

文章编号:0559-9350(2011)06-0685-07

## 基于马尔可夫状态切换理论的水环境系统随机模拟与风险评估

牛军宜<sup>1</sup>, 冯平<sup>2</sup>, 张伟<sup>3</sup>

(1. 郑州大学 水利与环境学院, 河南 郑州 450001; 2. 天津大学 建筑工程学院, 天津 300072;

3. 天津市水文水资源勘测管理中心, 天津 300061)

**摘要:** 针对水质时间序列一般具有结构复杂性和演变过程非平稳性的特征, 提出将马尔可夫状态切换-自回归模型(Markov Switching-Auto Regressive Model)用于水质时序的随机模拟和水环境系统的风险分析, 并探讨了运用MS-AR模型进行水质时序随机模拟与分析的关键步骤。最后运用该模型和蒙特卡罗(Monte Carlo)方法, 对天津果河某断面的总磷浓度时序(1999—2007年)进行了随机模拟和污染风险评价。结果表明, MS-AR模型有效识别出了该水质时序在演变过程中的不同结构模式, 模拟出的总磷浓度序列能从总体上反映实测总磷时序的演变特征和动态变化趋势, 克服了利用单纯统计理论进行水环境风险分析无外延性的不足。

**关键词:** 水质时序; 马尔可夫理论; 状态切换; Box-Cox变换; 自回归模型; 随机模拟

**中图分类号:** X143

**文献标识码:** A

## 1 研究背景

河流、水库水质状况的变化过程不但受降水、径流、温度等自然条件的影响, 而且人类活动诸如产业结构规划、经济发展策略、环境保护部门监管力度以及生产技术水平变化等诸多因素, 均对水体水质有显著影响, 因而水环境系统的演变过程是一个复杂多变的动态过程。水环境的风险评价是指通过评估水环境系统的质量状态超过给定水环境质量标准的控制限值的程度及其发生的概率, 然后提出相应的管理对策的过程<sup>[1]</sup>。针对水环境系统存在风险的根源及其他影响水环境质量的不确定因素, 人们相应地提出了随机模型<sup>[2]</sup>、灰色模型<sup>[3]</sup>、模糊模型<sup>[4-5]</sup>、信息熵模型<sup>[6]</sup>等以及与其它理论耦合在一起的水环境风险评价模型<sup>[7]</sup>。上述模型虽然在水环境风险评价与管理方面具有良好的适用性, 但在反映水质时序结构特征及水环境系统演进模式方面有一定欠缺。由于水质时间序列是水环境系统对自然和人类活动等确定与不确定因素的综合反映, 因此研究水质时间序列的演变特征和变化趋势对水环境系统的风险评估有着重要意义。基于上述考虑, 本文尝试运用马尔可夫状态切换-自回归模型(Markov Switching-Auto Regressive Model, MS-AR模型)来分析水质时序的动态结构特征以及水环境系统的演进模式, 然后通过随机模拟的方法对水环境系统的风险进行评估。

MS-AR模型<sup>[8-9]</sup>最初应用于研究经济学中的相依数据, 此后在经济金融领域得到了广泛应用<sup>[10-14]</sup>, 它是一种研究具有变结构动力特征的时间序列分析方法, 对模拟序列结构复杂、演进模式多变的动态时序过程有良好的适应性, 其基本思想是通过外部数据计算系统内部处于何种不可观测结构的概率来对系统进行建模分析。由于决定金融时序和大部分水质时序演变特征与趋势的主导因素均是人类活动, 两者相近, 因此, 本文将MS-AR模型运用于水质时序的建模分析, 并通过实例检验其实用性。

收稿日期: 2009-09-16

基金项目: 国家自然科学基金(50879051)

