



文章编号:1671-6833(2020)03-0072-07

# 基于拉丁超立方抽样的改进型多链 DRAM 算法 求解地下水污染反问题

张双圣<sup>1,3</sup>, 强 静<sup>2</sup>, 刘汉湖<sup>1</sup>, 刘喜坤<sup>3</sup>, 孙韶华<sup>4</sup>

(1.中国矿业大学 环境与测绘学院,江苏 徐州 221116; 2.中国矿业大学 数学学院,江苏 徐州 221116;  
3.徐州市城区水资源管理处,江苏 徐州 221018; 4.山东省城市供排水水质监测中心,山东 济南 250100)

**摘 要:** 针对运用贝叶斯统计方法求解地下水污染反问题时,经典 MCMC 算法(Metropolis 算法)求解结果受样本初始点影响且计算效率低的问题,提出了一种基于拉丁超立方抽样方法的改进型多链延迟拒绝自适应 Metropolis 算法(DRAM)。将贝叶斯统计方法与二维水质对流-扩散方程相耦合,建立地下水污染源识别模型。构建一个污染物在地下水含水层中瞬时排放的算例,分别运用 Metropolis 算法、多链 Metropolis 算法以及改进型多链 DRAM 算法对污染源信息(污染源强度、排放位置 $(x, y)$ 和排放时长)进行反求。算例研究表明, Metropolis 算法受样本初始点影响,容易出现反演结果局部最优或者反演结果难以收敛的问题;多链 Metropolis 算法虽然显著提高了反演结果的准确性,但是反演效率相对低下;改进型多链 DRAM 在保证反演准确性的条件下,可显著提高反演效率(相对于多链 Metropolis 算法提高 68%),实现反演结果准确性与效率的双提高。

**关键词:** 二维水质模型; 贝叶斯-马尔科夫链蒙特卡洛法; 拉丁超立方抽样; 延迟拒绝自适应 Metropolis 算法; 污染源识别

中图分类号: X703

文献标志码: A

doi:10.13705/j.issn.1671-6833.2019.02.016

## 0 引言

目前,国内外水污染溯源问题的研究较多,其本质是水质模型参数的识别。参数识别方法主要有马尔科夫链蒙特卡罗算法(Markov chain Monte Carlo, MCMC)<sup>[1-7]</sup>、微分进化算法<sup>[8-11]</sup>、遗传算法<sup>[12-13]</sup>和最小二乘算法<sup>[14]</sup>等。陈海洋等<sup>[4]</sup>运用经典 MCMC 算法实现了水体污染源项的识别,得到了污染源强度、污染源位置和污染源排放时长 3 个未知参数的估计值。闵涛等<sup>[12]</sup>利用遗传算法求解对流-扩散方程源项识别反问题。这些算法可分为确定性方法和不确定性方法,其中确定性方法未考虑误差因素,得到参数确定的估计值,但容易丢掉“真值”;不确定性方法具有较强的随机性,收敛速度慢,计算量随着参数的增多而呈指数增长,得到待估参数的一个小的可靠范围。其中, MCMC 方法应用较为广泛。Metropolis 算法<sup>[13-14]</sup>是一种经典的 MCMC 方法,应用非常广

泛。大多数 MCMC 算法都是通过改进 Metropolis 算法得到的,比如延迟拒绝算法(delayed rejection, DR)<sup>[15-16]</sup>、自适应 Metropolis 算法(AM)<sup>[17]</sup>、延迟拒绝自适应 Metropolis 算法(DRAM)<sup>[18]</sup>等。DRAM 算法是一种高效自适应 MCMC 算法,其本质是把 DR 算法和 AM 算法二者组合起来。DR 算法保证了马尔科夫链的局部自适应,AM 算法采用全局自适应调整策略,因此 DRAM 算法能够对 MCMC 链条进行全局和局部的自适应,DRAM 算法抽样效率优于 DR 算法和 AM 算法各自单独使用的情况<sup>[19]</sup>。但是 Metropolis 算法和 DRAM 算法都只有一条链,容易陷入到局部最优值<sup>[19-20]</sup>,计算结果给决策者造成一定的负面影响。

笔者运用二维地下水水质对流-扩散方程,通过贝叶斯公式建立水体污染源识别模型,得到污染源强度、排放位置 $(x, y)$ 和排放时长 4 个未知参数的后验概率密度函数,并运用拉丁超立方抽

收稿日期:2019-10-07;修订日期:2019-12-15

基金项目:国家自然科学基金资助项目(51774270);国家水体污染控制与治理科技重大专项基金资助项目(2015ZX07406005)

作者简介:张双圣(1983—),男,中国矿业大学博士研究生,主要从事地下水污染控制研究, E-mail: zhang\_shuangsheng@163.com。



样方法<sup>[21]</sup>优化 DRAM 算法,实现参数反演过程中准确性与效率的双提高,为突发性水污染反问题研究提供借鉴。

## 1 研究方法

### 1.1 贝叶斯公式

贝叶斯公式<sup>[22-23]</sup>如下:

$$p(\alpha|d) = \frac{p(d|\alpha)p(\alpha)}{p(d)} \propto p(d|\alpha)p(\alpha), \quad (1)$$

