

基于机器学习的总溶解气体预测模型构建与评估*

杨玲, 杨慧霞**, 王云云, 姚云波, 王顺

(贵州大学土木工程学院, 贵阳 550025)

摘要: 随着梯级水库的持续开发, 大坝泄流产生的过饱和总溶解气体在河道内难以消散, 这将导致鱼类患气泡病甚至死亡, 因此开发大坝下游总溶解气体预测模型对保护生物多样性具有重要意义。本文收集美国哥伦比亚河上 3 个大坝监测站点的日监测数据(水温、气压、流量、大坝溢流以及总溶解气体饱和度), 在此基础上利用 BP 神经网络、随机森林以及提升树 3 种机器学习算法预测总溶解气体饱和度, 并对 3 种算法的预测性能进行评价和对比。研究发现, 随着具有显著性相关性的输入变量个数增加, 各模型预测性能呈上升趋势, 且不同模型受输入因子的影响不同。在最佳输入变量方案下, 随机森林(平均绝对误差(MAE) = 1.273%, 均方根误差(RMSE) = 1.775%, $R^2 = 0.952$)和提升树(MAE = 1.268%, RMSE = 1.622%, $R^2 = 0.960$)的预测性能最佳。在模型验证阶段, 提升树模型可以将预测值与实测值的平均相对误差控制在 2.4% 以内。本研究所构建的模型能够快速准确地预测大坝泄流期间下游河道内总溶解气体饱和度, 有助于提前评估过饱和风险, 及时调整排放调度方案, 对于局部鱼类保护区提前采取防护措施, 从而减少总溶解气体对水生生态的影响。研究结果可为深入开展基于机器学习的总溶解气体预测模型提供一定的参考价值。

关键词: 总溶解气体过饱和; 机器学习; 预测; 梯级水库

Construction and evaluation of total dissolved gas prediction model based on machine learning*

Yang Ling, Yang Huixia**, Wang Yunyun, Yao Yunbo & Wang Shun

(College of Civil Engineering, Guizhou University, Guiyang 550025, P.R. China)

Abstract: The ongoing advancement of cascade reservoirs has resulted in the formation of supersaturated total dissolved gases, which are challenging to disperse in riverine environments. This phenomenon has the potential to precipitate the onset of gas bubble disease and, in extreme cases, may even result in fish mortality. Therefore, the development of a predictive model for total dissolved gases downstream of dams is important for biodiversity conservation. This paper collected data from three dam monitoring stations on the Columbia River in the United States, comprising measurements of water temperature, barometric pressure, flow, dam overflow, and total dissolved gas saturation. These data were used to train three machine learning algorithms, namely, BP neural networks, random forests, and boosting trees, which were then employed to predict total dissolved gas saturation. The performance of the three algorithms was evaluated and compared. It found that as the number of significantly correlated input variables increases, the predictive performance of each model showed an upward trend, and different models were affected differently by input factors. Random forest (MAE = 1.273%, RMSE = 1.775%, $R^2 = 0.952$) and boosting tree (MAE = 1.268%, RMSE = 1.622%, $R^2 = 0.960$) had the best prediction performance under the optimal input variable scheme. In the model validation phase, boosting tree and random forest showed higher accuracy, with average relative errors of 2.3% and 2.6% between their predicted and measured values. In the model validation phase, the boosting tree model can control the average relative error between the predicted and measured values within 2.4%. The model constructed in this study can rapidly and accurately predict total dissolved gas (TDG) saturation in the downstream channel during dam releases. This enables the risk of TDG saturation to be assessed in advance, thereby facilitating the timely adjustment of the discharge scheduling scheme. Protective measures are taken in advance for localized fish sanctuaries, thus reducing the impact of TDG on aquatic ecology. The results of the study can provide some reference value for in-

* 2024-05-07 收稿; 2024-08-20 收修改稿。

贵州省科技计划项目(QKHJ-[2019]1117)资助。

** 通信作者; E-mail: bt-yanghuixia@163.com。

depth machine learning-based total dissolved gas prediction modeling.

Keywords: Total dissolved gas supersaturation; machine learning; prediction; cascade reservoirs

大坝泄流往往会产生总溶解气体(total dissolved gas, TDG)过饱和现象,其产生原理是高坝泄水具有高水头高流速的特点,下泄水流与大量空气接触,在高压、强掺气的环境下,大量气体溶于水产生总溶解气体过饱和现象^[1]。所谓总溶解气体是指氧气、氮气、二氧化碳、水蒸气等,不同气体在对应的温度及压强下有着不同的溶解度,亦有着对应的平衡状态,但在某些特殊情况下,水中各种气体的浓度高于平衡状态时的浓度,此时的水体被称为总溶解气体过饱和水体^[2]。早在 20 世纪 60 年代,美国科学家就在哥伦比亚河首次发现由于大坝泄流产生的 TDG 过饱和水体导致大马哈鱼、虹鳟鱼死亡的现象^[3-4]。我国葛洲坝电站、新安江水库和三峡大坝运行初期也曾观测到由于水中 TDG 过饱和而导致鱼类患气泡病甚至死亡的现象^[5-6]。

