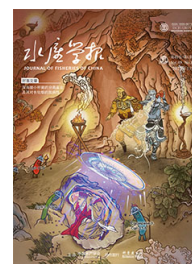


DOI: 10.11964/jfc.20220613532

基于不同投喂策略的循环水养殖系统氨氮预测模型



孙雪倩¹, 李丽^{1,2*}, 董双林^{1,2}, 田相利^{1,2}, 张盛坤^{1,2}

1. 中国海洋大学, 海水养殖教育部重点实验室, 山东 青岛 266003; 2. 青岛海洋科学与技术试点国家实验室, 海洋渔业科学与食物产出过程功能实验室, 山东 青岛 266237

摘要:

【目的】实现对循环水养殖系统水体中总氨氮 (TAN) 浓度的预测, 并研究投喂策略对 TAN 预测模型预测精度的影响。

【方法】本研究测定了斑石鲷养殖池内 7 个水质指标, 采用主成分分析 (PCA) 和 Pearson 相关性分析法对数据进行前处理并形成三种数据集: 原始数据集 (OD)、Pearson 数据集 (Pearson D) 和 PCA 数据集 (PCAD), 结合随机森林 (RF)、BP 神经网络 (BP)、门控循环单元 (GRU)、长短期记忆网络 (LSTM) 这 4 种模型, 对两种投喂策略下养殖水体中的 TAN 浓度进行预测, 并采用均方根误差 (RMSE)、均方误差 (MSE)、平均绝对误差 (MAE) 和 R 方值 (R^2 -score) 对模型进行评估。

【结果】RF 模型的预测效果最差, 随着投喂策略的改变, GRU 与 LSTM 模型预测精度较高且稳定, 而 BP 模型预测精度波动较大。不同投喂阶段筛选出的最优预测模型不同, 人工和自动化投喂阶段的最优模型分别为 Pearson D-BP 和 Pearson D-GRU 模型, 在整个实验周期中, PCAD-LSTM 模型、Pearson D-LSTM 模型和 Pearson D-GRU 模型预测性能较好。人工投喂阶段与自动化投喂阶段相比, Pearson D-LSTM 模型的 RMSE、MSE 和 MAE 分别降低了 0.007 2、0.001 9 和 0.003 6, R^2 -score 升高了 0.107 5; Pearson D-GRU 模型的 RMSE、MSE 和 MAE 分别降低了 0.003 0、0.000 8 和 0.003 0, R^2 -score 升高了 0.082 6。

【结论】投喂策略会影响 TAN 预测模型的预测精度, 结合 Pearson 分析的 GRU 或 LSTM 模型可很好地实现该系统养殖水体中 TAN 的预测, 该结果可为 RAS 氨氮预测技术的优化提供参考。

关键词:斑石鲷; 循环水养殖系统; 投喂策略; 总氨氮; 水质预测模型

第一作者: 孙雪倩, 从事水产养殖生态学研究, E-mail: 972086765@qq.com



通信作者: 李丽, 从事水产养殖生态学研究, E-mail: l_li@ouc.edu.cn

资助项目: 国家重点研发计划 (2020YFD0900201, 2017YFE0122100)

收稿日期: 2022-06-08

修回日期: 2022-09-14

文章编号:

1000-0615(2025)01-019611-12

中图分类号: S 969; TP 183

文献标志码: A

作者声明本文无利益冲突

©《水产学报》编辑部(CC BY-NC-ND 4.0)
Copyright © Editorial Office of Journal of Fisheries of China (CC BY-NC-ND 4.0)

随着水产养殖技术的不断提高, 规模化、集约化水产养殖蓬勃发展, 而高密度养殖、人工投饵量持续增加, 导致养殖水体中残饵和粪便积累, 并经微生物分解产生大量的氨氮等有害物质^[1]。养殖水体中的总氨氮 (TAN) 包含离子氨和非离子氨, 其中非离子氨对养殖生物毒性很大, 能够抑制其生长发育、降低其免疫力、损伤细胞等^[2-4]。因此,



高效、准确地检测氨氮对于防止养殖水质恶化和预防养殖动物病害发生具有重要意义。

检测氨氮浓度的主要方法有纳氏试剂比色法、水杨酸光度法、次溴酸钠氧化法、电化学法、分子吸收光谱法等^[5], 这些方法均为离线检测。目前有一些多参数水质分析仪, 如美国公司生产的 YSI 6820/6920、德国的 WTW Multi 9630 等可对水体氨氮浓度进行实时监测, 但存在成本高、维护困难等问题, 而且大部分仪器都仅适用于淡水而不是海水^[6]。

近年来, 机器学习方法包括时间序列法、人工神经网络法、支持向量机等, 被广泛用于水产养殖水质因子的预测和管理^[7-8]。养殖水体中氨氮浓度易受溶解氧 (DO)、pH 等环境因子的影响, 因此, 有学者使用机器学习方法利用多种环境因子如温度、pH、电导率、DO 等来预测水中的氨氮浓度, 并取得了较好的效果^[9-11]。Pearson 相关性分析和主成分分析 (PCA) 等多元统计方法能够进行数据解释和模式识别, 常用于评估水质的时空变化以及对大型复杂数据集的解释^[12-13]。研究指出, Pearson、PCA 等多元统计方法可用于探究多种水质参数之间的潜在关系, 并且能够在降低数据集维数的同时解释所有参数, 简化数据分析^[14-15]。部分学者在对养殖水体中氨氮进行预测时, 考虑了氨氮与其他环境因子之间的关系, 通过多元统计分析筛选出影响氨氮浓度的主要环境因子, 在此基础上建立模型预测氨氮浓度, 有助于提高模型预测精度及泛化性能^[16]。

以上关于水体氨氮预测模型的研究多通过优化算法提高模型预测精度, 而对养殖生产活动本身可能通过影响水体氨氮浓度变化规律进而对模型预测精度产生影响关注较少。养殖水体的氨氮浓度与养殖生物摄食、排泄和饲料投喂量有关, 而养殖生物对投喂饲料的消化吸收及排氮率、残饵量等与投喂策略紧密相关^[17]。因此, 投喂策略对养殖水体中氨氮的动态变化规律影响较大。目前, 国内关于投喂策略对养殖水体中氨氮预测模型预测精度的影响还未见报道, 因此, 本实验采用两种投喂策略, 探究投喂策略对氨氮预测模型预测精度的影响。通过测定可能与氨氮相关的 6 个环境变量, 利用 Pearson 相关性分析或 PCA 分析, 对环境变量的关联性进行研究, 筛选与氨氮浓度关联性较

大的环境变量, 实现对水质数据的降维处理; 然后选用 4 种常用预测模型, 对斑石鲷 (*Oplegnathus punctatus*) 循环水养殖系统养殖池内氨氮进行预测, 以期 RAS 氨氮预测技术的优化提供新视角。

1 材料与方法

1.1 数据采集与分析

本实验研究地点为山东省莱州明波水产有限公司的“循五”养殖车间斑石鲷循环水养殖系统 (RAS), 该系统由养殖池、弧形筛、生物滤池、紫外线消毒池和充氧装置 (液氧) 等单元组成。系统运行稳定, 共 7 个养殖池, 每个养殖池体积为 30 m³。系统总水量约为 460 m³, 日换水量占总水量的 5%, 循环次数为 16 次/d。系统 2019 年 1 月开始养殖斑石鲷, 初始养殖密度为 17.29 kg/m³。

