

# 基于 ISSA-SVR 算法的近海典型水质参数遥感反演

刘元杰<sup>1</sup>, 崔建勇<sup>1</sup>, 董文<sup>2</sup>, 万剑华<sup>1</sup>, 张杰<sup>1</sup>

(1 中国石油大学(华东)海洋与空间信息学院 青岛 266580;

2 中国科学院空天信息创新研究院 北京 100094)

**摘要:** 化学需氧量 (COD) 和叶绿素 a (Chl-a) 浓度作为与光谱相关的典型水质参数, 是反映水体污染程度和富营养化程度的重要指标。支持向量回归模型 (Support Vector Regression, SVR) 适用于小样本, 广泛用于近海典型水质参数的遥感反演, 但也存在模型参数选择困难、容易陷入局部最优解的问题。针对这一问题, 本文构建融合反向学习和模拟退火的改进麻雀算法 (Improved Sparrow Search Algorithm, ISSA), 通过改进麻雀算法对 SVR 模型的惩罚系数和核函数参数进行参数寻优, 提出了一种改进的支持向量回归模型 (ISSA-SVR)。通过该模型利用实测水面光谱与水质参数数据建立 COD 和 Chl-a 浓度反演模型。利用 Sentinel-2 卫星遥感光谱数据对模型的精度进行验证, 得到各水质参数浓度的反演精度。采用 ISSA 算法优化 SVR 建立的 COD 浓度预测模型和 Chl-a 浓度预测模型的平均相对误差 (MRE) 分别为 20.02% 和 30.17%。反演结果均优于其他模型 (线性回归、SVR 和 SSA-SVR 模型)。实验结果证实, ISSA-SVR 算法是实现 COD 浓度和 Chl-a 浓度遥感反演的有效方法, 可为我国近海典型水质参数遥感反演及后续水体科学管理提供参考。

**关键词:** COD; Chl-a; SVR; 麻雀搜索算法

中图分类号: P237; TP75 文献标志码: A 文章编号: 2095-1000(2024)03-0081-10

DOI: 10.12347/j.ycyk.20240120001

引用格式: 刘元杰, 崔建勇, 董文, 等. 基于 ISSA-SVR 算法的近海典型水质参数遥感反演[J]. 遥测遥控, 2024, 45(3): 81-90.

## Remote Sensing Retrieval of Coastal Water Quality Parameters Based on ISSA-SVR Method

LIU Yuanjie<sup>1</sup>, CUI Jianyong<sup>1</sup>, DONG Wen<sup>2</sup>, WAN Jianhua<sup>1</sup>, ZHANG Jie<sup>1</sup>

(1. College of Oceanography and Space Informatics, China University of Petroleum, Qingdao 266580, China;

2. Academy of Aerospace Information Innovation, Chinese Academy of Sciences, Beijing 100094, China)

**Abstract:** The chemical oxygen demand (COD) and chlorophyll a concentration, which are typical water quality parameters related to the spectrum, serve as important indicators for reflecting the degree of water pollution and eutrophication. Support Vector Regression (SVR) is suitable for small sample sizes and widely utilized in remote sensing retrieval of typical offshore water quality parameters; however, it faces challenges in model parameter selection and may easily fall into local optimal solutions. To address this issue, an Improved Sparrow Search Algorithm (ISSA) is developed by integrating reverse learning and simulated annealing. An enhanced support vector regression model (ISA-SVR) is proposed by refining the Sparrow algorithm to optimize the penalty coefficient and kernel parameters of the SVR model. Inversion models for COD and Chl-a concentrations are established using measured water spectra and data on water quality parameters. The accuracy of the model is validated using Sentinel-2 satellite remote sensing spectral data, yielding inversion accuracies for each water quality parameter concentration. The mean relative error (MRE) of the COD concentration prediction model and Chl-a concentration prediction model based on ISSA algorithm optimized SVR are 20.02% and 30.17%, respectively, outperforming other models such as linear regression, SVR, and SSA-SVR models. Experimental results demonstrate that ISA-SVR algorithm represents an effective approach for remotely sensed retrieval of COD and Chl-a concentrations while offering valuable insights for subsequent scientific management of offshore water quality.

**Keywords:** COD; Chl-a; SVR; Sparrow search algorithm

**Citation:** LIU Yuanjie, CUI Jianyong, DONG Wen, et al. Remote Sensing Retrieval of Coastal Water Quality Parameters Based on ISSA-SVR Method[J]. Journal of Telemetry, Tracking and Command, 2024, 45(3): 81–90.

## 0 引言

随着经济的快速发展, 水环境也持续恶化, 受到严重污染的近海水体威胁着中国的生态环境, 制约着我国经济的可持续发展。水质监测的全面精确是整治水体污染和确保水域生态系统健康的先决条件。有机物是海洋生态系统的重要组成部分之一, 它与重金属、有毒有机化合物等污染物汇合释放, 加剧了水体富营养化程度和污染程度<sup>[1,2]</sup>。化学需氧量 (Chemical Oxygen Demand, COD) 是用来表示水中有机物含量的重要指标, 叶绿素 a (Chlorophyll-a, Chl-a) 是表征水体富营养化的典型参数。确定上述两个有机物相关水质参数浓度对于治理水污染和环境保护具有重要意义, 近年来受到广泛的关注<sup>[3-6]</sup>。

水质参数遥感监测是通过水体水质参数浓度和光谱特征的对应关系, 构建反演算法。传统水质监测方法主要是在监测区域实地采样, 将水样带回实验室测量浓度来预估监测区的水质情况, 难以准确反映污染物大范围的空间分布和水质参数浓度变化情况。而遥感水质监测时效性高、监测范围广, 可进行周期性动态监测<sup>[7,8]</sup>。随着遥感技术的不断发展, 遥感技术在水质监测中得到更为广泛的应用。

