



(10) 授权公告号 CN 119577698 B

(45) 授权公告日 2025.04.18

(21) 申请号 202510135862.3

(22) 申请日 2025.02.07

(65) 同一申请的已公布的文献号  
申请公布号 CN 119577698 A

(43) 申请公布日 2025.03.07

(73) 专利权人 中国电建集团贵阳勘测设计研究  
院有限公司地址 550081 贵州省贵阳市观山湖区兴黔  
路16号专利权人 中电建生态环境设计研究有限公  
司(72) 发明人 张景峥 赵谊 罗超 魏浪  
陶维中 张庆 吴维兴 张小清(74) 专利代理机构 贵州派腾知识产权代理有限  
公司 52114

专利代理师 易昕

(51) Int.Cl.

G06F 18/27 (2023.01)

G06F 17/18 (2006.01)

G06Q 10/0631 (2023.01)

G06Q 50/26 (2024.01)

(56) 对比文件

CN 115496290 A, 2022.12.20

CN 116228497 A, 2023.06.06

审查员 王瑾香

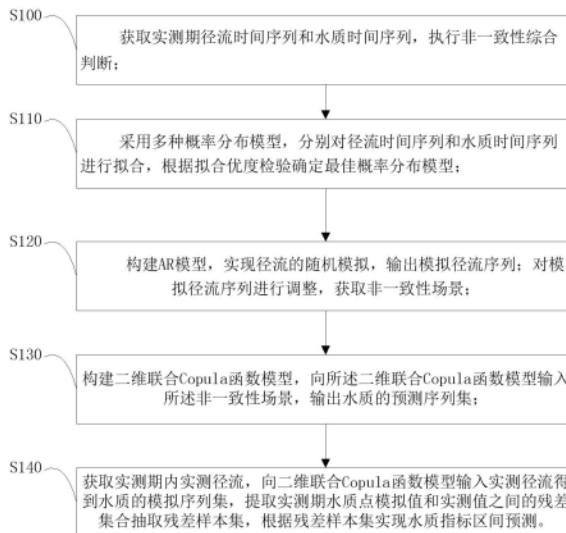
权利要求书2页 说明书11页 附图12页

## (54) 发明名称

一种非一致性场景下的水质指标区间预测  
方法

## (57) 摘要

本发明公开了一种非一致性场景下的水质指标区间预测方法,包括以下步骤:获取指定时间段内径流时间序列和水质时间序列,执行非一致性综合判断;采用多种概率分布模型,分别对径流时间序列和水质时间序列进行拟合,根据拟合优度检验确定最佳概率分布模型;构建AR模型,实现径流的随机模拟,输出模拟径流序列;对模拟径流序列进行调整,获取非一致性场景;构建二维联合Copula函数模型,向二维联合Copula函数模型输入未来非一致性场景,输出水质预测序列集;从水质的模拟序列集提取实测期水质点模拟值和实测值之间的残差集合抽取残差样本集,根据残差样本集实现水质指标区间预测。根据上述技术方案,可以得出非一致性背景下水质数据的点预测和水质区间预测。



1. 一种非一致性场景下的水质指标区间预测方法,其特征在于,包括以下步骤:

获取实测期径流时间序列和水质时间序列,执行非一致性综合判断;

采用多种概率分布模型,分别对所述径流时间序列和水质时间序列进行拟合,根据拟合优度检验确定最佳概率分布模型;

构建AR模型,实现径流的随机模拟,输出模拟径流序列;对模拟径流序列进行调整,获取非一致性场景;所述AR模型由随机扰动项 $\varepsilon_t$ 和时间序列的部分项目的加权之和组成:表

示为: $x_t = a_1 x_{t-1} + a_2 x_{t-2} + \cdots + a_p x_{t-p} + \varepsilon_t = \sum_{j=1}^p a_j x_{t-j} + \varepsilon_t$ ,其中, $x_t$ 为t时刻的径流模拟

值、p为自回归阶数、 $a_j$ 为自回归系数、 $\varepsilon_t$ 为噪声项;所述对模拟径流序列进行调整指:通过统计因子调整法和径流季节性变异模拟非一致场景;所述非一致场景包括:均值上升、均值下降、 $C_v$ 上升、 $C_v$ 下降、季节前移、季节滞后;

构建二维联合Copula函数模型,向所述二维联合Copula函数模型输入所述非一致性场景,可输出水质的预测的模拟序列集;

获取实测期内实测径流,向所述二维联合Copula函数模型,输入所述实测径流,得到水质的模拟序列集,提取实测期水质点模拟值和实测值之间的残差集合抽取残差样本集,根据所述残差样本集实现水质指标区间预测;

其中,所述统计因子调整法用于根据非一致性条件对序列统计特征参数的影响模拟非一致性场景,所述场景包括均值非一致性和 $C_v$ 非一致性;生成均值的场景时, $Q_i^{mean} = K Q_i^{AR}$ ,其中, $Q_i^{mean}$ 和 $Q_i^{AR}$ 分别为均值非一致性场景下的径流序列和AR模型得到的模拟径流序列,K为调整因子;生成 $C_v$ 的场景时,保证均值不变的情况下, $Q_i^{C_v} = K Q_i^{AR} + (1-K)\bar{Q}$ ,其中, $Q_i^{C_v}$ 表示 $C_v$ 非一致性场景下的径流序列; $\bar{Q}$ 为径流均值;

所述径流季节性变异用于体现非一致性条件对径流季节性的影响,通过累计时间分布函数实现径流在季节非一致性条件下的重构。

2. 根据权利要求1所述的水质指标区间预测方法,其特征在于,所述执行非一致性综合判断包括:

通过MK趋势分析方法获取各时间序列的变化趋势,包括:使用标准正态分布进行假设检验,以MK统计量大于或小于95%置信水平的临界值判断时间序列的趋势变化;

通过MK突变检验方法判断各时间序列的非一致性背景,包括:取95%的置信水平,判断检验量曲线和在置信区间内是否相交,如果相交,则判断时间序列在交点处产生了显著突变,呈现出非一致性;

通过相关性分析获取径流序列与水质序列间的相关程度,包括:当相关性系数大于0时,则表示两序列间存在正相关性,反之则存在负相关性;其绝对值越接近于1,相关性程度越大。

3. 根据权利要求1所述的水质指标区间预测方法,其特征在于,采用多种概率分布模型包括:确定径流时间序列和各水质时间序列的概率分布特征,所述概率分布特征包括:极值分布、指数分布、伽马分布、广义极值分布、广义帕累托分布、对数正态分布、正态分布和威布尔分布;

所述拟合优度包括:AIC值、RMSE值和KS检验的统计量;

所述根据拟合优度检验确定最佳概率分布模型的依据包括:KS检验的统计量等于1时说明满足假设的分布特征;RMSE值越小,拟合程度越好;AIC值越小,拟合程度越好。

4.根据权利要求1所述的水质指标区间预测方法,其特征在于,所述径流季节性变异实现径流在季节非一致性条件下的重构模型表示为:

$$TDF(j) = \frac{\sum_{i=1}^j Q_i}{\sum_{i=1}^n Q_i},$$

$$TDF'(i) = TDF(i) - TDF^{tri1}(i) + TDF^{tri2}(i), i = 1, 2, \dots, n,$$

$$Q'_i = \begin{cases} TDF'(i) \sum_{i=1}^n Q_i^{AR}, & i = 1, 2, \dots, n \\ (TDF'(i) - TDF'(i-1)) \sum_{i=1}^n Q_i^{AR}, & i = 1, 2, \dots, n' \end{cases}$$

其中, $TDF(j)$ 为时间序列中第j个累计时间分布函数,n为时间序列长度, $Q_i$ 和 $Q'_i$ 分别表示实测径流时间序列和季节非一致性场景下的径流时间序列, $TDF'(i)$ 、 $TDF^{tri1}(i)$ 和 $TDF^{tri2}(i)$ 分别表示非一致性场景下、非一致性变化前和非一致性变化后三角形入流过程线的累计时间分布函数。

5.根据权利要求1所述的水质指标区间预测方法,其特征在于,所述构建二维联合Copula函数模型包括:

确定模拟所用的水质指标序列v;径流u;

确定径流和水质指标序列的边缘分布;

选择Copula函数,描述径流和水质指标间的相关性结构。

6.根据权利要求1所述的水质指标区间预测方法,其特征在于,所述根据所述残差样本集实现水质指标区间预测指:在95%和90%的置信水平上,进行模拟残差的区间预测,计算模拟水质数据的上下限。

7.根据权利要求1所述的水质指标区间预测方法,其特征在于,所述抽取残差样本集时,采用Bootstrap方法,对实测期水质点模拟值和实测值之间的残差集合有放回的抽取N次,得到N组新的残差样本集。

## 一种非一致性场景下的水质指标区间预测方法

### 技术领域

[0001] 本发明涉及水质指标预测技术领域,具体而言,涉及一种非一致性场景下的水质指标区间预测方法。

### 背景技术

[0002] 随着社会发展,生活用水、工业用水和农业用水的用量在不断增加,对水资源的需求也日益增大。然而,日益严峻的水体污染成为目前我国面临的主要水问题之一。在此情况下,对表征水体污染状况的各项水质指标进行模拟和预测也成为水质研究的热点和关键。

[0003] 在实际使用中,常通过对采集水样来监测各项指标值(如pH值、高锰酸钾指数、氨氮、总磷、总氮等),并依据《地表水环境质量标准》中各项水质指标的限值将水体分为五类。

[0004] 现阶段,已有不少研究针对水质进行了模拟。这些研究大多采用统计模型、水文模型、机器学习和深度学习,以及多种方法耦合等手段,基于历史期的其他水文数据序列进行水质模拟,进而得出相应的水质指标序列。然而,变化环境下,全球范围内的水文非一致性过程愈加明显,传统的水文一致性已不适用实际环境,因此,基于历史期水文数据的水质模拟结果的合理性还需要进一步确认,并需要在各水文数据表现出非一致性的情况下,对与水文数据存在相关性的水质指标数据进行进一步分析和模拟。

[0005] 非一致性水文数据的特征表现在其统计特征和参数等随时间产生了变化,这导致传统观点认为的水文模拟过程中呈“稳态”的假定不再适用。而目前常用的方案为基于一致性假设进行水质点预测,即输入序列不考虑随时间变化的统计特征和参数,且不能突出水质模拟数据的不确定性。因此,需要一种水质模拟的方法,从非一致性背景出发,综合考虑多种非一致性输入和水质样本代表性中的不确定性,实现非一致性的水质数据的点预测和区间预测。

### 发明内容

[0006] 实现上述目的,本申请提供了一种非一致性场景下的水质指标区间预测方法,包括以下步骤:

[0007] 获取实测期内径流时间序列和水质时间序列,执行非一致性综合判断;

