

求 van Genuchten 模型参数的 AM-MCMC 方法^①

石晓蕾¹, 徐绍辉^{1*}, 廖凯华²

(1 青岛大学环境科学系, 青岛 266071; 2 南京大学水科学系, 南京 210093)

摘要: 采用基于自适应采样算法的马尔科夫链蒙特卡罗方法 (简称 AM-MCMC) 来估算描述土壤水分特征曲线的 van Genuchten 模型的参数, 并推求出模型参数的后验分布, 从而为模型参数的不确定性分析提供依据。结果表明, 对于 van Genuchten 模型而言, 采用 AM-MCMC 算法能得到模型参数后验均值和方差分布, 并且能推求出模型参数的置信区间, 所以用这种算法来求解 van Genuchten 方程的参数是行之有效的, 为求解 van Genuchten 模型参数提供了一种新的思路。

关键词: 土壤水分特征曲线; van Genuchten 模型; AM-MCMC 算法

中图分类号: S152

土壤水分特征曲线表征了土壤压力水头与水分含量之间的函数关系, 是土壤最重要的水力特性之一。在研究土壤水入渗、蒸发、滞留、土壤侵蚀及溶质运移过程中, 土壤水分特征曲线是推求各种水分运动参数的重要工具, 因此土壤水分特征曲线一直是土壤物理学家们关注的重点。

由于土壤水分特征曲线的影响因素复杂, 至今还没有从理论上建立土壤含水量和土壤基质势之间的关系, 通常用经验公式来描述。van Genuchten 提出的描述土壤水分特征曲线方程匹配大部分土壤水分特征曲线的形状, 因此得到广泛应用。然而, 因 van Genuchten 方程是一个复杂的非线性方程, 其中的参数较多, 并且参数拟合属于非线性问题, 许多学者对土壤水分特征曲线的拟合做过研究。如最小二乘法^[1]、非线性单纯形法^[1]、单纯形调优法^[2]等。但用上述这些方法会遇到求解停止或参数为负以及计算效率低等问题。

描述土壤水分特征曲线的模型, 在数值求解过程中存在很多不确定因素, 如模型参数和模型输入的不确定性、模型本身的不确定性、模型对复杂的实际问题进行简化而产生的不确定性以及观测资料的不确定性等。因此, 采用传统的方法对 van Genuchten 方程的参数进行求解, 往往由于人为因素和计算复杂等因素的影响使估计结果带有较大的误差, 不易得到最优解。

马尔科夫链蒙特卡罗 (Markov chain Monte Carlo, MCMC) 方法是近年来随着计算机技术以及贝叶斯理

论改进而发展起来的用于推求随机变量后验分布的行之有效的方法, 它能方便地处理非常复杂的模型。本文采用基于自适应算法的马尔科夫蒙特卡罗方法来估计 van Genuchten 方程的参数, 并对模型参数的后验分布进行了分析。

1 MCMC 方法

马尔科夫链蒙特卡罗方法是通过建立平稳分布为 $\pi(x)$ 的马尔科夫链来获得 $\pi(x)$ 的样本。基于随机采样的 MCMC 方法能够避开似然函数的解析式和归一化常数的求解, 利用随机抽取随机参数样本使其收敛到贝叶斯后验分布, 根据抽得的样本便可统计出参数的后验分布^[7]。

马尔科夫链蒙特卡罗方法 (MCMC) 作为一种随机模拟方法早已为许多学者应用到气象、物理、通信等领域。但是, MCMC 法用于模型参数不确定性分析是近几年来才发展起来的。Tyler 和 Lucy^[3]提出了 3 种 MCMC 方法, 分别用于水文模型并对这 3 种方法的模拟结果进行了比较。Yves 与 Jeffrey^[4]发表了关于使用一种随机步长的 Metropolis 算法得到参数的最优范围。邢贞相等^[5]建立了基于 AM-MCMC 算法的贝叶斯概率洪水预报模型, 并给出了预报的方差; 卫晓婧等^[6]以汉江玉带河流域的实例证明基于马尔科夫链-蒙特卡罗算法的改进通用似然不确定性估计方法能推求出性质优良的预测区间, 从而更真实合理地反映水文模型

