

DOI: 10. 20040/j. cnki. 1000-7709. 2023. 20222365

基于贝叶斯层次分析法的沙颍河生态流量研究

赵 焯¹, 张 翔¹, 邓梁堃¹, 付 湘¹, 胡 俊²

(1. 武汉大学水资源与水电工程科学国家重点实验室, 湖北 武汉 430072;
2. 水利部中国科学院水工程生态研究所, 湖北 武汉 430079)

摘要: 维持河道内生态流量是协调水资源开发利用与河流生态保护的重要举措, 现有的生态流量计算方法大多受限于生态数据不足、生态建模困难。为此, 建立了基于浮游植物生物量的贝叶斯层次模型, 研究探讨了河流水文—生态响应关系, 以沙颍河为例, 分析计算了汛前、汛期、汛后的浮游植物生物量与流量、水温的关系。结果表明, 贝叶斯层次模型可识别水文—生态的复杂响应关系, 分层分析产生的优势借用提高了短系列数据的可用性; 生物量—流量关系参数 β 汛前大于0的概率为0.683, 汛期小于0的概率为0.577, 汛前河道内流量的保证及汛期泄流的控制对浮游植物生物量有正向影响; 在符合设计要求的生态流量条件下, 预测生物量最大和最小改善分别发生在河段7(汛前+32.05%, 汛期+871.80%, 汛后+81.79%) 和河段3(汛前+3.91%, 汛期+1.16%, 汛后+2.89%)。

关键词: 贝叶斯层次模型; 水文—生态响应关系; 生态流量; 浮游植物生物量; 沙颍河

中图分类号: X143; P333; TV214

文献标志码: A

文章编号: 1000-7709(2023)11-0048-05

1 引言

经济社会的发展需求促使人类对自然资源的开发利用不断增强, 闸坝、水库等工程的修建极大地改变了河流的水文情势, 并被认为是诸多水生态问题产生的主要原因之一^[1]。生态流量概念的提出正是为了协调水资源开发利用与河流生态保护之间的矛盾^[2]。常见的生态流量确定方法包括水文学法、水力学法、生物栖息地法。水文学法基于水文过程, 对生态考虑较少^[3]; 水力学法和生物栖息地法通过中间媒介构建水文—生态关系, 缺少对生态的直接反应^[3]。而诸如流量等水文要素与生物量、物种分布等生物要素相互作用, 构成了复杂的水文—水生态系统。贝叶斯层次模型形成的主要原因即为数据结构复杂, 这使得它在描述水文—生态关系中复杂且多层次未知的因果关系时非常奏效^[4]。此外, 贝叶斯层次分析法可利用时空数据之间的相关性, 让空间(时间)邻近的因变量共用随机效应来弥补生态数据不足的缺陷^[5], 这使得贝叶斯层次模型在生态学中的应用

愈加广泛: CHEN W 等^[6]利用贝叶斯理论分析了流量对本地鱼类与入侵鱼类的影响; WEBB J A 等^[7]利用贝叶斯法发现了胡瓜鱼丰度与夏季流量的相关关系。然而, 国内少有将其应用于探究水文—生态响应关系, 且汛期非汛期的差异性也并未被纳入考虑。为此, 本文基于贝叶斯层次分析法构建了水文—生态模型, 研究了沙颍河流域内浮游植物对河道内流量、水温变化的动态响应, 为生态流量的确定及河流健康评估与管理提供了科学依据。

2 研究方法

2.1 贝叶斯层次方法

2.1.1 贝叶斯层次理论

设 Y_1, Y_2, \dots, Y_n 为来自总体 Y 的样本, Y_1, Y_2, \dots, Y_n 的联合概率密度函数 $f(Y, \theta)$ 为:

$$f(Y, \theta) = f(Y_1, Y_2, \dots, Y_n, \theta) \quad (1)$$

式中, θ 为未知参数。

当产生具体样本 $y = (y_1, y_2, \dots, y_n)$ 后, $y = (y_1, y_2, \dots, y_n)$ 的联合概率密度函数即为在给定