目前, 国内外许多学者采用机器学习方法研究利用水体光谱反射率得到水体水质参数浓度, 然而部分水质参数光谱响应低, 机器学习算法参数选择困难, 往往要大量重复实验才能确定用于建模反演精度高的参数。SVR 适用于小样本, 具有泛化能力强, 预测精度高的特点, 已经广泛应用于多种水质参数的反演过程。Chegoonian A M 利用加拿大布法罗庞德湖的 Chl-a 浓度数据建立 SVR 模型, 通过周边其他六个湖泊的水质数据验证得到 SVR 模型精度优于传统的半经验模型<sup>[9]</sup>。Ye 等人利用遗传算法优化 SVR 模型参数, 建立悬浮物浓度反演模型, 通过 GF-1 影像对龙泉湖悬浮物浓度进行反演<sup>[10]</sup>。Tan 等利用实际水质测量数据和 Sentinel-2 遥感影像, 建立对总氮浓度和总磷浓度的 SVR 反演模型反演岷江水水质参数浓度<sup>[11]</sup>。Valera 等应用 SVR 建立适用于小型沿海海湾中的近岸溶

解氧浓度反演模型<sup>[12]</sup>。刘梅等基于高光谱无人机影像, 建立达到养殖池塘预测精度要求的总悬浮物、总氮、总磷和 COD 的 SVM 反演模型, 快速反演养殖小区水质空间分布图, 实现养殖池塘及尾水处理池水水质信息的快速准确获取<sup>[13]</sup>。

选择合适的参数是建立 SVR 模型的首要任务。国内外许多学者使用各种优化算法, 例如遗传算法<sup>[14,15]</sup>、蛙跳算法<sup>[16]</sup>、飞蛾优化算法<sup>[17]</sup>、粒子群优化算法<sup>[18,19]</sup>、萤火虫算法<sup>[20]</sup>、灰狼优化算法<sup>[21]</sup>等, 选择支持向量回归模型的惩罚系数和核函数参数的最优解, 对支持向量回归模型进行优化, 以解决支持向量回归模型的参数选择问题, 进一步实现水质参数的高精度反演。

麻雀搜索算法 (Sparrow Search Algorithm, SSA) 于 2020 年由 Xue 等<sup>[22]</sup>首次提出, 是一种模拟麻雀觅食和反捕食行为的新型智能优化算法。麻雀搜索算法与其他算法相比, 具备快速收敛和高求解精度的特点, 但随着算法迭代的进行, 可能会面临种群多样性降低的问题。这是因为搜索速度过快可能导致算法陷入局部极值。为克服传统 SSA 算法的不足, 基于双混沌映射系统初始化麻雀种群, 丰富种群多样性。同时, 通过反向学习和模拟退火的扰动变异能力使算法在迭代后期更容易跳出局部最优解。

本文针对 SVR 反演水质参数时的模型参数选取困难以及 SSA 算法参数寻优易陷入局部最优解的问题, 根据双混沌映射系统具备更佳的混沌特性以及反向学习和模拟退火的扰动变异能力, 构建融合反向学习和模拟退火的改进麻雀算法 (Improved Sparrow Search Algorithm, ISSA), 通过 ISSA 算法对 SVR 模型的惩罚系数和核函数参数进行参数寻优, 提高支持向量机的预测精度, 建立 ISSA-SVR 预测模型。利用实测水面光谱与水质参数数据构建反演 COD 和 Chl-a 浓度的 ISSA-SVR 模型, 并通过 Sentinel-2 卫星数据验证模型精度。

## 1 研究方法

SVR 是一种建立在 SVM 基础上的函数回归算法, 采用非线性单元隐藏层的前馈网络结构<sup>[23]</sup>。在水质参数反演方面, SVR 的特性可在一定程度上

规避模型的过度拟合,具有良好的预测精度。然而,SVR模型存在参数选择难的问题,这直接影响了模型的学习和泛化能力。

### 1.1 麻雀搜索算法

麻雀搜索算法是模拟麻雀觅食和反捕食行为的新型智能搜索算法。麻雀种群内共有三种行为类型:发现者、加入者、预警者。发现者寻找食物,加入者跟随发现者寻找食物,预警者发出警报避免危险。其中,发现者和加入者可以相互转换,且预警者在种群中随机产生。

假设整个麻雀种群为 $N$ ,则第 $i$ 只麻雀在 $D$ 维空间中的位置 $X_i = [x_{i1}, x_{i2}, \dots, x_{iD}]$ ,  $i = 1, 2, \dots, N$ 。一般情况下,发现者占总数的10%~20%,每次迭代中发现者的位置更新公式如下:

$$X_i^{t+1} = \begin{cases} X_i^t \cdot \exp\left(\frac{i}{\alpha \cdot T}\right), & R_2 < ST \\ X_i^t + Q \cdot L, & R_2 \geq ST \end{cases} \quad (1)$$

式中, $\alpha \in (0,1)$ 的随机数; $T$ 为最大迭代次数; $t$ 是当前迭代次数; $X_i^t$ 为第 $i$ 只麻雀在第 $t$ 次迭代时的位置; $Q$ 是服从正态分布的随机数; $L$ 为元素均为1的 $1 \times D$ 维矩阵, $ST$ 和 $R_2$ 分别为安全值和预警值。当 $R_2 < ST$ 时,表示环境安全,发现者可进行大范围搜索; $R_2 \geq ST$ 时,表示种群中部分个体发现了环境危险并进行了预警;此时,所有个体飞到安全区域进行搜索。

除了发现者之外,所有留在种群中的都是加入者。更新公式如下:

$$X_i^{t+1} = \begin{cases} Q \cdot \exp\left(\frac{X_{\text{worst}} - X_i^t}{i^2}\right), & i > \frac{n}{2} \\ X_p^{t+1} + |X_i^t - X_p^{t+1}| \cdot A^+ \cdot L, & i \leq \frac{n}{2} \end{cases} \quad (2)$$

