

# 基于 DeepAR – RELM 的池塘溶解氧时空预测方法研究

樊宇星<sup>1,2</sup> 任妮<sup>3</sup> 田港陆<sup>1,2</sup> 段青玲<sup>1,2</sup>

(1. 中国农业大学信息与电气工程学院, 北京 100083; 2. 中国农业大学国家数字渔业创新中心, 北京 100083;  
3. 江苏省农业科学院农业信息研究所, 南京 210014)

**摘要:** 水体溶解氧(Dissolved oxygen, DO)是养殖水产品健康生长的重要生态因子。池塘溶解氧易受多种因素的影响,会产生时间和空间上分布的差异,现有的溶解氧预测方法大多是针对单监测点的时间序列预测,无法描述池塘溶解氧的空间分布,因此,对池塘溶解氧进行时间和空间预测非常重要。本文提出一种基于自回归循环神经网络(Autoregressive recurrent neural network, DeepAR)和正则化极限学习机(Regularized extreme learning machine, RELM)的池塘溶解氧时空预测方法。首先采用样本熵(Sample entropy, SE)衡量各个监测点溶解氧序列的波动程度,采用最大互信息系数(Maximum mutual information coefficient, MIC)衡量监测点溶解氧序列之间的相关性,综合选取出溶解氧序列波动程度较小且与各个监测点相关性较大的监测点作为中心监测点,并以中心监测点为原点,建立池塘空间坐标系;其次采用 DeepAR 算法构建中心监测点的溶解氧时间序列预测模型,实现中心监测点溶解氧时间序列预测;最后采用 RELM 算法构建中心监测点与池塘各位置溶解氧之间的空间映射关系模型,结合中心监测点溶解氧时间序列预测值和池塘空间坐标,实现对未来时刻池塘溶解氧的空间预测。该方法在提高时间序列预测精度的同时,实现了对未来时刻池塘溶解氧空间状态的预测。在真实的数据集上进行测试,预测未来 24 h 的池塘空间溶解氧值,均方根误差(RMSE)为 1.263 3 mg/L、平均绝对误差(MAE)为 0.975 5 mg/L、平均绝对百分比误差(MAPE)为 14.873 2%。并与标准极限学习机(Extreme learning machine, ELM)、径向基神经网络(Radial basis function neural network, RBFNN)、梯度提升回归树(Gradient boosting regression tree, GBRT)和随机森林(Random forest, RF)4种预测方法进行对比,各评价指标的性能均有较大幅度提升,表明该方法有较好的预测精度和泛化能力,能够较准确地实现池塘溶解氧时空预测。

**关键词:** 池塘养殖; 溶解氧; 时空预测; 自回归循环神经网络

中图分类号: TP391 文献标识码: A 文章编号: 1000-1298(2020)S1-0405-08

## Spatio-temporal Prediction Method of Dissolved Oxygen in Ponds Based on DeepAR – RELM

FAN Yuxing<sup>1,2</sup> REN Ni<sup>3</sup> TIAN Ganglu<sup>1,2</sup> DUAN Qingling<sup>1,2</sup>

(1. College of Information and Electrical Engineering, China Agricultural University, Beijing 100083, China  
2. National Innovation Center for Digital Fishery, China Agricultural University, Beijing 100083, China  
3. Institute of Agricultural Information, Jiangsu Academy of Agricultural Sciences, Nanjing 210014, China)

**Abstract:** Dissolved oxygen (DO) in water is an important ecological factor for the healthy growth of aquaculture products. Dissolved oxygen in ponds is susceptible to many factors, which would cause differences in temporal and spatial distribution. Most of the existing dissolved oxygen prediction methods are time series predictions for a single monitoring point, and it cannot describe the spatial distribution of dissolved oxygen in the pond. Therefore, it is very important to predict the spatial and temporal dissolved oxygen in ponds. A spatio-temporal prediction method of dissolved oxygen in ponds based on autoregressive recurrent neural network (DeepAR) and regularized extreme learning machine (RELM) was proposed. Firstly, according to the sample entropy (SE) of the original dissolved oxygen sequence of each monitoring point and the maximum mutual information coefficient (MIC) between the sequences, a

收稿日期: 2020-08-03 修回日期: 2020-09-21

基金项目: 江苏省农业科技自主创新资金项目(CX(19)1003)和山东省重大科技创新工程项目(2019JZZY010703)

作者简介: 樊宇星(1997—),男,硕士生,主要从事水产养殖大数据挖掘研究, E-mail: 1271804066@qq.com

通信作者: 段青玲(1967—),女,教授,博士生导师,主要从事智能信息处理研究, E-mail: dqling@cau.edu.cn

monitoring point with a smaller entropy value and a greater correlation with each point was selected as the central monitoring point, and the pond spatial coordinate system was established with the central monitoring point as the origin. Secondly, the DeepAR algorithm was used to predict the time series of dissolved oxygen in the central monitoring point. Finally, the RELM algorithm was used to construct the spatial mapping relation model between the central monitoring point and the dissolved oxygen in each location of the pond, and the spatial prediction of the dissolved oxygen in the pond in the future was realized by combining the predicted value of the time series of the dissolved oxygen at the central monitoring point and the spatial coordinates of the pond. This method not only improved the accuracy of time series prediction, but also realized the spatial prediction of dissolved oxygen in ponds. Tested on a real dataset predicting the dissolved oxygen value of the pond space in the next 24 hours, the root mean square error (RMSE) was 1.263 3 mg/L, the average absolute error (MAE) was 0.975 5 mg/L, and the average absolute percentage error (MAPE) was 14.873 2%. Compared with common prediction methods, the performance of each evaluation index was greatly improved, which could more accurately realize the spatio-temporal prediction of dissolved oxygen in ponds.

