

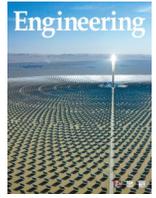


ELSEVIER

Contents lists available at ScienceDirect

Engineering

journal homepage: www.elsevier.com/locate/eng



Research
Artificial Intelligence—Article

一种基于多因素分析和多模型集成的海洋溶解氧浓度时间序列预测混合神经网络模型

刘辉*, 杨睿, 段铸, 吴海平

Institute of Artificial Intelligence and Robotics (IAIR), Key Laboratory of Traffic Safety on Track of Ministry of Education, School of Traffic and Transportation Engineering, Central South University, Changsha 410075, China

ARTICLE INFO

Article history:

Received 4 January 2020
Revised 25 July 2020
Accepted 18 October 2020
Available online 19 March 2021

关键词

溶解氧浓度预测
时间序列多步预测
多因素分析
经验小波变化分解
多模型优化集成

摘要

溶解氧是水产养殖的重要指标,准确预测溶解氧浓度可有效提高水产品质量。本文提出了一种新的溶解氧混合预测模型,该模型包括多因素分析、自适应分解和优化集成三个阶段。首先,考虑到影响溶解氧浓度的因素复杂繁多,采用灰色关联度法筛选出与溶解氧关系最密切的环境因素,多因素的考虑使得模型融合更加有效。其次,运用经验小波变换方法自适应地将溶解氧、水温、盐度和氧饱和度等序列分解为子序列。然后,利用5个基准模型对经验小波变换分解出的子序列进行预测,这5个子预测模型的集成权重通过粒子群优化和引力搜索算法计算得出。最后,通过加权分配得到溶解氧多因素集成模型。来自太平洋岛屿海洋观测系统在希洛的WQB04站收集的时间序列数据验证了该模型的性能。实验的评价指标包括Nash-Sutcliffe效率系数、Kling-Gupta效率系数、平均绝对百分比误差、误差标准差和决定系数。实例分析表明:①所提出的模型能够获得优异的溶解氧预测结果;②该模型优于文中其他对比模型;③预测模型可用于分析溶解氧变化趋势,便于管理者能够做出更好的决策。

© 2021 THE AUTHORS. Published by Elsevier LTD on behalf of Chinese Academy of Engineering and Higher Education Press Limited Company. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

1. 引言

溶解氧 (DO) 在沿海城市水产养殖和海洋生态的可持续发展中发挥着重要作用。如今,随着工业的快速发展,大量的温室气体排放正在日益加剧全球变暖[1],导致海水温度升高,DO从海水中逸出[2]。如果海水中的DO浓度过低,鱼虾等海洋生物将会缺少氧气,从而导致生长缓慢甚至窒息[3]。对于近海养殖户来说,大规模的水产品死亡将带来巨大的经济损失[4]。更严重的是,DO很难扩散到海洋深处,缺氧会迫使水生生物改变它们的栖

息地[5],海洋生态系统可能因此发生重大变化。一旦生态系统的良性循环被打破,海洋生物的生存将受到极大的威胁[6]。准确的DO含量预测对于海洋生态平衡和沿海城市的水产养殖业发展非常必要。它可以提供DO的未来趋势变化分析,并通过帮助决策者提前采取有效措施来促进管理[7-8]。受营养物质、气候、生态环境和水生生物生命活动等因素的影响,海水DO含量呈非线性变化,具有相当大的时滞性,这一特性使得DO难以被预测。此外,传感器获取的数据在传输过程中容易丢失,且此类数据常包含的不确定性异常值也对预测的准确性提出了更大的

* Corresponding author.

E-mail address: csuliuhui@csu.edu.cn (H. Liu).

挑战[9]。

1.1. 相关工作

近年来,人工智能(AI)技术的兴起催生了AI模型浪潮。在此基础上,DO预测模型可分为两类:物理模型和数据驱动模型。数据驱动模型包括AI模型和统计模型。为了全面评估DO模型的最新成果,我们总结了相关统计模型和AI模型的文献。统计模型具有计算简便的优点。此外,统计模型可以通过挖掘DO序列中的潜在关系来预测未知情况。在实际应用中,DO序列是非线性和时间延迟的。Faruk [10]使用自回归综合移动平均(ARIMA)模型进行DO预测,以处理时间序列的不稳定性和非线性。在Li等[11]的实验中,他们提出灰色模型来预测DO的趋势项。Huan等[12]通过贝叶斯证据框架得到模型的最优参数值。Khan等[13]对比了新贝叶斯回归模型和新自回归修正模糊线性回归方法。对比结果表明,基于模糊数的方法能较好地捕捉城市河流环境中DO的变化。Khan等[14]通过构建多元线性回归(MLR)模型来预测DO浓度的变化,DO变化的不确定性被有效地表征和传播。在Kisi和Parmar [15]后续的研究中,MLR模型被用作实验比较,以探索其他预测模型的性能。总的来说,这些基于统计的模型获得了理想的DO预测效果,但预测的准确性仍有待提高。

与基于统计的模型相比,AI组件模型具有更好的预测性能。本节的核心重点是描述带有AI组件的模型。常用的AI算法包括多层感知器(MLP) [16]、径向基神经网络(RBNN) [17]、反向传播神经网络(BPNN)、极限学习机(ELM) [18]、最小二乘支持向量机(LSSVM) [19]和模糊神经网络(FNN) [20-21]等[22]。由于不同参数的选择对AI模型的性能有很大的影响,因此有必要通过优化算法来确定最优参数。Ren等[23]使用遗传算法(GA)来优化FNN的参数。此外,粒子群优化算法(PSO) [24]、柯西粒子群优化算法(CPSO) [25]、萤火虫算法(FFA) [26]等优化算法被用于进行智能模型[27]的参数选择。随着深度学习技术的快速发展,其以优异的适应能力得到了广泛的研究。Ma等[28]通过深度矩阵因子分解解决了稀疏矩阵的问题。Ren等[29]使用深度置信网络(DBN)强大的特征提取和函数表示能力处理高度复杂的非线性DO时间序列数据。然而,考虑到DO变化的复杂规律和许多其他因素,简单的机器学习基础预测器和参数优化方法不足以实现准确的DO预测。为了进一步提高AI模型的预测性能,越来越多的学者正在使用特征选择、分解和集成方法进行DO预测的最新研究。有关这类研究的

进一步详情如下所述。

(1) 海洋和淡水环境中DO浓度的变化不是孤立的,它受到许多因素的影响,包括温度、浊度、pH、叶绿素和电导率等[30],这些因素的变化对DO浓度有着无形的影响。为了有效地利用这些因素,还需要研究其具体影响。Shi等[22]利用聚类方法对水质时间序列进行分割,最终提高了DO的预测精度。Ren等[23]通过相关分析整理了与DO呈正相关和负相关的因素。结果表明,将与DO强相关的因素引入模型后,该模型展现出较好的预测效果[31]。从DO产生和消耗的角度来看,应该将这些影响因素有机地结合起来,而不是直接粗暴地输入预测模型。这样考虑是不全面的,多因素预测序列的直接叠加可能会导致模型的DO预测精度降低。需要分析各影响因素的权重分配,从而确定最佳组合方法[32]。因此,探索DO预测的多因素分析方法具有重要的研究价值[33]。

(2) 原始DO序列的非平稳性质不利于预测。DO序列的极端波动使其难以被预测,分解方法的出现很大程度上解决了这个问题。分解后的子序列具有独立的振荡分量,更加稳定,也更容易被预测[34-35]。为了将这种理论效果应用到实际问题中,许多AI模型的数据预处理分解方法已经被提出。广泛应用的分解算法包括集合经验模态分解(EEMD) [12]、离散小波变换(DWT)和变分模态分解(VMD) [36]。分解后的DO序列可预测性更强。然而,目前主流的分解方法要求使用者根据经验来确定分解层数。这种做法不可避免地引入了人为误差[37]。为了提高预测模型的性能,需要研究一种能自适应确定分解层数的分解算法。

(3) 经过多年的研究和发展,一些集成模型可以用于DO的智能预测。集成模型可以综合多种预测模型的优点,达到更好的预测效果[35]。通过分层集成,Zhu等[38]成功解决了无法同时实现准确预测高、低层次不同浓度成分的问题。他们采用迭代逐步多元线性回归(ISMLR)对数据进行预处理,并运用人工神经网络(ANN)和MLP进行分层预测。最后,利用折衷规划(CP)方法对优化结果进行评估和选择。Kisi等[39]提出了一种新的贝叶斯模型平均(BMA)方法作为集合模型。通过与ELM、ANN、自适应神经模糊推理系统(ANFIS)、分类回归树(CART)和MLR进行对比实验,发现BMA集成方法对DO预测非常有效。集成模型的实质是利用多个模型的互补优势来提高预测精度和应用范围。不幸的是,上述集成方法仅使用多个模型进行分层预测,然后简单地叠加结果。直接集成多个模型会大大增加集成模型的复杂性,容易导致过拟合。与此同时,单个模型的缺

陷是其他模型无法弥补的。因此，选择合理的模型集合是非常重要的。不同的机器学习预测模型具有各自的特点，能够处理不同状态下的时间序列。如果能够将它们有机地结合起来，那么集成模型将实现优势互补。适当的权重分配可以掩盖基准模型的预测缺陷，增强基准模型的优势。因此，有必要探索科学的多模型集成方法。

上述DO预测模型已被证明是有效的。然而，仍然有一些科研空白需要被填补以进一步提高DO的预测性能。

表1 [10-19,21-30,36,38-40]总结了使用数据驱动方法进行DO预测的最新研究成果。

1.2. 本研究的创新性

综上所述，在DO预测研究中，对于多因素分析、自适应分解分析和科学集成分析的对比研究较少。针对上述局限性，本文提出了一种新的DO预测模型：MF-RNN-EWT-BEGOE模型。MF代表所提出的多因素分析方法，