收稿日期: 2022-11-10, **修回日期:** 2023-02-14

基金项目: 国家自然科学基金重点项目(U21A2002)

作者简介: 赵焯(1998-), 女, 博士研究生, 研究方向为生态水文, E-mail: zhaoyeah@whu.edu.cn

通讯作者: 张翔(1969-), 男, 博士、教授、博导, 研究方向为生态水文、城市水文, E-mail: zhangxiang@whu.edu.cn

θ 的条件概率密度函数 $f(y|\theta)$:

$$f(y|\theta) = f(y_1, y_2, \dots, y_n, \theta) \quad (2)$$

式(2)为似然函数,其综合了总体信息和样本信息。贝叶斯方法将未知参数(θ)视为具有先验分布特征的随机变量,假设其概率密度函数为 $P(\theta)$,则样本 y 和参数 θ 的联合概率密度函数 $P(y, \theta)$ 为:

$$P(y, \theta) = f(y|\theta)P(\theta) \quad (3)$$

在给定一组观测值 y 的情况下,可根据 $P(y, \theta)$ 来推断参数 θ 。令

$$P(y) = \int_{\Theta} P(y, \theta) d\theta = \int_{\Theta} f(y|\theta)P(\theta) d\theta \quad (4)$$

式中, Θ 为参数 θ 的取值范围; $P(y)$ 为样本 y 的边缘概率密度函数。

进而,样本 y 和参数 θ 的联合概率密度函数 $P(y, \theta)$ 可表示为:

$$P(y, \theta) = P(\theta|y)P(y) \quad (5)$$

其中

$$P(\theta|y) = \frac{P(y, \theta)}{P(y)} = \frac{f(y|\theta)P(\theta)}{\int_{\Theta} f(y|\theta)P(\theta) d\theta} \quad (6)$$

式中, $P(\theta|y)$ 为给定样本观测值 y 的情况下参数 θ 的条件概率密度函数。

$\int_{\Theta} f(y|\theta)P(\theta) d\theta$ 与参数 θ 无关,因此式(6)可记为:

$$P(\theta|y) \propto f(y|\theta)P(\theta) \quad (7)$$

假设先验分布 $P(\theta)$ 受参数 λ 的影响,在贝叶斯理论中, λ 也可被视为是符合某种分布的随机变量。参数 θ 先验分布涉及的参数 λ 的分布参数被称为超参数(ω)^[8],其分布被称为超分布^[9]。根据式(7)贝叶斯层次模型中的后验分布可被记作:

$$f(\theta|y) \propto f(y|\theta)f(\theta|\lambda)f(\lambda|\omega) \quad (8)$$

与未知参数分布相关的参数也服从某一分布,这很自然的导致贝叶斯模型中的某些参数处于不同层级^[9],这样的模型被称为贝叶斯层次模型(图 1)。

2.1.2 先验分布

先验分布的指定直接影响着参数的后验结果,一般是基于认识主体丰富的经验和对事件的了解程度而确定的,尤其需要关注先验期望和先验方差。在生态学的应用中,常常无法获得可用的先验信息,这种情况下需指定一个不会在很大程度上影响后验结果的先验分布以充分采纳样本数据信息,这种分布被称为“无信息分布”或“模糊分布”,常用的是设定 $\theta \sim N(\bar{\theta}, \sigma^2)$ ^[6]。

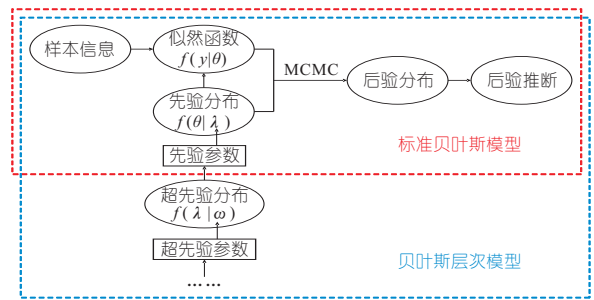


图 1 标准贝叶斯、层次贝叶斯模型示意图

Fig. 1 Schematic diagram of the standard and hierarchical Bayesian model

2.1.3 后验分布

马尔科夫链蒙特卡罗方法(MCMC)将贝叶斯推断过程中复杂的积分计算分解为一些依赖于模型中每个参数条件分布的简单问题,有效解决了贝叶斯方法中后验分布计算困难的问题。MCMC以马尔科夫链 $\{\theta^{(1)}, \theta^{(2)}, \dots, \theta^{(T)}\}$ 为基础,马尔科夫链满足:

$$P(\theta^{(t+1)} | \theta^{(t)}, \dots, \theta^{(1)}) = P(\theta^{(t+1)} | \theta^{(t)}) \quad (9)$$

即在给定 $t+1$ 时刻前所有的 θ 值时, $t+1$ 时刻 θ 的条件概率分布仅仅依赖 t 时刻 θ 的值。为从后验分布 $P(\theta|y)$ 中产生样本,需先构建马尔科夫链,且要保证构建的马尔科夫链的平衡分布即为 $P(\theta|y)$ 的后验分布。构建完成后,赋予马尔科夫链运行初始值,生成样本值直至到达平衡状态。

2.2 基于贝叶斯层次法的水文—生态响应建模

2.2.1 指标选取

浮游植物是水生态系统的重要组成部分,它可通过光合作用固碳释氧增加水体溶氧量,抑制有害菌生长,减轻水体污染;然而,单一浮游植物的泛滥也会引起河湖生态功能的下降。因此,保证浮游植物生物量处于合理区间内对保护河流生态多样性、维系河流生态健康十分重要。

大量研究表明,浮游植物的生长繁殖与流量^[10]及水温^[11]密切相关。流量限制着浮游植物的分布,包括流速和剪应力对浮游植物种类的影响,还包括高低流量、水位变化对河道植物种类的影响。水温决定浮游植物的萌发和休眠,影响浮游植物的代谢。考虑到上述情况,可根据浮游植物对流量及水温的预期反应制定一个简单的概念模型。假设浮游植物生物量是河道内流量及水温的函数,非汛期的河道内流量的保证一定程度上考虑了维持水生生物量的作用。此外,由于浮游植物的部分流动性和群落性,其丰度往往表现出较大的空间差异,在河段尺度上更能检测到它对水文变异的响应。为此需在河段尺度内组合数

据,这与贝叶斯层次模型的第二层次结构相吻合。

2.2.2 层次模型构建

根据研究区域水文水力特性划分河段,对研究区域内水文、水生态指标进行多次采样,建立基于贝叶斯层次分析的水文—生态响应模型:

$$\begin{cases} y_{ij} \sim N(\theta_{ij}, \tau_c) \\ \theta_{ij} = \alpha_i + \beta_i Q_j + \gamma_i T_j \end{cases} \quad (10)$$

式中, y_{ij} 为每个河段(j)内每个站点(i)的浮游植物生物量样本数据; Q_j 、 T_j 分别为采样当天河段的流量、水温; θ_{ij} 为生物量样本数据的平均值, τ_c 为该分布的方差,假设其服从经典的贝叶斯分析中常用的形状参数为 0.001、尺度参数为 0.001 的 gamma 分布 $G(0.001, 0.001)$; α_i 为考虑流量和温度影响后的模拟生物量样本均值的截距; β_i 、 γ_i 分别为生物量与流量、水温的关系系数。

α_i 、 β_i 、 γ_i 3 个参数在河段层面建模:

$$\begin{cases} \alpha_i \sim N(\alpha_c, \tau_\alpha) \\ \beta_i \sim N(\beta_c, \tau_\beta) \\ \gamma_i \sim N(\gamma_c, \tau_\gamma) \end{cases} \quad (11)$$

式中, α_c 、 β_c 、 γ_c 、 τ_α 、 τ_β 、 τ_γ 分别为 α_i 、 β_i 、 γ_i 3 个参数的均值和方差,因为对这 6 个参数效应的方向和强度无确切的先验预期,所以这 6 个超参数被赋予模糊先验分布:(0,10 000)区间上的无信息先验均匀分布,6 个参数在河流层次建模(图 2)。