**Key words:** pond farming; dissolved oxygen; spatio-temporal prediction; DeepAR model

## 0 引言

溶解氧(Dissolved oxygen, DO)是影响水质的重要指标之一,其含量直接关系水生生物的生存与繁殖<sup>[1]</sup>。池塘溶解氧易受多种因素的影响,会产生时间和空间上分布的差异,即在同一位置不同时刻溶解氧状态不同,不同位置溶解氧状态也不同<sup>[2]</sup>。养殖水产品栖息在不同的水层中,因此对池塘溶解氧进行时空精准预测,及时准确掌握未来池塘空间内溶解氧的变化情况,对于实现池塘溶解氧精细化管理、提高养殖效益具有重要意义。

池塘溶解氧时空预测分为时间序列预测和空间预测,即对池塘某一位置未来时刻溶解氧变化进行预测以及某一时刻池塘空间不同位置处的溶解氧状态进行预测。目前,国内外学者开展了溶解氧预测研究,在溶解氧时间序列预测<sup>[3-10]</sup>方面,朱成云等<sup>[4]</sup>构建了基于混沌变异的分布估计(Distribution estimation algorithm of chaotic mutation, CMEDA)优化最小二乘支持向量机回归(Least square support vector regression, LSSVR)的溶解氧预测模型,实现鱼塘养殖溶解氧的预测。REN等<sup>[7]</sup>采用变分模态分解(Variational mode decomposition, VMD)方法分解原始数据,并将分解后的数据作为深度信念网络(Deep belief network, DBN)的输入构建预测模型,实现对循环水养殖中连续4 h溶解氧含量的预测。CAO等<sup>[8]</sup>建立了基于门控递归单元(Gated recurrent unit, GRU)的溶解氧时间序列预测模型,对不同时间区间内水产养殖水体的溶解氧含量进行预测。LIU等<sup>[9]</sup>提出了基于注意力机制(Attention)和递归神经网络(Recurrent neural network, RNN)的溶解氧多步预测模型,实现对不同时间段溶解氧含量的预测。目前的研究大多只

在溶解氧短期预测方面取得良好结果,随着预测时长的增加,预测误差逐渐增大,无法满足溶解氧预测的更高要求。

溶解氧空间预测研究相对较少。已有研究人员运用流体力学等机理方法建立空间水动力模型,但目前空间水动力模型多应用于污水排放、洪水管理等领域<sup>[11-15]</sup>。机理方法需要采集池塘空间的生物呼吸量、光合产氧量以及有机物含量等多种参数,建立数学模型,这些参数通常较难获取。随着水产养殖大数据技术的发展,基于数据的建模技术成为水产养殖大数据分析的趋势<sup>[16]</sup>。CHEN等<sup>[17]</sup>采用径向基神经网络(RBFNN)建立养殖池塘溶解氧空间插值模型,实现对整个养殖池塘各位置溶解氧的插值预测。空间预测存在的问题是,对未来池塘空间溶解氧插值预测时需先对各个监测点建立模型进行溶解氧的预测,各个监测点预测结果的累积误差较大,无法做到长时间的空间预测。

针对上述问题,本文提出一种基于单监测点的池塘溶解氧时空预测方法。首先根据各个监测点原始溶解氧序列样本熵(SE)以及序列之间的最大互信息系数(MIC)综合选取中心监测点,建立空间坐标系;其次采用自回归循环神经网络(DeepAR)算法对中心监测点溶解氧序列进行预测;最后采用正则化极限学习机(RELM)算法构建中心监测点与池塘各个位置溶解氧之间的空间映射关系模型,实现对未来时刻池塘溶解氧的空间预测。

## 1 材料和方法

### 1.1 数据采集

以江苏省宜兴市高滕镇物联网水产养殖基地的蟹塘为实验场地,池塘中部署了溶解氧等多种水质传感器,池塘边安置一个小型气象站,采集数据类型

如表 1 所示。本文采集 2019 年 7 月 9—22 日 8 个溶解氧监测传感器共 15 536 条监测数据。蟹塘中 8 个溶解氧监测点的位置分布俯视图和剖面图如图 1 所示,在水平方向呈现三角分布,在垂直方向分为上、中、下 3 层,可以在得到池塘溶解氧更大监测范围的同时,掌握不同位置不同水层溶解氧的变化。在点  $E$ 、 $F$  处于垂直方向分别放置 3 个传感器,分别距水面 0.2、0.4、0.6 m,点  $D$  处水位较浅,放置 2 个传感器,分别距水面 0.2、0.4 m,其中点  $E_2$  处传感器

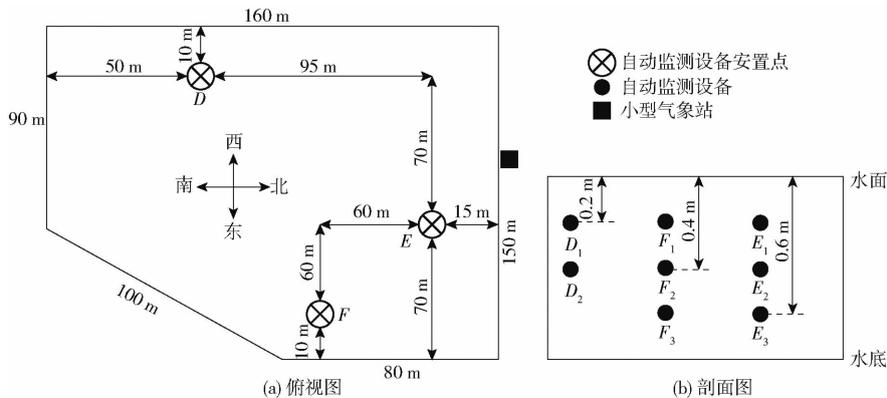


图 1 池塘监测点分布示意图

Fig. 1 Distribution diagram of pond monitoring points

系统采集数据时,受到浑浊水质以及网络抖动等影响,会造成采集样本中存在异常值和缺失值,处理方法如下:

采用线性插值方法对异常值进行填充,假设  $x_k$  和  $x_{k+c}$  时的溶解氧分别为  $y_k$  和  $y_{k+c}$ ,时刻  $x_k$  和  $x_{k+c}$  之间存在异常值,利用过点  $(x_k, y_k)$  和  $(x_{k+c}, y_{k+c})$  的直线  $L(x)$  近似表示原时序关系,线性插值公式如下

$$L(x) = y_k + \frac{y_{k+c} - y_k}{x_{k+c} - x_k}(x - x_k) \quad (1)$$

式中  $k$ ——时刻

$c$ ——上一个正常时刻到下一个正常时刻的间隔

$x \in [x_k, x_{k+c}]$ ,输入样本异常时刻,将结果填充到异常值位置处。

采用均值法对缺失值进行填充,假设时刻  $k-1$  和  $k+1$  之间存在缺失值,缺失值计算公式如下

$$y_k = \frac{y_{k-1} + y_{k+1}}{2} \quad (2)$$

## 1.2 溶解氧时空预测总体流程

为了准确预测未来时刻池塘空间溶解氧状况,本文提出了基于 DeepAR-RELM 的池塘溶解氧时空预测方法,流程图如图 2 所示。

具体步骤如下:①中心监测点选取。根据各个监测点原始溶解氧序列的样本熵以及序列之间的最大互信息系数综合选取中心监测点,并以中心监测

采集的水质参数包括溶解氧含量、水温、pH 值、叶绿素含量、浊度,其余点只采集溶解氧参数。

表 1 采集数据类型

Tab. 1 Type of data collected

类型	采集参数	采集周期/ (min·次 <sup>-1</sup> )
水质	溶解氧含量、水温、pH 值、叶绿素含量、浊度	10
气象	大气温度、大气湿度、太阳辐射、风速、风向、大气压强、雨量	1

⊗ 自动监测设备安置点

● 自动监测设备

■ 小型气象站

水面

水底

0.2 m

0.4 m

0.6 m

图 2 池塘溶解氧时空预测方法流程图

Fig. 2 Flow chart of spatio-temporal prediction method of dissolved oxygen in pond

点为原点建立池塘空间坐标系。②时间序列预测。采用 DeepAR 算法构建中心监测点的溶解氧时间序列预测模型,实现中心监测点溶解氧时间序列预测。③空间预测。采用 RELM 算法构建中心监测点与池塘各点溶解氧之间的空间映射关系模型,实现池塘溶解氧的空间预测。

### 1.3 中心监测点选取方法

本文为了描述池塘空间各点位置信息,需要选取中心监测点,建立池塘空间坐标系,并对中心监测点溶解氧进行时间序列预测,作为空间预测模型的输入。

由于各个监测点所在位置和环境不同,导致各个监测点的溶解氧序列波动性以及它们之间的相关性存在差异。样本熵 (SE) 是一种用于度量时间序列波动程度的方法,熵越小,序列波动程度则越小,越容易对其进行预测<sup>[18]</sup>。最大互信息系数 (MIC)

可以衡量两个序列之间的相关性,相关性反映了序列之间的关联程度,相关性越大,越容易从一个序列中得到另一个序列的信息<sup>[19]</sup>。因此,本文采用样本熵衡量每个监测点溶解氧序列的波动程度,采用最大互信息系数衡量监测点溶解氧序列之间的相关性,综合选取出序列波动程度较小且与各个监测点相关性较大的监测点作为中心监测点。

假设池塘监测点个数为  $N$ ,其中第  $i$  个监测点溶解氧序列为  $X_i = \{x_{it}\} (t = 1, 2, \dots, n; i = 1, 2, \dots, N)$ ,定义

$$B^m(r) = \frac{1}{n-m+1} \sum_{j=1}^{n-m+1} \frac{K_j}{n-m+1} \quad (3)$$

式中  $n$ ——第  $i$  个监测点溶解氧时间序列长度

$m$ ——向量序列维数

$B^m(r)$ ——2 个  $m$  维向量在相似容限  $r$  下匹配  $m$  个点的概率

$K_j$ ——2 个向量之间距离小于等于  $r$  的数目

第  $i$  个监测点溶解氧序列的样本熵为

$$S_{Ei}(n, m, r) = -\ln \frac{B^{m+1}(r)}{B^m(r)} \quad (4)$$

式中  $S_{Ei}$ ——第  $i$  个监测点溶解氧序列的样本熵

将 2 个监测点溶解氧序列  $X_i$  和  $X_j$  的值域分别划分为  $x$  段和  $y$  段,获取  $xy$  个网格划分结果,则监测点  $i$  和  $j$  溶解氧序列  $X_i$  和  $X_j$  之间的最大互信息系数  $M_{IC}(X_i, X_j)$  为

$$M_{IC}(X_i, X_j) = \max_{xy < R(n)} \frac{I^*(D, x, y)}{\lg \min\{x, y\}} \quad (5)$$

式中  $D$ ——2 个溶解氧序列的有序对集合

$I^*(D, x, y)$ ——网格划分后各个网格中互信息的最大值

$R(n)$ ——网格划分数量  $xy$  上限值

本文计算第  $i$  个监测点与各个监测点溶解氧序列的最大互信息系数的和以及第  $i$  个监测点溶解氧序列的样本熵  $S_{Ei}$ ,选取两者差值  $M_{Si}$  最大的监测点作为中心监测点,计算公式如下

$$\begin{cases} M_{Si} = \sum_{j=1}^N M_{IC}(X_i, X_j) - S_{Ei} \\ \max\{M_{Si}\} \rightarrow i \end{cases} \quad (6)$$

本文中,共有 8 个监测点,最后选取  $D_1$  监测点作为中心监测点。 $m = 2, r = 0.25s_{id}, s_{id}$  为序列的标准差,  $R(n) = n^{0.6}$ 。

