

鱼类产卵行为与生态水文指标响应关系研究 ——以长江四大家鱼为例

班学君^{1,2}, 樊博³, 刘瀚¹, 俞立雄⁴, 林俊强¹, 夏继红², 张迪¹

(1. 中国水利水电科学研究院, 北京 100038;

2. 河海大学农业与科学工程学院, 江苏 南京 211100;

3. 水利部科技推广中心, 北京 100038;

4. 中国水产科学研究院长江水产研究所, 湖北 武汉 430223)

摘要: 四大家鱼作为我国重要淡水经济鱼类, 近年来产卵规模大幅衰减, 水文条件作为影响家鱼自然繁殖的重要因素, 确定四大家鱼产卵行为与生态水文指标的关系对保护四大家鱼早期资源具有重要意义。以长江中游宜昌段为研究区域, 采用ROC曲线、分类回归树、Logistic回归等方法将生态水文指标对家鱼产卵刺激划分为“无效、低效、高效”多分类问题, 分析家鱼产卵行为与生态水文指标间的响应关系, 并对比研究不同方法间的性能与优缺点。结果显示, ROC曲线判别准确率、预测准确率、模型鲁棒性均最优, 是判定生态水文指标对四大家鱼产卵规模影响的最优方法; 当流量日增量大于 $1\ 370\ \text{m}^3/\text{s}$ 且两次洪峰间隔时间大于7.5 d时可以引起四大家鱼高效产卵响应。辨识出高效刺激家鱼产卵的生态水文指标及其阈值, 对生态调度的优化改进具有重要的现实意义。

关键词: 四大家鱼; 生态水文指标; ROC曲线; 生态调度

中图分类号: Q958.12 **文献标志码:** A **文章编号:** 1674-3075(2024)01-0067-08

水文情势改变是引起河流生态系统改变的主要驱动力(Richter et al., 2003), 水利水电工程建设运行不可避免地改变河流水文情势, 鱼类繁殖季节洪峰过程坦化将减弱产漂流性卵鱼类繁殖所需的涨水水流刺激, 进而影响鱼类自然繁殖。从众多生态水文指标中识别出关键指标, 建立鱼类产卵行为与水文指标响应关系, 并量化相应指标阈值, 是开展生态调度人为营造洪水过程的重要基础。目前国内外学者应用系统重构分析法、冗余分析法、遗传规划法分析了四大家鱼、铜鱼等产漂流性卵鱼类产卵行为与生态水文指标的相关性, 筛选了一些代表性水文指标并量化了相关阈值作为水库生态调度的参考依据(周雪等, 2019; 班璇等, 2019; 陈诚等, 2020)。已有研究多关注生态水文指标是否可以有效刺激鱼类产卵, 即把生态水文指标与鱼类行为的响应问题概化为“是、否”刺激产卵的

“二分类”问题。随着水库生态调度实践的深入推进, 在有效刺激鱼类产卵的基础上, 如能通过水文指标与鱼类行为的响应研究, 进一步辨识出低效、高效刺激鱼类产卵的生态水文指标及其阈值, 对生态调度的优化改进具有重要的现实意义。

本文以四大家鱼为研究对象, 选择受三峡水库生态调度影响最为显著的四大家鱼宜昌产卵场江段为研究区域, 基于2013–2023年(除2016、2021年)的水文和鱼类繁殖活动监测数据, 把生态水文指标“是、否”可以刺激鱼类产卵的“二分类”问题延拓为“无效、低效、高效”刺激鱼类产卵的“多分类”问题, 应用ROC法、分类回归树、logistic回归等不同数据挖掘方法, 建立鱼类产卵与关键水文指标的响应关系模型, 并对比研究了不同方法间的性能与优缺点, 期为三峡水库生态调度方案的优化提供依据。

1 材料与方法

1.1 数据收集

本研究使用的鱼类早期资源数据源于2013–2023年(2016、2021年未监测)中国水产科学研究院长江水产研究所与中国水利水电科学研究院在四大家鱼宜昌产卵场开展的鱼类繁殖期联合生态监测(周雪等, 2021)。研究使用的水文数据源于宜昌水文站逐日流量和水位数据。

收稿日期: 2023-10-31

基金项目: 国家自然科学基金联合基金项目(U2240214); 国家重点研发计划(2021YFC3200304)。

作者简介: 班学君, 1998年生, 男, 硕士研究生, 主要从事生态水文学研究。E-mail: 1021188239@qq.com

通信作者: 林俊强, 1984年生, 男, 正高级工程师, 主要从事生态水文学研究。E-mail: junquang-lin@hotmail.com

1.2 水文及生物指标

以往研究表明,四大家鱼产卵行为发生在产卵场水温达到 18°C 以上(Zhang et al,2000),且亲鱼感受到水文信号的刺激时。由于本次研究选取的5-7月长江宜昌段水温均可满足 18°C 以上这一要求,故不考虑水温对于产卵的影响。据统计,四大家鱼产卵大多数发生在涨水过程中,仅少数在涨水期受到涨水信号刺激后于高流量维持期完成产卵。结合前人研究经验,本次研究选取涨水持续天数、初始流量、洪峰流量、流量日增长量、流量总增长量、初始水位、洪峰水位、水位日上涨率、前后两次洪峰过程间隔时间、前后两个洪峰过程的水位差异等10个生态水文指标描述涨水过程(Zhang et al,2000;彭期冬,2011;周雪等,2019)。利用这些生态水文指标可有效描述一个涨水过程的强度、持续时间、强度变化率,用以探究其对四大家鱼产卵的影响。相关研究表明,四大家鱼产卵在不同流速下多发生在涨水开始的0.5~2 d后(李建等,2013),因此本研究中将连续涨水时间大于2 d的涨水过程认定为有效涨水过程。

