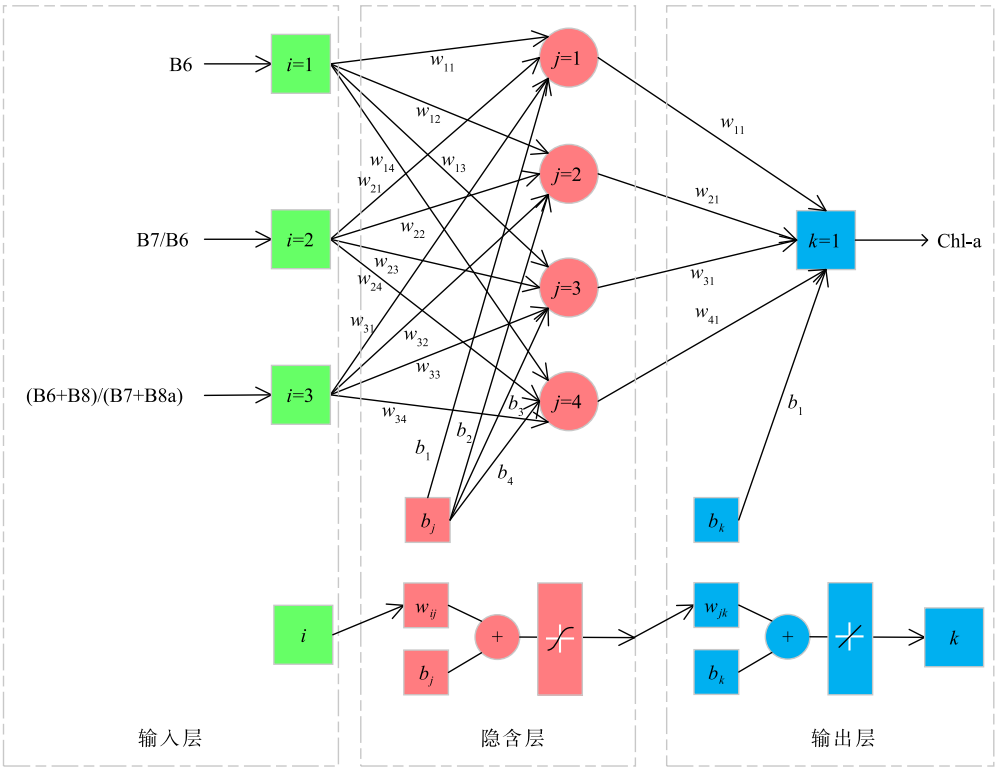
数据的 Shapiro-Wilk 检验结果显示，B9、B5/B4、B7/B6、B8a/B8、(1/B4-1/B5)×B6、(1/B4-1/B5)×

B7 没有呈现显著性 ($p > 0.05$)，说明以上 6 组数据符合正态分布。其余 19 组数据的峰度绝对值小于 10 并且偏度绝对值小于 3，这些数据的正态检验直方图也都显示为钟形 (中间高两端低)，说明这些数据基本符合正态分布。根据本研究的实测叶绿素 a 质量浓度数据与对应水质样品采样点的 Sentinel-2 影像反射率的相关分析结果，建立模型所采用的对叶绿素 a 质量浓度敏感的波段都处于 0.001 的显著水平，分别是在单波段、单波段比值、双波段比值中相关系数最大的波段和波段组合 B6、B7/B6 和 (B6+B8)/(B7+B8a)。根据表 3，B5/B4、(1/B4-1/B5)×B6、(1/B4-1/B5)×B7 和 (1/B4-1/B5)×B8 与叶绿素 a 质量浓度的相关系数很小，分别为−0.009、−0.128、−0.086、−0.098。这说明以上波段组合与叶绿素 a 质量浓度存在弱相关关系或几乎没有相关关系，与但雨生等[16] 的研究存在差异。在但雨生等[16] 的研究中，采用与叶绿素 a 实测数据准同步的 Sentinel-2 影像，在秋季，采用了同样的相关分析方法，因此，出现这种差异可能由不同研究区的水体光学特征的差别所引起，也就是说可能由于不同研究区水体的固有光学特征和表观光学特征的差异，导致同一卫星传感器的所接收的水体反射信息存在差别。此外，NDVI 与叶绿素 a 质量浓度的相关系数为 0.489，P 值为 0.006，说明 NDVI 与叶绿素 a 质量浓度存在 0.01 水平的显著性。