作者简介: 牛军宜(1978-), 河南宜阳人, 博士, 讲师, 主要从事水资源与水环境的研究。E-mail: njy1230@126.com

## 2 研究方法

**2.1 MS-AR模型基本原理**<sup>[8-9]</sup> 对于演变过程如式(1)的离散时间序列 $\{x_t, t=1, 2, \dots, T\}$ , 其在 $t=1, 2, \dots, t_0$ 期间为一种演变结构模式, 在 $t=t_0+1, t_0+2, \dots, T$ 期间为另一种演变结构模式, 即时序 $\{x_t\}$ 在 $t_0$ 时刻发生了结构性变化, 传统的方法是在 $t \leq t_0$ 和 $t \geq t_0$ 时段分别建模

$$\begin{cases} \text{模型1: } x_t = c_1 + \varphi_1 x_{t-1} + \varepsilon_t, & 1 \leq t \leq t_0 \\ \text{模型2: } x_t = c_2 + \varphi_1 x_{t-1} + \varepsilon_t, & t_0 + 1 \leq t \leq T \end{cases} \quad (1)$$

但如果时间序列 $\{x_t\}$ 在 $t=1, 2, \dots, T$ 期间时而以模型1的形式演变, 时而以模型2的形式演变, 上述分段建模的方法就难以适应该时序 $\{x_t\}$ 的模拟及预测, 而且识别其时序结构变点也是一项繁琐的工作。

MS-AR模型可以解决上述难题, 它的建模思路是: 对于具有复杂结构的时序 $\{x_t\}$ , 假定它的演变过程可用 $M$ 个 $p$ 阶自回归过程来模拟, 然后引入状态变量 $S_t$ ,  $S_t$ 代表时间序列在时刻 $t$ 所处的结构模式, 它是具有状态空间 $J=(1, 2, \dots, M)$ 的一阶马尔可夫过程, 即当前状态仅与前一个状态有关, 与过去其他状态无关。这里假定模型的残差服从正态分布, 则基于马尔可夫状态切换的自回归模型如式(2)所示

$$\begin{cases} x_t = c^{S_t} + \sum_{k=1}^p \varphi_k^{S_t} x_{t-k} + \sigma^{S_t} \varepsilon_t, & S_t \in J \\ \varepsilon_t \sim i.i.d.N(0, 1) \\ Pr(S_t = j | S_{t-1} = i) = P_{ij} \\ \sum_{j=1}^M P_{ij} = 1, & (i, j \in J) \end{cases} \quad (2)$$

模型中待定参数为集合 $\theta = (c^{S_t}, \varphi_k^{S_t}, \sigma^{S_t}, P_{ij} | S_t \in J; k=1, 2, \dots, p; i, j \in J)$ , 其中:  $c^{S_t}$ 为系统处于状态 $S_t$ 下自回归模型的截距项;  $\varphi_k^{S_t}$ 为状态 $S_t$ 下自回归系数;  $\sigma^{S_t}$ 为在状态 $S_t$ 下模型残差的均方差;  $\{\varepsilon_t, t=1, 2, \dots, T\}$ 为服从同一正态分布的零均值白噪声序列。(式中: *i.i.d.N*为independent and identically distributed in Normal的缩写, 是用来表示高斯白噪声的符号;  $Pr$ 为系统在 $t-1$ 时刻由状态 $i$ 转移到状态 $j$ 的概率;  $P$ 为该马尔可夫过程的状态转移概率矩阵。)

对于参数集 $\theta$ 采用最大似然估计来求取, 为表述方便, 这里用 $\zeta_{i, t-1}$ 表示系统在时刻 $t-1$ 所处的状态为 $i$  ( $i \in J$ )的概率, 最大似然估计求取 $\theta$ 的过程如下:

$$\zeta_{i, t-1} = Pr(S_{t-1} = i | \Omega_{t-1}; \theta) \quad (i \in J) \quad (3)$$

式中 $\Omega_{t-1} = \{x_{t-1}, x_{t-2}, \dots, x_1\}$ 为在 $t-1$ 时刻所观测到的所有序列样本的集合, 它可以被看作到 $t$ 时刻的信息集。

由式(2)可知,  $t$ 时刻系统处于第 $j$ 个状态的条件密度函数

$$\eta_{j, t} = f(x_t | S_t = j, \Omega_{t-1}; \theta) = \frac{1}{\sqrt{2\pi} \sigma^{S_t}} \exp \left[ -\frac{\left( x_t - c^{S_t} - \sum_{k=1}^p \varphi_k^{S_t} x_{t-k} \right)^2}{2(\sigma^{S_t})^2} \right] \quad (4)$$