表1 当前最先进的数据驱动DO预测方法总结

Category	Research content	Reference	Year of publication	Contribution	
Statistics-based models	ARIMA model	[10]	2010	The ARIMA method was applied to study stationary and nonstationary time series	
	Grey model	[11]	2018	The grey model was proposed to forecast the trend term of DO	
	Bayesian model	[12]	2018	A bayesian evidence framework was applied to obtain optimal parameter values	
		[13]	2017	The performance of the new bayesian regression model and the new autoregressive modified fuzzy linear regression method in DO prediction were compared	
MLR model	[14]	2013	The MLR model was used to characterize and propagate the uncertainty of DO changes		
		2016	The MLR model was utilized to evaluate the performance of multiple models, such as LSSVM, multivariate adaptive regression splines (MARS), and the M5 model tree (M5Tree)		
	AI-component models	Basic prediction models	[21]	2016	The improved fuzzy neural network method was used to predict low DO events
			[18]	2020	The prediction performance of ELM and MLP on daily DO concentration was compared
		[17]	2012	The MLP and RBNN models based on ANNs were used as a comparison with MLR models based on statistical methods	
		[30]	2013	ANNs, ANFIS, and gene expression programming (GEP) were used to compare the prediction performance of DO concentration	
		[16]	2014	The <i>k</i> -means clustering method and MLP were combined to improve the prediction accuracy of the model	
		[19]	2018	Several data-driven methods were used for modeling a comparison of daily DO concentration prediction, such as LSSVM, MARS, and M5Tree	
		Optimization method for neural networks	[23]	2018	The GA algorithm was adopted to optimize the center and width of the FNN
	[24]		2018	The best parameters of the BPNN were determined by the PSO algorithm	
[27]	2019		The dual-scale DO soft-sensor modeling method was applied to improve forecasting performance		
Deep learning methods	[25]	2014	The CPSO was employed to optimize the kernel parameter and the regularization parameter of the LSSVR model		
	[26]	2017	The FFA was used to optimize the three parameters of MLP.		
	[28]	2020	Deep matrix factorization was adopted for DO forecasting		
Optimization method for decomposition algorithms	[29]	2020	A deep belief network was applied to improve forecasting performance		
	[40]	2019	The decomposed series was reconstructed by the sample entropy (SE) method to make forecasting easier		
Ensemble methods	[38]	2018	ISMLR + MLR + ANN was used to hierarchically predict high and low concentration data		
	[39]	2020	By a comparison with multiple models such as ELM, ANNs, ANFIS, CART, and MLR, the application potential of the BMA ensemble model was verified		
Feature-selection methods	[22]	2019	<i>k</i> -medoids clustering based on segmentation was applied to select features		
	[23]	2018	Correlation analysis was adopted for feature selection		
Decomposition methods	[36]	2019	DWT and VMD were adopted to decompose the original data		
	[12]	2018	EEMD was utilized to improve DO forecasting performance		

RNN代表复制因子神经网络的离群点检测方法，EWT代表经验小波变换，BEGOE代表BFGS-ENN-GRNN-ORELM-ELM。该模型可分为多因素分析、自适应分解和基于优化的集成三个阶段。4个因素下的各子层采用同一类型的基准模型进行预测，最后合并输出得到模型的预测结果。在优化集成阶段，采用粒子群算法和引力搜索算法(PSOGSA)对Broyden-Fletcher-Goldfarb-Shanno(BFGS)模型、elman神经网络(ENN)模型、广义回归神经网络(GRNN)模型、离群鲁棒极限学习机(ORELM)模型和ELM模型进行优化，得到BEGOE模型。第2节提供了所提出的模型的具体细节。为了充分了解模型各部分的作用，我们进行了很多对比实验。

以下是本研究的创新和贡献。

(1) 设计了一种混合三阶段海洋DO预测集成模型。该模型考虑了多因素分析、自适应分解分析和科学集成分析，从而填补了1.1节中所述的三个研究空白。多因素分析合理地考虑了温度、盐度、浊度、叶绿素和氧饱和度对海洋DO浓度的影响。通过降低DO序列的非平稳性，EWT分解自适应地提高了时间序列的可预测性。采用PSOGSA优化方法对多个基准模型的权重进行优化。基准模型的互补进一步提高了混合模型的性能。这三个模块协同工作，实现了对DO的准确预测。

(2) 本文合理地考虑了影响海洋中DO浓度的几个因素。多因素分析方法不仅考虑了单因素和多因素对模型精度的影响，还考虑了多因素之间的关联性。这样可以获得融合性能更好的混合模型，避免单一因素的单边影响和局部劣势。这种方法也弥补了直接多因素输入的不足，因为直接地多因素输入可能会产生负面影响。该方法可以从多方面考虑导致DO浓度变化的原因。

(3) 自适应数据预处理方法在分解效果方面取得了理想的结果。与主流分解方法相比，自适应分解避免了人为选择分解层数所引入的误差，在这种方法下，可以更容易地预测分解层。此外，EWT分解方法丢弃分解模式中的残差信号，从海洋DO时间序列中提取有意义的信息。EWT方法还在一定程度上补偿了一些分解算法对噪声和采样的敏感性。数学理论的支持使该机器学习算法的性能得到有效的发展。

(4) 多基准模型的有机结合填补了集成模型的研究空白。采用元启发式优化算法对BFGS、ENN、GRNN、ORELM和ELM进行权重分配。集成方法可以在实现多基准模型优点的同时减少单一模型的不利影响。该集成模型解决了一些预测模型鲁棒性差的问题，充分体现了性能互补的优越性。因此，基于元启发式优化算法的多模型互补

集成框架具有重要的应用价值。

2. 所提出的多因素预测模型

如前所述，所提出MF-RNN-EWT-BEGOE模型的建模过程可分为三个阶段：多因素特征提取、EWT分解和多模型优化集成。图1展示了建模流程。值得注意的是，该模型采用了时间序列预测中常用的多输入多输出策略(MIMOS)[41]。与递归策略(RS)相比，MIMOS的累积误差更小[42]。图2是RS和MIMOS的原理图。

2.1. 第一阶段——多因素分析方法

2.1.1. 子阶段1.1——复制因子神经网络异常点检测

离群值剔除是DO数据处理的一个重要环节。预处理使数据看起来更有序，更容易被预测模型学习，而无需关注异常值的来源。值得注意的是，只有训练集被修正，而测试集保持不变，才是验证模型有效性的一种更有说服力的方式。RNN用于检测原始海洋DO系列的异常值，这是一个多层前馈神经网络[43]。

假设RNN的 k 层中第 i 个神经元的输出为 $S_k(I_{ki})$ 。 I_{ki} 表示 k 层中第 i 个神经元的输入， S_k 表示第 k 层中使用的激活函数。神经元的输入可以表示如下：

$$\theta = I_{ki} = \sum_{j=0}^{L_{k-1}} w_{kij} Z_{(k-1)j} \quad (1)$$

式中， w_{kij} 是连接 k 层中第 i 个神经元和 $k-1$ 层中第 j 个神经元的权重； Z_{kj} 是 k 层中第 j 个神经元的输出； L_k 是 k 层中的神经元个数。在本例中，当 $k=2,4$ 时，激活函数可以表示如下：

$$S_k(\theta) = \tanh(a_k \theta) \quad (k=2,4) \quad (2)$$

式中， a_k 为1。对于 $k=3$ 的中间层，其激活函数可以表示为一类阶跃函数，激活函数可以表示如下：

$$S_3(\theta) = \frac{1}{2} + \frac{1}{4} \sum_{j=1}^{N-1} \tanh \left[a_3 \left(\theta - \frac{j}{N} \right) \right] \quad (3)$$

式中， N 代表步数； a_3 代表晋升到下一层的速率。

一般情况下，进入第三层的连续数据通过RNN的 S_3 逐步激活函数被转换为一批离散值。这相当于将序列样本映射到 N 个集群。最后，RNN可以计算单个离群值和小簇离群值，从而完成对WQB04站点海洋数据的异常值检测。附录A中的算法S1给出了RNN离群点检测方法的伪代码。

2.1.2. 子阶段1.2——灰色关联度法的多因素分析

为了筛选出对DO预测最有利的因素，采用灰色关联

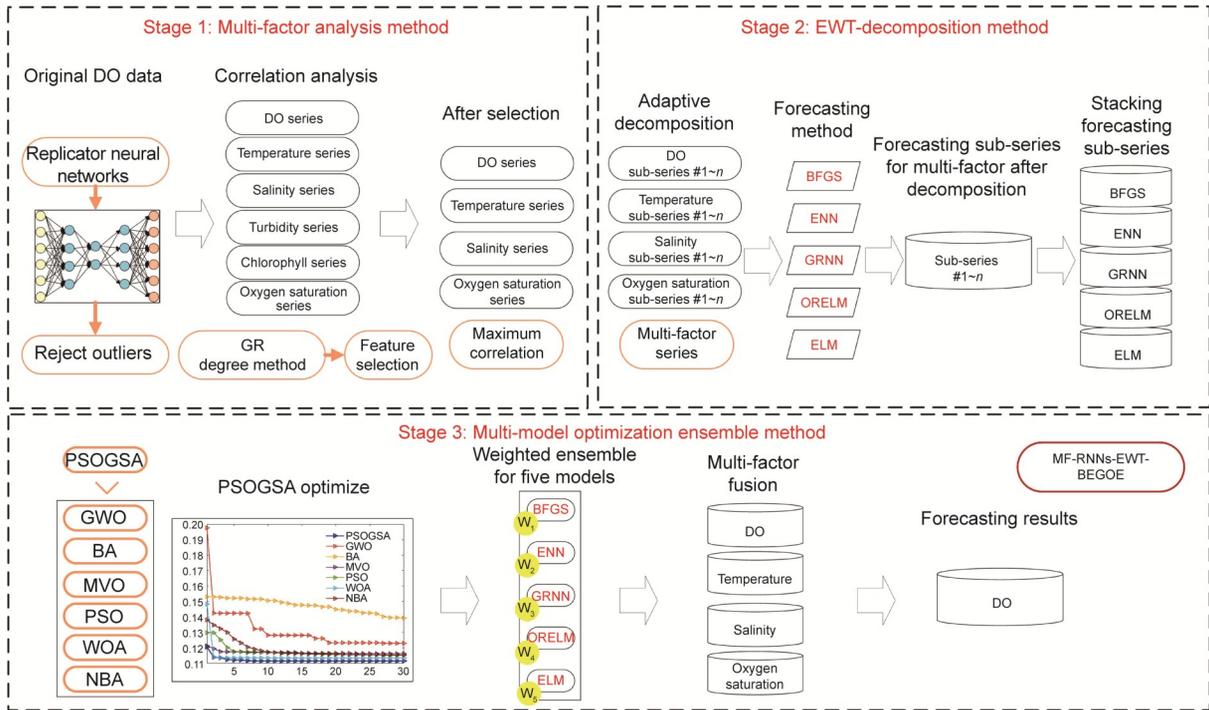


图1. 所提出的三阶段模型的建模流程。GR: 灰色关联; GWO: 灰狼优化算法; BA: 蝙蝠算法; MVO: 多元宇宙优化算法; WOA: 鲸鱼优化算法; NBA: 新蝙蝠算法。

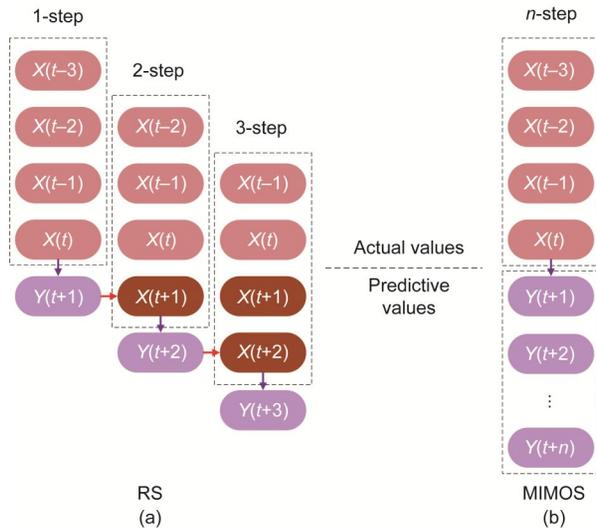


图2. 多步预测策略示意图。

度法选取原始多因素序列的特征。在本例中，灰色关联度法采用于计算温度、盐度、浊度、叶绿素、氧饱和度等环境因素的相关性。时间序列之间的灰色关联度越大，说明两个序列之间的距离越近。利用特征选择确定与DO关联度最大的序列。灰色关联度的具体计算方法如下[44-45]：