[0008] 采用多种概率分布模型,分别对径流时间序列和水质时间序列进行拟合,根据拟合优度检验确定最佳概率分布模型;

[0009] 构建AR模型,实现径流的随机模拟,输出预测的模拟径流序列;对模拟径流序列进行调整,获取非一致性场景;

[0010] 构建二维联合Copula函数模型,向二维联合Copula函数模型输入非一致性场景,输出水质的预测序列集;

[0011] 获取实测期内实测径流,向所述二维联合Copula函数模型,输入实测径流,得到水质的模拟序列集,提取实测期水质点模拟值和实测值之间的残差集合抽取残差样本集,根据残差样本集实现水质指标区间预测。

[0012] 其中,执行非一致性综合判断包括:

[0013] 通过MK趋势分析方法获取各时间序列的变化趋势,包括:使用标准正态分布进行假设检验,以MK统计量大于或小于95%置信水平的临界值判断时间序列的趋势变化;

[0014] 通过MK突变检验方法判断各时间序列的非一致性背景,包括:取95%的置信水平,判断检验量曲线和在置信区间内是否相交,如果相交,则判断时间序列在交点处产生了显著突变,呈现出非一致性;

[0015] 通过相关性分析获取径流序列与水质序列间的相关程度,包括:当相关性系数大于0时,则表示两序列间存在正相关性,反之则存在负相关性;其绝对值越接近于1,相关性程度越大。

[0016] 进一步的,采用多种概率分布模型包括:确定径流时间序列和各水质时间序列的概率分布特征,所述概率分布特征包括:极值分布、指数分布、伽马分布、广义极值分布、广义帕累托分布、对数正态分布、正态分布和威布尔分布;

[0017] 拟合优度包括:AIC值、RMSE值和KS检验的统计量;

[0018] 根据拟合优度检验确定最佳概率分布模型的依据包括:KS检验的统计量等于1时说明满足假设的分布特征;RMSE值越小,拟合程度越好;AIC值越小,拟合程度越好。

[0019] 其中,AR模型由随机扰动项 $\varepsilon_t$ 和时间序列的部分项目的加权之和组成:表示为:

$$[0020] \quad x_t = a_1 x_{t-1} + a_2 x_{t-2} + \cdots + a_p x_{t-p} + \varepsilon_t = \sum_{j=1}^p a_j x_{t-j} + \varepsilon_t, \text{其中, } x_t \text{ 为 } t \text{ 时刻}$$

的径流模拟值、 $p$  为自回归阶数、 $a_j$  为自回归系数、 $\varepsilon_t$  为噪声项;

[0021] 对模拟径流序列进行调整指:通过统计因子调整法和径流季节性变异模拟非一致场景;所述非一致场景包括:均值上升、均值下降、 $C_v$  上升、 $C_v$  下降、季节前移、季节滞后;

[0022] 进一步的,统计因子调整法用于根据非一致性条件对序列统计特征参数的影响模拟非一致性场景,所述场景包括均值非一致性和 $C_v$  非一致性;

[0023] 生成均值的场景时, $Q_i^{mean} = K Q_i^{AR}$ ,其中, $Q_i^{mean}$  和 $Q_i^{AR}$  分别为均值非一致性场景下的径流序列和AR模型得到的模拟径流序列, $K$  为调整因子;

[0024] 生成 $C_v$  的场景时,保证均值不变的情况下, $Q_i^{C_v} = K Q_i^{AR} + (1-K) \bar{Q}$ ,其中, $Q_i^{C_v}$  表示 $C_v$  非一致性场景下的径流序列; $\bar{Q}$  为径流均值。

[0025] 进一步的,径流季节性变异用于体现非一致性条件对径流季节性的影响;

[0026] 径流季节性变异通过累计时间分布函数实现径流在季节非一致性条件下的重构模型,重构模型表示为:

$$[0027] \quad TDF(j) = \frac{\sum_{i=1}^j Q_i}{\sum_{i=1}^n Q_i},$$

$$[0028] \quad TDF'(i) = TDF(i) - TDF^{tri1}(i) + TDF^{tri2}(i) \quad , i = 1, 2, \dots, n,$$

$$[0029] \quad Q'_i = \begin{cases} TDF'(i) \sum_{i=1}^n Q_i^{AR} & , i = 1, 2, \dots, n \\ (TDF'(i) - TDF'(i-1)) \sum_{i=1}^n Q_i^{AR} & , i = 1, 2, \dots, n' \end{cases}$$

[0030] 其中,  $TDF(j)$  为时间序列中第  $j$  个累计时间分布函数,  $n$  为时间序列长度,  $Q_i$  和  $Q'_i$  分别表示实测径流时间序列和季节非一致性场景下的径流时间序列,  $TDF'(i)$ 、 $TDF^{tri1}(i)$  和  $TDF^{tri2}(i)$  分别表示非一致性场景下、非一致性变化前和非一致性变化后三角形入流过程线的累计时间分布函数。

[0031] 构建二维联合Copula函数模型包括:

[0032] 确定模拟所用的水质指标序列  $\mathbf{v}$ ; 径流  $\mathbf{u}$ ;

[0033] 确定径流和水质指标序列的边缘分布;

[0034] 选择合适的Copula函数, 描述径流和水质指标间的相关性结构。

[0035] 进一步的, 抽取残差样本集时, 采用Bootstrap方法, 对实测期水质点模拟值和实测值之间的残差集合有放回的抽取  $N$  次, 得到  $N$  组新的残差样本集。

[0036] 根据所述残差样本集实现水质指标区间预测指: 在95%和90%的置信水平上, 进行模拟残差的区间预测, 计算模拟水质数据的上下限。

[0037] 本发明从非一致性背景出发, 通过场景模拟考虑多种非一致性输入, 进而得出代表非一致性的水质数据的点预测; 其次, 通过考虑水质样本代表性中的不确定性, 对水质数据做出区间预测。利用本发明得到的水质模拟数据综合考虑了非一致性的影响以及样本代表性中的不确定性, 具有一定的实际价值和理论价值。

## 附图说明

[0038] 图1是根据本发明实施例提供的非一致性场景下的水质指标区间预测方法步骤图;

[0039] 图2是根据本发明实施例提供的实测径流序列的非一致性综合诊断结果示意图;

[0040] 图3是根据本发明实施例提供的径流与水质序列间的相关性热图;

[0041] 图4是根据本发明实施例提供的各非一致性场景下的径流模拟序列示意图;

[0042] 图5是根据本发明实施例提供的场景1下的水质模拟结果示意图;

[0043] 图6是根据本发明实施例提供的场景2下的水质模拟结果示意图;

[0044] 图7是根据本发明实施例提供的场景3下的水质模拟结果示意图;

[0045] 图8是根据本发明实施例提供的场景4下的水质模拟结果示意图;

[0046] 图9是根据本发明实施例提供的场景5下的水质模拟结果示意图;

[0047] 图10是根据本发明实施例提供的场景6下的水质模拟结果示意图;

[0048] 图11是根据本发明实施例提供的场景1中95%置信水平下4项水质的点和区间预测

示意图；

[0049] 图12是根据本发明实施例提供的场景1中90%置信水平下4项水质的点和区间预测示意图。

### 具体实施方式

[0050] 本发明从非一致性背景出发,通过场景模拟多种非一致性输入,进而得出代表非一致性的水质数据的点预测;其次,通过分析水质样本中的不确定性,对水质数据做出区间预测。具体实现时,首先通过非一致性诊断明确径流数据的非一致性背景,同时采用8种不同的概率分布模型,根据实际需求确定需要模拟的水质指标,通过拟合优度检验确定出适合研究区水质数据及径流数据的最佳概率分布;其次,基于径流的自相关性,通过AR模型进行径流随机模拟,再利用非一致性场景模拟技术生成6种不同的径流非一致性情景;最后,通过Copula函数构建二维模型,以不同非一致性径流情景为输入,模拟出不同非一致性情景下的水质变化情况,并利用Bootstrap方法对实测期模拟数据和观测数据的残差进行有放回抽样,最终得出不同置信水平下,水质模拟数据的点预测和区间预测。

[0051] 下面结合说明书附图对本发明的具体实现方式做详细描述。

[0052] 图1提供了本发明提出的非一致性场景下的水质指标区间预测方法,包括以下步骤:

[0053] 步骤S100:获取实测期内径流时间序列和水质时间序列,执行非一致性综合诊断;

[0054] 在本发明中,首先需要明确径流数据的非一致性背景,为后续步骤提供理论依据。本发明采用综合检验方法对径流时间序列和水质时间序列进行平稳性检验以实现非一致性综合诊断,主要包括:

[0055] 1)通过MK趋势分析方法获取各时间序列的变化趋势,包括:使用标准正态分布进行假设检验,以MK统计量大于或小于95%置信水平的临界值判断时间序列的趋势变化;

[0056] MK趋势分析统计量的算法包括:

[0057] 
$$S = \sum_{i=1}^{n-1} \sum_{j=i+1}^n \text{sgn}(x_j - x_i),$$
 其中, $S$ 表示MK趋势检验的中间统计量, $i$ 和 $j$ 表示时间序列的序号, $n$ 为序数长度, $x_i$ 、 $x_j$ 为序列指定序号的径流值或水质值;

[0058] 其中, $x_i$ 、 $x_j$ 的数据关系表示为:

[0059] 
$$\text{sgn}(x_j - x_i) = \begin{cases} +1, & x_j - x_i > 0 \\ 0, & x_j - x_i = 0 \\ -1, & x_j - x_i < 0 \end{cases};$$

[0060] 进一步的, $\sigma^2 = \frac{n(n-1)(2n+5)}{18},$

$$[0061] \quad Z = \begin{cases} \frac{S-1}{\sigma}, S > 0 \\ 0, S = 0 \\ \frac{S+1}{\sigma}, S < 0 \end{cases}, \sigma^2 \text{表示} S \text{的方差}, Z \text{表示MK趋势检验的结果统计量}。$$

[0062] 2) 通过MK突变检验方法判断各时间序列的非一致性背景, 包括: 取95%的置信水平, 判断检验量曲线向前统计量 $UF_k$ 和向后统计量 $UB_k$ 在置信区间内是否相关相交, 如果相交, 则判断时间序列在交点处产生了显著突变, 呈现出非一致性;

[0063] 具体计算表示为:

$$[0064] \quad s_k = \sum_{i=1}^k r_i,$$

$$[0065] \quad r_i = \begin{cases} +1, x_i > x_j \\ 0, x_i \leq x_j \end{cases},$$

$$[0066] \quad E(s_k) = \frac{n(n+1)}{4},$$

$$[0067] \quad \text{var}(s_k) = \frac{n(n-1)(2n+5)}{72},$$

$$[0068] \quad UF_k = \frac{[s_k - E(s_k)]}{\sqrt{\text{var}(s_k)}},$$

$$[0069] \quad UB_k = -UF_k,$$

[0070] 其中,  $s_k$  表示MK突变检验的中间统计量,  $E(s_k)$  和  $\text{var}(s_k)$  分别表示序列的均值和方差,  $UF_k$  和  $UB_k$  分别表示MK突变检验的向前统计量和向后统计量。