3.3 BP 神经网络与多元线性回归模型

本研究选用多元线性回归模型以及可以模拟各种非线性关系的 BP 神经网络模型来进行湖泊水叶绿素 a 质量浓度的反演。从对 2 种模型的检验结果来看，BP 神经网络模型的 MAPE 为 21.36%，小于线性回归模型 27.90%；RMSE 为 0.002 8，小于线性回归模型的 0.004 5。从 2 种模型的决定系数 R^2 来看，本研究的多元线性回归模型的 3 个自变量 (B6、B7/B6 和 (B6+B8)/(B7+B8a)) 只能用来解释因变量 (叶绿素 a 质量浓度) 方差的 78.8%，远低于 BP 神经网络模型 ($R^2 = 0.925$)。这说明 BP 神经网络模型的预测精度要高于线性回归模型。这与赵玉芹等[36] 基于 SPOT 5 遥感影像建立了多元线性回归模型、BP 神经网络模型和径向基神经网络模型，对渭河陕西段水域的 4 种水质参数进行反演所得出的结论几乎一致，均认为 BP 神经网络模型的预测精度最高。同样，岳佳佳等[37] 以黄石磁湖为研究区，利用 IKONOS 遥感影像和水质采样数据构建了多元线性回归模型、BP 神经网络模型和径向基神经网络模型，对比后，亦得出神经网络模型反演结果优于多元线性回归模型。此外，杨柳等[38] 也使用 ETM+数据和准同步实测的 2 种水质数据，建立多个隐含层为 1 的 BP 神经网络模型，得出了 BP 神经网络方法反演水质参数结果好于线性回归方法的结论。

在水质参数反演中，BP 神经网络虽然较多元线性回归模型有较高的精度，但是也存在一些缺点。BP 神经网络的设计通常需要确定网络的层数、输入层的节点数、隐含层的节点数、输出层神经元个数、网络的初始权值，需要选择不同的传递函数和训练方法。本研究使用 feedforwardnet 函数建立神经网络，该函数自动对数据进行归一化处理，可以根据 train 函数自动确定网络的输入层和输出层数。如图 2 所示，本研究的 BP 神经网络输入层数为 3，分别为 Sentinel-2 卫星数据的 B6、B7/B6 和 (B6+B8)/(B7+B8a)3 个特征数据。将这 3 个特征数据按照不同的权值 w_{ij} 分别输入到隐含层各个神经元，再和各神经元阈值 b_j 求和，之后激活传递函数 (双曲正切函数 sigmoid)，接着通过一定的权值 w_{jk} 由隐层进入输出层，与输出层阈值 b_k 求和后激活输出函数 (线性函数 purelin)，随后按减少误差的原则，从输出层经过隐含层，回到输入层，不断调整各连接权值，进行反复训练。虽然这样可以不断地提高正确率，但 BP 神经网络的一些参数的选择依然没有依据，隐含层的神经元个数也只能依据经验公式进行试凑，网络的权值是通过训练样本和学习率参数得到的，这也导致了 BP 神经网络的不稳定。除此之外，BP 神经网络每次训练的初始权值是随机给定的，导致了其不可重现性；网络还存在样本依赖性，主要依赖于选择的训练样本是否典型，所以，应避免样本集合代表性差、矛盾样本多、存在冗余样本等问题；再者，网络容易陷入局部最优，需要实时检测误差率的变化以确定最佳训练次数。相比之下，径向基网络模型具有结构可靠、不依赖初始权