由式(3)和式(4)可得到 $t$ 时刻的状态混合密度函数

$$f(x_t | \Omega_{t-1}; \theta) = \sum_{i=1}^M \sum_{j=1}^M P_{ij} \zeta_{i, t-1} \eta_{j, t} \quad (5)$$

从而由贝叶斯公式可得到系统在时刻 $t$ 所处状态为 $j$ 的概率

$$\zeta_{j, t} = \frac{\sum_{i=1}^M P_{ij} \zeta_{i, t-1} \eta_{j, t}}{f(x_t | \Omega_{t-1}; \theta)}, \quad (i \in J) \quad (6)$$

由上述公式的计算结果可以对参数  $\theta$  采用迭代法进行最大似然估计。待估参数集  $\theta$  的条件对数似然函数为

$$L(\theta | \Omega_T) = \sum_{t=1}^T \lg f(x_t | \Omega_{t-1}; \theta) \quad (7)$$

具体的迭代估计步骤如下：

(1) 首先给定  $t=1, 2, \dots, p$  时刻处于各状态的概率  $\zeta_{i, t}$  作为初值，初值的选取方法可采用限制性无条件概率，即

$$\zeta_{i, t} = \frac{1 - P_{ii}}{\sum_{j=1}^M (1 - P_{jj})}, \quad (t=1, 2, \dots, p, i \in J) \quad (8)$$

然后令  $t=p$ ；

(2) 将各个初值代入式(4)和式(5)得到的系统在  $t=t+1$  时刻状态空间的条件概率密度  $\eta_{j, t+1}$  ( $j \in J$ )；再将  $\eta_{j, t+1}$  和  $\zeta_{i, t}$  ( $t=t, t-1, \dots, t-p+1$ ) 代入式(5)得到  $f(x_{t+1} | \Omega_t; \theta)$ ；

(3) 将  $f(x_{t+1} | \Omega_t; \theta)$ 、 $\zeta_{i, t}$  和  $\eta_{j, t+1}$  代入式(6)得到  $\zeta_{i, t+1}$  ( $i \in J$ )；如果  $t+1=T$ ，结束迭代，转到下一步；否则令  $t=t+1$ ，转到第(2)步继续迭代。

(4) 将  $f(x_{t+1} | \Omega_t; \theta)$  ( $t=p+1, p+2, \dots, T$ ) 代入式(7)，通过数值搜索寻优方法求取参数集  $\theta$  的最大似然估计  $\hat{\theta}$ 。通过以上步骤求得  $\hat{\theta}$  后，可由式(6)计算任一时刻  $t$  处的各个状态的概率  $\hat{\zeta}_{j, t}$ 。

由于采用极大似然法估计自回归模型参数时，要求所模拟的时间序列  $\{x_t, t=1, 2, \dots, T\}$  必须为一高斯过程<sup>[15]</sup>，而实际的水质时间序列并不总是能满足这个条件，因此在对已有水质时序建立 MS-AR 模型前，首先应对水质时序数据进行高斯性检验，如果水质时序过程是非高斯的，就需要寻求某种变换把原水质时序变换为服从高斯分布的时间序列，经作者对天津于桥水库流域的大量实测水质时序进行高斯变换试验，发现 Box-Cox 变换<sup>[16]</sup>是一种很好的方法。Box-Cox 变换是 Box 和 Cox 于 1964 年提出的一种基于极大似然法的幂转换模型，其转换模型如下

$$x_t^{(\lambda)} = \begin{cases} \frac{x_t^\lambda - 1}{\lambda}, & \lambda \neq 0 \\ \ln x, & \lambda = 0 \end{cases} \quad (9)$$

式中  $\lambda$  为一个待定变换参数，对于  $\lambda$  值的选择，可以通过极大似然法来估计，其似然函数及估计方法参见文献[16]。

**2.2 MS-AR 模型的随机模拟原理** 基于 MS-AR 模型的随机模拟是在给定模型初始状态的基础上，按照状态转移矩阵，借助蒙特卡罗法模拟各个状态之间的转换过程。它的随机模拟过程包括两部分内容，即状态切换过程的随机模拟和状态值的随机模拟：

