

张新, 翁晓姚. 多维度的水厂日供水量规律性评价方法[J]. 净水技术, 2026, 45(1): 63–72, 83.

ZHANG X, WENG X Y. Evaluation method of multi-dimensional regularity for daily water supply capacity of WTPs [J]. Water Purification Technology, 2026, 45(1): 63–72, 83.

多维度的水厂日供水量规律性评价方法

张 新, 翁晓姚*

(上海城投水务<集团>有限公司, 上海 200082)

摘 要 【目的】 在城市现代化建设的进程中, 供水系统的复杂特征和动态特征日益显现, 水厂的日供水量作为供水系统的重要指标且受到多种因素的影响, 可有效反映这些特征。【方法】 本文从时域、频域及复杂性 3 个角度, 提出了一种多维度量化评价水厂日供水量规律性和非规律性的方法, 推动日供水量预测模型选择、建立的透明化。规律性量化评价方面, 包含时域分析中的自相关系数绝对值的算术平均值 (MAAC)、变异性补 (1-CV) 以及基于季节与趋势分解 (STL) 法的趋势强度 (TS) 与季节强度 (SS)、趋势季节强度与噪声强度之比 (R_0) 4 项指标; 非规律性量化评价方法, 包含时域分析中基于 STL 法的噪声强度 (NS)、频域分析中基于功率谱的归一化谱熵 (H_{norm}) 以及复杂性分析中的归一化样本熵 (NSE) 和归一化赫斯特指数 ($H_{\text{E, norm}}$) 4 项指标。【结果】 进而, 以上海中心城区浦西区域的 10 家水厂日供水量为研究对象, 基于供水运行规律性分析和相关研究, 设置了自相关分析、STL 分解和样本熵的相关参数后, 分别计算规律性指标、非规律性指标, 然后通过算术平均值得到规律性总得分、非规律性总得分。评价结果显示, 本研究提出方法评价的 10 家水厂的供水规律性和不规律性得分可相互佐证。【结论】 结合评价结果, 选用广义自回归条件异方差 (GARCH) 模型和极端梯度提升 (XGBoost) 模型开展日水量预测与规律性分析; 针对高规律性、中规律性、低规律性和复杂性 4 类水厂, 推荐水量预测模型, 辅助提高了水量预测的可解释性。

关键词 供水系统 日供水量 规律性评价 时域分析 频域分析 复杂性分析 水量预测

中图分类号: TU991 文献标志码: A 文章编号: 1009-0177(2026)01-0063-11

DOI: 10.15890/j.cnki.jsjs.2026.01.008

Evaluation Method of Multi-Dimensional Regularity for Daily Water Supply Capacity of WTPs

ZHANG Xin, WENG Xiaoyao*

(Shanghai Chengtouw Water <Group> Co., Ltd., Shanghai 200082, China)

Abstract [Objective] In the process of urban modernization, the complexity and dynamic characteristics of water supply systems are becoming increasingly evident, and the daily water supply volume of water treatment plants (WTPs), as an important indicator of the water supply system, can effectively reflect the overall characteristics. [Methods] This paper proposed a multi-dimensional quantitative evaluation method from temporal, frequency, and complexity domains to assess both regularity and irregularities in daily water supply patterns of WTPs, while simultaneously enhancing the transparency in model selection and development for daily water supply prediction through systematic regularity quantification. In terms of regularity quantification, it included four indices in time domain analysis: the mean absolute value of autocorrelation coefficients (MAAC), the complement of variability (1-CV), and two indices based on seasonal-trend decomposition procedure using LOESS (STL) method: the trend strength and seasonality strength (TS+SS), and the ratio of trend seasonality strength to noise strength (R_0). For irregularity quantification, it included four indices: noise strength (NS) based on STL in time domain analysis, normalized spectral entropy (H_{norm}) based on power spectrum in frequency domain analysis, and two indices in complexity analysis: normalized sample entropy (NSE) and normalized Hurst exponent ($H_{\text{E, norm}}$). [Results] Subsequently, taking the daily water supply volumes of 10 WTPs in the Puxi area of Shanghai's central urban district as

[收稿日期] 2024-12-06

[作者简介] 张新 (1980—), 男, 高级工程师, 主要从事管网模型、机器学习模型与供排水系统运营管理工作, Email: zhangxin@shanghaiwater.com。

[通信作者] 翁晓姚 (1972—), 女, 正高级工程师, 主要从事供排水生产运营管理工作, E-mail: wengxiaoyao@shanghaiwater.com。

the research object, based on the analysis of water supply operation regularity and related research, the relevant parameters for autocorrelation analysis, STL decomposition, and sample entropy were set. Regularity and irregularity indices were calculated, and then the total scores for regularity and irregularity were obtained through arithmetic mean, conducting a quantification evaluation of the regularity and irregularity of daily water supply volumes. The evaluation results demonstrated that the regularity and irregularity scores of 10 WTPs assessed by the method proposed in this study for different water treatment plants can mutually corroborate each other.

[**Conclusion**] Based on these evaluation outcomes, GARCH and XGBoost models are subsequently employed to conduct daily water demand prediction and regularity analysis. For WTPs categorized into four types-high regularity, medium regularity, low regularity, and complexity-specific water demand prediction models are recommended. This method ological framework can effectively enhance the interpretability of water demand forecasting.

Keywords water supply system daily water supply capacity regularity evaluation time-domain analysis frequency-domain analysis complexity analysis water demand prediction

供水系统作为韧性城市建设的关键一环,近年来正面临着前所未有的挑战^[1]。在现代城市化进程中,供水系统的复杂性和动态特征日益显现。水厂的日供水量作为供水系统的重要指标,受到多种因素的影响;比如城市供水量受新冠疫情封控影响较大,下降时间和封控措施的实施同步,日均供水量的损失约为正常情况下的 7%^[2],其他因素还包括自然环境变化、社会经济发展水平、人口增长以及供水管理调度策略等^[3]。在供水量预测领域,近年研究^[4-6]主要聚焦提高预测精度为目标的混合模型和深度学习应用方面,但对水量预测模型落地应用时,面临的模型选择、透明度不高、可解释性不足等方面问题研究较少。

本研究通过综合时域、频域和复杂性分析方法,提出了一种针对水厂日供水量的规律性、非规律性的多维度分析评价方法。在对上海中心城区浦西区域内 10 家水厂多年的日供水量数据进行评价分析的基础上,选用广义自回归条件异方差(GARCH)模型和极端梯度提升(XGBoost)模型开展日水量预测与规律性分析;针对高规律性、中规律性、低规律性和复杂性水厂,推荐了水量预测模型。