#### 1.4 基于 DeepAR 的溶解氧预测方法

溶解氧序列数据具有非线性和非平稳的特点。自回归循环神经网络(DeepAR)是一种适用于时间序列预测的监督学习算法,具有较强的非线性拟合能力,它在递归神经网络(RNN)基础上能够根据时

间序列的过去生成时间序列未来的点预测和概率预测<sup>[20]</sup>。因此,本文采用 DeepAR 算法对中心监测点溶解氧时间序列进行预测。

设中心监测点溶解氧序列为  $Z = \{z_t\} (t = 1, 2, \dots, t_0)$ ,模型目标为估计后续  $T$  个时间步长的概率分布  $p$ 。

$$p(z_{t_0+t_0+T} | z_{1:t_0}) = \prod_{t=t_0}^{t_0+T} l(z_t | \theta(h_t, \Theta)) \quad (7)$$

其中  $h_t = H(h_{t-1}, z_{t-1}, \Theta)$  (8)

式中  $l$ ——似然函数

$\theta$ ——似然函数  $l$  的似然参数

$H$ ——RNN 神经元内部函数

$\Theta$ ——模型参数

$h_t$ ——隐藏层输出

$t_0$ ——中心监测点溶解氧序列长度

$z_t$ ——第  $t$  个时刻溶解氧值

本文使用 DeepAR 进行单变量的时间序列预测,模型训练和预测过程如图 3 所示。

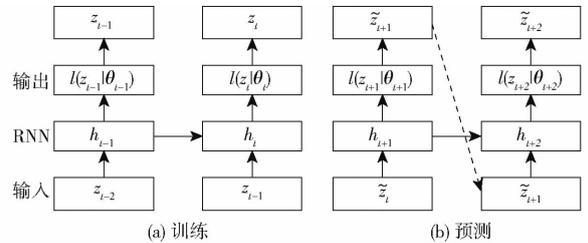


图 3 DeepAR 模型训练和预测过程示意图

Fig. 3 Schematic of DeepAR model training and prediction process

模型训练阶段:在每个时刻  $t$ ,网络的输入包括上一个时刻的取值  $z_{t-1}$  以及上一个时刻隐藏层的输出  $h_{t-1}$ ,通过最大化对数似然函数学习模型的参数,对数似然函数的计算公式如下

$$L(\theta) = \sum_{t=1}^{t_0} \ln l(z_t | \theta(h_t)) \quad (9)$$

式中  $L(\theta)$ ——对数似然函数

预测阶段:在每个时刻  $t$  使用蒙特卡洛采样<sup>[21]</sup>得到符合  $l(\cdot | \theta_t)$  的采样值  $\tilde{z}_t$ ,将采样值  $\tilde{z}_t$  输入下一个时间步,重复这个过程,得到一系列  $t_0 \sim t_0 + T$  时刻的采样值,利用这些采样值可以计算所需的目标值,如分位数、期望等。

$\theta(h_t)$  的具体形式取决于似然函数  $l(z | \theta)$ ,本文使用高斯似然函数  $l_c(z | \mu, \sigma)$  预测实值数据的概率,  $\theta = (\mu, \sigma)$ ,计算公式如下

$$l_c(z | \mu, \sigma) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(z-\mu)^2}{2\sigma^2}} \quad (10)$$

$$\mu(h_t) = wh_t + b \quad (11)$$

$$\sigma(h_t) = \lg(1 + \exp(wh_t + b)) \quad (12)$$

式中  $w$ ——权重矩阵  $b$ ——偏置矩阵  
 $\mu(h_i)$ 、 $\sigma(h_i)$ ——高斯分布函数的均值和标准差

本文取  $T = 144$ , 取每个时刻蒙特卡洛采样后所有取值的均值作为中心监测点溶解氧时间序列的预测结果。

### 1.5 基于 RELM 的溶解氧空间预测方法

池塘空间不同位置的溶解氧分布并不均匀, 在同一水层, 不同位置的溶解氧浓度不尽相同, 在同一垂直面, 不同水层的溶解氧浓度也有所不同。中心监测点溶解氧与空间各点溶解氧之间是一种非线性关系, 而正则化极限学习机 (RELM) 具有处理复杂非线性函数的能力<sup>[22]</sup>, 因此, 为了准确获得池塘空间不同位置的溶解氧状态, 本文采用 RELM 构建中心监测点与池塘各点溶解氧之间的空间映射关系模型。

RELM 是一种前馈神经网络, 它在标准极限学习机 (ELM) 的基础上增加正则项系数, 提高模型的泛化能力<sup>[23]</sup>。本文构建 3 层 RELM 模型, 由输入层、隐藏层和输出层以及层间节点连接组成。

假设池塘中的监测点数量为  $N$ , 选取的监测点溶解氧序列长度为  $K$ , 空间预测模型训练集  $s = \{(\mathbf{a}_m, \mathbf{o}_m)\} (m = 1, 2, \dots, M), M = NK, \mathbf{a}_m = (a_{m1}, a_{m2}, a_{m3}, a_{m4}), a_{m1}$  表示选取的中心监测点溶解氧值,  $a_{m2}$ 、 $a_{m3}$ 、 $a_{m4}$  表示池塘监测点的空间坐标值,  $\mathbf{o}_m$  为空间坐标点处的溶解氧值。RELM 模型的表示如下

$$\mathbf{o}_m = \sum_{i=1}^L \beta_i g(\alpha_i \mathbf{a}_m + \mathbf{b}_i) \quad (13)$$

式中  $g(\cdot)$ ——激活函数  
 $L$ ——隐藏层节点的数量  
 $\alpha_i, \mathbf{b}_i$ ——第  $i$  个隐藏层节点输入权重和偏置  
 $\beta_i$ ——第  $i$  个隐藏层节点的输出权重  
 $\mathbf{a}_m, \mathbf{o}_m$ ——输入、输出