假设 $R_0=(r_0(1), r_0(2), r_0(3), \dots, r_0(n))$ 是一组系统特性行为， $R_i=(r_i(1), r_i(2), r_i(3), \dots, r_i(n))$ ($1 < i < N$) 是一组相关行为。因此，两个序列之间的相关系数可以通过以下计算方法得到：

$$\alpha_i(k) = \frac{\min_i \min_k |r_0(k) - r_i(k)| + \beta \max_i \max_k |r_0(k) - r_i(k)|}{|r_0(k) - r_i(k)| + \beta \max_i \max_k |r_0(k) - r_i(k)|} \quad (4)$$

式中， $r_0(k)$ 和 $r_i(k)$ 的绝对差值为点 k ； $\min_i \min_k |r_0(k) - r_i(k)|$ 是两层之间的最小绝对差异； $\max_i \max_k |r_0(k) - r_i(k)|$ 是两层之间的最大绝对差异； β (辨率因子) 通常为 0.5。

序列 r_0 与 r_i 之间的关系 ρ 如下：

$$\rho(r_0, r_i) = \frac{\sum_{k=1}^n \alpha_i(k)}{n}, \quad (i = 1, 2, 3, \dots, n) \quad (5)$$

2.2. 第二阶段——EWT分解方法

在多因素分析方法之后，将EWT分解用于自适应地分解DO、水温、盐度、氧饱和度序列。图3为时间序列数据的分解方案。在训练阶段，对 $[X_1^s, \dots, X_{1247}^s, X_{1248}^s]$ 进行分解， X 表示DO数据的样本点，第 s 个预测器的标记为 $\#s$ 。在预测阶段，将数据划分为验证集 $[X_{1249}^s, \dots, X_{1343}^s, X_{1344}^s]$ 和测试集 $[X_{1345}^s, \dots, X_{1439}^s, X_{1440}^s]$ ，验证集用于集成模型优化，测试集用于测试所提出模型的性能。在此过程中，利用BFGS、ENN、GRNN、ORELM和ELM模型对子序列进行预测，得到序列1249~1440的预测结果。分解过程包括1步至3步的预测结果。附录A中的算法S2给出了EWT分解方法的伪代码。

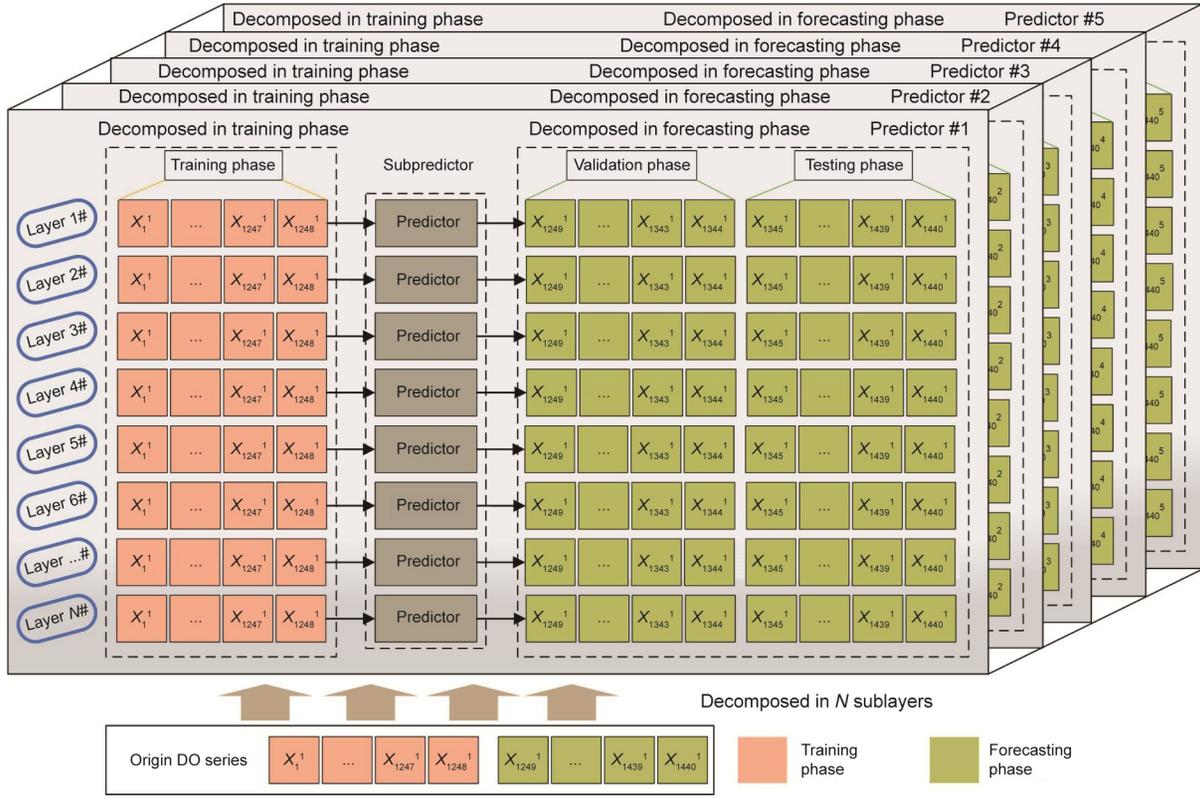


图3. EWT分解和逐层预测。

2.3. 第三阶段——多模型优化集成方法

为了进一步提高预测性能，将BFGS、ENN、GRNN、ORELM和ELM相结合，得到BEGOE模型。利用PSOGSA算法对集成权重进行优化。加权优化方案如图4所示。PSOGSA的优化目标函数是最小化平均绝对百分比误差（MAPE）。MAPE是指不同样本的实际值和预测值之间的平均绝对百分比误差。目标函数在验证集数据中进行计算，具体目标函数如下所示：

$$\text{Object}_{\text{MAPE}} = \min \left\{ \sum_{p=1}^P E_t \left[\left(E_D[\hat{y}_p(t, D)] - E_\varepsilon[Y(t, \varepsilon)] \right)^2 \right] / P \right\}$$

$$\text{s.t. } \hat{y}_p(t, D) = \frac{\sum_{a=1}^5 (\hat{X}_p^a(t, D) \times w_a)}{(w_1 + w_2 + w_3 + w_4 + w_5)}$$

$$0 \leq w_a \leq 1$$
(6)

式中， $Y(t, \varepsilon)$ 表示在 t 时刻带有噪声 ε 的真实DO序列； $\hat{y}(t, D)$ 表示 t 时刻BEGOE模型的预测结果，其中BFGS、ENN、GRNN、ORELM和ELM使用样本 D 进行训练； E_t 、 E_D 和 E_ε 表示带有噪声 ε 的训练样本随着时间推移的期望值。对于 E_ε ， $E_\varepsilon[Y(t, \varepsilon)]$ 相当于无噪声的真实DO序列。然而，真实序列的噪声是无法消除的，因为真实DO序列不是人工序列。在实际操作中为了与 $E_\varepsilon[Y(t, \varepsilon)]$ 近似，将

$E_\varepsilon[Y(t, \varepsilon)]$ 考虑为 $Y(t, \varepsilon)$ [46]。此外， P 为最大预测步数； $[w_1, w_2, w_3, w_4, w_5]$ 为集成权重； $\hat{X}_p^1, \hat{X}_p^2, \hat{X}_p^3, \hat{X}_p^4, \hat{X}_p^5, \hat{y}_p$ 为BFGS、ENN、GRNN、ORELM、ELM和BEGOE在验证集数据中的第 p 步预测结果。

考虑本例中涉及DO、水温、盐度、氧饱和度序列，故BEGOE模型的输入为4个序列，输出为一个DO序列。通过最小化目标函数MAPE，得到合适的集成权重。水温、盐度、氧饱和度的影响被纳入DO序列。最后，得到BEGOE的预测结果。附录A中的算法S3给出了多模型优化集成方法的伪代码。

3. 实验分析

本节通过多次对比实验验证了所提出模型的有效性。3.1节介绍了海洋观测系统采集的6组真实时间序列集。3.2节展示了评价预测结果的性能指标。3.3节解释了模型的实施细节和模型每个阶段的贡献。3.4节设置了两个实验，实验一横向对比了模型各组件的贡献，实验二对比了几个最先进的模型，以证明所提出的模型的优越性。

3.1. 数据描述

本节利用太平洋岛屿海洋观测系统在希洛的WQB04

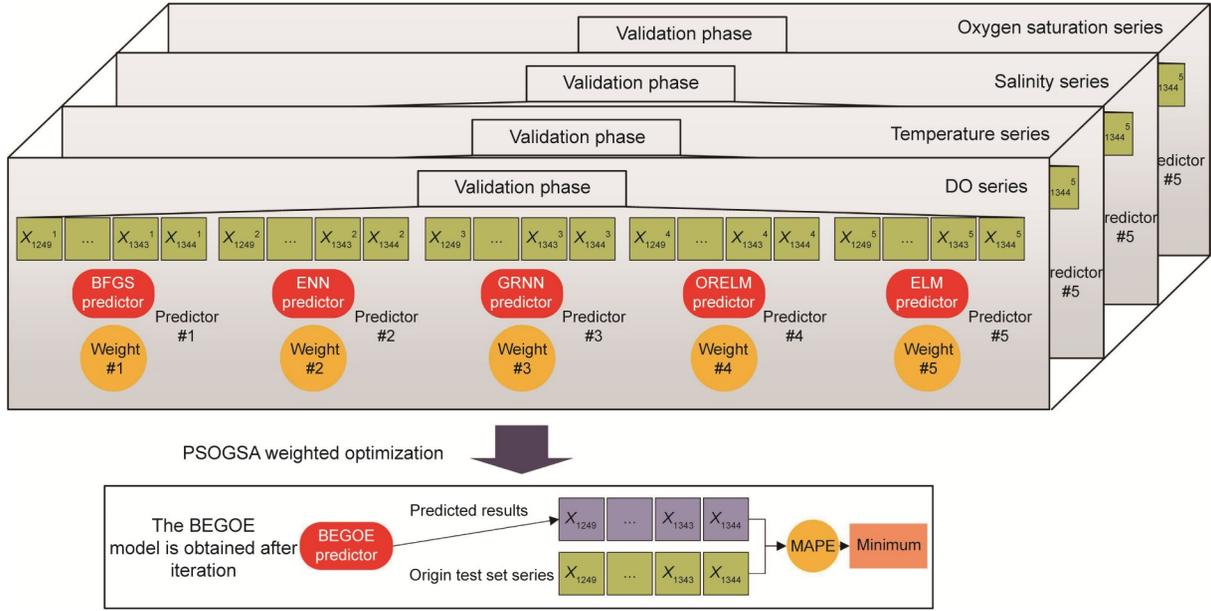


图4. 多因素分析下的PSO-GSA优化集成过程。

站收集的DO、温度、盐度、浊度、叶绿素、氧饱和度等时间序列数据进行实验研究。6组时间序列数据的时间间隔均为15 min。为了有效地构建后续模型，将数据集分为三个部分：训练集、验证集和测试集。6个因素对应的数据均选自2016年12月1—30日的时间段。其中，12月16—30日的数据作为数据集#1，12月1—15日的数据作为数据集#2。每个影响因素对应的序列总长度均为1440。训练集包含第1~1248的数据，验证集包含第1249~1344的数据，测试集包含第1345~1440的数据。附录A中的图S1和表S1是经过RNN离群值处理后的数据集信息。