(1) 状态切换过程的随机模拟：状态切换过程的随机模拟是在初始状态已知的情况下，按照状态转移概率矩阵，借助蒙特卡罗方法<sup>[17]</sup>模拟模型在各个状态之间切换的过程。如果模型在  $t$  时刻的状态为  $i$ ，那么采用蒙特卡罗法模拟它从  $t$  时刻转移到  $t+1$  时刻的状态过程为，首先由随机数发生器产生一个 0 到 1 之间服从均匀分布的伪随机数  $Rand$ ，然后将  $Rand$  与  $p_{ii}$  相比，如果  $Rand \leq p_{ii}$ ，模型在  $t+1$  时刻将继续停留在状态  $i$  向前演进；如果  $Rand > p_{ii}$ ，模型将向其它状态切换，计算出  $(Rand - p_{ii})$ ，再拿  $(Rand - p_{ii})$  分别与  $\{p_{i, j}, j \in J, j \neq i\}$  ( $J$  为模型的状态空间) 相比，如果  $p_{i, k}$  的数值最接近于  $(Rand - p_{ii})$ ，那么  $t+1$  时刻，模型将切换到状态  $k$  向前演进。

(2) 状态值的随机模拟：如果模型在  $t$  时刻的状态为  $S_t$ ，那么用随机数算法产生一个服从均值为 0、均方差为  $\sigma^{S_t}$  正态分布的伪随机数  $Randn$ ，这是因为我们的参数估计过程是建立在残差服从正态分布基础上的， $\sigma^{S_t}$  为模型在状态  $S_t$  下模型残差的均方差。因为水环境系统中的质量状态变量一般来说有最小可能取值、最可能取值和最大可能取值，因此模型的误差也对应应有取值范围，由于模型残差近似服从正态分布，这里参照文献[7]中的方法，规定  $Randn$  的可能取值区间为  $\left[ \overline{Res} - 2\sigma^{S_t}, \overline{Res} + 2\sigma^{S_t} \right]$ ，其中的  $\overline{Res}$  为状态为  $S_t$  时序列  $\{x_i\}$  经过 Box-Cox 变换并标准化后的序列  $\{y_i\}$  的 MS-AR 模型残差的均值， $\sigma^{S_t}$  对应残差的均方差。那么  $y_t$  的模拟值为

$$z_t = c^{S_t} + \sum_{k=1}^p \varphi_k^{S_t} z_{t-k} + Randn \quad (10)$$

然后对序列  $\{z_i\}$  作标准化反变换及 Box-Cox 反变换即可得到序列  $\{x_i\}$  的模拟值。

**2.3 水环境系统风险评估<sup>[7]</sup>** 对水环境系统的质量状态变量  $\{x_i\}$  如果给定水环境质量标准的控制限值为  $x_c$ ，那么水环境风险可定义为

$$R_z = Pr(x_i \geq x_c) \quad (11)$$

式中： $x_i$  可以为水质断面的某一项监测指标，也可以为通过水质综合评价模型将多项监测指标归一化的综合评价指标； $Pr$  为事件  $x_i \geq x_c$  的概率函数。

然后通过所建的 MS-AR 模型采用随机模拟的方法得到水环境系统的质量状态变量  $\{x_i\}$  的模拟序列  $\{z_n, n=1, 2, \dots, N\}$ ， $N$  为随机模拟的试验次数。研究表明，试验次数越多， $z$  的频率分布越接近于其真实的概率分布，在实用上  $N$  一般取频率分布收敛时所对应的试验次数。若序列  $\{z_n\}$  中超过水环境质量标准限值  $x_c$  的数目为  $N_z$ ，则水环境风险的估计值为