实验于 2020 年 7 月 24 日—10 月 9 日进行, 养殖过程中投喂斑石鲷配合饲料。2020 年 7 月 24 日—8 月 21 日采用人工投喂, 每日 07:30 和 16:30 各投喂 1 次; 8 月 22 日—10 月 9 日, 使用自动化投喂机投喂饲料, 从早上 7:30 开始, 每隔 1 小时投喂 1 次, 每日 7 次。实验期间投喂量保持不变, 两种投喂策略下系统日投喂量每池均为 4.2 kg。

实验期间每周选择一天连续 24 h 采集养殖池内水样 (8:30 至次日 7:30, 每小时采集 1 次), 共采集 12 次。水样经 0.45 μm 玻璃纤维滤膜过滤后放入 50 mL 聚乙烯瓶中, 立即带回中国海洋大学水产养殖生态实验室分别测定过滤水样中 TAN、亚硝酸盐氮 (NO₂⁻-N)、硝酸盐氮 (NO₃⁻-N) 和溶解活性磷 (SRP)。TAN、NO₂⁻-N、NO₃⁻-N 和 SRP 使用全自动间断化学分析仪 (德国, CleverChem380) 测量, 使用水质监测仪 (帮邦水产) 现场测定养殖池溶解氧 (DO)、pH 和温度 (T)。

本实验共采集 7 个变量 (TAN、NO₂⁻-N、NO₃⁻-N、SRP、DO、pH、T), 288 组样本, 共 288×7 组数据。其中, 前 120 组样本为人工投喂阶段获取, 后 168 组样本为自动化投喂阶段获取。为保证不同投喂阶段的数据量相同, 选择人工投喂 120×7 组、自动化投喂最后的 120×7 组及整个实验周期的 288×7 组数据, 分别建立氨氮预测模型, 取数据前 80% 为训练集,

后 20% 为测试集。

1.2 模型选择

BP 神经网络 (back propagation, BP) 人工神经网络 (ANN) 是一种灵活的建模工具, 是基于生物驱动计算的数学模型, 具有自适应、自学习、容错等特点, 一些用数学公式无法解决的复杂问题, 可以使用 ANN 模型, 通过神经计算方法进行处理^[18-19]。常用的 ANN 模型是 BP 神经网络, BP 能够通过预测值与实测值之间误差的反向传播来更新模型中的权重和偏差, 其输入值和输出值之间的关系如式 (1)^[20]:

$$y = f\left(\sum wx + b\right) \quad (1)$$

式中, w 为相邻层的权重矩阵; b 为相邻层的偏置向量; f 为激活函数。

BP 神经网络由 3 层组成, 分别为输入层、隐藏层和输出层, 3 层的神经元连接在一起, 而同一层的神经元是分开的, 其中, 多个隐藏层可以提高模型的准确性, 但这也也会使模型的计算复杂度上升, Kolmogorov 定理证明单个隐藏层能够让 ANN 逼近任何复杂的非线性函数, 并在输入和输出层之间建立非线性映射^[21-22]。因此, 本研究建立了一个 3 层的 BP 神经网络, 网络参数的输入节点数量与输入变量数相同, 隐藏节点的数量层设置为 8, 输出层的节点数设置为 1, 数据集训练 100 次。

随机森林 (random forest, RF) RF 是一种组合学习算法, 通过生成许多决策树及其组合计算输出值, 每棵决策树所得到的输出值是由随机向量决定的, 该随机向量通过对森林中同分类决策树独立采样。与其他集成方法相比, RF 计算量较小, 且数据噪声、过度训练、异常值等对其影响较小, 常用于分类和回归预测^[23]。随机森林算法公式^[24]:

$$bt(x, y) = a_n I[I_n(x) = y] - \max_{j \neq y} a_n I[I_n(x) = j] \quad (2)$$

$$GE^* = G_{X, Y} [bt(x, y) < 0] \quad (3)$$

$$G_{X, Y} [(G_\theta(I(x, \theta) = y) - \max_{j \neq y} G_\theta[I(x, \theta) = j]) < 0] \quad (4)$$

式中, $I_n(x)$ 为分类器, $n = 1, 2, \dots, N$; x 为输入向量; y 为输出值; I 为指示函数; a_n 为平均值; $bt(x, y)$ 为特定值分类正确与分为其他类的平均票数之间的差值; X, Y 为 X, Y 上空间

概率; GE^* 为泛化误差; θ 为决策树的随机向量; $I(x, \theta)$ 为基于 x 和 θ 的输出。

长短期记忆网络 (long short-term memory, LSTM) 循环神经网络 (RNN) 包括 LSTM、门控循环单元 (GRU) 等, 常被用作时间序列预测^[25]。Hochreiter 等^[26] 首次提出 LSTM, 是 RNN 的优化网络, 结构包含记忆单元的隐藏层单元, 能够克服 RNN 的缺陷, 其记忆单元由 3 个门组成, 分别为遗忘门、输出门和输入门, 输入门和输出门主要用于控制记忆单元输入和输出到网络其余部分的信息, 遗忘门将具有高权重的输出信息从前一个神经元传递到下一个神经元^[27]。其表达式^[25]:

$$h_i = \sigma(W_{nh}n_{i-1} + W_{xh}x_i + e_h) \quad (5)$$

$$s_i = \sigma(W_{ns}n_{i-1} + W_{xs}x_i + e_s) \quad (6)$$

$$o_i = \sigma(W_{no}n_{i-1} + W_{xo}x_i + e_o) \quad (7)$$

$$C'_i = \tanh(W_{nc}n_{i-1} + W_{xc}x_i + e_c) \quad (8)$$

$$C_i = h_i * C_{i-1} + s_i * C'_i \quad (9)$$

$$n_i = o_i * \tanh(C_i) \quad (10)$$

$$f = \sigma(W_f n_i + e_f) \quad (11)$$

式中, x_i 为 i 时刻的输入序列; n_{i-1} 为 $i-1$ 时刻的输出; $W_{xh}, W_{xs}, W_{xo}, W_{nh}, W_{ns}, W_{no}, W_{nc}, W_{xc}, W_f$ 为相应的权重矩阵; e_h, e_s, e_o 为遗忘门、输入门和输出门的误差; e_f 为预测偏差; h_i, s_i, o_i 为遗忘门、输入门和输出门的状态; σ 为 sigmoid 激活函数; C'_i 为 i 时刻输入的临时状态; C_i 为当前单元格状态; \tanh 为 \tanh 激活函数; n_i 为当前单元格输出; f_i 为 i 时刻的预测值; “*” 为哈达玛积。

门控循环单元 (gated recurrent unit, GRU)

Cho 等^[28] 基于 LSTM 提出了结构更加紧凑简单的 GRU, 在 GRU 中, LSTM 记忆单元的 3 个门被简化为 2 个门 (更新门和重置门), 其表达式^[25]:

$$r_i = \sigma(W_{nr}n_{i-1} + W_{xr}x_i + e_r) \quad (12)$$

$$z_i = \sigma(W_{nz}n_{i-1} + W_{xz}x_i + e_z) \quad (13)$$

$$n'_i = \tanh(W_{xn'}x_i + W_{nn'}n_{i-1} * r_i + e_{n'}) \quad (14)$$

$$n_i = (1 - z_i) * n_{i-1} + z_i * n'_i \quad (15)$$

$$f_i = \sigma(W_f n_i + e_f) \quad (16)$$

式中, “*”为哈达玛积; W_{xr} 、 W_{xz} 、 W_{xm} 为“输入→状态”的权重矩阵; W_{nr} 、 W_{nz} 、 W_{nm} 为“状态→状态”的权重矩阵; W_f 为从隐藏层到输出层的权重矩阵; e_r 、 e_z 、 e_n 、 e_f 为偏移量; σ 为 sigmoid 激活函数。