式中:  $\alpha$  为模型的未知参数;  $d$  为实测数据;  $p(\alpha|d)$  为参数的后验概率密度函数;  $p(\alpha)$  为参数的先验概率密度函数;  $p(d|\alpha)$  为条件概率密度函数;  $p(d)$  为归一化的积分常数。

假设模型中未知参数共有  $m$  个,则  $\alpha = (\alpha_1, \alpha_2, \dots, \alpha_m)$ 。环境水力学的模型参数都分布在一个特定的范围内,可以认为每个参数都服从均匀分布,且  $\alpha_1, \alpha_2, \dots, \alpha_m$  相互独立。模型参数  $\alpha_i$  的先验概率密度函数可定义为:

$$p(\alpha_i) = \begin{cases} \frac{1}{B_i - A_i}, & \alpha_i \in [A_i, B_i]; \\ 0, & \text{其他。} \end{cases} \quad (2)$$

总的先验分布  $p(\alpha)$  可表示为:

$$p(\alpha) = \prod_{i=1}^m p(\alpha_i). \quad (3)$$

假设模型中实测值共有  $n$  个,即  $d = (d_1, d_2, \dots, d_n)$ 。  $d_i$  表示第  $i$  个实测数据,  $C_i(x, y, t|\alpha)$  表示相应的第  $i$  个预测值,则  $\varepsilon_i = d_i - C_i(x, y, t|\alpha)$  为测量误差,  $i = 1, 2, \dots, n$ 。假设测量误差服从均值为 0、标准偏差为  $\sigma$  的正态分布,且相互独立,则条件概率密度函数  $p(d|\alpha)$  可表示如下:

$$p(d|\alpha) = \frac{1}{(2\pi\sigma^2)^{n/2}} \cdot \exp\left\{-\sum_{i=1}^n \frac{[d_i - C_i(x, y, t|\alpha)]^2}{2\sigma^2}\right\}. \quad (4)$$

联合式(1)~(4),可得后验概率密度函数  $p(\alpha|d)$  为:

$$p(\alpha|d) = \frac{\prod_{i=1}^m p(\alpha_i)}{(2\pi\sigma^2)^{n/2} p(d)} \cdot \exp\left\{-\sum_{i=1}^n \frac{[d_i - C_i(x, y, t|\alpha)]^2}{2\sigma^2}\right\}. \quad (5)$$

由于  $\frac{\prod_{i=1}^m p(\alpha_i)}{(2\pi\sigma^2)^{n/2} p(d)}$  是与参数  $\alpha$  的选取无关的固定数,可记为  $\lambda$ 。故式(5)可写为:

$$p(\alpha|d) = \lambda \exp\left\{-\sum_{i=1}^n \frac{[d_i - C_i(x, y, t|\alpha)]^2}{2\sigma^2}\right\}. \quad (6)$$

在实测数据  $d$  固定的条件下,式(6)是关于参数  $\alpha$  的函数。式(6)通过积分求解参数  $\alpha$  的概率分布,计算较为复杂,难以得出明确的解析表达式。并且随未知参量维数的增加,数值积分算法的计算量将呈指数增长,实现复杂且难度较大。

### 1.2 延迟拒绝自适应 Metropolis 算法

式(6)一般可采用马尔科夫链蒙特卡罗方法(MCMC)近似求解。MCMC 方法的核心是 Monte Carlo 模拟方法和 Markov Chain 抽样方法。当“概率事件”样本点足够多时,频率可以近似代表概率,这是 Monte Carlo 模拟方法的本质,而 Markov Chain 抽样方法能够保证 Markov 链花更多的时间在概率大的区域,节省 Monte Carlo 模拟方法的工作量。

Metropolis 算法是一种经典的 MCMC 算法,应用非常广泛,具体算法参见文献[13-14],其提议分布一旦设定便保持不变,当提议分布远离目标分布时,马尔科夫链收敛速度会变慢。为了提高马尔科夫链收敛效率,Haario 等<sup>[18]</sup>将 DR 算法和 AM 算法结合提出了高效自适应 DRAM 算法,其具体步骤如下。

#### 1.2.1 非自适应阶段

(1)在模型未知参数  $\alpha$  先验范围内随机产生初始的参数样本  $X_1(t=1)$ ,非自适应阶段迭代次数为  $N_0$ ,提议分布为高斯分布  $q_t = N(X_t, C_0)$ ,其中  $C_0$  为协方差矩阵,要求  $C_0$  满足对称正定条件,并且高斯分布满足对称随机游走。

(2)从提议分布  $N(X_t, C_0)$  中抽取一个参数样本  $X^*$ ;产生  $[0, 1]$  上的随机数  $u$ ;计算接受概率  $\min\left\{1, \frac{p(X^*|d)}{p(X_t|d)}\right\}$ ,其中  $p(X^*|d)$  与  $p(X_t|d)$  由式(6)计算得到。如果  $u < \min\left\{1, \frac{p(X^*|d)}{p(X_t|d)}\right\}$ ,接受  $X^*$ ,即  $X_{t+1} = X^*$ ;否则,拒绝  $X^*$ ,即  $X_{t+1} = X_t$ 。