选择鱼类早期资源监测得到的四大家鱼日均产卵量作为生物指标以反映产卵规模,依据每次涨水过程四大家鱼产卵量占全年产卵量的百分比,将四大家鱼产卵规模划分为无效产卵、低效产卵、高效产卵。其中,将产卵占比低于全年5%的产卵行为界定为无效产卵(刘瀚等,2023),在2013-2023年(除2016、2021年)66次有效涨水过程中,有33次涨水过程刺激了四大家鱼有效产卵,有33次未引起有效产卵。将引起四大家鱼有效产卵的33次涨水过程中四大家鱼产卵占比的中位数19.59%作为划分低效产卵和高效产卵的阈值,一次产卵事件产卵量占比小于等于全年产卵量19.59%且高于5%的认为是低效产卵,高于19.59%的认为是高效产卵。根据该划分标准,共有17次涨水过程被界定为可以引起四大家鱼低效产卵,16次涨水过程被界定为可以引起四大家鱼高效产卵。

1.3 建立鱼类产卵与生态水文指标响应关系模型

本文选择了在数据挖掘领域具有代表性的ROC曲线、分类回归树和logistic回归方法,将生态水文指标作为自变量,四大家鱼产卵规模作为因变量,研究生态水文指标与四大家鱼产卵行为的响应关系。

1.3.1 ROC曲线 受试者工作特征曲线(receiver operator characteristic curve, ROC曲线)是一种用来评价分类器或者诊断系统性能的图形工具,它可以显示在不同的阈值下,分类器或者诊断系统的真阳性

率(灵敏度)和假阳性率(1-特异度)之间的关系。ROC曲线最初应用于雷达性能评价,后广泛应用于机器学习、医疗诊断等领域(Schmugge et al,2007; Obuchowski & Bullen,2018; Wu et al,2018)。ROC曲线对类别的改变不敏感,当样本的正负比例发生改变时,其曲线形状几乎保持不变,这使得其可以更加客观地评价分类效果的好坏。ROC曲线是1种主要用于评价二分类问题模型的工具,当需要绘制多分类问题的ROC曲线时,常用的方法是将多分类问题转化为多个二分类问题(李亭亭等,2022),绘制多个ROC曲线,再将这些曲线整合分析,综合评价多分类问题模型。

本文中,涨水过程对四大家鱼产卵刺激的“无效、低效、高效”是一个三分类问题,因此,需要将这一问题划分为“无效”与“低效或高效”对比、“无效或低效”与“高效”对比两个二分类问题,以选定的生态水文指标作为检验变量,以产卵规模“无效”“低效或高效”产卵和“无效或低效”“高效”产卵分别作为状态变量绘制两组ROC曲线。在每组ROC曲线中,选择分类效果最好的指标用于阈值的确定。最终整合两组ROC曲线得出的最优阈值,确定该三分类问题最终的阈值及分类效果。绘制ROC曲线时,统计每一检验变量不同阈值下的真阳性率、假阳性率,作为ROC曲线的纵坐标、横坐标,最佳阈值的确定采用“尤登指数”,即“敏感性+特异性-1”,该指数值的最大值处就是最佳的阈值(Raghavan et al,2016)。

对于每组ROC曲线,由于选取了多个生态水文指标作为检验变量,因此可以绘制出多条ROC曲线。综合比较ROC曲线下面积(AUC)和判别准确率,选取1个分类效果最佳的曲线用于最终阈值的确定,其中AUC的值越接近1,表示模型的效果越好。

1.3.2 分类回归树法 分类回归树(Classification and regression trees analysis, CART)是一种无需假设数据分布的非参统计方法,它通过二元递归分离法,根据一定的划分准则,将数据集不断地划分为两个子集,使得每次划分后的两个子集之间的差异最大(黎明政等,2019)。分类回归树的构造过程包括树的生成、树的终止、树的剪枝和最优树的选择4个步骤。分类回归树算法包括分类树和回归树,其中分类树和回归树的特征选择采用的是不同的方法,分类树使用基尼系数,回归树使用均方差。本次研究中以选定的生态水文指标作为预测变量,将一个涨水过程内的四大家鱼产卵规模无效、低效、高效作为响应变量,分别定义为“无、低、高”,构建分类树。