已有研究发现,下游河道内有些鱼类可以利用补偿深度躲避 TDG 的危害^[7-9],即鱼类通过寻求 TDG 低饱和度的庇护区,以避免 TDG 的威胁。不仅如此,高饱和度与低饱和度 TDG 河流的交汇处往往更适合鱼类躲避 TDG 的危害^[10-11],Shen 等^[12]提出在交汇处通过河床开挖或者增加一些阻力障碍来扩大适合鱼类栖息的低饱和度 TDG 区域。近年来,不少学者提出通过合理的水库调度也能有效降低过饱和 TDG 的生成^[13],优化后的调度措施能够减缓大坝高流量泄水时间,从而在一定程度上减轻 TDG 的累计水平。且随着梯级水库的持续开发,其对 TDG 的累积效应日益突出^[14]。Ma 等^[15]通过数值模拟大坝下游 TDG 的产生及耗散过程,提出一种能够降低 TDG 的梯级水库调度方案。基于此,快速准确地预测大坝泄流期间下游河道内 TDG 饱和度,有助于提前评估过饱和风险,及时调整排放调度方案,对于局部鱼类保护区提前采取防护措施,从而减少 TDG 对水生生态的影响。

目前关于 TDG 预测模型的建立,主要基于气液界面传质理论和流体动力学等通过室内实验以及数值模拟得出^[13,16]。Politano 等^[17]建立了一个三维两相流 TDG 预测模型,通过模型预测提出将大部分泄洪流量集中在一个海湾,促进气泡向自由表面输移,从而降低 TDG 的生成。Ma 等^[18]建立了一维非常定 TDG 模型,预测不同流量模式下高坝下游过饱和 TDG 的运移和耗散,并建议大坝采取不连续排放模式。Feng 等^[19]采用非定常三维两相流计算流体力学模型预测了 TDG 的分布。Wan 等^[20]开发了一个具有洪水脉冲流量模式的动态多目标过饱和 TDG 管理模型。与传统的实验或者数值模拟方法建立的 TDG 预测模型相比,机器学习算法不需要复杂的数学模型,只是通过对众多数据的挖掘,来寻找出数与数之间的映射关系。对于刚修建的大坝可能不太适合用机器学习算法建立 TDG 预测模型,但随着科技的发展,越来越多的大坝已经积累了庞大的数据库,在建立 TDG 预测模型方面,机器学习算法则是高效而省时的一种选择。

随着人工智能的快速发展,机器学习算法已被广泛应用于各种领域,如经济贸易^[21-22]、生物医学^[23-25]、交通运输^[26-27]、环境工程^[28-29]等。Moon 等^[30]选择 pH、水温等 9 种输入变量运用 AdaBoost 算法模型来预测溶解氧浓度。Adnan 等^[31]提出将集中机器学习方法相结合形成一种新的混合方法来预测巴基斯坦北部曼格拉流域的月径流预测,总结出混合模型比独立机器学习模型精度更高。近年来,TDG 监测数据日益丰富,成为机器学习算法用于预测 TDG 的坚实基础。Heddam 等^[32-33]采用克里格插值法(KIM)、响应面法(RSM)以及广义回归神经网络(GRNN)等方法预测大坝下游 TDG 饱和度。不仅如此,支持向量回归(SVR)和极限学习机(ELM)两种机器学习方法也常被用于预测 TDG^[34-35]。由于机器学习算法种类繁多,因此选择一种更高效、更精确的算法来预测 TDG 至关重要。

本文通过美国地质勘探局的网站,选取美国哥伦比亚河上邦纳维尔大坝、约翰戴大坝以及达尔斯大坝监测站点的日监测数据,包括水温(T)、气压(P)、流量(D)、大坝溢流(SFD)以及 TDG 饱和度。利用邦纳维尔大坝的日监测数据,通过 BP 神经网络(BP)、随机森林(RF)以及提升树(BT)3 种机器学习算法建立 TDG 预测模型。为了对比 3 种模型的预测性能,采用平均绝对误差(MAE)、均方根误差(RMSE)、决定系数(R^2)3 种评价指标对模型的性能进行评价。并利用约翰戴大坝以及达尔斯大坝的日监测数据对所建模型进行检验。本文为机器学习算法在开发大坝下游 TDG 预测方向提供了一种新方法,对水电开发河段的水生态和水环境保护具有重要意义。

1 数据和方法

1.1 数据来源

本文所采用的原位观测数据均来源于美国地质勘探局的网站 (<https://waterdata.usgs.gov>)。美国和加拿大境内的哥伦比亚河上梯级水电站开发建设最完善,故选取美国哥伦比亚河上3个连续大坝监测站点的日监测数据,包括水温(T)、气压(P)、流量(D)、大坝溢流(SFD)以及TDG饱和度。其中,邦纳维尔大坝(Bonneville Dam)位于华盛顿和俄勒冈交界的哥伦比亚河上,1938年6月6日投产,装机容量105万kW,大坝为混凝土重力坝,坝高约60m,平均年泄流量5190.47 m³/s;达尔斯大坝(The Dalles Dam)位于邦纳维尔大坝的上游,1957年5月13日投产,装机容量178万kW,大坝为座混凝土重力坝,坝高约61m,平均年泄流量5037.56 m³/s;约翰戴大坝(John Day Dam)位于达尔斯大坝的上游,1968年7月16日投产,装机容量21.6万kW,大坝为座混凝土重力坝,坝高约56m,平均年泄流量4881.82 m³/s。