①基金项目: 国家自然科学基金项目 (41171183) 和水利部公益性行业科研专项经费项目 (201201024) 资助。

* 通讯作者 (xsh631201@126.com)

作者简介: 石晓蕾 (1985—), 女, 山东威海人, 硕士研究生, 主要从事地下环境中水流和溶质运移的数值模拟。E-mail: sxlei69@126.com

的不确定性;陆乐等^[7]利用基于马尔科夫链(MCMC)采样法,结合地下水数值模拟的 MODFLOW 软件,提出贝叶斯方法,用于水文地质参数识别这一复杂的非线性优化问题。

采样算法对 MCMC 方法有很大的影响。常用的采样算法有 Metropolis-Hastings 算法^[8]、Gibbs 采样^[9]和 Adaptive Metropolis (简称 AM) 算法^[10]等。本文采用的是 AM 算法。如果完全进行随机采样,需要的样本量非常大,该算法通过控制采样方式,明显提高了计算效率。

AM 算法是 Haario 提出的一种改进的 MCMC 采样器。与传统的 Metropolis-Hastings 算法和 Gibbs 采样相比,AM 算法不再需要事先确定参数的先验分布,在抽样过程中依据马尔科夫链的历史抽样信息自适应地调整协方差矩阵,并且可以并行运算。它是由后验参数的协方差矩阵来估算。

AM 算法将先验分布定义为参数空间的多维正态分布的形式,后验参数的协方差矩阵(先验密度)在每次迭代后自适应地调整。这样,计算生成的新的参数样本 $\theta^* \sim N(\theta_i, C_i)$ 。

2 van Genuchten 模型

描述土壤水分特征曲线的 van Genuchten 模型,其基本表达式为:

$$\frac{\theta(h) - \theta_r}{\theta_s - \theta_r} = \left[\frac{1}{1 + (\alpha h)^n} \right]^m \quad (1)$$

式中, $\theta_{(h)}$ 是土壤体积含水量 (cm^3/cm^3); θ_s 和 θ_r 分别为土壤饱和含水量和残余含水量; h 为压力水头 (cm); α 是与进气吸力相关的参数; m 和 n 是土壤水分特征曲线的形状系数 ($m = 1 - 1/n$)。

3 求解 van Genuchten 模型参数的 AM-MCMC 算法

用 AM-MCMC 算法来求解 van Genuchten 方程的参数可分为以下步骤:

对所选的模型方程进行分析,把方程中的 θ_s 、 θ_r 、 α 、 n 看作变量,公式可写成: $\theta = f(\theta_r, \theta_s, \alpha, n, h)$, $h = f(\theta_r, \theta_s, n, h)$, 根据以往的相关资料确定方程各参数的初始取值范围如表 1。

表 1 van Genuchten 模型参数的取值范围

Table 1 Value ranges of van Genuchten model parameters

参数	取值范围
$\theta_r (\text{cm}^3/\text{cm}^3)$	0 ~ 0.3
$\theta_s (\text{cm}^3/\text{cm}^3)$	0.3 ~ 0.6
$\alpha (\text{cm})$	0 ~ 1.0, 通常为 0.005 ~ 0.05
n	1 ~ 10.0, 通常为 1.1 ~ 3.5

(2) 似然函数的选择是 AM-MCMC 算法中的一个比较重要的问题,因为它影响到结果的精度和收敛的速度。似然函数的形式灵活多样,一般针对不同的问题进行不同的选择。为了方便与模型进行耦合,本文采用的似然函数为

$$s = 1 / [(\theta - \theta_i)^2 + (h - h_i)^2] \quad (2)$$

(3) 利用 AM-MCMC 采样方法进行初始化, $i = 0$ 。

(4) 采集一个关于 4 个变量的初始样本,结合模型与真值,得到 θ 和 h 的初始估计值。

(5) 参数样本的协方差的计算:如式(3)。在初始 $I \leq i_0$ 次迭代中,协方差矩阵 C_i 取固定值 C_0 。我们可以根据先验分布来确定参数初始协方差矩阵 C_0 的值,之后协方差 C_i 可自适应更新。

$$C_i = \begin{cases} C_0 & i \leq i_0 \\ S_d \text{Cov}(\theta_0, \dots, \theta_{i-1}) + S_d \varepsilon I_d & i \geq i_0 \end{cases} \quad (3)$$