3.2. 性能评估

将一组指标用于评估模型的确定性预测性能，包括决定系数 (R^2)、Nash-Sutcliffe 效率系数[47]、Kling-Gupta 效率系数[48]、误差标准差 (SDE) 和 MAPE。这5个指标的表达式如下所示：

$$R^2 = \frac{\sum_{t=1}^{N_y} (Y_t - \bar{Y}_t)(\hat{Y}_t - \bar{\hat{Y}}_t)}{\sqrt{\left(\sum_{t=1}^{N_y} (Y_t - \bar{Y}_t)^2 \sum_{t=1}^{N_y} (\hat{Y}_t - \bar{\hat{Y}}_t)^2 \right)}} \quad (7)$$

$$NSE = 1 - \frac{\sum_{t=1}^{N_y} (Y_t - \hat{Y}_t)^2}{\sum_{t=1}^{N_y} (Y_t - \bar{Y}_t)^2} \quad (8)$$

$$KGE = 1 - \sqrt{(R-1)^2 + (\alpha-1)^2 + (\beta-1)^2}$$

$$\alpha = \frac{\sum_{t=1}^{N_y} (\hat{Y}_t - \bar{\hat{Y}}_t)^2}{\sum_{t=1}^{N_y} (Y_t - \bar{Y}_t)^2} \quad (9)$$

$$\beta = \frac{\bar{\hat{Y}}_t}{\bar{Y}_t}$$

$$SDE = \sqrt{\frac{1}{N_y} \sum_{t=1}^{N_y} \left((Y_t - \hat{Y}_t) - (\bar{Y}_t - \bar{\hat{Y}}_t) \right)^2} \quad (10)$$

$$MAPE = \left(\sum_{t=1}^{N_y} \left| \frac{Y_t - \hat{Y}_t}{Y_t} \right| \right) / N_y \quad (11)$$

式中， \hat{Y} 为预测结果； Y 为真实数据； N_y 为真实数据的个数。

3.3. 建模分析

3.3.1. 多因素分析

海洋中的DO受多种因素的影响，本文分析了影响DO浓度增减的各种因素。图5为系统动力学模型。

如图5所示，影响DO浓度的因素很多。通常，大气扩散、水交换、机械有氧运动、水生光合作用等是DO的主要来源。在自然条件下，海洋中的DO含量与空气中的氧气分压和水温密切相关。当溶入水中的氧的速率等于从水中逸出的氧的速率时，溶解达到动态平衡。海洋中DO的平衡还受到盐度、浊度、pH值等因素的影响。此外，

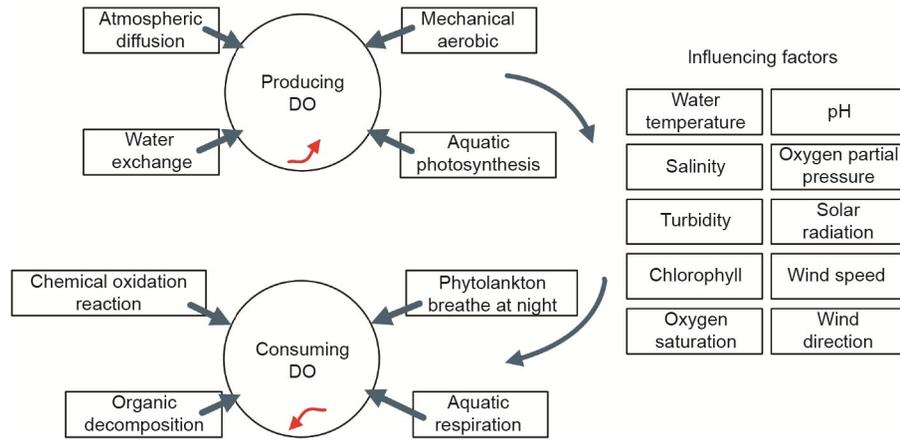


图5. DO变化的系统动力学模型图。

化学氧化反应、有机分解、浮游植物夜间呼吸和水生呼吸是海洋DO消耗的主要因素。总之，影响DO含量变化的因素很多，其分布变化规律很难用简单的机制来解释。

本实验采用灰色关联度法计算温度、盐度、浊度、叶绿素、氧饱和度与DO浓度之间的关系。具体结果如表2所示。由表2可知，温度、盐度、氧饱和度与DO的关联度最高。相关指标分别达到0.61、0.68、0.87，均超过设置的阈值0.5。值得一提的是，盐度中含有营养成分，而营养成分在DO浓度中起着重要作用。考虑到高浓度的浊度通过影响植物在水中的光合作用来影响产氧量，也将浊度作为DO相关因子输入模型。实验采用多因素变量输入和单因素变量输出的形式。在实际重复实验对比中发现，当温度、盐度和氧饱和度作为DO的影响因素输入模型时，预测精度最好。实验结果表明，如果在模型中输入浊度、叶绿素、pH等低相关性因素，预测精度会相应降低。在这种情况下，叶绿素指数总是保持为 -8.8×10^{-7} 。因此，它没有表现出与DO变化的相关性，所以叶绿素与DO的相关系数不是一个数字。

3.3.2. 分解模型分析

在本节中，利用EWT自适应分解来降低DO序列、水温序列、盐度序列和氧饱和度序列的非平稳性。值得一提的是，只对训练集数据进行分解。对原始DO序列和高相关性序列进行分解后，子序列因具有更稳定的特征而更容易预测。主预测器仅需集中在这些子序列中的某个频带

上。最后，通过对子预测器的预测结果进行叠加，得到经过数据预处理后的最终结果。本节通过分解输入序列对比了预测模型的性能变化。

3.3.3. 优化方法分析

在本节中，将PSOGSA与灰狼优化算法(GWO)、蝙蝠算法(BA)、多元宇宙优化算法(MVO)、粒子群优化算法(PSO)、鲸鱼优化算法(WOA)和新蝙蝠算法(NBA)等多种备选方法进行了对比。这些优化算法大多基于启发式仿生算法，从理论角度难以明确说明哪一种算法更适合。因此，需要通过实验验证每种优化算法在这种情况下的DO预测效果。

3.4. 对比研究

3.4.1. 实验一——与所提出模型的每个组成部分进行比较

在实验一中，基于所提出模型各组件构建了5个基准模型，用于评估所提出的模型。表3列出了它们的具体表达式。为了比较实际计算中环境参数对DO的影响，我们建立了一个多因素分析实验。模型1~3没有进行多因素分析。在同样的情况下，将多因素分析加入模型4和模型5与所提出模型中。换句话说，模型4和模型5以及所提出的模型都是利用DO序列、温度序列、盐度序列、氧饱和度序列等环境因素进行训练的。为了比较优化集成的效果，建立了多模型集成对比实验。与模型1和模型4相比，5个基准模型被优化、合并进模型2和模型5。为了验

表2 WQB04数据集中环境因子与DO的相关系数

Indicator	Correlation coefficients				
	Temperature	Salinity	Turbidity	Chlorophyll	Oxygen saturation
Dissolved oxygen	0.61	0.68	0.32	NaN	0.87

NaN: not a number.

证EWT分解方法的优越性,还进行了一组有无数据预处理的实验对比。更具体地说,模型3和所提出模型增加了一个EWT数据预处理模块,以降低原始序列的不稳定性。表4、表5列出了DO测试数据集的预测结果。此外,为了体现优化算法的效果,在多模型优化集成模块中使用了3.3.3节所述的7种优化算法。图6为PSOGSA和其他优化算法下的第三步的测试集预测结果。

表3 基准模型和所提出模型的实验设置

Model	Experimental setups
Model 1	RNNs-BFGS/ENN/GRNN/ORELM/ELM
Model 2	RNNs-BEGOE
Model 3	RNNs-EWT-BEGOE
Model 4	MF-RNNs-BFGS/ENN/GRNN/ORELM/ELM
Model 5	MF-RNNs-BEGOE
Proposed model	MF-RNNs-EWT-BEGOE

从表4、表5和图6中可以得出如下结论:

(1) 无论是在数据集#1还是数据集#2中,模型2(RNN-BEGOE)均显著优于模型1(RNN-BFGS/ENN/GRNN/ORELM/ELM),表明多模型集成方法是有效的。该方法结合了5个基准模型的优势。优化算法经过多次迭代,通过最优权重分配完成各模型的组合。集成模型综合

了各基准模型的优点,同时摒弃了它们的负面性能。以数据集#1的第一步预测结果为例,RNN-BFGS、RNN-ENN、RNN-GRNN、RNN-ORELM、RNN-ELM和RNN-BEGOE的MAPE分别为1.76%、1.51%、2.74%、1.49%、1.61%和1.32%。NSE、KGE、SDE、 R^2 等评价指标均表明模型2的预测效果优于模型1。

(2) 模型3(RNN-EWT-BEGOE)显著优于模型2(RNN-BEGOE),说明基于分解的数据预处理方法对提高模型预测性能有积极作用。EWT自适应地将原始序列分解为多个子序列。分解后的序列具有独立的振动分量,稳定性强。EWT在消除不同子序列之间相互作用的同时,有完整的数学理论支撑。这些优点的结合使得EWT成为一个很好的数据预处理候选方案。以数据集#1的第一步预测结果为例,模型2的NSE、KGE、MAPE、SDE和 R^2 分别为0.70、0.75、1.32%、0.10 $\text{mg}\cdot\text{L}^{-1}$ 和0.83,模型3的分别为0.97、0.96、0.38%、0.04 $\text{mg}\cdot\text{L}^{-1}$ 和0.97。

(3) 模型4(MF-RNN-BFGS/ENN/GRNN/ORELM/ELM)、模型5(MF-RNN-BEGOE)以及所提出的多因素分析模型与未进行多因素分析的模型1~3相比,性能显著改善。这项研究表明,温度、盐度和氧饱和度等环境因素对DO浓度有影响。它们之间存在复杂的非线性关系,仅

