

基于机器学习预测模型的水库叠梁门措施对水库下泄水温调节作用的验证研究

朴虹奕, 周湘山, 徐劲草, 张磊, 秦甦

(中国电建集团成都勘测设计研究院有限公司, 四川 成都 611130)

摘要:通过使用 R 语言编程软件对原始数据的归一化处理 and 参数的筛选与成分相关性分析,建立了基于机器学习 5 种算法预测模型。结果表明:通过对模型的训练和修正,5 种基于机器学习的水库下泄水温预测模型中,支持向量机模型的效果最佳,模型预测值的准确度可以达到 0.78。支持向量机水库水温预测模型可以实现对水温的监测和控制,从而对水库的生态环境影响及时采取优化措施,并进行结果的预测验证。

关键词:机器学习;统计预测模型;水温;鱼类繁殖

中图分类号:P332.6

文献标志码:B

文章编号:1001-2184(2024)06-0116-04

Verification Study on the Effect of Reservoir Stoplog Measures on Reservoir Discharge Water Temperature Regulation Based on Machine Algorithm Prediction Model

PIAO Hongyi, ZHOU Xiangshan, XU Jingcao, ZHAN Lei, QIN Su

(PowerChina Chengdu Engineering Corporation Limited, Chengdu Sichuan 611130)

Abstract: By using R language programming software to first normalize the original data, screen parameters and screen component correlation, five prediction models based on machine learning algorithms have been established. The results show that through the training and modification of the model, the support vector machine model has the best effect among the 5 kinds of machine learning-based prediction models of reservoir drainage water temperature, and the accuracy of this model prediction value has reached to 0.78. Support vector machine reservoir water temperature prediction model can not only realize the monitoring and control of water temperature, but also to take timely optimization measures and to get immediate validation for the ecological environment of the reservoir, and verify the prediction results.

Key words: Machine learning; Statistical prediction model; Water temperature; Fish reproduction

1 概述

水库的蓄水运行会造成水温分层现象,直接对水库水下泄会导致其影响区域内河道的水温异于河道天然水温^[1]。这种影响会使天然河道里动植物生长发育、数量分布以及栖息地等重要指标变化,在鱼类繁殖产卵关键时期的春季,水温变化对生态环境的影响更加不容忽视^[2]。因此,科学的监测水库修建后所造成的生态环境影响并及时采取针对性措施显得至关重要。

为及时了解泄水对水库下游生态环境的影响并采取应对措施,可以对水库周边的生态环境进行实时监测,并根据监测的数据建立模型进行预测。目前,国内外学者已经开展了大量水库水温

预测模型的研究,部分结果已得到应用^[3-4]。垂直一维、二维、三维水库水温数值模型在我国多处水库得到广泛应用^[5-8],但绝大多数的模型都需要经过大量的计算和参数进行校准,验证过程耗时长。相比传统的物理水温预测模型,统计模型的优点在于模型相对简单,易操作,输入数据边界限制少,避免了传统机理模型在预测中耗费大量时间,并且模拟工况较为固定,无法满足复杂条件下水库生态调度快速决策需求的困难。

利用已有的锦屏一级水电站水库的水温、气温以及水文数据建立了五种统计模型,并对得到的预测结果进行了验证分析与对比评价。由此探索通过统计模型算法对水库下生态环境改良措施的效果进行验证及预测的可行性及效果,为水库

收稿日期:2024-10-13

运行者和管理者对水库的生态环境进行更有利的监测和把控提供思路。

2 预测模型的参数确定

2.1 数据的处理

该研究将使用 R 语言作为数据的处理和模型的构建。R 语言是一种数学编程语言,主要用于统计分析、绘图、数据挖掘。为了避免一些极端的数据(过大或者过小)影响整体样本的精确度,而导致一些相关性高的因子无法体现,在开始正式建模前,把收集到的数据进行归一化操作,再通过中心极限定理,可以知道误差的分布并将数据中出现的较大偏差值(标有数字的值)筛选出来,从而清除数据的残留值,使数据更具有统一性和代表性。