进行CART计算时,树的最大深度max_depth设置为2,以防止树过拟合。树的分类权重class_weight设置为“balanced”,也就是在每次分裂时考虑样本的类别分布,给少数类更高的权重,这样可以处理不平衡数据集的问题。每个叶子节点最少样本min_samples_leaf设置为4,这样可以防止树过于细分数据,造成过拟合。树的分裂标准criterion选择“gini”,即采用基尼系数进行特征选择,基尼系数计算公式如下:

$$\text{Gini}(D) = 1 - \sum_{k=1}^K \left(\frac{|C_k|}{|D|} \right)^2 \quad (1)$$

式中: D 为样本集合, K 为类别个数, C_k 为第 k 个类别的样本子集, $\frac{|C_k|}{|D|}$ 为第 k 个类别的样本占比。基尼系数越小,表示样本集合的纯度越高。

1.3.3 Logistic 回归 逻辑回归(Logistic 回归)是用于解决二分类问题的机器学习方法,它可以估计某个事件发生的可能性,也可以根据一个阈值来划分不同的类别。逻辑回归的基本思想是将线性回归的结果通过S形的函数(sigmoid函数或逻辑函数)映射到0~1,然后用最大似然估计来求解模型的参数。逻辑回归虽然名字中有“回归”,但实际上是一种分类方法,它可以用于二分类或多分类问题(王济川和郭志刚,2001)。

对本文进行的“无效”“低效”“高效”三分类问题,常用的方法是有序回归。有序多分类的Logistic回归原理是将因变量的多个分类依次分割为多个二元的Logistic回归。但直接使用多个二元的Logistic回归进行分析相较于使用有序回归,具有实现简单,易于理解的优势,因此本文将涨水过程对鱼类产卵无效、低效、高效刺激的三分类问题转化为“无效”与“低效或高效”对比、“无效或低效”与“高效”对比两个二分类问题。

以所选的生态水文指标为自变量,产卵规模为因变量,分别利用逻辑回归对其进行计算,最后将两次分类结果整合,得出最终的结果。对于这两个二分类问题,二项分布成功的概率 p 为:

$$\text{logit}(p) = \ln\left(\frac{p}{1-p}\right) = \beta_1 \times X_1 + \beta_2 \times X_2 + \dots + \beta_n \times X_n + \beta_0 \quad (2)$$

式中: p 为二项分布试验成功的概率, X 为协变量, β 为回归系数。

利用尤登值确定 p 的阈值 $p_{\text{threshold}}$,当 p 大于

$p_{\text{threshold}}$ 时,表明二项分布成功。例如在“无效”与“低效或高效”对比这个二分类问题中,当计算出的某次涨水过程的 $p > p_{\text{threshold}}$,说明该涨水过程对四大家鱼产卵产生了有效刺激,反之为无效刺激。

1.4 训练集、测试集划分

为验证模型性能,提供可靠的评估结果,将数据集按照7:3的比例划分为训练集、测试集。利用训练集数据对模型进行训练,利用测试集数据对模型进行验证。

2 结果与分析

2.1 指标独立性

指标独立性分析结果(图1)显示,初始水位与初始流量、洪峰水位与洪峰流量、水位日上涨率与流量日上涨率间具有高度线性关系,将其剔除。将筛选后的涨水过程涨水持续时间、初始流量、洪峰流量、流量日增长量、流量总增长量、两次洪峰间隔时间、两次洪峰水位差异等7项生态水文指标用于后续分析。

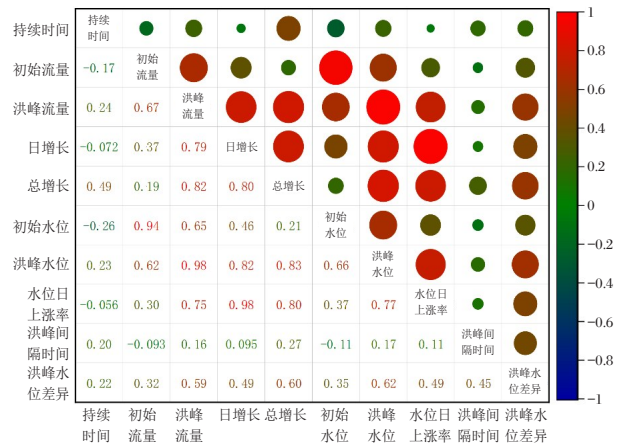


图1 指标独立性分析

Fig.1 Results of indicator independent analysis

2.2 识别结果

2.2.1 ROC 曲线 以指标独立分析筛选后的7个生态水文指标为检验变量,分别以产卵规模“无效”“低效或高效”以及“无效或低效”“高效”为状态变量,绘制ROC曲线,如图2所示。描述各指标分类效果的曲线下面积AUC、渐进显著性 P ,以及各指标的最佳判别阈值和该阈值下的判别准确率列入表1中。

结果显示,对于“无效”与“低效或高效”对比问题,涨水持续时间和初始流量的AUC值较小,分别为0.524、0.587,且对分类结果影响不显著($P > 0.05$),说明这两个指标的分类效果较差。洪峰流量、流量日增