式中, $\text{Cov}(\theta_0, \dots, \theta_{i-1})$ 为已有的所有样本向量的协方差矩阵; S_d 为比例因子,依赖于样本的维数 d ,确保接受概率保持在一个合适范围内,按照 Gelman 等^[20]的建议, S_d 的取值为 $(2.4)^2/d$; ε 为一个较小的数,以确保样本向量的各分量的方差不会降为 0; i_0 是初始抽样次数; I_d 为 d 维单位矩阵。

对于第 $i+1$ 次迭代的协方差的计算公式可以由式(3)推导出来。

$$C_{i+1} = \frac{i-1}{i} C_i + \frac{sd}{i} (\overline{i\theta_{i-1}\theta_{i-1}^T} - (i+1)\overline{\theta_i\theta_i^T} + \theta_i\theta_i^T + dI) \quad (4)$$

式中, $\overline{\theta_{i-1}}$ 和 $\overline{\theta_i}$ 分别表示前 $i-1$ 次和 i 次的参数抽样的均值。产生新的参数样本 $\theta^* \sim N(\theta_i, C_i)$ 。

(6) 接受概率 a 的计算:

$$a = \min \left\{ 1, \frac{p(y/\theta^*)p(\theta^*)}{p(y/\theta_i)p(\theta_i)} \right\} \quad (5)$$

式中, $p(y/\theta)$ 是参数的似然函数, $p(\theta)$ 是参数的先验分布。

(7) 随机数 u 的生成: $u \sim U[0,1]$, 若 $u < a$ 则接受 $\theta_{i+1} = \theta^*$, 否则 $\theta_{i+1} = \theta_i$ 。

(8) 重复以上 (5) ~ (7) 步, 直到产生足够的样本数为止。

(9) 对采样结果进行分析, 求出不同水头压力下的土壤体积含水量, 并与实测值进行比较。

4 数值算例

本文试验所用的土样采自青岛平度农田的棕壤。用压力膜仪分别测量了 25、60、100、300、1 000、3 000、15 000 cm 压力水头下的体积含水量。所用样本的国际制质地类型是黏壤土, 砂粒含量为 494.9 g/kg, 粉粒含量为 332.9 g/kg, 黏粒含量为 172.2

g/kg, 有机质含量为 14.1 g/kg, 体积质量为 1.46 g/cm³。

该算法中参数的设置: 初始迭代次数为 1 000, 观测值为上述 7 个压力水头下分别测得的 100 个不同原状土样体积含水量的数据, 共 100 个样本。AM 算法的运行次数为 2 000 次。

5 结果与分析

图 1 为土壤残余含水量 θ_r 、土壤饱和含水量 θ_s 、与进气吸力相关的参数 α 和形状系数 n 的 2 000 个样本的采样过程图。从图中我们可以看出参数取值遍历了每个参数的取值空间, 不会局限在某个小范围内, 遍历性很好。