注：i为输入层特征变量，i=1，2，3，分别表示B6、B7/B6、(B6+B8)/(B7+B8a)，共3个输入变量；w_{ij}为第i个特征变量输入到隐含层第j个神经元的权值，j为神经元个数，j=1，2，3，4；w_{ij}表示权值，从上至下各权值w_{11}、w_{12}、…、w_{34}分别为−1.424 7、−1.337 6、−0.665 1、2.037 6、−0.309 4、1.135 4、1.205 6、−1.048 0、−1.705 1、−0.195 2、1.582 4、−0.972 8；b_j为隐含层第j个神经元阈值，阈值b_1、b_2、b_3、b_4分别为2.188 0、1.090 9、−0.436 3、2.015 1；k为输出层变量，k=1，表示Chl-a；w_{jk}为隐含层第j个神经元输出到第k个变量的权值，各权值w_{11}、w_{21}、w_{31}、w_{41}分别为−0.420 7、−0.505 0、0.432 7、0.649 4；b_k为输出层阈值，b_1=0.068 3。

图 2　BP 神经网络结构示意图

Fig. 2　Schematic diagram of BP neural network structure

值等优点。吴倩等[39]认为径向基网络模型是值得推广的光谱反演叶绿素模型。同时需要注意的是，吴倩等[39]的研究采用地物光谱仪测定采样点水体反射光谱，波谱区间为 325 ~ 1 075 nm，光谱采样间隔为 1.6 nm，光谱分辨率为 3.5 nm，光谱分辨率较高，且径向基网络模型同时也存在丢失信息、数据不充分时无法工作等缺点。本研究旨在探索适用于秋季洱海流域的叶绿素 a 质量浓度的遥感反演模型，采用的是 Sentinel-2 多光谱数据，波段配置不同也会影响叶绿素 a 的反演[11]，因此，依然使用了在各个地区利用遥感反演水质参数比较成熟的 BP 神经网络模型。

3.4　叶绿素 a 质量浓度的空间分布

如表 5 所示，BP 神经网络的模型预测误差要比多元线性回归模型小，采用构建好的 BP 神经网络模型对洱海秋季 2019 年 10 月 12 日、11 月 9 日 2 d 的叶绿素 a 质量浓度进行反演，结果见图 3 和图 4。由反演结果可以看出，利用 Sentinel-2 影像可从宏观上反演叶绿素 a 质量浓度，并且可以非常直观地展现叶绿素 a 质量浓度的空间分布，这也是遥感监测的优势体现。利用这一优势，结合地理信息系统 (geographic information system, GIS) 以及人工智能 (artificial intelligence, AI) 技术，构建蓝藻水华及富营养化监测预警系统是未来的研究方向之一[40]。如以 10 mg·m^{-3} 叶绿素 a 质量浓度作为藻华发生的临界值[41]，利用以上技术结合适用于当地的 BP 神经网络模型，可以构建藻华发生的预警系统，这也是本研究的实际意义之一。

由洱海 2 d 的反演结果可以看出，叶绿素 a 质量浓度均呈现洱海北部高于南部的分布状态，且洱海北部沿岸地区叶绿素 a 质量浓度值最大。10—11 月，叶绿素 a 质量浓度出现扩散，北部较高

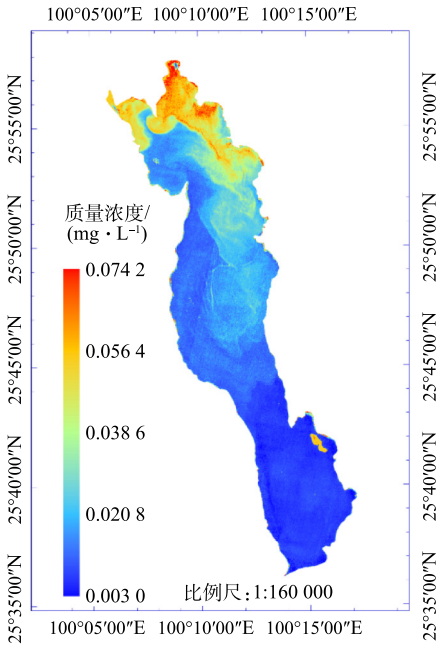


图 3 洱海 2019 年 10 月 12 日叶绿素 a 质量浓度分布图

Fig. 3 Distribution of chlorophyll a mass concentration in Erhai lake on October 12th, 2019