WRNO 站点位于邦纳维尔大坝下游,获取的是2014—2023年的日监测数据,用于模型构建,去除无效数据后,总计收集3575条有效数据。JHAW 站点和TDDO 站点分别位于约翰戴大坝和达尔斯大坝的下游,收集了2023年一整年的日监测数据,用于对本文所建立模型的检验,去除无效数据,有效数据分别为349条和351条(表1)。对收集数据进行统计分析,绘制小提琴图,如图1所示。小提琴图中的中间原点表示中位数,中间的箱线盒表示下四分位点到上四分位点的数据,从其延伸的黑线代表1.5倍四分位距的数据,外部形状表示核密度估计量。从图中可以看出,对于不同的影响因素,3个站点数据集的小提琴分布类似,且数据不存在离群值或者离群值较少,说明数据集构建合理。

表1 3个监测站点的具体信息

Tab.1 Specific information of the three monitoring sites

位置	站点标号	USGS ID	经度	纬度	收集数据时间段	有效数据
邦纳维尔大坝	WRNO	453630122021400	122°02'14"E	45°36'30"N	2014-1-1-2023-12-31	3575
约翰戴大坝	JHAW	454249120423500	120°42'35"E	45°42'49"N	2023-1-1-2023-12-31	349
达尔斯大坝	TDDO	14105700	121°10'20"E	45°36'27"N	2023-1-1-2023-12-31	351

1.2 模型构建方法

1.2.1 BP神经网络 BP(back propagation)神经网络由输入层、隐含层和输出层3层组成。BP神经网络的过程主要分为两个阶段,第一阶段是信号的正向传播,从输入层经过隐含层,最后到达输出层;第二阶段是误差的反向传播,从输出层到隐含层,最后到输入层,依次调节隐含层到输出层的权重和偏置,输入层到隐含层的权重和偏置。

1.2.2 随机森林 随机森林(random forest)是一种监督学习算法,是在决策树的基础上使用bagging方式集成的一种算法。其构建过程主要是将许多决策树并行起来对数据进行处理,且每棵树都是独立的,互不干扰,由于每棵树使用的样本都是随机的,故每棵树产生的结果也就不同。对于分类问题,按照多棵树分类器投票决定最终分类的结果;对于回归问题,由多棵树预测值的均值决定最终预测结果。

1.2.3 提升树 提升树(boosting tree)也是一种集成算法,由于单一的决策树难以解决某些复杂问题,因此通过集成算法将模型进行组合。提升树就是以决策树为基函数,使用Boosting方式集成的一种算法。与随机森林不同的是,这种方法是将多个决策树串联起来,首先第一棵树对数据进行训练,第二棵树再对第一棵树的不足进行改进,以此类推,最后一棵树则是把前面的树作为一个整体,再进行适当的改进。总的来说,后面加入的树模型只会比前面的树模型越来越强。提升树方法实际采用加法模型,其构建过程在于第一个回归树预测的效果可能一般,但是第二个回归树把第一个预测错的残差作为输入,也就是说,如果第一个点的值被预测错误,那么在下一个回归树里面模型的权值就会变大,相当于吸取前一个模型的经验,从而提高模型的效果。

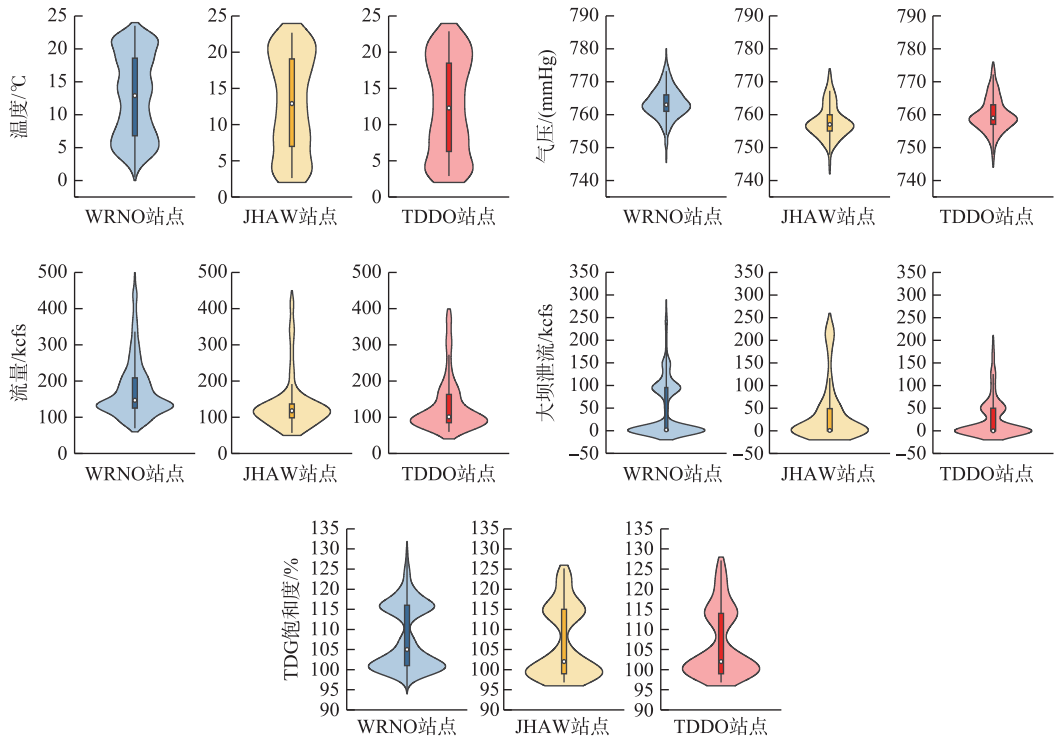


图1 原位观测数据集小提琴图

Fig.1 Violin graph of *in situ* observation data set

2 模型构建

2.1 数据预处理

在模型构建时,若输入的计算因子与 TDG 饱和度相关性太弱,就会导致模型计算量大,降低计算精度,不利于模型的构建。故本文采取 Spearman 法分析各变量与 TDG 饱和度之间的相关性。Spearman 系数为正时表示 2 种因子正相关,Spearman 系数为负时表示 2 种因子负相关,当系数越接近 1 或者-1 时,则表示相关性越强。如图 2 所示,SFD 与 TDG 饱和度的相关性最高($r=0.90$),其次相关性按大小排列依次是流量($r=0.54$)、水温($r=0.46$)、水压($r=-0.39$),最终确定 SFD、D、T、P 4 种输入变量。同时为了研究不同输入变量对 TDG 预测性能的影响,按照相关性大小确定了 4 种输入变量组合方案,如图 2 所示。