(3)重复过程(2),直至达到迭代次数  $N_0$ 。

#### 1.2.2 自适应阶段

(1)通过步骤(1)已抽取到的样本点  $X_1, X_2, \dots, X_{N_0}$ ,此时  $t = N_0$ 。下一步提议函数  $q_t^1(X_t, \cdot) = N(X_t, C_t^1)$ ,式中  $C_t^1 = s_m \text{cov}(X_1, X_2, \dots, X_t) + s_m \varepsilon I_m$ ;  $m$  表示参数  $\alpha$  的维数,  $s_m = 2.38/m^{2[18]}$ ;  $\varepsilon$



表示一个很小的正数,可取为 $10^{-6}$ ,目的是保证 $\mathbf{C}_t^{-1}$ 为正定矩阵; $\mathbf{I}_m$ 是一个 $m$ 维单位矩阵。自适应阶段迭代次数为 $N$ 。

(2)马尔科夫链的当前状态为 $\mathbf{X}_t$ ,从提议分布 $q_{t+1}$ 中抽取一个样本 $\mathbf{X}_1^*$ ;计算接受概率 $A_t^1(\mathbf{X}_t, \mathbf{X}_1^*) = \min\left\{1, \frac{p(\mathbf{X}_1^*|d)}{p(\mathbf{X}_t|d)}\right\}$ ;产生 $[0,1]$ 上随机数 $u_1$ 。

如果 $u_1 < A_t^1(\mathbf{X}_t, \mathbf{X}_1^*)$ ,接受 $\mathbf{X}_1^*$ ,即 $\mathbf{X}_{t+1} = \mathbf{X}_1^*$ ,转入(3)。

如果 $u_1 \geq A_t^1(\mathbf{X}_t, \mathbf{X}_1^*)$ ,设定 $2 \leq i \leq 3, \gamma = [1, 0.2, 0.01], \mathbf{C}_t^2 = \gamma_2 \mathbf{C}_t^1$ ,从 $q_t^2(\mathbf{X}_t, \cdot) = N(\mathbf{X}_t, \mathbf{C}_t^2)$ 中抽取一个样本 $\mathbf{X}_2^*$ ;计算接受概率 $A_t^2(\mathbf{X}_t, \mathbf{X}_1^*, \mathbf{X}_2^*) =$

$\min\left\{1, \frac{p(\mathbf{X}_2^*|d)q_t^1(\mathbf{X}_1^*, \mathbf{X}_2^*)[1 - A_t^1(\mathbf{X}_2^*, \mathbf{X}_1^*)]}{p(\mathbf{X}_t|d)q_t^1(\mathbf{X}_t, \mathbf{X}_1^*)[1 - A_t^1(\mathbf{X}_t, \mathbf{X}_1^*)]}\right\}$ ;  
产生 $[0,1]$ 上随机数 $u_2$ ;如果 $u_2 < A_t^2(\mathbf{X}_t, \mathbf{X}_1^*, \mathbf{X}_2^*)$ ,接受 $\mathbf{X}_2^*$ ,即 $\mathbf{X}_{t+1} = \mathbf{X}_2^*$ ,转入(3)。

如果 $u_2 \geq A_t^2(\mathbf{X}_t, \mathbf{X}_1^*, \mathbf{X}_2^*)$ , $\mathbf{C}_t^3 = \gamma_3 \mathbf{C}_t^1$ ,从 $q_t^3(\mathbf{X}_t, \cdot) = N(\mathbf{X}_t, \mathbf{C}_t^3)$ 中抽取一个样本 $\mathbf{X}_3^*$ ;计算接受概率 $A_t^3(\mathbf{X}_t, \mathbf{X}_1^*, \mathbf{X}_2^*, \mathbf{X}_3^*) =$

$\min\left\{1, \frac{p(\mathbf{X}_3^*|d)q_t^1(\mathbf{X}_3^*, \mathbf{X}_2^*)q_t^2(\mathbf{X}_2^*, \mathbf{X}_1^*)}{p(\mathbf{X}_t|d)q_t^1(\mathbf{X}_t, \mathbf{X}_1^*)q_t^2(\mathbf{X}_t, \mathbf{X}_2^*)}\right\}$   
 $\frac{[1 - A_t^1(\mathbf{X}_3^*, \mathbf{X}_2^*)][1 - A_t^2(\mathbf{X}_3^*, \mathbf{X}_2^*, \mathbf{X}_1^*)]}{[1 - A_t^1(\mathbf{X}_t, \mathbf{X}_1^*)][1 - A_t^2(\mathbf{X}_t, \mathbf{X}_1^*, \mathbf{X}_2^*)]}$   
产生区间 $[0,1]$ 上随机数 $u_3$ ;如果 $u_3 < A_t^3(\mathbf{X}_t, \mathbf{X}_1^*, \mathbf{X}_2^*, \mathbf{X}_3^*)$ ,接受 $\mathbf{X}_3^*$ ,即 $\mathbf{X}_{t+1} = \mathbf{X}_3^*$ ,转入(2)。

(3)重复过程(2),直至达到迭代次数 $N$ 。

### 1.3 拉丁超立方抽样

采用 Metropolis 算法和 DRAM 算法求解多维参数反演问题,Markov Chain 容易受到样本初始点的影响,可能在局部最优处达到稳定或者产生难以收敛的问题。