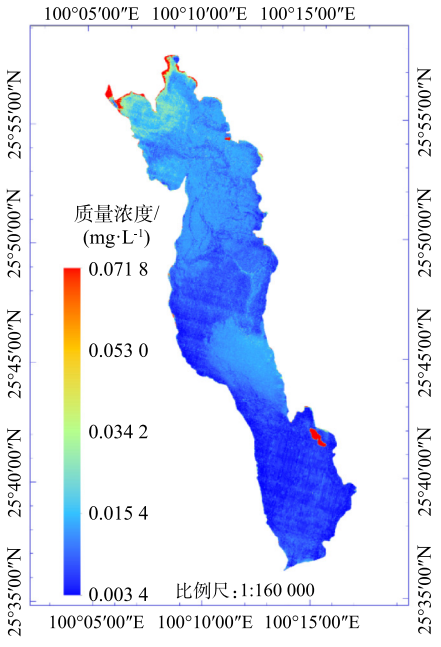


图 4 洱海 2019 年 11 月 9 日叶绿素 a 质量浓度分布图

Fig. 4 Distribution of chlorophyll a mass concentration in Erhai lake on November 9th, 2019

质量浓度逐渐向洱海中部扩散，叶绿素 a 质量浓度最大值降低，分布面积扩大。从叶绿素 a 质量浓度来看，10 月最大值为 0.074 2 mg·L^{-1}，最小值为 0.003 0 mg·L^{-1}；11 月最大值为 0.071 8 mg·L^{-1}，最小值为 0.003 4 mg·L^{-1}。10 月与 11 月相比较而言，洱海叶绿素 a 质量浓度区间基本没有发生变化。根据 11 月的反演结果，洱海叶绿素 a 质量浓度仍然呈北部大于南部的趋势。这与祁兰兰等[22]利用 GF-1 卫星数据的研究结果 (2014—2019 年 11 月，洱海叶绿素 a 质量浓度在空间上均呈现南部>中部>北部) 相反。

为了进一步探寻造成此结果的原因，首先想到的是将叶绿素 a 质量浓度与水体富营养化联系起来。这是因为叶绿素 a 质量浓度是藻类生物量的指标，可以很容易的与水体富营养化联系，而营养化水平与叶绿素 a 质量浓度又是成正相关的[42]，因此，将 2019 年 11 月 9 日 Sentinel-2A 的真彩色合成影像通过 3% 的线性拉伸显示 (图 5)，通过目视解译，可以看出洱海北部的水华现象。其次，由于本研究使用的是 Sentinel-2 数据，而通过查询，没有与祁兰兰等[22]研究中相同日期 (2019 年 11 月 22 日) 的 Sentinel-2 数据，未能与其做对比实验。相应地，在 11 月 9 日，也没有查询到高分一号 WFV 洱海区域的数据，也不能利用祁兰兰等[22]采用的反演模型对本研究结果做对比验证。但是，WANG 等[41]的研究显示，洱海叶绿素 a 质量浓度在 2016 年 11 月、2017 年 11 月均为北部

图 5 洱海 2019 年 11 月 9 日 Sentinel-2A 真彩色合成图

Fig. 5 Sentinel-2A true-color composite image of Erhai lake on November 9, 2019

高于南部；TAN 等[42] 的研究同样表明 2016 年 11 月洱海北部叶绿素 a 质量浓度高于南部。这种结果的差异可能是祁兰兰等[22] 研究所采用模型的区域性限制，该模型最早是利用环境一号卫星 CCD 数据以太湖为研究区建立的[43]，后来被朱利等[44] 证明高分一号 WFV 数据与环境一号 CCD 数据在波段设置上具有很强的一致性，并将该模型用于高分一号 WFV 数据的太湖叶绿素 a 质量浓度反演。

4　结论

1) 在秋季的洱海流域，对于 Sentinel-2 卫星数据，与叶绿素 a 质量浓度具有显著的相关关系 ($P < 0.001$)，且在单波段、单波段比值和双波段比值中相关系数最大的分别为 B6、B7/B6 和 (B6+B8)/(B7+B8a)。

