

DOI: 10.20040/j.cnki.1000-7709.2023.20230103

# 基于 M-GA-BP 模型的乌梁素海悬浮物浓度反演研究

吴陈昊, 付学良, 李宏慧, 扈 华

(内蒙古农业大学计算机与信息工程学院, 内蒙古 呼和浩特 010018)

**摘要:** 为了解乌梁素海水质状况, 提出一种基于 M-GA-BP 的乌梁素海悬浮物浓度反演方法, 该方法以 Sentinel-2 遥感卫星影像为数据源, 考虑研究区域存在的时空特性, 提出将月份数据作为悬浮物浓度反演的一个特征, 利用遗传算法(GA)优化 BP 神经网络的权值与阈值, 建立 GA-BP 神经网络模型, 并与传统 BP 神经网络模型进行比较。结果表明, 引入月份特征模型有效降低了模型复杂度, 提高了模型反演精度, 其中, M-GA-BP 模型反演精度最高, 训练集、测试集决定系数分别为 0.916、0.903, 训练集、测试集均方根误差分别为 0.049、0.057  $\mu\text{g/L}$ , 研究结果为乌梁素海悬浮物浓度反演提供了新思路。

**关键词:** 月份特征; 遗传算法; BP 神经网络; 悬浮物; Sentinel-2; 乌梁素海

**中图分类号:** P333; TP79; TP391.9

**文献标志码:** A

**文章编号:** 1000-7709(2023)12-0049-04

## 1 引言

悬浮物是衡量湖泊湿地水质状况的重要指标, 其浓度不仅会影响水下光场的分布, 还直接影响到水体初级生产力和水域生态环境<sup>[1]</sup>。但随着社会的发展, 湖泊湿地水质问题逐渐显现, 严重影响了区域生态环境。因此, 开展湖泊湿地悬浮物浓度监测尤为重要。遥感技术因其具有快速、方便、成本低等优势, 已成为湖泊湿地环境监测的重要手段<sup>[2]</sup>。在研究湖泊湿地悬浮物的过程中, 发现悬浮物浓度在时间和空间上的分布具有一定的特征性, 这种时空特征不具有固定模式, 但会因研究区域的时间及地理位置的变化影响着湖泊悬浮物的浓度<sup>[3]</sup>。传统的遥感反演方法主要分为经验法、半分析法和分析法。但经验法<sup>[4]</sup>适应性通常较差, 半分析法和分析法<sup>[5]</sup>模型相对复杂, 不适合广泛应用。机器学习算法已逐渐应用于遥感反演中, 而神经网络作为能适应十分复杂的非线性关系的机器学习方法之一<sup>[6]</sup>, 也因其存在的优势在悬浮物浓度遥感反演中具有较强的适用性。乌梁素海位于内蒙古巴彦淖尔市乌拉特前旗境内, 湖泊面积为 325.31  $\text{km}^2$ , 其中, 仅 123.11  $\text{km}^2$  为开

阔水域, 剩余为芦苇区域, 库容量约为  $2.5 \times 10^8 \sim 3.0 \times 10^8 \text{m}^3$ ; 芦苇面积大, 故分析悬浮物浓度对湖泊的影响就具有十分重要的意义。鉴此, 本文以乌梁素海为例, 结合 Sentinel-2 遥感卫星和实测数据, 将月份数据与模型相结合, 提出一种 M-GA-BP 神经网络模型, 并利用该模型对乌梁素海悬浮物浓度进行了反演, 验证了模型的性能。

## 2 材料与方法

### 2.1 实测数据获取

实测数据由内蒙古农业大学国家级研究团队“河湖湿地水环境保护与修复技术研究团队”所提供, 研究区域采样点见图 1。乌梁素海属于寒旱区湖泊, 湖体大约每年 11 月上旬开始结冰, 冻融

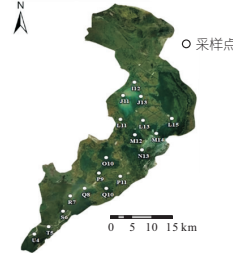


图 1 研究区域采样点

Fig. 1 Study area sampling sites

**收稿日期:** 2023-01-30, **修回日期:** 2023-04-17

**基金项目:** 国家自然科学基金项目(62041211, 61962047); 国家重点研发计划(2019YFC049205); 内蒙古自然科学基金(2020MS06011, 2019MS06015); 内蒙古高校科技重点项目(NJZZ23044)