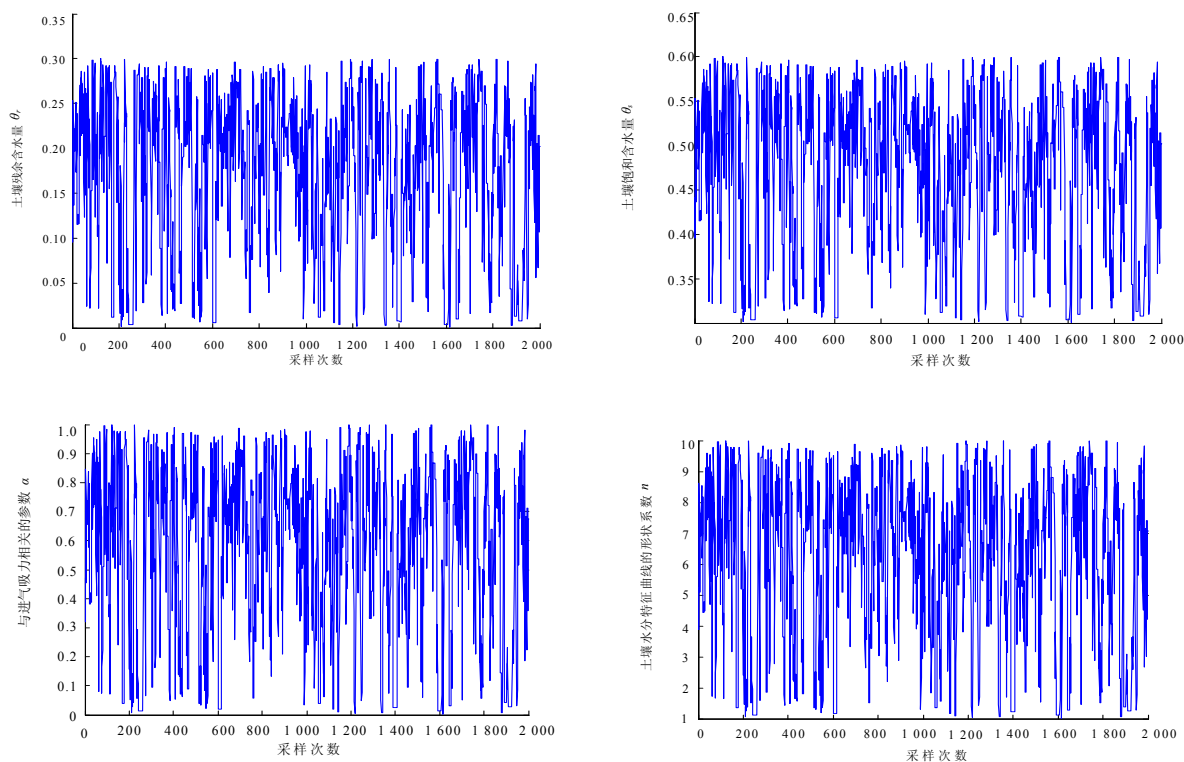


图 1 各参数采样过程

Fig. 1 Sampling processes of parameters

基于 AM-MCMC 算法的模拟不受参数的先验抽样区间的影响, 不依赖参数的推荐分布, 因此算法的遍历性和较高的抽样效率能够得到保证, 图 1 很好地验证了这一理论。

图 2 分别为参数 θ_s 、 θ_r 、 α 、 n 的后验分布平均值和方差的变化过程图。当迭代次数小于 200 时, 各参数的均值和方差波动很大; 大于 1 000 以后, 各参数的

平均值和方差基本达到稳定。从图 1 以及图 2 可以得出这样一个结论: 所获得的参数样本的分布已经收敛到参数的后验分布。

通过运行 AM 算法, 进行 2 000 次迭代, 得到 4 个参数的 2 000 个样本, 并且能够得到它们各自的似然值。图 3 为经 AM 采样方法获得的每组参数所对应的似然值。