模型训练过程中, 隐藏层节点输入权重矩阵  $\alpha$  和隐藏层偏置矩阵  $\mathbf{b}$  均为随机生成, 因此仅需计算隐藏层输出权重矩阵  $\beta$ , 表示如下

$$\beta = \left( \mathbf{H}^T \mathbf{H} + \frac{\mathbf{I}}{\lambda} \right)^{-1} \mathbf{H}^T \mathbf{S} \quad (14)$$

式中  $\mathbf{H}$ ——隐藏层的输出矩阵  
 $\mathbf{S}$ ——模型训练阶段输入的参数矩阵  
 $\mathbf{I}$ ——单位矩阵  
 $\lambda$ ——正则化系数

预测阶段, 根据训练阶段得到的隐藏层节点输入权重矩阵  $\alpha$  和隐藏层偏置矩阵  $\mathbf{b}$  以及隐藏层输出权重矩阵  $\beta$ , 输入预测参数  $\mathbf{a}$ , 得到溶解氧空间预测结果  $\mathbf{o}$  为

$$\mathbf{o} = \beta g(\alpha \mathbf{a} + \mathbf{b}) \quad (15)$$

### 1.6 模型性能和评估方法

实验开发环境如下: 编程语言是 Python 3.7 (64 位), 开发环境为 Anaconda 3。

本文搭建含有两层长短期记忆网络 (Long short-term memory, LSTM) 隐藏层的 DeepAR 模型, 隐藏层神经元个数均为 40, 批处理尺寸为 32, 优化器选择 Adam, 学习率为 0.01, Dropout 参数为 0.1, 模型迭代次数为 500。RELM 模型的隐藏层激活函数为 Sigmoid 函数, 利用网格搜索方法<sup>[24]</sup> 确定正则化系数  $\lambda = 0.2$  和隐藏层节点数量  $L = 10$ , 其网络拓扑结构为 4-10-1。RELM 模型的部分训练数据如表 2 所示。

表 2 训练数据  
 Tab.2 Training data

输入 $\mathbf{a}_m$				输出 $\mathbf{o}_m /$
$a_{m1} / (\text{mg} \cdot \text{L}^{-1})$	$a_{m2} / \text{m}$	$a_{m3} / \text{m}$	$a_{m4} / \text{m}$	$(\text{mg} \cdot \text{L}^{-1})$
9.28	70	95	0	8.54
9.28	70	95	0.2	8.32
9.28	70	95	0.4	7.20
9.58	130	35	0.4	7.04

选用均方根误差 (RMSE)、平均绝对误差 (MAE) 和平均绝对百分比误差 (MAPE) 作为模型评价指标。

## 2 实验

### 2.1 中心监测点选取

本文共有 8 个监测点, 根据式 (6) 计算每个监测点与各个监测点溶解氧序列的最大互信息系数  $M_{ic}$  的和以及每个监测点溶解氧序列的样本熵  $S_E$ , 并计算两者的差值  $M_s$ , 结果如图 4 所示。

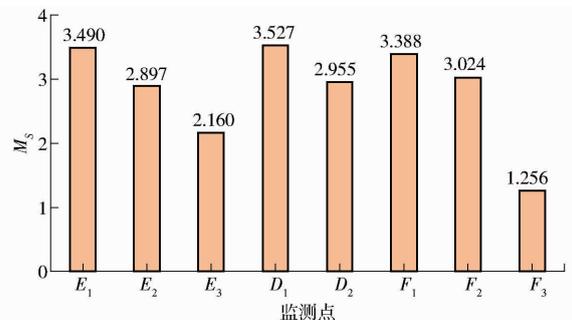


图 4 各个监测点溶解氧序列  $M_s$  计算结果

Fig.4  $M_s$  calculation results of dissolved oxygen sequence at each monitoring point

从图中可以看出,  $D_1$  监测点的  $M_s$  值最大, 说明其溶解氧序列样本熵小, 序列随机程度低, 较容易对其未来状态进行准确预测, 且与其余监测点的相关性大, 可以从其序列中得到更多其他监测点溶解氧

的信息。因此选择点  $D_1$  为中心监测点,进行时间序列预测,并以此点为原点,池塘东向为  $X$  轴正方向,北向为  $Y$  轴正方向,由水面向水底为  $Z$  轴正方向,建立空间坐标系,空间坐标系示意图如图 5 所示。

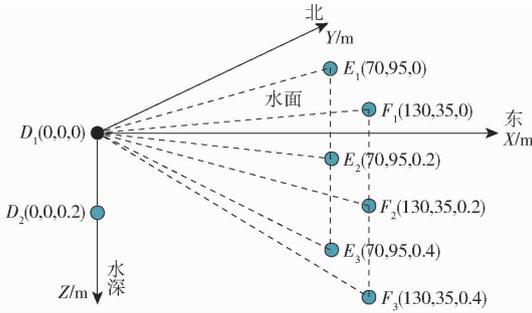


图 5 池塘空间坐标系示意图

Fig. 5 Coordinate diagram system of pond space

## 2.2 中心监测点溶解氧预测结果

将 2019 年 7 月 9—21 日  $D_1$  监测点 1 798 组溶解氧数据作为时间序列预测模型的训练集,7 月 22 日 144 组溶解氧数据作为测试集。

选用 LSTM、GRU 和 Attention - GRU 模型与 DeepAR 模型进行比较。这些模型都是常用的时间序列预测算法模型,具有一定的代表性。每个模型都使用相同的训练数据集和测试数据集。图 6 为 4 种模型在 2019 年 7 月 22 日  $D_1$  监测点的溶解氧预测曲线。