[0071] 3) 通过相关性分析明确径流序列与水质序列间的相关程度, 本发明采用相关系数法进行相关性分析, 包括: 当相关性系数大于0时, 则表示两序列间存在正相关性, 反之则存在负相关性; 其绝对值越接近于1, 相关性程度越大。

[0072] 本专利采用Pearson、Kendall、Spearman三种不同的相关性系数对径流序列和多种水质序列进行相关性分析。

$$[0073] \quad \text{Pearson分析方法表示为: } r = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2 \sum_{i=1}^n (y_i - \bar{y})^2}},$$

[0074] Kendall方法表示为: 
$$T_{au} = \frac{C - D}{\frac{1}{2}n(n-1)},$$

[0075] Spearman方法表示为: 
$$\rho = \frac{\sum_{i=1}^n (R(x_i) - \bar{R}(x))(R(y_i) - \bar{R}(y))}{\sqrt{\sum_{i=1}^n (R(x_i) - \bar{R}(x))^2 \sum_{i=1}^n (R(y_i) - \bar{R}(y))^2}},$$

[0076] 式中,  $r$ 、 $T_{au}$  和  $\rho$  分别表示Pearson相关系数、Kendall相关系数和Spearman相关系数;  $x_i$  和  $y_i$  分别表示长度为  $n$  的时间序列,  $\bar{x}$  和  $\bar{y}$  分别表示其均值;  $C$  和  $D$  分别表示两个时间序列中有和没有一致性的元素对数;  $R(x_i)$  和  $R(y_i)$  分别表示第  $i$  个值在序列中的大小排序号,  $\bar{R}(x)$  和  $\bar{R}(y)$  分别表示序号均值。

[0077] 本发明提供了具体的案例:采用某地水文站2007年10月至2018年9月间的月尺度径流数据,及同期的4项月尺度水质数据:高锰酸钾指数(CODMn, mg/L)、溶氧量(DO, mg/L)、氨氮量(NH4N, mg/L)、PH值(PH)。

[0078] 首先,对案例数据进行非一致性检验,结果如图2所示,表明径流在实测期内多处存在非一致性变化,即非一致性背景现实存在。

[0079] 在非一致性检验过程中,对径流数据和4项水质数据进行相关性分析,结果如图3所示,表明径流序列与CODMn序列、NH4N序列、PH序列存在负相关性,与DO序列存在正相关性。

[0080] 步骤S110:采用多种概率分布模型,分别对所述径流时间序列和水质时间序列进行拟合,根据拟合优度检验确定最佳概率分布模型;

[0081] 在本步骤中,定义径流时间序列和各水质时间序列多种概率分布特征,具体包括:极值分布(EV)、指数分布(EXP)、伽马分布(GAM)、广义极值分布(GEV)、广义帕累托分布(GP)、对数正态分布(LOGN)、正态分布(NORM)、威布尔分布(WBL)等8种不同的概率分布;

[0082] 后续根据Kolmogorov-Smirnov (KS) 检验、均方根误差准则(RMSE)、Akaike information criterion (AIC信息准则)等3种不同的拟合优度检验确定各时间序列最优的概率分布模型。

[0083] 具体的, EV分布满足概率密度函数: 
$$f(x | \mu, \sigma) = \frac{1}{\sigma} e^{-\frac{x-\mu}{\sigma}} e^{-\frac{x-\mu}{\sigma}};$$

[0084] EXP分布满足概率密度函数: 
$$f(x | \mu) = \frac{1}{\mu} e^{-\frac{x}{\mu}};$$

[0085] GAM分布满足概率密度函数: 
$$f(x | a, b) = \frac{1}{b^a \Gamma(a)} x^{a-1} e^{-\frac{x}{b}};$$

[0086] GEV分布满足概率密度函数:  $f(x | \mu, \delta, \varepsilon) = e^{-(1+\varepsilon)\frac{x-\mu}{\delta} - \frac{1}{\varepsilon}}, 1 + \varepsilon \frac{x-\mu}{\delta} > 0;$

[0087] GP分布满足概率密度函数:  $f(x | \delta, \varepsilon, \mu) = \frac{1}{\delta} (1 + \varepsilon \frac{x-\mu}{\delta})^{(-\frac{1}{\varepsilon}-1)};$

[0088] LOGN分布满足概率密度函数:  $f(x | \mu, \sigma) = \frac{1}{x\sigma\sqrt{2\pi}} e^{-\frac{(\ln x - \mu)^2}{2\sigma^2}};$

[0089] NORM分布满足概率密度函数:  $f(x | \mu, \sigma) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}};$

[0090] WBL分布满足概率密度函数:  $f(x | a, b) = \frac{b}{a} (\frac{x}{a})^{b-1} e^{-(\frac{x}{a})^b}.$

[0091] 根据拟合优度检验确定最佳概率分布模型的依据包括:KS检验的统计量等于1时说明满足假设的分布特征;RMSE值越小,拟合程度越好;AIC值越小,拟合程度越好。

[0092] 具体的:

[0093] KS检验的统计量为  $D$ , 当其等于1时表示在给定的显著性水平上满足假设的分布特征, 反之则不满足, 表示为:  $D = \max |F_{\exp}(x) - F_{obs}|$ , 式中,  $F_{\exp}(x)$  表示序列的累计经验分布函数,  $F_{obs}$  表示经验频率;

[0094] RMSE值越小, 拟合程度越好, 表示为:  $RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (p - p_i)^2}$ , 式中,  $p$  表示经验频率,  $p_i$  表示理论频率;

[0095] AIC值越小, 拟合程度越好, 表示为:  $AIC = m \ln(\frac{1}{n} \sum_{i=1}^n (p - p_i)^2) + 2k$ , 式中,  $k$  表示模型参数个数。

[0096] 根据本发明提供的案例, 采用AIC信息准则、RMSE、KS检验等三种不同的拟合优度检验方法确定径流序列和4项水质序列的最佳概率分布模型, 分布结果如表1所示:

[0097] 表1

[0098]

| 序列    | 指标   | EV      | EXP     | GAM     | GEV      | GP      | LOGN     | NORM    | WBL     |
|-------|------|---------|---------|---------|----------|---------|----------|---------|---------|
| 径流    | AIC  | -569.29 | -928.56 | -891.22 | -815.66  | -971.28 | -864.48  | -671.19 | -901.74 |
|       | RMSE | 0.12    | 0.03    | 0.03    | 0.04     | 0.02    | 0.04     | 0.08    | 0.03    |
|       | KS   | 1       | 0       | 0       | 0        | 0       | 0        | 1       | 0       |
| CODMn | AIC  | -684.09 | -427.65 | -824.63 | -831.14  | -567.31 | -862.83  | -746.16 | -771.40 |
|       | RMSE | 0.07    | 0.20    | 0.04    | 0.04     | 0.11    | 0.04     | 0.06    | 0.05    |
|       | KS   | 1       | 1       | 0       | 0        | 1       | 0        | 0       | 0       |
| DO    | AIC  | -591.86 | -406.60 | -809.74 | -1004.29 | -461.23 | -890.95  | -698.97 | -709.46 |
|       | RMSE | 0.10    | 0.21    | 0.05    | 0.02     | 0.17    | 0.03     | 0.07    | 0.07    |
|       | KS   | 1       | 1       | 0       | 0        | 1       | 0        | 1       | 1       |
| NH4N  | AIC  | -552.41 | -598.28 | -849.94 | -1016.31 | -679.94 | -1033.19 | -638.26 | -799.02 |
|       | RMSE | 0.12    | 0.10    | 0.04    | 0.02     | 0.07    | 0.02     | 0.09    | 0.05    |
|       | KS   | 1       | 1       | 0       | 0        | 1       | 0        | 1       | 0       |
| PH    | AIC  | -542.57 | -332.55 | -420.92 | -378.54  | -315.17 | -413.96  | -437.23 | -495.58 |
|       | RMSE | 0.13    | 0.28    | 0.20    | 0.23     | 0.30    | 0.21     | 0.19    | 0.15    |
|       | KS   | 1       | 1       | 1       | 1        | 1       | 1        | 1       | 1       |

结果表明,径流序列的最优概率分布为GP、CODMn序列的最优概率分布为LOGN、DO序列的最优概率分布为GEV、NH4N序列的最优概率分布为LOGN、PH序列的最优概率分布为EV。

[0099] 步骤S120:构建AR模型,实现径流的随机模拟,输出模拟径流序列;对模拟径流序列进行调整,获取非一致性场景;

[0100] 由于径流具有自相关性和随机性,因此采用AR模型对径流进行随机模拟。

[0101] AR模型由随机扰动项 $\varepsilon_t$ 和时间序列的部分项目的加权之和组成:表示为:

$$[0102] \quad x_t = a_1 x_{t-1} + a_2 x_{t-2} + \cdots + a_p x_{t-p} + \varepsilon_t = \sum_{j=1}^p a_j x_{t-j} + \varepsilon_t, \text{其中, } x_t \text{ 为 } t \text{ 时刻}$$

的径流模拟值、 $p$  为自回归阶数、 $a_j$  为自回归系数、 $\varepsilon_t$  为噪声项;

[0103] 各时刻的径流模拟值构成径流模拟序列。

[0104] 获得径流模拟序列后,对模拟径流序列进行调整,生成多种非一致性场景。

[0105] 在本发明中,对模拟径流序列进行调整指:通过统计因子调整法和径流季节性变异模拟非一致性场景;非一致性场景包括:均值上升(场景1)、均值下降(场景2)、 $C_v$  上升(场景3)、 $C_v$  下降(场景4)、季节前移(场景5)、季节滞后(场景6);

[0106] 其中,统计因子调整法用于根据非一致性条件对序列统计特征参数的影响模拟非一致性场景,所述场景包括均值非一致性和 $C_v$  非一致性;

[0107] 生成均值的场景时,可以在保证其他参数不变的情况下,由AR模型得到的模拟序列在5%水平上放大或缩小,这样就得到均值非一致性场景下的径流输入: $Q_i^{mean} = KQ_i^{AR}$ ,其中, $Q_i^{mean}$  和 $Q_i^{AR}$  分别为均值非一致性场景下的径流序列和AR模型得到的模拟径流序

列,  $K$  为调整因子;

[0108] 生成  $C_v$  的场景时, 保证均值不变的情况下, 将由AR模型得到的模拟序列的  $C_v$  值在 5% 水平上放大或缩小, 这样就得到  $C_v$  非一致性场景下的径流输入, 表示为:

$Q_i^{C_v} = KQ_i^{AR} + (1-K)\bar{Q}$ , 其中,  $Q_i^{C_v}$  表示  $C_v$  非一致性场景下的径流序列;  $\bar{Q}$  为径流均值。

[0109] 径流季节性变异主要关注于径流时间序列的季节性变化, 用于体现一致性条件对径流季节性的影响; 本发明中主要生成前移非一致性和滞后非一致性的场景。

[0110] 具体的, 径流季节性变异通过累计时间分布函数 ( $TDF$ ) 进行时间尺度上的调整, 通过对季节性因子进行调整, 实现径流在季节非一致性条件下的重构模型, 重构模型表示为:

$$[0111] \quad TDF(j) = \frac{\sum_{i=1}^j Q_i}{\sum_{i=1}^n Q_i},$$

$$[0112] \quad TDF'(i) = TDF(i) - TDF^{tri1}(i) + TDF^{tri2}(i) \quad , i = 1, 2, \dots, n,$$

$$[0113] \quad Q_i' = \begin{cases} TDF'(i) \sum_{i=1}^n Q_i^{AR} & , i = 1, 2, \dots, n \\ (TDF'(i) - TDF'(i-1)) \sum_{i=1}^n Q_i^{AR} & , i = 1, 2, \dots, n' \end{cases}$$

[0114] 其中,  $TDF(j)$  为时间序列中第  $j$  个累计时间分布函数,  $n$  为时间序列长度,  $Q_i$  和  $Q_i'$  分别表示实测径流时间序列和季节非一致性场景下的径流时间序列,  $TDF'(i)$ 、 $TDF^{tri1}(i)$  和  $TDF^{tri2}(i)$  分别表示非一致性场景下、非一致性变化前和非一致性变化后三角形入流过程线的累计时间分布函数。

[0115] 对于本发明提供的案例, 利用AIC和贝叶斯信息准则确定出最大自回归阶数, 进而构建径流AR模型, 通过AR模型对径流序列进行模拟, 模拟序列长度设为100。在获得径流模拟序列后, 通过统计因子调整法和季节性变异法, 对模拟径流序列进行调整, 进而得到了6种径流的非一致性场景 (如图4所示)。其中, 场景1-4代表了径流序列统计特征 (均值、 $C_v$ ) 非一致性, 其变异幅度为5%; 场景5-6代表了径流季节非一致性, 分别代表了前移和滞后。

[0116] 六种非一致性场景构成后, 即可在二维联合Copula函数模型中生成水质的模拟序列。

[0117] 步骤S130: 构建二维联合Copula函数模型, 向所述二维联合Copula函数模型输入所述非一致性场景, 可输出水质的预测的模拟序列集;

[0118] Copula函数可以将多个变量的边缘分布和它们之间的相关性结构分开研究,且对边缘分布的类型没有任何限制,它既可以描述非线性相关关系,又可以描述非对称相关关系。Copula函数能基于变量间的复杂关系为水质预测提供一种灵活有效的模拟方法。

[0119] 本发明中采用的Copula函数包括Gumbel Copula、Clayton Copula和Frank Copula三种类型:

[0120] Gumbel Copula表示为:  $C(u, v) = e^{-[(-\ln u)^\theta + (-\ln v)^\theta]^{\frac{1}{\theta}}}$ ,

[0121] Clayton Copula表示为:  $C(u, v) = (u^{-\theta} + v^{-\theta} - 1)^{\frac{1}{\theta}}$ ,

[0122] Frank Copula表示为:  $C(u, v) = -\frac{1}{\theta} \ln[1 + \frac{(e^{-\theta u} - 1)(e^{-\theta v} - 1)}{e^{-\theta} - 1}]$ , 其中,  $\theta$  为

Copula函数的参数。

[0123] 基于Sklar定理,设  $H(\cdot, \cdot)$  为具有边缘分布  $F(\cdot)$  和  $G(\cdot)$  的联合分布函数,则存在一个Copula函数  $C(\cdot, \cdot)$  满足公式 (34), 若  $F(\cdot)$  和  $G(\cdot)$  连续,则  $C(\cdot, \cdot)$  就被唯一确定。

[0124] 因此,在本发明中,首先确定模拟所用的水质指标序列  $v$ ; 径流  $u$ ;

[0125] 确定径流和水质指标序列的边缘分布,表示为:  $H(u, v) = C(F(u), G(v))$ ;

[0126] 选择合适的Copula函数,描述径流和水质指标间的相关性结构。

[0127] 最后,基于最优的Copula函数和构建好的二维联合Copula函数模型,将非一致性场景输入到模型中,得出各项水质的模拟序列。

[0128] 在本发明提供的案例中,使用Copula函数构建非一致性二维水质联合随机模拟模型,利用3种不同的Copula函数,分别对径流数据和水质数据进行拟合,确定出Frank Copula为最优Copula函数。

[0129] 进一步的,将6种非一致性场景输入到构建好的Frank Copula模型中,可得到场景1至场景6的4项水质数据的模拟结果,如图5至图10所示。

[0130] 步骤S140: 获取实测期内实测径流,向二维联合Copula函数模型,输入实测径流,得到水质的模拟序列集,提取模拟序列集中实测期水质点模拟值和实测值之间的残差集合抽取残差样本集,根据残差样本集实现水质指标区间预测。

[0131] 具体的,首先,通过Copula函数模拟得出实测期各水质数据的点模拟值,并计算与实测值的残差样本集合;

[0132] 接下来,通过Bootstrap方法,对实测期水质点模拟值和实测值之间的残差集合有放回的抽取  $N$  次,得到  $N$  组新的残差样本集。具体的,Bootstrap方法不需要对总体分布做出假设,仅基于原始样本序列进行再抽样,并不断预测抽出样本的参数值,进而推断出未知总体的参数特征并可定量描述参数预测的不确定性。对于实测残差样本集  $\{X_i, i = 1, 2, \dots, n\}$ , 设  $X_i \sim F(x)$ , 其总体分布参数为  $\theta(F_n)$ , Bootstrap方法可表示为:

[0133]  $R_n^{*(j)} = \hat{\theta}(F_n^{*(j)}) - \theta(F_n), \quad j = 1, 2, \dots, N,$

$$[0134] \quad \theta^{(j)}(F) = \hat{\theta}(F_n) - T_n \cong \hat{\theta}(F_n) - R_n^{*(j)}, \quad j=1,2,\dots,N,$$

[0135] 式中,  $R_n^{*(j)}$  表示总体参数预测和总体分布参数差值间Bootstrap统计量;  $\hat{\theta}(F_n^{*(j)})$  和  $\hat{\theta}(F_n)$  表示第  $j$  个再生样本和总体的参数预测;  $\theta^{(j)}(F)$  表示对于每一个  $R_n^{*(j)}$  所得到的参数预测近似值。

[0136] 最后,根据所述残差样本集实现水质指标区间预测:即分别在95%和90%的置信水平上对新样本集进行抽样,得出各水质模拟序列的区间预测。具体的,对  $N$  组新的残差样本集分别在95%和90%的置信水平上再次抽样,进而得出各水质序列在95%和90%的置信水平上模拟值的置信上界和下界,最终可得到各水质模拟值的区间预测。

[0137] 在本发明案例中,首先通过构建好的二维Copula水质模拟模型,做出实测期4项水质数据的模拟序列,进而得到水质点模拟值和实测值之间的残差集合。再通过Bootstrap方法,将残差集合有放回的抽取1000次,得到了1000组新的残差样本集。取95%和90%的置信水平,对1000组再生残差样本进行抽样,得到了对应置信水平下的水质模拟值残差的上界和下界,进而得到水质模拟的区间预测,预测结果如表2所示:

[0138] 表2

[0139]

| 置信水平  | CODMn (mg/L) | DO (mg/L) | NH4N (mg/L) | PH    |
|-------|--------------|-----------|-------------|-------|
| 95%上界 | 0.13         | 0.40      | 0.23        | 0.10  |
| 95%下界 | -0.38        | -0.14     | -0.26       | -0.26 |
| 90%上界 | 0.09         | 0.38      | 0.20        | 0.08  |
| 90%下界 | -0.33        | -0.09     | -0.22       | -0.23 |

[0140] 其中,以95%和90%置信水平下,场景1中4项水质数据模拟的点预测和区间预测结果如图11和图12所示。

[0141] 针对目前常用的水质预测的技术方案,本发明充分考虑了非一致性条件对水质模拟的影响,并且给出了具体场景下的量化模拟结果。在模拟环境中,采用非一致性场景模拟,使用统计因子调整和季节性变异得出了具有代表性的6类不同的非一致性场景,并将这些非一致性场景作为水质模拟的输入序列,克服了以往研究中不考虑场景模拟直接进行后续的水质模拟的基础性不足;进一步的,本发明将非一致性条件作为水质模拟的主要考虑对象,将6种非一致性场景数据作为输入,通过二维水质模拟模型进而得到了非一致性条件下的水质变化情况,并与未产生非一致性时的模拟序列进行对比,改进了以往研究中直接使用历史数据或与历史数据具有一致性的数据作为输入进行模拟的方法,克服了以往的方案使结果不能代表实际情况的缺陷;另一方面,本发明充分考虑了水质样本代表性中的不确定性,并以点预测和区间预测的形式进行量化,通过Bootstrap方法对实际残差样本集进行重抽样,并将获得的再生残差样本叠加在水质点预测上,得出了水质的区间预测,相对传统技术方案中不考虑非一致性场景的水质预测,更加提高了预测结果的准确性。