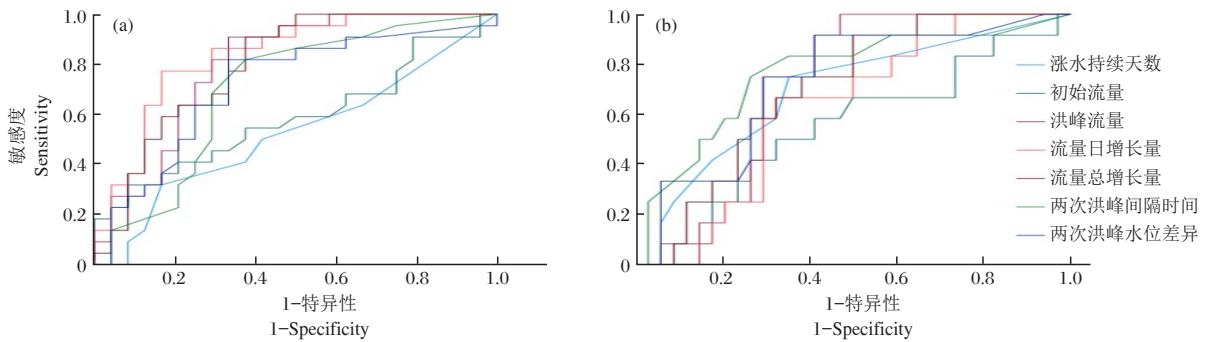


图 2 针对“无效”与“低效或高效”对比问题(a)和针对“无效或低效”与“高效”对比问题(b)绘制 ROC 曲线
Fig.2 ROC curves for the problem of “ineffective vs low effective + high effective” in (a)
and “ ineffective + low effective vs high effective” in (b)

表 1 针对“无效”与“低效或高效”对比问题(a)和针对“无效或低效”与“高效”对比问题(b)
各生态水文指标最佳阈值及模型性能参数

Tab.1 Optimal thresholds and model performance parameters of eco-hydrological indicators for the problem
of (a) “ineffective vs low effective + high effective” and (b) “ ineffective + low effective vs high effective”

检验变量	AUC		渐近显著性 P		最佳判别阈值		判别准确率	
	(a)	(b)	(a)	(b)	(a)	(b)	(a)	(b)
涨水持续时间/d	0.524	0.692	0.783	0.05*	6.5	4.5	58.70%	67.39%
初始流量/ $\text{m}^3 \cdot \text{s}^{-1}$	0.587	0.55	0.312	0.608	16470	14470	63.04%	63.04%
洪峰流量/ $\text{m}^3 \cdot \text{s}^{-1}$	0.799	0.708	0.001*	0.033*	17600	17600	78.26%	65.22%
流量日增长量/ $\text{m}^3 \cdot \text{s}^{-1}$	0.835	0.632	<0.001*	0.177	1370	1463.75	80.43%	67.39%
流量总增长量/ $\text{m}^3 \cdot \text{s}^{-1}$	0.797	0.701	0.001*	0.04*	2810	3325	76.09%	60.87%
两次洪峰间隔时间/d	0.701	0.761	0.02*	0.008*	5.5	7.5	71.74%	73.91%
两次洪峰水位差异/m	0.717	0.727	0.012*	0.021*	0.46	0.5	73.91%	67.39%

注: *指渐进显著性小于等于 0.05。

Note: * denotes the progressive significance less than or equal to 0.05.

长量、流量总增长量、两次洪峰间隔时间、两次洪峰水位差异的 AUC 值均大于 0.7, 表明模型的区分效果较好, 综合考虑各指标的 AUC 值和最佳阈值下的判别准确率, 选择流量日增长量作为判别四大家鱼产卵规模是否有效的指标, 最佳判别阈值为 $1370 \text{ m}^3/\text{s}$, 即当 1 次涨水过程的流量日增长量大于 $1370 \text{ m}^3/\text{s}$ 时, 认为该次涨水过程可以刺激四大家鱼有效产卵。

对于“无效”与“低效或高效”对比问题, 涨水持续时间、初始流量、流量日增量作为分类指标的 AUC 值均较小 (<0.7), 且渐进显著性均大于等于 0.05, 表明以这些指标进行分类的模型区分效果较差, 分类效果不显著。洪峰流量、流量总增长量、两次洪峰间隔时间、两次洪峰水位差异作为分类指标的 AUC 均大于 0.7 ($P < 0.05$), 表明模型的区分效果较好, 综合考虑 AUC 值和判别准确率, 选择以两次洪峰间隔时间作为判别四大家鱼产卵规模是否高效的指标, 最

佳判别阈值为 7.5 d, 即当 1 次涨水过程距上次涨水过程的洪峰间隔时间大于 7.5 d 时, 认为其可以刺激四大家鱼高效产卵。

结合上述两个分类标准, 当一次涨水过程的流量日增量小于 $1370 \text{ m}^3/\text{s}$ 时, 认为此次涨水过程不能刺激四大家鱼进行有效产卵; 当流量日增量大于 $1370 \text{ m}^3/\text{s}$ 时, 认为此次涨水过程可以刺激四大家鱼进行有效产卵; 当流量日增量大于 $1370 \text{ m}^3/\text{s}$ 且两次洪峰间隔时间大于 7.5 d 时, 认为此次涨水过程可以刺激四大家鱼进行高效产卵。在训练集 46 次涨水过程中, 模型最终判别正确 33 场次, 模型判别准确率达 71.74%。