为了保证样本初始点的随机性和均匀性,本研究基于拉丁超立方抽样方法<sup>[21]</sup>对 DRAM 算法进行改进。拉丁超立方抽样方法是一种多维分层随机抽样方法,具有良好的散布均匀性和代表性。假设要在 $m$ 维模型参数 $\alpha$ 先验范围 $[A_i, B_i](i = 1, 2, \dots, m)$ 内抽取 $q$ 组样本,具体步骤如下:

(1)把 $m$ 维模型参数的 $m$ 个先验范围 $[A_i, B_i](i = 1, 2, \dots, m)$ 都均分成 $q$ 个小区间,小区间可记为 $[A_{ij}, B_{ij}](i = 1, 2, \dots, m; j = 1, 2, \dots, q)$ ,总共产生 $m \times q$ 个小区间。

(2)在任一小区间 $[A_{ij}, B_{ij}]$ 中随机抽取一个数,记为 $\alpha_{ij}$ ,总共产生 $m \times q$ 个数,组成矩阵为

$$\Psi_{mq} = \begin{bmatrix} \alpha_{11} & \alpha_{12} & \cdots & \alpha_{1q} \\ \alpha_{21} & \alpha_{22} & \cdots & \alpha_{2q} \\ \vdots & \vdots & \vdots & \vdots \\ \alpha_{m1} & \alpha_{m2} & \cdots & \alpha_{mq} \end{bmatrix}。$$

(3)对矩阵 $\Psi_{mq}$ 中行向量 $[\alpha_{i1}, \alpha_{i2}, \dots, \alpha_{iq}](i = 1, 2, \dots, m)$ 进行随机排序,所得向量记为 $[\beta_{i1}, \beta_{i2}, \dots, \beta_{iq}](i = 1, 2, \dots, m)$ ;由此得到矩阵

$$\Phi_{mq} = \begin{bmatrix} \beta_{11} & \beta_{12} & \cdots & \beta_{1q} \\ \beta_{21} & \beta_{22} & \cdots & \beta_{2q} \\ \vdots & \vdots & \vdots & \vdots \\ \beta_{m1} & \beta_{m2} & \cdots & \beta_{mq} \end{bmatrix}。$$

(4)矩阵 $\Phi_{mq}$ 中每个列向量是一组样本,共抽取得到 $q$ 组样本。

### 1.4 改进型多链 DRAM 算法

改进型多链 DRAM 算法具体步骤:

(1)运用拉丁超立方抽样方法,在模型参数先验范围内随机抽取 $q$ 组初始样本。

(2)分别以 $q$ 组初始样本作为初始点采用 DRAM 算法进行迭代运算,得到 $q$ 条 Markov Chains。

(3)对上面 $q$ 条 Markov Chains 求平均值作为最终结果。

## 2 算例应用

### 2.1 算例概述

均质各向同性地下含水层中,二维地下水水质对流-扩散方程<sup>[24-25]</sup>可以表述为:

$$\frac{\partial}{\partial x}\left(D_x \frac{\partial C}{\partial x}\right) + \frac{\partial}{\partial y}\left(D_y \frac{\partial C}{\partial y}\right) - \frac{\partial}{\partial x}(uC) = \frac{\partial C}{\partial t}, \quad (7)$$

式中: $C$ 为测点 $(x, y)$ 在 $t$ 时长的污染物的浓度,mg/L; $t$ 为污染物排放后开始预测的时长,d; $D_x$ 、 $D_y$ 分别为纵向、横向的扩散系数, $\text{m}^2/\text{d}$ ; $u$ 为地下含水层水流平均流速, $\text{m}/\text{d}$ 。

对于污染物瞬时点源排放模式,方程的解析解可表示为:

$$C(x, y, t) = \frac{M}{4\pi kh \sqrt{D_x D_y t}} \exp\left[-\frac{(x-ut)^2}{4D_x t} - \frac{y^2}{4D_y t}\right], \quad (8)$$

式中: $M$ 为污染物的排放量,g; $h$ 为含水层的厚度,m; $x$ 、 $y$ 分别为预测点距排放点的纵向、横向距离,m; $k$ 为含水层的有效孔隙率。

假设研究区域为均质各向同性地下含水层(建立坐标系,如图1)。含水层厚度为 $h = 1 \text{ m}$ ,

孔隙率  $k=0.3$ ,  $x$  轴代表含水层中水流的方向, 水流速度为  $u=5\text{ m/d}$ ; 纵向弥散系数为  $1.5\text{ m}^2/\text{d}$ , 横向弥散系数为  $0.3\text{ m}^2/\text{d}$ 。待求参数真值: 污染物泄漏点坐标为  $(X_0=200\text{ m}, Y_0=200\text{ m})$ , 污染源强度  $M=1\ 000\text{ g}$ , 污染源排放时长  $T_0=110\text{ d}$ 。观测点  $D$  坐标为  $(X_1=800\text{ m}, Y_1=215\text{ m})$ 。

2.2 监测数据

以构建的算例模型的计算值作为观测值, 设定 2017 年 5 月 1 日上午 10:00 点作为初始监测时间, 其后每隔 2 d 测量一次, 直至 2017 年 5 月 21 日上午 10:00 点结束, 将研究区水文地质参数带入