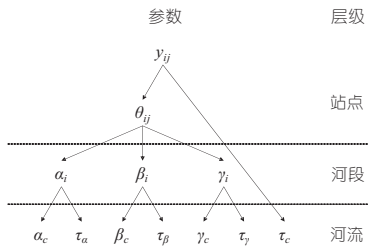


图 2 贝叶斯层次模型层级结构

Fig. 2 Bayesian hierarchical model hierarchy

2.2.3 后验分布计算

结合采样信息及先验分布利用 MCMC 计算站点、河段、河流层面的参数后验分布 ($\theta^{(i)}$, $\alpha^{(i)}$, $\beta^{(i)}$, $\gamma^{(i)}$, $\alpha_c^{(i)}$, $\tau_\alpha^{(i)}$, $\beta_c^{(i)}$, $\tau_\beta^{(i)}$, $\gamma_c^{(i)}$, $\tau_\gamma^{(i)}$, $\tau_c^{(i)}$, $i = 1,$

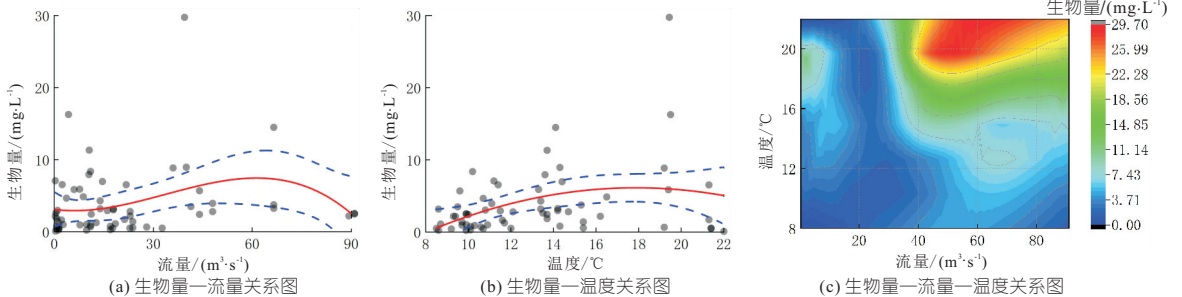


图 4 生物量—流量,生物量—温度关系

Fig. 4 Biomass-flow, biomass-temperature relationships

$2, \dots, I_m$), I_m 为迭代产生的样本个数。在足够长的预热期后^[12],利用 GELMAN A 等^[13]提出的 PSRF(Potential Scale Reduction Factor)判断马尔科夫链的收敛性,直到变量对应的 PSRF 小于 1.2 后认为样本已收敛。

3 实例研究

3.1 样本分析研究

沙颍河发源于河南省登封县高山,径流量年际变化大、年内分配不均、河流生态脆弱(图 3)。综合沙颍河水文水力特性、河道开发建设情况、河道外取水行为,将其分为 7 个河段进行研究。沙颍河 2013 年汛前(4 月 17 日)、汛期(9 月 13 日)、汛后(11 月 10 日)3 个时段 7 个河段 21 个采样点浮游植物的采样数据,流量及水温数据来自 ECMWF ERA5-Land 数据库发现,汛前、汛期和汛后流域内水库的蓄放导致流量数据分散并间接影响到生物量,因此后续分析将对汛前、汛期、汛后的水文—生态关系分别建模分类探讨。