**作者简介:** 吴陈昊(1998-), 男, 硕士研究生, 研究方向为智能计算与数据挖掘, E-mail: 3206319322@qq.com

**通讯作者:** 付学良(1966-), 男, 博士、教授、博导, 研究方向为智能计算与数据挖掘, E-mail: fuxl@imau.edu.cn

期可长达 5 个月,因此采样时间主要集中在 6~9 月,并固定为每月中下旬进行采样,采样深度为垂直向下距水面 0.5 m 处。悬浮物浓度实测数据于 2015~2018 年共采集 92 组数据,并随机分为训练集 64 组、测试集 28 组。

### 2.2 遥感数据获取与处理

Sentinel-2 卫星由欧洲航天局(ESA)研发,可用于监测陆地生态、内河和沿海区域环境等遥感影像资料。其所搭载一枚多光谱成像仪(MSI),具有 13 个光谱波段,并具有 10、20、60 m 三种空间分辨率。具体卫星数据见表 1。影像数据可从欧洲航天局官网(<https://scihub.copernicus.eu/dhus/#/home>)下载。

表 1 Sentinel-2 传感器光谱特征信息

Tab. 1 Sentinel-2 sensor spectral characteristic information

波段	Sentinel-2A			波段	Sentinel-2A		
	中心波长/nm	带宽/nm	分辨率/m		中心波长/nm	带宽/nm	分辨率/m
B1	433.9	27	60	B8	835.1	145	10
B2	496.6	98	10	B8A	864.8	33	20
B3	560.0	45	10	B9	945.0	26	60
B4	664.5	38	10	B10	1 373.5	75	60
B5	703.9	19	20	B11	1 613.7	143	20
B6	740.2	18	20	B12	2 202.4	242	20
B7	782.5	28	20				

通过欧洲航天局(ESA)下载与实测悬浮物浓度数据日期同步或相差一天的准同步 Sentinel-2L1C 级遥感影像数据,并在下载前设置筛选条件为云量小于 20%,以保证所获取乌梁素海遥感影像的清晰;遥感影像之间拍摄间隔尽量大,以保证试验结果更具代表性。Sentinel-2 遥感影像数据预处理包括裁剪、重采样、几何校正及波段反射率提取等,以获得各波段遥感反射率。数据进行大气校正时,利用 ESA 提供的 Sen2Cor 插件获取 Sentinel-2L2A 级影像数据。数据进行重采样时,利用 SNAP 软件中 S2 Resampling Processor,使用最邻近像元法,设置分辨率为 20 m。

### 2.3 悬浮物浓度反演模型

#### 2.3.1 BP 神经网络模型

BP 神经网络(BPNN)<sup>[7]</sup>包含输入层、隐含层、输出层三个部分,是一种信号前向传递、误差反向传播的多层前馈网络。

在 BP 神经网络中,信号前向传播和误差反向传播是两个过程。正向传播时,信号从输入层传入,经隐含层处理后至输出层输出,当得到的输出不满足所需的期望时,则进行反向传播。误差从输出层按原来的路线反向传播至输入层,并对各层节点权值和阈值进行调整,随之,再开始下一

步的正向传播,不断调整使误差达到最小。BP 神经网络具体步骤如下。

**步骤 1** 初始化参数,即隐含层权值  $\omega_{ij}$ 、阈值  $\theta$  及输出层权值  $v_{jk}$ 、阈值  $\lambda$ 。

**步骤 2** 计算隐含层节点输出,具体公式为:

$$q_j = f\left(\sum_{i=1}^n \omega_{ij}x_i - \theta_j\right) \quad (1)$$

式中,  $q_j$  为第  $j$  个隐含层节点输出值;  $n$  为输入层节点数;  $f$  为隐含层激活函数;  $x_i$  为第  $i$  个输入信号。