2.2 确定参数

涉及的预测模型分为对天然水温的预测模型

和下泄水温的预测模型。为了找出影响天然水温和水库下泄水温最直接、显著的原因从而保证模型的精度,需要将所有的影响因子与天然水温和水库下泄水温进行相关性筛选,然后分析每一个自变量与因变量的 R 值以及是否是正向相关性。天然水温与入库水温、天然入库流量和气温之间的相关性很强,均高于 0.6,于是选择入库水温、将 2005 年到 2021 年雅砻江中下游锦屏一级水文站以及气象站的资料放进 R,进行运行,得到天然入库流量和气温作为坝址天然水温预测模型的参数;建库后的下泄水温与入库流量、入库水温和气温之间的相关性很强,均高于 0.7,于是选择入库流量、入库水温和气温作为下泄水温预测模型的参数,天然水温模型预测参数相关性分析和下泄水温模型预测参数相关性分析见表 1、2。各个指标皮尔逊相关系数见图 1。

天然水温	1.00	0.70	0.88	0.80	0.89	0.60	0.65	0.53	0.34	0.32
下泄水温	0.70	1.00	0.76	0.80	0.42	0.76	0.82	0.53	0.32	0.23
入库水温	0.88	0.76	1.00	0.75	0.83	0.66	0.68	0.65	0.33	0.31
库尾水温	0.80	0.80	0.75	1.00	0.52	0.76	0.77	0.45	0.32	0.31
天然入库流量	0.89	0.42	0.83	0.52	1.00	0.78	0.87	0.45	0.32	0.21
流量	0.60	0.76	0.66	0.76	0.78	1.00	0.67	0.78	0.32	0.21
气温	0.62	0.82	0.68	0.77	0.87	0.67	1.00	0.23	0.31	0.21
水位	0.53	0.53	0.65	0.45	0.45	0.78	0.23	1.00	0.11	0.21
风速	0.34	0.32	0.33	0.32	0.32	0.32	0.31	0.11	1.00	0.80
风向	0.32	0.23	0.31	0.31	0.21	0.21	0.21	0.21	0.80	1.00
	天然水温	下泄水温	入库水温	库尾水温	天然入库流量	流量	气温	水位	风速	风向

图 1 各个指标皮尔逊相关系数

表 1 天然水温模型预测参数相关性分析

涉及因素	参数相关性		
	相关系数 R	决定系数 R ²	
天然入库流量	0.882 3	0.912 4	
坝址天然水温	入库水温	0.824 2	0.853 4
	气温	0.763 4	0.823 2
	水位	0.554 3	0.602 4
	风速	0.532 1	0.553 2
	风向	0.354 3	0.393 2

表 2 下泄水温模型预测参数相关性分析

锦屏站	涉及因素	参数相关性	
		相关系数 R	决定系数 R ²
下泄水温	气温	0.931 9	0.965 3
	库尾水温	0.874 3	0.892 4
	流量	0.785 4	0.834 2
	水位	0.654 3	0.688 4
	风速	0.522 1	0.543 2
	风向	0.324 3	0.383 2

3 模型的建立、训练以及筛选

3.1 模型的建立

将确定的参数代入初代模型,其中包括测试线性模型、神经网络模型、朴素贝叶斯模型、随机森林模型和支持向量机模型共计 5 种模型。通过设定模型初始参数值,数据拟合度以及模型适配后,再对已有的数据进行模型的训练和修正。以支持向量为例,在目前的研究中,关于支持向量机参数优化的问题采用的是网格搜索与交叉验证相结合的方法。首先将模型参数(C, g)分别取值的 m, n 个组进行组合评分,选出评分最高的组合作为模型的参数(网络搜索法),然后再把原始数据分为 5 组,将其中的 4 组作为训练集用于模型的训练和修正,剩下的一组将作为测试模型以测试模型的准确度,5 次结果的均值作为对模型精度(交叉验证法)。