式中 $X_{\text{worst}}$ 是当前全局最差的位置; $X_p$ 是发现者当前的最佳位置; $A^+ = A^T(AA^T)^{-1}$ ,其中 $A$ 是一个 $1 \times D$ 维矩阵,每个元素随机分配值1或-1。当 $i > \frac{n}{2}$ 时,适应度较低的第 $i$ 个个体的搜索位置较差,需要飞到其他地方进行搜索。

麻雀种群中的预警者一般占种群总数的10%~20%,预警者的初始位置是在种群形成的初始阶段随机产生的。位置更新公式如下:

$$X_i^{t+1} = \begin{cases} X_{\text{best}}^t + \beta \cdot |X_i^t - X_{\text{best}}^t|, & f_i \neq f_g \\ X_i^t + K \cdot \left(\frac{|X_i^t - X_{\text{worst}}^t|}{(f_i - f_w) + \varepsilon}\right), & f_i = f_g \end{cases} \quad (3)$$

式中, $X_{\text{best}}$ 为当前全局最优位置; $\beta$ 是服从正态分布的随机数,均值为0,方差为1,表示步长控制参数; $K \in [-1, 1]$ 的随机数; $f_i$ 为第 $i$ 个个体的适应度值; $f_g$ 为当前全局最优适应度值; $f_w$ 是目前全局最差的适应度值; $\varepsilon$ 是最小的常数。当 $f_i \neq f_g$ 时,表示第 $i$ 个个体处于种群边缘,容易受到攻击;当 $f_i = f_g$ 时,意味着处于群体中心的个体意识到危险,需要接近其他个体以避免危险。

虽然麻雀搜索算法具有收敛速度快、稳定性高、调整参数少、计算简单等优点,但与其他智能搜索算法一样,多次迭代后也会有局部最优解的出现。为此,本文对麻雀搜索算法进行了改进。

### 1.2 双混沌映射系统初始化种群

混沌映射通常被应用于智能优化算法的种群初始化。然而现有的部分混沌映射会出现混沌范围不连续的问题。同时由于混沌映射结构简单,参数的微小变化导致映射非混沌。因而本文基于Cao等人的 $n$ 维混沌映射模型<sup>[24]</sup>,利用Logisti混沌映射和Fraction混沌映射构建二维混沌映射系统,将两个混沌映射作为种子,代数运算后执行正弦变换生成新的混沌映射结构,并将其应用于麻雀搜索算法的种群初始化。二维混沌映射系统公式如下:

$$\begin{cases} x_{i+1} = \sin\left(\pi\left(4p_1x_i(1-x_i) + \frac{1}{y_i^2+0.1} - p_2y_i\right)\right) \\ y_{i+1} = \sin\left(\pi\left(4p_1y_i(1-y_i) + \frac{1}{x_i^2+0.1} - p_2x_i\right)\right) \end{cases} \quad (4)$$

式中, $p_1, p_2$ 为Logisti混沌映射和Fraction映射的控制参数, $p_1, p_2 \in [1, 100]$ 。

三种映射方式的分布图如图1所示。对比可发现,二维混沌映射值分布更为均匀,因此,利用二维混沌映射系统映射初始化种群可以有效丰富种群多样性。

### 1.3 融合反向学习和模拟退火策略

反向学习是Tizhoosh提出的新方法,通过反向学习机制,利用已知解求出对应的反向解,比较得到的两个解并保留更优的那个。将反向学习策略融入麻雀算法中。对发现者公式更新如下:

$$\begin{aligned} X_{\text{best}}^t &= ub + r \oplus (lb - X_{\text{best}}^t) \\ X_{i,j}^{t+1} &= X_{\text{best}}^t + b_1 \oplus (X_{\text{best}}^t - X_i^t) \end{aligned} \quad (5)$$

式中, $X_{\text{best}}^t$ 为第 $t$ 代最优解的反向解, $ub, lb$ 分别是上下界, $r$ 是服从 $(0,1)$ 标准均匀分布的 $1 \times D$ 维的



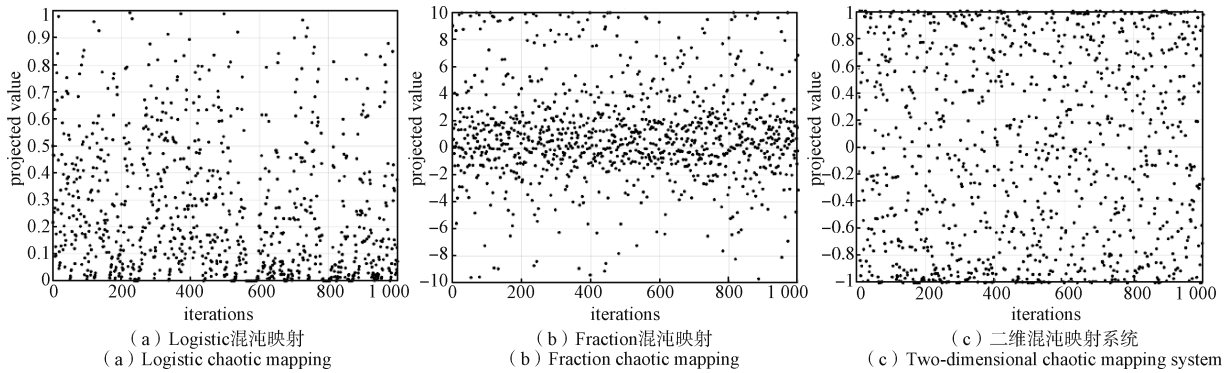


图1 混沌值分布情况

Fig.1 Distribution of mapping values

随机数矩阵,  $b_1$ 表示信息交换控制参数。

同时, 利用模拟退火算法在寻优过程中的突跳性提高麻雀搜索算法的全局寻优能力。退火温度  $T$ 与适应度和接受概率相关, 计算公式如下:

$$T = \frac{(f_{\max} - f_{\min})}{\ln p} = -\frac{|\Delta f|}{\ln p} \quad (6)$$

