

唐毅, 徐全, 杜彬, 等. 基于 SARIMA-VMD-LSSVM 的水产养殖溶解氧质量浓度预测[J]. 江苏农业学报, 2024, 40(8): 1473-1482.  
doi:10.3969/j.issn.1000-4440.2024.08.012

# 基于 SARIMA-VMD-LSSVM 的水产养殖溶解氧质量浓度预测

唐毅, 徐全, 杜彬, 王磊, 袁瑞豪, 袁禹

(西华大学机械工程学院, 四川 成都 610039)

**摘要:** 为了充分利用溶解氧质量浓度的数据特征, 进一步提高水产养殖中溶解氧质量浓度预测的准确性, 提出“线性与非线性”与“分解-预测-集成”相结合的溶解氧质量浓度预测模型。该模型首先由季节性差分自回归滑动平均(SARIMA)模型对溶解氧质量浓度随着时间变化而组成的数据序列(简称溶解氧质量浓度的时间序列)进行线性拟合, 使用变分模态分解(VMD)对残差序列进行分解, 然后将各残差分量代入经改进的灰狼算法(IGWO)优化的最小二乘支持向量机模型(LSSVM)中, 得到非线性分量的预测结果。最后集成线性与非线性预测结果, 得到最终的溶解氧质量浓度预测值。结果表明, 与 SARIMA、LSSVM、VMD-LSSVM 模型相比, 基于 SARIMA-VMD-LSSVM 模型对溶解氧质量浓度进行预测的精度显著提高, 预测的均方根误差(RMSE)为0.078 7, 平均相对误差(MAPE)为0.022 6, 说明该组合模型可有效提取溶解氧质量浓度的时间序列的多尺度特征, 从而更精准地进行溶解氧质量浓度的预测。

**关键词:** 水产养殖; 溶解氧; 变分模态分解; 组合预测方法; 改进的灰狼算法

**中图分类号:** TP391;S912 **文献标识码:** A **文章编号:** 1000-4440(2024)08-1473-10

## Prediction of dissolved oxygen mass concentration in aquaculture based on SARIMA-VMD-LSSVM

TANG Yi, XU Quan, DU Bin, WANG Lei, YUAN Ruihao, YUAN Yu

(School of Mechanical Engineering, Xihua University, Chengdu 610039, China)

**Abstract:** In order to make full use of the data characteristics of dissolved oxygen mass concentration and further improve the accuracy of dissolved oxygen mass concentration prediction in aquaculture, a dissolved oxygen mass concentration prediction model combining "linear and nonlinear" and "decomposition-prediction-integration" was proposed. Firstly, the seasonal auto regressive integrated moving average (SARIMA) model was used to linearly fit the dissolved oxygen mass concentration time series, and the residual sequence was decomposed using variational mode decomposition (VMD). Then, each residual component was substituted into the least square support vector machine (LSSVM) model optimized by the improved gray wolf algorithm (IGWO) to obtain the prediction results of the nonlinear component. Finally, the linear and nonlinear prediction results were integrated to obtain the final dissolved oxygen mass concentration prediction value. Experimental

results showed that compared with SARIMA, LSSVM, and VMD-LSSVM models, the prediction accuracy of SARIMA-VMD-LSSVM model was significantly improved. The root mean square error (RMSE) was 0.078 7, and the mean absolute percentage error (MAPE) was 0.022 6, indicating that the combined model could effectively extract the multi-scale features of the time series of dissolved oxygen mass concentration, and achieve more accurate predic-

收稿日期: 2023-05-25

基金项目: 四川省科技成果转化示范项目(2020ZHCG0076); 工业控制技术国家重点实验室开放课题(ICT2022B45)

作者简介: 唐毅(1997-), 男, 四川江油人, 硕士, 主要研究方向为机器学习、智能水产养殖。(Tel) 13320889755; (E-mail) tangyi970516@163.com

通讯作者: 徐全, (E-mail) quanxnjd@sina.com

tion.

**Key words:** aquaculture; dissolved oxygen; variational mode decomposition; combinatorial forecasting methods; improved gray wolf algorithm

在水产养殖中,溶解氧含量是养殖户关心的重要参数之一,其含量过高或不足,都会严重影响水生生物的生长与健康,决定着养殖对象的产量和品质。目前,调节溶解氧含量常用的方法有实时或定时调节,然而由于数据传输及增氧机的启动与停止都存在滞后性,仅依据现有数据进行溶解氧含量的调节,不仅难以改善水质,还可能引起水质指标的波动,不利于水生生物的健康<sup>[1]</sup>。因此,溶解氧含量的预测,能够帮助养殖户及时调控水质、优化养殖管理,最大限度地发挥水生生物的生产潜力,提高生产效益。

目前,统计预测、机器学习预测及组合模型预测等方法是对溶解氧含量进行预测的主要方法。其中,差分自回归滑动平均 (ARIMA) 模型是统计预测中最常使用的模型,如 Park 等<sup>[2]</sup>使用 ARIMA 模型对洛东江口水质进行预测。随着机器学习技术的不断发展,其在数据预测领域的应用也越来越广泛,在对水质的预测中也取得了不错的效果,例如,张梦迪等<sup>[3]</sup>用反向传播 (BP) 神经网络模型实现了对多水质参数的预测;龚怀瑾等<sup>[4]</sup>使用变尺度混沌搜索优化的最小二乘支持向量机 (LSSVM) 模型,实现了对溶解氧含量的预测。然而,单一模型难以全面有效地表征溶解氧质量浓度随着时间变化而组成的数据序列 (简称溶解氧质量浓度的时间序列) 的特征。因此,针对溶解氧含量的非线性与非平稳特征,“分解-预测-集成”的组合预测模型被广泛应用于水质预测中。刘晨等<sup>[5]</sup>为了降低溶解氧质量浓度时间序列的复杂性,使用集合经验模态分解 (EEMD),再结合 SVM 实现了最终的预测。白雯睿等<sup>[6]</sup>将变分模态分解方法 (VMD) 与 CNN-LSTM 模型相结合,提升了溶解氧含量的预测精度。然而,上述方法未考虑溶解氧含量的数据特征,且在模型的选择上未考虑预测的时效性。

为了弥补上述单一模型与组合模型在水质预测中的不足,本研究提出 1 种基于季节性差分自回归滑动平均 (SARIMA)-VMD-LSSVM 的短时溶解氧质量浓度组合预测模型。SARIMA 模型能够识别具有周期性与非平稳时间序列的线性规律,且结构简单,使用少量数据便可进行预测。VMD 分解算法可用于充分削弱序列的非线性特征,同时能起到降噪的

作用。LSSVM 模型的训练速度快,适用于非线性回归问题,且对少量数据建模具有一定的优势。因此,本研究拟通过将 3 种算法进行串联组合用于水产养殖中溶解氧质量浓度的预测,并引入改进的灰狼算法优化 LSSVM 模型,以期避免 LSSVM 参数选择的盲目性,进一步提升预测精度。