由于收集的数据不同指标量纲不同,为避免梯度消失或梯度爆炸问题,把所有数据都进行最大、最小值归一化处理,即全都转换为 0~1 范围内的数据。利用 WRNO 站点十年的监测数据,随机抽取 80% ($n=2860$) 的数据作为模型的训练集,20% ($n=715$) 的数据作为模型的测试集。

2.2 模型评价指标

为了评估 3 种模型的性能,选择平均绝对误差 (MAE)、均方根误差 (RMSE) 和决定系数 (R^2) 3 种性能指标对所建模型进行评价。其中 MAE 表征预测值与真实值之间的偏离程度,可以准确反映出实际预测误差的大小,MAE 越接近 0,表示模型拟合较好。RMSE 即误差平方和平均值的算术平方根,RMSE 越接近 0,表示模型拟合越好。 R^2 表示预测值与实际值的拟合程度, R^2 的范围在 0~1 之间, R^2 接近 1,则表示模型拟合程度好。

2.3 模型对比分析

为了对比不同机器学习算法和不同输入变量对 TDG 预测模型精度的影响,本文利用 BP 神经网络、随

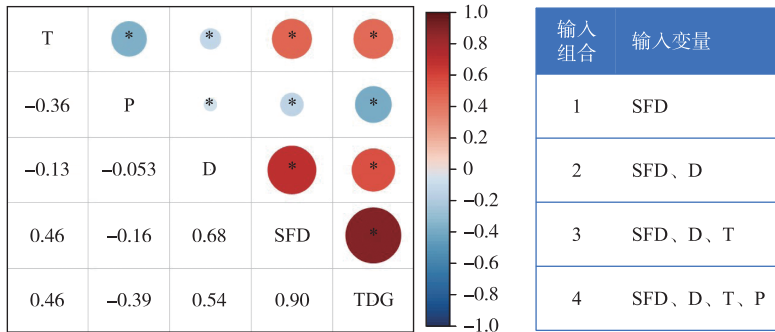


图 2 TDG 与其他参数的 Spearman 相关系数矩阵以及各模型输入变量方案

Fig.2 Spearman's correlation coefficient matrix of TDG with other parametric indicators and schematic diagram of each model input variable

机森林、提升树 3 种机器学习算法各构建了 4 种不同组合输入变量模型,总计构建了 12 组 TDG 预测模型,各模型评价指标参数如图 3 所示。

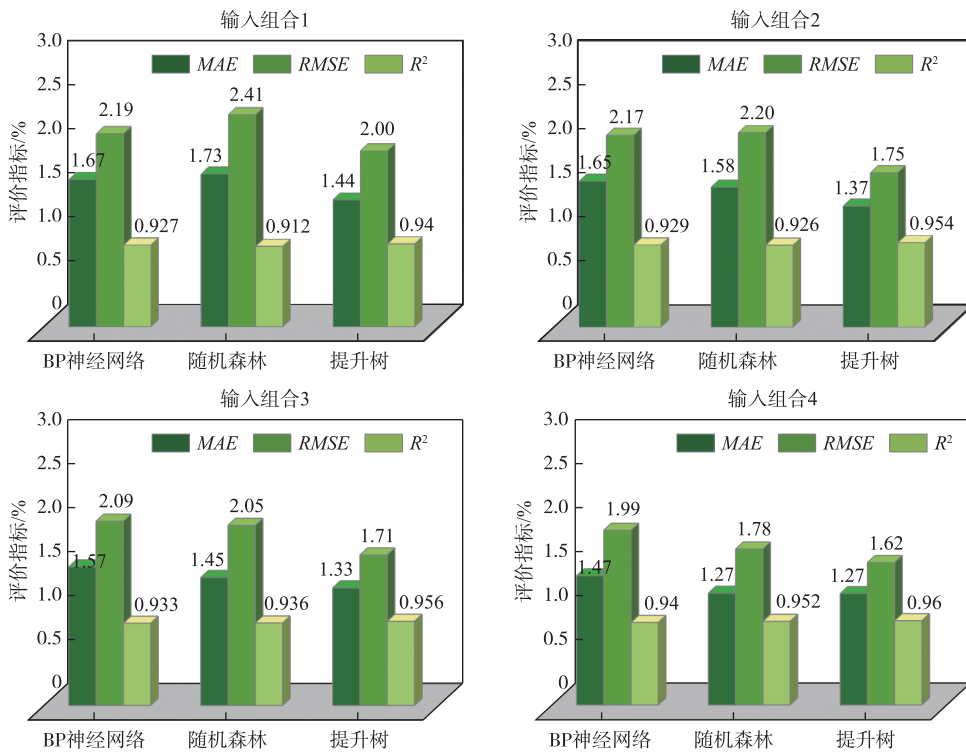


图 3 4 种输入组合下的测试集预测性能评估

Fig.3 Performance evaluation of test set prediction under four input combinations