本评价方法可作为水厂日供水量预测模型选择的前置分析工具,通过量化时间序列的规律性特征(趋势强度、周期稳定性等)和非规律性特征(随机波动、混沌特性等),辅助为不同特性的水厂匹配预测模型,从而提升预测效率,提高预测工作的可解释性,提升供水系统运营管理的科学化和精细化水平^[7]。

1 研究方法

水厂供应量作为典型的时间序列数据,可从时

域、频域 2 个角度开展分析。同时,水厂供水作为供水系统中的一个重要环节,其水厂出水量的时间序列数据包含了用水、运维、调度及异常处理等工况的综合信息,具备模式多样性的特征。因此,也可从复杂性角度分析。

本研究在水厂日供水量时域、频域和复杂性分析的基础上,提出了规律性、非规律性指标分析框架,如图 1 所示。

1.1 时域分析

1.1.1 统计分析

1) 变异性指标

变异系数(CV)是衡量时间序列相对变异程度的指标,可直接比较具有不同单位或量级的时间序列。在应用中,CV 的值越高,意味着时间序列的波动性或不确定性越大。相反,CV 值越低,则表明时间序列数据越稳定,变异较小。

本研究中应用 1-CV 作为变异性补,重点描述水厂出水量时间序列中的稳定性、确定性部分。

2) 自相关性指标

自相关分析是通过计算时间序列滞后值之间的自相关函数,量化时间序列的相关性,揭示其中的周期性、趋势性和随机性^[7]。为了量化时间序列整体自相关强度,本研究采用自相关系数绝对值的算术平均值(MAAC),在式中表示为 M_{ACC} ,如式(1)。

$$M_{ACC} = \frac{1}{K} \sum_{k=1}^K |\rho_k| \quad (1)$$

其中: ρ_k ——滞后 k 的自相关系数;

K ——自相关系数数量;

M_{ACC} ——自相关系数绝对值的算术平均值。

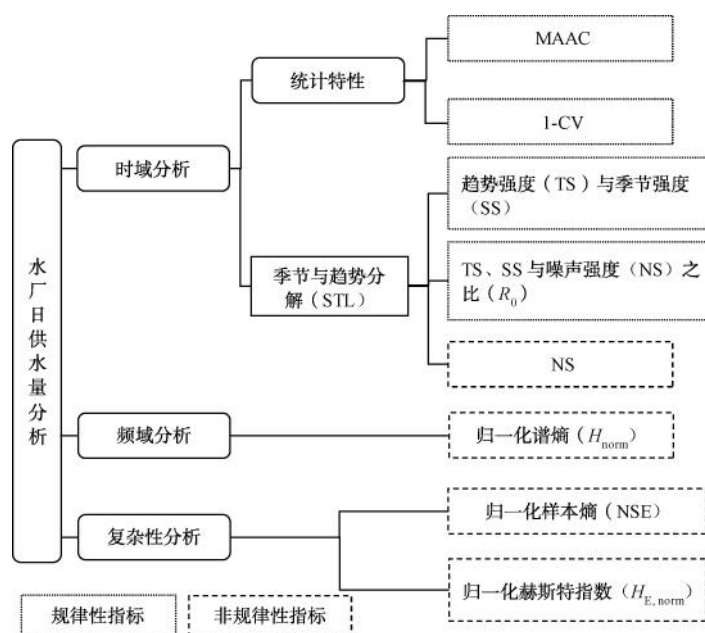


图1 规律性与非规律性分析框架

Fig. 1 Analysis Framework of Regularity and Irregularity

M_{AAC} 越高,表明时间序列中存在更显著的趋势或周期性成分。

1.1.2 STL 分解

STL 广泛用于非平稳时间序列的趋势和季节性模式分析^[8],其利用局部回归 (LOESS) 对时间序列逐步分解为 3 部分:趋势、季节性和残差,可揭示时间序列中的 TS 和 SS^[9]。

1) TS 与 SS 指标

基于 STL 分解将时间序列拆解为趋势、季节性成分和残差成分后,通过计算趋势分项、季节性成分占总方差的比例,可得出 TS、SS,在式中分别表示为 T_s 、 S_s 如式(2)~式(3)。

$$T_s = \frac{\text{Var}(T_t)}{\text{Var}(X_t)} \quad (2)$$

$$S_s = \frac{\text{Var}(S_t)}{\text{Var}(X_t)} \quad (3)$$

其中: T_s ——趋势强度;

S_s ——季节强度;

Var——方差;

T_t ——STL 分解 X_t 得到的趋势时间序列数据;

X_t ——总体时间序列数据;

S_t ——STL 分解 X_t 得到的季节时间序列

数据。

两者的数值在 $[0, 1]$, 值越高, 分别表明趋势成分、季节性成分的占比越大。趋势强度高的序列通常适合长期分析, 而季节性强度高的序列适合周期性预测。两者均体现了时间序列信号的规律性。

2) R_0

R_0 用于量化趋势和季节性等规律性部分对随机波动的影响, 如式(4)。

$$R_0 = \frac{\text{Var}(T_t + S_t)}{\text{Var}(R_t)} \quad (4)$$

其中: R_t ——STL 分解 X_t 得到的噪声时间序列数据。

R_0 值越高, 表明趋势和季节性等规律性部分是时间序列变化的主要驱动力; 反之, 随机波动发挥的作用就越大。

为了便于比较, 按式(5)将 R_0 进行归一化。

$$R_0 = \frac{R_0}{1 + R_0} \quad (5)$$

3) NS 指标

NS 是衡量随机波动成分对时间序列总变化贡献程度的重要指标, 如式(6)。

$$N_s = \frac{\text{Var}(R_t)}{\text{Var}(X_t)} \quad (6)$$

其中: N_s ——噪声强度。

N_s 值在 $[0, 1]$, 值越高, 表明时间序列的随机性越高, 规律性越差。

1.2 频域分析与 H_{norm}

频域分析利用傅里叶变换将时间序列从时间域转化为频域, 可揭示时间序列中周期性信号的频率分布。谱熵是根据香农原理和信息熵原理计算时间序列信号的频谱功率分布和规律性^[10], 如式(7)~式(8)。

$$H_{\text{norm}} = \frac{-\sum_i \hat{P}(f_i) \log[\hat{P}(f_i)]}{\log N} \quad (7)$$

$$\hat{P}(f_i) = \frac{P(f_i)}{\sum_j P(f_j)} \quad (8)$$

其中: $P(f_i)$ ——频率 f 对应的功率谱密度;

$\hat{P}(f_i)$ ——归一化功率谱密度;

N ——频率点数。

归一化后的谱熵值为 $[0, 1]$ 。 H_{norm} 越接近 0, 表明时间序列对应频谱中某些频率分量占主导地位, 此时间序列具有较强的周期性; H_{norm} 越接近 1 则表明频率分布均匀, 序列呈现随机性。