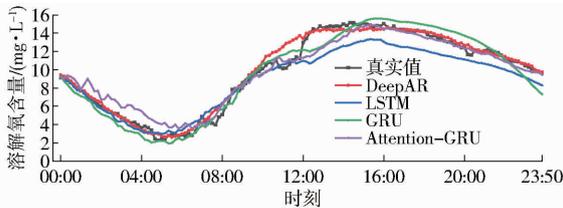


图 6 4 种时间序列预测模型预测曲线

Fig. 6 Result curves of four time series prediction models

计算出预测 24 h 共 144 个时间点整体的 RMSE、MAE 和 MAPE,其预测结果如表 3 所示。

在预测未来 24 h 溶解氧时,本文模型的 3 种指标比 Attention - GRU 模型分别下降了 0.298 8 mg/L、0.295 4 mg/L、6.313 6 个百分点,比 GRU 模型分别下降了 0.318 6 mg/L、0.393 1 mg/L、4.353 1 个百分

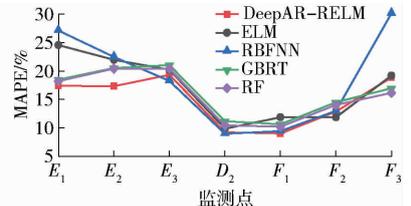
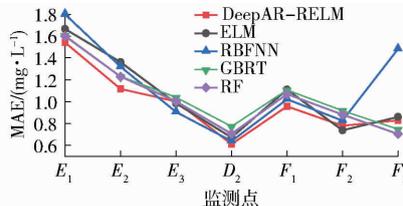
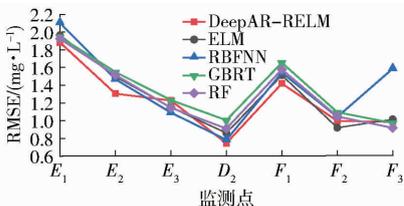


图 7 5 种空间预测模型在各监测点处预测结果曲线

Fig. 7 Prediction result curves of five spatial prediction models at each monitoring point

表 3 4 种模型预测结果对比

Tab. 3 Comparison of prediction results of four models

模型	RMSE/( $\text{mg}\cdot\text{L}^{-1}$ )	MAE/( $\text{mg}\cdot\text{L}^{-1}$ )	MAPE/%
DeepAR	0.681 7	0.449 9	5.950 1
LSTM	1.426 0	1.161 9	12.418 5
GRU	1.000 3	0.843 0	10.303 2
Attention - GRU	0.980 5	0.745 3	12.263 7

点,比 LSTM 模型分别下降了 0.744 3 mg/L、0.712 0 mg/L、6.468 7 个百分点,说明本文模型比这 3 种模型具有更好的预测效果,更加适合长时间的溶解氧时间序列预测,也为池塘溶解氧空间预测提供了更加精准的输入。

## 2.3 溶解氧空间预测结果

将各个监测点 14 384 组溶解氧数据作为空间模型的训练集,将 7 月 22 日除  $D_1$  监测点外的其余监测点 1 008 组溶解氧数据作为测试集。

为了评估本文所提出的空间预测模型的性能,选用标准极限学习机 (ELM)、径向基神经网络 (RBFNN)、梯度提升回归树 (GBRT) 和随机森林 (RF) 算法与本文模型在相同的数据集上进行比较。每个模型的输入为 DeepAR 模型在  $D_1$  监测点处的溶解氧预测值以及各点的空间坐标。图 7 为 5 种模型的 RMSE、MAE、MAPE 预测误差曲线。由 5 种模型的误差曲线可发现,ELM 模型和 RBFNN 模型的波动性较大,GBRT 模型和 RF 模型次之,RELM 模型虽然在某些位置处的预测结果不如其他模型,但是预测误差表现最为稳定。其主要原因是 RELM 模型保留了 ELM 模型的优良泛化能力,同时正则化减少了 ELM 模型的过拟合,使得预测结果更加稳定。

5 种模型对所有监测点预测的 RMSE、MAE 和 MAPE 的结果如表 4 所示。

由表 4 可以看出,本文提出的模型具有更好的准确性,在各种性能标准 (RMSE、MAE、MAPE) 方面都优于其他模型。本文模型的 RMSE、MAE 和 MAPE 比其他模型都小,说明误差较大的点少于其他模型,该模型预测值更接近真实值。

为了进一步比较各空间预测模型的预测效果,

表 4 5 种预测模型的预测结果对比

Tab. 4 Comparison of prediction results of five prediction models

模型	RMSE/(mg·L <sup>-1</sup> )	MAE/(mg·L <sup>-1</sup> )	MAPE/%
DeepAR - RELM	1.263 3	0.975 5	14.873 2
ELM	1.320 7	1.055 5	17.088 4
RBFNN	1.425 9	1.141 7	18.487 9
GBRT	1.386 8	1.059 6	16.195 4
RF	1.333 3	1.026 8	15.693 7

本文计算了 5 种模型在所有监测点预测结果的绝对百分比误差(APE),公式如下

$$A_{PE} = \left| \frac{y - \hat{y}}{y} \right| \times 100\% \quad (16)$$

式中  $y$ ——真实值  $\hat{y}$ ——预测值

将 APE 分为 3 个范围,0 ~ 20% 是非常满意和可使用的误差范围,20% ~ 40% 是误差相对较大,但仍可接受,超过 40% 时结果不可用。结果如图 8 所示,本文模型中,只有 56 个点大于 40%,有 219 个点在 20% ~ 40% 内,均小于其他模型,有 733 个点小于 20%,多于其他模型。显然,本文模型具有较高的预测精度。