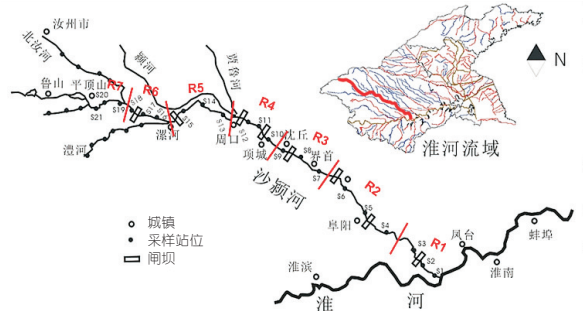


图 3 沙颍河采样点位及分段概化图

Fig. 3 Generalized map of sampling points and segments of the Shaying River

图 4(a)、(b)为生物量与流量、温度的关系散点图,该散点图拟合了非参数广义相加模型曲线及 5%和 95%的置信区间。由图 4(a)、(b)可看出,流量的增大和温度的升高在某一范围内(60 m³/s, 18 °C)对浮游植物生物量的增加起正向推动作用,考虑到沙颍河流域浮游植物物种丰度较

高,且并无单一物种泛滥趋势,浮游植物的释氧抑菌作用相比其呼吸耗氧作用明显,因此在低温低流量的汛前及汛后期保证河道内的生态流量对维持浮游植物生长繁殖是十分必要的;而在高温多水的汛期,合理控制河道内流量以保证浮游植物不被冲刷对维持河道生态系统稳定也很有意义。从生物量与流量、温度的等高线图(图 4(c))可看出,在 18~22 °C、50~80 m³/s 的环境条件下浮游植物生物量最大,超过 26 mg/L。

3.2 沙颍河水文—生态响应关系

按照本文的建模方法建立沙颍河水文—生态贝叶斯层次模型,利用 MCMC 分汛前、汛期、汛后三个时间段各产生 7 000(7 个河段×1 000)个样本。为探究沙颍河水文—生态响应关系,本文主要关注的参数为 β 与 γ ,即生物量与流量及温度的关系参数,表 1 为 7 个河段在汛前、汛期及汛后的后验参数 β 、 γ 均值偏差及蒙特卡罗误差。分析各河段汛期与非汛期 β 、 γ 的分布图,发现参数 β 在非汛期除 R2 河段汛后 β 的均值略小于 0 外,其他所有河段非汛期的 β 的均值都大于 0,且蒙特卡罗误差均在 0.007 以内,这表明非汛期水生生物量的增加与河道内流量的增加有关,这也体现在非汛期 β_c 的分布上(表 2)(汛前大于 0 的概率为 0.683,汛后大于 0 的概率为 0.471)。对于汛期,所有河段 β 的均值都小于 0,这表明汛期较大的流量对水生生物的生长繁殖不利,这一结论与非参数广义相加模型拟合结果一致。对于参数 γ ,无论是汛期还是非汛期,各河段 γ 的均值都大于 0,这表明温度的升高对浮游植物生物量的增加有正向影响。

表 1 各河段分期 β 、 γ 统计参数

Tab. 1 Statistical parameters of β and γ for each period in each river section

参 数	河 段	汛前		汛期		汛后	
		均值±偏差	MC 误差	均值±偏差	MC 误差	均值±偏差	MC 误差
β	R1	0.039±0.051	0.004	-0.080±0.325	0.049	0.007±0.050	0.007
	R2	0.003±0.056	0.005	-0.254±0.557	0.090	-0.001±0.053	0.007
	R3	0.033±0.051	0.004	-0.033±0.241	0.032	0.011±0.052	0.007
	R4	0.028±0.051	0.004	-0.001±0.259	0.034	0.002±0.052	0.007
	R5	0.005±0.054	0.004	-0.046±0.309	0.039	0.008±0.052	0.007
	R6	0.130±0.073	0.007	-0.015±0.297	0.038	0.013±0.051	0.007
	R7	0.004±0.055	0.005	-0.080±0.257	0.035	0.004±0.053	0.007
γ	R1	0.485±0.198	0.029	0.919±0.543	0.082	0.988±0.316	0.054
	R2	0.484±0.199	0.030	1.065±0.492	0.070	1.046±0.316	0.054
	R3	0.471±0.193	0.029	0.801±0.431	0.063	1.028±0.318	0.054
	R4	0.440±0.200	0.031	0.860±0.441	0.063	0.977±0.311	0.053
	R5	0.423±0.195	0.030	0.722±0.497	0.075	1.037±0.317	0.054
	R6	0.467±0.211	0.031	0.820±0.456	0.068	0.999±0.311	0.053
	R7	0.493±0.196	0.029	0.740±0.446	0.065	0.969±0.304	0.051