$$R_z = N_z / N \quad (12)$$

最后，根据风险的估计值和模拟序列  $\{z_n\}$  的统计特征，可提出水环境风险的相应对策。

### 3 实例应用

**3.1 基本数据及其预处理** 果河是天津于桥水库的主要支流，自引滦入津工程建成通水后，成为引滦入津的重要输水通道，其年均径流量约占于桥水库年入库总流量的 85% 左右，而且果河桥断面是海河流域重点水质监测断面，该断面的水质状况反映了天津市于桥水库上游来水水质的总体状况。监测资料显示，果河桥断面总磷超标较为严重，在 1999 年 11 月—2007 年 12 月间的监测记录中，超过国家 II 类水质标准约占 52.58%。本文基于上述研究方法，以该断面该期间的总磷浓度时序  $\{x_i, i=1, 2, \dots, 98\}$  为研究对象，对其建立 MS-AR 模型并分析该水质时序的动态演进特征，然后通过随机模拟的方法对于桥水库上游来水水质安全风险做出评价。图 1 为果河桥断面 1999 年 11 月—2007 年 12 月期间 98 个月份的总磷浓度过程线。经检验  $\{x_i\}$  不服从正态分布，故首先对其进行 Box-Cox 变换 ( $\lambda = -0.446$ ) 并对其进行标准化处理得到序列  $\{y_i\}$ ，采用 Kolmogorov-Smirnov 检验方法对序列  $\{y_i\}$  进行正态性检验， $\{y_i\}$  通过正态假设检验的伴随概率为 84.23%，序列  $\{y_i\}$  的变化过程线如图 2 所示。

**3.2 参数估计** 从图 2 中难以判断出  $\{y_i\}$  的演变过程有几种状态，为了便于对比说明，这里取  $S_{\max} = 3$ 。对于  $\{y_i\} (i=1, 2, \dots, 98)$  的最大回归阶数，可根据自回归模型自相关函数拖尾、偏自相关函数截尾的性质来初步估计它的最大回归阶数定为 5。表 1 为序列  $\{y_i\}$  的 MS-AR 模型在不同状态数量及回归阶数组合  $(s, p)$  下的残差白噪声检验及模型优选评价指标计算成果，表中的  $P$  值为残差通过白噪声假设

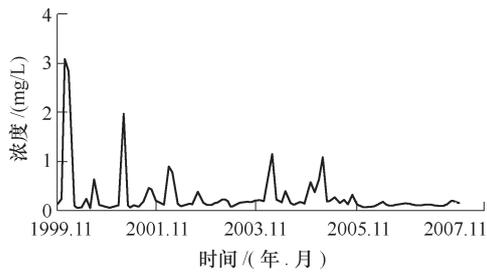


图1 总磷浓度时序变化过程线

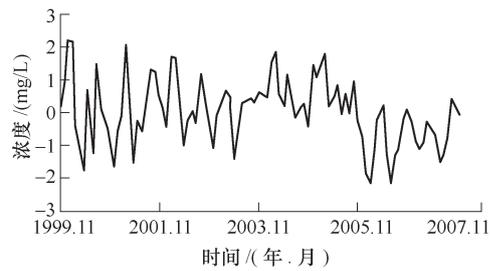


图2 序列{y\_t}变化过程线

检验的伴随概率向量中的最小值，当  $s=1$  时，MS-AR 模型退化为 AR 模型。

表1 序列  $\{y_t\}$  的模型残差白噪声检验及模型优选评价指标

| $\varphi$ | s=1   |        |       | s=2          |               |              | s=3   |        |       |
|-----------|-------|--------|-------|--------------|---------------|--------------|-------|--------|-------|
|           | P 值   | AIC    | $R^2$ | P 值          | AIC           | $R^2$        | P 值   | AIC    | $R^2$ |
| 1         | 0.180 | -25.85 | 0.240 | 0.163        | -12.66        | 0.266        | 0.278 | -7.03  | 0.422 |
| 2         | 0.315 | -25.54 | 0.253 | 0.929        | -71.18        | 0.432        | 0.236 | -5.88  | 0.177 |
| 3         | 0.470 | -28.79 | 0.299 | <b>0.877</b> | <b>-78.18</b> | <b>0.475</b> | 0.595 | -39.09 | 0.653 |
| 4         | 0.660 | -30.73 | 0.341 | 0.635        | -27.54        | 0.433        | 0.882 | -15.04 | 0.448 |
| 5         | 0.661 | -37.60 | 0.385 | 0.824        | -56.17        | 0.595        | 0.570 | -7.99  | 0.321 |

由表1可知，2种结构状态的3阶自回归模型为序列  $\{y_t\}$  的最优 MS-AR 模型。对序列  $\{y_t\}$  的拟合值进行标准化逆变换及 Box-Cox 逆变换，得到原序列  $\{x_t\}$  的 MS-AR 模型，采用  $\chi^2(m)$  检验该模型残差序列，不能拒绝残差序列为白噪声的假设，并且通过检验的伴随概率向量中最小值为 96.87%，模型的可决系数  $R^2$  为 0.498，模型参数计算结果见表2。