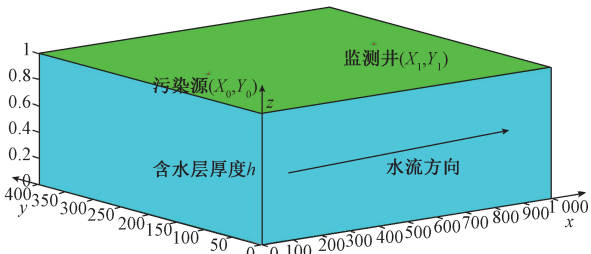


图 1 算例模型示意图  
Figure 1 Sketch of example model

式(8), 可计算得出观测点  $D$  的污染物浓度  $d_i$ , 以此作为观测点  $D$  的污染物的观测值, 如表 1。

表 1 观测点 $D$ 处污染物观测浓度序列						
Table 1 Sequence value of pollutant concentration at observation point $D$						
时刻	5 月 1 日	5 月 3 日	5 月 5 日	5 月 7 日	5 月 8 日	5 月 10 日
质量浓度/( $\text{mg}\cdot\text{L}^{-1}$ )	0.014 80	0.061 20	0.179 6	0.381 1	0.594 0	0.690 7
时刻	5 月 12 日	5 月 15 日	5 月 17 日	5 月 19 日	5 月 21 日	
质量浓度/( $\text{mg}\cdot\text{L}^{-1}$ )	0.608 0	0.410 6	0.215 5	0.088 90	0.029 16	

2.3 结果及比较

根据先验信息, 污染源强度  $M$ 、污染源位置坐标  $(X_0, Y_0)$  和污染源排放时长  $T_0$  4 个待求参数的取值范围分别为:  $400\text{ g}\leq M\leq 1\ 600\text{ g}$ ,  $-200\text{ m}\leq X_0\leq 600\text{ m}$ ,  $185\text{ m}\leq Y_0\leq 215\text{ m}$ ,  $60\text{ d}\leq T_0\leq 160\text{ d}$ 。运用 Matlab 软件编程, 对比分析不同算法的求解结果。

2.3.1 Metropolis 算法求解结果

采用经典 Metropolis 算法构建马尔科夫链对模型进行求解, 马尔科夫链选取初始点不同, 4 个待估参数  $M, X_0, Y_0, T_0$  的迭代曲线差别很大。以待估参数  $M$  为例, 随机选取两组不同初始样本, 如样本点 1:  $(M, X_0, Y_0, T_0)=(827.44, 592.05, 190.59, 80.77)$ ; 样本点 2:  $(M, X_0, Y_0, T_0)=(687.50, 141.15, 213.89, 146.97)$ , 待估参数  $M$  迭代曲线见图 2。

由图 2(a)可知, 在初始样本点为  $M_1=827.44\text{ g}$  时, Metropolis 算法迭代到 5 000 次时, 参数  $M$  值逐渐趋于稳定, 并最终在  $M=1\ 450\text{ g}$  处收敛, 但是该值与真值 ( $M=1\ 000\text{ g}$ ) 相差较大, 误差达 45%。另由图 2(b)可知, 在样本初始点为  $M_2=687.50\text{ g}$  时, Metropolis 算法迭代到 50 000 次时, 参数  $M$  值仍未收敛。因此运用 Metropolis 算法求解污染源强时, 求解结果受样本初始点的影响较大, 容易出现局部最优或者难以收敛的问题。同理, 在求解另外 3 个待求参数  $X_0, Y_0, T_0$  时也会出现类似问题。

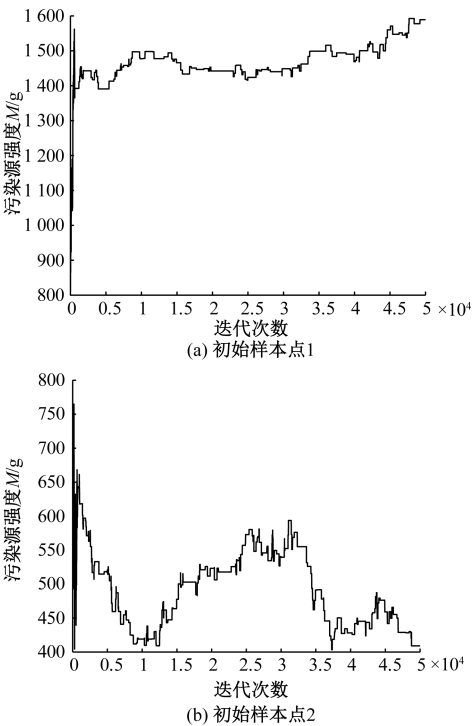


图 2 基于 Metropolis 算法的不同初始点  
模型参数  $M$  迭代曲线图

Figure 2 Iterative curves of model parameter  $M$  based on Metropolis algorithm with different initial values

2.3.2 多链 Metropolis 算法求解结果

为防止反演结果的局部最优, 构建基于拉丁超立方抽样的多链 Metropolis 算法求解算例 (设定 40 条链, 单链长度为 50 000)。污染源强度  $M$ 、污染源位置坐标  $(X_0, Y_0)$  和污染源排放时长  $T_0$  的反演迭代曲线见图 3。