本研究中 LSTM 与 GRU 模型的隐藏层均为 8, 训练次数均为 100。

1.3 分析软件和模型性能评估

以上模型均使用 Python(3.9.0) 中 statsmodels 库实现。另外采用 PCA 和 Pearson 相关性分析对数据进行前处理, PCA 和相关分析采用 R 语言软件实现。

采用均方根误差 (root mean square error, RMSE)、均方误差 (mean square error, MSE)、平均绝对误差 (mean absolute error, MAE) 和 R 方值 (R^2 -score) 这 4 种指标来评估模型预测性能。

$$\text{RMSE} = \left[1/k \sum_{n=1}^k (a_n - a'_n)^2 \right]^{1/2} \quad (17)$$

$$\text{MSE} = 1/k \sum_{n=1}^k (a_n - a'_n)^2 \quad (18)$$

$$\text{MAE} = 1/k \sum_{n=1}^k |a_n - a'_n| \quad (19)$$

$$R^2\text{-score} = 1 - \sum_{n=1}^k (a_n - a'_n)^2 / (a_n - \bar{a}_n)^2 \quad (20)$$

式中, k 为数量; a_n 为实测值; a'_n 为预测值; \bar{a}_n 为实测平均值。

RMSE 为误差分布的程度, MSE 为真实值与预测值差平方的期望, MAE 为每个预测值与真实值之间绝对误差绝对值的平均值, 以上 3 个指标数值越小, 模型预测效果越好; R^2 -score 为信息与拟合复发线的接近程度的事实比例, 也被称为保证系数或多次复发的各种保证系数, 数值始终介于 0~1, 数值越大, 模型预测效果越好^[29]。

2 结果

2.1 Pearson 相关性分析与 PCA

采用 Pearson 相关性分析, 分析了养殖水

体 TAN 浓度与其他 6 个水质指标的相关性。结果显示, TAN 与 pH 和 SRP 呈极显著正相关 ($P < 0.01$), 与 NO_2^- -N 呈显著正相关 ($P < 0.05$), 与 DO 呈负相关 ($P < 0.05$) (图 1)。因此, 选择 NO_2^- -N、SRP、DO 与 pH 4 个水质指标作为氨氮预测模型的变量。

PCA 是一种用于减少多元数据集维度的统计工具, 在降维的同时保持输入数据的最大信息值不变^[30]。本实验对除 TAN 外其他 6 个水质指标 (NO_2^- -N、 NO_3^- -N、SRP、DO、pH 和 T) 进行 PCA 分析, 结果显示, 前 4 个主成分为有效主成分, 总方差解释率为 81.51% (表 1)。所有样本在 4 个主成分上的得分如表 2 所示, 选用前 4 个主成分得分作为变量对 TAN 进行预测。

2.2 氨氮预测模型

本实验采用 BP、RF、LSTM 和 GRU 共 4 种分析方法, 基于人工投喂、自动化投喂阶段及整个实验周期的原始数据集 (OD) (包括 7 个变量: TAN、 NO_2^- -N、 NO_3^- -N、SRP、DO、pH 和 T)、Pearson 数据集 (Pearson D) (包括 5 个变量: TAN、 NO_2^- -N、SRP、DO 和 pH) 及 PCA 数据集 (PCAD) (包括 5 个变量: TAN、PC1、PC2、PC3 和 PC4) 对养殖水体 TAN 进行预测。

人工投喂阶段氨氮预测模型性能评估 在对 OD 预测结果中, RF 模型的预测效果最差, R^2 -score 仅为 0.119 6, Pearson D-RF 模型 R^2 -score 仅达到了 0.363 5, PCAD-RF 模型 R^2 -score 为负值 (-0.011 2) (图 2, 表 3)。

BP 模型对 OD 预测效果较好 (RMSE = 0.134 5, MSE=0.018 1, MAE=0.111 7, R^2 -score=0.697 2), Pearson D-BP 的预测精度和拟合度最高 (RMSE=0.117 0, MSE=0.013 7, MAE=0.092 9, R^2 -score=0.770 8)。但 PCAD-BP 模型的预测效果很差, 其 R^2 -score 为负值 (-0.125 7)。

LSTM 与 GRU 模型的预测效果相差不大, 其 OD 和 Pearson D 的 R^2 -score 均达到了 0.7 以上, 其中, Pearson D-LSTM (RMSE=0.132 2, MSE=0.017 5, MAE=0.098 3, R^2 -score=0.726 8) 与 Pearson D-GRU 模型 (RMSE=0.130 2, MSE=0.016 9, MAE=0.094 0, R^2 -score=0.735 0) 的预测精度与拟合度最高。PCAD-LSTM 模型 (RMSE=0.159 7, MSE=0.025 5, MAE=0.108 9,

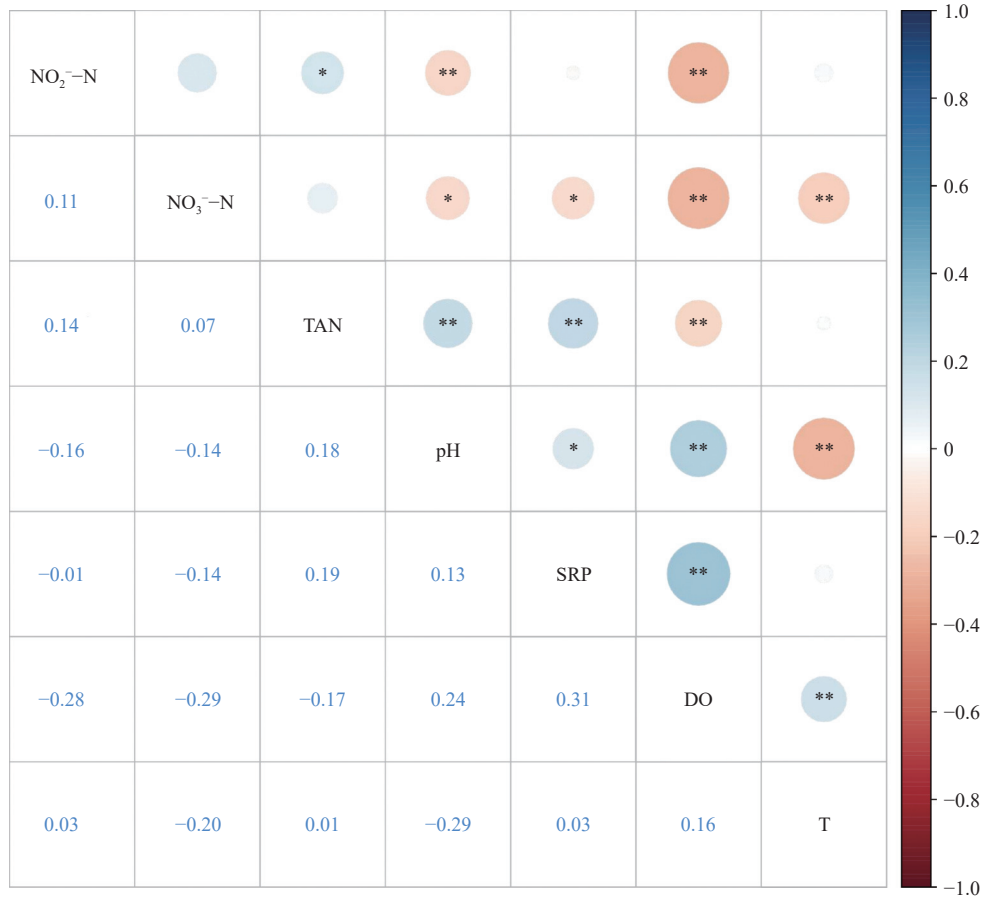


图 1 7 个水质变量的 Pearson 相关系数

“*” 存在显著差异, $P < 0.05$; “***” 存在极显著差异, $P < 0.01$ 。圆圈大小代表所对应的相关系数绝对值大小, 绝对值越大, 圆圈越大, 相关性越显著。