表 2 分期 β_c 、 γ_c 统计参数

Tab. 2 Statistical parameters of β_c , γ_c for sub-periods

参 数	期 间	均 值	97.5%置信	97.5%置信	中 位 数	t-value	P(>t)
			区间下限	区间上限			
β_c	汛前	0.034	-0.047	0.124	0.031	0.000	0.683
	汛期	-0.047	-0.589	0.413	-0.058	0.000	0.423
	汛后	0.007	-0.064	0.104	-0.003	0.000	0.471
γ_c	汛前	0.468	0.109	0.773	0.516	0.000	0.983
	汛期	0.855	0.004	1.613	0.840	0.000	0.952
	汛后	1.007	0.333	1.528	1.032	0.000	0.999

根据模型结果,拟定沙颍河流域年内生态流量过程在非汛期不低于 20 m³/s,汛期不超过 80 m³/s,在符合设计要求的生态流量条件下,所有河段的水生生物量模拟结果的中位数均高于样本中捕获的水生生物平均量(表 3)。然而,预测模型的不确定性意味着可信区间与采样数据重叠,最大的预期改善发生在河段 7(汛前+32.03%,汛期+872.03%,汛后+81.85%)即河流最上游,这可能是由于河流上游流量较小且受到人类活动影响较大,浮游植物生物量对流量条件的改善较为敏感;最小的预期改善发生在河段 3(汛前+3.92%,汛期+1.16%,汛后+2.88%)。

表 3 各河段生物量样本均值及设计条件下模拟中位数
Tab. 3 Mean values of biomass for each river section and median simulated values under design conditions

河 段	汛前		汛期		汛后	
	样本 均值	模拟中位数 (相对变幅)	样本 均值	模拟中位数 (相对变幅)	样本 均值	模拟中位数 (相对变幅)
R1	4.666	4.995(+7.05%)	10.329	12.054(+16.70%)	1.137	1.580(+38.98%)
R2	3.496	3.927(+12.33%)	12.439	12.789(+2.81%)	3.492	3.534(+1.20%)
R3	3.866	4.017(+3.91%)	4.648	4.702(+1.16%)	3.258	3.352(+2.89%)
R4	2.300	2.912(+26.61%)	6.904	7.408(+7.30%)	0.685	0.842(+22.92%)
R5	1.240	1.482(+19.52%)	1.908	2.865(+50.16%)	3.142	3.303(+5.12%)
R6	6.144	6.838(+11.30%)	3.963	7.090(+78.90%)	1.747	2.240(+28.22%)
R7	3.535	4.668(+32.05%)	0.773	7.512(+871.80%)	0.390	0.709(+81.79%)

总的来说,尽管贝叶斯分层模型以线性方式(式(10))模拟了流量、温度对浮游植物生物量的非线性响应(图 4),但其优势在于它为每个河段设置了不同的模型参数,根据河段的特性定制了模型,但也将同一河段内的站点数据视为可交换的。它在本质上假设同一河段的站点参数服从相同的先验分布,从而汇集了来自相近站点的信息,这种信息的共享减少了站点层次上的偏差,也减少了模型误差的方差。这种方法尤其适用于需要研究水文—生态响应关系但缺少大量采样数据的河湖。

4 结论

a. 贝叶斯层次模型可用于产生面向概率的

推理从而识别水文—生态的复杂响应关系,缓解生态数据不足、建模困难等难题。

b. 沙颍河非汛期河道内流量的保证及汛期水库泄流的控制对浮游植物生物量有正向影响。

c. 设计生态流量条件下,沙颍河浮游植物生物量最大和最小改善分别发生在河段 7(汛前 + 32.05%, 汛期 + 871.80%, 汛后 + 81.79%) 和河段 3(汛前 + 3.91%, 汛期 + 1.16%, 汛后 + 2.89%)。