[0142] 以上公开的仅为本发明的几个具体实施例,但是,本发明并非局限于此,任何本领域的技术人员能思之的变化都应落入本发明的保护范围。

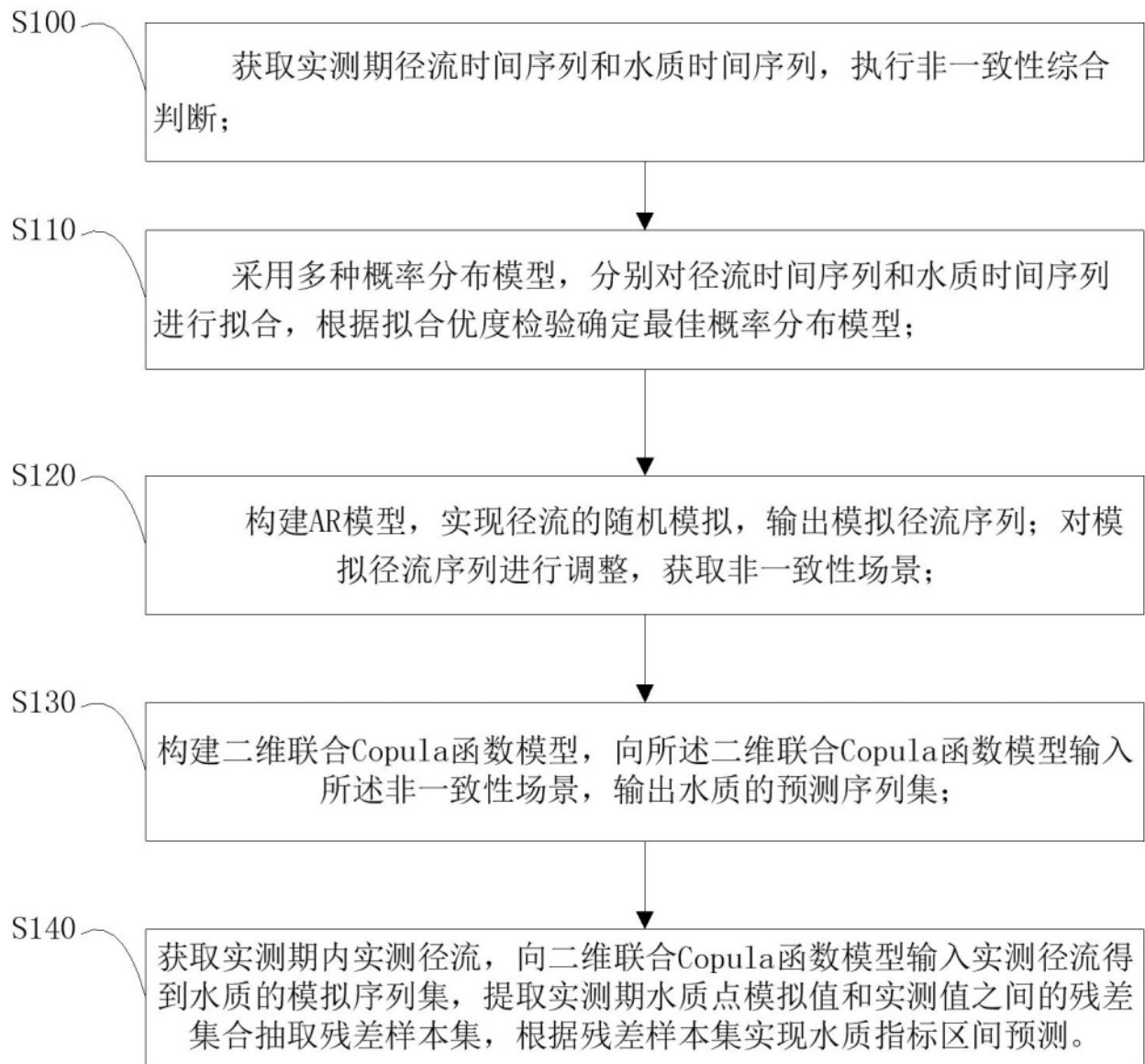


图1

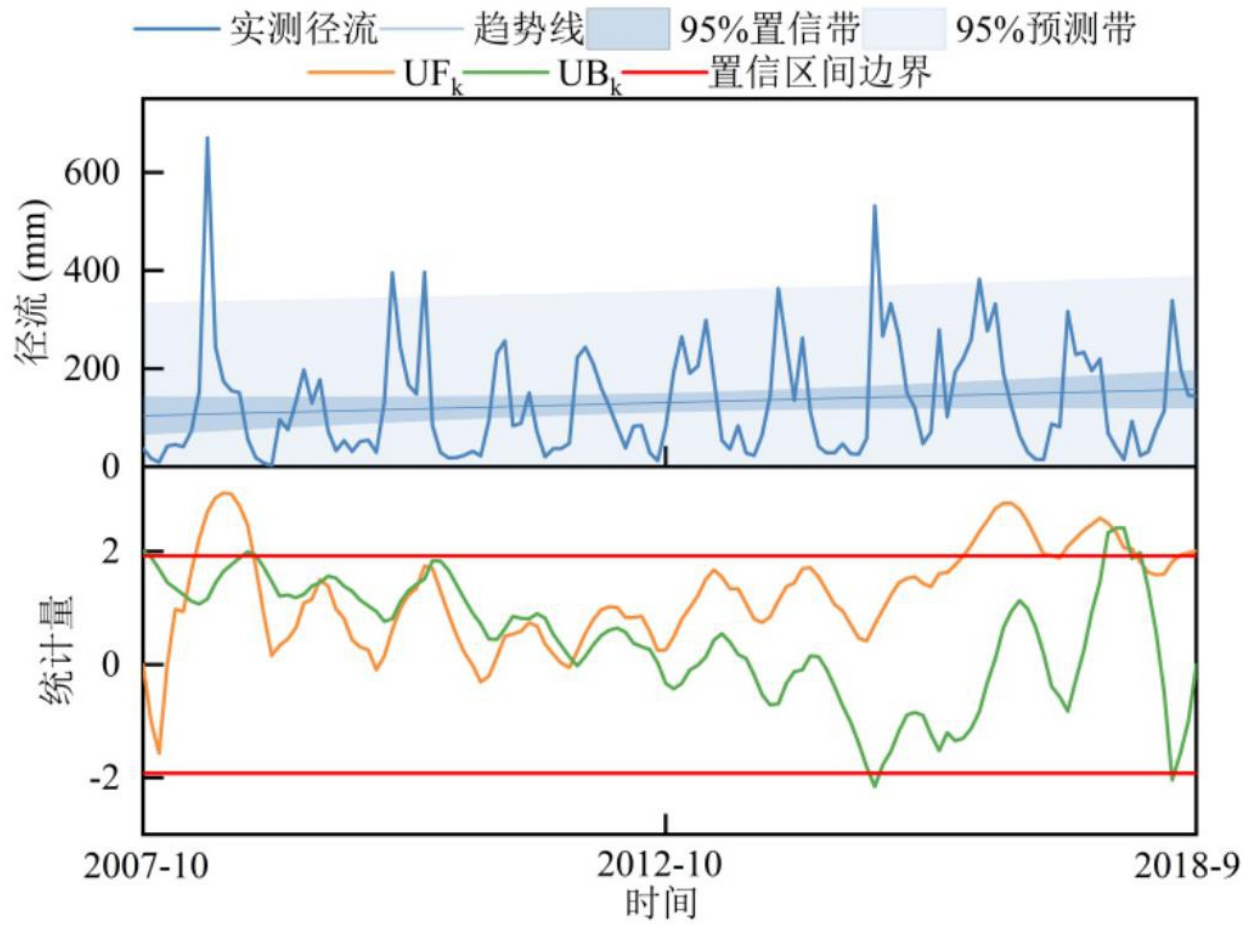