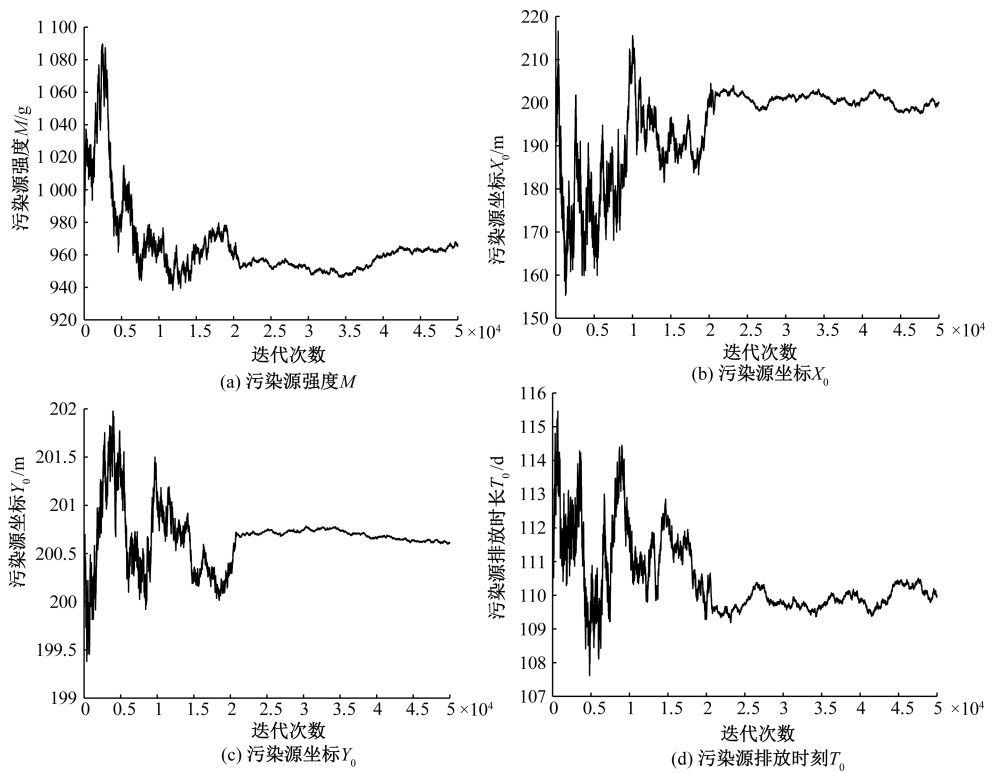


图 3 基于多链 Metropolis 算法的模型参数迭代曲线

Figure 3 Iterative curves of model parameters based on multi-chain Metropolis algorithm

由图 3 可知,当基于拉丁超立方抽样的多链 Metropolis 算法迭代到 25 000 次时,4 个参数反演结果均基本在真值处达到稳定且收敛,而且通过程序验证迭代结果不受样本初始点选取的影响,有效解决了经典 Metropolis 算法受初始点的影响导致不收敛或局部最优的问题,表明该算法具有较强的稳定性。剔除前 25 000 次不稳定结果,对剩余的 25 000 次计算结果进行后验统计分析,见表 2。

由表 2 可知,基于拉丁超立方抽样的多链 Metropolis 算法得出的 4 个待估参数  $M$ 、 $X_0$ 、 $Y_0$ 、 $T_0$  的均值与真值的误差分别为 4.36%、0.18%、0.35%、0.08%,中值误差分别为 4.30%、0.15%、0.35%、0.06%,表明基于拉丁超立方抽样方法的 Metropolis 算法可实现计算结果的全局最优,有效提高反演结果的准确性。

2.3.3 改进型多链 DRAM 算法求解结果

多链 Metropolis 算法虽然有效提高了反演结

果的准确性,但是迭代次数需达到 25 000 次时才能收敛,反演效率相对偏低,为此笔者构建基于拉丁超立方抽样的多链 DRAM 算法(设定 40 条链,单链长度为 12 000)进行参数求解。4 个参数的反演遍历均值图见图 4。

由图 4 可知,改进型多链 DRAM 算法迭代到 8 000 次时,4 个参数取值均基本达到稳定,与基于拉丁超立方抽样的多链 Metropolis 算法(25 000 次时迭代稳定且收敛)相比,速率提高 68%,并且通过程序验证迭代结果不受样本初始点的影响,表明基于拉丁超立方抽样的 DRAM 方法具有较高的反演效率及稳定性。

对于改进型多链 DRAM 算法的反演结果,剔除前 10 000 次不稳定结果,对剩余的 2 000 次模型参数进行后验统计分析,见表 3。由表 3 可知,该算法在提高反演效率的前提下,同样可保证反演结果的准确性。

表 2 基于多链 Metropolis 算法的模型参数后验统计结果

Table 2 Posterior statistical results of model parameters based on multi-chain Metropolis algorithm

项目	均值	均值误差/%	中值	中值误差/%
污染源强度 $M/g$	956.38	4.36	957.02	4.30
污染源坐标 $X_0/m$	200.36	0.18	200.29	0.15
污染源坐标 $Y_0/m$	200.69	0.35	200.69	0.35
泄漏时间距离首次监测时间 $T_0/min$	109.91	0.08	109.93	0.06