2) 使用 BP 神经网络，建立了适用于秋季洱海流域的叶绿素 a 质量浓度反演模型。对比使用同样数据建立的多元线性回归模型，BP 神经网络模型的平均绝对误差百分比 (MAPE) 和均方根误差 (RMSE) 均小于多元线性回归模型，决定系数 R^2 大于多元线性回归模型。总体来说，BP 神经网络模型的精度高于线性回归模型，可使用 Sentinel-2 数据，利用本研究构建的 BP 神经网络模型反演叶绿素 a 质量浓度，能得到较可靠的反演结果。

3) 通过构建的具有 4 个隐含层神经元节点的 3 层 BP 神经网络模型，反演洱海 2019 年 10 月 12 日、11 月 9 日叶绿素 a 质量浓度，结果均显示洱海北部叶绿素 a 质量浓度明显高于南部。因此，基于 Sentinel-2 影像和 BP 神经网络模型可以宏观监测叶绿素 a 质量浓度的空间分布。

参 考 文 献

[1] 彭文启, 王世岩, 刘晓波. 洱海水质评价[J]. 中国水利水电科学研究院学报, 2005，3(3): 192-198.

[2] 项继权. 湖泊治理: 从“工程治污”到“综合治理”: 云南洱海水污染治理的经验与思考[J]. 中国软科学, 2013(2): 81-89.

[3] 马巍, 苏建广, 杨洋, 等. 洱海水质演变特征及主要影响因子分析[J/OL]. 中国水利水电科学研究院学报, 1-9(2021-06-23)[2022-06-10].https://kns.cnki.net/kcms/detail/11.5020.TV.20210623.1038.001.html. 10.13244/j.cnki.jiwhr.20200248

[4] 石宏博, 黄玥, 李杰, 等. 洱海水质评价及污染物来源分析[J]. 水电能源科学, 2021, 39(10): 72-75.

[5] 张亚鹏. 真菌棘孢木霉SHS3对微囊藻的溶藻活性及其溶藻机制[D]. 南京: 南京大学, 2018.

[6] CUI T W, ZHANG J, WANG K, et al. Remote sensing of chlorophyll a concentration in turbid coastal waters based on a global optical water classification system[J]. ISPRS Journal of Photogrammetry and Remote Sensing, 2020, 163: 187-201.

[7] 周博天, 张雅燕, 施坤. 湖泊营养状态遥感评价及其表征参数反演算法研究进展[J]. 遥感学报, 2022, 26(1): 77-91.

[8] 王波, 黄津辉, 郭宏伟, 等. 基于遥感的内陆水体水质监测研究进展[J]. 水资源保护, 2022, 38(3): 117-124.

[9] 王林, 白洪伟. 基于遥感技术的湖泊水质参数反演研究综述[J]. 全球定位系统, 2013, 38(1): 57-61.

[10] 吴煜晨. 基于MODIS遥感数据源的内陆水体叶绿素a浓度反演算法综述[J]. 江西水利科技, 2017, 43(1): 14-18.

[11] 罗婕纯一, 秦龙君, 毛鹏, 等. 水质遥感监测的关键要素叶绿素a的反演算法研究进展[J]. 遥感技术与应用, 2021, 36(3): 473-488.

[12] JOHANSEN R, BECK R, NOWOSAD J, et al. Evaluating the portability of satellite derived chlorophyll-a algorithms for temperate inland lakes using airborne hyperspectral imagery and dense surface observations[J]. Harmful Algae, 2018, 76: 35-46.

[13] 潘鑫, 杨子, 杨英宝, 等. 基于高分六号卫星遥感影像的太湖叶绿素a质量浓度反演[J]. 河海大学学报(自然科学版), 2021, 49(1): 50-56.

[14] 郑震. 基于OLI遥感影像的叶绿素a质量浓度反演研究[J]. 灌溉排水学报, 2017, 36(3): 89-93.

[15] 陈命男. 基于Landsat 8数据的淀山湖叶绿素a浓度遥感反演研究[J]. 中国资源综合利用, 2021, 39(2): 44-46.