表2 MS-AR 模型自回归系数的估计值

|             | 状态 I   | 状态 II  |          | 状态 I   | 状态 II |
|-------------|--------|--------|----------|--------|-------|
| $\varphi_1$ | 0.804  | -0.305 | $c$      | -0.165 | 0.682 |
| $\varphi_2$ | -0.338 | 0.433  | $\sigma$ | 0.652  | 0.703 |
| $\varphi_3$ | 0.047  | -0.030 | $P_{ii}$ | 0.838  | 0.504 |

所建 MS-AR 模型拟合原序列  $\{x_t\}$  的情况如图3所示，模型的所处状态的概率转移过程如图4所示，该模型基本上能反映出总磷时序在此期间内的变化特征。

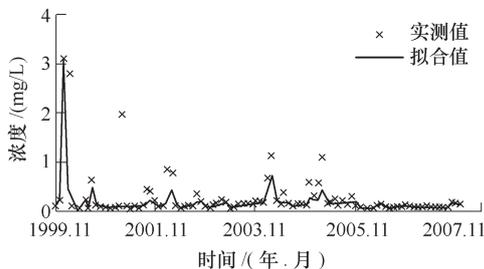


图3 模型拟合原序列  $\{x_t\}$  情况

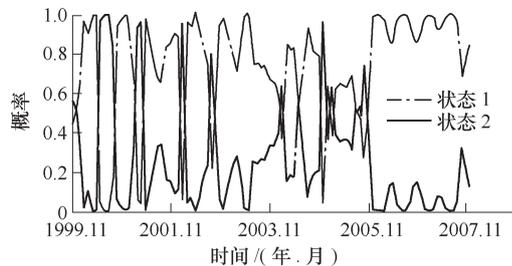


图4 系统两种状态概率切换过程

如果视  $Pr(S_t = 1 | \Omega_{t-1}) > 0.5$  ( $\Omega_{t-1}$  为  $t$  时刻的信息集) 的情况下该断面的总磷浓度时序是在状态 I 下演进，当  $Pr(S_t = 2 | \Omega_{t-1}) > 0.5$  时，认为该断面的总磷浓度时序是在状态 II 下演进，那么由图4我们可看出，果河桥断面的总磷时序在 1999 年 11 月—2007 年 12 月期间，其演变过程大部分时间处于状态 I，但仍存在两种结构状态切换现象，状态 I 的持续性要比状态 II 好。由状态转移矩阵和公式

$E(L_i) = 1/(1 - P_{ii})$ ，我们可以计算出该总磷时序在其演进过程中处于各个状态下所持续时长的期望值，经计算模型连续处于状态 I 下的时长期望值为 6.2 个月，连续处于状态 II 下的时长期望值为 2 个月。作用于状态 I 背后的动力因子具有稳定性和持续性，例如，管理部门对水质的常规防控措施、生活污水的排放等影响因子就具有持续性；作用于状态 II 背后的动力机制具有突发性和暂态性，例如工业污染源的偷排偷放、气候情况异常(暴雨洪水引起农业面源污染负荷短时间内大量汇入河流)，以及有关部门针对突发性污染事件启动相应的应急预案及控制措施等。

**3.3 果河桥断面总磷浓度超标风险分析** 利用上述模型对果河桥总磷时序进行随机模拟，表 3 给出了不同模拟次数总磷浓度的平均值、最小值、最大值，以及按照式(12)计算出的总磷浓度超过国家水质标准中 II 类水质的控制准限值的风险概率，结果见表 3。

表 3 总磷模拟序列的统计特征及风险评估结果

| 实验次数   | 总磷模拟浓度统计特征(mg/L) |       |      | 风险的估计值 |
|--------|------------------|-------|------|--------|
|        | 平均值              | 最小值   | 最大值  |        |
| 500    | 0.156            | 0.037 | 2.86 | 53.80% |
| 2 000  | 0.151            | 0.033 | 3.53 | 54.70% |
| 5 000  | 0.144            | 0.033 | 4.57 | 53.40% |
| 10 000 | 0.146            | 0.031 | 4.29 | 53.37% |
| 20 000 | 0.146            | 0.030 | 4.47 | 53.33% |
| 40 000 | 0.146            | 0.030 | 6.22 | 53.34% |