表 3 基于改进型多链 DRAM 算法的模型参数后验统计结果  
Table 3 Posterior statistical results of model parameters based on improved multi-chain DRAM algorithm

项目	均值	均值误差%	中值	中值误差%
污染源强度 $M/\text{g}$	969.93	3.01	1013.83	2.62
污染源坐标 $X_0/\text{m}$	197.99	1.01	186.77	0.93
污染源坐标 $Y_0/\text{m}$	200.84	0.42	200.38	0.41
泄漏时间距离首次监测时间 $T_0/\text{d}$	109.60	0.37	107.37	0.35

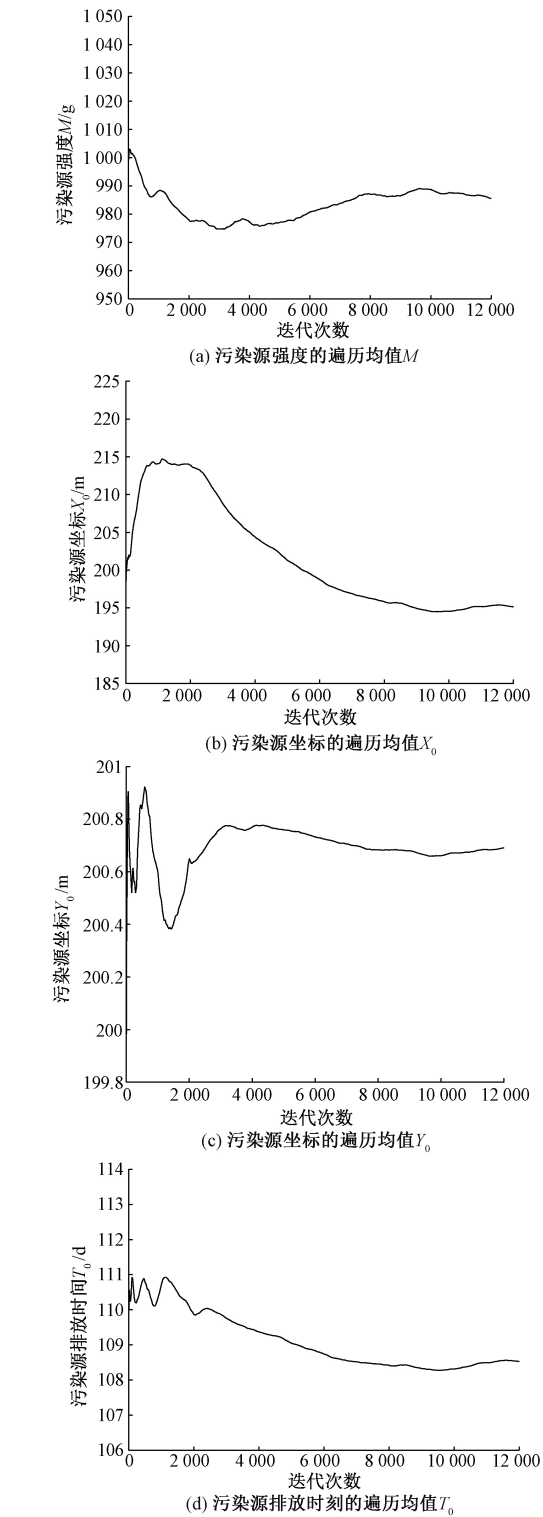


图 4 基于改进型多链 DRAM 算法的模型参数遍历均值图  
Figure 4 Ergodic mean plots of model parameters based on improved multi-chain DRAM algorithm

3 结论

(1)利用 Bayesian 公式以概率语言解决突发性地下水污染反问题,可有效获取污染源信息(污染源强度、污染源位置和污染源排放时长)。

(2)运用经典 Metropolis 算法求解污染源信息时,求解结果受样本初始点的影响较大,容易出现局部最优或者难以收敛的问题,反演得到的未知参数估计值误差较大。

(3)基于拉丁超立方抽样的 Metropolis 算法可实现反演结果的全局最优,有效提高反演结果的准确性,但是效率相对低下。基于拉丁超立方抽样的 DRAM 算法在保证反演结果准确性的前提下,可显著提高反演效率,其可靠性和稳定性均优于经典 Metropolis 算法。

参考文献:

[1] HAARIO H, SAKSMAN E, TAMMINEN J. An adaptive Metropolis algorithm [J]. Bernoulli, 2001, 7 (2): 223-242.

[2] ZENG L Z, SHI L S, ZHANG D X, et al. A sparse grid based Bayesian method for contaminant source identification[J]. Advances in water resources, 2012, 37: 1-9.

[3] 陈海洋,腾彦国,王金生,等. 基于 Bayesian-MCMC 方法的水体污染识别反问题[J]. 湖南大学学报(自然科学版),2012,39(6): 74-78.

[4] 杨海东,肖宜,王卓民,等. 突发性水污染事件溯源方法[J]. 水科学进展, 2014, 25(1): 122-128.

[5] 顾文龙,卢文喜,张宇,等. 基于贝叶斯推理与改进的 MCMC 方法反演地下水污染源释放历史[J]. 水利学报,2016, 47(6): 772-779.

[6] 朱嵩,毛根海,刘国华,等. 改进的 MCMC 方法及其应用[J]. 水利学报,2009,40(8):1019-1023.