2.2.2 分类回归树 以筛选后的生态水文指标作为预测变量, 产卵规模作为响应变量, 进行分类回归树分析。CART 分析结果显示, 流量总增长量和两次洪峰间隔时间是影响四大家鱼产卵规模的重要指标。分类树显示(图 3), 当涨水过程的流量总增长量小于

等于 $2\ 145\ \text{m}^3/\text{s}$ 时,该涨水过程对四大家鱼产卵行为刺激效果为“无效”;当流量总增长量大于 $2\ 145\ \text{m}^3/\text{s}$,且两次洪峰间隔时间大于 $11.5\ \text{d}$ 时,该涨水过程对四大家鱼产卵行为刺激效果为“高效”;当流量总增长量大于 $2\ 145\ \text{m}^3/\text{s}$,且两次洪峰间隔时间小于 $11.5\ \text{d}$ 时,该涨水过程对四大家鱼产卵行为刺激效果为“低效”。在训练集所有的 46 次涨水过程中,经分类回归树模型判别,有 12 场次涨水对四大家鱼产卵行为刺激效果为“无效”,27 场次涨水对四大家鱼产卵行为刺激效果为“低效”,7 场次涨水对四大家鱼产卵行为刺激效果为“高效”,判别正确 28 场次,模型的判别准确率为 60.87%。涨水持续天数、初始流量、洪峰流量、流量总增长量、两次洪峰的水位差异对四大家鱼的产卵规模不产生影响。

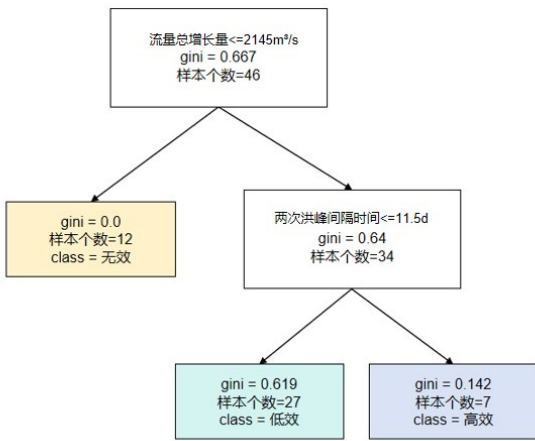


图3 四大家鱼产卵规模与生态水文指标间关系的分类树分析

Fig.3 Classification tree analysis of the relationship between spawning scale of the four major Chinese carps and eco-hydrological indicators

2.2.3 Logistic 回归 以筛选后的 7 个生态水文指标为协变量,分别以产卵规模“无效”“低效或高效”以及“无效或低效”“高效”作为状态变量,利用 Logistic 回归对其进行分析,得出经过筛选后进入模型的变量及其参数,如表 2 所示。

表 2 进入 Logistic 回归模型中指标及参数

Tab.2 Indicators and parameters used in the logistic regression model

状态	进入模型指标	回归系数 B	显著性	Exp(B)
“无效”与“低效或高效”对比	流量日增长量/ $\text{m}^3 \cdot \text{s}^{-1}$	0.001	0.003	1.001
	常量	-1.778	0.004	0.169
“无效或低效”与“高效”对比	涨水持续天数/d	0.318	0.059	1.374
	两次洪峰间隔时间/d	0.179	0.018	1.196
	常量	-4.16	0.002	0.016

对于“无效”与“低效或高效”对比问题,分析结果显示,涨水过程的流量日增长量对该涨水过程是否可以刺激四大家鱼有效产卵产生影响。涨水过程可以刺激四大家鱼进行有效产卵的概率 p_1 可通过下式计算得出:

$$\text{logit}(p_1) = 0.001 \times dQ - 1.778 \quad (3)$$

$$p_1 = \frac{e^{0.001 \times dQ - 1.778}}{1 + e^{0.001 \times dQ - 1.778}} \quad (4)$$

式中: p_1 为涨水过程可以刺激四大家鱼进行有效产卵的概率, dQ 为涨水过程的流量日增长量。

对于“无效或低效”与“高效”对比问题,结果显示,涨水过程的涨水持续天数和两次洪峰间隔时间是刺激四大家鱼高效产卵的重要生态水文指标。涨水过程可以刺激四大家鱼进行高效产卵的概率 p_2 可通过下式计算得出:

$$\text{logit}(p_2) = 0.318 \times t_1 + 0.179 \times t_2 - 4.16 \quad (5)$$

$$p_2 = \frac{e^{0.318 \times t_1 + 0.179 \times t_2 - 4.16}}{1 + e^{0.318 \times t_1 + 0.179 \times t_2 - 4.16}} \quad (6)$$