1.3 复杂性分析

1) 样本熵 (SE)

随着非线性理论的发展, 20 世纪 90 年代初, Pincus 首次提出用近似熵计算时间序列复杂度, 并在各领域得到广泛应用。有研究者^[11]对近似熵进行修正, 提出 SE, 旨在降低近似熵的误差, 是一种与近似熵类似但精度更好的方法。SE 可通过评估供水系统中相关时间序列中相似模式的出现概率, 量化时间序列的规则性和复杂性。SE 值越低, 时间序列自我相似性越高, 产生新模式的概率越低, 时间序列越简单; 反之, 时间序列自我相似性越低, 产生新模式的概率越高, 时间序列越复杂^[10]。

为了对不同长度或尺度的时间序列进行比较, 将 SE 进行归一化得到归一化样本熵 (NSE), 在式中表示为 N_{SE} , 定义如式(9)。

$$N_{\text{SE}} = \frac{S_E(m, r, N)}{\log N} \quad (9)$$

其中: $S_E(m, r, N)$ ——样本熵;

$\log N$ ——样本熵可能达到的最大值;

m ——嵌入维度, 用于定义局部子序列模式;

r ——时间序列标准差的比例, 用于判断 m 维向量是否相似。NSE 的值限定在 $[0, 1]$ 。

2) $H_{E, \text{norm}}$

赫斯特指数 (H_E) 是时间序列分析中用于衡量依赖性和持久性的关键指标, 其可应用于水资源中的复杂性和随机过程的动态特性研究^[12]。 H_E 基于时间序列的分形性质, 通过研究时间序列自相似性, 来评估系统的稳定性和调节能力。

H_E 与 0.5 的偏离程度, 可用于评估时间序列的随机性与确定性特性, 本研究通过计算其与 0.5 的距离并归一化形成 $H_{E, \text{norm}}$, 定义如式(10)~式(11)。

$$H_{E, \text{norm}} = 1 - |H_E - 0.5| \times 2 \quad (10)$$

$$H_E = \frac{\log(R/S)}{\log L} \quad (11)$$

其中: R/S ——时间序列的极差与标准差之比;

$H_{E, \text{norm}}$ ——归一化后的赫斯特指数;

H_E ——赫斯特指数;

L ——滞后长度。

$H_{E, \text{norm}}$ 在 $[0, 1]$ 。当 $H_E = 0.5$ 时, $H_{E, \text{norm}}$ 的值为 1, 表明序列接近随机性。 H_E 接近 0 或 1 时, $H_{E, \text{norm}}$ 的值趋近于 0, 表明序列中包含显著的强持久性特征。

1.4 规律性、非规律性指标

基于上述时域分析、频域分析和复杂性分析所形成的规律性、非规律性指标, 分别计算算术平均值 (MAAC), 可形成规律性得分 (RS)、非规律性得分 (IS), 在式中分别表示为 R_s 、 I_s , 如式(12)和式(13)。

$$R_s = \frac{(1 - C_v) + M_{\text{ACC}} + (T_s + S_s) + R_0}{4} \quad (12)$$

$$I_s = \frac{N_s + H_{\text{norm}} + N_{\text{SE}} + H_{E, \text{norm}}}{4} \quad (13)$$

其中: R_s ——规律性得分;

I_s ——非规律性得分;

C_v ——变异系数。

2 实例研究

2.1 实例概况

针对上海中心城区浦西供水服务区域内 10 家

水厂 2017 年 1 月 1 日—2024 年 6 月 29 日的日供水量,开展规律性、非规律性量化评价。从 10 家水厂日供水量箱线图(图 2)中可以观察到,不同水厂的水量分布具有显著差异。如标记为 C、Y2 和 T 的水厂显示出较高的中位数和较宽的四分位距,表明这些水厂的日供水量较大且波动较大。相反,标记为 L 和 W 的水厂则显示出较低的水量中位数和较窄

的四分位距,表明其日供水量较小且相对稳定。

规律性与波动性之间既有区别,又有联系。日供水量波动大的水厂,可能受趋势、季节等规律性影响因素,其数值变化范围虽然较大,但规律性可能较强。日供水量波动较小的水厂,可能因调度措施、维修等非规律性因素影响,导致其非规律性较强。

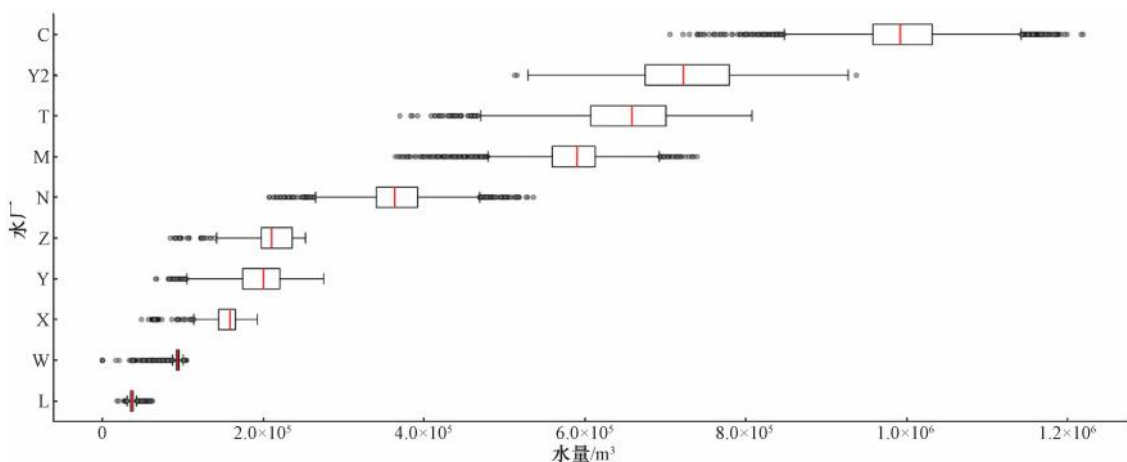


图 2 浦西供水服务区域 10 家水厂水量箱线图

Fig. 2 Boxplot of Water Volume for 10 WTPs in Puxi Water Supply Service Area

水厂日供水量主要受 3 个关键运营环节的综合影响:水量需求、水厂运行状况以及调度方式。这 3 个因素既包含了规律性的要素,也包含了非规律性的要素。

规律性要素主要与水量需求的变化特征相关,包括季节性波动和趋势的稳定性。例如,夏季高温天气通常会导致用水需求显著增加,而冬季则相对减少,总体呈现年度季节周期性变化。此外,水厂日供水量还可能受到经济发展水平、城市化进程以及人口变化等长期趋势的影响。