3.2 模型的训练和筛选

如表 3、4 所示,参数优化过后的模型平均值

误差(MAE)和均方根误差(RMSE)都比默认参数减少并更趋近于 0(当预测值与真实值完全吻合时等于 0),表明优化后的参数使真实值与预测值的偏差减小了,而且模型的拟合度 R^2 也提升并达到了 0.8,说明支持向量机模型的预测准确程度有 80%。

表 3 支持向量机预测模型参数优化

默认参数	C	1
	g	auto
	C 搜索范围	0~10 000
优化后参数	C	20
	g 搜索范围	0.000 1~10.000 0
	g	0.008

表 4 支持向量机预测模型参数优化后结果

指标\模型	默认参数	优化参数
RMSE	0.145	0.132
MAE	0.112	0.103
MAPE	7.963	8.020
R^2	0.757	0.789

5 种模型分别训练和修正后得到的算法模型精度测试结果见表 5。通过测试模型预测的准确度(整体精度)和模型预测值与实际值的一致性(Kappa),作为检验一致性的指标。当指标趋近于 1 时,说明预测模型的有效度高,而指标趋近于 0 时,则证明预测模型有效度低。结果发现线性模型的预测精度最低,基本可以判定为水温与其相关因子呈现的是非线性关系,且预测与实际值的一致性也极低,表明结果的偏向性很高,这代表预测的结果分布很不均匀。而所有模型中支持向量机算法模型的整体精度和 Kappa 精度最高,这表明这个算法的预测准确率最高,和数据所呈现的规律和运算模式很契合,而且一致性也在 0.5 以上,说明预测值与实际值有较高的一致性。因此确定支持向量机模型作为鱼类产卵期水库下泄水温的算法模型。

3.3 支持向量机算法理论介绍

支持向量机的方程为

$$W^T x + b = 0 \quad (1)$$

式中: W^T , b 为支持向量机中的无量纲参数。如公式大于零,则参数与下泄水温判定为属于同一

表 5 算法模型精度测试结果

识别方法	整体精度	Kappa
线性模型	0.34	0.23
神经网络模型	0.46	0.33
朴素贝叶斯模型	0.52	0.42
随机森林模型	0.59	0.43
支持向量机模型	0.78	0.64

类别,如公式小于零,则判定该参数与下泄水温不为同一类别。

$$d = \frac{(|W^T x + b|)}{(\|w\|)} \quad (2)$$

根据支持向量机参数之间的几何距离,来判断模型的预测效果, d 为参数与预测值之间的精度距离,距离越小说明模型预测精度越高。

4 基于支持向量机模型的叠梁门措施效果预测

锦屏一级水电站于 2012 年 12 月蓄水,2014 年 7 月全部机组投产发电,将库尾水温和入库水温数据与天然下泄水温逐月进行对比可以发现,自水库建成后,库区存在显著的水温分层现象。3~8 月,随着入库水温、气温及太阳辐射强度等增高,表层水体逐渐升温,密度减小且表、底温差不断变大;此外,该时段入、出库流量也会显著增大,受电站进水口附近水体紊动影响,库区出现双温跃层,温度差引起的密度分层使热量难以向下传递,由此导致水库下泄水温较干流入库水温低 $0.3 \sim 1.4 \text{ }^\circ\text{C}$,而 3~6 月恰为鱼类产卵期的重要时段。由此可以得出结论:水库的修建确实会对河流下泄水温产生影响,从而对鱼类的繁殖产卵环境造成影响。

(1)坝址天然水温的预测模型。通过将天然水温作为预测 x 值,把锦屏一级水电站的流量、库尾水温和气温作为 y 值当作模型构建的参数,带入到支持向量机预测模型中获得的预测值为未受到水库修建影响的天然水温值。

(2)下泄水温预测模型。通过将下泄水温作为预测 x 值,把锦屏一级水电站投入使用后的流量、库尾水温和气温作为 y 值当作模型构建的参数,带入到支持向量机预测模型中获得的预测值为未实施措施时水库下泄水温值。