$f_{\max}$ 、 $f_{\min}$ 和 $\Delta f$ 为麻雀种群最大、最小适应度值及差值,  $p$ 表示退火系数, 取值为[0.8,1)。当进行全局寻优时,  $\Delta f > 0$ , 则根据公式判断:

$$\exp\left(\frac{-\Delta f}{T}\right) \geq \text{rand}(0, 1) \quad (7)$$

成立, 则接受新位置。

### 1.4 ISSA 算法流程

- ① 初始化参数, 包括种群数量、发现者和预警者的比例、预警值和安全值、初始退火温度等。
- ② 根据式(4)映射初始化种群。
- ③ 计算麻雀个体的适应度值并排序, 找出最优和最差适应度值的个体及其位置。
- ④ 根据式(1)进行发现者位置的更新。
- ⑤ 根据式(2)进行加入者位置的更新。
- ⑥ 根据原麻雀算法的预警者更新公式(3)进行预警者的位置更新。
- ⑦ 通过反向学习策略, 根据式(5)对当前最优解进行扰动, 产生新解。
- ⑧ 将新种群与原种群的适应度值进行比较, 进行个体位置的更新。
- ⑨ 由模拟退火机制, 根据式(7)进行全局最优位置的更新。
- ⑩ 若达到阈值, 则停止循环, 否则, 再次循环, 直到达到条件。
- ⑪ 算法执行结束, 输出并记录最优结果。

## 2 实验

### 2.1 研究区域与数据来源

研究区域位于中国山东省青岛市唐岛湾(120.18°E, 35.95°N)。该研究区域涵盖 1 637.6 公顷, 包括近岸沿海湿地以及人工湿地两个主要类别。该区域湿地类型多样, 生态系统保存较好。在科学和保护价值上具有重要意义。近年来, 黄岛区的快速发展对唐岛湾地区的环境状况产生了一定影响。因此, 对唐岛湾地区进行大规模的水质监测和保护工作势在必行。

#### 2.1.1 实测数据

研究人员在 2022 年 9 月 18 日和 2023 年 5 月 16 日分别采集了 20 个点, 总计 40 个点的光谱数据和水质参数数据。采样时间内, 天气相对晴朗, 阳光强烈。图 2 标注了两次实验的采样点位。

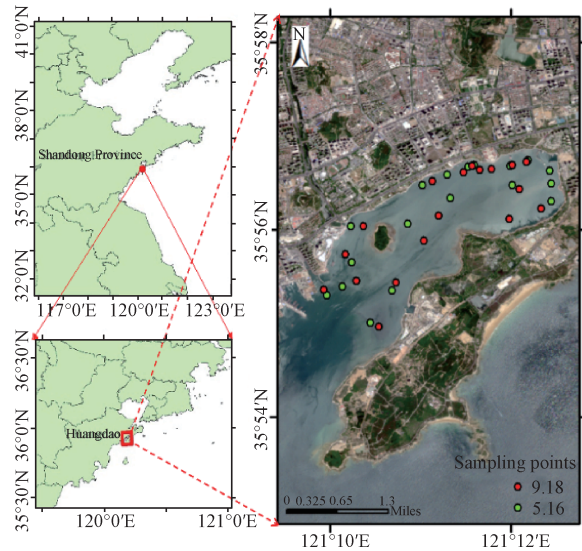


图2 采样点

Fig. 2 Sampling stations

在2022年9月18日的实验时间内,潮汐的范围为210 cm~328 cm。在2023年5月16日的实验时间内,潮汐的范围为181 cm~324 cm。这两个日期的潮汐值大致相等,可排除潮汐变化的影响。

采用AWRMMS(水面高光谱辐射移动测量系统)进行了水面光谱测量。在测量过程中,测量系统的伸缩臂被定位在与太阳入射面成135°角度的位置,这种设置使得三个传感器能够高效地捕捉太阳辐亮度、水面辐亮度和天空辐照度三个参数,并实时查看光谱轮廓。采用Mobley描述的方法,处理、计算得到离水辐亮度和遥感反射率数据。在每个采样点进行了15次光谱测量,计算测量数据的平均值为该点的遥感反射率。

在每个采样点进行了水质监测,以获取水质参数数据。使用EXO多参数水质分析仪在现场收集叶绿素a浓度。研究人员在每个采样点,进行了25次测量,并使用平均值确定采样点的叶绿素a浓度。此外,在每个采样点收集了水样,使用分光光度法在实验室中确定了COD浓度。

#### 2.1.2 遥感数据

Sentinel-2 有两颗卫星: Sentinel-2A 和 Sentinel-2B, 分别于2015年和2017年发射, 重复访问周期为5天。Sentinel-2 遥感卫星携带多光谱传感器(MSI)。MSI有13个波段, 分别为10 m处有4个可见光波段, 20 m处有6个近红外波段, 60 m处有3个波段<sup>[25]</sup>, 涵盖了从可见光到短波红外波长(0.44  $\mu\text{m}$ ~2.2 $\mu\text{m}$ )的电磁光谱。地面轨道带长约290 km。

本研究从欧洲航天局哥白尼开放获取中心下载了2022年9月18日和2023年5月16日各1景Sentinel-2B-MSI Level-1C数据。在2022年9月18日, 卫星过境时间是11:40, 采样时间从10:00到13:40。而在2023年5月16日, 卫星过境时间是10:00, 采样时间从9:00到12:30。在唐岛湾海域水质参数和水面实测光谱数据采集的时段内天气晴朗无云, 遥感图像与对应的实测数据一致性较好, 受天气、环境、观测条件变化影响较小。用软件SNAP4.0.2版对Sentinel-2数据进行辐射定标和大气校正的预处理<sup>[26]</sup>。C2RCC是一种基于人工神经网络(Artificial Neural Network, ANN)的大气校正算法, 最初由Doerffer R提出<sup>[27]</sup>, 该算法利用辐射传输模拟得到的大气校正后水体反射率与大气顶层辐射亮度数据集进行训练, 经过训练的神经