分析表 3 可得出以下结论：

(1)对果河桥断面的总磷时序采用所建的 MS-AR 模型模拟 40 000 次时，模拟系列的均值、最小值和风险估计值已收敛，此时该断面处总磷浓度的平均值为 0.146mg/L，最小值为 0.030mg/L，污染风险估计值为 53.34%。模拟浓度系列的均值超过国家 II 类水质标准限值(0.1mg/L)0.46 倍，说明该断面总磷在今后一段时间内的污染形势比较严峻，于桥水库的污染项目中，总磷污染仍有较大的输入性污染风险。

(2)随机模拟结果中，总磷浓度模拟系列中的最小值(0.030mg/L)与实测系列的最小值(0.030mg/L)相等，而模拟系列最大值(6.22mg/L)为实测系列的最大值(3.09mg/L)2 倍左右，从水环境系统动力机制的角度审视，果河桥断面处的实测总磷月平均浓度不可能有如此之大，为小概率事件，从统计学的观点出发，我们可以不予考虑，绝大部分的模拟值是在一个合理范围之内，因此，可以认为该模型是稳定可靠的。

(3)按照常规的统计方法，果河桥断面实测总磷浓度系列的均值(0.250mg/L)大于模拟系列的均值(0.146mg/L)，而实测总磷浓度系列超过 II 类水质标准限值的频率(52.58%)却低于模拟系列的风险估计值(53.34%)，这说明实测总磷浓度系列的离差比模拟系列的离差大，模拟系列的统计特征并不完全和实测系列相同。我们分析图 1 可知，果河桥总磷浓度时序在计算时段内，它的均值、极大值、以及波动程度均有减小的趋势。需要注意的是，只要系列的均值超过水质标准限值，那么波动程度减小势必引起超标次数增加，使得超标频率增大。然而单纯的统计结果是静态的，并不能反映总磷浓度时序的这些变化趋势。

运用本文所建的 MS-AR 模型，借助 Monte Carlo 随机模拟方法，得到的总磷模拟系列是动态的，它不但具有该段时间内的实测总磷浓度系列的大部分统计特征，而且反映了实测总磷浓度时序演变过程中的结构特征和变化趋势，因此该风险评估模型具有动态性和外延性，更加适合于水环境系统演变趋势的分析和风险评估。

## 4 结论

运用文中所建的MS-AR模型对于桥水库上游控制断面—果河桥断面总磷的随机模拟表明, 如果以1999年11月—2007年12月这段时间内的总磷浓度时序作为估计未来总磷污染风险的信息集, 那么果河桥断面的总磷浓度均值及波动方差均有减小的趋势, 但超标的风险依然比较大, 说明于桥水库仍然有较大的输入性总磷污染风险, 为保障天津市安全供水风险, 仍需进一步加强于桥水库上游流域总磷污染源的防控措施。MS-AR模型在自动识别时间序列的演变模式方面有较好的效果, 适合对受人类活动影响显著、演变过程复杂的水质时间序列进行随机模拟, 在水环境系统的风险分析和评估方面有着良好的运用前景。另外, 由于MS-AR模型在建模前需要对数据进行正态变换, 建模后进行模拟预测时, 需要对所得的模拟值进行反变换, 在此过程中可能会出现较大误差<sup>[18]</sup>, 那么结合待分析数据的特性, 寻求恰当的正态变换方法对MS-AR模型应用于水质时序模拟分析有着重要意义。