(3)实施了叠梁门的下泄水温预测模型。通过将实施了叠梁门措施的下泄水温作为预测 x

值,把实施了叠梁门措施的坝下水位、库尾水温和气温作为 y 值当作模型构建的参数,带入到支持向量机预测模型中获得的预测值为实施了叠梁门措施的水库下泄水温值。

锦屏一级水电站下泄水温模型预测结果见表6,通过对比3~6月天然水温的预测值、水库下泄水温预测值和实施了叠梁门的下泄水温预测值,可以看到实施后的3~6月的坝下水温有明显的改善。其中,天然水温的预测月平均值达到13.9℃,而水坝建成后的下泄水温只有11.8℃。实施叠梁门措施后,下泄水温与原始天然水温差值不大于1℃,3~6月平均值也少于1℃。从模型的预测结果可以得知:在3~6月的时间段增加叠梁门措施可以使水库的下泄水温更接近天然水温,从而减少因为水库的修建,对下游鱼类繁殖产卵重要时段所造成的影响,减少对生态的破坏。

表6 锦屏一级水电站下泄水温模型预测结果

站点	时期	/℃				
		3月	4月	5月	6月	3~6月 平均值
锦屏一级 水电站 下泄水温	天然	11.10	13.90	16.50	17.80	14.83
	实施前	10.50	11.80	13.00	15.20	12.63
	实施后	10.30	12.90	13.00	16.00	13.05
	天然水温 对比实施后 的差值	0.80	1.00	3.50	1.80	0.05

5 结 语

以统计模型为基础,根据锦屏一级水电站的水库下泄水温、流量、水位、气温和库尾水温和入库水温等,建立锦屏一级水电站的天然水温预测模型和下泄水温预测模型,实现了锦屏一级水电站坝下河段的生态、水温的模拟和下泄水温的调控。

(1)将锦屏一级水电站的水温、气温和流量等数据进行归一化操作,将数据中出现的较大偏差值筛选出来,使数据更具有统一性和代表性。将参数与参数之间的相关性进行对比分析,并筛选出流量、库尾水温和气温为影响天然水温和下泄水温的决定性影响因子。将确定好的参数带入测试线性模型、神经网络模型、朴素贝叶斯模型、随机森林模型和支持向量机模型,通过优化默认参

数修正和训练,并对优化的参数进行模型预测的验证,得到模型预测值与实际值的拟合度来判定最终的水温预测模型。

(2)5种机器算法构建的水温预测模型中,支持向量机的精度是所有模型中最高的,模型预测准确度达到0.78,主要原因在于选择的是集中在3~6月的数据,样本量偏小,有利于支持向量机在有限的数量下得出最佳规律。除此之外,由于支持向量机很善于处理非线性问题,特别是在该研究中决定性系数之间的非线性关系,可以通过支持向量机将样本从低维度映射到高维功能,将非线性问题转化为线性问题,从而提高模型预测的精度值。

(3)通过增加叠梁门措施的数据比较分析以及支持向量机模型的预测结果得出:增加了叠梁门措施后水库下泄水温比没有做措施前的水温有明显的升高,使3~6月的下泄水温更加接近天然水温。

(4)相比传统的物理水温预测模型,机器学习方法可以有效利用已有的实测数据对未来的水库天然水温情况或者是对实施了措施的水库水温状况进行精度较高的预测。机器学习模型相对简单,好操作,需要的数据量少且不用对水库水温数据进行持续的跟踪收集,模型的预测值与实际值拟合度很高,因此可以有效地对水库的水温环境进行监控和及时采取措施。这种创新性的水温预测模型对下游水生生态环境的保护具有重要意义。

参考文献:

- [1] 刘有志,相建方,陈文夫,等.狭长型水库蓄水至初期运行阶段水温演化规律研究[J].水利学报,2020,51(11):1412-1422.
- [2] 王悦,叶绿,高千红.三峡库区175m试验性蓄水期间水温变化分析[J].人民长江,2011,42(15):5-8.
- [3] 胡平,刘毅,唐忠敏,等.水库水温数值预测方法[J].水利学报,2010,41(9):1045-1053. DOI:10.13243/j.cnki.slx.2010.09.013.
- [4] 阮娅,脱友才,邓云,等.乌东德水库水温预测及低温水减缓措施[J].长江流域资源与环境,2017,26(11):1912-1918.
- [5] 邹家祥.环境影响评价技术手册:水利水电工程[M].北京:中国环境科学出版社,2009:212-213.
- [6] 孙祥,朱广伟,宣文怡,等.天目湖沙河水库热分层变化及其对水质的影响[J].环境科学,2018,39(6):2632-2640.