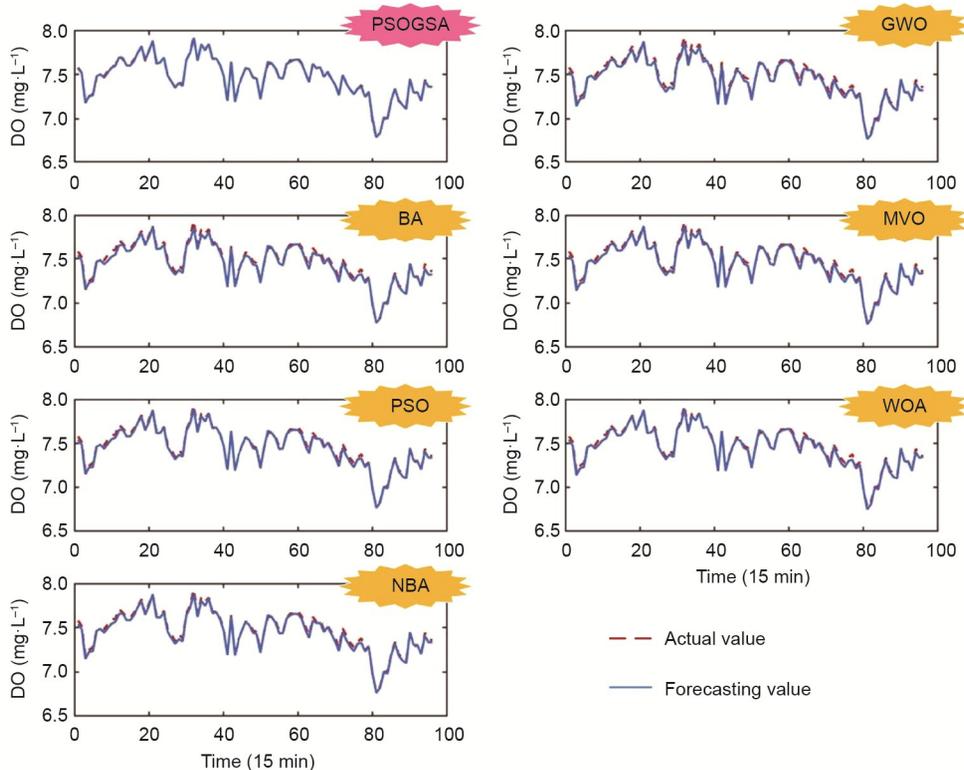


图6. PSOGSA和其他优化算法在测试集中的第三步预测结果(上述提及的不同优化方法均基于多因素-EWT-BEGOE集成模型以进行控制变量对比实验)(数据集#1)。Time(15min)表示数据的采样间隔为15min。

表4 实验一对比模型的评价指标(数据集#1)

Horizon	Model	NSE	KGE	MAPE (%)	SDE (mg·L ⁻¹)	R ²
1-step	RNNs-BFGS	0.47	0.69	1.76	0.16	0.50
	MF-RNNs-BFGS	0.61	0.77	1.45	0.14	0.62
	RNNs-ENN	0.55	0.69	1.51	0.15	0.56
	MF-RNNs-ENN	0.62	0.75	1.41	0.14	0.63
	RNNs-GRNN	-0.24	0.19	2.74	0.19	0.35
	MF-RNNs-GRNN	0.52	0.56	1.63	0.15	0.53
	RNNs-ORELM	0.58	0.77	1.49	0.14	0.62
	MF-RNNs-ORELM	0.62	0.77	1.41	0.14	0.63
	RNNs-ELM	0.53	0.71	1.61	0.15	0.55
	MF-RNNs-ELM	0.60	0.76	1.44	0.14	0.61
	RNNs-BEGOE	0.70	0.75	1.32	0.10	0.83
	MF-RNNs-BEGOE	0.91	0.90	0.67	0.07	0.91
	RNNs-EWT-BEGOE	0.97	0.96	0.38	0.04	0.97
	MF-RNNs-EWT-BEGOE	1.00	1.00	0.11	0.01	1.00
2-step	RNNs-BFGS	0.28	0.57	2.00	0.19	0.36
	MF-RNNs-BFGS	0.43	0.66	1.78	0.17	0.47
	RNNs-ENN	0.38	0.59	1.75	0.17	0.41
	MF-RNNs-ENN	0.47	0.63	1.72	0.16	0.48
	RNNs-GRNN	-0.33	0.15	2.84	0.20	0.28
	MF-RNNs-GRNN	0.32	0.41	1.99	0.18	0.33
	RNNs-ORELM	0.36	0.67	1.91	0.17	0.48
	MF-RNNs-ORELM	0.45	0.66	1.76	0.16	0.48
	RNNs-ELM	0.39	0.58	1.85	0.17	0.41
	MF-RNNs-ELM	0.46	0.65	1.72	0.16	0.48
	RNNs-BEGOE	0.53	0.65	1.69	0.13	0.68
	MF-RNNs-BEGOE	0.81	0.82	1.03	0.09	0.82
	RNNs-EWT-BEGOE	0.92	0.92	0.67	0.06	0.92
	MF-RNNs-EWT-BEGOE	0.99	0.99	0.14	0.02	1.00
3-step	RNNs-BFGS	0.02	0.46	2.28	0.22	0.22
	MF-RNNs-BFGS	0.24	0.54	2.06	0.19	0.32
	RNNs-ENN	0.23	0.47	1.99	0.19	0.29
	MF-RNNs-ENN	0.29	0.51	1.99	0.19	0.33
	RNNs-GRNN	-0.41	0.11	2.92	0.20	0.23
	MF-RNNs-GRNN	0.21	0.34	2.15	0.19	0.24
	RNNs-ORELM	0.09	0.56	2.31	0.19	0.33
	MF-RNNs-ORELM	0.27	0.54	2.04	0.19	0.33
	RNNs-ELM	0.22	0.46	2.12	0.20	0.27
	MF-RNNs-ELM	0.27	0.53	1.99	0.19	0.33
	RNNs-BEGOE	0.47	0.60	1.80	0.14	0.64
	MF-RNNs-BEGOE	0.71	0.76	1.24	0.11	0.75
	RNNs-EWT-BEGOE	0.80	0.88	1.07	0.10	0.80
	MF-RNNs-EWT-BEGOE	0.98	0.98	0.17	0.04	0.99

考虑单个DO因子不够全面。以数据集#1的第一步预测结果为例,模型2的NSE、KGE、MAPE、SDE和R²分别为

0.70、0.75、1.32%、0.10 mg·L⁻¹和0.83,模型5的分别为0.91、0.90、0.67%、0.07 mg·L⁻¹和0.91。实验结果表明,

表5 实验一对比模型的评价指标(数据集#2)

Horizon	Model	NSE	KG E	MAPE (%)	SDE (mg·L ⁻¹)	R ²
1-step	RNNs-BFGS	0.77	0.85	1.20	0.12	0.78
	MF-RNNs-BFGS	0.80	0.89	1.12	0.11	0.80
	RNNs-ENN	0.75	0.76	1.39	0.12	0.75
	MF-RNNs-ENN	0.79	0.85	1.17	0.11	0.79
	RNNs-GRNN	0.07	0.35	2.84	0.19	0.46
	MF-RNNs-GRNN	0.39	0.63	2.31	0.19	0.45
	RNNs-ORELM	0.79	0.89	1.15	0.11	0.80
	MF-RNNs-ORELM	0.80	0.86	1.13	0.11	0.80
	RNNs-ELM	0.76	0.83	1.26	0.12	0.76
	MF-RNNs-ELM	0.78	0.82	1.24	0.12	0.78
	RNNs-BEGOE	0.87	0.78	1.04	0.09	0.90
	MF-RNNs-BEGOE	0.95	0.94	0.60	0.06	0.95
	RNNs-EWT-BEGOE	0.98	0.99	0.35	0.03	0.98
	MF-RNNs-EWT-BEGOE	1.00	0.98	0.18	0.01	1.00
2-step	RNNs-BFGS	0.60	0.74	1.68	0.15	0.63
	MF-RNNs-BFGS	0.64	0.81	1.57	0.15	0.66
	RNNs-ENN	0.58	0.63	1.86	0.16	0.59
	MF-RNNs-ENN	0.64	0.75	1.65	0.15	0.64
	RNNs-GRNN	0.00	0.33	2.95	0.20	0.40
	MF-RNNs-GRNN	0.23	0.56	2.61	0.21	0.34
	RNNs-ORELM	0.62	0.81	1.68	0.15	0.66
	MF-RNNs-ORELM	0.66	0.76	1.59	0.14	0.66
	RNNs-ELM	0.60	0.72	1.73	0.16	0.60
	MF-RNNs-ELM	0.62	0.69	1.78	0.15	0.62
	RNNs-BEGOE	0.78	0.71	1.36	0.12	0.81
	MF-RNNs-BEGOE	0.88	0.91	0.95	0.08	0.89
	RNNs-EWT-BEGOE	0.94	0.95	0.62	0.06	0.95
	MF-RNNs-EWT-BEGOE	0.99	0.97	0.28	0.02	0.99
3-step	RNNs-BFGS	0.46	0.67	1.96	0.17	0.51
	MF-RNNs-BFGS	0.47	0.73	1.95	0.18	0.53
	RNNs-ENN	0.47	0.54	2.14	0.18	0.49
	MF-RNNs-ENN	0.50	0.65	1.99	0.17	0.51
	RNNs-GRNN	-0.06	0.30	3.06	0.20	0.35
	MF-RNNs-GRNN	0.18	0.54	2.71	0.22	0.31
	RNNs-ORELM	0.44	0.72	2.13	0.18	0.53
	MF-RNNs-ORELM	0.53	0.67	1.90	0.17	0.54
	RNNs-ELM	0.47	0.64	2.02	0.18	0.49
	MF-RNNs-ELM	0.49	0.61	2.07	0.17	0.50
	RNNs-BEGOE	0.69	0.65	1.60	0.14	0.73
	MF-RNNs-BEGOE	0.77	0.84	1.39	0.12	0.78
	RNNs-EWT-BEGOE	0.87	0.91	0.95	0.09	0.89
	MF-RNNs-EWT-BEGOE	0.96	0.94	0.58	0.04	0.98

多因素考虑是正确的。将多种因素合理地输入预测模型中,可以得到更准确、更科学的预测结果。

(4) 综合所有的实验数据集、预测步数和不同的模型评价指标,所提出的混合模型的预测效果最为突出。所提

出的模型具有良好的鲁棒性和准确性,使其预测结果非常接近实际的DO值。以数据集#1的第一步预测结果为例,所提出的模型的NSE、KGE、MAPE、SDE和 R^2 分别为1.00、1.00、0.11%、 $0.01 \text{ mg}\cdot\text{L}^{-1}$ 和1.00。以数据集#2的第一步预测结果为例,所提出的模型的NSE、KGE、MAPE、SDE和 R^2 分别为1.00、0.98、0.18%、 $0.01 \text{ mg}\cdot\text{L}^{-1}$ 和1.00。两组实验表明,所提出的模型具有良好的持久性。