## 参 考 文 献:

- [ 1 ] 耿福明, 薛联青, 陆桂华, 等. 饮用水源水质健康危害的风险度评价[J]. 水利学报, 2006, 37(10): 1242-1245.
- [ 2 ] 王文圣, 金菊良, 李跃清, 等. 水文水资源随机模拟技术[M]. 成都: 四川大学出版社, 2007.
- [ 3 ] 周林飞, 许士国, 孙万光. 基于灰色聚类法的扎龙湿地水环境质量综合评价[J]. 大连理工大学学报, 2007, 47(2): 240-245.
- [ 4 ] 李如忠. 基于不确定信息的城市水源水环境健康风险评价[J]. 水利学报, 2007, 38(8): 895-900.
- [ 5 ] 李如忠, 洪天求, 金菊良. 河流水质模糊风险评价模型研究[J]. 武汉理工大学学报, 2007, 29(2): 43-46.
- [ 6 ] 金菊良, 魏一鸣. 复杂系统广义智能评价方法与应用[M]. 北京: 科学出版社, 2008.
- [ 7 ] 金菊良, 吴开亚, 李如忠. 水环境风险评价的随机模拟与三角模糊数耦合模型[J]. 水利学报, 2008, 39(11): 1257-1261.
- [ 8 ] Hamilton James D. A new approach to the economic analysis of non-stationary time series and the business cycle [J]. *Econometrica*, 1989, 57 (2): 357-384.
- [ 9 ] Krolzig H M. Markov-switching Vector Autoregressions: Modelling, Statistical Inference and Application to Business Cycle Analysis[M]. Berlin: Springer, 1998.
- [ 10 ] Liu Wen-Hsien, Chyi Yih-Luan. A Markov regime-switching model for the semiconductor industry cycles[J]. *Economic Modelling*, 2006, (23): 569-578.
- [ 11 ] 苏涛, 詹原瑞, 刘家鹏. 基于马尔科夫状态转换下的CAPM实证研究[J]. 系统工程理论与实践, 2007, 27(6): 21-26.
- [ 12 ] Alessandro Cologni, Matteo Manera. The asymmetric effects of oil shocks on output growth: A Markov-Switching analysis for the G-7 countries[J]. *Economic Modelling*, 2009, 26(1): 1-29.
- [ 13 ] Liu Ji-Chun. Stationarity for a Markov-switching Box-Cox transformed threshold GARCH process[J]. *Statistics & Probability Letters*, 2007, 77(13): 1428-1438.
- [ 14 ] Francq C, Zakořan J M. Stationarity of multivariate Markov-switching ARMA models[J]. *Journal of Econometrics*, 2001, 102(2): 339-364.
- [ 15 ] Hamilton J D. 时间序列分析[M]. 刘明志译. 北京: 中国社会科学出版社, 1999.
- [ 16 ] Toshimitsu Hamasaki, SeoYoung Kim. Box and Cox power-transformation to confined and censored non-normal responses in regression[J]. *Computational Statistics & Data Analysis*, 2007, 51(8): 3788-3799.
- [ 17 ] Smith A F M, Robert G O. Bayesian computation via the Gibbs sampler and related Markov chain Monte Carlo methods[J]. *Journal of Royal Statistical Society Series B*, 1993, 55: 3-23.
- [ 18 ] 梁忠民, 戴昌军. 水文分析计算中两种正态变换方法的比较研究[J]. 水电能源科学, 2005, 23(2): 1-3.

(下转第699页)

## Random simulation and environmental risk assessment for water system based on Markov Switching theory

NIU Jun-yi<sup>1</sup>, FENG Ping<sup>2</sup>, ZHANG Wei<sup>3</sup>

(1. Zhengzhou University, Zhengzhou 450001, China; 2. Tianjin University, Tianjin 300072, China;

3. Tianjin survey and Management Center of Hydrology and Water Resources, Tianjin 300061, China)

**Abstract:** Focusing on the non-stationary and structural complexity of water quality time series, the Markov Switching-Auto Regressive Model (MS-AR) was introduced for water quality time series random simulation, and then the main procedures of environmental risk assessment for water system by using MS-AR model was discussed. Finally, by means of this model and Monte Carlo method, the total phosphorus (TP) concentration time series in a river section in Guohe River, Tianjin, was randomly simulated, and the risk of TP exceeding national standard limits was assessment. The result shows that the MS-AR model can identify two evolution patterns of the series, the simulated series reflect the evolution features and the changing trend of the actual time series, which overcame the disadvantage of extensionality and prediction capability of using statistical method only to the environmental risk analysis of water system.

**Key words:** water quality time series; Markov theory; Regime Switching; Box-Cox transformation; autoregressive model; random simulation

(责任编辑: 韩 昆)