为确定输入变量对 TDG 预测模型的重要性,在输入组合 1 的基础上分析流量、水温、气压 3 种因子对各模型的影响。对于 BP 神经网络模型,流量、水温、气压分别使模型的 RMSE 降低了 1.0%、3.4%、4.9%,说明气压的影响最大,流量的影响最小。对于随机森林模型,流量、水温、气压分别使模型的 RMSE 降低了 8.5%、7.0%、13.2%,说明气压的影响最大,水温的影响最小。对于提升树模型,流量、水温、气压分别使模型的

$RMSE$ 降低了 12.5%、2.1%、5.1%,说明流量的影响最大,水温的影响最小。因此,不同模型受输入因子的影响各不相同。

随着输入因子的增加,3种模型的预测精度都有所提高(图3),在最佳输入变量组合下,模型预测精度从大到小分别为提升树($MAE = 1.268\%$, $RMSE = 1.622\%$, $R^2 = 0.96$)、随机森林($MAE = 1.273\%$, $RMSE = 1.775\%$, $R^2 = 0.952$)、BP神经网络模型($MAE = 1.47\%$, $RMSE = 1.99\%$, $R^2 = 0.94$)。当BP神经网络模型从输入组合1增加到输入组合4时,其模型的 MAE 和 $RMSE$ 分别降低了10.8%和9.2%, R^2 增加了1.4%。当随机森林模型从输入组合1增加到输入组合4时,其模型的 MAE 和 $RMSE$ 分别降低了26.5%和26.2%, R^2 则增加了4.5%。当提升树模型从输入组合1增加到输入组合4时,其模型的 MAE 和 $RMSE$ 分别降低了11.9%和18.7%, R^2 增加了2.2%。说明随机森林模型受输入因子数的影响较大,即当更多的影响因子作为模型的输入变量时,它呈现出最好的改进,而BP神经网络模型受输入因子数的影响较小。

使用4个变量作为输入因子,3个模型都获得了最佳精度。图4是不同模型输入组合4时测试集实测值与预测值的散点图,可以明显看出,与其他2个模型相比,提升树模型的拟合线方程更接近精确线($y=x$)。

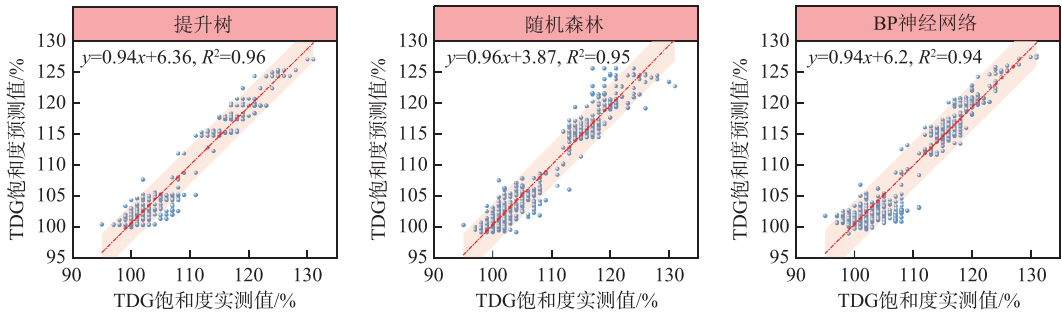


图4 输入组合4的各模型测试集线性拟合(粉色区域为 $y=x$ 的95%置信区间)

Fig.4 Linear fit to each model test set for input combination 4
(The pink area represents the 95% confidence interval of $y=x$)

3 模型验证和结果分析

基于以上分析结果,得出提升树在预测TDG方面表现最佳,其次是随机森林和BP神经网络模型。为了验证所建立模型的适用性,将3个TDG预测模型用于预测哥伦比亚河邦纳维尔大坝上游的约翰戴和达尔斯大坝监测站点的TDG饱和度,收集JHAW站点和TDDO站点2023年一整年的日监测数据(T、P、D、SFD),经过同样的归一化处理,作为所建立模型的输入因子,从而预测出约翰戴大坝和达尔斯大坝下游的TDG饱和度,结果如图5所示。

图5展示了2个监测站点2023年TDG饱和度的实测值与预测值,其中,提升树与随机森林模型的预测值与实测值的拟合度较高,几乎所有预测值都在实测值的5%误差范围内,预测结果呈现出整体上的一致性。但从细节上仍然表现出一定的差异性,在TDG饱和度低于115%时,提升树与随机森林的预测结果较为一致,但当TDG饱和度高于115%时,随机森林模型的预测误差较大。相比之下,提升树的预测结果相对更接近实测值。两次验证阶段,BP神经网络预测值与实测值误差较大,原因分析认为,该模型在预测TDG饱和度时出现了“过拟合”现象。一般情况下,BP神经网络模型的训练能力与预测能力呈正比,但这种趋势不是固定的,而是有一个极限,当达到此极限时,随着训练能力的提高,预测能力反而会下降,即出现“过拟合”现象,或称为模型的泛化能力差。BP神经网络算法是一种基于经验风险最小化原则的神经网络方法,它对训练数据的拟合程度过于敏感,导致模型参数选择复杂、优化困难、泛化能力差和稳定性低。