非规律性要素主要指一些偶发性或不可预测的因素。这类因素包括节假日期间水量的异常波动,以及大型公众活动可能导致区域短时间内用水需求急剧上升;水厂运行的突发情况,例如设备故障导致的供水能力下降;调度人员的主观调度偏好以及监测设备或网络传输的异常等。

这些规律性和非规律性要素的影响都包含在水量的时间序列中。

2.2 规律性分析评价

时域分析中的自相关性、STL 分解和复杂性分析中的相关初始参数,需要结合供水业务相关运行

规律及历史相关研究成果来设置。

水费征收以月度为账单周期,促使用户行为及供水运行管理存在明显的月度周期,其代表了供水运营的有效周期,可作为自相关分析中捕捉短期序列依赖性和周期模式的合适尺度;因此,设置自相关性分析的滞后 $k=30$,从而确保自相关函数可有效捕捉月度行为模式。

由于需水量受气候周期(如季节性温度变化和降水模式)以及年度社会和经济活动的重大影响,在日供水量 STL 分解中选择季节性周期为 365 d,可确保 STL 分解在保持趋势估算和残差分析完整性的同时,有效地分离出年度季节性。

复杂性分析的 $S_E(m, r, N)$ 计算中,结合相关研究^[13-14],为能够有效捕捉时间序列中的复杂性,同时避免过高维度带来的计算复杂性,本研究选择嵌入维度 $m=2$; r 的设置与时间序列的标准差有关,为获得最佳的区分度和敏感性,本研究选择 $r=0.2$ 。

经上述参数初始设置,规律性指标、非规律性指标各分项计算结果和 RS、IS,如图 3 和图 4 以及表 1 所示。

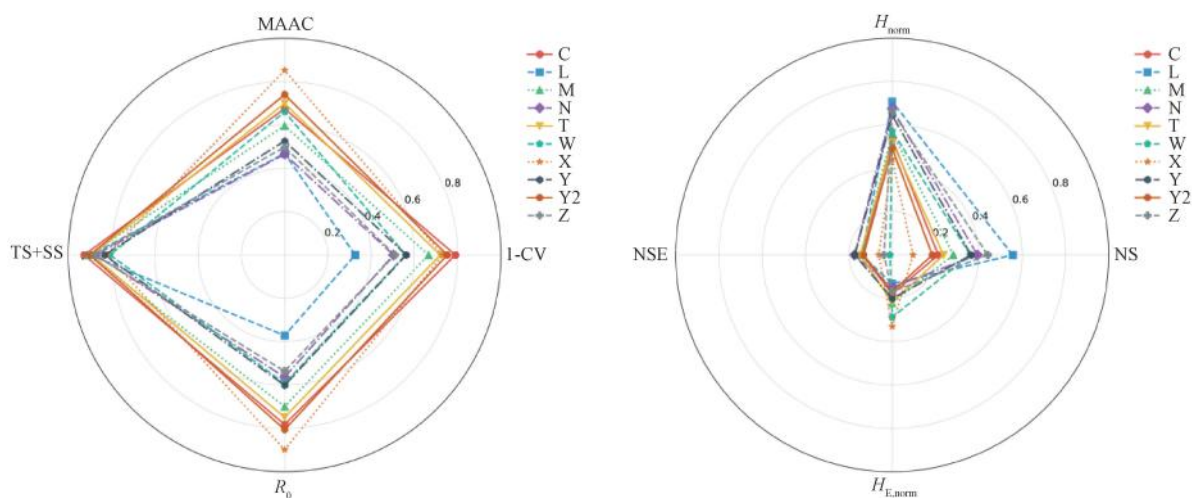


图 3 浦西 10 家水厂规律性、非规律性指标雷达分布示意图

Fig. 3 Radar Plot of Regularity and Irregularity Indices for 10 WTPs in Puxi

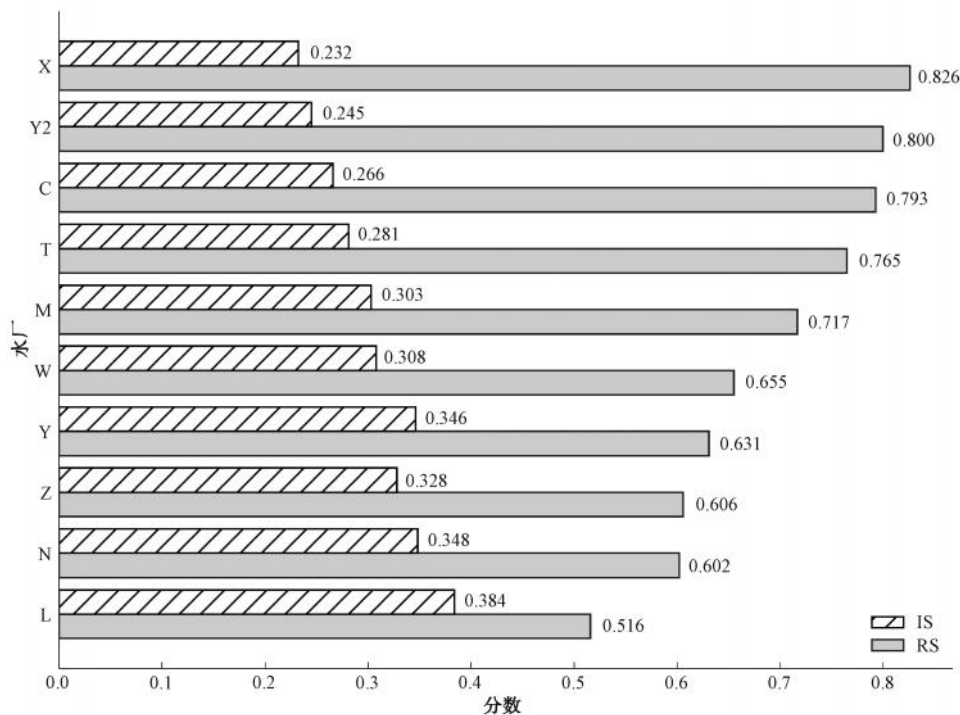


图 4 规律性、非规律得分示意图

Fig. 4 Schematic Diagram of Regularity and Irregularity Scores

由图 3 和图 4 以及表 1 可知,总体规律性指标普遍高于不规律性指标,说明各水厂总体上表现出较强的规律性,符合水厂供水以规律性较强的需求侧为主,以调度调节、水厂及监测设备运行异常为辅的总体情况。同时,不同水厂在规律性和不规律性特征上,表现出较大的差异。由图 3 可知,10 家水厂 MAAC 均呈现较高水平,特别是 X 和 Y2 水厂,表