[16] 但雨生, 周忠发, 李韶慧, 等. 基于Sentinel-2的平寨水库叶绿素a浓度反演[J]. 环境工程, 2020, 38(3): 180-185.

[17] 马丰魁, 姜群鸥, 徐黎丹, 等. 基于BP神经网络算法的密云水库水质参数反演研究[J]. 生态环境学报, 2020, 29(3): 569-579.

[18] 徐鹏飞, 程乾, 金平斌. 基于神经网络模型的千岛湖清洁水体叶绿素a遥感反演研究[J]. 长江流域资源与环境, 2021, 30(7): 1670-1679.

[19] 张宏建, 王冰, 周健, 等. 基于BP神经网络的内陆河流水质遥感反演[J]. 华中师范大学学报(自然科学版), 2022, 56(2): 333-341.

[20] 张娇, 陈莉琼, 陈晓玲. 基于FAI方法的洱海蓝藻水华遥感监测[J]. 湖泊科学, 2016, 28(4): 718-725.

[21] 杜芳芳. 湖泊流域土地利用变化与湖泊水质关系研究[D]. 昆明: 昆明理工大学, 2011.

[22] 祁兰兰, 王金亮, 农兰萍, 等. 基于GF-1卫星数据的洱海干季水质时空变化监测[J]. 人民长江, 2021, 52(9): 24-31.

[23] 毕顺, 李云梅, 吕恒, 等. 基于OLCI数据的洱海叶绿素a浓度估算[J]. 湖泊科学, 2018, 30(3): 701-712.

[24] KRAVITZ J, MATTHEWS M, BERNARD S, et al. Application of Sentinel 3 OLCI for chl-a retrieval over small inland water targets: Successes and challenges[J]. Remote Sensing of Environment, 2020, 237: 111562.

[25] 冼翠玲, 张艳军, 张明琴, 等. 基于高分辨率多光谱影像的温瑞塘河水质反演模型研究[J]. 中国农村水利水电, 2017(3): 90-95.

[26] 杨斌, 李丹, 王磊, 等. 基于Sentinel-2A岷江上游地表生物量反演与植被特征分析[J]. 科技导报, 2017, 35(21): 74-80.

[27] European Space Agency. Level-2A algorithm overview[EB/OL]. (2022-03-19)[2022-04-10].https://sentinel.esa.int/web/sentinel/technical-guides/sentinel-2-msi/level-2a/algorithm.

[28] 赵姝雅. 基于遥感的白洋淀水质参数反演研究[D]. 廊坊: 北华航天工业学院, 2019.

[29] 韩秀珍, 郑伟, 刘诚, 等. 基于MERSI和MODIS的太湖水体叶绿素a含量反演[J]. 地理研究, 2011, 30(2): 291-300.

[30] 董舜丹, 何宏昌, 付波霖, 等. 基于Landsat-8陆地成像仪与Sentinel-2多光谱成像仪传感器的香港近海海域叶绿素a浓度遥感反演[J]. 科学技术与工程, 2021, 21(20): 8702-8712.

[31] 谢小红, 魏虹, 李昌晓, 等. 水淹胁迫下枫杨(*Pterocarya stenoptera* C. DC.)幼苗叶片高光谱特征的研究[J]. 西南大学学报(自然科学版), 2011, 33(4): 93-98.

[32] 张德丰. MATLAB R2020a神经网络典型案例分析[M]. 北京: 电子工业出版社, 2021.

[33] BUCKTON D, O'MONGAIN E, DANAHER S. The use of neural networks for the estimation of oceanic constituents based on the MERIS instrument[J]. International Journal of Remote Sensing, 1999, 20(9): 1841-1851.

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

[35] 张丽华, 林茂森, 田英, 等. 环境因子对大伙房水库叶绿素α的影响研究[J]. 环境科学与技术, 2020, 43(4): 230-236.

[36] 赵玉芹, 汪西莉, 蒋赛. 渭河水质遥感反演的人工神经网络模型研究[J]. 遥感技术与应用, 2009, 24(1): 63-67.