为进一步分析预测性能,计算出3个模型在2个站点的预测值与实测值之间的平均相对误差来评价预测精度,结果如图6所示。对于JHAW站点的预测,提升树模型、随机森林模型以及BP神经网络模型的预测值与实测值之间的平均相对误差分别为2.34%、2.69%、7.08%;对于TDDO站点的预测,提升树模型、随机

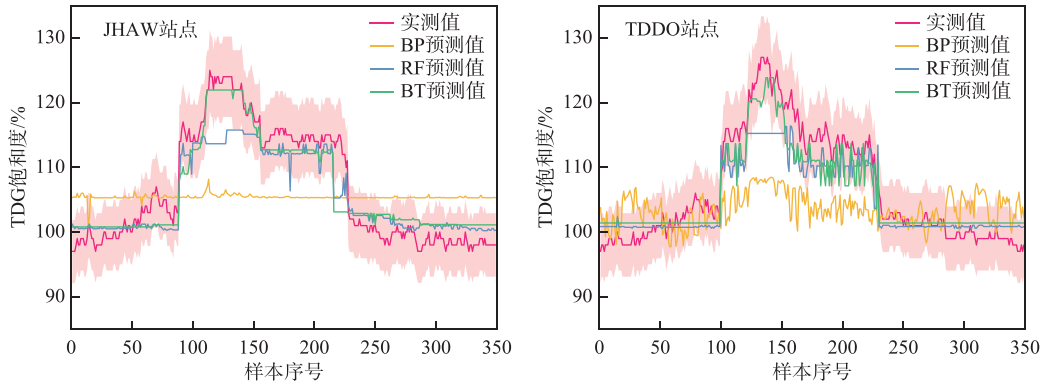


图 5 JHAW 站点和 TDDO 站点不同模型预测值与实测值对比图(粉色区域代表实测值的 5% 误差范围)
 Fig.5 Comparison of predicted and measured values of different models at JHAW site and TDDO site
 (The pink area represents a 5% error range of the measured value)

森林模型以及 BP 神经网络模型的预测值与实测值之间的平均相对误差分别为 2.31%、2.64%、5.81%。分析可知,提升树的预测精度最高,其次是随机森林模型。总的来看,提升树与随机森林模型预测精度没有太大的差异,平均相对误差保持在 2.7% 以内,对于 JHAW 站点,随机森林预测精度波动较小,对于 TDDO 站点,提升树模型预测精度波动较小。本文旨在建立高精度的 TDG 饱和度预测模型,因此,从模型的建立到验证阶段,无论是输入因子数量、各种评价指标、预测值与实测值的对比及相对误差大小,提升树都表现出最好的预测结果。故与本文其他 2 种模型相比,提升树模型更适用于预测大坝下游 TDG 饱和度。

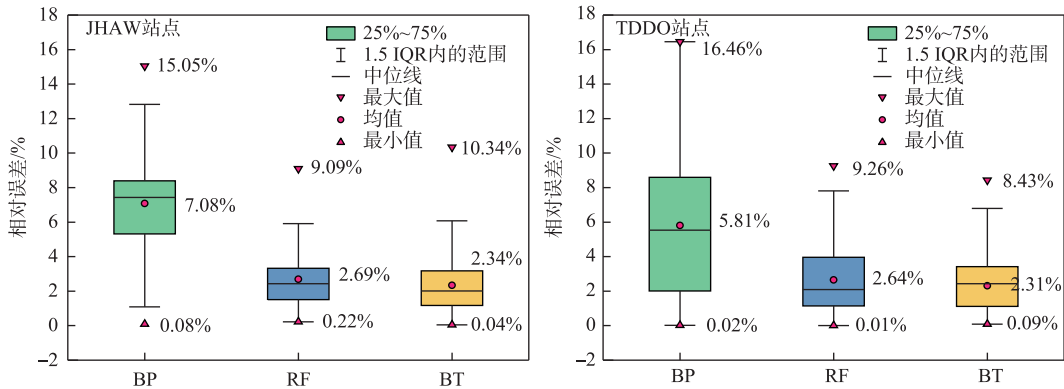


图 6 最佳输入变量组合下 3 种模型预测值与实测值的相对误差
 Fig.6 Relative error between predicted values and measured values of three models
 under the optimal combination of input variables

4 结论

本文采用哥伦比亚河上邦纳维尔大坝下游 WRNO 监测站点近十年的日监测数据,以水温、气压、流量、大坝溢流 4 个指标作为输入变量,建立了 3 种基于机器学习算法的 TDG 预测模型。并采用平均绝对误差、均方根误差和决定系数对其进行评估。此外,选择位于哥伦比亚河上约翰戴大坝和达尔斯大坝下游的 JHAW 站点和 TDDO 站点的数据对模型进行检验。主要结论如下:

1) 随着输入变量的增加,3 种模型的预测性能都有所提高。各模型对不同输入因子的影响各不相同,大坝溢流和气压对 BP 神经网络和随机森林模型影响最大,大坝溢流和流量对提升树模型的影响最大。

2)对比BP神经网络和随机森林算法,提升树算法在预测大坝下游TDG饱和度方面表现出了更高的预测精度。

基于各种精度评价及模型验证,提升树算法构建的TDG预测模型,可将TDG饱和度的平均相对误差控制在2.4%以内,实现了大坝下游河道TDG的快速精准预测,可为基于机器学习的TDG预测模型提供相关借鉴。由于目前大坝下游TDG以及其他影响因子的监测数据较少,机器学习模型的预测性能受到限制,未来还应开展具有更多参数指标的机器学习算法的研究,尝试选择更先进的机器学习方法,进一步提高TDG预测模型的稳定性和准确性。