**步骤 3** 计算输出层节点输出,具体公式为:

$$Z_k = f\left(\sum_{j=1}^m v_{jk}q_j - \lambda_k\right) \quad (2)$$

式中,  $m$  为隐含层节点数;  $f$  为输出层激活函数。

**步骤 4** 权值更新,具体公式为:

$$\omega_{ij}(t+1) =$$

$$\omega_{ij}(t) + \eta q_j(1 - q_j)x_i \sum_{k=1}^h v_{jk}(y_k - Z_k) \quad (3)$$

$$v_{jk}(t+1) = v_{jk}(t) + \eta q_j(y_k - Z_k) \quad (4)$$

式中,  $\eta$  为学习率,取值在  $[0, 1]$  之间。

**步骤 5** 阈值更新,具体公式为:

$$\theta_j(t+1) =$$

$$\theta_j(t) + \eta q_j(1 - q_j)x_i \sum_{k=1}^h v_{jk}(y_k - Z_k) \quad (5)$$

$$\lambda_k(t+1) = \lambda_k(t) + (y_k - Z_k) \quad (6)$$

**步骤 6** 判断是否迭代结束,若未结束,则执行步骤 2。

#### 2.3.2 基于月份特征下遗传算法优化 BP 神经网络

图 2 为 2015~2018 年乌梁素海悬浮物浓度随月份的变化。由图 2 可看出,悬浮物浓度在 6 月最小并开始逐渐增至 7 月浓度最高,随之逐渐降低,乌梁素海悬浮物浓度在不同时间差异较大。因此,考虑乌梁素海悬浮物浓度存在的时空特性,将月份作为模型的特征进行输入,使遥感数据和实测数据可在模型中按月份分组进行反演。

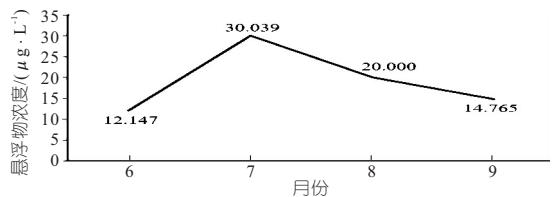


图 2 2015~2018 年不同月份平均悬浮物浓度  
Fig. 2 Average suspended matter concentrations by month, 2015-2018

BP 神经网络虽然拥有很强的非线性映射能力,却很容易陷入局部最优,反演效果较差。遗传算法(GA)<sup>[8]</sup>可从不同方向随机生成起始点并在解空间中通过适应度函数的引导进行自适应搜

索,且具有较强的全局搜索能力。因此,利用 GA 的优势,对 BP 神经网络进行优化,构建了基于月份特征的 GA-BP 模型(M-GA-BP 模型)。

M-GA-BP 具体步骤如下。

**步骤 1** 确定网络拓扑结构。将遥感波段数据(B1, B2, ..., B12)与月份特征数据(6、7、8、9 月)相结合作为模型输入层,将悬浮物浓度作为模型输出层。

**步骤 2** 初始化种群。设置种群数量  $M$ , 初始化网络中的权值、阈值,并将其编码为一组染色体。

**步骤 3** 定义适应度函数。将模型的悬浮物浓度反演值和悬浮物浓度实测值的绝对误差之和的倒数作为适应度函数  $F$ ,  $F$  越大,适应度越高,即:

$$F = 1 / \sum_{i=1}^n (y_i - \hat{y}_i)^2 \quad (7)$$

式中,  $y_i$ 、 $\hat{y}_i$  分别为悬浮物浓度实测值、反演值。

**步骤 4** 选择、交叉、变异操作。利用轮盘赌选择法确定个体选择概率,并用单点交叉和均匀变异法进行交叉、变异操作。

**步骤 5** 计算个体适应度。判断是否满足停止条件,若满足,则执行步骤 6;否则,执行步骤 4。

**步骤 6** 解码。将 GA 输出的权值、阈值作为 BP 神经网络的初始权值、阈值,并计算误差,当满足条件后,输出最终反演结果。

### 2.3.3 反演评价指标

为验证模型反演效果,选用决定系数( $R^2$ )和均方根误差( $R_{RMSE}$ )对反演结果进行评价,其中  $R^2$  越趋近于 1,效果越好; $R_{RMSE}$  越小,效果越好。