参考文献:

- [1] 谢悦,夏军,张翔,等. 基于淮河流域中游鱼类不同等级生境保护目标的生态需水[J]. 南水北调与水利科技, 2017, 15(5): 76-81, 133.
- [2] 王俊钊,张翔,吴绍飞,等. 基于生径比的淮河流域中上游典型断面生态流量研究[J]. 南水北调与水利科技, 2016, 14(5): 71-77.
- [3] 董哲仁,张晶,赵进勇. 生态流量的科学内涵[J]. 中国水利, 2020(15): 15-19.
- [4] 葛金金,张汶海,彭文启,等. 水文-生态响应关系构建方法综述[J]. 水利水电技术, 2020, 51(11): 23-29.
- [5] MELLIOS N K, MOE S J, LASPIDOU C. Using Bayesian hierarchical modelling to capture cyanobacteria dynamics in Northern European lakes[J]. Wa-

ter research, 2020, 186:116356.

- [6] CHEN W, OLDEN J D. Designing flows to resolve human and environmental water needs in a dam-regulated river[J]. Nature communications, 2017, 8(1): 2158.
- [7] WEBB J A, STEWARDSON M J, KOSTER W M. Detecting ecological responses to flow variation using Bayesian hierarchical models[J]. Freshwater biology, 2010, 55(1): 108-126.
- [8] 茆诗松,汤银才. 贝叶斯统计[M]. 北京: 中国统计出版社, 2012.
- [9] 宫秀军. 贝叶斯学习理论及其应用研究[D]. 北京: 中国科学院研究生院, 2002.
- [10] 林秋奇,胡韧,韩博平. 流溪河水库水动力学对营养盐和浮游植物分布的影响[J]. 生态学报, 2003(11): 2278-2284.
- [11] 杨东方,陈生涛,胡均,等. 光照、水温和营养盐对浮游植物生长重要影响大小的顺序[J]. 海洋环境科学, 2007(3): 201-207.
- [12] GILKS W R, RICHARDSON S, SPIEGELHALTER D J. Markov Chain Monte Carlo in Practice [M]. London, Chapman & Hall, 1996.
- [13] GELMAN A, RUBIN D B. Inference from iterative simulation using multiple sequences[J]. Statistical science, 1992, 7(4): 457-472.

Ecological Flow Response in Shaying River Based on Bayesian Hierarchical Model

ZHAO Ye¹, ZHANG Xiang¹, DENG Liang-kun¹, FU Xiang¹, HU Jun²

(1. State Key Laboratory of Resources and Hydropower Engineering Science, Wuhan University, Wuhan 430072, China; 2. Institute of Hydroecology, MWR & CAS, Wuhan 430079, China)

Abstract: Maintaining ecological flow in a river is an important initiative to coordinate water resources development and utilization with river ecological protection. Most of the existing ecological flow calculation methods are limited by insufficient ecological data, difficult ecological modeling. In this paper, a Bayesian hierarchical model based on phytoplankton biomass was established to investigate the hydrological-ecological response relationship. The relationship between phytoplankton biomass and flow and water temperature in pre-flood, flood and post-flood periods was analyzed by taking the Shaying River as a case study. The results show that the Bayesian hierarchical model can identify the complex hydrological-ecological response relationship, and the advantage of hierarchical analysis improves the availability of short series data; The probability of biomass-flow covariate parameter β greater than 0 in pre-flood is 0.683, and the probability of less than 0 in flood is 0.577, and the guarantee of flow in the river channel in pre-flood and the control of discharge in flood have positive effects on phytoplankton biomass; Under the ecological flow conditions that meet the design requirements, the predicted maximum and minimum biomass improvements occur in river section 7 (+32.05% before flood, +871.80% during flood, and +81.79% after flood) and river section 3 (+3.91% before flood, +1.16% during flood, and +2.89% after flood), respectively.

Key words: Bayesian hierarchical model; hydrological-ecological response relationship; ecological flow; phytoplankton biomass; Shaying River