本文利用训练好的模型生成 2019 年 7 月 22 日 12:00 养殖池塘空间中 0.2 m 和 0.4 m 水层平面的

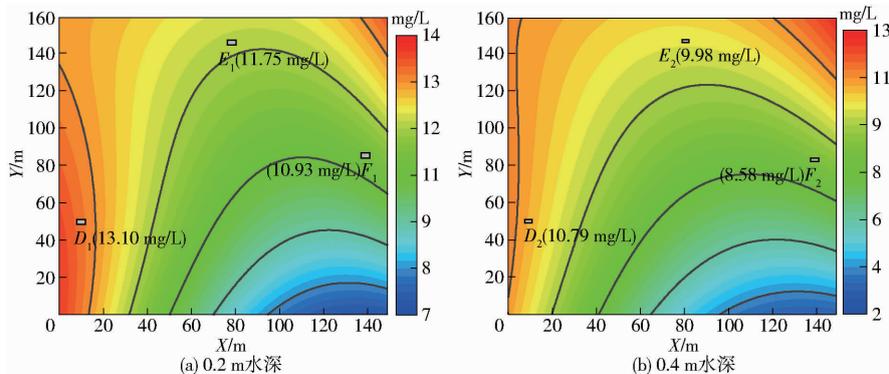


Fig. 9 Prediction results of dissolved oxygen in different water layers

表 5 气象数据

Tab. 5 Meteorological data

时间	相对湿度/%	温度/°C	压强/kPa	风速/(m·s <sup>-1</sup> )	风向/(°)	太阳辐射/(W·m <sup>-2</sup> )	雨量/mm
2019-07-22 11:57:08	78.9	32.2	100.14	2.1	142.0	766.0	0
2019-07-22 11:58:13	78.9	32.0	100.14	1.0	122.0	777.0	0
2019-07-22 11:59:08	78.6	32.2	100.13	1.4	107.0	770.0	0
2019-07-22 12:00:07	78.5	32.2	100.11	1.5	106.0	749.0	0

### 3 结束语

本文提出一种基于 DeepAR - RELM 的池塘溶解氧时空预测方法,用于预测未来时刻的池塘溶解氧含量空间分布。首先选取中心监测点,建立

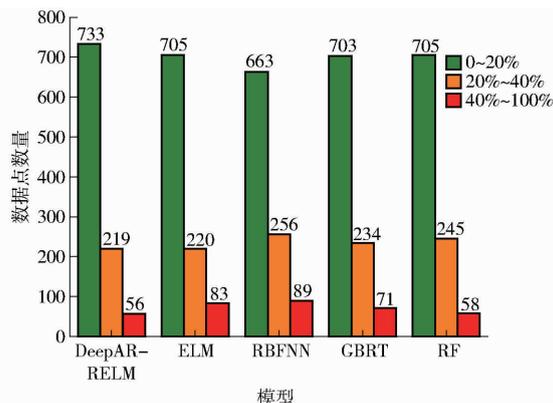


图 8 APE 结果统计

Fig. 8 APE result statistics

150 × 160 个坐标点溶解氧值,结果如图 9 所示。池塘附近的部分气象数据如表 5 所示。可以看出在养殖池塘中,溶解氧含量分布不均匀。在同一水层,点 D 和点 E 相对于点 F 处于池塘下风口处,而点 D 和点 E 溶解氧含量较高,点 F 溶解氧含量较低,可以看出在风力作用下,溶解氧含量高的表层水移到下风口沿岸,致使下风口处的溶解氧含量相对较高。在不同水层,溶解氧含量的分布受水层所接收到的光照和温度影响,呈现出从上到下垂直递减状态。因此,点 D<sub>2</sub>、E<sub>2</sub>、F<sub>2</sub> 处的溶解氧含量比点 D<sub>1</sub>、E<sub>1</sub>、F<sub>1</sub> 处的低。

池塘空间坐标系;其次采用 DeepAR 算法构建中心监测点的溶解氧含量时间序列预测模型,实现中心监测点溶解氧时间序列预测;最后采用 RELM 模型构建中心监测点与池塘各点溶解氧含量之间的空间映射关系模型,实现对未来时刻池塘溶解