## 3 基于 M-GA-BP 模型的乌梁素海悬浮物浓度反演

### 3.1 模型参数设置

充分考虑研究区域所存在的时空特性,提出一种基于月份特征的 GA-BP 模型,并设置 BP 网络结构为 16 个输入层、8 个隐含层和 1 个输出层,迭代次数为 50。GA 算法具体参数设置见表 2。

表 2 GA 参数设置

Tab. 2 GA parameter setting

相关参数	取值(概率)	相关参数	取值(概率)
进化次数	50	交叉概率	0.02
种群数量	300	变异概率	0.02
选择概率	0.02		

### 3.2 BP 神经网络模型与 GA-BP 模型反演结果对比

表 3 为 BP 神经网络模型与 GA-BP 模型反演结果对比。

表 3 BP 与 GA-BP 反演结果对比

Tab. 3 Comparison of BP and GA-BP inversion results

模型	训练		测试	
	$R^2$	$R_{RMSE}/(\mu\text{g} \cdot \text{L}^{-1})$	$R^2$	$R_{RMSE}/(\mu\text{g} \cdot \text{L}^{-1})$
BP	0.596	0.112	0.534	0.125
GA-BP	0.876	0.064	0.850	0.071

由表 3 可知,BP 神经网络模型反演效果不太理想,而经 GA 优化后的 BP 神经网络模型,训练集和测试集决定系数分别达 0.876、0.850,训练集和测试集均方根误差分别达 0.064、0.071  $\mu\text{g}/\text{L}$ ,具有较好的反演效果。由于 BP 神经网络模型主要是利用误差函数进行单调上升或单调下降实现对网络中权值和阈值的更新,而乌梁素海悬浮物浓度反演是一个非线性问题,存在多个极值,当 BP 神经网络模型遇到局部最优时,将停止对权值和阈值的调整,而 GA 算法具有很强的全局搜索能力,可使种群中个体不在单一化方向发展。因此,GA-BP 模型可较好地反演乌梁素海悬浮物浓度。

### 3.3 基于月份特征下不同模型反演结果对比

表 4 为引入月份特征下不同模型反演结果对比。由表 4 可知,引入月份特征后,BP 神经网络模型精度与 GA-BP 模型精度均有所提升,其中 M-GA-BP 反演效果最好,训练集和测试集决定系数分别为 0.916、0.903,训练集和测试集均方根误差分别为 0.049、0.057  $\mu\text{g}/\text{L}$ 。由于乌梁素海不同月份间悬浮物浓度存在固有差异,且乌梁素海是内蒙古自治区河套灌区灌排水系统极其重要的组成部分,每到春灌秋浇之际,大量农田退水将通过排灌渠流入乌梁素海,使乌梁素海悬浮物浓度在不同月份差异加大,而将数据按月份分组进行反演,可消除部分乌梁素海因外界因素和内在原因形成的时空特性对模型反演性能的影响。因此,引入月份特征后,可进一步提高模型反演的准确度和稳定性,在乌梁素海悬浮物浓度反演中具有较大的优势和应用潜力。

表 4 引入月份特征下不同模型反演结果对比

Tab. 4 Comparison of inversion results of different models with introduction of month feature

模型	训练		测试	
	$R^2$	$R_{RMSE}/(\mu\text{g} \cdot \text{L}^{-1})$	$R^2$	$R_{RMSE}/(\mu\text{g} \cdot \text{L}^{-1})$
M-BP	0.724	0.096	0.652	0.108
M-GA-BP	0.916	0.049	0.903	0.057

为了更直观地反映引入月份特征模型的反演效果,绘制了 4 种模型的实测值与反演值对比图见图 3。图 3 中斜线为 1:1 线。由图 3 可看出,引入月份特征后,实测值与反演值间的差值相对小,更加接近 1:1 线。