Fig. 1 Pearson correlation coefficient of 7 water quality parameters

“*” significant difference at $P < 0.05$, “***” extremely significant difference at $P < 0.01$. The circle size represents the absolute value of the corresponding correlation coefficient, the larger the absolute value, the larger the circle and the more significant the correlation.

表 1 前 4 个主成分的总方差解释

Tab. 1 Total variance explanation of the first four principal components

主成分 principal components	初始特征值 initial eigenvalue	方差贡献率/% percentage of variance explained	累计方差贡献率/% percentage of total variance explained
1	1.769	29.475	29.475
2	1.313	21.888	51.363
3	0.979	16.319	67.683
4	0.829	13.822	81.505

表 2 所有样本在前 4 个主成分得分

Tab. 2 Score of all the sampling points on the first four principal components

样本序号 sample serial no.	主成分得分 score of the first four principal components			
	PC1	PC2	PC3	PC4
1	-0.045 5	-0.330 8	0.415 3	-0.192 5
2	-0.144 5	-0.296 4	0.481 9	-0.155 1
3	-0.194 0	-0.287 6	0.406 3	-0.014 4
4	-0.063 0	-0.394 1	0.186 5	-0.226 8
5	0.092 1	-0.354 5	0.403 9	-0.025 4
.....
284	0.652 1	-0.126 5	-0.064 8	-0.476 9
285	0.843 4	-0.134 7	0.155 7	-0.508 1
286	0.730 4	-0.125 8	-0.115 6	-0.494 5
287	0.773 3	-0.032 9	0.797 5	-0.560 9
288	0.741 5	-0.080 6	0.179 5	-0.582 9

R^2 -score=0.601 3) 与 PCAD-GRU 模型 (RMSE=0.161 9, MSE=0.026 2, MAE=0.110 0, R^2 -score=0.589 9) 预测精度与拟合度相对较低。

由上可知, 在人工投喂阶段, Pearson D-BP 模型对 TAN 浓度的预测效果最好。其次为

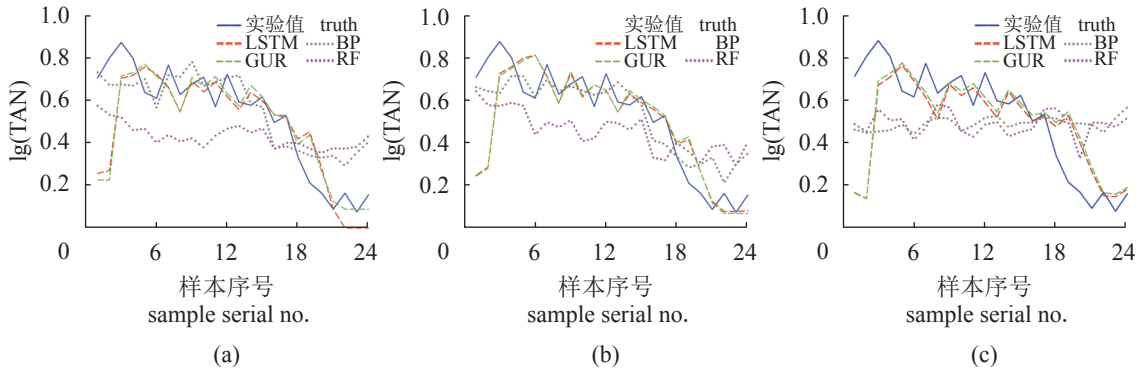


图2 人工投喂阶段结合4种分析方法的总氮预测模型预测值与真实值对比

(a) 原始数据集 (OD), (b) Pearson 数据集 (Pearson D), (c) PCA 数据集 (PCAD); 下同。

Fig. 2 Comparison between predicted and true values of TAN prediction models built by four methods under artificial feeding strategy

(a) original data set (OD), (b) Pearson data set (Pearson D), (c) PCA data set (PCAD); the same below.

表3 人工投喂阶段结合4种分析方法的总氮预测模型评估指标

Tab. 3 Evaluation indexes of TAN prediction models built by four methods under artificial feeding strategy

数据集 data sets	评估指标 evaluation indexes	RF	BP	LSTM	GRU
OD	RMSE	0.229 3	0.134 5	0.135 8	0.137 9
	MSE	0.052 6	0.018 1	0.018 4	0.019 0
	MAE	0.213 9	0.111 7	0.093 2	0.105 2
	R ² -score	0.119 6	0.697 2	0.711 8	0.702 7
Pearson D	RMSE	0.195 0	0.117 0	0.132 2	0.130 2
	MSE	0.038 0	0.013 7	0.017 5	0.016 9
	MAE	0.181 0	0.092 9	0.098 3	0.094 0
	R ² -score	0.363 5	0.770 8	0.726 8	0.735 0
PCAD	RMSE	0.245 7	0.259 3	0.159 7	0.161 9
	MSE	0.060 4	0.067 2	0.025 5	0.026 2
	MAE	0.219 8	0.226 9	0.108 9	0.110 0
	R ² -score	-0.011 2	-0.125 7	0.601 3	0.589 9

Pearson D-LSTM 模型与 Pearson D-GRU 模型。

自动化投喂阶段氮氮预测模型评估性能 RF 模型与 BP 模型对 TAN 的预测效果较差, 其 R²-score 均为负值。GRU 模型对 TAN 的预测效果最好, 其对 OD 预测的精度与拟合度较高 (RMSE=0.141 4, MSE=0.020 0, MAE=0.102 6, R²-score=0.608 3), Pearson D-GRU 模型 (RMSE=0.133 2, MSE=0.0177, MAE=0.0970, R²-score=0.6524) 与 PCAD-GRU 模型 (RMSE=0.133 3, MSE=0.017 8, MAE=0.097 1, R²-score=0.651 8) 对 TAN 的预测

精度和拟合度最高 (图3, 表4)。

LSTM 模型对 TAN 的预测效果不如 GRU 模型。Pearson D-LSTM 模型 (RMSE=0.139 4, MSE=0.019 4, MAE=0.101 9, R²-score=0.619 3) 的预测精度较好, 其 R²-score 超过了 0.6。而 LSTM 模型在 OD 和 PCAD 上的 R²-score 相对较低, 均为 0.59 左右。