氧的空间预测。实验结果表明,本文方法可以较准确地实现池塘溶解氧含量时空预测,与常见预测模型进行对比,各评价指标的性能均有较大幅度提升。

## 参 考 文 献

- [1] RAHMAN A, DABROWSKI J, MCCULLOCH J. Dissolved oxygen prediction in prawn ponds from a group of one step predictors [J]. *Information Processing in Agriculture*, 2020, 7(2): 307-317.
- [2] 饶伟, 杨卫中, 位耀光, 等. 鱼菜共生水体溶解氧时空变化规律及其影响因素研究[J/OL]. *农业机械学报*, 2017, 48(增刊): 374-380.  
RAO Wei, YANG Weizhong, WEI Yaoguang, et al. Temporal and spatial variability of water dissolved oxygen with influence factors in aquaponics system [J/OL]. *Transactions of the Chinese Society for Agricultural Machinery*, 2017, 48(Supp.): 374-380. [http://www.jcsam.org/jcsam/ch/reader/view\\_abstract.aspx?flag=1&file\\_no=2017s057&journal\\_id=jcsam](http://www.jcsam.org/jcsam/ch/reader/view_abstract.aspx?flag=1&file_no=2017s057&journal_id=jcsam). DOI:10.6041/j.issn.1000-1298.2017.S0.057. (in Chinese)
- [3] 刘双印, 徐龙琴, 李道亮, 等. 基于蚁群优化最小二乘支持向量回归机的河蟹养殖溶解氧预测模型[J]. *农业工程学报*, 2012, 28(23): 167-175.  
LIU Shuangyin, XU Longqin, LI Daoliang, et al. Dissolved oxygen prediction model of eriocheir sinensis culture based on least squares support vector regression optimized by ant colony algorithm [J]. *Transactions of the CSAE*, 2012, 28(23): 167-175. (in Chinese)
- [4] 朱成云, 刘星桥, 李慧, 等. 工厂化水产养殖溶解氧预测模型优化[J/OL]. *农业机械学报*, 2016, 47(1): 273-278.  
ZHU Chengyun, LIU Xingqiao, LI Hui, et al. Optimization of prediction model of dissolved oxygen in industrial aquaculture [J/OL]. *Transactions of the Chinese Society for Agricultural Machinery*, 2016, 47(1): 273-278. [http://www.j-csam.org/jcsam/ch/reader/view\\_abstract.aspx?file\\_no=20160137&flag=1&journal\\_id=jcsam](http://www.j-csam.org/jcsam/ch/reader/view_abstract.aspx?file_no=20160137&flag=1&journal_id=jcsam). DOI:10.6041/j.issn.1000-1298.2016.01.037. (in Chinese)
- [5] YAN J X, YU L J, MAO W W, et al. Study on prediction model of dissolved oxygen about water quality monitoring system based on BP neural network [J]. *Advanced Materials Research*, 2014, 3137: 1407-1411.
- [6] KISI O, ALIZAMIR M, DOCHESHMEH G A. Dissolved oxygen prediction using a new ensemble method. [J]. *Environmental Science and Pollution Research International*, 2020, 27(9): 9589-9603.
- [7] REN Q, WANG X Y, LI W S, et al. Research of dissolved oxygen prediction in recirculating aquaculture systems based on deep belief network [J]. *Aquacultural Engineering*, 2020, 90: 102085.
- [8] CAO X K, LIU Y R, WANG J P, et al. Prediction of dissolved oxygen in pond culture water based on K-means clustering and gated recurrent unit neural network [J]. *Aquacultural Engineering*, 2020, 91: 102122.
- [9] LIU Y Q, ZHANG Q, SONG L H, et al. Attention-based recurrent neural networks for accurate short-term and long-term dissolved oxygen prediction [J]. *Computers and Electronics in Agriculture*, 2019, 165: 104964.
- [10] CAO W J, HUAN J, LIU C, et al. A combined model of dissolved oxygen prediction in the pond based on multiple-factor analysis and multi-scale feature extraction [J]. *Aquacultural Engineering*, 2019, 84: 50-59.
- [11] ZHANG C, GAO X, WANG L, et al. Analysis of agricultural pollution by flood flow impact on water quality in a reservoir using a three-dimensional water quality model [J]. *Journal of Hydroinformatics*, 2013, 15(4): 1061-1072.
- [12] LEE H, CHUNG S, RYU I, et al. Three-dimensional modeling of thermal stratification of a deep and dendritic reservoir using ELCOM model [J]. *Journal of Hydro-environment Research*, 2013, 7(2): 124-133.
- [13] PRESTON A, HANNOUN I A, LIST E J, et al. Three-dimensional management model for Lake Mead, Nevada, Part 2: findings and applications [J]. *Lake and Reservoir Management*, 2014, 30(3): 303-319.
- [14] 李自立, 刘奥琦, 莫旭涛. 一种基于表面状态参数的三维海洋动力学模型数值求解方法 [J]. *热带海洋学报*, 2020, 39(1): 12-19.  
LI Zili, LIU Aoqi, MO Xutao. A numerical method for solving the three-dimensional ocean dynamics model based on surface state parameters [J]. *Journal of Tropical Oceanography*, 2020, 39(1): 12-19. (in Chinese)
- [15] 周正印, 杨楠. 基于三维数值模型的河道水质动态模拟研究 [J]. *环境保护科学*, 2019, 45(6): 108-113.  
ZHOU Zhengyin, YANG Nan. Study on dynamic simulation of river water quality based on three-dimensional numerical model [J]. *Environmental Protection Science*, 2019, 45(6): 108-113. (in Chinese)
- [16] 段青玲, 刘怡然, 张璐, 等. 水产养殖大数据技术研究进展与发展趋势分析 [J/OL]. *农业机械学报*, 2018, 49(6): 1-16.  
DUAN Qingling, LIU Yiran, ZHANG Lu, et al. State-of-the-art review for application of big data technology in aquaculture [J/OL]. *Transactions of the Chinese Society for Agricultural Machinery*, 2018, 49(6): 1-16. [http://www.j-csam.org/jcsam/ch/reader/view\\_abstract.aspx?file\\_no=20180601&flag=1&journal\\_id=jcsam](http://www.j-csam.org/jcsam/ch/reader/view_abstract.aspx?file_no=20180601&flag=1&journal_id=jcsam). DOI:10.6041/j.issn.1000-1298.2018.06.001. (in Chinese)
- [17] CHEN Y Y, YU H H, CHENG Y J, et al. A hybrid intelligent method for three-dimensional short-term prediction of dissolved oxygen content in aquaculture [J]. *PloS one*, 2018, 13(2): e0192456.
- [18] YIN J, XIAO P X, LI J, et al. Parameters analysis of sample entropy, permutation entropy and permutation ratio entropy for RR interval time series [J]. *Information Processing and Management*, 2020, 57(5): 102283.
- [19] 梁吉业, 冯晨娇, 宋鹏. 大数据相关分析综述 [J]. *计算机学报*, 2016, 39(1): 1-18.  
LIANG Jiye, FENG Chenjiao, SONG Peng. A survey on correlation analysis of big data [J]. *Chinese Journal of Computers*, 2016, 39(1): 1-18. (in Chinese)
- [20] SALINAS D, FLUNKERT V, GASTHAUS J, et al. DeepAR: probabilistic forecasting with autoregressive recurrent networks [J]. *International Journal of Forecasting*, 2020, 36(3): 1181-1191.
- [21] YOUNESI A, SHAYEGHI H, SAFARI A, et al. Assessing the resilience of multi microgrid based widespread power systems against natural disasters using monte carlo simulation [J]. *Energy*, 2020, 207: 118220.
- [22] LIU H, WU H, LI Y. Smart wind speed forecasting using EWT decomposition, GWO evolutionary optimization, RELM learning and IEWT reconstruction [J]. *Energy Conversion and Management*, 2018, 161: 266-283.
- [23] SUN W, LIU M. Wind speed forecasting using FEEMD echo state networks with RELM in Hebei, China [J]. *Energy Conversion and Management*, 2016, 114: 197-208.
- [24] WANG X S, GONG G H, LI N, et al. Detection analysis of epileptic EEG using a novel random forest model combined with grid search optimization. [J]. *Frontiers in Human Neuroscience*, 2019, 13: 52.