## 1 基本理论与模型的构建

### 1.1 SARIMA 模型

SARIMA 模型是由 Jenkins 与 Box 提出的一种时间序列预测模型,由自回归模型 (AR) 与移动平均模型 (MA) 逐步发展而来,这类模型融合了时间序列分析和回归分析的优点,并在此基础上进一步发展,适用于存在周期性与非平稳性的序列,其一般形式如下:

$$\Delta^d \Delta_s^D x_t = \frac{\Theta_q(L) B_Q(L^s) e_t}{\Phi_p(L) A_P(L^s)} \quad (1)$$

式中,  $\Delta^d$ 、 $\Delta_s^D$  分别表示非周期部分的  $d$  阶差分、周期部分的  $D$  阶差分;  $x_t$  表示时间序列;  $e_t$  表示误差,服从正态分布;  $\Phi_p(L)$ 、 $A_P(L^s)$  分别表示非周期部分、周期部分的自回归多项式,其中  $\Phi_p(L) = (1 - \varphi_1 L - \varphi_2 L^2 - \dots - \varphi_p L^p)$ ,  $A_P(L^s) = (1 - \alpha_1 L^s - \alpha_2 L^{2s} - \dots - \alpha_P L^{Ps})$ ,  $\varphi_p$ 、 $\alpha_P$  分别表示滞后  $p$  阶、 $P$  阶多项式的自相关系数,其中  $p$ 、 $P$  分别表示自回归阶数、周期性自回归阶数;  $\Theta_q(L)$ 、 $B_Q(L^s)$  分别表示非周期部分、周期部分的移动平均多项式,其中  $\Theta_q(L) = (1 - \theta_1 L + \theta_2 L^2 + \dots + \theta_q L^q)$ ,  $B_Q(L^s) = (1 + \beta_1 L^s + \beta_2 L^{2s} + \dots + \beta_Q L^{Qs})$ ;  $\theta_q$ 、 $\beta_Q$  分别表示滞后  $q$  阶、 $Q$  阶多项式系数,其中  $q$ 、 $Q$  分别表示移动平均阶数、周期性移动平均阶数。

最终的模型可以表示为 SARIMA( $p, d, q$ )( $P, D, Q$ ), SARIMA 模型的建模流程包括平稳性检验、差分、周期性检验、季节性差分、模型定阶和预测。

### 1.2 变分模态分解方法

变分模态分解 (VMD) 方法是由 Dragomiretskiy 等<sup>[7]</sup>提出的新型时频分析方法,该方法将时间序列信号分解为一系列不同频率与振幅的模态分量。该方法具有消除边界效应、避免模态混叠及较强的抗干扰能力的特点。

VMD 方法将信号  $f(x)$  分解为  $K$  个模态分量  $(\mu_k)$ , 每个模态分量以  $\omega_k$  为中心频率进行分解, 变成约束变分问题, 公式如下:

$$\begin{cases} \min_{\{u_k, \omega_k\}} \left\{ \sum_{k=1}^K \left\| \partial_t \left[ \left( \delta(t) + \frac{j}{\pi t} \right) u_k(t) \right] e^{-j\omega_k t} \right\|_2^2 \right\} \\ \text{s.t. } \sum_{k=1}^K u_k(t) = f(t) \quad k=1, 2, \dots, K \end{cases} \quad (2)$$

$$L(\{u_k\}, \{\omega_k\}, \lambda) = \alpha \sum_{k=1}^K \left\| \partial_t \left[ \left( \delta(t) + \frac{j}{\pi t} \right) u_k(t) \right] e^{-j\omega_k t} \right\|_2^2 + \left\| f(t) - \sum_{k=1}^K u_k(t) \right\|_2^2 + \langle \lambda(t), f(t) - \sum_{k=1}^K u_k(t) \rangle \quad (3)$$

式中,  $\alpha$  可以确保在高斯噪声的情况下信号重构的准确度;  $\lambda$  可用于维护严格的约束条件;  $\langle \rangle$  表示求内积。

VMD 分解步骤如下:

步骤 1: 对初始参数进行初始化, 如模态分量  $\{u_k^1\}$ 、中心频率  $\{\omega_k^1\}$  及拉格朗日乘数  $(\lambda^1)$ 。并选取适当的模态分量数量 ( $K$ )。

步骤 2: 根据公式 (4) 与公式 (5) 分别更新模态  $(u_k)$  和中心频率  $(\omega_k)$  的取值, 公式如下:

$$\hat{u}_k^{n+1}(\omega) = \frac{\hat{f}(\omega) - \sum_{i \neq k} \hat{u}_i(\omega) + \left[ \frac{\hat{\lambda}(\omega)}{2} \right]}{1 + 2\alpha(\omega - \omega_k)^2} \quad (4)$$

式中,  $\hat{u}_k^{n+1}(\omega)$ 、 $\hat{f}(\omega)$ 、 $\hat{u}_i(\omega)$ 、 $\hat{\lambda}(\omega)$  表示各变量的傅立叶变换,  $\omega$  表示中心频率。

$$\omega_k^{n+1} = \frac{\int_0^\infty \omega |\hat{u}_k^{n+1}(\omega)|^2 d\omega}{\int_0^\infty |\hat{u}_k^{n+1}(\omega)|^2 d\omega} \quad (5)$$

步骤 3: 根据公式 (6), 更新  $\lambda$ , 相关公式如下:

$$\hat{\lambda}^{n+1}(\omega) = \hat{\lambda}^n(\omega) + \tau [\hat{f}(\omega) - \sum_k \hat{u}_k^{n+1}(\omega)] \quad (6)$$

式中,  $\tau$  为更新因子。

步骤 4: 给定判定精度的阈值  $(\varepsilon) > 0$ , 若满足公式 (7) 的条件, 则停止迭代, 否则返回步骤 2。

$$\sum_k \frac{\|\hat{u}_k^{n+1} - \hat{u}_k^n\|_2^2}{\|\hat{u}_k^n\|_2^2} < \varepsilon \quad (7)$$

### 1.3 LSSVM 模型

LSSVM 模型<sup>[8]</sup>是在支持向量机 (SVM) 的基础上进行改进得到的, 该模型将 SVM 的二次规划问题转变成线性方程组求解问题, 大大提升了算法的求解速度。对于非线性时间序列样本集  $\{x_i, y_i\}, i=1, 2, \dots, N$ , 将输入样本映射到高维特征空间, 然后在特征空间中进行线性回归, 假设最优回归函数为:

$$\hat{y} = f(x) = \omega^T \phi(x) + b \quad (8)$$

式中,  $\omega$  为权重向量,  $\phi(x)$  为处理非线性问题

式中,  $\delta(t)$  表示单位脉冲函数;  $\partial_t$  表示对函数求时间  $t$  的偏导;  $\|\cdot\|$  表示  $L^2$  范数;  $j^2 = -1$ 。