网络模型从总卫星辐亮度 $L_t(\lambda)$ 直接反演得到离水辐亮度 $L_w(\lambda)$ , 该方法适用于多种水体, 并且理论上被设计为适用于所有的水色卫星数据。本次研究采用SNAP工具箱v.8.0中的C2RCC工具进行大气校正<sup>[28,29]</sup>。根据现场采样点的经纬度, 用ArcGIS软件提取每个采样点对应的遥感反射率图像像素值。

#### 2.2 改进SVR的水质参数高光谱遥感反演模型

本研究以唐岛湾COD和Chl-a的遥感反演为例, 评估ISSA算法对性能的改进。首先, 提取用于建模的光谱数据。利用这些敏感波段, 建立ISSA-SVR反演模型。随后, 将这些模型应用于Sentinel-2卫星数据, 以远程估算海域中的COD和Chl-a浓度, 为唐岛湾水质保护提供数据支持。

##### 2.2.1 光谱数据与水质参数浓度的相关性分析

在实测水体反射光谱的过程中, 光谱数据的绝对数值可能受到太阳辐射强度及太阳高度角变化的影响以及光谱仪移动的影响。通过标准化, 可以在一定程度上消除环境因素的影响, 使得在不同时间和采样点采集的光谱数据具有可比性<sup>[30]</sup>。实测光谱遥感反射率如图3(a)所示, 归一化光谱遥感反射率如图3(b)所示。归一化后的光谱曲线比原光谱有更明显的反射率峰和波谷。将归一化后的光谱数据通过Sentinel-2的卫星波段响应函数转化为波段等效遥感反射率。

Person相关分析是从统计学的角度研究两个或多个随机变量之间相关性强度的一种方法。通过对光谱遥感反射率和水质参数浓度的相关性分析确定用于建模反演的光谱波段。已有研究表明, 波段组合可以有效地衰减其他环境因素的影响, 如水面平整度和微波<sup>[31]</sup>, 从而提高了模型的反演精度。因此, 在进行Person相关分析时, 以光谱波段比值、波段差值、波段乘积等波段组合方式输入。最终在多种波段组合中选择相关系数较大的几个作为最终反演COD和Chl-a的敏感因子。选择的敏感因子及其相关系数见表1。

综上所述, 筛选出了反演COD浓度和Chl-a浓度的敏感因子, 为后续构建反演模型提供了基础。

##### 2.2.2 模型构建与检验

根据选择的最优反演波段组合分别构建ISSA-SVR模型、SSA-SVR模型、SVR模型和多元线性回归模型。为了减少单个模型随机性和预设训练

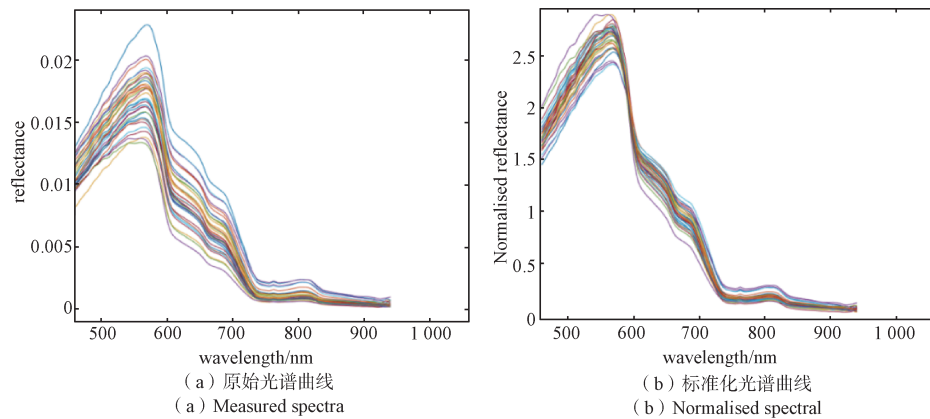


图3 光谱遥感反射率

Fig. 3 Spectral remote sensing reflectance

表1 各参数的敏感因子

Table 1 Sensitivity factors of each parameter

水质参数	敏感因子	相关系数
COD	B1/B2 B7/B6	0.772 0.736
Chl-a	B1/B3 B3/B2 B4×B3	0.678 0.663 0.638

集造成的干扰, 获得 ISSA-SVR 模型的稳定性能, 实验采用交叉验证的方式。从 2022 年 9 月 18 日和 2023 年 5 月 16 日两天获得的实测数据中, 每日随机抽取 5 个采样点的数据共 10 个数据组成测试样本, 两日剩余 30 个采样点的数据作为模型的训练样本。在随机获得的训练集中训练所有模型, 同时用对应的测试集进行验证。重复实验 100 次, 并对结果进行平均, 平均值定义为模型的最终预测结果。在 ISSA-SVR 模型预测中, 首先设置模型的初始参数: 种群规模为 50, 最大迭代次数为 200。用决定系数  $R^2$  确定模型拟合度, 基于平均相对误差 (Mean Relative Error, MRE) 评估模型的反演精度。决定系数的取值在 0 到 1 之间; 越接近 1, 模型的拟合越好。平均相对误差越小, 对应的反演模型的精度越高。COD 浓度和 Chl-a 浓度的反演结果见表 2。

从  $R^2$  看, COD 的 ISSA-SVR 反演模型的决定系数为 0.734, Chl-a 的 ISSA-SVR 反演模型决定系数为 0.797。结果显示 ISSA-SVR 模型的决定系数显著高于传统的 SVR 模型和 SSA-SVR 模型, 拟合效果更好。从相对误差来看, COD 和 Chl-a 的 ISSA-SVR 反演模型的平均相对误差分为 20.02% 和 30.17%。ISSA-SVR 模型在测试集中的平均相对误差远小于其他模型, 对上述三种水质参数的反演精度更高。通过对比 ISSA-SVR 模型、SSA-SVR 模