[37] 岳佳佳, 庞博, 张艳君, 等. 基于神经网络的宽浅型湖泊水质反演研究[J]. 南水北调与水利科技, 2016, 14(2): 26-31.

[38] 杨柳, 韩瑜, 汪祖茂, 等. 基于BP神经网络的温榆河水质参数反演模型研究[J]. 水资源与水工程学报, 2013, 24(6): 25-28.

[39] 吴倩, 林蕾, 王学军, 等. 福海叶绿素含量的人工神经网络反演模型[J]. 地理与地理信息科学, 2004(4): 27-30.

[40] LIN S S, SHEN S L, ZHOU A, et al. Sustainable development and environmental restoration in Lake Erhai, China[J]. Journal of Cleaner Production, 2020, 258: 120758.

[41] WANG X, DENG Y, TUO Y, et al. Study on the temporal and spatial distribution of chlorophyll a in Erhai Lake based on multispectral data from environmental satellites[J]. Ecological Informatics, 2021, 61: 101201.

[42] TAN W, LIU P, LIU Y, et al. A 30-year assessment of phytoplankton blooms in Erhai Lake using landsat Imagery: 1987 to 2016[J]. Remote Sensing, 2017, 9(12): 1265.

[43] 徐祎凡, 李云梅, 王桥, 等. 基于环境一号卫星多光谱影像数据的三湖一库富营养化状态评价[J]. 环境科学学报, 2011, 31(1): 81-93.

[44] 朱利, 李云梅, 赵少华, 等. 基于GF-1号卫星WFV数据的太湖水质遥感监测[J]. 国土资源遥感, 2015, 27(1): 113-120.

(责任编辑: 郑晓梅)

Mass concentration inversion for chlorophyll a in Erhai lake based on Sentinel-2

XIE Enhong[1,2], WU Junen[1,2], YANG Kun[1,2,*]

1. Faculty of Geography, Yunnan Normal University, Kunming 650500, China; 2. Engineering Research Center of GIS Technology in Western China, Ministry of Education, Kunming 650500, China

*Corresponding author, E-mail: kmdcynu@163.com

Abstract　In order to dynamically monitor eutrophic pollutants in Erhai lake, the remote sensing technology was used to invert the chlorophyll-a mass concentration, the core parameter reflecting eutrophication of water. The inversion model suitable for the local season was established to conduct the macro monitoring of the mass concentration of chlorophyll-a in water. Based on Sentinel-2 images and measured mass concentration data of chlorophyll a in Erhai lake in autumn, the inversion bands were selected by parameter correlation analysis method, then BP neural network model and multiple linear regression model were established. Seven sample points were randomly selected to cross-verify the two models, and then the mass concentration of chlorophyll a in Erhai lake was inverted. The results showed that a significant correlation occurred between Sentinel - 2 data and the mass concentration of chlorophyll a (the absolute value of Pearson's product moment correlation coefficient was higher than 0.7, $P < 0.001$), and the bands or band combinations with the largest correlation coefficient in single band, single band ratio and dual band ratio were B6, B7 / B6 and (B6 + B8) / (B7 + B8a), respectively; the three-layer BP neural network model with four neuron nodes in hidden layer had the smallest root mean square error and the largest determination coefficient, which were 0.002 8 and 0.925, respectively. On October 12th and November 9th, 2019, the spatial distribution of mass concentration of chlorophyll a in the northern part of Erhai lake was higher than that in the southern part. The mean absolute percentage error of BP neural network model was 21.36%, the root mean square error was 0.002 8, and the coefficient of determination was 0.925. The mean absolute percentage error of multiple linear regression model was 27.90%, the root mean square error was 0.004 5, and the coefficient of determination was 0.788. In general, the inversion accuracy of mass concentration of chlorophyll-a by BP neural network model was higher than that by multiple linear regression model. The results of this study can provide a reference for relevant departments to dynamically monitor water quality of Erhai lake and formulate water quality protection measures of Erhai lake.

Keywords　chlorophyll inversion; Sentinel-2; Erhai lake; BP neural network