为了求解上述问题的最优解, 引用二次惩罚因子  $(\alpha)$  和拉格朗日乘子  $(\lambda)$  将问题转化为非约束变分问题进行求解, 得到增广变分问题, 公式如下:

的核函数, 可将输入空间映射到高维特征空间。

将上述回归问题转化为求解最优问题:

$$\begin{cases} \min J(\omega, b, e) = \frac{1}{2} \omega^T \omega + \frac{\gamma}{2} \sum_{i=1}^n e_i^2 \\ \text{s.t. } y_i = \omega^T \phi(x_i) + b + e_i \end{cases} \quad (9)$$

式中,  $J$  为损失函数;  $\omega$  为权重向量;  $b$  为偏置量;  $e_i$  为误差;  $\gamma$  为正则化参数。

为了求解上述优化问题, 引入拉格朗日乘子向量  $(a_i)$  和 KKT 条件, 将优化问题转化为二元问题, 构建如下 Lagrange 函数:

$$L(\omega, b, e, a) = \frac{1}{2} \omega^T \omega + \gamma \sum_{i=1}^l e_i^2 - \sum_{i=1}^l a_i [\omega^T \phi(x_i) + b - y_i + e_i] \quad (10)$$

分别对  $L(\omega, b, e, a)$  求  $\omega$ 、 $b$ 、 $e$ 、 $a$  的偏微分, 公式如下:

$$\begin{cases} \frac{\partial L}{\partial \omega} = 0 \rightarrow \sum_{i=1}^n a_i \phi(x_i) = 0 \\ \frac{\partial L}{\partial b} = 0 \rightarrow \sum_{i=1}^n a_i = 0 \\ \frac{\partial L}{\partial e_i} = 0 \rightarrow a_i + \gamma e_i = 0 \\ \frac{\partial L}{\partial a_i} = 0 \rightarrow \omega^T \phi(x_i) + b - y_i + e_i = 0 \end{cases} \quad (11)$$

根据公式 (11), 消去公式 (10) 中的  $\omega$ 、 $e_i$ , 最终得到 LSSVM 的决策函数, 公式如下:

$$f(x) = \sum_{i=1}^n a_i K(x_i, x_j) + b \quad (12)$$

式中,  $K(x_i, x_j)$  为核函数, 本研究选择 RBF (高斯径向基函数) 为核函数, 公式如下:

$$K(x_i, x_j) = \exp \left( -\frac{\|x_i - x_j\|}{2\sigma^2} \right) \quad (13)$$

式中,  $\sigma$  为径向基宽度参数。

### 1.4 改进的灰狼优化算法 (GWO) 优化 LSSVM

1.4.1 灰狼优化算法 灰狼优化算法 (GWO)<sup>[9]</sup>是

一种基于自然界中灰狼行为的启发式优化算法。寻优过程分为包围、猎捕和攻击 3 个过程。

(1) 包围。灰狼群体会先按照公式 (14) 包围猎物。

$$X(t+1) = X_p(t) - A |CX_p(t) - X(t)| \quad (14)$$

$$\begin{cases} A = 2ar_1 - a \\ C = 2r_2 \\ a = 2 - 2t/t_{\max} \end{cases} \quad (15)$$

式中,  $X_p(t)$  为猎物的位置;  $X(t)$  为当前灰狼的位置;  $t$  为当前迭代次数;  $t_{\max}$  为最大迭代次数;  $r_1$  与  $r_2$  为  $[0, 1]$  的随机数;  $a$  为收敛因子。

(2) 猎捕。该过程由  $\alpha$  (最优解)  $\beta$  (次优解) 与  $\delta$  (次次优解) 引导更新个体位置:

$$\begin{cases} X_1 = X_\alpha - A_1 |C_1 X_\alpha - X(t)| \\ X_2 = X_\beta - A_2 |C_2 X_\beta - X(t)| \\ X_3 = X_\delta - A_3 |C_3 X_\delta - X(t)| \end{cases} \quad (16)$$

$$X(t+1) = (X_1 + X_2 + X_3) / 3 \quad (17)$$

式中,  $X_\alpha$  为最优解位置;  $X_\beta$  为次优解位置;  $X_\delta$  为次次优解位置。

(3) 攻击。通过收敛因子 ( $a$ ) 从 2 到 0 的线性递减实现攻击。群体在迭代结束后, 获得最优解。

1.4.2 改进的灰狼优化算法 GWO 作为一种元启发式优化算法, 可以用来优化 LSSVM 的惩罚参数  $\gamma$ 、 $\sigma$ 。然而, GWO 算法与大多数智能算法一样, 存在后期收敛速度慢、易陷入局部最优解等问题。因此, 本研究引入非线性收敛因子, 并结合麻雀算法中的探索者以提升算法跳出局部最优的能力与收敛速度。

(1) 非线性收敛因子。在 GWO 算法中, 收敛因子  $a$  从 2 到 0 线性递减, 无法对全局与局部搜索能力进行平衡。非线性收敛因子呈抛物线状递减, 有利于提升全局搜索能力, 在迭代后期快速递减, 有利于提升局部搜索能力。

$$a = 2 - 2 \left( \frac{\sin \frac{t}{t_{\max}}}{\sin 1} \right)^\mu \quad (18)$$

式中,  $\mu$  为调节系数,  $\mu = 0.5$ 。

(2) 麻雀算法优化。本研究将麻雀算法中探索者的大范围搜索能力引入灰狼算法<sup>[10]</sup>, 用于进一步提升 GWO 算法的搜索范围与收敛能力。具体步骤如下:

步骤 1: 根据比例系数 ( $\zeta$ ), 将狼群划分为探索者狼群与跟随者狼群。

$$\zeta = \frac{PN}{N - PN} \quad (19)$$

式中,  $N$  为灰狼总数;  $PN$  表示位置较好的探索者灰狼数量。

步骤 2: 探索者灰狼位置更新。

$$X(t+1) = X(t) e^{\left(\frac{-\zeta}{\zeta \cdot t_{\max}}\right)} \quad (20)$$

式中,  $\zeta$  为  $[0, 1]$  范围内的随机值。

步骤 3: 其他灰狼仍按照原灰狼优化算法进行位置更新。

1.4.3 改进的灰狼算法 (IGWO) 优化 LSSVM 采用 IGWO 优化 LSSVM 的惩罚参数 ( $\gamma$ ) 和径向基宽度 ( $\sigma$ ), 并以 LSSVM 最小均方差函数 ( $MSE$ ) 作为目标函数, 其表达式见公式 (21)。

$$\min f(\gamma, \sigma) = \frac{1}{m} \sum_{i=1}^m (x_i - \hat{x}_i)^2 \quad (21)$$