明供水模式具有显著的时间依赖性;1-CV 反映了供水量的整体稳定性,其中 X 和 Y2 水厂表现最为突出; R_0 和 TS+SS 的分布形态,说明 10 家水厂在长期变化特征上具有一定的一致性。图 3 的 4 个不规律性指标,10 家水厂的 H_{norm} 均普遍较高(0.4~0.7),表明供水量在频域上具有显著的复杂性;NS 表现出明显的差异性,反映了 10 家水厂供水波动的不同特

表 1 浦西 10 家水厂规律性指标、非规律性指标结果明细
Tab. 1 Detailed Calculation Results of Regular and Irregular Indices for 10 WTPs in Puxi

水厂	1-CV	MAAC	TS+SS	R_0	RS	NS	H_{norm}	NSE	$H_{E,norm}$	IS
X	0.730	0.852	0.821	0.899	0.826	0.096	0.441	0.062	0.330	0.232
Y2	0.752	0.740	0.900	0.806	0.800	0.185	0.492	0.132	0.170	0.245
C	0.789	0.675	0.927	0.782	0.793	0.213	0.535	0.137	0.180	0.266
T	0.720	0.699	0.892	0.750	0.765	0.237	0.520	0.148	0.218	0.281
M	0.665	0.596	0.907	0.699	0.717	0.282	0.564	0.143	0.222	0.303
W	0.560	0.661	0.807	0.593	0.655	0.361	0.572	0.010	0.287	0.308
Y	0.562	0.526	0.832	0.602	0.631	0.364	0.645	0.173	0.201	0.346
Z	0.510	0.498	0.878	0.537	0.606	0.441	0.662	0.039	0.170	0.328
N	0.503	0.465	0.874	0.564	0.602	0.393	0.676	0.174	0.149	0.348
L	0.326	0.468	0.898	0.372	0.516	0.556	0.707	0.141	0.133	0.384

征;NSE 均普遍较低 (<0.2),说明局部时间尺度上仍存在一定的规律性,复杂性、随机性较低; $H_{E,norm}$ 均相对集中 ($0.1\sim0.3$),反映了 10 家水厂日供水量时间序列的长期相关性特征比较稳定。

各水厂 RS 和 IS 的相关系数达到 0.95 且呈负相关性,如图 5 所示,辅助印证了本评价方法的合理性。

2.3 水量预测分析

基于上述规律性分析、非规律性分析,从水量预测模型选择、预测结果评价 2 方面,开展各水厂水量预测分析。本研究选用源于金融领域的统计学模型 GARCH^[15] 和机器学习模型 XGBoost^[16],如表 2 所示。利用这 2 个模型,针对上述 10 家水厂开展预测周期为 30~60 d 的水量研究。在完成必要的特征工程和划分训练集、测试集的基础上,经模型训练

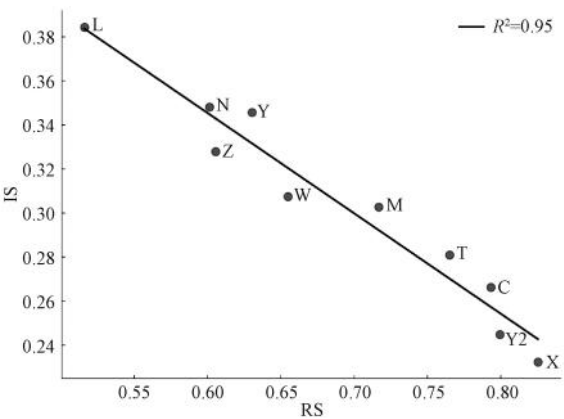


图 5 RS、IS 相关性示意图

Fig. 5 Correlation Diagram of RS and IS

(含参数优化)、开展预测评价,模型整体训练、预测框架如图 6 所示;各水厂 30~60 d 的预测结果的平均绝对百分比误差 (MAPE),如图 7 所示。

表 2 GARCH 和 XGBoost 模型对比
Tab. 2 Comparison of GARCH and XGBoost Models

比较方面	GARCH 模型	XGBoost 模型
基本假设	假设时间序列的条件方差依赖于过去的扰动(误差)和方差,误差项服从特定分布(如正态分布或 t 分布)	作为一种基于集成学习的梯度提升方法,它通过组合多个决策树的预测来创建学习器;可以处理各种数据类型,不假设特定的数据分布
参数估计	通常使用最大似然估计(MLE)方法	采用梯度提升框架顺序构建树,优化带有正则化项的损失函数;参数调整通常涉及交叉验证等技术以防止过拟合
统计检验	使用拉格朗日乘数检验来确定条件异方差性的存在;通过残差诊断(例如检查残差的自相关性)来评估模型的适用性	通常依赖于交叉验证评估模型性能,缺乏严格的统计推断
解释能力	该模型直接提供了时间序列波动性动态的洞察;参数可以解释为波动性冲击的持续性和幅度	由于模型的集成性质,可解释性相对较弱;可通过特征重要性排序等方法了解变量的影响程度
稳定性	模型稳定性要求与滞后平方残差(ARCH 项)和滞后条件方差(GARCH 项)相关的参数之和小于 1,否则可能导致非平稳的波动性过程;对异常值较为敏感	通过正则化技术避免过拟合,提高模型的泛化能力;对异常值具有一定的鲁棒性

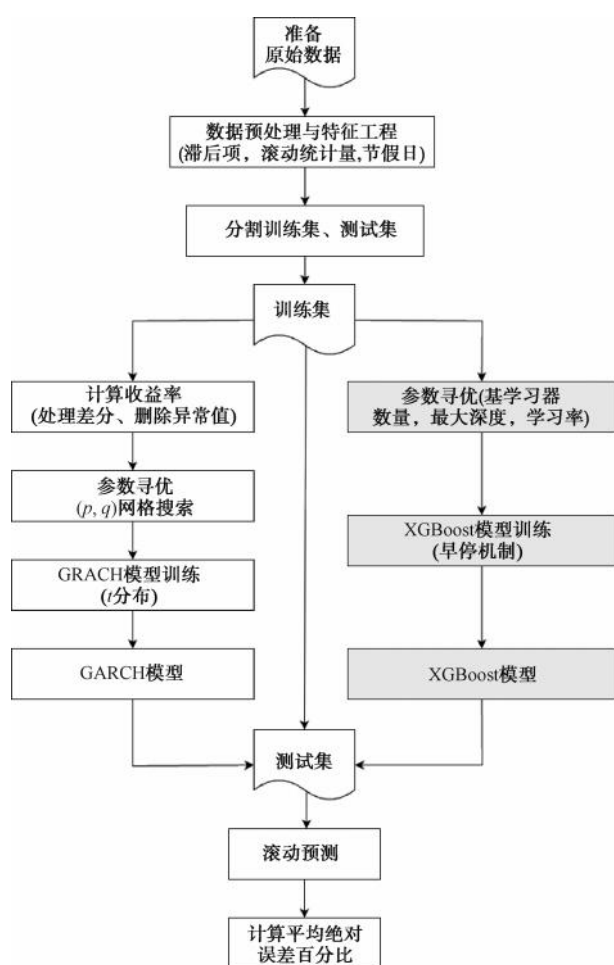


图 6 GARCH 模型和 XGBoost 模型训练与预测
流程示意图

Fig. 6 Schematic Diagram of Training and Prediction
Workflow of GARCH and XGBoost Models