总的来说,要提高混合模型的性能,多因素分析、自适应分解分析和科学的集成模型分析是必不可少的。该混合模型框架具有较强的可解释性,为DO预测的改进提供了方向。

3.4.2. 实验二——与现有模型比较

近年来,人们提出了许多基于AI的DO预测模型。在实验二中,我们复现了2018—2020年发表的三个最先进的模型,并将它们与所提出的混合模型进行比较,以验证DO预测性能。这些模型都比较复杂,基本的机器学习模型也可以取得满意的结果。为了完成科学合理的模型评价,我们增加了基于ANN的支持向量机(SVM)模型和基于深度学习的DBN模型进行比较。具体模型介绍如下:

Li等[31]使用递归神经网络、长短期记忆(LSTM)网络和门控循环单元(GRU)分别构建了三个DO预测模型,以确定最合适的一个。在实验中,计算pH、浊度、温度、NH₄⁺-N、DO之间的相关系数,然后将这些序列均匀输入预测模型。与Li等[31]的直接输入所有参数相比,我们设置了阈值和实验筛选方法以最终确定对DO预测有积极影响的因素。Li等[31]的方法有降低准确性的风险。事实上,影响因素的相关系数只能作为参考指标,而这些环境因素与DO之间的真实关系是非常复杂的。

Huan等[12]将EEMD数据预处理方法与LSSVM相结合,提高了模型预测DO的性能。运用EEMD对DO序列进行分解,然后利用LSSVM分别对这些本征模函数(IMF)分量进行预测。通过叠加得到综合预测结果。在最终DO预测结果生成之前,利用BPNN对预测结果进行重构,从而消除了错误分析。遗憾的是,EEMD在数学定义方面缺乏足够的理论支持,导致了分解缺陷。我们使用了一种新的信号处理工具EWT来克服这一缺点。同时,EWT的自适应分解特性可以确定最优的分解层数,使分解效果最大化。

Ren等[23]利用遗传算法优化FNN的中心层的中心和宽度,以确定最佳组合,提高DO预测性能。在所提出的模型中,我们将优化算法应用于多个基准模型的集成。这种方法的优点是合理地结合了多个AI模型的优越性能,同时丢弃了较差的性能。与单一参数优化的预测器相比,

多模型的互补性可以使模型更加成熟,使其获得更充分的训练。在所使用的模型中,通过多次迭代得到的合理权重分配起着关键作用。

更具体地,表6、表7和图7、图8展示了所提出模型和最先进模型的DO测试集预测结果。

从图7、图8和表6、表7可以得出如下结论:

综合所有的数据集、预测步数和不同的时间序列评价指标,与文中其他最先进的模型相比,所提出的模型具有最佳的预测性能。以数据集#1的第一步预测结果为例,SVM、DBN、Li等[31]的模型、Huan等[12]的模型、Ren等[23]的模型和所提出的模型的MAPE分别为1.41%、1.44%、0.52%、0.56%、0.51%和0.11%。所提出的模型具有较好的预测鲁棒性,即使在较高的预测步骤中也能令人满意地进行预测。以数据集#1的第3步预测结果为例,SVM、DBN、Li等[31]的模型、Huan等[12]的模型、Ren等[23]的模型和所提出的模型的MAPE分别为1.99%、2.06%、1.05%、0.89%、1.21%和0.17%。基于其他持久性指标,所提出模型的性能非常出色。以数据集#1的第3步预测结果为例,所提出的模型的NSE、KGE、MAPE、SDE和 R^2 分别为0.98、0.98、0.17%、 $0.04 \text{ mg}\cdot\text{L}^{-1}$ 和0.99。以数据集#2的第3步预测结果为例,所提出模型的NSE、KGE、MAPE、SDE和 R^2 分别为0.96、0.94、0.58%、 $0.04 \text{ mg}\cdot\text{L}^{-1}$ 和0.98。

综上所述,不同数据集的预测结果表明,所提出的混合预测模型框架合理,预测效果令人满意。这主要是因为该模型融合了多因素分析、自适应分解分析和科学的集成模型分析方法。这种组合有助于提高所提出模型的预测精度和鲁棒性。

4. 潜在应用

海洋环境中DO浓度水平是由两种效应引起的:耗氧效应,降低DO浓度;氧回收效应,增加DO浓度。水生生物的生命活动消耗一定数量的氧气,从而降低了DO浓度。水生植物的光合作用和空气中氧气的扩散会增加DO浓度。图9(a)为真实DO时间序列数据的变化以及第1步、第2步和第3步预测结果。以下事实如图9所示:①随着温度的升高,DO浓度降低;随着温度的降低,DO浓度增加。这是因为更高的温度将迫使水中的DO逃逸到表面[图9(b)]。②盐度与DO呈负相关关系,与耗氧量有关。充足的营养会促进水生生物的生命活动,导致呼吸和耗氧量增加。盐度的增加也会促使DO浓度水平接近饱和[图9(c)]。③高浊度会降低水生植物的光合作用,影

表6 所提出模型和最先进模型的DO测试集预测结果(数据集#1)

Horizon	Model	NSE	KGE	MAPE (%)	SDE (mg·L ⁻¹)	R ²
1-step	SVM	0.62	0.75	1.41	0.14	0.63
	DBN	0.61	0.74	1.44	0.14	0.62
	Li's model [31]	0.95	0.94	0.52	0.05	0.95
	Huan's model [12]	0.95	0.96	0.56	0.05	0.95
	Ren's model [23]	0.95	0.96	0.51	0.05	0.95
	Proposed model	1.00	1.00	0.11	0.01	1.00
2-step	SVM	0.47	0.63	1.72	0.16	0.48
	DBN	0.45	0.61	1.75	0.16	0.48
	Li's model [31]	0.84	0.83	0.90	0.09	0.85
	Huan's model [12]	0.87	0.92	0.83	0.08	0.87
	Ren's model [23]	0.86	0.91	0.87	0.08	0.86
	Proposed model	0.99	0.99	0.14	0.02	1.00
3-step	SVM	0.29	0.51	1.99	0.19	0.33
	DBN	0.26	0.47	2.06	0.18	0.32
	Li's model [31]	0.79	0.76	1.05	0.10	0.80
	Huan's model [12]	0.85	0.91	0.89	0.08	0.86
	Ren's model [23]	0.73	0.85	1.21	0.12	0.74
	Proposed model	0.98	0.98	0.17	0.04	0.99

表7 所提出模型和最先进模型的DO测试集预测结果(数据集#2)

Horizon	Model	NSE	KGE	MAPE (%)	SDE (mg·L ⁻¹)	R ²
1-step	SVM	0.79	0.85	1.17	0.11	0.79
	DBN	0.79	0.89	1.14	0.11	0.80
	Li's model [31]	0.96	0.96	0.51	0.04	0.97
	Huan's model [12]	0.97	0.96	0.50	0.04	0.97
	Ren's model [23]	0.96	0.96	0.56	0.05	0.96
	Proposed model	1.00	0.98	0.18	0.01	1.00
2-step	SVM	0.64	0.75	1.65	0.15	0.64
	DBN	0.63	0.81	1.61	0.15	0.66
	Li's model [31]	0.91	0.91	0.81	0.07	0.92
	Huan's model [12]	0.95	0.94	0.62	0.06	0.95
	Ren's model [23]	0.92	0.94	0.77	0.07	0.92
	Proposed model	0.99	0.97	0.28	0.02	0.99
3-step	SVM	0.50	0.65	1.99	0.17	0.51
	DBN	0.46	0.72	1.95	0.18	0.53
	Li's model [31]	0.86	0.85	1.06	0.09	0.86
	Huan's model [12]	0.94	0.94	0.68	0.06	0.94
	Ren's model [23]	0.82	0.88	1.19	0.11	0.82
	Proposed model	0.96	0.94	0.58	0.04	0.98