表2 训练集和测试集的参数评估结果

Table 2 Parameter evaluation results for the training set and test set

水质参数	模型	训练集		测试集	
		$R^2$	MRE%	$R^2$	MRE%
COD	线性回归	0.537	43.67	0.442	40.20
	SVR	0.760	25.07	0.660	24.60
	SSA-SVR	0.626	35.58	0.545	38.90
	ISSA-SVR	0.908	16.52	0.734	20.02
Chl-a	线性回归	0.555	48.15	0.429	51.60
	SVR	0.641	45.41	0.590	58.99
	SSA-SVR	0.761	31.18	0.631	33.92
	ISSA-SVR	0.874	27.32	0.797	30.17

型、SVR 模型和多元线性模型的拟合和平均相对误差结果, ISSA-SVR 模型是实现 COD 和 Chl-a 有效反演的最佳模型。

### 2.2.3 采用 ISSA-SVR 模型对 Sentinel-2 卫星数据进行水质参数浓度反演

为了进一步验证 ISSA-SVR 模型在 Sentinel-2 卫星图像上的精度, 确保该模型具有实际应用价值。通过与采样点相同时间和位置的 Sentinel-2 卫星图像进行验证, 并分析采样点预测值与实测值的相关性。采样点预测值与实测值的散点图如图 4 所示, 大部分都分布在数据偏差线附近。

验证结果见表 3。COD 浓度和 Chl-a 浓度的实测值与预测值拟合的决定系数为 0.71 和 0.74。COD 预测值最小相对误差为 2.49%, 平均相对误差为 23.77%。Chl-a 预测值最小相对误差为 13.44%, 平均相对误差为 26.42%。对比 SVR 模型和 SSA-SVR 模型, ISSA-SVR 模型的拟合更好、精度更高、具有较好的数据相关性和预测一致性, 说明



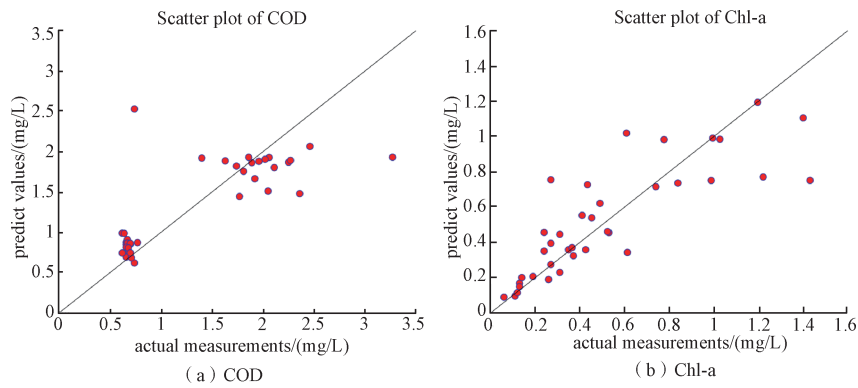


图4 样本点预测值和实测值的散点图

Fig. 4 The scatter plot of sample point predict values and actual measurements

该模型可以实现对典型水质参数浓度的有效反演。

表3 卫星图像验证结果

Table 3 Satellite image validation results

水质参数	模型	R <sup>2</sup>	MRE
COD	SVR	0.61	27.90%
	SSA-SVR	0.59	35.62%
	ISSA-SVR	0.71	23.77%
Chl-a	SVR	0.6	49.81%
	SSA-SVR	0.65	31.36%
	ISSA-SVR	0.74	26.42%

利用2022年9月18日获取的Sentinel-2卫星遥感数据，结合构建的ISSA-SVR模型进行反演，计算出卫星遥感图像每一个像元对应的水质参数浓度，获得的唐岛湾海域的各水质参数浓度空间反演效果图，如图5所示。

由图可知，COD和Chl-a浓度在河口和沿岸较高。结合唐岛湾周边地理情况及查阅资料发现，唐岛湾周边存在一定的排污口以及大量的住宅区，

一定的生活污水是导致河口和沿岸COD浓度和Chl-a浓度偏高的原因之一。

### 2.3 讨论

如图4所示，基于Sentinel-2卫星遥感数据的部分采样点反演结果与实测浓度略有差异值，可能是因为两者之间数据质量的差异。用于收集测量光谱的TRIOS传感器与Sentinel-2卫星传感器不同，并且该传感器具有不同的参数集，例如信噪比。结果表明，采样点应尽可能远离海岸，以尽量减少非水像元对唐岛湾海岸的影响，并且在光谱采集过程中船舶的轻微摇摆导致部分反演结果与实测值存在显著差异。

此外，本实验通过相关性分析选择用于建模的光谱特征波段，在选择COD反演波段时，B7/B6相关系数为0.736，但是COD对光的吸收率随波长指数级下降，在近红外波段对反射的贡献很低。所以，选择更好的波段选择算法是下一步研究的重点。

为了研究水污染和水体富营养化问题对生态

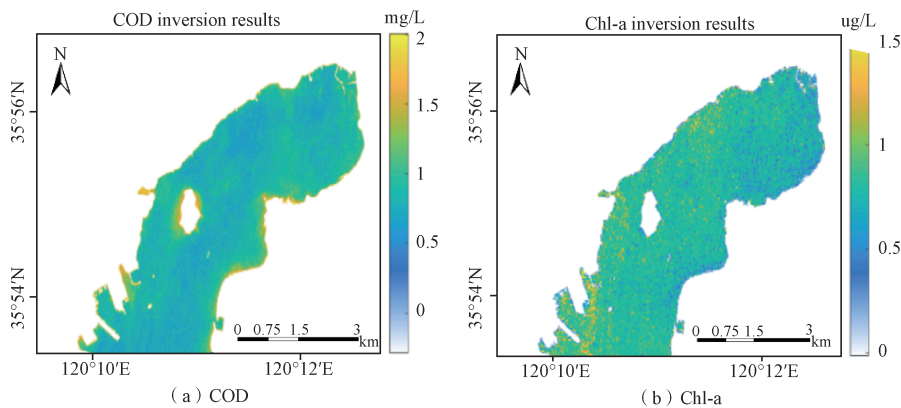


图5 唐岛湾反演结果

Fig. 5 Spatial inversion results of water quality parameters in Tangdao Bay

系统的影响, 实现 COD 浓度和 Chl-a 浓度的高精度遥感反演。本研究利用 2022 年 9 月 18 日和 2023 年 5 月 16 日在唐岛湾水域测量的水质参数数据和光谱数据, 构建 ISSA-SVR 模型预测唐岛湾的 COD 浓度和 Chl-a 浓度。评价结果表明, ISSA-SVR 混合优化模型在该区域具有良好的可行性和预测性。与 SVR 模型的平均相对误差相比, ISSA-SVR 模型的预测误差分别降低了 4.13% 和 23.39%。结果表明, ISSA-SVR 模型对该地区上述两种水质参数浓度的预测精度均优于 SVR 模型。