式中: p_2 为涨水过程可以刺激四大家鱼进行高效产卵的概率, t_1 为涨水持续时间(d), t_2 为该涨水过程距上一涨水过程的间隔时间(d)。

利用尤登值最大原则确定 p_1 、 p_2 的最佳阈值, p_1 的最佳阈值为 45.10%, p_2 的最佳阈值为 12.53%。 p_1 最佳阈值对应的流量日增长量为 $1\ 366\ \text{m}^3/\text{s}$, p_2 最佳阈值对应的涨水持续时间为 5 d、两次洪峰间隔时间为 3.5 d。即当流量日增长量大于 $1\ 366\ \text{m}^3/\text{s}$ 时,可引起四大家鱼有效产卵;当流量日增长量大于 $1\ 366\ \text{m}^3/\text{s}$ 、涨水持续时间大于 5 d 且两次洪峰间隔时间大于 3.5 d 时,可引起四大家鱼高效产卵。以此作为判别标准,在训练集 46 次涨水过程中,模型判别正确 33 场次,准确率达 71.74%。

2.3 比较分析

2.3.1 模型判别效果比较 利用 ROC 曲线、Logistic 回归判别某一涨水过程是否可以刺激四大家鱼有效产卵判别准确率均达 71.74%。利用分类回归树进行判别时,判别准确率达 60.87%。

2.3.2 模型预测效果验证 利用测试集数据对上文所建模型进行验证,在对测试集 20 场次涨水过程刺激四大家鱼产卵效果预测中,ROC 曲线、Logistic 回归均预测正确 15 次,预测准确率达 75%,分类回归树预测正确 11 次,预测准确率达 55%。

2.3.3 模型稳定性分析 为验证各模型的鲁棒性,对总体数据集按 7:3 的比例随机划分训练集和测试集,共划分 10 组。分别对各组数据利用上述 3 种方法进行计算,比较不同训练集、测试集划分下,各种方法的判别准确率、预测准确率。结果显示,利用 ROC 曲线进行分类的判别准确率、预测准确率均为最高水平,认为 ROC 曲线是鲁棒性较好的模型(表 3)。

表 3 3 种模型鲁棒性分析 %

Tab.3 Robust stability analysis of the three models

组别	ROC		分类回归树		Logistic 回归	
	判别准确率	预测准确率	判别准确率	预测准确率	判别准确率	预测准确率
1	69.57	75	65.21	55	67.39	70
2	67.39	70	67.39	60	67.39	75
3	67.39	70	69.57	60	67.39	75
4	71.74	65	80.43	65	69.57	65
5	67.39	65	63.04	55	65.22	75
6	71.74	70	65.22	55	67.39	65
7	73.91	60	71.74	60	67.39	65
8	69.57	75	65.22	55	67.39	55
9	69.57	75	71.74	60	67.39	60
10	73.91	65	71.74	60	73.91	45
均值	70.22	69	69.13	58.5	68.043	65

2.3.4 综合比较分析 综合比较 ROC 曲线、分类回归树、Logistic 回归应用于识别影响四大家鱼产卵规模的关键生态水文指标效果,对比分析各方法下的判别准确率、预测准确率、模型鲁棒性。结果显示,ROC 曲线判别准确率、预测准确率、模型稳定性在 3 种方法中均处于最高水平,认为 ROC 曲线是本次研究中判定生态水文指标对四大家鱼产卵规模影响的最优方法。

3 讨论

3.1 影响四大家鱼产卵的生态水文指标

四大家鱼产卵规模与其产卵环境的生态水文指标密切相关,已有研究表明,涨水持续时间、流量日增长率、水位日增长率、两次洪峰间隔时间等指标可以对四大家鱼产卵规模产生显著影响(徐薇等,2020)。本研究中,将连续涨水设为先决条件,分析描述涨水过程的生态水文指标对四大家鱼产卵规模的影响,结果显示,长江中游四大家鱼产卵规模受到流量日增长量、两次洪峰间隔时间影响,与已有研究结果(徐薇等,2020)一致。

流量日增长量影响四大家鱼能否进行有效产卵,流速加大的过程会刺激成熟的亲鱼,促进亲鱼产

卵排精(陈永柏等,2009),较大流量增长率会加快水流流速增长过程,进而促进四大家鱼的产卵行为。两次洪峰间隔时间影响四大家鱼能否进行高效产卵,前后两次洪峰间隔时间越长,产卵场内等待水流信号刺激的成熟亲鱼越多,当下一次洪峰过程到来,这些亲鱼感受到水流信号刺激完成大规模产卵。此外,有研究(李博等,2021)表明,刺激四大家鱼产卵的洪峰过程断面初始流量需达到某一定值。但在本次研究中,存在一些初始流量较小的可以刺激四大家鱼有效产卵的涨水场次,原因在于这些涨水场次的流量日增长量和涨水持续时间较大,初始流量虽然较小,但流量在涨水过程内可以达到较高水平,从而导致最终结果中初始流量与四大家鱼产卵规模相关性不大。