综合表 1 和图 7 可知, GARCH 模型在时间序列呈现较强自相关 ($MAAC > 0.70$) 时, 能有效捕捉波动率的长期持续性。X 厂 0.852 的 $MAAC$ 强度 (全样本最高), 其残差平方序列满足 GARCH 建模条件; 同时, 其 $1-CV = 0.730$ 与 $R_0 = 0.899$ 表明: X 厂水量数据的噪声较低, 从而, 使其 $MAPE$ 显著小于 XGBoost。Y2 厂虽然 $TS+SS = 0.900$ 接近强季节阈值, 但 $IS = 0.245$ (全样本最低) 与中等稳定性 ($1-CV = 0.752$) 保障了 GARCH 有效性。GARCH 模型对伪趋势 ($NSE < 0.05$ 且 $TS+SS > 0.8$) 场景具较强的适应性, 突出表现在: W 厂因其伪趋势特征明显 (低复杂度 $NSE = 0.01$ 与高趋势与季节强度 $TS+SS = 0.807$ 并存) 及 $H_{E, norm} = 0.208$ 表明 W 厂水量序列数据, 是简单波动模式占据主导, 此类序列满足

GARCH 过程, 使其 GARCH 的预测效果优于 XGBoost。Z 厂与 W 厂具备类似的伪趋势特征 ($NSE = 0.039$ 且 $TS+SS = 0.878$) 使 GARCH 的 $MAPE$ 较 XGBoost 低。Y 厂 ($MAAC = 0.526$) 虽然自相关水平未达到 GARCH 优势阈值, 但其 $IS = 0.346$ 与 $R_0 = 0.602$ 的组合特征表明其非平稳性主要源于方差波动而非趋势变化, 使 GARCH 的 $MAPE$ 较 XGBoost 低。

GARCH 的线性假设在复杂场景中表现出明显局限, 在趋势主导 ($TS+SS > 0.85$ 且 $1-CV > 0.65$) 与混沌噪声场景 ($H_{norm} > 0.65$ 且 $R_0 < 0.6$) 中, 较 XGBoost 模型 $MAPE$ 更高。C 厂 ($TS+SS = 0.927$ 全样本最高且 $1-CV = 0.789$) 表现为强趋势与季节耦合 ($TS+SS > 0.9$) 且高稳定性 ($1-CV > 0.75$) 形成确定性分量主导结构。T 厂 ($R_0 = 0.750$ 且 $H_{norm} = 0.520$) 表现中等噪声环境与频谱复杂度平衡 ($H_{norm} = 0.52$), 使树模型分裂准则可有效解析时序依赖。M 厂在明显不符合强自相关 ($MAAC > 0.70$) 与伪趋势 ($NSE < 0.05$ 且 $TS+SS > 0.8$), 但符合强季节模式 ($TS+SS = 0.907$) 的情况下, XGBoost 可准确分离真实趋势。N 厂 ($H_{norm} = 0.676$ 且 $R_0 = 0.564$) 表现为明显的混沌系统特性 ($H_{norm} > 0.65$) 与中等噪声 ($R_0 = 0.564$) 组合, 构成非线性映射优势空间。L 厂 ($NS = 0.556$, 全样本最高且 $H_{norm} = 0.707$) 高噪声强度与高混沌性, XGBoost 通过缺失值补偿机制和深度树结构 [最大深度 (max_depth) = 7] 实现噪声免疫, 使得 $MAPE$ 低于 GARCH。

总体上, 上述 10 家水厂, 可分为高规律性、中规律性、低规律性和复杂性 4 级。RS ≥ 0.8 的作为高规律性水厂 (X 和 Y2) 主要由于日水量时间序列数据较强的自相关性, 其序列波动呈现显著持续性和平稳残差结构。当自相关性指标 $MAAC > 0.70$ 时, 推荐可有效捕捉波动率长期持续性的 GARCH、FIGARCH^[17] 等模型。

RS ≥ 0.7 且 RS < 0.8 的为中规律性水厂 (C、T 和 M 水厂), 以突出趋势-季节耦合特征为标志 ($TS+SS > 0.85$), 兼具中等稳定性 ($1-CV > 0.65$)。推荐可有效模拟非线性关系的 XGBoost、长短时记忆网络 (LSTM)^[18] 等模型, 以及可捕捉季节性的先知模型 (Prophet^[19])、三角函数季节模型 (TBATS^[20]) 等模型。