式 (21) 中,  $x_i$  表示真实值;  $\hat{x}_i$  表示预测值。

使用 IGWO 优化 LSSVM 模型的具体步骤如下:

步骤 1: 初始化参数。生成 20 只灰狼, 每只灰狼对应 1 组 LSSVM 的  $\gamma$ 、 $\sigma$ , 设置其取值范围为 (0.01 ~ 1 000.00), 设置最大迭代次数为 100。

步骤 2: 寻找最佳猎物。从第 1 只灰狼开始, 根据公式 (21) 计算适应度, 并按照适应度高低划分灰狼等级, 适应度较佳的前 3 匹狼分别表示为  $\alpha_{\text{狼}}$ 、 $\beta_{\text{狼}}$ 、 $\delta_{\text{狼}}$ , 剩下的狼则全为  $\omega_{\text{狼}}$ 。

步骤 3: 位置更新。根据公式 (20) 对探索者灰狼进行位置更新, 根据公式 (17) 对其他灰狼进行位置更新。

步骤 4: 判断是否达到最大迭代次数, 若达到, 则输出最优参数组合。

## 1.5 SARIMA-VMD-LSSVM 模型的构建

溶解氧质量浓度时序数据具有线性与非线性特征<sup>[11]</sup>, 同时也存在周期性与非平稳性特征。由于单一模型无法适应所有特征结构的数据, 因此拟合结果不佳。为了兼顾溶解氧质量浓度的所有数据特征, 本研究提出基于 SARIMA-VMD-LSSVM 模型的水产养殖溶解氧质量浓度组合预测模型。其中, SARIMA 模型能够较好地体现时间序列的线性与周期性特征, 且结构简单; LSSVM 模型训练速度快, 适用于非线性回归问题。因此, 本研究将 SARIMA 模型与 LSSVM 模型进行串联结合, 以充分发挥各模型

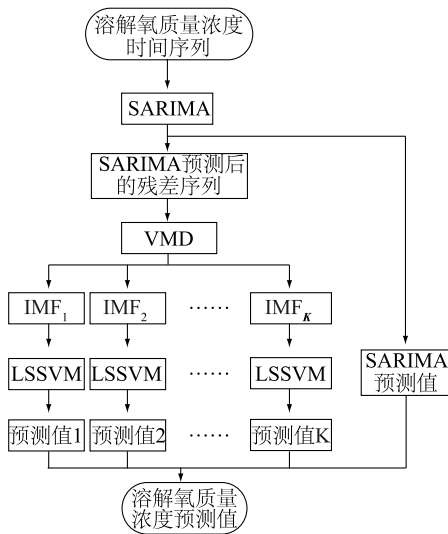


的能力。然而, SARIMA 模型拟合后的残差序列汇聚了溶解氧质量浓度时间序列的非线性特征, 且整体呈现出白噪声特性, 因此在进行 LSSVM 模型训练之前, 本研究引入 VMD 分解算法对残差序列进行进一步的特征提取, 以降低其非线性带来的影响, 增强可预测性。SARIMA-VMD-LSSVM 组合模型的结构如图 1 所示。

SARIMA-VMD-LSSVM 模型的预测步骤具体如下:

步骤 1: SARIMA 建模及预测。根据 SARIMA 模型建模步骤, 确定 SARIMA 模型参数, 并对溶解氧质量浓度的时间序列( $x_t$ )进行拟合与预测, 得到拟合序列 $\{\hat{l}_t\}$ 与预测值( $\hat{l}$ ), 借助公式(22)得到残差序列 $\{e_t\}$ :

$$e_t = x_t - \hat{l}_t \quad (22)$$



SARIMA: 季节性差分自回归滑动平均; VMD: 变分模态分解;  $IMF_1 \sim IMF_k$ : 本征模态分量; LSSVM: 最小二乘支持向量机模型。

图1 季节性差分自回归滑动平均-变分模态分解-最小二乘支持向量机模型 (SARIMA-VMD-LSSVM) 组合预测模型的结构

Fig.1 The structure of combined predictive model seasonal auto-regressive integrated moving average-variational mode decomposition-least square support vector machine (SARIMA-VMD-LSSVM)

步骤 2: VMD 分解。对 $\{e_t\}$ 进行 VMD 分解, 得到  $K$  个残差分量 $\{IMF_1\}$ 、 $\{IMF_2\}$ 、 $\dots$ 、 $\{IMF_k\}$ 。

步骤 3: LSSVM 建模及预测。以 $\{IMF_1\}$ 、 $\{IMF_2\}$ 、 $\dots$ 、 $\{IMF_k\}$ 作为样本集, 对 LSSVM 进行训练和预测, 得到各残差分量的预测值 $\{\hat{IMF}_1\}$ 、 $\{\hat{IMF}_2\}$ 、 $\dots$ 、 $\{\hat{IMF}_k\}$ 。

步骤 4: 结合 SARIMA 预测值和各残差分量 LSSVM 预测值, 得到最终预测值( $\hat{Y}$ ), 即 $\hat{Y} = \hat{l} + \hat{IMF}_1 + \hat{IMF}_2 + \dots + \hat{IMF}_k$ 。

## 1.6 评价指标

对在本研究中模型的预测精度与误差使用平均绝对百分比误差 (MAPE)、均方根误差 (RMSE) 进行评估, 其值越小, 表明模型的性能越好。

$$MAPE = \frac{1}{m} \sum_{i=1}^m \left| \frac{x_i - \hat{x}_i}{x_i} \right| \quad (23)$$

$$RMSE = \sqrt{\frac{1}{m} \sum_{i=1}^m (x_i - \hat{x}_i)^2} \quad (24)$$

## 2 实例分析

### 2.1 数据来源

本研究采用的数据集来自文献[12]中报道的海南省陵水黎族自治县水产养殖基地收集的真正溶解氧质量浓度数据, 并对溶解氧质量浓度数据集进行下采样处理, 每隔 30 min 采集 1 次, 试验数据共计 2 985 条。其中连续 6 d 的溶解氧质量浓度的时间序列如图 2 所示。划分训练集与测试集的比例为 8 : 2, 其中测试集用于进一步验证经训练集训练后的溶解氧质量浓度预测模型的性能。

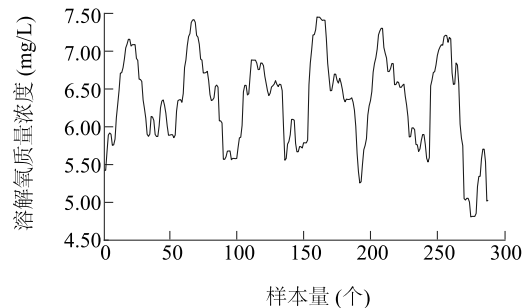


图2 连续 6 d 溶解氧质量浓度的原始时间序列

Fig.2 Original time series of dissolved oxygen mass concentration for six consecutive days