图2

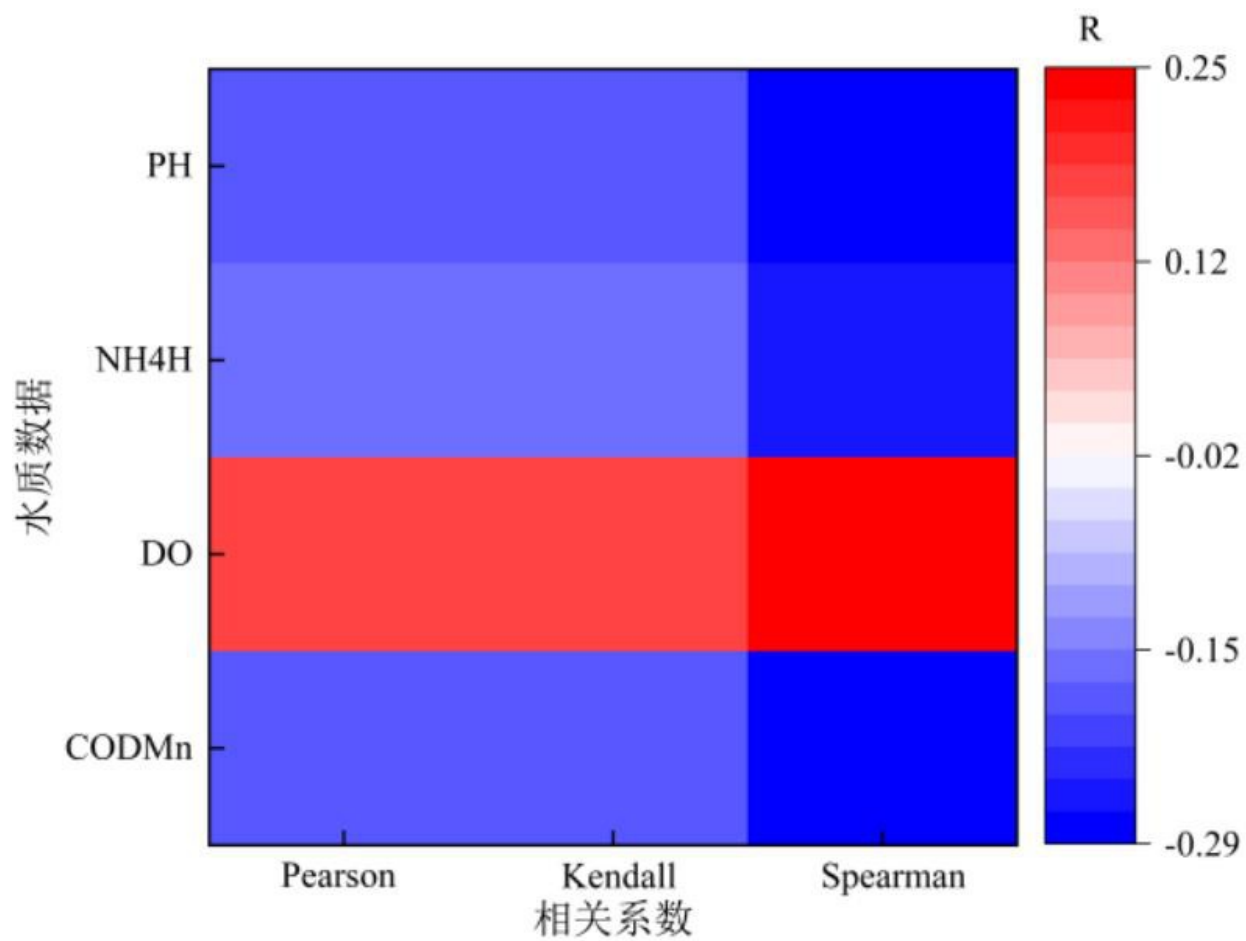


图3

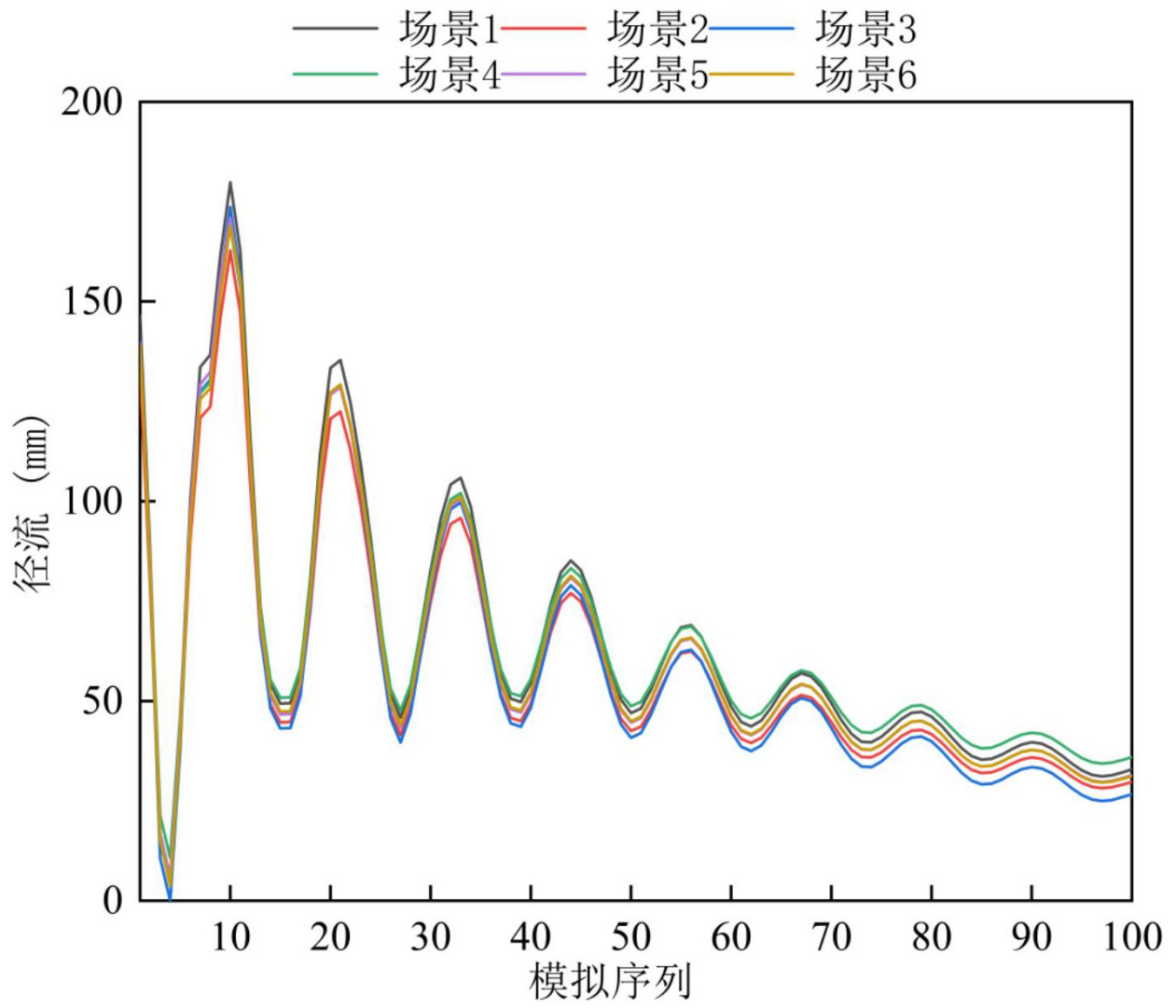


图4

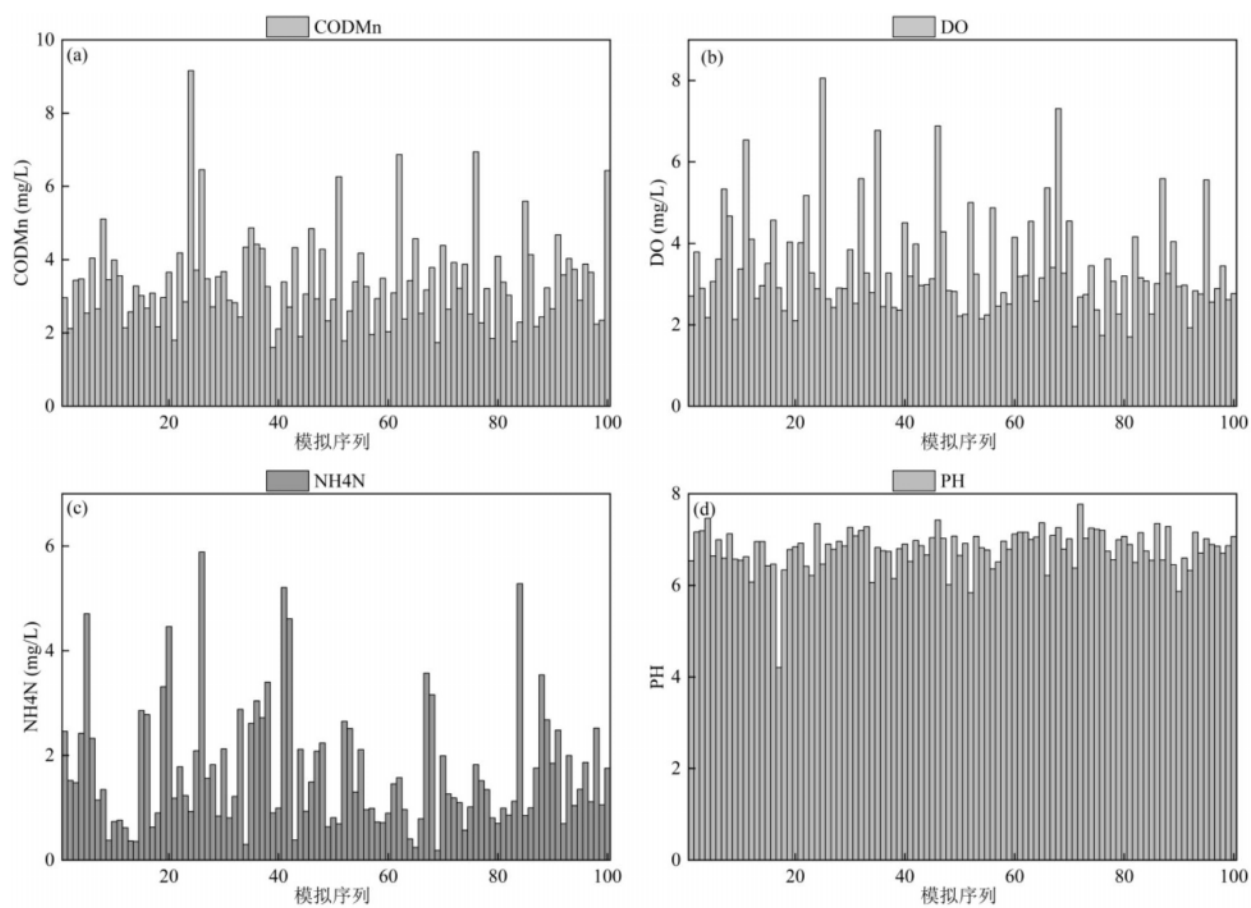


图5