RS ≥ 0.6 且 RS < 0.7 的为低规律性水厂 (W、Y、

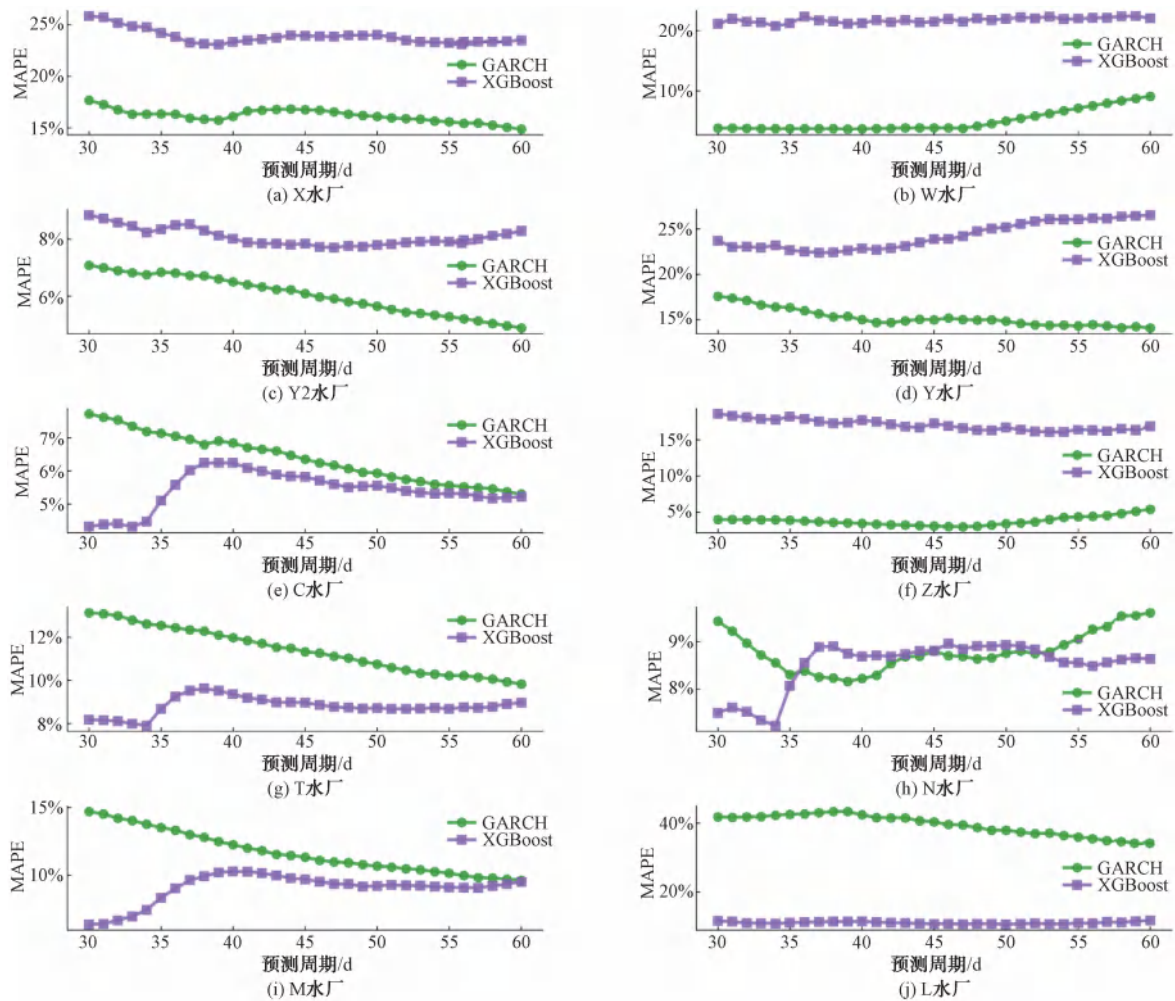


图 7 GARCH 模型和 XGBoost 模型预测结果 MAPE 示意图

Fig. 7 Schematic Diagram of MAPE of Prediction Results from GARCH and XGBoost Models

Z 和 N 水厂), 特征为高噪声环境 ($R_0 < 0.6$ 且 $T R_0 > 0.5$), 其 H_{norm} 普遍超过 0.6。推荐使用 GARCH、LightGBM^[21]、CatBoost^[22] 等可适应强噪声的模型。

$RS < 0.6$ 且复杂性水厂 (L 水厂), 具有极端噪声 ($R_0 < 0.4$) 与混沌频谱特征 ($H_{\text{norm}} > 0.7$), 推荐可适应非线性情况的 DeepAR^[23]、N-BEATS^[24] 等深度学习模型。

3 结论与建议

本研究从时域、频域和复杂性 3 个角度, 分别形成了规律性、非规律性的 4 项指标; 通过算术平均值形成 RS、IS。以浦西区域 10 家水厂多年积累的日供应量数据为基础, 开展规律性、非规律性的量化评价和水量预测评价。结合 10 家水厂供水量 30~60 d 预测结果的分析, 形成高规律性、中规律性、低规律性和复杂性 4 个类别, 并分别推荐水量预测模型。

本研究的规律性、非规律性量化评价可为水量预测模型的选择和应用, 提供量化依据。在一定程度上缓解了机器学习模型、深度学习类水量预测模型应用于与生产实践时面临的解释性差、透明度不高等问题, 进而可提升水量预测的可靠性。具备一定可靠性的水量预测, 在供水系统优化调度方面, 可辅助统筹调整各厂水泵运行模式, 避免低效运行或设备过载, 辅助降低能源消耗、设备维修率; 在提升应急响应能力方面, 动态可靠的水量预测可辅助快速识别供需缺口, 辅以实时压力监测, 可及时调配各水厂富余供水能力, 高效应对用水突增、大口径管道爆管等突发情况。

本研究的规律性、非规律性指标算法, 辅以水厂日供应量计算采集程序, 可动态量化评价水厂运行状况, 动态选择水量预测模型开展预测并及时评价

预测结果可靠性。水量预测模型结果,可以应用程序编程接口方式,为调度、水厂运行管理等相关系统提供关键输入,推动供水系统运营管理从“经验驱动”转向“数据驱动”,提升运营管理精细化、科学化水平。