需要说明的是, 本研究采用的预测方式是单步预测, 即通过历史时刻对下个时刻的预测值进行预测。由于样本点过多, 为了更好地展示溶解氧质量浓度序列的处理过程, 下文中每个窗口仅展示其中连续的 500 个样本点。同时, 为了更好地对比最终预测结果与实测值之间的差异, 预测结果的图窗仅展示 200 个连续样本点, 但仍然使用全部测试集作为模型最终的评价标准。

仿真平台具体参数: Window 10 操作系统, 4 核

1.80 GHz CPU, MatlabR 2022b。

## 2.2 SARIMA 线性预测

**2.2.1 平稳化处理** 首先通过 MATLAB 中的 `adftest` 函数对溶解氧质量浓度的时间序列进行单位根检测(ADF),此时 $P=0.054\ 0$  ( $P>0.05$ ),可知此时原溶解氧质量浓度的时间序列为非平稳序列。对溶解氧质量浓度的时间序列进行 1 阶差分处理后, $P=0.001\ 0$  ( $P<0.05$ ),此时溶解氧质量浓度的时间序列整体表现平稳。如图 3 所示,1 阶差分处理后的溶解氧质量浓度的时间序列仍存在周期性变化,因此还需对其进行 1 阶季节性差分处理,结果如图 4 所示。

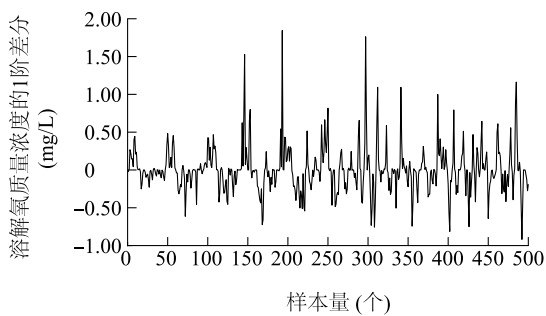


图 3 经 1 阶差分处理后的溶解氧质量浓度的时间序列

Fig.3 Time series of dissolved oxygen mass concentration after first-order differencing treatment

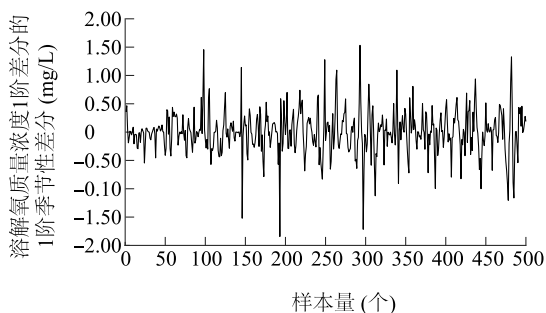


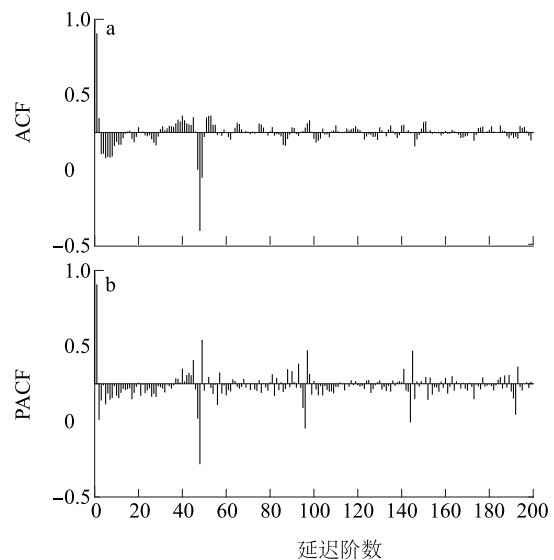
图 4 经 1 阶差分与 1 阶季节性差分处理后的溶解氧质量浓度的时间序列

Fig.4 Time series of dissolved oxygen mass concentration after 1st order differencing treatment and 1st order seasonal differencing treatment

**2.2.2 SARIMA 模型参数定阶** 对溶解氧质量浓度时间序列进行 1 阶差分与 1 阶季节性差分处理,则 $d=0$  或 1,  $D=0$  或 1。溶解氧质量浓度采样频率为 1 次间隔 0.5 h,则 $s=48$ 。由图 5a 可以看出,在延迟阶数为 1 阶、48 阶时溶解氧质量

浓度的时间序列数据与历史数据具有较强的相关性,即 $q=0$  或 1,  $Q=0$  或 1。由图 5b 可以看出,在延迟阶数为 1 阶、2 阶时具有较强相关性,即 $p=0\sim 2$ ,在 48 阶、96 阶、144 阶时溶解氧质量浓度的时间序列数据与历史数据具有较强的相关性,即  $P$  的取值范围为 0~3。

根据参数  $p$ 、 $d$ 、 $q$ 、 $P$ 、 $D$ 、 $Q$  的初步定阶结果,用 SPSS 中的正态贝叶斯信息量准则(BIC)确定最佳取值。正态 BIC 同时考虑模型拟合度与模型复杂度,其数值越小,表明模型越佳。最小正态 BIC 为-3.408,此时对应的最佳拟合度的模型为 SARIMA(2,1,0)(1,1,1)<sub>48</sub>。



ACF:自相关系数;PACF:偏自相关系数。

图 5 溶解氧质量浓度时间序列经 1 阶差分、1 阶季节性差分处理后的 ACF 与 PACF

Fig.5 ACF and PACF of the dissolved oxygen mass concentration time series after first-order differencing treatment and first-order seasonal differencing treatment

**2.2.3 SARIMA 对溶解氧质量浓度的预测** 用 SARIMA(2,1,0)(1,1,1)<sub>48</sub>对溶解氧质量浓度的时间序列进行预测,结果如图 6 所示。同时得到预测后剩余的残差序列,整体呈现白噪声特征(图 7)。

## 2.3 VMD 分解

本研究依据最小峭度准则<sup>[13]</sup>,得到溶解氧质量浓度残差序列的最佳分解层数为 6。由图 8 可以看出,各分量数据虽依旧保持较高频率,但 VMD 分解依旧能够分解出不同频率的信号,且更具有规律性,

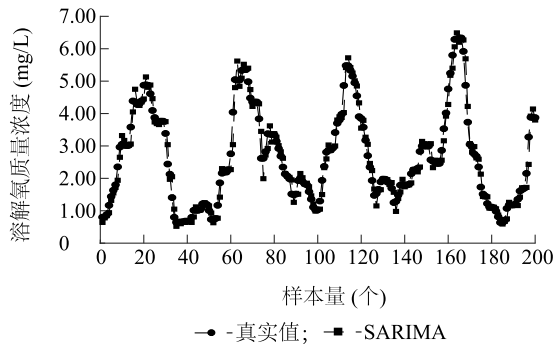


图6 基于季节性差分自回归滑动平均 (SARIMA) 模型的溶解氧质量浓度的预测结果

Fig.6 Forecasting results of dissolved oxygen mass concentration based on seasonal auto regressive integrated moving average (SARIMA) model