为了解决 SSA 算法种群多样性低的问题, 本文提出的 ISSA 算法在麻雀种群初始化阶段引入了基于 Logistic 映射和 Fraction 映射的二维混沌映射系统, 与传统混沌映射相比, 该映射系统的映射值更为分散, 可以有效增加麻雀种群的多样性。

为了解决 SSA 算法种群容易陷入局部最优解的问题, 本文提出的 ISSA 算法系统引入反向学习策略, 改进了 SSA 算法的位置更新公式, 对最优解进行扰动产生新解。同时引入模拟退火算法, 当新位置的适应度值大于当前全局最优位置的适应度值时, 仍有保留的可能。反向学习策略和模拟退火算法显著地提高了 SSA 算法的全局搜索能力, 避免在迭代过程中陷入局部最优, 有效避免了局部最优解的出现。

总体而言, ISSA 算法可以有效解决 SSA 算法容易陷入局部最优的问题, 在典型水质参数的反演中具有更好的预测精度。

### 3 结束语

本文引入了二维混沌映射系统, 模拟退火算法和反向学习算法, 设计了一种混合优化算法 ISSA 算法, 解决麻雀搜索算法易陷入局部最优解的问题。

本文引入 ISSA 算法对 SVR 模型参数进行优化, 并与多元线性回归、SVR、SSA-SVR 等多种预测方法进行对比分析。结果表明, ISSA 算法的反演精度明显优于其他模型。该模型应用于 Sentinel-2 卫星遥感数据时, 仍然具有良好的预测性能。证明基于实测光谱和卫星影像的水质参数反演方法具备实际应用于遥感水质监测的能力。

基于实测光谱的 ISSA-SVR 模型在唐岛湾水域表现良好, 但其在其他复杂水环境中的应用性

能仍有待进一步研究。因此, 下一步需要收集更大范围的水体光谱数据和其他水质参数, 对 ISSA-SVR 模型方法进行更全面的评价。接下来的工作将继续进行水质实验, 以扩大样本库, 并且利用其他卫星遥感数据(如 MODIS、GOCI 等)在多个不同水域进行实验, 以提高模型的适用性。

### 参考文献

- [1] SINGH S, D'SA E J, SWENSON E M. Chromophoric dissolved organic matter (CDOM) variability in Barataria Basin using excitation-emission matrix (EEM) fluorescence and parallel factor analysis (PARAFAC)[J]. *Science of the Total Environment*, 2010, 408(16): 3211-3222.
- [2] CORY R M, MCKNIGHT D M. Fluorescence spectroscopy reveals ubiquitous presence of oxidized and reduced quinones in dissolved organic matter[J]. *Environmental Science & Technology*, 2005, 39(21): 8142-8149.
- [3] 苏杭, 黄海军, 刘艳霞, 等. 基于遥感反演的浒苔绿潮暴发对有色溶解有机物与叶绿素时空分布特征的影响[J]. *海洋科学*, 2023, 47(4): 9-24.  
SU Hang, HUANG Haijun, LIU Yanxia, et al. Effects of *Enteromorpha prolifera* blooms on the spatial and temporal distributions of CDOM and chlorophyll by remote sensing[J]. *Marine Sciences*, 2023, 47(4): 9-24.
- [4] HAN Zenglei, XIAO Min, WANG Zhongliang. Application of Landsat data in analyzing the influence mechanism of dissolved organic carbon in Taihu Lake[J]. *Chinese Journal of Ecology*, 2020, (7): 2446 -2455.
- [5] 黄全佳. ElNiño 事件对河流入海 COD 的调控机制-基于高分辨率时间序列观测的证据[J]. *海洋学报*, 2021, 43(6): 62-70.  
HUANG Quanjia. Regulation of El Niño event on riverine COD export: Evidence from high frequency time-series buoy monitoring[J]. *Haiyang Xuebao*, 2021, 43(6): 62-70.
- [6] SAGAN V, PETERSON K T, MAIMAITIJIANG M, et al. Monitoring inland water quality using remote sensing: Potential and limitations of spectral indices, bio-optical simulations, machine learning, and cloud computing[J]. *Earth-Science Reviews*, 2020, 103187.
- [7] GHOLIZADEH M, MELESSE A, REDDI L. A comprehensive review on water quality parameters estimation us-