由上可知, 在自动化投喂阶段, Pearson D-GRU 模型预测效果最好, 其次为 PCAD-GRU 模型和 Pearson D-LSTM 模型。

整个实验周期氮氮预测模型评估性能

RF 模型与 BP 模型对 TAN 的预测效果很差, 其 R²-score 均为负值 (图4, 表5)。LSTM 模型与 GRU 模型相比, 其对 TAN 的预测效果最好。在对 OD 的预测中, LSTM 模型 (RMSE=0.138 7, MSE=0.019 2, MAE=0.103 8, R²-score=0.606 6) 的预测精度与拟合度高于 GRU 模型 (RMSE=0.151 2, MSE=0.022 9, MAE=0.116 5, R²-score=0.532 2)。PCAD-LSTM 模型 (RMSE=0.133 6, MSE=0.017 9, MAE=0.105 6, R²-score=0.634 6) 的预测精度与拟合度最好, 其次为 Pearson D-LSTM 模型 (RMSE=0.134 5, MSE=0.018 1, MAE=0.104 0, R²-score=0.629 6) 和 Pearson D-GRU 模型 (RMSE=0.135 7, MSE=0.0184, MAE=0.105 1, R²-score=0.623 2)。

综上, 对整个实验周期的 TAN 浓度预测结果显示, PCAD-LSTM 模型预测效果最好, 其次为 Pearson D-LSTM 模型和 Pearson D-GRU 模型。

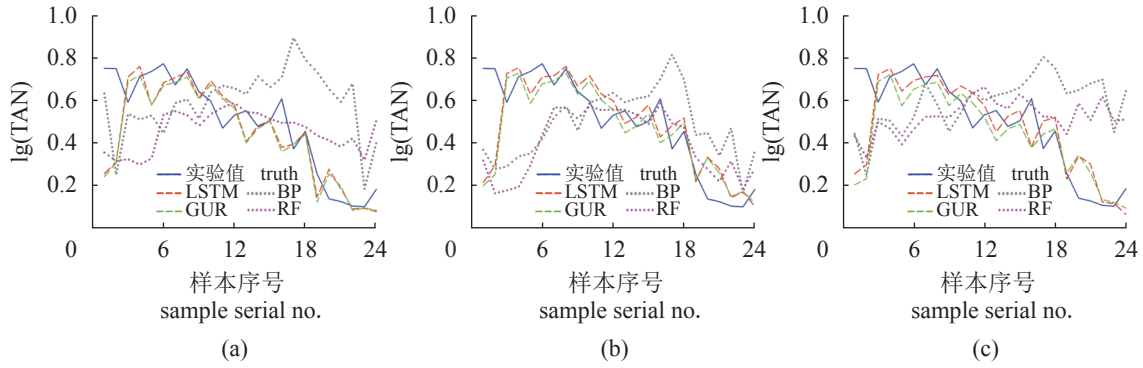


图3 自动化投喂阶段结合4种分析方法的总氮氮预测模型预测值与真实值对比

Fig. 3 Comparison between predicted and true values of TAN prediction models built by four methods under automatic feeding strategy

表4 自动化投喂阶段结合四种分析方法的总氮氮预测模型评估指标

Tab. 4 Evaluation indexes of TAN prediction models built by four methods under automatic feeding strategy

数据集 data sets	评估指标 evaluation indexes	RF	BP	LSTM	GRU
OD	RMSE	0.238 1	0.297 0	0.143 4	0.141 4
	MSE	0.056 7	0.088 2	0.020 6	0.020 0
	MAE	0.198 3	0.244 5	0.101 9	0.102 6
	R ² -score	-0.134 6	-0.764 6	0.596 9	0.608 3
Pearson D	RMSE	0.245 2	0.259 5	0.139 4	0.133 2
	MSE	0.060 1	0.067 4	0.019 4	0.017 7
	MAE	0.174 7	0.224 5	0.101 9	0.097 0
	R ² -score	-0.202 7	-0.347 8	0.619 3	0.652 4
PCAD	RMSE	0.261 5	0.314 3	0.143 6	0.133 3
	MSE	0.068 4	0.098 8	0.020 6	0.017 8
	MAE	0.219 5	0.265 0	0.098 8	0.097 1
	R ² -score	-0.368 3	-0.976 9	0.595 8	0.651 8

3 讨论

随机森林 (RF) 可以使用决策树的组合建立稳定的回归预测模型, 能够对非线性和非正态数据进行处理, 且较小的数据量也能得到良好的训练, 目前该方法已被广泛应用于多种区域水环境的水质预测^[31]。但在本实验中, RF 模型对氨氮的预测效果较差, 并出现过拟合现象。过拟合现象是指在训练集上评估时模型的准确度很高, 但在单独的测试集上评估时准确度低^[32]。过拟合现象产生可能是因为随机噪声和偶尔的异常值被错误地视为结构模型的一部分, 因此, 需要加强对数据噪声和异常值的处理^[33]。神经网络被证实是一种有效的水质建模方法^[34-36]。Ma 等^[37]建立 BP 模型, 并通过温度、pH、TAN、亚硝氮、硝氮、磷酸盐、叶绿素-a、化学需氧量 (COD) 和 5 天生化需氧量 (BOD₅) 共 9 种不同参数预测集约化养虾池中的水质变

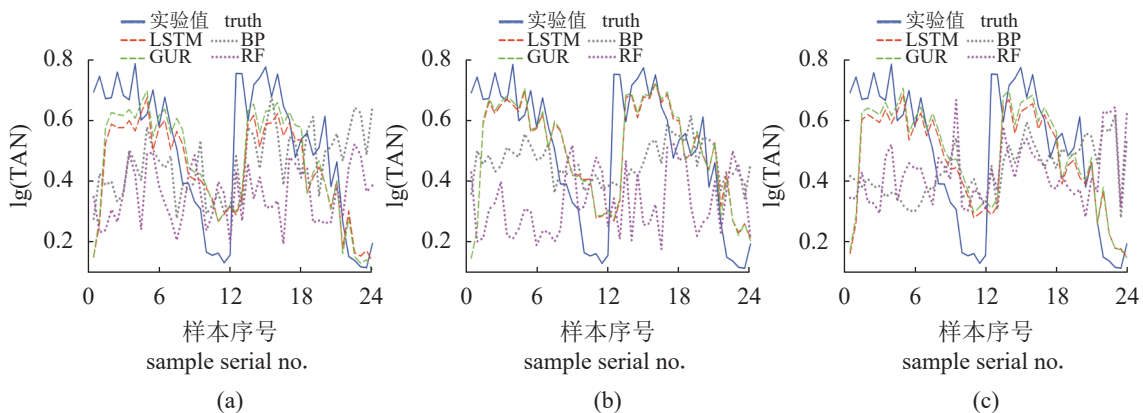


图4 整个实验期间结合4种分析方法的总氮氮预测模型预测值与真实值对比

Fig. 4 Comparison between predicted and true values of TAN prediction models built by four methods under the whole experimental period

表 5 整个实验期间结合 4 种分析方法的
总氮氮预测模型评估指标

Tab. 5 Evaluation indexes of TAN prediction models
built by four methods under
the whole experimental period

数据集 data sets	评估指标 evaluation indexes	RF	BP	LSTM	GRU
OD	RMSE	0.287 1	0.245 7	0.138 7	0.151 2
	MSE	0.082 4	0.060 4	0.019 2	0.022 9
	MAE	0.251 4	0.205 6	0.103 8	0.116 5
	R^2 -score	-0.717 1	-0.257 5	0.606 6	0.532 2
Pearson D	RMSE	0.323 3	0.208 5	0.134 5	0.135 7
	MSE	0.104 5	0.043 5	0.018 1	0.018 4
	MAE	0.281 9	0.183 3	0.104 0	0.105 1
	R^2 -score	-1.176 5	0.095 0	0.629 6	0.623 2
PCAD	RMSE	0.259 5	0.255 0	0.133 6	0.138 0
	MSE	0.067 4	0.065 0	0.017 9	0.019 1
	MAE	0.228 7	0.223 3	0.105 6	0.109 4
	R^2 -score	-0.402 8	-0.354 3	0.634 6	0.610 1