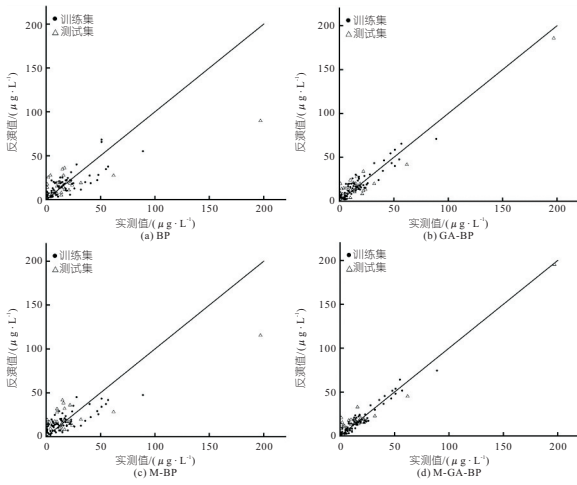


图3 不同模型实测值与反演值对比

Fig. 3 Comparison of actual measurements and inversions of different models

### 4 结论

a. 与传统BP神经网络模型相比,GA-BP模型的决定系数与均方根误差均有所提升,具有更好的效果,不易陷入局部最优,GA-BP模型可很好地反演乌梁素海湖泊水体中悬浮物的浓度。

b. 将月份作为特征之一进行反演时,可有效消除部分时空特性的影响,提高模型反演效果,引入月份特征建立的模型更符合研究区域的实际情况,有助于治理区域水环境系统和生态平衡。

### 参考文献:

[1] 曹引,冷运涛,赵红莉,等. 南四湖水体实测高光谱与悬浮物浓度及浊度关系分析[J]. 水电能源科学, 2016,34(1):40-44.

[2] 黄李童,陈江,朱渭宁,等. 基于 Landsat-8 的城市湖泊水体总悬浮物吸收系数的遥感反演——以杭州西湖为例[J]. 环境科学学报, 2018, 38(10): 4073-4082.

[3] 吴晓东,王露,晁建颖,等. 太湖流域漏湖水体悬浮物分布特征及其影响因素[J]. 湖泊科学, 2020, 32(6):1848-1857.

[4] 赵兰兰,袁杰,李磊,等. 地貌瞬时单位线在滦河中小流域的应用[J]. 水电能源科学, 2020, 38(5): 31-34.

[5] CHU Q, ZHANG Y, MA R, et al. MODIS-based remote estimation of absorption coefficients of an inland turbid lake in China[J]. Remote sensing, 2020,12(12):1940.

[6] 王麟,赵涛. 基于BP神经网络的人工加糙明渠糙率预测模型[J]. 水电能源科学, 2021, 39(9): 125-127, 124.

[7] 翟来铮,高宇龙,李清富,等. 基于GA-BP模型的卵石混凝土抗折强度预测[J]. 人民黄河, 2022, 44(11):137-141.

[8] HOLLAND J H. Adaptation in natural and artificial systems; an introductory analysis with application to biology, control and artificial intelligence [M]. Michigan: University of Michigan Press, 1975.

## Study on Inversion of Suspended Matter in Wuliangsu Lake Based on M-GA-BP

WU Chen-hao, FU Xue-liang, LI Hong-hui, HU Hua

(College of Computer and Information Engineering, Inner Mongolia Agricultural University, Hohhot 010018, China)

**Abstract:** In order to understand the water quality of Wuliangsu Lake, an inversion method of total suspended matter concentration based on M-GA-BP was proposed. Using Sentinel-2 remote sensing satellite images as the data source and considering the spatial and temporal characteristics existing in the study area, the monthly data was considered as a feature for the inversion of TSM concentration. The GA-BP model was built by optimizing the weights and thresholds of the BP neural network using genetic algorithm (GA), and comparing with the traditional BP neural network model. The results show that the introduction of the monthly feature model effectively reduces the model complexity and improves the model inversion accuracy, among which the M-GA-BP model has the highest inversion accuracy with the coefficients of determination of 0.916 and 0.903 for the training and test sets, respectively, and the root mean square errors of 0.049  $\mu\text{g/L}$  and 0.057  $\mu\text{g/L}$  for the training and test sets, respectively. The study can provide a new idea for the inversion of TSM concentration in the Wuliangsu Lake.

**Key words:** month feature; genetic algorithm; BP neural network; suspended solid; Sentinel-2; Wuliangsu Lake