- ing remote sensing techniques[J]. *Sensors*, 2016, 16(8): 1298.
- [8] DUAN H, MA R, ZHANG Y, et al. Remote-sensing assessment of regional inland lake water clarity in northeast China[J]. *Limnology*, 2009, 10(2): 135-141.
- [9] CHEGOONIAN A M, PETER R L, KIANA Z, et al. Regional upscaling of Chlorophyll-A retrieval from Small Eutrophic Lakes Via Sentinel-2[J]. 2022 IEEE International Geoscience and Remote Sensing Symposium (IGARSS), 2022: 5622-5625.
- [10] YE Xuehui, LI Yuxia, LING Tong, et al. Remote sensing retrieval of suspended solids in Longquan Lake based on GA-SVM model[J]. 2017 IEEE International Geoscience and Remote Sensing Symposium (IGARSS), 2017: 5501-5504.
- [11] TAN Z, REN J, LI S, et al. Inversion of nutrient concentrations using machine learning and influencing factors in Minjiang River[J]. *Water*, 2023, 15(7): 1398.
- [12] VALERA M, WALTER R K, BAILEY B A, et al. Machine learning based predictions of dissolved oxygen in a small coastal embayment[J]. *Journal of Marine Science and Engineering*, 2020, 8(1007).
- [13] 刘梅, 马启良, 原居林, 等. 基于无人机高光谱遥感技术对内陆养殖池塘水质监测的研究[J]. *海洋与湖沼*, 2022, 53(1): 195-205.
- LIU Mei, MA Qiliang, YUAN Julin, et al. Water quality monitoring of inland aquaculture ponds based on UAV hyperspectral remote sensing technology[J]. *Oceanologia et Limnologia Sinica*, 2022, 53(1): 195-205.
- [14] XUE T L, ZHAO D F, HAN F. Research on water quality prediction model of SVR based on GA optimization[J]. *Environ*, 2020, 38: 123-127.
- [15] 崔海, 余鑫磊, 庞继伟, 等. 采用BP-ANN和改进SVR的进水BOD软测量模型[J]. *哈尔滨工业大学学报*, 2022, 54(2): 59-66.
- CUI Hai, YU Xinlei, PANG Jiwei, et al. Influent BOD soft sensing models based on BP-ANN and improved SVR[J]. *Journal of Harbin Institute of Technology*, 2022, 54(2): 59-66.
- [16] MAHMOUDI N, OROUJI H. Integration of shuffled frog leaping algorithm and support vector regression for prediction of water quality parameters[J]. *Water Resour*, 2016, 30: 2195-2211.
- [17] MIRJALILI S. Moth-flame optimization algorithm: A novel nature-inspired heuristic paradigm[J]. *Knowledge-based Systems*, 2015, 89: 228-249.
- [18] 盛辉, 池海旭, 许明明, 等. 改进SVR的内陆水体COD高光谱遥感反演[J]. *光谱学与光谱分析*, 2021, 41(11): 3565-3571.
- SHENG Hui, CHI Haixu, XU Mingming, et al. Inland water chemical oxygen demand estimation based on improved SVR for hyperspectral data[J]. *Spectroscopy and Spectral Analysis*, 2021, 41(11): 3565-3571.
- [19] 贺振霞, 鲍学英. 基于改进PSO优化SVR的地下水水质综合评价研究[J]. *水文*, 2021, 41(6): 26-32.
- HE Zhenxia, BAO Xueying. Comprehensive evaluation of groundwater quality based on SVR optimized by improved PSO[J]. *Journal of China Hydrology*, 2021, 41(6): 26-32.
- [20] 郭利进, 李博仑. 基于GNFA-SVR污水出水总氮软测量研究[J]. *工业水处理*, 2022, 42(10): 111-117.
- GUO Lijin, LI Bolun. Research on soft measurement of effluent total nitrogen based on GNFA-SVR[J]. *Industrial Water Treatment*, 2022, 42(10): 111-117.
- [21] 张燕君, 康成龙, 柳雅倩, 等. 基于表面增强拉曼光谱技术和GWO-SVR算法快速实现水中总氮总磷含量检测[J]. *光谱学与光谱分析*, 2021, 41(10): 3147-3152.
- ZHANG Yanjun, KANG Chenglong, LIU Yaqian, et al. Rapidly detection of total nitrogen and phosphorus content in water by surface enhanced Raman spectroscopy and GWO-SVR algorithm[J]. *Spectroscopy and Spectral Analysis*, 2021, 41(10): 3147-3152.
- [22] XUE J K, SHEN B. A novel swarm intelligence optimization approach: Sparrow search algorithm[J]. *Systems Science & Control Engineering*, 2020, 8(1): 22-34.
- [23] KEERTHI S S, LIN C J. Asymptotic behaviors of support vector machines with Gaussian kernel[J]. *Neural Computation*, 2003, 15(7): 1667-1689.
- [24] CAO Weijia, CAI Hang, HUA Zhongyun. N-dimensional chaotic map with application in secure communication[J]. *Chaos, Solitons & Fractals*, 2022, 163: 112519.
- [25] AGUILERA M A Z. Classification of land-cover through machine learning algorithm for fusion of Sentinel-2a and planet scope imagery[C]//2020 IEEE Latin American GRSS & ISPRS Remote Sensing Conference

- (LAGIRS),2020: 246-253.
- [26] WANG J, DING J, YU D, et al. Capability of Sentinel-2 MSI data for monitoring and mapping of soil salinity in dry and wet seasons in the Ebinur Lake region, Xinjiang, China[J]. *Geoderma*, 2019, 353: 172-187.
- [27] DOERFFER R, SCHILLER H. The MERIS case 2 water algorithm[J]. *International Journal of Remote Sensing*, 2007, 28 (3-4): 517-535.
- [28] MU J, LIU C, MENG X. Atmospheric correction of inland lake waters based on Sentinel-2 satellite data[J]. *Mapping Bulletin*, 2021, 09: 9-14.
- [29] SCHILLER H, ROLAND D. 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.
- [30] 刘善伟, 武钰林, 许明明, 等. 无人机多光谱遥感反演近海 fDOM 浓度[J]. *海洋技术学报*, 2021, 40(6): 33-39. LIU Shanwei, WU Yulin, XU Mingming, et al. Retrieval of offshore fDOM concentration by UAV multispectral remote sensing[J]. *Journal of Ocean Technology*, 2021, 40(6): 33-39.
- [31] PARIDA B R, KUMARI A. Mapping and modeling mangrove biophysical and biochemical parameters using Sentinel-2A satellite data in Bhitarkanika National Park, Odisha[J]. *Modeling Earth Systems and Environment*, 2021, 7: 2463-2474.

#### [作者简介]

刘元杰 1999年生, 硕士研究生。

崔建勇 1976年生, 讲师。

董文 1984年生, 助理研究员。

万剑华 1966年生, 教授, 博士生导师。

张杰 1963年生, 教授, 博士生导师。

(本文编辑: 杨秀丽)