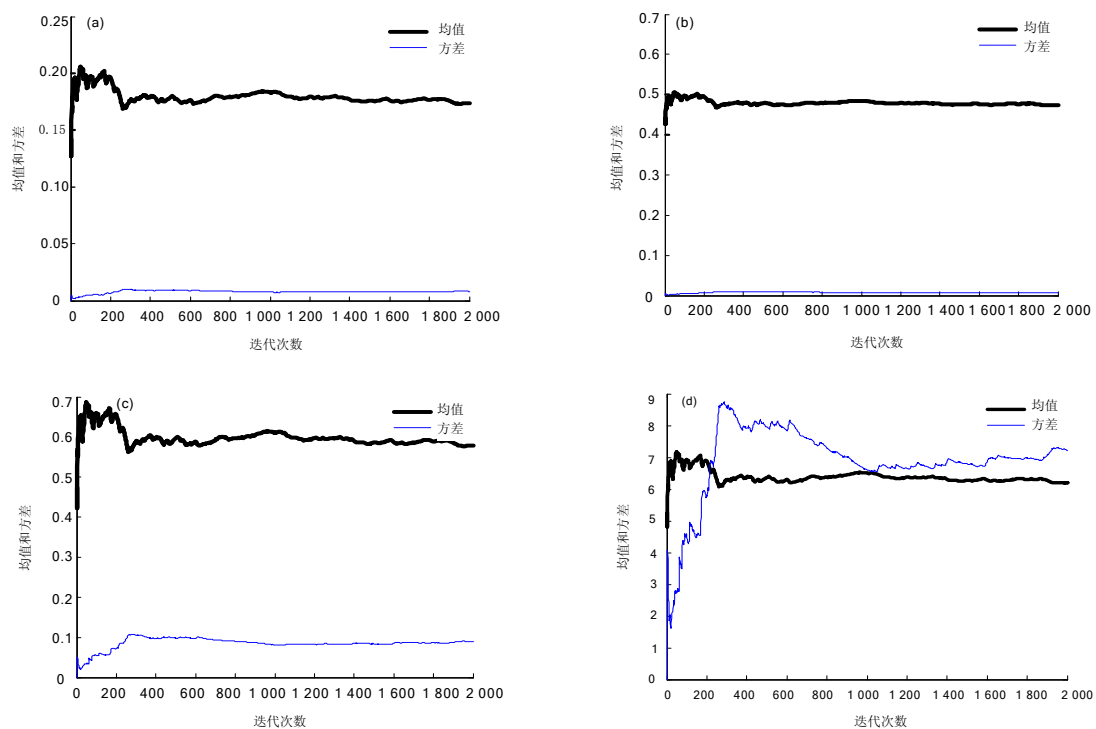


图 2 各参数后验分布平均值与方差的变化过程 (a. θ_r ; b. θ_s ; c. α ; d. n)

Fig. 2 Change processes of posterior distribution means and variances of parameters

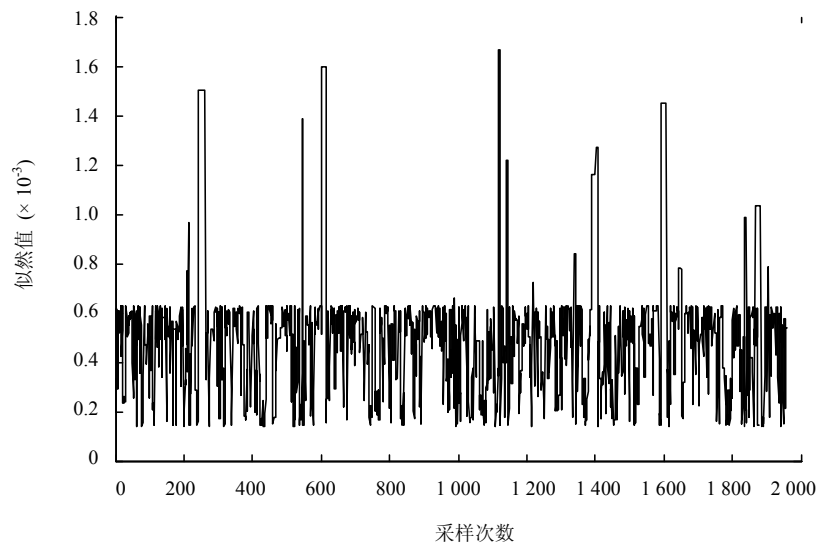


图 3 各组参数的似然值

Fig. 3 Likelihood value of each group parameters

我们知道,在观测条件相同的情况下,对某个变量进行多次重复观测,依据偶然误差的特性,当观测次数 n 无限增大时,则 $\lim_{n \rightarrow \infty} \frac{|\Delta|}{n} = 0$ 。式中 Δ 为观测量的真值与观测值直接的差值,即观测值的真误差。则

$n \rightarrow \infty$ 时算术平均值趋近于真值。但是在实际测量工作中,观测次数总是有限的,但算术平均值较观测值仍更接近于真值。最接近于真值的算术平均值称为最或然值或最可靠值。所以,本文采用算术平均值作为

真值用来检测拟合结果。

图 4 为将 AM-MCMC 算法求解的一组参数代入 van Genuchten 方程所得的拟合曲线、众数以及最小均方误差拟合的曲线与最可靠值的拟合结果。从图 4 可以看出，最大似然值对应的参数组拟合的曲线与最小均方误差处对应点拟合的曲线是重