响DO的产生,因此浊度对DO的增加有一定的抑制作用[图9(d)]。④溶解在水中的氧含量又称氧饱和度。当水

与大气之间的氧交换达到平衡时,水中的DO浓度与氧饱和度的波动最为相似[图9(e)]。数学建模与AI技术相

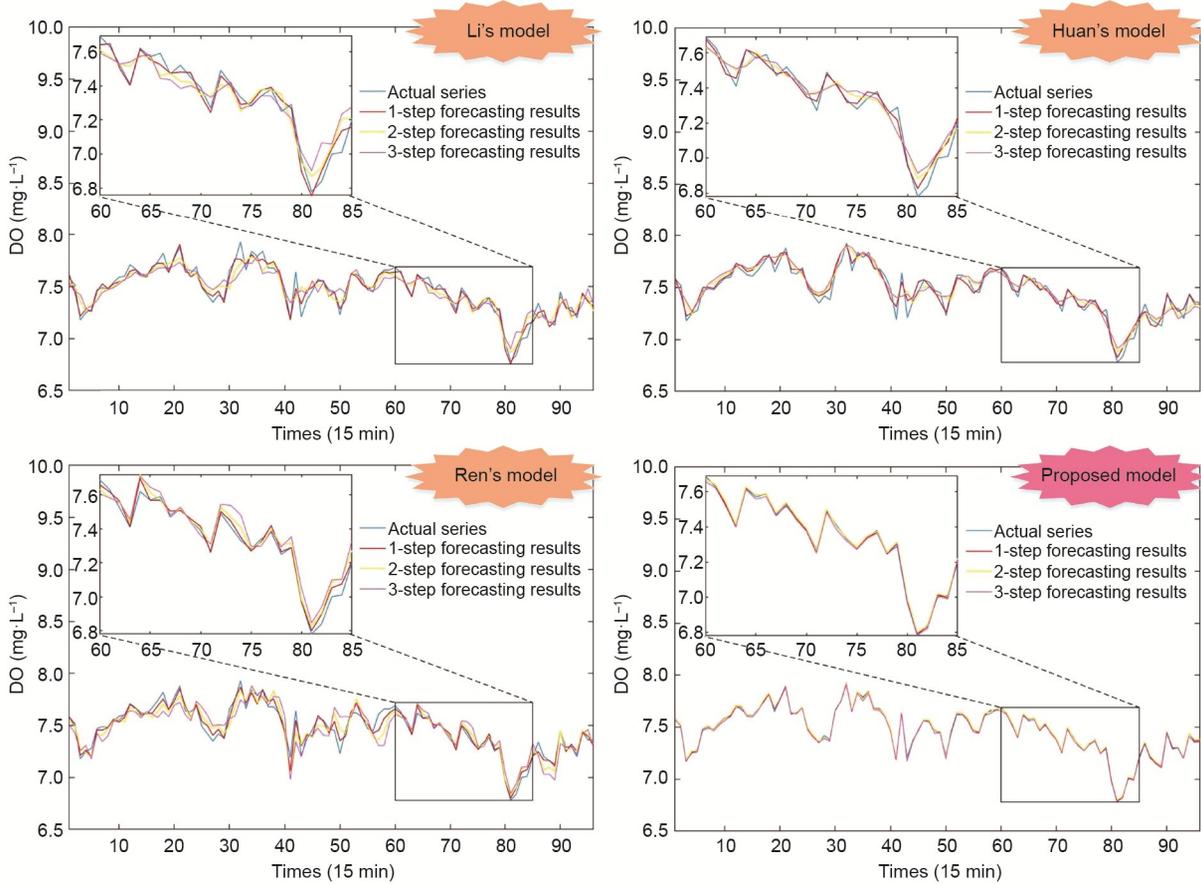


图7. 所提出模型 (MF-RNN-EWT-BEGOE) 和其他模型的DO测试集预测结果 (数据集#1)。

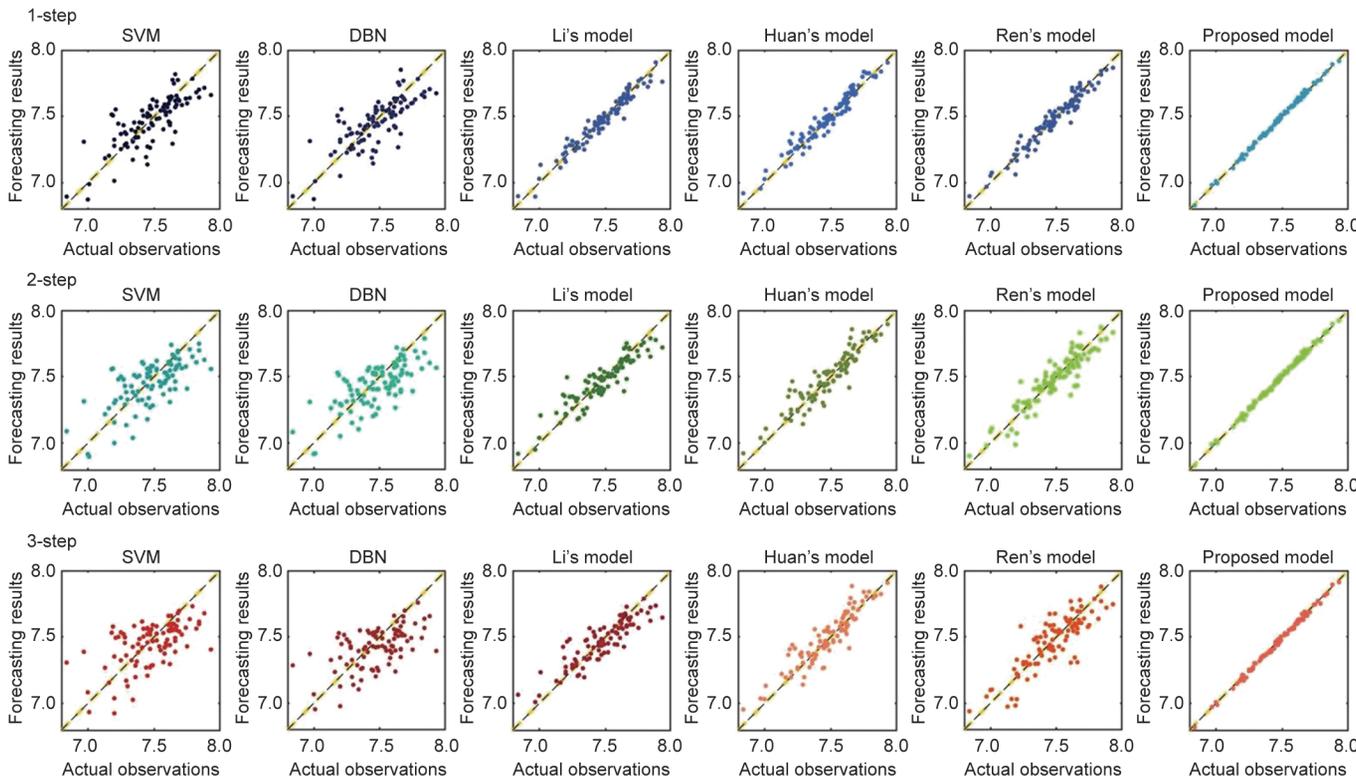


图8. 实际观测值和所提出模型与其他模型所生成的对应预测结果的散点图。

结合，可以相对科学地预测 DO 浓度的变化。该预测可为相关行业的从业人员提供有益的指导。

本研究提出一种新的混合模型，可以得到可靠的 DO 预测结果。该模型以温度、盐度、浊度、氧饱和度等诸多与 DO 密切相关的因素为输入变量，最终输出 DO 预测结果。DO 数据与相关影响因素数据的划分情况如图 9 (f) 所示。具体的多输入单输出预测结果如图 9 所示。所提出

的模型对用户的潜在应用如下：

- 水产养殖户可以利用所提出的 DO 预测模型来确定池塘和近海养殖区域未来 DO 浓度的变化。他们可以以此采取相应的应急措施以减少经济损失。

- 渔业部门管理者可以通过准确的预测，充分了解 DO 的情况和趋势。这可以为改善海洋生物生存环境提供科学的决策依据，有利于生态水质的恢复和调节。

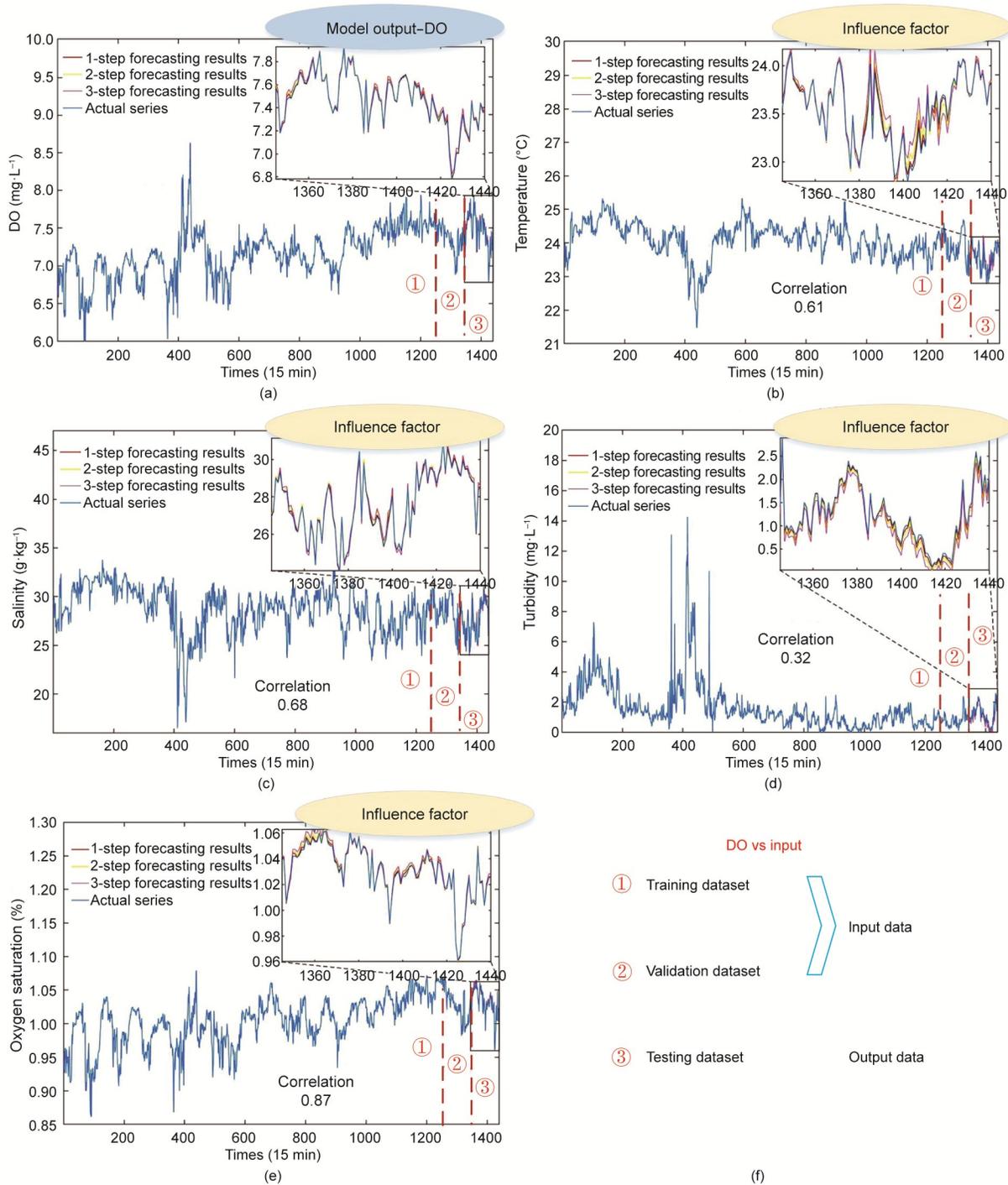


图 9. 多输入单输出模型的 DO 与其他输入数据的变化趋势图。

5. 结论

DO 预测技术的发展为海产品养殖户及相关产业带来了更多的可能性。精准实时的 DO 预测, 可以为管理者提供科学的决策指导。不幸的是, 海洋 DO 序列的不稳定性和其他各种潜在因素给 DO 预测带来了挑战。在本研究中, 提出了一种用于 DO 预测的新型混合模型——MF-RNN-EWT-BEGOE。该模型的主要特点是多因素分析、自适应分解和多模型集成。通过对案例研究中的实验结果进行分析, 可以得出以下结论:

(1) 考虑到影响海洋中 DO 浓度的因素很多, 提出了一种多因素分析方法来综合这些因素对 DO 浓度的影响。大量的物理、化学和生物系统影响使得 DO 波动机制更加复杂。合理的多因素考虑将使得模型预测更加准确。

(2) 采用自适应数据预处理分解方法降低 DO 序列的不稳定性, 使子序列更易于预测。实验结果表明, 对分解后的原始序列进行预测的方法是有效的。

(3) 提出了一种新的多模型集成方法来完成多个模型的组合, 吸纳各基准模型的优秀品质, 使得集成模型更加成熟。多个基准模型的对比实验结果表明, 多模型优势互补可以使集成模型的性能朝着更好的方向发展。

(4) 所提出的混合模型的每个组成部分都能显著提高模型对 DO 的预测性能。通过 RNN 检测和剔除异常值, 可以减少极端异常值对模型的负面影响。通过灰色关联多因素相关性分析, 可以更全面地考虑影响 DO 变化的环境因素。EWT 自适应分解有效地提高了模型的预测精度, 减少了序列非平稳性的影响。BEGOE 集成模型结合了多个基准模型的优点, 使模型的预测性能最大化。因此, 本文提出的混合模型明显优于其他用于对比的模型。

鉴于第 4 节中描述的应用潜力, 很明显, 准确的 DO 预测可以为水产养殖户减少经济损失提供科学的指导。准确的 DO 预测也可以为渔业部门管理者帮助改善池塘和海洋生物的生存环境提供科学依据, 还可以帮助决策者调节和恢复生态水质。该混合模型具有优异的多步预测性能。MIMOS 机制为相关行业人员提供了充分的关于未来 DO 波动的信息, 从而能够指导其进行有效管理。

致谢

该研究得到了国家自然科学基金项目(61873283)、长沙市科技计划项目(KQ1707017)和中南大学创新驱动项目(2019CX005)的大力支持。

Compliance with ethics guidelines

Hui Liu, Rui Yang, Zhu Duan, and Haiping Wu declare that they have no conflict of interest or financial conflicts to disclose.

Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.eng.2020.10.023>.

References

- [1] Yang Z. Watershed ecology and its applications. *Engineering* 2018;4(5):582–3.
- [2] Hoogakker BAA, Lu Z, Umling N, Jones L, Zhou X, Rickaby RE, et al. Glacial expansion of oxygen-depleted seawater in the eastern tropical Pacific. *Nature* 2018;562(7727):410–3.
- [3] McClanahan TR, Ateweberhan M, Muhando CA, Maina J, Mohammed MS. Effects of climate and seawater temperature variation on coral bleaching and mortality. *Ecol Monogr* 2007;77(4):503–25.
- [4] Gimpel A, Stelzenmüller V, Grote B, Buck BH, Floeter J, Núñez-Riboni I, et al. A GIS modelling framework to evaluate marine spatial planning scenarios: collocation of offshore wind farms and aquaculture in the German EEZ. *Mar Policy* 2015;55:102–15.
- [5] Keller AA, Ciannelli L, Wakefield WW, Simon V, Barth JA, Pierce SD. Occurrence of demersal fishes in relation to near-bottom oxygen levels within the California Current large marine ecosystem. *Fish Oceanogr* 2015; 24(2): 162–76.
- [6] Addanki SC, Venkataraman H. Greening the economy: a review of urban sustainability measures for developing new cities. *Sustainable Cities Soc* 2017; 32:1–8.
- [7] Schmidtko S, Stramma L, Visbeck M. Decline in global oceanic oxygen content during the past five decades. *Nature* 2017;542(7641):335–9.
- [8] Culbertson SD, Piedrahita RH. Aquaculture pond ecosystem model: temperature and dissolved oxygen prediction—mechanism and application. *Ecol Modell* 1996;89(1–3):231–58.
- [9] Liu S, Xu L, Li D, Li Q, Jiang Yu, Tai H, et al. Prediction of dissolved oxygen content in river crab culture based on least squares support vector regression optimized by improved particle swarm optimization. *Comput Electron Agric* 2013;95:82–91.
- [10] Faruk DÖ. A hybrid neural network and ARIMA model for water quality time series prediction. *Eng Appl Artif Intell* 2010;23(4):586–94.
- [11] Li C, Li Z, Wu J, Zhu L, Yue J. A hybrid model for dissolved oxygen prediction in aquaculture based on multi-scale features. *Inf Process Agric* 2018; 5(1): 11–20.
- [12] Huan J, Cao W, Qin Y. Prediction of dissolved oxygen in aquaculture based on EEMD and LSSVM optimized by the Bayesian evidence framework. *Comput Electron Agric* 2018;150:257–65.
- [13] Khan U, Valeo C, Comparing A. Bayesian and fuzzy number approach to uncertainty quantification in short-term dissolved oxygen prediction. *J Environ Inform* 2017;30(1):1–16.
- [14] Khan UT, Valeo C, He J. Non-linear fuzzy-set based uncertainty propagation for improved DO prediction using multiple-linear regression. *Stochastic Environ Res Risk Assess* 2013;27(3):599–616.
- [15] Kisi O, Parmar KS. Application of least square support vector machine and multivariate adaptive regression spline models in long term prediction of river water pollution. *J Hydrol* 2016;534:104–12.
- [16] Ay M, Kisi O. Modelling of chemical oxygen demand by using ANNs, ANFIS and k-means clustering techniques. *J Hydrol* 2014;511:279–89.
- [17] Ay M, Kisi O. Modeling of dissolved oxygen concentration using different neural network techniques in Foundation Creek, El Paso County, Colorado. *J Environ Eng* 2012;138(6):654–62.

- [18] Zhu S, Heddam S. Prediction of dissolved oxygen in urban rivers at the Three Gorges Reservoir, China: extreme learning machines (ELM) versus artificial neural network (ANN). *Water Quality Res J* 2020;55(1):106–18.
- [19] Heddam S, Kisi O. Modelling daily dissolved oxygen concentration using least square support vector machine, multivariate adaptive regression splines and M5 model tree. *J Hydrol* 2018;559:499–509.
- [20] Ay M, Kisi Ö. Estimation of dissolved oxygen by using neural networks and neuro fuzzy computing techniques. *KSCE J Civ Eng* 2017;21(5):1631–9.
- [21] Khan UT, Valeo C. Dissolved oxygen prediction using a possibility theory based fuzzy neural network. *Hydrol Earth Syst Sci* 2016;20(6):2267–93.
- [22] Shi P, Li G, Yuan Y, Huang G, Kuang L. Prediction of dissolved oxygen content in aquaculture using Clustering-based Softplus Extreme Learning Machine. *Comput Electron Agric* 2019;157:329–38.
- [23] Ren Q, Zhang L, Wei Y, Li D. A method for predicting dissolved oxygen in aquaculture water in an aquaponics system. *Comput Electron Agric* 2018;151:384–91.
- [24] Wu J, Li Z, Zhu L, Li G, Niu B, Peng F. Optimized BP neural network for dissolved oxygen prediction. *IFAC-PapersOnLine* 2018;51(17):596–601.
- [25] Liu S, Xu L, Jiang Yu, Li D, Chen Y, Li Z. A hybrid WA-CPSO-LSSVR model for dissolved oxygen content prediction in crab culture. *Eng Appl Artif Intell* 2014;29:114–24.
- [26] Raheli B, Aalami MT, El-Shafie A, Ghorbani MA, Deo RC. Uncertainty assessment of the multilayer perceptron (MLP) neural network model with implementation of the novel hybrid MLP-FFA method for prediction of biochemical oxygen demand and dissolved oxygen: a case study of Langat River. *Environ Earth Sci* 2017;76(14):503.
- [27] Yang H, Csukás B, Varga M, Kucska B, Szabó T, Li D. A quick condition adaptive soft sensor model with dual scale structure for dissolved oxygen simulation of recirculation aquaculture system. *Comput Electron Agric* 2019;162:807–24.
- [28] Ma J, Ding Y, Cheng JCP, Jiang F, Xu Z. Soft detection of 5-day BOD with sparse matrix in city harbor water using deep learning techniques. *Water Res* 2020;170:115350.
- [29] Ren Q, Wang X, Li W, Wei Y, An D. Research of dissolved oxygen prediction in recirculating aquaculture systems based on deep belief network. *Aquacult Eng* 2020;90:102085.
- [30] Kisi O, Akbari N, Sanatipour M, Hashemi A, Teimourzadeh K, Shiri J. Modeling of dissolved oxygen in river water using artificial intelligence techniques. *J Environ Inform* 2013;22(2):92–101.
- [31] Li W, Wu H, Zhu N, Jiang Y, Tan J, Guo Y. Prediction of dissolved oxygen in a fishery pond based on gated recurrent unit (GRU). *Inf Process Agric* 2020;8(1):185–93.
- [32] Ta X, Wei Y. Research on a dissolved oxygen prediction method for recirculating aquaculture systems based on a convolution neural network. *Comput Electron Agric* 2018;145:302–10.
- [33] Mandal S, Debnath M, Ray S, Ghosh PB, Roy M, Ray S. Dynamic modelling of dissolved oxygen in the creeks of Sagar island, Hooghly-Matla estuarine system, West Bengal, India. *Appl Math Model* 2012;36(12):5952–63.
- [34] Liu H, Yang R, Duan Z. Wind speed forecasting using a new multi-factor fusion and multi-resolution ensemble model with real-time decomposition and adaptive error correction. *Energy Convers Manage* 2020;217:112995.
- [35] Liu H, Yang R, Wang T, Zhang L. A hybrid neural network model for short-term wind speed forecasting based on decomposition, multi-learner ensemble, and adaptive multiple error corrections. *Renew Energy* 2021;165:573–94.
- [36] Zounemat-Kermani M, Seo Y, Kim S, Ghorbani MA, Samadianfard S, Naghshara S, et al. Can decomposition approaches always enhance soft computing models? Predicting the dissolved oxygen concentration in the St. Johns River, Florida. *Appl Sci* 2019;9(12):2534.
- [37] Gilles J. Empirical wavelet transform. *IEEE Trans Signal Process* 2013;61(16):3999–4010.
- [38] Zhu JJ, Kang L, Anderson PR. Predicting influent biochemical oxygen demand: balancing energy demand and risk management. *Water Res* 2018;128:304–13.
- [39] Kisi O, Alizamir M, Gorgij AD. Dissolved oxygen prediction using a new ensemble method. *Environ Sci Pollut Res Int* 2020;27(9):9589–603.
- [40] Cao W, Huan J, Liu C, Qin Y, Wu F. A combined model of dissolved oxygen prediction in the pond based on multiple-factor analysis and multi-scale feature extraction. *Aquacult Eng* 2019;84:50–9.
- [41] Chatfield C, Weigend AS. Time series prediction: forecasting the future and understanding the past: Neil A. Gershenfeld and Andreas S. Weigend, 1994, 'The future of time series', in: A. S. Weigend and N. A. Gershenfeld, eds., (Addison-Wesley, Reading, MA), 1-70. *Int J Forecast* 1994;10(1):161–3.
- [42] Chatfield C. The future of the time-series forecasting. *Int J Forecast* 1988;4(3):411–9.
- [43] Hawkins S, He H, Williams G, Baxter R. Outlier detection using replicator neural networks. In: *Proceedings of International Conference on Data Warehousing and Knowledge Discovery*; Kinsdale, Ireland. Berlin: Springer; 2002. p. 170–80.
- [44] Hu J, Wang J, Zhang X, Fu Z. Research status and development trends of information technologies in aquacultures. *Nongye Jixie Xuebao* 2015; 46(7):251–63.
- [45] Ip WC, Hu BQ, Wong H, Xia J. Applications of grey relational method to river environment quality evaluation in China. *J Hydrol* 2009;379(3–4):284–90.
- [46] Valentini G, Dietterich TG. Bias-variance analysis of support vector machines for the development of SVM-based ensemble methods. *J Mach Learn Res* 2004; 5:725–75.
- [47] Nash JE, Sutcliffe JV. River flow forecasting through conceptual models part I —a discussion of principles. *J Hydrol* 1970;10(3):282–90.
- [48] Knoben WJM, Freer JE, Woods RA. Inherent benchmark or not? Comparing Nash-Sutcliffe and Kling-Gupta efficiency scores. *Hydrol Earth Syst Sci* 2019; 23(10):4323–31.