5 参考文献

- [1] Cheng XJ, Chen YC. Theoretical analysis of dissolved gas supersaturation downstream of sluicing dam and application. *Advances in Water Science*, 2007, **18**(3): 346-350. [程香菊, 陈永灿. 大坝泄洪下游水体溶解气体超饱和理论分析及应用. 水科学进展, 2007, **18**(3): 346-350.]
- [2] Wang HR, Zhou ZL, Xing YN et al. On supersaturation of total dissolved gas for hydraulic projects. *Advances in Science and Technology of Water Resources*, 2010, **30**(5): 12-15. [王皓冉, 周卓灵, 行亚楠等. 水利工程总溶解气体过饱和问题探讨. 水利水电科技进展, 2010, **30**(5): 12-15.]
- [3] Politano MS, Carrica PM, Turan C et al. A multidimensional two-phase flow model for the total dissolved gas downstream of spillways. *Journal of Hydraulic Research*, 2007, **45**(2): 165-177. DOI: 10.1080/00221686.2007.9521757.
- [4] Brown RS, Carlson TJ, Gingerich AJ et al. Quantifying mortal injury of juvenile Chinook salmon exposed to simulated hydro-turbine passage. *Transactions of the American Fisheries Society*, 2012, **141**(1): 147-157. DOI: 10.1080/00028487.2011.650274.
- [5] Tan D, Ni Z, Zhen Y et al. Dissolved gas supersaturation downstream of dam and its effects on fish. *Freshwater Fisheries*, 2006, **36**(3): 56-59.
- [6] Li R, Li KF, Feng JJ et al. Review on the effect of dissolved gas supersaturation of dam spill on fishes and its mitigation measures. *Advanced Engineering Science*, 2023, **55**(4): 91-101. DOI: 10.15961/j.jsuese.202200117. [李然, 李克锋, 冯镜洁等. 水坝泄水气体过饱和对鱼类影响及减缓技术研究综述. 工程科学与技术, 2023, **55**(4): 91-101.]
- [7] Beeman JW, Maule AG. Migration depths of juvenile Chinook salmon and steelhead relative to total dissolved gas supersaturation in a Columbia River reservoir. *Transactions of the American Fisheries Society*, 2006, **135**(3): 584-594. DOI: 10.1577/T05-193.1.
- [8] Wang YM, Liang RF, Tuo YC et al. Tolerance and avoidance behavior towards gas supersaturation in rock carp *Procypris rabaudi* with a history of previous exposure. *North American Journal of Aquaculture*, 2015, **77**(4): 478-484. DOI: 10.1080/15222055.2015.1059913.
- [9] Pleizier NK, Algera D, Cooke SJ et al. A meta-analysis of gas bubble trauma in fish. *Fish and Fisheries*, 2020, **21**(6): 1175-1194. DOI: 10.1111/faf.12496.
- [10] Shen X, Hodges BR, Li R et al. Factors influencing distribution characteristics of total dissolved gas supersaturation at confluences. *Water Resources Research*, 2021, **57**(6): e2020WR028760. DOI: 10.1029/2020WR028760.
- [11] Shen X, Li R, Hodges BR et al. Experiment and simulation of supersaturated total dissolved gas dissipation: Focus on the effect of confluence types. *Water Research*, 2019, **155**: 320-332. DOI: 10.1016/j.watres.2019.02.056.
- [12] Shen X, Li R, Huang JP et al. Shelter construction for fish at the confluence of a river to avoid the effects of total dissolved gas supersaturation. *Ecological Engineering*, 2016, **97**: 642-648. DOI: 10.1016/j.ecoleng.2016.10.055.
- [13] Fu XL, Li D, Zhang XF. Simulations of the three-dimensional total dissolved gas saturation downstream of spillways under unsteady conditions. *Journal of Hydrodynamics*, 2010, **22**(4): 598-604. DOI: 10.1016/S1001-6058(09)60093-7.
- [14] Ma Q, Li R, Feng JJ et al. Cumulative effects of cascade hydropower stations on total dissolved gas supersaturation. *Environmental Science and Pollution Research*, 2018, **25**(14): 13536-13547. DOI: 10.1007/s11356-018-1496-2.
- [15] Ma Q, Li R, Feng JJ et al. Ecological regulation of cascade hydropower stations to reduce the risk of supersaturated total dissolved gas to fish. *Journal of Hydro-Environment Research*, 2019, **27**: 102-115. DOI: 10.1016/j.jher.2019.10.002.
- [16] Politano M, Carrica P, Weber L. A multiphase model for the hydrodynamics and total dissolved gas in tailraces. *International Journal of Multiphase Flow*, 2009, **35**(11): 1036-1050. DOI: 10.1016/j.ijmultiphaseflow.2009.06.009.
- [17] Politano M, Arenas Amado A, Bickford S et al. Evaluation of operational strategies to minimize gas supersaturation downstream of a dam. *Computers & Fluids*, 2012, **68**: 168-185. DOI: 10.1016/j.compfluid.2012.08.003.
- [18] Ma Q, Liang RF, Li R et al. Operational regulation of water replenishment to reduce supersaturated total dissolved gas in riverine wetlands. *Ecological Engineering*, 2016, **96**: 162-169. DOI: 10.1016/j.ecoleng.2016.03.019.
- [19] Feng JJ, Wang L, Li R et al. Operational regulation of a hydropower cascade based on the mitigation of the total dissolved gas supersaturation. *Ecological Indicators*, 2018, **92**: 124-132. DOI: 10.1016/j.ecolind.2017.04.015.
- [20] Wan H, Tan Q, Li R et al. Incorporating fish tolerance to supersaturated total dissolved gas for generating flood pulse discharge patterns