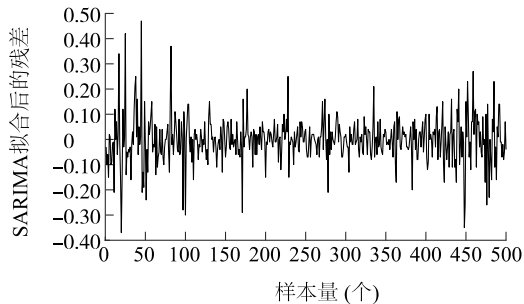


图7 用季节性差分自回归滑动平均 (SARIMA) 模型的拟合后残差序列

Fig.7 Residual series obtained after fitting with seasonal auto regressive integrated moving average (SARIMA) model

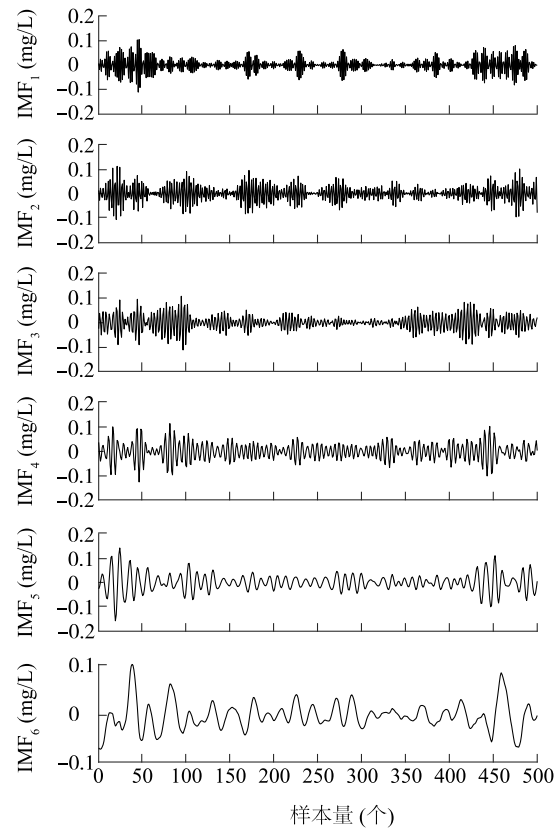
由此提升了残差数据的可预测性。

#### 2.4 IGWO-LSSVM 残差预测

为了避免数据冗余带来的影响,本研究用偏自相关系数(PACF)确定各分量的最佳嵌入维数。如图9所示,前3个延迟(Lag)的偏自相关系数绝对值都大于0.6,说明 $IMF_1$ 时间序列在任意时刻的值都与其前3个历史时刻的值密切相关,因此取最佳嵌入维数( $m$ )=3。同理,得到 $IMF_2 \sim IMF_6$ 的最佳嵌入维数,分别为4、4、4、4、3。

通过IGWO确定各残差分量的最佳LSSVM模型参数,详见表1。对各残差分量进行LSSVM建模预测,并将各分量的预测结果叠加,得到最终残差序列的预测结果,如图10所示。

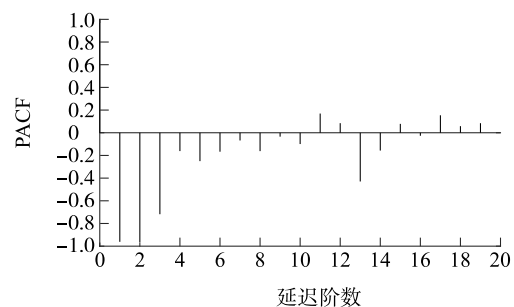
基于上述SARIMA模型,得到线性预测结果,其中通过LSSVM模型预测得到非线性结果,只需将2



$IMF_1 \sim IMF_6$ :溶解氧质量浓度残差序列VMD后的本征模态分量。

图8 溶解氧质量浓度的残差序列变分模态分解(VMD)结果

Fig.8 Variational mode decomposition (VMD) results of residual series of dissolved oxygen mass concentration



PACF:偏自相关系数。

图9 溶解氧质量浓度残差第1模态分量( $IMF_1$ )的偏自相关系数值(PACF)

Fig.9 The partial autocorrelation coefficient (PACF) of the residual first mode component ( $IMF_1$ ) of dissolved oxygen mass concentration

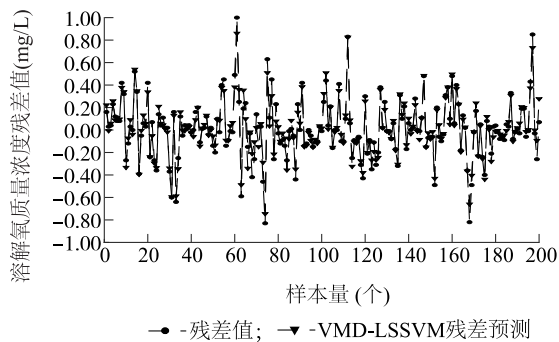
个分量进行集成,便可得到最终溶解氧质量浓度的预测值,结果见图11。

表 1 改进的灰狼算法 (IGWO) 优化后的最小二乘支持向量机模型 (LSSVM) 参数

Table 1 Parameters of least square support vector machine (LSSVM) model optimized by improved gray wolf algorithm (IGWO)

分量	$\gamma$	$\sigma$
IMF <sub>1</sub>	930.93	92.86
IMF <sub>2</sub>	215.01	99.85
IMF <sub>3</sub>	39.41	112.99
IMF <sub>4</sub>	430.97	99.76
IMF <sub>5</sub>	774.14	48.16
IMF <sub>6</sub>	99.73	11.30

IMF<sub>1</sub>~IMF<sub>6</sub>:溶解氧质量浓度残差序列的本征模态分量; $\gamma$ :惩罚参数; $\sigma$ :径向基宽度参数。



VMD、LSSVM 见图 1 注。

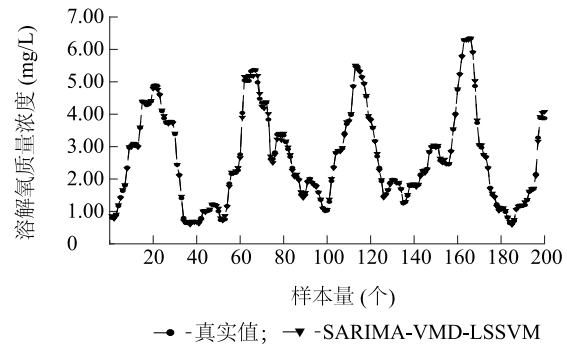
图 10 溶解氧质量浓度残差序列预测结果

Fig.10 Forecasting results of residual series of dissolved oxygen mass concentration