[7] RUZEK B, KVASNICKA M. Differential evolution algorithm in the earthquake hypocenter location [J]. Pure and applied geophysics, 2001, 158: 667-693.

[8] 牟行洋.基于微分进化算法的污染源项识别反问题研究[J].水动力学研究与进展(A 辑),2011,26 (1):24-30.

[9] 张双圣,强静,刘喜坤,等. 基于贝叶斯-微分进化算法的污染源识别反问题 [J]. 山东大学学报(工学版),2018,48(1):131-136.

[10] HOLLAND J. Adaptation in Natural and Artificial Sys-



- terms [M]. Ann Arbor, MI: University of Michigan Press, 1975.
- [11] 闵涛,周孝德,张世梅,等. 对流-扩散方程源项识别反问题的遗传算法[J]. 水动力学研究与进展, A辑, 2004, 19(4): 520-524.
- [12] 王秀峰,卢桂章. 系统建模与辨识[M]. 北京: 电子工业出版社, 2004.
- [13] METROPOLIS N, ROSENBLUTH A W, ROSENBLUTH M N, et al. Equation of state calculations by fast computing machines[J]. The journal of chemical physics, 1953, 21(6): 1087-1092.
- [14] HASTINGS W K. Monte Carlo sampling methods using Markov chains and their applications[J]. Biometrika, 1970, 57(1): 97-109.
- [15] TIERNEY L, MIRA A. Some adaptive Monte Carlo methods for bayesian inference[J]. Statistics in medicine, 1999(18): 2507-2515.
- [16] MIRA A. Ordering and improving the performance of Monte Carlo Markov Chains[J]. Statistical science, 2002(16): 340-350.
- [17] HAARIO H, SAKSMAN E, TAMMINEN J. An adaptive Metropolis algorithm[J]. Bernoulli, 2001, 7(2): 223-242.
- [18] HAARIO H, LAINE M, MIRA A. DRAM: Efficient adaptive MCMC [J]. Statistics and computing, 2006, 16(4): 339-354.
- [19] 张江江. 地下水污染源解析的贝叶斯监测设计与参数反演方法 [D]. 杭州: 浙江大学, 2017.
- [20] VRUGT J A. Markov chain Monte Carlo simulation using the DREAM software package: Theory, concepts, and Matlab implementation [J]. Environmental modeling & software, 2016(75): 273-316.
- [21] 戴英彪. 基于拉丁超立方试验设计的事故再现结果不确定性分析[D]. 长沙: 长沙理工大学, 2011.
- [22] 师黎, 王艺. 基于贝叶斯估计的初级视皮层光栅朝向编码研究 [J]. 郑州大学学报(工学版), 2012, 33(6): 1-4.
- [23] 刘厂, 赵俊翔, 胡海. 基于双贝叶斯估计的动态威胁运动状态估计 [J]. 郑州大学学报(工学版), 2017, 38(2): 55-60.
- [24] 陈崇希, 李国敏. 地下水溶质运移理论及模型 [M]. 武汉: 中国地质大学出版社, 1996.
- [25] 郑春苗, BENNETT G D. 地下水污染物迁移模拟 [M]. 2 版. 北京: 高等教育出版社, 2009.

## Improved Multi-chain DRAM Algorithm Based on Latin Hypercube Sampling for Inverse Problems of Underground Water Pollution

ZHANG Shuangsheng<sup>1,3</sup>, QIANG Jing<sup>2</sup>, LIU Hanhu<sup>1</sup>, LIU Xikun<sup>3</sup>, SUN Shaohua<sup>4</sup>

(1.School of Environment Science and Spatial Informatics, China University of Mining and Technology, Xuzhou 221116, China; 2.School of Mathematics, China University of Mining and Technology, Xuzhou 221116, China; 3.Xuzhou City Water Resource Administrative Office, Xuzhou 221018, China; 4.Shandong province urban water supply and drainage monitoring center, Jinan 250100, China)

**Abstract:** Aiming at the problem of the result affected by samples' initial values with classical MCMC algorithm (Metropolis algorithm), when the inverse problems of underground water pollution were solved by Bayesian statistical methods, an improved multi-chain delayed rejection and adaptive Metropolis (DRAM) algorithm based on latin hypercube sampling was presented. An underground water pollution source identification model was built by coupling Bayesian statistical methods to two-dimensional water quality convection-diffusion equation. An example of a pollutant in the underground aquifer discharged instantly was put forward, and the pollution source information including source's position, intensity and discharging time was solved by Metropolis algorithm, multi-chain Metropolis algorithm and improved multi-chain DRAM algorithm respectively. The example showed that the inversion results affected by initial values with Metropolis algorithm were locally optimal or difficult to convergence; While the multi-chain Metropolis algorithm could significantly improve the accuracy of the inversion results, the inversion efficiency was relatively low; On the contrary, the improved multi-chain DRAM could significantly improved the inversion efficiency under the condition of accuracy (improved by 68% compared with the multi-chain Metropolis algorithm), realizing double improvement of inversion accuracy and efficiency.

**Key words:** two-dimensional water quality model; Bayesian-Markov chain Monte Carlo simulation; latin hypercube sampling; delayed rejection and adaptive Metropolis algorithm; pollution source identification