3.2 鱼类产卵有效性多分类问题计算方法

对于鱼类产卵有效性的分类问题常用计算方法有 ROC 曲线法(刘瀚等,2023)、分类回归树(黎明政等,2019)、支持向量机、Logistic 回归(崔雪森等,2016)、广义加性模型(胡丽娟等,2021;孙浩奇等,2022)等方法。支持向量机是一种黑箱模型,计算复杂度较高,利用支持向量机只能对鱼类产卵规模进行分类,不能得出分类阈值。广义加性模型变量间的关系是非参数形式,无法通过简单的数学公式来描述,因此,观测变量间依赖性关系的常用方式是采用图形化的平滑曲线。但对平滑参数的选择和平滑曲线的解读需结合实际情况进行,否则容易产生背离实际的曲解。因此,本次研究最终选择了方法简单易用,结果直观明了的 ROC 曲线法、分类回归树、Logistic 回归用于分析计算。

本次研究面向的是产卵规模无效、低效、高效的三分类问题,而 ROC 曲线、Logistic 回归常用于二分类问题。当 ROC 曲线、Logistic 回归用于多分类问题时,可以将多分类问题转化为多个二分类问题,如本研究关注的三分类问题可以转化为“无效”与“低效或高效”对比、“无效或低效”与“高效”对比两个二分类问题,分别确定分类的阈值,最终将两个二分类的结果整合起来,得出分类结果。经上文验证,使用这种方法进行的类的判别准确率、预测准确率均保持在较高水平。

3.3 面向鱼类产卵需求的生态流量泄放方法

为满足四大家鱼产卵的生态水文需求,需要充分发挥上游水库的生态调控能力,调整水库出库流量(Baumgartner et al,2020)。三峡水库自 2011 年以来,已连续 13 年开展生态调度,但并非每次生态调度过程都能诱发四大家鱼大规模产卵(李朝达等,

2021)。为对四大家鱼产卵的有效刺激,需要依据其产卵需求对生态调度方式不断优化。已有研究表明,当水文指标满足较高的水位上涨率和流量日增长率,较长的涨水时间,前后两次洪峰过程间隔时间较长这些条件时,会诱发大规模产卵(Zhang et al, 2000)。徐薇等(2020)基于长江中游沙市段四大家鱼产卵数据及水文数据开展研究,得出流量日增长量平均大于 $2\ 000\ \text{m}^3/\text{s}$,与前一次洪峰的间隔时间在 $5\ \text{d}$ 可以增加四大家鱼产卵量。黎明政等(2019)认为三峡库区忠县至涪陵江段四大家鱼繁殖主要发生在流量日增量大于 $840\ \text{m}^3/\text{s}$ 。上述研究结果与本文研究结果相仿。本次研究将洪水过程对于四大家鱼产卵刺激划分为无效、低效、高效,进一步量化了相关生态水文指标,即流量日增长量大于 $1\ 370\ \text{m}^3/\text{s}$ 且两次洪峰间隔时间大于 $7.5\ \text{d}$ 的涨水过程可以刺激四大家鱼高效、大规模产卵,为后续生态调度方案制定提供了依据。

参考文献

班璇, Panayiotis D, 吕晓蓉, 等, 2019. 长江葛洲坝下游鱼类资源量的关键水文指标识别[J]. 水利水电科技进展, 39(1): 15-20.

陈诚, 黎明政, 高欣, 等, 2020. 长江中游宜昌江段鱼类早期资源现状及水文影响条件[J]. 水生生物学报, 44(5): 1055-1063.

陈永柏, 廖文根, 彭期冬, 等, 2009. 四大家鱼产卵水文水动力特性研究综述[J]. 水生态学杂志, 30(2): 130-133.

崔雪森, 伍玉梅, 周爱忠, 等, 2016. 基于 Logistic 回归模型的非沿海欧洲沙丁鱼渔场与环境因素关系模型的构建[J]. 大连海洋大学学报, 31(2): 211-218.

胡丽娟, 宋超, 耿智, 等, 2021. 长江口凤鲚仔稚鱼主要繁殖季节的时空分布特征[J]. 中国水产科学, 28(9): 1152-1161.

李博, 邵星晨, 黄涛, 等, 2021. 三峡水库生态调度对长江中游宜昌江段四大家鱼自然繁殖影响分析[J]. 长江流域资源与环境, 30(12): 2873-2882.

李朝达, 林俊强, 夏继红, 等, 2021. 三峡水库运行以来四大家鱼产卵的生态水文响应变化[J]. 水利水电技术(中英文), 52(5): 158-166.

李建, 夏自强, 戴会超, 等, 2013. 三峡初期蓄水对典型鱼类栖息地适宜性的影响[J]. 水利学报, 44(8): 892-900.

黎明政, 马琴, 陈林, 等, 2019. 三峡水库产漂流性卵鱼类繁殖现状及水文需求研究[J]. 水生生物学报, 43(S1): 84-96.

李亭亭, 吕王勇, 周娇, 等, 2022. 基于类间距离的 Logistic 序贯多分类的成都空气质量污染分析[J]. 科学技术与工程, 22(1): 409-415.

刘瀚, 林俊强, 秦鑫, 等, 2023. 基于生物水文响应模型的刺激四

大家鱼繁殖生态调度目标量化方法研究——以三峡水库为例[J]. 湖泊科学, 35(5): 1741-1751.