- based on a simulation-optimization approach. *Water Resources Research*, 2021, **57**(9): e2021WR030167. DOI: 10.1029/2021WR030167.
- [21] Zhang C, Wang FR, Yu PW *et al.* Personalized dynamic promotion based on deep augmented learning. *Management World*, 2023, **39**(5): 160-178. [张诚, 王富荣, 郁培文等. 基于深度增强学习的个性化动态促销. 管理世界, 2023, **39**(5): 160-178.]
- [22] Zhou Y, Duan YR. Retail products sales forecast based on clustering and machine learning. *Computer System Applications*, 2021, **30**(11): 188-194. [周雨, 段永瑞. 基于聚类与机器学习的零售商品销量预测. 计算机系统应用, 2021, **30**(11): 188-194.]
- [23] Zhou C, Jin XQ, Guo ZL *et al.* Comparison of cognitive frailty risk prediction models for community older adults based on machine learning algorithms. *Journal of Nursing Science*, 2023, **38**(19): 1-5, 11. [周闯, 金学勤, 郭正丽等. 不同机器学习算法的社区老年人认知衰弱风险预测模型比较. 护理学杂志, 2023, **38**(19): 1-5, 11.]
- [24] Shu LM, Li QF, Gu HM *et al.* Research progress of machine learning in the early identification and nursing of neonatal sepsis. *Chinese Journal of Emergency and Critical Care Nursing*, 2023, **4**(6): 515-518. [束礼梅, 李秋芳, 顾慧敏等. 机器学习在新生儿败血症早期识别和护理中的研究进展. 中华急危重症护理杂志, 2023, **4**(6): 515-518.]
- [25] Gu J, Qian YR, Wang LL *et al.* Survey of artificial intelligence in functional magnetic resonance imaging data for autism. *Computer Engineering and Applications*, 2023, **59**(22): 57-68. [顾剑, 钱育蓉, 王兰兰等. 人工智能在功能磁共振成像数据中的自闭症研究综述. 计算机工程与应用, 2023, **59**(22): 57-68.]
- [26] 申志豪, 李娜, 尹世豪等. 基于 TPA-Transformer 的机票价格预测. 数据与计算发展前沿, 2023, **5**(6): 115-125.
- [27] Pan ZT. Prediction and analysis of China freight transportation based on machine learning time series. *China Storage & Transport*, 2022, (1): 85-86. [潘正桐. 基于机器学习时间序列的中国货物运输量预测分析. 中国储运, 2022, (1): 85-86.]
- [28] Ouyang CY, Qin Y, Liu Z *et al.* Prediction of CO₂, CH₄ diffusion fluxes at the water-air interface and analysis on its influencing factors using machine learning algorithms in the Three Gorges Reservoir. *J Lake Sci*, 2023, **35**(2): 449-459. DOI: 10.18307/2023.0206. [欧阳常悦, 秦宇, 刘臻等. 基于机器学习的水-气界面 CO₂、CH₄扩散通量预测及影响因素分析——以三峡水库为例. 湖泊科学, 2023, **35**(2): 449-459.]
- [29] Shi CY, Liu F, Zhu K *et al.* Combining long-term remote sensing monitoring and machine learning algorithms to predict spatiotemporal changes in algal proliferation risk. *J Lake Sci*, 2024, **36**(3): 670-684. DOI: 10.18307/2024.0311. [时晨熤, 刘凤, 祝铠等. 结合长时序遥感监测和机器学习算法预测藻类增殖风险时空变化. 湖泊科学, 2024, **36**(3): 670-684.]
- [30] Moon J, Lee J, Lee S *et al.* Urban river dissolved oxygen prediction model using machine learning. *Water*, 2022, **14**(12): 1899. DOI: 10.3390/w14121899.
- [31] Adnan RM, Mostafa RR, Kisi O *et al.* Improving streamflow prediction using a new hybrid ELM model combined with hybrid particle swarm optimization and grey wolf optimization. *Knowledge-Based Systems*, 2021, **230**: 107379. DOI: 10.1016/j.knsys.2021.107379.
- [32] Heddam S. Generalized regression neural network based approach as a new tool for predicting total dissolved gas (TDG) downstream of spillways of dams: A case study of Columbia River Basin dams, USA. *Environmental Processes*, 2017, **4**(1): 235-253. DOI: 10.1007/s40710-016-0196-5.
- [33] Heddam S, Keshtegar B, Kisi O. Predicting total dissolved gas concentration on a daily scale using Kriging interpolation, response surface method and artificial neural network: Case study of Columbia River Basin dams, USA. *Natural Resources Research*, 2020, **29**(3): 1801-1818. DOI: 10.1007/s11053-019-09524-2.
- [34] AlOmar MK, Hameed MM, Al-Ansari N *et al.* Data-driven model for the prediction of total dissolved gas: Robust artificial intelligence approach. *Advances in Civil Engineering*, 2020, **2020**(1): 1-20. DOI: 10.1155/2020/6618842.
- [35] Wang ZH, Feng JJ, Liang MY *et al.* Prediction model and application of machine learning for supersaturated total dissolved gas generation in high dam discharge. *Water Research*, 2022, **220**: 118682. DOI: 10.1016/j.watres.2022.118682.