## 2.5 模型对比分析

2.5.1 不同模型的对比分析 为了验证本研究所提模型的预测性能,将 SARIMA-VMD-LSSVM 模型与单一模型 SARIMA、LSSVM 及组合模型 VMD-LSSVM 进行对比分析。如图 12 所示,4 种模型都能较好地预测溶解氧质量浓度的变化趋势,但是也能明显看出,SARIMA 与 LSSVM 模型在转折点的精度较差,这是由于单一模型难以学习到转折点的信息。此外还可以看出,VMD-LSSVM 与 SARIMA-VMD-LSSVM 的拟合程度较好。

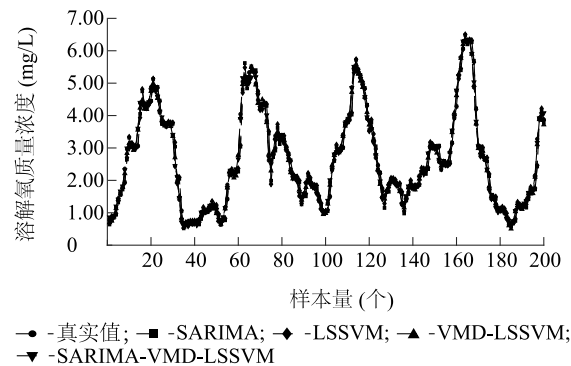
进一步计算出 4 种模型的平均相对误差 (MAPE)、均方根误差 (RMSE) 并记录某一时刻所需的运行时间。如表 2 所示,与 SARIMA、LSSVM 模型相比,VMD-LSSVM、SARIMA-VMD-LSSVM 组合模型的 MAPE、RMSE 明显降低,其中 SARIMA-VMD-LSSVM 模型相较于 VMD-LSSVM 模型有更好的预测精度,说明本研究提出的模型能更加充分地挖掘溶解



SARIMA:季节性差分自回归滑动平均模型;LSSVM:最小二乘支持向量机模型;VMD:变分模态分解。

图 11 季节性差分自回归滑动平均-变分模态分解-最小二乘支持向量机模型 (SARIMA-VMD-LSSVM) 溶解氧质量浓度预测结果

Fig.11 The dissolved oxygen mass concentration prediction results based on seasonal auto regressive integrated moving average-variational mode decomposition-least square support vector machine (SARIMA-VMD-LSSVM) model



SARIMA:季节性差分自回归滑动平均模型;LSSVM:最小二乘支持向量机模型;VMD:变分模态分解。

图 12 不同预测模型与真实溶解氧质量浓度的预测结果对比

Fig.12 Comparison between predicted results of different prediction models and real dissolved oxygen mass concentration

氧质量浓度的数据特征。由表 2 还可以看出,SARIMA-VMD-LSSVM 组合模型整体预测时间仅有 12.957 2 s,可见该组合模型适用于溶解氧质量浓度的短时间预测。

2.5.2 不同优化算法的对比 为了验证改进灰狼算法的有效性,本研究选用粒子群优化算法 (PSO)、蚁群优化算法 (ALO)、灰狼优化算法 (GWO)、改进灰狼优化算法 (IGWO) 分别对 LSSVM 惩罚参数和径向基宽度参数进行寻优,并用优化后的 LSSVM 模



表 2 季节性差分自回归滑动平均-变分模态分解-最小二乘支持向量机模型 (SARIMA-VMD-LSSVM) 与单一模型预测精度的对比

Table 2 Comparison of prediction accuracy between seasonal autoregressive integrated moving average-variational mode decomposition-least square support vector machine (SARIMA-VMD-LSSVM) model and single model

模型	MAPE	RMSE	时间 (s)
SARIMA	0.081 6	0.292 6	0.385 0
LSSVM	0.084 4	0.293 2	1.883 3
VMD-LSSVM	0.034 8	0.100 8	8.805 8
SARIMA-VMD-LSSVM	0.022 6	0.078 7	12.957 2

SARIMA:季节性差分自回归滑动平均模型;LSSVM:最小二乘支持向量机模型;VMD:变分模态分解;MAPE:平均相对误差;RMSE:均方根误差。

型对 SARIMA 拟合后的溶解氧质量浓度残差序列进行预测。由于残差序列中存在值为 0 的情况,因此仅采用 RMSE 作为评价指标。在本研究中,设立种群数量为 20 个,最大迭代次数为 100 次。

对用 4 种优化方法预测溶解氧质量浓度残差序列的结果与真实残差值(表 3)进行比较发现,使用固定超参数的 LSSVM 模型的 RMSE 明显较大,拟合程度不高,说明不同残差分量所需的 LSSVM 模型各不相同。加入优化算法后, RMSE 减小,整体预测效果提升。由于 LSSVM 超参数较少,仅为二维,因此优化后的 LSSVM 模型预测效果相近,但仍然能看出 IGWO-LSSVM 模型的 RMSE 低于其他模型,其原因是对于其他模型而言,某些残差分量难以找到最优

LSSVM 模型,从而陷入局部最优。

表 3 残差序列的 4 种最小二乘支持向量机 (LSSVM) 模型的预测结果

Table 3 Prediction results of residual series based on four least square support vector machine (LSSVM) models

模型	RMSE
LSSVM	0.125 8
PSO-LSSVM	0.054 7
ACO-LSSVM	0.054 5
GWO-LSSVM	0.054 6
IGWO-LSSVM	0.052 4

LSSVM:最小二乘支持向量机;PSO:粒子群优化算法;ACO:蚁群优化算法;GWO:灰狼优化算法;IGWO:改进灰狼优化算法;RMSE:均方根误差。

为了进一步验证 IGWO 的有效性,选取 Sphere 单峰函数、Ackely 多峰函数和 Griewank 多峰函数进行测试<sup>[14]</sup>,并将测试结果与 PSO、ALO、GWO 算法所得结果进行比较,设各算法的种群规模为 50 个,最大迭代次数为 100 次。由表 4 可以看出,IGWO 在不同测试函数中整体上有相对较好的寻优结果,其不同维度下的 Sphere 函数的寻优精度较 PSO、ALO 和 GWO 算法高 1~14 个数量级。在不同维度下,Ackely 函数 IGWO 算法的寻优精度比 PSO、ALO 和 GWO 算法高 1~17 个数量级。在 Griewank 函数中,IGWO 算法在 10~100 维范围均获得理论最优值 0;在 300 维,其理论最优值相较于 PSO、ALO、GWO 至少提升了 7 个数量级。由此可见,IGWO 具有更好的全局搜索能力与收敛速度。