参考文献

- [1] 顾晨. 超大城市供水系统韧性评估与应用实践——以上海受咸潮入侵影响为例[J]. 净水技术, 2024, 43(5): 1-9.
GU C. Resilience assessment and application for megacity water supply system: Case of Shanghai salt-tide intrusion[J]. Water Purification Technology, 2024, 43(5): 1-9.
- [2] 周骅. 新冠疫情封控管理期间特大型城市用水量影响分析[J]. 净水技术, 2022, 41(9): 137-142, 149.
ZHOU H. Influence analysis of lockdown measure during COVID-19 epidemic on water consumption for a mega city[J]. Water Purification Technology, 2022, 41(9): 137-142, 149.
- [3] 姚建旭. 江西某水厂供水量组合预测方法及应用[D]. 扬州: 扬州大学, 2024.
YAO J X. Combined forecasting method and application of water supply in a water plant in Jiangxi Province[D]. Yangzhou: Yangzhou University, 2024.
- [4] 周欣磊, 顾海挺, 刘晶, 等. 基于集成学习与深度学习的日供水量预测方法[J]. 浙江大学学报(工学版), 2023, 57(6): 1120-1127.
ZHOU X L, GU H T, LIU J, et al. Daily water supply prediction method based on integrated learning and deep learning[J]. Journal of Zhejiang University (Engineering Science), 2023, 57(6): 1120-1127.
- [5] 刘怀利, 王铭铭, 查淳膺, 等. 基于多维度输入的水厂日取水量卷积长短期记忆网络预测[J]. 水电能源科学, 2024, 42(5): 72-75.
LIU H L, WANG M M, ZHA C Y, et al. Convolutional long and short-term memory network prediction of daily water intake from waterworks based on multidimensional input [J]. Water Resources and Power, 2024, 42(5): 72-75.
- [6] 刘奔. 天津市双水源供水系统用水量预测及优化调度研究[D]. 天津: 天津大学, 2022.
LIU B. Study on water consumption forecasting and optimal dispatching of double water supply system in Tianjin [D]. Tianjin: Tianjin University, 2022.
- [7] PIRES C, MARTINS M V. Enhancing water management: A comparative analysis of time series prediction models for distributed water flow in supply networks[J]. Water, 2024, 16(13): 1827. DOI: 10.3390/w16131827.
- [8] CLEVELAND R B, CLEVELAND W S, MCRAE J E, et al. STL: A seasonal-trend decomposition procedure based on loess [J]. Journal of Official Statistics, 1990, 6(1): 3-73.
- [9] WANG X, SMITH K A, HYNDMAN R J. Characteristic-based clustering for time series data[J]. Data Mining and Knowledge Discovery, 2006, 13(3): 335-364.
- [10] ZHANG H, WANG H, YAN Y, et al. Weighted dynamic transfer network and spectral entropy for weak nonlinear time series detection [J]. Nonlinear Dynamics, 2023, 111(10): 9345-9359.
- [11] 薛联青, 刘远洪, 张梦泽, 等. 基于样本熵的降雨和径流时间序列突变检验[J]. 地球科学与环境学报, 2015, 37(3): 75-80.
XUE L Q, LIU Y H, ZHANG M Z, et al. Mutation test on rainfall and runoff time series based on sample entropy [J]. Journal of Earth Sciences and Environment, 2015, 37(3): 75-80.
- [12] PIZARRO A, ACUÑA P, DIMITRIADIS P, et al. Past, present, and future of the hurst-kolmogorov dynamics in stochastics: A bibliometric analysis of the last 50 years in water resources[J]. Journal of Hydrology, 2024, 643: 132008. DOI: 10.1016/j.jhydrol.2024.132008.
- [13] CHOU C M. Complexity analysis of rainfall and runoff time series based on sample entropy in different temporal scales [J]. Stochastic Environmental Research and Risk Assessment, 2014, 28(6): 1401-1408.
- [14] ZHOU Y, ZHANG Q, LI K, et al. Hydrological effects of water reservoirs on hydrological processes in the East River (China) basin: Complexity evaluations based on the multi-scale entropy analysis[J]. Hydrological Processes, 2012, 26(21): 3253-3262.
- [15] BOLLERSLEV T. Generalized autoregressive conditional heteroskedasticity[J]. Journal of Econometrics, 1986, 31(3): 307-327.
- [16] CHEN T Q, GUESTRIN C. XGBoost: A scalable tree boosting system [C]// Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. USA: Association for Computing Machinery, 2016: 785-794.
- [17] BAILLIE R T, BOLLERSLEV T, MIKKELSEN H O. Fractionally integrated generalized autoregressive conditional heteroskedasticity[J]. Journal of Econometrics, 1996, 74(1): 3-30.
- [18] HOCHREITER S, SCHMIDHUBER J. Long short-term memory [J]. Neural Computation, 1997, 9(8): 1735-1780.
- [19] TAYLOR S J, LETHAM B. Forecasting at scale [J]. The American Statistician, 2018, 72(1): 37-45.
- [20] DE LIVERA A M, HYNDMAN R J, SNYDER R D. Forecasting time series with complex seasonal patterns using exponential smoothing[J]. Journal of the American Statistical Association, 2011, 106(496): 1513-1527.

(下转第 83 页)

- Technology, 2020, 39(12): 13–18.
- [18] 中国工程建设标准化协会. 建筑给水排水设计规范[M]. 北京: 中国建筑工业出版社, 2007.
China Association for Engineering Construction Standardization. Code for design of building water supply drainage[S]. Beijing: China Architecture & Building Press, 2007.
- [19] 傅雪军, 马绍赛, 朱建新, 等. 封闭式循环水养殖系统水处理效率及半滑舌鳎养殖效果分析[J]. 环境工程学报, 2011, 5(4): 745–751.
FU X J, MA S S, ZHU J X, et al. Analysis of the efficiency of water treatment and the cultured effect of *Cynoglossus semilaevis* Günther in closed re-circulating aquaculture systems [J]. Chinese Journal of Environmental Engineering, 2011, 5(4): 745–751.
- [20] 刘鑫, 康建平, 林岗, 等. 东星斑循环水养殖系统的水体营养盐及微生物群落结构特征[J]. 福建农业学报, 2022, 37(6): 754–764.
LIU X, KANG J P, LIN G, et al. Nutrients and microbial community in recirculating aquaculture system for plectropomus leopardus aquaculture [J]. Fujian Journal of Agricultural Sciences, 2022, 37(6): 754–764.
- [21] 蒋磊, 刘铭羽, 李希, 等. 曝气对生物基质技术处理养殖废水脱氮除磷效果的影响特征及机理[J]. 水处理技术, 2021, 47(9): 71–76.
JIANG L, LIU M Y, LI X, et al. Characteristics and mechanism of the effect of aeration on nitrogen and phosphorus removal from swine wastewater treated by biological matrix technology [J]. Technology of Water Treatment, 2021, 47(9): 71–76.
- [22] 张正, 王印庚, 曹磊, 等. 海水循环水养殖系统生物膜快速挂膜试验[J]. 农业工程学报, 2012, 28(15): 157–162.
ZHANG Z, WANG Y G, CAO L, et al. Pilot-scale test for biofilm rapid formation in biofilter of recirculating mariculture system[J]. Transactions of the Chinese Society of Agricultural Engineering, 2012, 28(15): 157–162.
- [23] CHEN S L, LING J, BLANCHETON J P. Nitrification kinetics of biofilm as affected by water quality factors [J]. Aquacultural Engineering, 2006, 34(3): 179–197.
- [24] BAO W J, ZHU S M, GUO S R, et al. Particle size distribution mathematical models and properties of suspended solids in a typical freshwater pond [J]. Environmental Pollution, 2018, 241: 164–171. DOI: 10.1016/j.envpol.2018.05.063.
- [25] DAVIDSON J, SUMMERFELT S. Solids flushing, mixing, and water velocity profiles within large (10 and 150 m³) circular ‘Cornell-type’ dual-drain tanks [J]. Aquacultural Engineering, 2004, 32(1): 245–271.
- [26] LABATUT R A, EBELING J M, BHASKARAN R, et al. Effects of inlet and outlet flow characteristics on mixed-cell raceway (MCR) hydrodynamics [J]. Aquacultural Engineering, 2007, 37(2): 158–170.
- [27] PLEW D R, KLEBERT P, ROSTEN T W, et al. Changes to flow and turbulence caused by different concentrations of fish in a circular tank [J]. Journal of Hydraulic Research, 2015, 53(3): 364–383.

(上接第 72 页)

- [21] KE G, MENG Q, FINLEY T, et al. LightGBM: A highly efficient gradient boosting decision tree[J]. Advances in Neural Information Processing Systems, 2017(30): 3146–3154.
- [22] PROKHORENKOVA L, GUSEV G, VOROBEOV A, et al. CatBoost: Unbiased boosting with categorical features[J]. arXiv preprint arXiv:1706.09516, 2018.
- [23] 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.
- [24] BORIS O N, CARP D, CHAPADOS N, et al. N-BEATS: Neural basis expansion analysis for time series forecasting[J]. arXiv preprint arXiv:1905.10437, 2020.