(下转第133页)



图6 堆积体圆弧型滑塌变形

层及三叠系上统砂板岩为主。工程区位于“雅江—理塘”菱形断块内部,区域构造稳定性相对较好,局部区域构造较为强烈,不良地质较发育。

(2)两河口库区常见变形岸坡类型可分为倾倒变形岸坡、顺层岩质岸坡、岩土混合岸坡、堆积体岸坡四类,其中,倾倒变形岸坡及堆积体岸坡的变形破坏是库区蓄水过程中关注的重点。

(3)根据现场调研,两河口库区岸坡变形破坏类型主要有五类,分别为弯曲—拉裂型、滑移—弯

曲型、滑移—拉裂型、楔形体滑移、圆弧型滑塌破坏,其中弯曲—拉裂型破坏及堆积体岸坡(类)圆弧型滑塌在库区滑坡中占多数。

参考文献:

- [1] 张雄,杨泽艳. 我国水电工程边坡治理若干问题的思考与研究[J]. 水力发电,2018,44(8): 23-27,46.
- [2] 段亮. 黄草坪倾倒岸坡变形特征与库水位变化关系研究[D]. 成都:成都理工大学,2021.
- [3] 王平,胡明军,黄波林,等. 三峡库区巫峡剪刀峰顺层岩质岸坡破坏模式分析[J]. 中国地质灾害与防治学报,2021,32(5):52-61.
- [4] 吴章雷,钟雨田. 两河口水电站多源复杂防渗土料勘察研究[J]. 四川水力发电,2022,41(6): 28-34.
- [5] 向芸辛. 三峡库区顺层岩质滑坡治理效果综合评价—以天德村滑坡为例[D]. 重庆:重庆三峡学院,2023.

作者简介:

- 陈亚洲(1992-),男,河南商丘人,工程师,硕士,从事岩土工程勘察与设计工作;
- 杨鹏(1990-),男,四川营山人,工程师,本科,从事岩土工程勘察与设计工作;
- 曹廷(1987-),男,四川南充人,高级工程师,硕士,从事岩土工程勘察与设计工作。

(编辑:吴永红)

(上接第119页)

- [7] Obenour D R, Michalak A M, Zhou Y T, et al. Quantifying the impacts of stratification and nutrient loading on hypoxia in the northern Gulf of Mexico[J]. Environmental Science & Technology, 2012, 46(10): 5489-5496.
- [8] Lee Y G, Kang J H, Ki S J, et al. Factors dominating stratification cycle and seasonal water quality variation in a Korean estuarine reservoir[J]. Journal of Environmental Monitoring, 2010, 12(5): 1072-1081.

作者简介:

朴虹奕(1995-),女,四川成都人,助理工程师,研究生,从事水土保持监测、生态环境监测、环境保护信息化系统平台建设等工作;

- 周湘山(1986-),男,四川成都人,高级工程师,博士,从事水土保持监测、无人机航测及三维建模、生态环境监测、环境保护信息化系统平台建设等工作;
- 徐劲草(1982-),男,四川成都人,高级工程师,博士,从事水温监测、生态流量监测、生态环境监测、环境保护信息化系统平台建设等工作;
- 张磊(1982-),男,四川成都人,高级工程师,本科,从事水土保持监测、无人机航测及三维建模、生态环境监测、环境保护信息化系统平台建设等工作;
- 秦甦(1973-),男,四川成都人,正高级工程师,本科,从事水土保持监测、无人机航测及三维建模、生态环境监测、环境保护信息化系统平台建设等工作。

(编辑:吴永红)