合的。而众数拟合的曲线与实测值有一定的偏差，这也为进行参数不确定性分析提供了一定的依据。

分别统计了参数后验分布的平均值、最小值、最大值以及 7 个百分点的数据，参数后验分布的统计结果见表 2。

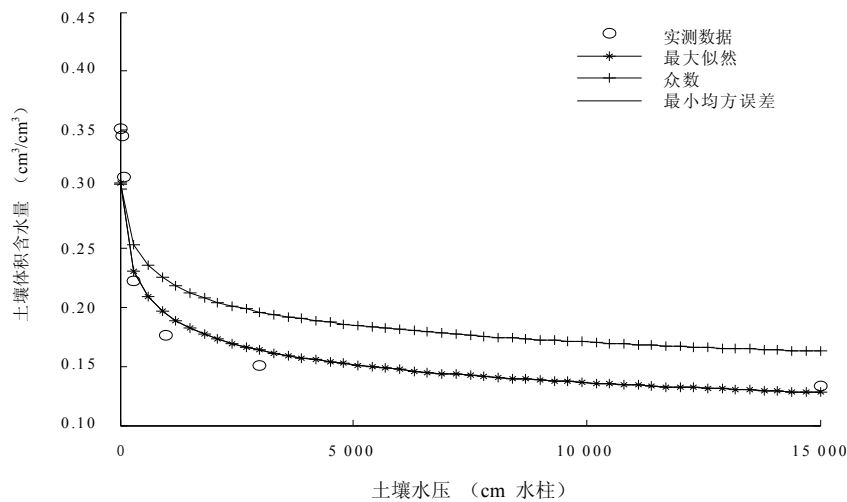


图 4 土壤水分特征曲线拟合

Fig. 4 Fitting of soil moisture characteristic curve

表 2 van Genuchten 模型参数的后验分布

Table 2 Posterior statistics of van Genuchten model parameters

参数	样本数	平均值	最小值	最大值	百分点						
					5%	10%	25%	50%	75%	90%	95%
θ_r	2 000	0.173 7	0.000 3	0.300 0	0.007 6	0.015 3	0.107 5	0.188 9	0.245 6	0.274 7	0.286 3
θ_s	2 000	0.473 7	0.300 3	0.600 0	0.307 6	0.317 0	0.408 4	0.492 1	0.546 2	0.578 4	0.589 0
α	2 000	0.578 9	0.001 1	0.999 9	0.025 2	0.050 9	0.358 5	0.629 6	0.818 8	0.915 7	0.954 4
n	2 000	6.210 5	1.010 1	9.999 1	1.226 5	1.462 7	4.229 6	6.687 4	8.373 2	9.269 5	9.603 8

利用 AM-MCMC 算法的优点之一就是能提供一定置信水平下的预测区间，从而进行参数的不确定性分析，这是使用其他传统方法所不能达到的。所以上表给出了各百分点数据，有助于获得参数的置信区间，从而为不确定性评价和进行风险分析提供依据。

从表 2 还可以看出最小值和最大值与表 1 中参数的取值范围基本一致，说明利用 AM-MCMC 算法能使参数空间得到有效的搜索。因此，可以用 AM-MCMC 算法来对 van Genuchten 模型参数的不确定性问题进行分析。

6 结论

本文采用的 AM-MCMC 算法能得到 van Genuchten

模型参数的后验分布，可以获得某一置信水平的区间预测，也可以依据参数的后验分布对模型参数的不确定性进行研究，从而能更合理地对模型进行贝叶斯预报。因此，AM-MCMC 算法能为模型参数不确定性的定量评价及风险分析提供了有效的途径。运用 AM-MCMC 算法来推求 van Genuchten 方程参数的后验分布是一种新的思路，实例表明这种算法也是合理可靠的。

参考文献：

- [1] 王金生, 杨志峰, 陈家军, 王志明. 包气带土壤水分滞留特征研究. 水利学报, 2000, 31(2): 1-7
- [2] 李春友, 任理, 李保国. 利用优化方法求算 Van Genuchten 方