化。但在本实验中, BP 模型只对人工投喂阶段的 TAN 原始数据集 (OD) 和 Pearson 数据集 (Pearson D) 预测结果较好, 在对其他阶段 TAN 浓度的预测中也出现了过拟合现象。

RNN 模型在水质评价方面应用也较为广泛。Zhang 等^[38]在对污水溢流的水位预测结果中发现, LSTM 和 GRU 在多步时间序列预测中表现出很强的能力, 且 GRU 的学习曲线更快。Jiang 等^[39]在对下水道关键水质指标 BOD₅、COD、TAN、总氮 (TN) 和总磷 (TP) 进行预测时发现, LSTM 和 GRU 具有良好的预测性能, 其中 GRU 整体性能更高。在水产养殖水质评估方面, 近年来 RNN 模型也有较多应用。Dabrowski 等^[40]比较了 LSTM、ANN 和线性动力系统 (LDS) 对养殖池塘溶解氧的预测效果, 结果表明 LSTM 模型优于其他模型。Li 等^[41]使用 RNN、LSTM 和 GRU 预测养殖池塘溶解氧, 发现 GRU 与 LSTM 的预测性能相似, 但用于 GRU 的时间成本和参数数量远低于 LSTM。本实验结果与以上研究相似, 本实验预测结果显示, 除了整个实验周期中 OD-GRU 模型的 R^2 -score 为 0.5322 以外, 其他阶段该模型 R^2 -score 均保持在约 0.6, 在人工投喂阶段达到了 0.7 以上。结果表明, GRU 与 LSTM 模型对 TAN 的预测性能较为稳定。

水质模型中, 输入变量会影响模型预测的输出和精度, 若与氨氮关联性较强的变量未被输入会使预测结果产生较大误差, 而与氨氮无显著相关的变量输入会降低模型的计算效率^[16]。有研究在使用遗传算法神经网络 (GA-LM) 模型对水产养殖水体中氨氮浓度进行预测时, 结合 PCA 技术确定了与氨氮具有高关联性的环境因子, 发现 GA-LM-PCA 模型的预测精度显著高于 GA-LM 模型^[42]。本实验中, 与 OD 相比, 部分模型在对 PCA 数据集 (PCAD) 预测时, 模型精度有所提高, 例如自动化投喂阶段 GRU 模型 RMSE 降低了 0.008 1, MSE 降低了 0.002 2, MAE 降低了 0.005 5, R^2 -score 升高了 0.043 5。有研究在使用改进极限学习机 (SPLS-ELM) 预测养殖水体 DO 时, 采用 Pearson 相关性分析提取了与溶解氧关联性强的因子, 使模型获得了较高的预测精度^[43]。本实验中, 相对于 OD 和 PCAD 的预测, Pearson 相关性处理后的数据集提高了氨氮预测模型的预测精度和拟合度, 例如在对整个实验周期的 TAN 预测中, Pearson-GRU 模型 RMSE 降低了 0.015 5, MSE 降低了 0.004 4, MAE 降低了 0.011 4, R^2 -score 升高了 0.091。

在水产养殖中, 合理的投喂策略能够提高养殖生物的饲料利用率和饲料转化率, 减少残饵量, 而残饵分解及养殖生物的废物排泄是养殖水体中氨氮的主要产生途径, 因此, 投喂策略会影响养殖水体中氨氮浓度变化, 从而可能会影响氨氮水质模型的预测精度^[44]。本实验结果显示, 与自动化投喂相比, 人工投喂阶段模型的预测性能较好。例如对 OD 的预测结果显示, 人工投喂相比自动化投喂, BP 模型的 RMSE 降低了 0.162 5, MSE 降低了 0.070 1, MAE 降低了 0.132 7, R^2 -score 升高了 1.461 8; 且在人工投喂阶段, BP、LSTM 和 GRU 三种模型对 TAN 的预测结果 R^2 -score 均达到了 0.7 以上。因此本研究初步表明, 投喂策略会对该系统氨氮预测模型的预测性能产生影响。目前, 关于投喂策略对氨氮预测模型的影响还未见报道。但有研究发现, 投喂策略会影响养殖生物的摄食、消化、排泄等代谢过程, 从而对养殖水质产生影响。Godoy-Olmos 等^[17]研究发现, 养殖生物的耗氧量和氨排泄量受摄食和投喂策略的影响, 人工投喂时, 鱼类的排氨率更高,

同时发现, 投喂策略可影响 RAS 水处理单元的氨氮去除率, 从而对水体中 TAN 浓度产生影响。此外, 不同投喂策略下养殖水体中 TAN 浓度的日变化规律不同, 人工投喂阶段 TAN 的日变化规律性更明显, 而自动化投喂阶段 TAN 浓度日间波动较小^[45]。本研究中, 人工投喂阶段模型精度更高, 可能是由于人工投喂采用多量少次投喂的方式, 使得每次投喂后残饵较多, 水中 TAN 浓度日变化波动较大, 数据周期性变化较为明显, 这种周期性的变化可能会影响到模型的精度。

RMSE、MSE、MAE 和 R^2 -score 四个指标常用于模型性能评价。有学者使用 GA-LM 模型预测养殖水体氨氮, 模型 RMSE 值为 0.186 4, MAE 为 0.121 8, 后通过优化算法使 RMSE 降低至 0.006 5, MAE 值降低至 0.004 7^[42]。汪翔等^[46]分别使用 BP 模型和简单递归神经网络 (Elman) 预测养殖水体氨氮, 得到的 R^2 值分别为 0.66 和 0.74。本研究中人工投喂阶段 Pearson-BP 模型预测 RMSE 为 0.117 0, MSE 为 0.013 7, MAE 为 0.092 9, R^2 -score 为 0.770 8, 且人工投喂阶段 GRU 和 LSTM 模型预测结果 R^2 -score 超过 0.7, 因此模型可用。但本实验氨氮数据样本量较小、所采集的环境因子较少, 模型预测精度还有待提高。后续研究可结合多种优化算法改进模型, 并增加氨氮数据样本量、采集环境因子数量, 以提高模型预测精度。

4 结论

投喂策略会影响模型的预测精度, 并且不同投喂阶段筛选出的最优预测模型不同。在人工投喂阶段, Pearson D-BP 模型对总氨氮 (TAN) 浓度的预测性能最好; 在自动化投喂阶段, Pearson D-GRU 模型预测性能最好; 在整个实验周期中, PCAD-LSTM 模型预测性能最好; 四种模型中, GRU 和 LSTM 模型对 TAN 的预测较为稳定。系统中水质因子数据经 Pearson 分析后可有效提高 TAN 预测模型的预测性能。

参考文献 (References):

[1] Martins C I M, Eding E H, Verdegem M C J, *et al.* New developments in recirculating aquaculture systems in Europe: a perspective on environmental sustainability[J]. *Aquacultural*

Engineering, 2010, 43(3): 83-93,

[2] Chandra J, Samali A, Orrenius S. Triggering and modulation of apoptosis by oxidative stress[J]. *Free Radical Biology and Medicine*, 2000, 29(3-4): 323-333,

[3] 张武肖, 孙盛明, 戈贤平, 等. 急性氨氮胁迫及毒后恢复对团头鲂幼鱼鳃、肝和肾组织结构的影响 [J]. *水产学报*, 2015, 39(2): 233-244.