表 4 不同测试函数的寻优精度

Table 4 Optimization accuracy of different test functions

函数	范围	最优解	算法	10 维	30 维	50 维	100 维	300 维
Sphere	[-100,100]	0	PSO	$3.14\times10^{-2}$	$1.44\times10^1$	$6.91\times10^1$	$5.99\times10^2$	$1.35\times10^4$
	[-100,100]	0	ALO	$3.19\times10^{-6}$	$6.38\times10^2$	$5.48\times10^3$	$2.15\times10^4$	$3.42\times10^5$
	[-100,100]	0	GWO	$4.96\times10^{-10}$	$2.19\times10^3$	$4.71\times10^{-2}$	$1.14\times10^1$	$2.69\times10^2$
	[-100,100]	0	IGWO	$1.90\times10^{-16}$	$7.41\times10^{-7}$	$6.88\times10^4$	$1.72\times10^{-2}$	4.12
Ackely	[-32,32]	$8.88\times10^{-16}$	PSO	0.49	4.57	6.15	8.65	$1.61\times10^1$
	[-32,32]	$8.88\times10^{-16}$	ALO	1.16	$1.05\times10^1$	$1.27\times10^1$	$1.44\times10^1$	$1.60\times10^1$
	[-32,32]	$8.88\times10^{-16}$	GWO	$1.12\times10^{-6}$	$3.60\times10^{-3}$	$4.07\times10^{-2}$	$5.33\times10^{-1}$	7.10
	[-32,32]	$8.88\times10^{-16}$	IGWO	$3.22\times10^{-9}$	$4.84\times10^{-5}$	$6.74\times10^{-4}$	$2.30\times10^{-2}$	$4.44\times10^{-16}$
Griewank	[-600,600]	0	PSO	$6.23\times10^{-1}$	4.93	$1.10\times10^1$	$3.39\times10^1$	$1.22\times10^2$
	[-600,600]	0	ALO	$3.70\times10^{-2}$	$1.34\times10^1$	$4.50\times10^1$	$2.02\times10^2$	$1.08\times10^3$
	[-600,600]	0	GWO	$3.26\times10^{-11}$	$5.04\times10^{-4}$	$5.60\times10^{-2}$	1.09	$1.84\times10^1$
	[-600,600]	0	IGWO	0	0	0	0	$6.20\times10^{-6}$

PSO:粒子群优化算法;ALO:蚁群优化算法;GWO:灰狼优化算法;IGWO:改进的灰狼优化算法。

### 3 结 论

为了充分利用溶解氧质量浓度的数据特征,进一步提升溶解氧质量浓度的预测精度,本研究将 SARIMA 模型、VMD 模型与 LSSVM 模型相结合,同时引入改进的灰狼优化算法(IGWO),构建了用于预测水产养殖中溶解氧质量浓度的模型,得出以下结论:(1)通过改进 GWO 的收敛因子,并结合麻雀优化算法,有效提升了灰狼优化算法的寻优能力。避免了 LSSVM 参数设定的盲目性,减少了人工选择参数的主观影响。(2)将 SARIMA 模型、VMD 模型与 LSSVM 模型 3 种模型结合,有效地表征了溶解氧质量浓度时间序列的周期性、线性、非线性与非平稳性特征。结果显示,SARIMA-VMD-LSSVM 模型的评价指标优于其单一模型与 VMD-LSSVM 模型,表明该组合模型更能有效挖掘溶解氧质量浓度的数据特征。(3)本研究中组合模型的主要预测原理为 SARIMA、LSSVM,两者都有较快的计算速度。试验结果表明,组合模型总用时为 12.957 2 s,适合用于溶解氧质量浓度的短时预测。本研究主要针对溶解氧质量浓度单因子进行研究,未来将尝试将其与其他水质因子结合,以进一步提升预测精度。此外,还可以结合嵌入式、网络通信和自动控制等技术,实现溶解氧质量浓度的在线预测与管控。

#### 参考文献:

- [1] 谢雨茜,李 路,朱 明,等. 基于 EMD 与 K-means 的 ILSTM 模型在池塘溶解氧预测中的应用[J]. 华中农业大学学报(自然科学版),2022,41(3):200-210.
- [2] PARK K D, KANG D H, SO Y, et al. Water quality prediction using the ARIMA time series analysis model in the Nakdong River estuary[C]. San Francisco: AGU Fall Meeting Abstracts,2019.
- [3] 张梦迪,徐 庆,刘振鸿,等. 基于动态滑动窗口 BP 神经网络的水质时间序列预测[J]. 环境工程技术学报,2022,12(3):809-815.
- [4] 龚怀瑾,毛 力,杨 弘. 基于变尺度混沌 QPSO-LSSVM 的水质溶氧预测建模[J]. 计算机与应用化学,2013,30(3):315-318.
- [5] 刘 晨,李 莎,丛孙丽,等. 基于 EEMD 和萤火虫算法优化 SVM 的溶解氧预测[J]. 计算机仿真,2021,38(1):359-365.
- [6] 白雯睿,杨毅强,郭 辉,等. 基于 VMD-CNN-LSTM 的珠江流域水质多步预测模型研究[J]. 四川轻化工大学学报(自然科学版),2022,35(4):66-74.
- [7] DRAGOMIRETSKIY K, ZOZZO D. Variational mode decomposition[J]. IEEE Transactions on Signal Processing,2014,62(3):531-544.
- [8] LUO H, PAAL S G. Metaheuristic least squares support vector machine-based lateral strength modelling of reinforced concrete columns subjected to earthquake loads[J]. Structures,2021,33:748-758.
- [9] MIRJALILI S, MIRJALILI S M, LEWIS A. Grey wolf optimizer[J]. Advances in Engineering Software,2014,69(3):46-61.
- [10] 高 兵,郑 雅,秦 静,等. 基于麻雀搜索算法和改进粒子群优化算法的网络入侵检测算法[J]. 计算机应用,2022,42(4):1201-1206.
- [11] 胡衍坤,王 宁,刘 枢,等. 时间序列模型和 LSTM 模型在水质预测中的应用研究[J]. 小型微型计算机系统,2021,42(8):1589-1573.
- [12] LIU J T, YU C, HU Z H, et al. Accurate prediction scheme of water quality in smart mariculture with deep Bi-S-SRU learning network[J]. Ieee Access,2020,8(99):24784-24798.
- [13] 郑 圆,胡建中,贾民平,等. 一种基于参数优化变分模态分解的滚动轴承故障特征提取方法[J]. 振动与冲击,2020,39(21):195-202.
- [14] 崔东文,袁树堂. 基于 WPD-AHA-ELM 模型的水质时间序列多步预测[J]. 三峡大学学报(自然科学版),2023,45(1):6-13.

(责任编辑:徐 艳)