彭期冬, 2011. 三峡工程对四大家鱼自然繁殖条件影响研究[D]. 北京: 中国水利水电科学研究院.

孙浩奇, 蒋日进, 陈峰, 等, 2022. 瓯江口凤鲚鱼卵、仔稚鱼的时空分布及其与环境因子关系[J]. 水生生物学报, 46(7): 963-973.

王济川, 郭志刚, 2001. Logistic 回归模型[M]. 北京: 高等教育出版社.

徐薇, 杨志, 陈小娟, 等, 2020. 三峡水库生态调度试验对四大家鱼产卵的影响分析[J]. 环境科学研究, 33(5): 1129-1139.

周雪, 王珂, 陈大庆, 等, 2019. 三峡水库生态调度对长江监利江段四大家鱼早期资源的影响[J]. 水产学报, 43(8): 1781-1789.

周雪, 汪登强, 段辛斌, 等, 2021. 长江中游宜昌江段鱼卵时空分布特征研究[J]. 长江流域资源与环境, 30(4): 861-868.

Baumgartner M T, Pitégoras A P, Baumgartner G, et al, 2020. Storage or Run-of-river Reservoirs: Exploring the Ecological Effects of Dam Operation on Stability and Species Interactions of Fish Assemblages[J]. Environmental Management, 65: 220-231.

Obuchowski N A, Bullen J A, 2018. Receiver operating characteristic (ROC) curves: review of methods with applications in diagnostic medicine[J]. Physics in Medicine & Biology, 63(7): 07TR01.

Raghavan R, Ashour F S, Bailey R, 2016. A Review of Cutoffs for Nutritional Biomarkers[J]. Advances in Nutrition, 7(1): 112-120.

Richter B D, Mathews R, Wigington H R, 2003. Ecologically sustainable water management: Managing river flows for ecological integrity[J]. Ecological Applications, 13(1): 206-224.

Schmugge S J, Jayaram S, Shin M C, et al, 2007. Objective evaluation of approaches of skin detection using ROC analysis[J]. Computer Vision and Image Understanding, 108(1): 41-51.

Wu W, Li A D, He X H, et al, 2018. A comparison of support vector machines, artificial neural network and classification tree for identifying soil texture classes in southwest China[J]. Computers & Electronics in Agriculture, 144: 86-93.

Zhang G, Chang J, Shu G, 2000. Applications of factor-criteria system reconstruction analysis in the reproduction research on grass carp, black carp, silver carp and bighead in the Yangtze River[J]. International Journal of general systems, 29(3): 419-428.

Relationship between Fish Spawning Behavior and Eco-hydrological Indicators: A Case Study of the Four Major Chinese Carps in the Yangtze River

BAN Xue-jun^{1,2}, FAN Bo³, LIU Han¹, YU Li-xiong⁴, LIN Jun-qiang¹, XIA Ji-hong², ZHANG Di¹

(1. State Key Laboratory of Simulation and Regulation of Water Cycle in River Basin,
China Institute of Water Resources and Hydropower Research, Beijing 100038, P.R. China;

2. College of Agricultural Science and Engineering, Hohai University, Nanjing 211100, P.R. China;

3. Science and Technology Promotion Center of the Ministry of Water Resources,
Beijing 100038, P.R. China;

4. Yangtze River Fisheries Research Institute, Chinese Academy of Fishery Sciences,
Wuhan 430223, P.R. China)

Abstract: As important economic freshwater fish in China, the spawning scale of the four major Chinese carps has declined significantly in recent years. Hydrological conditions play an important role in the natural reproduction of the four major Chinese carps. Thus, to protect early stage resources of these four species, it is important to determine the relationship between spawning behavior and eco-hydrological indicators. In this study, the Yichang section of the middle Yangtze River was selected for research, and we analyzed the relationship between the spawning behavior of the four major Chinese carps and eco-hydrological indicators using a receiver operating characteristic (ROC) curve, classification regression tree and logistic regression analysis. The performance, advantages and disadvantages of the three data mining methods were then compared. The study was based on monitoring data for hydrology and reproduction in the investigated section during the period 2013–2023 (except 2016, 2021). Stimulation of spawning behavior by ecohydrological indicators was analyzed for each of the four species and divided into three levels (no response, low response, and high response). The discrimination accuracy rates for each method were as follows: ROC curve (71.74%), logistic regression analysis (71.74%) and classification regression tree (60.87%). The prediction accuracy rates of the three methods were, respectively 75%, 75% and 55%, and the ROC curve was the most robust. Therefore, we recommend the ROC curve for estimating the impact of eco-hydrological indicators on the spawning scale of the four major Chinese carps. The best eco-hydrological conditions for effective spawning of the four major Chinese carps was when the daily increment of flow was greater than 1 370 m³/s and the interval time between two flood peaks was more than 7.5 days. In conclusion, our research identified the eco-hydrological indicators and thresholds that effectively stimulate spawning of the four major Chinese carps. This has important practical significance for optimizing and improving the ecological operation of Three Gorges Reservoir.

Key words: four major Chinese carps; eco-hydrological indicators; ROC curve; ecological operation