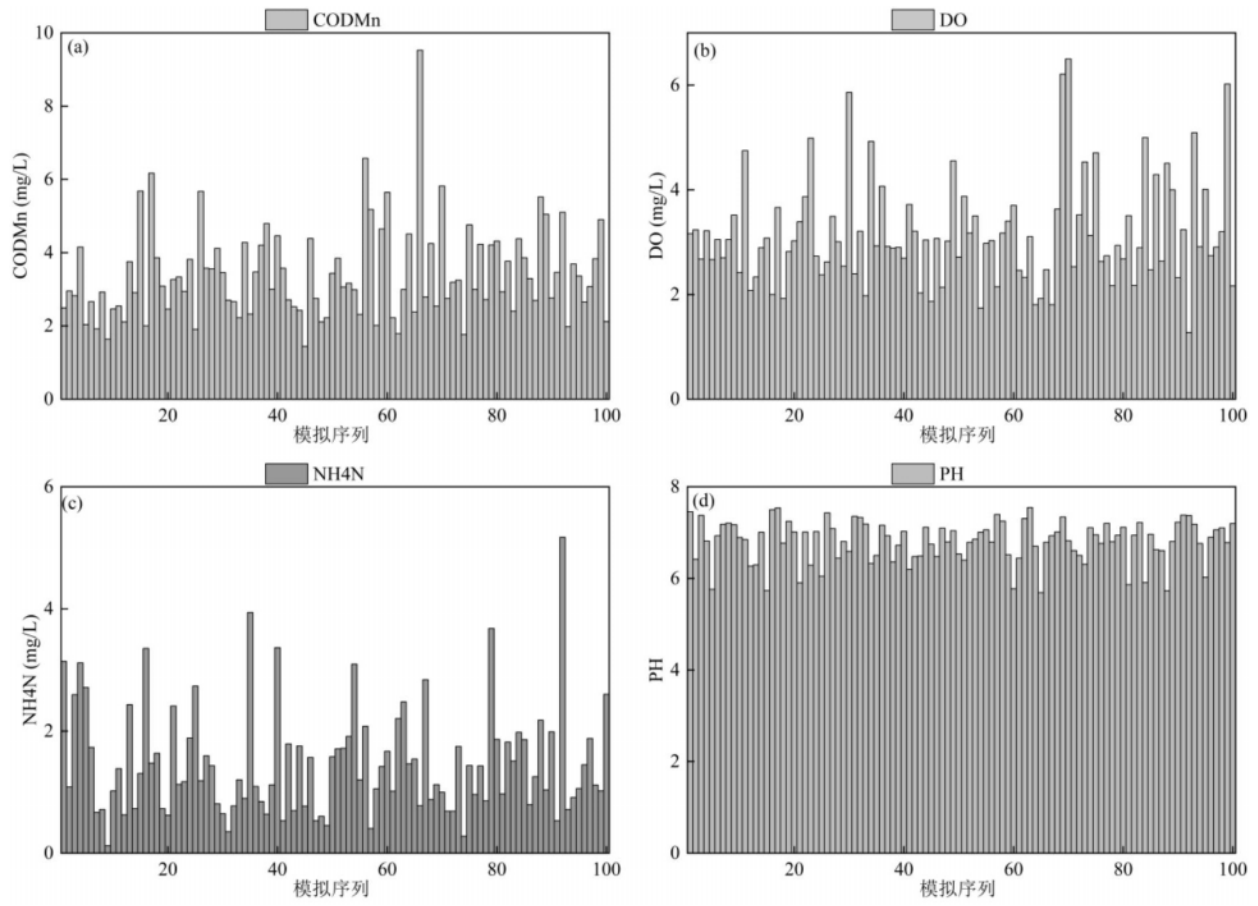


图6

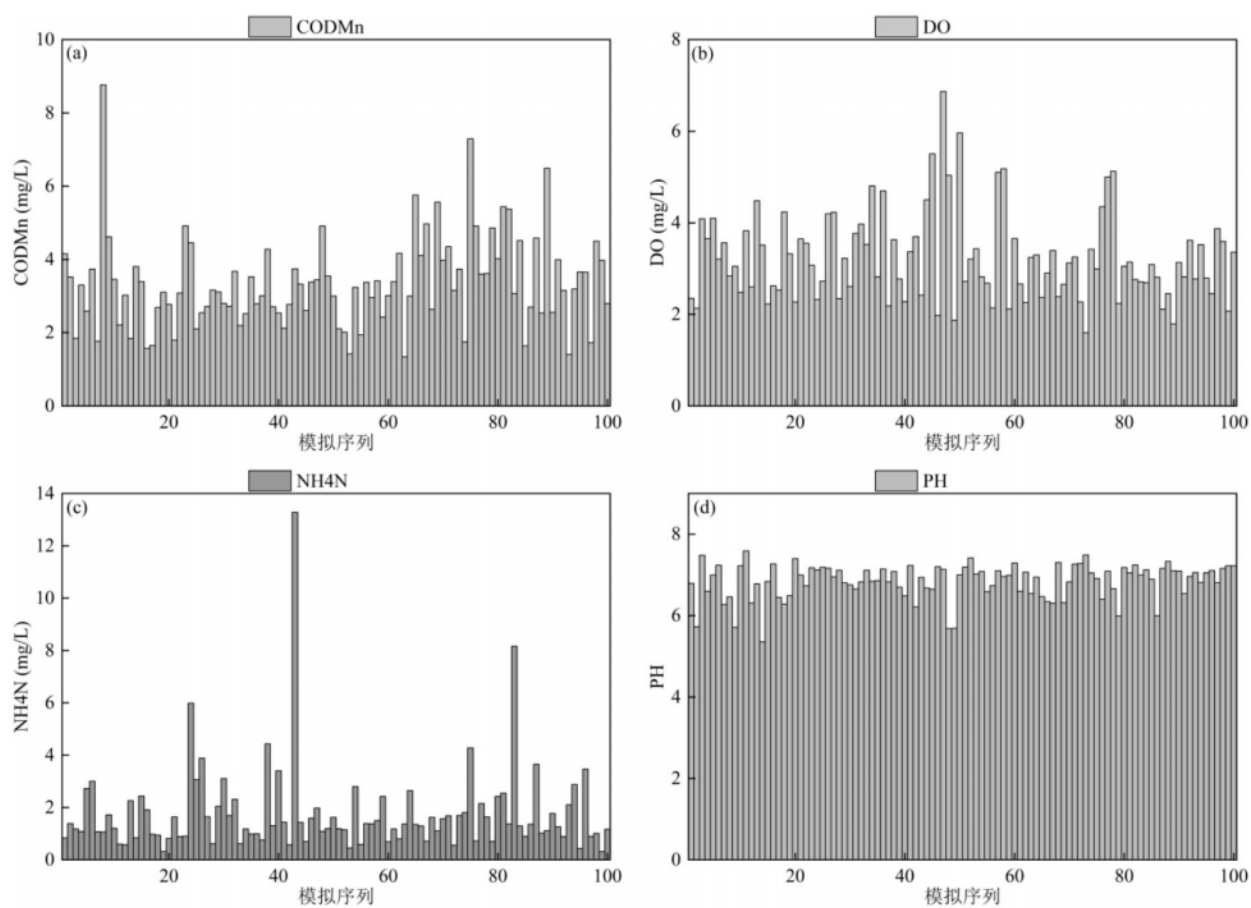


图7

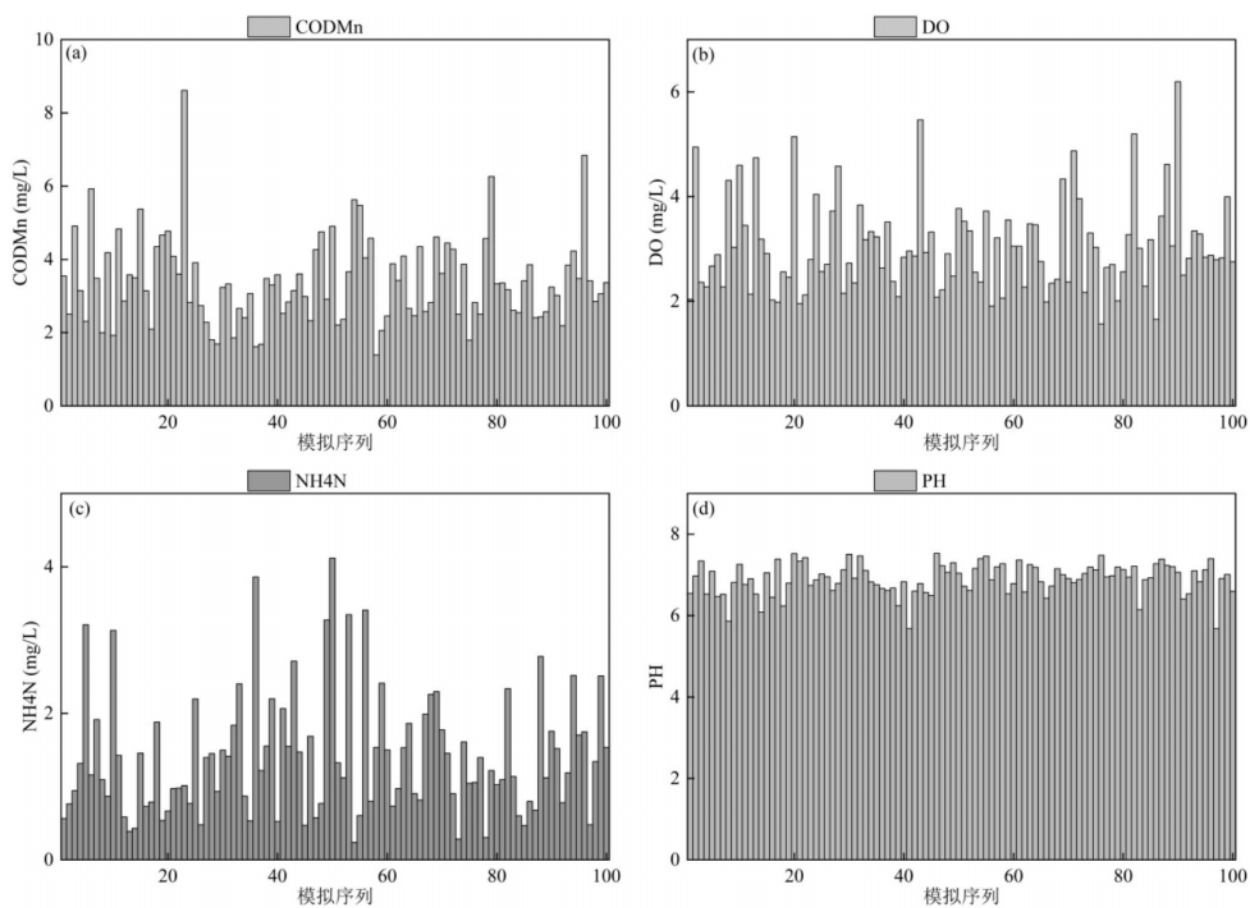


图8

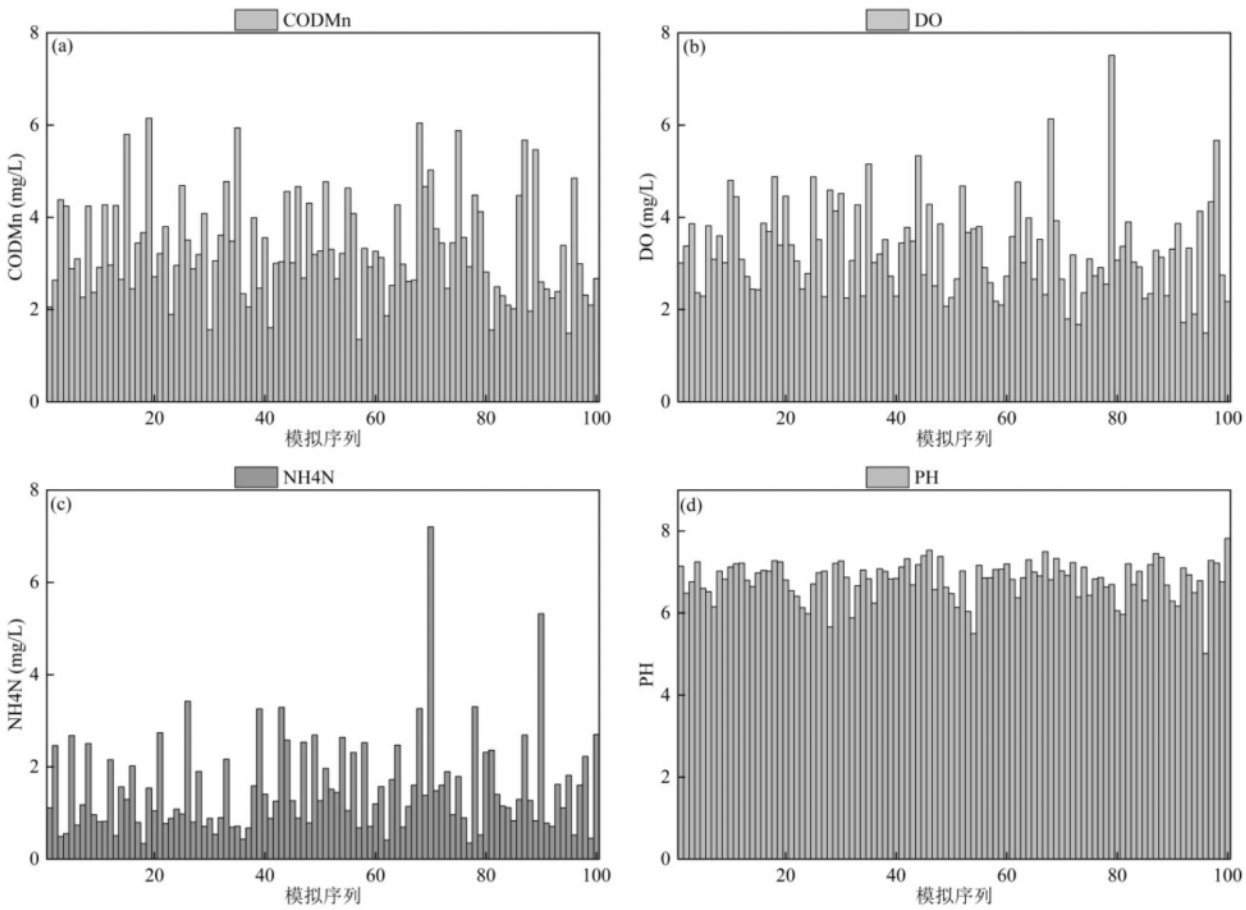