- 程参数. 水科学进展, 2001, 12 (4): 473-478
- [3] Smith TJ, Marshall LA. Bayesian methods in hydrologic modeling: A study of recent advancements in Markov chain Monte Carlo techniques. *Water Resour. Res.*, 2008, 44: 1-9
- [4] Atchade YF, Rosenthal JS. On adaptive Markov chain Monte Carlo algorithms. *Bernoulli*, 2005, 11(5): 815-828
- [5] 邢贞相, 芮孝芳, 崔海燕, 余美. 基于 AM-MCMC 算法的贝叶斯概率洪水预报模型. *水利学报*, 2007, 38(12): 1 500-1 506
- [6] 卫晓婧, 熊立华, 万民, 刘攀. 融合马尔科夫链-蒙特卡洛算法的改进通用似然不确定性估计方法在流域水文模型中的应用. *水利学报*, 2009, 40(4): 464-480
- [7] 陆乐, 吴吉春, 陈景雅. 基于贝叶斯方法的水文地质参数识别. *水文地质工程地质*, 2008, 35(5): 58-63
- [8] Chib S, Greenberg E. Understanding the Metropolis-Hasting algorithm. *American Statistician*, 1995, 49(4): 327-335
- [9] Smith AFM, Roberts GO. Bayesian computation via the Gibbs sampler and related Markov chain Monte Carlo methods. *Journal of The Royal Statistical Society Series B*, 1993, 55: 3-23
- [10] Haario H, Saksman E, Tamminen J. An adaptive Metropolis algorithm. *Bernoulli*, 2001, 7(2): 223-242
- [11] 刘建立, 徐绍辉, 刘慧. 估计土壤水分特征曲线的间接方法研究进展. *水利学报*, 2004, 2(2): 68-76
- [12] 邵明安, 王全九, 黄明斌. 土壤物理学. 北京: 高等教育出版社, 2006: 71
- [13] 魏义长, 刘作新, 康玲玲, 王云璋, 时明立. 土壤持水曲线 van Genuchten 模型求参的 Matlab 实现. *土壤学报*, 2004, 41(3): 380-386
- [14] 王建平, 程声通, 贾海峰. 基于 MCMC 法的水质模型参数不确定性研究. *环境科学*, 2006, 27(1): 24-30
- [15] 陆乐, 吴吉春. 地下水数值模拟不确定性的贝叶斯分析. *水利学报*, 2010, 41(3): 264-271
- [16] 熊立华, 卫晓婧, 万民. 水文模型两种不确定性研究方法的比较. *武汉大学学报 (工学版)*, 2009, 42(2): 137-142
- [17] 肖建英, 李永涛, 王丽. 利用 Van Genuchten 模型拟合土壤水分特征曲线. *地下水*, 2007, 29(5): 46-47
- [18] 夏卫生, 雷廷武, 刘贤赵, 刘纪根, 潘英华. 土壤水分特征曲线的推算. *土壤学报*, 2003, 40(2): 311-315
- [19] 杨改强, 霍丽娟, 杨国义, 李一菲, 钱天伟. 利用 MATLAB 拟合 van Genuchten 方程参数的研究. *土壤*, 2010, 42(2): 268-274
- [20] Gelman AG. Efficient metropolis jumping rules// Bernardo JM, Berger JO, David AF. *Bayesian Statistics*. New York: Oxford University Press, 1996: 599-608

AM-MCMC Approach to Estimate van Genuchten Model Parameters

SHI Xiao-lei¹, XU Shao-hui¹, LIAO Kai-hua²

(1 *Department of Environmental Science, Qingdao University, Qingdao 266071, China;*

2 *Department of Hydrosience, Nanjing University, Nanjing 210093, China)*

Abstract: This paper adopted the adaptive Metropolis algorithm of Markov chain Monte Carlo approach (AM-MCMC) to estimate the parameters of van Genuchten model, which describes the soil water characteristic curve. The posterior statistics of the van Genuchten model parameters were obtained to estimated the model uncertainty. It was found that the distribution of the posterior means and variances of the model parameters could be obtained. Confidence intervals of the model parameters could be ascertained. It was proved effective and novel in solving van Genuchten equation with this algorithm.

Key words: Soil water characteristic curve, van Genuchten model, AM-MCMC algorithm