Zhang W X, Sun S M, Ge X P, *et al.* Acute effects of ammonia exposure on histopathology of gill, liver and kidney in juvenile *Megalobrama amblycephala* and the post-exposure recovery[J]. *Journal of Fisheries of China*, 2015, 39(2): 233-244 (in Chinese).

[4] Zhang W X, Xia S L, Zhu J, *et al.* Growth performance, physiological response and histology changes of juvenile blunt snout bream, *Megalobrama amblycephala* exposed to chronic ammonia[J]. *Aquaculture*, 2019, 506: 424-436,

[5] Zuo J, Dai S H, Wei B M, *et al.* Comparison of three methods for determination of ammonia nitrogen in surface water[J]. *Meteorological and Environmental Research*, 2019, 10(3): 90-91,97,

[6] Lin K N, Zhu Y, Zhang Y B, *et al.* Determination of ammonia nitrogen in natural waters: recent advances and applications[J]. *Trends in Environmental Analytical Chemistry*, 2019, 24: e00073,

[7] Liu S Y, Xu L Q, Jiang Y, *et al.* A hybrid WA-CPSO-LSSVR model for dissolved oxygen content prediction in crab culture[J]. *Engineering Applications of Artificial Intelligence*, 2014, 29: 114-124,

[8] Yu H H, Chen Y Y, Hassan S G, *et al.* Dissolved oxygen content prediction in crab culture using a hybrid intelligent method[J]. *Scientific Reports*, 2016, 6(1): 27292,

[9] Li K, Wang W, Lin S H. Soft measurement of ammonia nitrogen concentration based on GA-SCN[C]//IEEE. 2018 IEEE Symposium on Product Compliance Engineering-Asia (ISPCEN). Shenzhen: IEEE, 2018: 1-4.

[10] Pedersen L F, Suhr K I, Dalsgaard J, *et al.* Effects of feed loading on nitrogen balances and fish performance in replicated recirculating aquaculture systems[J]. *Aquaculture*, 2012, 338-341: 237-245,

[11] Ren W J, Li L, Dong S L, *et al.* Effects of C/N ratio and light on ammonia nitrogen uptake in *Litopenaeus vannamei* culture tanks[J]. *Aquaculture*, 2019, 498: 123-131,

[12] Masoud A A. Groundwater quality assessment of the shallow

- aquifers west of the Nile Delta (Egypt) using multivariate statistical and geostatistical techniques[J]. *Journal of African Earth Sciences*, 2014, 95: 123-137,
- [13] Bu H M, Song X F, Zhang Y. Using multivariate statistical analyses to identify and evaluate the main sources of contamination in a polluted river near to the Liaodong Bay in Northeast China[J]. *Environmental Pollution*, 2019, 245: 1058-1070,
- [14] Barakat A, El Baghdadi M, Rais J, *et al.* Assessment of spatial and seasonal water quality variation of Oum Er Rbia River (Morocco) using multivariate statistical techniques[J]. *International Soil and Water Conservation Research*, 2016, 4(4): 284-292,
- [15] Silva L C B, Lopes B, Pontes M J, *et al.* Fast decision-making tool for monitoring recirculation aquaculture systems based on a multivariate statistical analysis[J]. *Aquaculture*, 2021, 530: 735931,
- [16] 陈英义, 成艳君, 杨玲, 等. 基于改进深度信念网络的池塘养殖水体氨氮预测模型研究 [J]. *农业工程学报*, 2019, 35(7): 195-202.
- Chen Y Y, Cheng Y J, Yang L, *et al.* Prediction model of ammonia-nitrogen in pond aquaculture water based on improved multi-variable deep belief network[J]. *Transactions of the Chinese Society of Agricultural Engineering*, 2019, 35(7): 195-202 (in Chinese).
- [17] Godoy-Olmos S, Jauralde I, Monge-Ortiz R, *et al.* Influence of diet and feeding strategy on the performance of nitrifying trickling filter, oxygen consumption and ammonia excretion of gilt-head sea bream (*Sparus aurata*) raised in recirculating aquaculture systems[J]. *Aquaculture International*, 2022, 30(2): 581-606,
- [18] Lippmann R. Book review: "neural networks, a comprehensive foundation", by Simon Haykin[J]. *International Journal of Neural Systems*, 1994, 5(4): 363-364.
- [19] Yang T T, Asanjan A A, Faridzad M, *et al.* An enhanced artificial neural network with a shuffled complex evolutionary global optimization with principal component analysis[J]. *Information Sciences*, 2017, 418-419: 302-316,
- [20] Cui Y Q, Liu H F, Wang Q L, *et al.* Investigation on the ignition delay prediction model of multi-component surrogates based on back propagation (BP) neural network[J]. *Combustion and Flame*, 2022, 237: 111852,
- [21] Guo Y P. Credit risk assessment of P2P lending platform towards big data based on BP neural network[J]. *Journal of Visual Communication and Image Representation*, 2020, 71: 102730,
- [22] Zhang D, Lin J Q, Peng Q D, *et al.* Modeling and simulating of reservoir operation using the artificial neural network, support vector regression, deep learning algorithm[J]. *Journal of Hydrology*, 2018, 565: 720-736,
- [23] Breiman L. Random forests[J]. *Machine Learning*, 2001, 45(1): 5-32,
- [24] 刘银超. 基于 BP 神经网络和随机森林的空气污染物浓度预测研究 [D]. 秦皇岛: 燕山大学, 2017.
- Liu Y C. Air quality forecasting using BP neural network and random forest model[D]. Qinhuangdao: Yanshan University, 2017 (in Chinese).
- [25] Liu M D, Ding L, Bai Y L. Application of hybrid model based on empirical mode decomposition, novel recurrent neural networks and the ARIMA to wind speed prediction[J]. *Energy Conversion and Management*, 2021, 233: 113917,
- [26] Hochreiter S, Schmidhuber J. Long short-term memory[J]. *Neural Computation*, 1997, 9(8): 1735-1780,
- [27] Shahid F, Zameer A, Muneeb M. Predictions for COVID-19 with deep learning models of LSTM, GRU and Bi-LSTM[J]. *Chaos, Solitons & Fractals*, 2020, 140: 110212
- [28] Cho K, Van Merriënboer B, Gulcehre C, *et al.* Learning phrase representations using RNN encoder-decoder for statistical machine translation[C]//Association for computational Linguistics.proceedings of the 2014 conference on Empirical Methods in natural language processing. Doha: Association for Computational Linguistics, 2014: 1724-1734.
- [29] Khan M A, Khan R, Algarni F, *et al.* Performance evaluation of regression models for COVID-19: a statistical and predictive perspective[J]. *Ain Shams Engineering Journal*, 2021, 13(2): 101574,
- [30] Tripathi M, Singal S K. Use of principal component analysis for parameter selection for development of a novel water quality index: a case study of river Ganga India[J]. *Ecological Indicators*, 2019, 96: 430-436,
- [31] Li B, Yang G S, Wan R R, *et al.* Combining multivariate statistical techniques and random forests model to assess and diagnose the trophic status of Poyang Lake in China[J]. *Ecological Indicators*, 2017, 83: 74-83,
- [32] Subramanian J, Simon R. Overfitting in prediction models-is it a problem only in high dimensions?[J]. *Contemporary Clinical Trials*, 2013, 36(2): 636-641,