图9

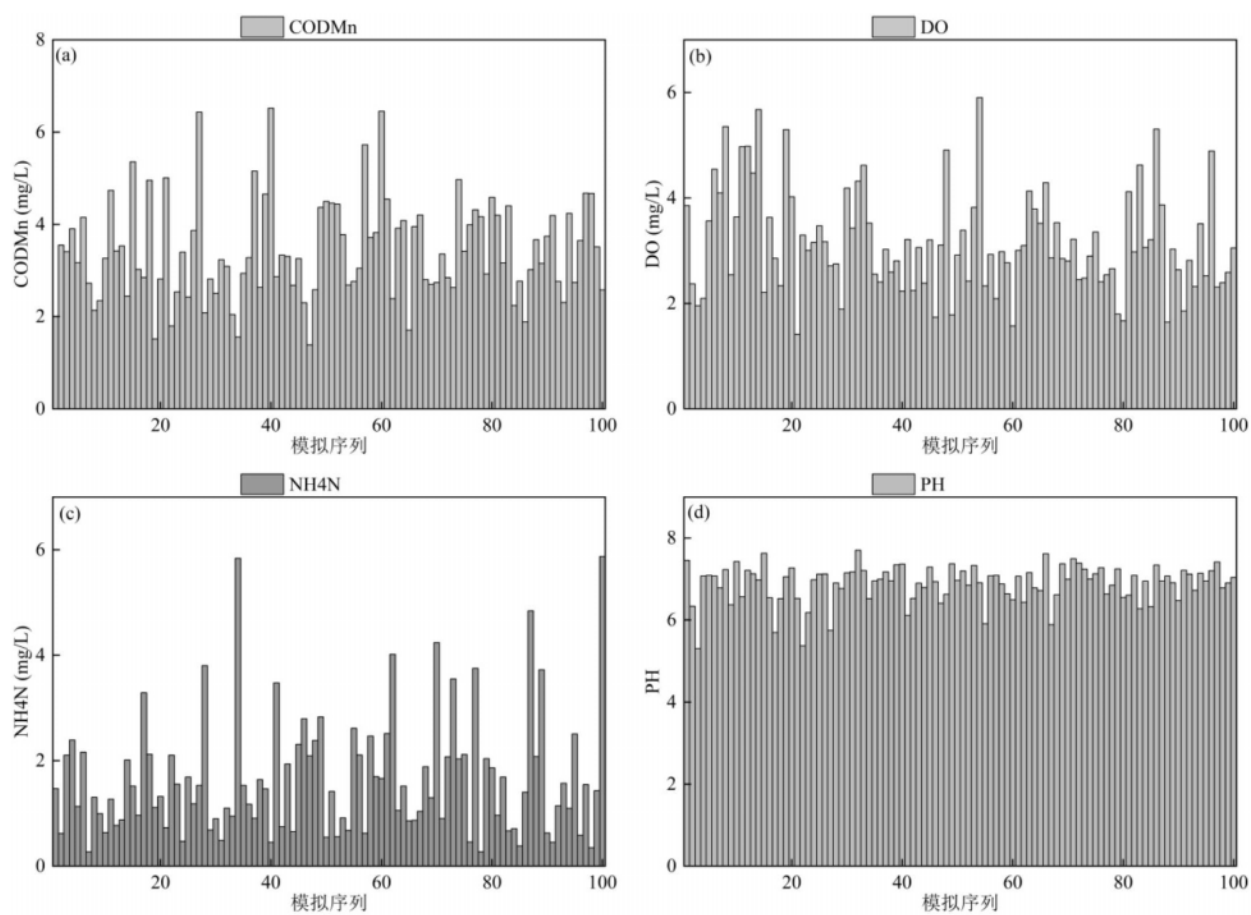


图10

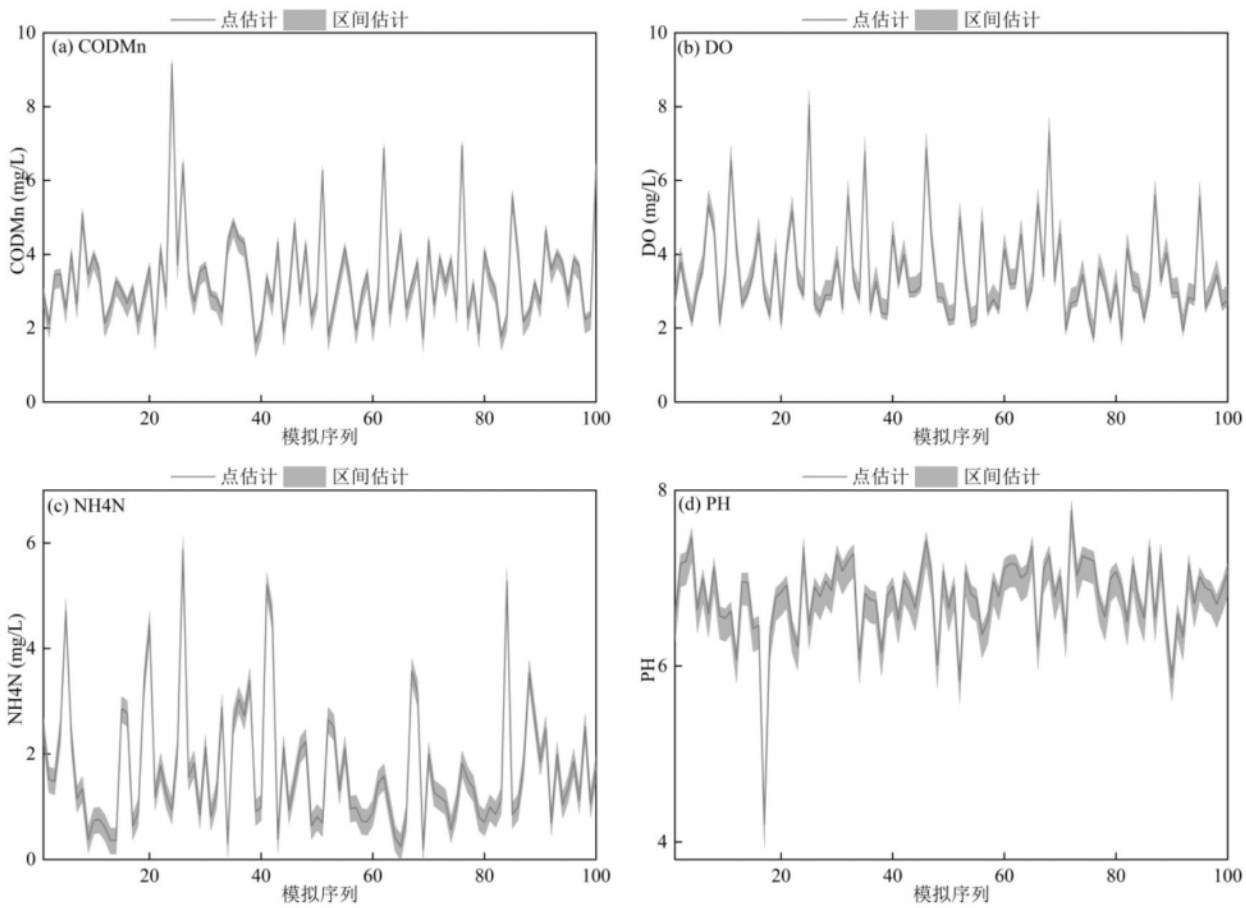


图11

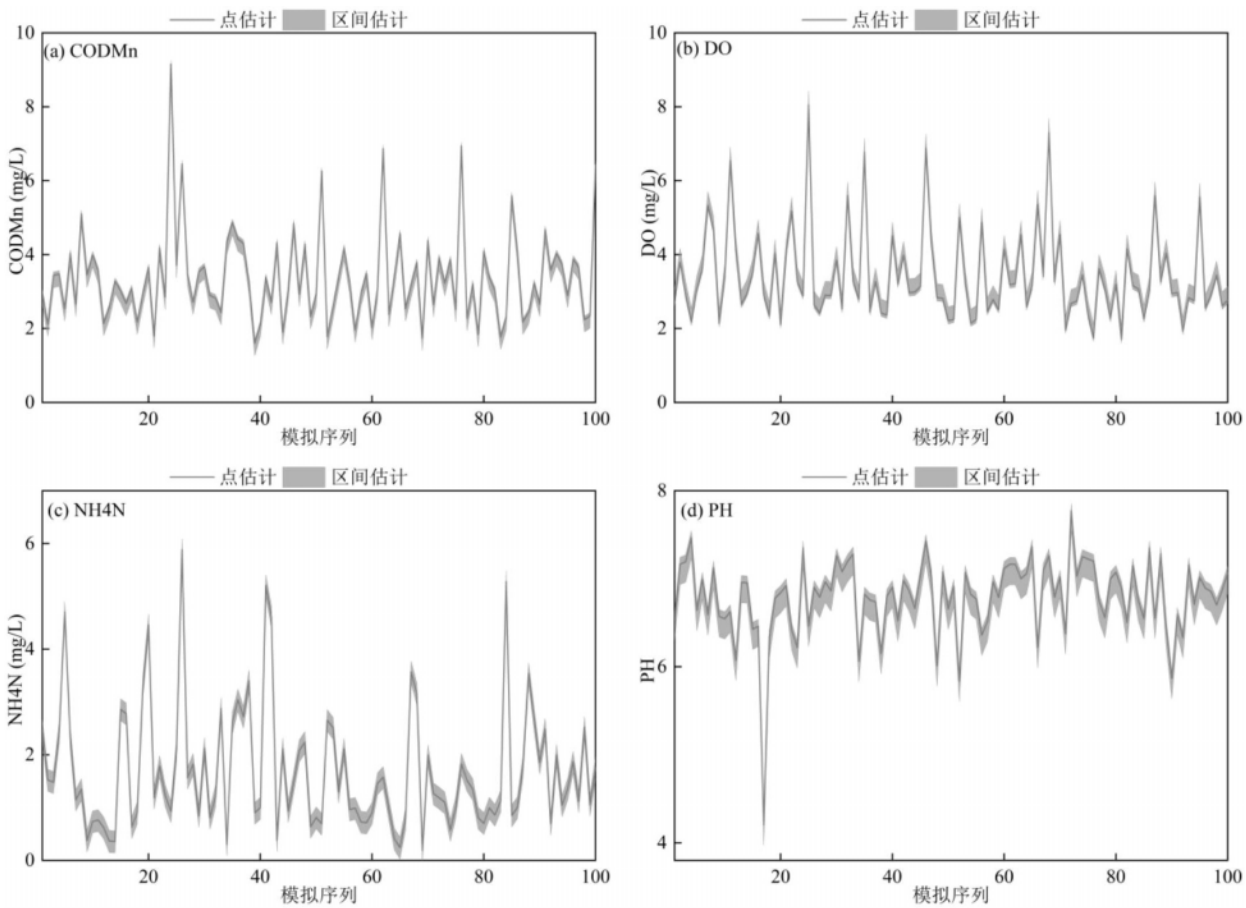


图12