- [33] Ord K. Data adjustments, overfitting and representativeness[J]. *International Journal of Forecasting*, 2020, 36(1): 195-196.
- [34] Maier H R, Jain A, Dandy G C, *et al.* Methods used for the development of neural networks for the prediction of water resource variables in river systems: current status and future directions[J]. *Environmental Modelling & Software*, 2010, 25(8): 891-909.
- [35] Yue X Y, Guo Y J, Wang J R, *et al.* Water pollution forecasting model of the back-propagation neural network based on one step secant algorithm[C]//IEEE. International conference on information computing and applications. Tangshan: Springer, 2010: 458-464.
- [36] Melesse A M, Ahmad S, McClain M E, *et al.* Suspended sediment load prediction of river systems: an artificial neural network approach[J]. *Agricultural Water Management*, 2011, 98(5): 855-866.
- [37] Ma Z, Song X F, Wan R, *et al.* Artificial neural network modeling of the water quality in intensive *Litopenaeus vannamei* shrimp tanks[J]. *Aquaculture*, 2014, 433: 307-312.
- [38] Zhang D, Lindholm G, Ratnaweera H. Use long short-term memory to enhance Internet of Things for combined sewer overflow monitoring[J]. *Journal of Hydrology*, 2018, 556: 409-418.
- [39] Jiang Y Q, Li C L, Sun L, *et al.* A deep learning algorithm for multi-source data fusion to predict water quality of urban sewer networks[J]. *Journal of Cleaner Production*, 2021, 318: 128533.
- [40] Dabrowski J J, Rahman A, George A. Prediction of dissolved oxygen from pH and water temperature in aquaculture prawn ponds[C]//Association for Computing Machinery. Proceedings of the Australasian Joint Conference on Artificial Intelligence-Workshops. Wellington: Association for Computing Machinery, 2018: 2-6.
- [41] Li W Y, Wu H, Zhu N Y, *et al.* Prediction of dissolved oxygen in a fishery pond based on gated recurrent unit (GRU)[J]. *Information Processing in Agriculture*, 2021, 8(1): 185-193.
- [42] 姚启, 缪新颖. 基于主成分分析及 GA-LM 的水产养殖环境溶解氧和氨氮含量预测 [J]. *大连海洋大学学报*, 2021, 36(5): 851-858.
- Yao Q, Miao X Y. Prediction of dissolved oxygen and ammonia nitrogen concentrations in aquaculture environment based on PCA and GA-LM[J]. *Journal of Dalian Fisheries University*, 2021, 36(5): 851-858 (in Chinese).
- [43] 施珮, 匡亮, 袁永明, 等. 基于改进极限学习机的水体溶解氧预测方法 [J]. *农业工程学报*, 2020, 36(19): 225-232.
- Shi P, Kuang L, Yuan Y M, *et al.* Dissolved oxygen prediction for water quality of aquaculture using improved ELM network[J]. *Transactions of the Chinese Society of Agricultural Engineering*, 2020, 36(19): 225-232 (in Chinese).
- [44] Gichana Z M, Liti D, Waidbacher H, *et al.* Waste management in recirculating aquaculture system through bacteria dissimilation and plant assimilation[J]. *Aquaculture International*, 2018, 26(6): 1541-1572.
- [45] 孙雪倩, 李丽, 董双林, 等. 基于不同投喂模式的循环水养殖系统中三级生物滤池运行效率研究 [J]. *中国海洋大学学报*, 2022, 52(5): 32-43.
- Sun X Q, Li L, Dong S L, *et al.* Research on operation efficiency of three-stage biofilter in a recirculating aquaculture system under different feeding modes[J]. *Periodical of Ocean University of China*, 2022, 52(5): 32-43 (in Chinese).
- [46] 汪翔, 何吉祥, 余磊, 等. Elman 网络在养殖水体氨氮预测中的应用研究 [J]. *安徽农业科学*, 2015, 43(31): 365-367.
- Wang X, He J X, She L, *et al.* Application of Elman neural network in aquaculture water NH₃-N prediction[J]. *Journal of Anhui Agricultural Sciences*, 2015, 43(31): 365-367 (in Chinese).

Ammonia nitrogen prediction model for recirculating aquaculture system based on different feeding strategies

SUN Xueqian¹, LI Li^{1,2*}, DONG Shuanglin^{1,2}, TIAN Xiangli^{1,2}, ZHANG Shengkun^{1,2}

1. Key Laboratory of Mariculture, Ministry of Education, Ocean University of China, Qingdao 266100, China;

2. Function Laboratory for Marine Fisheries Science and Food Production Processes, Qingdao National Laboratory for Marine Science and Technology, Qingdao 266235, China

Abstract: The prediction and warning of total ammonia nitrogen (TAN) in aquaculture are crucial. Current optimizations of TAN prediction models primarily rely on improved algorithms. However, various management strategies, such as feeding strategies, may be implemented during the aquaculture production process, potentially affecting the prediction performance of these models. This study aims to model and predict TAN concentrations in recirculating aquaculture system (RAS) and investigate the impact of different feeding strategies on the prediction performance of TAN models. We measured 7 water quality parameters in the tanks of *Oplegnathus punctatus* within an RAS. The data were pre-processed using principal component analysis (PCA) or Pearson correlation analysis, resulting in three datasets: the original dataset (OD), the Pearson dataset (PearsonD) and the PCA dataset (PCAD). These datasets were then integrated with random forest (RF), back propagation neural network (BP), gated circulation unit (GRU), and long short-term memory (LSTM) to forecast TAN concentrations under two distinct feeding strategies. The performance of the models was evaluated using root mean square error (RMSE), mean square error (MSE), mean absolute error (MAE), and R-square value (R^2 -score). The RF model exhibited the poorest prediction performance. The GRU and LSTM models demonstrated acceptable and stable prediction accuracy, while the accuracy of the BP model varied. Optimal prediction models differed under the two feeding strategies: under artificial feeding, Pearson correlation analysis combined with BP yielded higher accuracy, whereas under automatic feeding, Pearson correlation analysis combined with GRU performed better. Pearson correlation analysis combined with LSTM or GRU, and PCA combined with LSTM, showed superior performance throughout the experimental period. Compared to the automatic feeding strategy, under artificial feeding, the RMSE, MSE and MAE of the models built with Pearson correlation analysis and LSTM decreased by 0.007 2, 0.001 9 and 0.003 6 respectively, while the R^2 -score increased by 0.107 5. Similarly, the RMSE, MSE and MAE of the models built with Pearson correlation analysis and GRU decreased by 0.003 0, 0.000 8 and 0.003 0 respectively, and the R^2 -score increased by 0.082 6. Feeding strategy significantly influences the prediction accuracy of TAN models, and Pearson correlation analysis combined with GRU or LSTM could be employed to predict TAN in RAS effectively. This study provides a reference for the optimization of ammonia nitrogen prediction technology in RAS.

Key words: *Oplegnathus punctatus*; recirculating aquaculture system; feeding strategy; total ammonia nitrogen; water quality prediction model

Corresponding author: LI Li. E-mail: l_li@ouc.edu.cn

Funding projects: National Key Research and Development Program of China (2020YFD0900201, 2017YFE0122100)