

# 随机 GP 法在概念性水文模型参数优选中的应用\*

郝振纯<sup>1</sup>, 谷军方<sup>1,2</sup>, 杜富慧<sup>1</sup>

(1. 河海大学水文水资源与水利工程科学国家重点实验室, 江苏南京 210098;  
2. 邯郸市水利局, 河北邯郸 056000)

**摘要:** 在基于近似梯度及模式搜索法的基础上, 提出了复合上述两种方法的 GP 局部优化方法。以 Nash 确定性系数为目标函数, 对水文模型的参数空间随机搜索后采用 GP 方法优化, 运用参数空间筛选策略, 以获得全局最优解集。上述方法结合导数信息和随机性质的算法, 使优化过程脱离局部极小解从而达到近似全局最优解集。杨楼单元流域应用新安江模型的实例研究结果表明, 随机的 GP 优化方法可以成功的率定概念性水文模型参数。

**关键词:** GP 优化; 随机优化; 参数率定; 新安江模型

**中图分类号:** P349 **文献标识码:** A **文章编号:** 0529-6579 (2009) 06-0018-06

## Application of Stochastic GP Algorithms Optimization to Conceptual Hydrologic Model Parameters

HAO Zhenchun<sup>1</sup>, GU Junfang<sup>2</sup>, DU Fuhui<sup>1</sup>

(1. State Key Laboratory of Water Resources and Hydropower Engineering Science, Hohai University, Nanjing 210098, China;  
2. Bureau of Water Resources of Handan Municipality, Handan 056000, China)

**Abstract:** Combining the approximate gradient-based steepest descent algorithm and the pattern search algorithm, the GP algorithm, a new local optimization algorithm for conceptual hydrologic model parameters is presented. With Nash facticity coefficient as the target function the random search techniques is used for searching parameter space, then optimize the selected parameter set using GP algorithm. The global optimization parameter is achieved by filtering parameter space strategy. The above-mentioned method comprise the derivative information and stochastic properties, make the optimization set escaping the local maximum to the global set. The practical efficiency is verified by using a case in YandLou unite drainage basin. It is shown that parameters of hydrologic model can be automatically calibrated successfully.

**Key words:** GP algorithm; stochastic optimization; parameters calibration; Xinyanjiang model

水文模型参数优选按方式可分为手工优选(如试错法)、自动优选和手工加自动的混合优选方法。手工参数优选算法最终优选的参数值因人而异, 与经验和所受的训练以及对模型结构的理解程度有关, 该方法主要的不足之处是难以判断优选过

程什么时候达到最优以及所得到的解是否为最优解。随着计算机技术的迅猛发展, 参数自动优选方法得到了普遍的应用。

水文模型参数自动优化按优选方法或策略可分为基于导数的优化、基于非导数的优化等方法。

\* 收稿日期: 2008-11-04

基金项目: 国家自然科学基金重点资助项目(40830639); 国家自然科学基金资助项目(40801012); 水利部公益性资助项目(200801027); 教育部“长江学者和创新团队发展计划资助”资助项目(IRT0717)

作者简介: 郝振纯(1958年生), 男, 教授; E-mail: hzchun@hhu.edu.cn

基于导数的优化方法有 Newton 法、共轭梯度法<sup>[1]</sup>、步长加速法及最速下降法等。这些传统的基于导数的优化方法得到的优化结果不可靠。这是由于目标函数的响应表面在多峰的、不连续的, 优化结果对初值有依赖性。

基于非导数的优化有直接搜索法、模式搜索法、单纯形法、Rosen Brock 法、遗传算法<sup>[2]</sup>、加速遗传算法<sup>[3]</sup>、GLUE 方法<sup>[4]</sup>、SCE-UA 复合进化方法<sup>[5-6]</sup>、模拟退火法及方法对比研究等<sup>[7]</sup>。由于基于导数的优化方法存在前面所说的诸多问题, 因此基于非导数的优化方法, 特别是仿生或仿自然优化算法在最近的研究中得到了很大的发展。基于非导数的优化方法虽然能够得到近似的全局最优解, 但上述方法都需要占用很大的计算机资源及较长的计算时间。

基于导数的优化方法由于目标函数高度粗糙且不连续性的响应表面, 使得这种方法在模型参数优化时受到了很大的限制, 但这种方法收敛速度快, 节省计算时间, 此外, 在响应表面连续的情况下, 基于导数的梯度方向可以确定的减少模拟结果与实际水文过程的误差, 而不是无方向的搜索, 使计算过程冗长。

基于非导数的优化方法在运用随机及加入仿生或仿自然方法后, 虽可达到近似的全局最优解集, 但以上方法与基于导数的方法相比, 收敛速度慢, 占用计算机资源多, 对计算机的性能要求高。

因此, 如果将基于导数的优化方法与基于非导数的优化方法结合起来, 选择合适的时机使用随机方法与基于导数的方法, 可以在避免陷入局部最优解集的同时, 减少计算时间, 目前, 这类复合方法研究较少。本文建议采用一种包括上述两种方法的复合搜索策略对概念性水文模型参数进行自动优化, 其优化策略是在模型参数具有物理意义的参数空间随机选取模型参数初值, 在响应表面连续时, 运用梯度法和模式搜索法的复合方法使其快速收敛到最优解集, 此后, 再对这些优化后的局部最优解集进一步优化, 具体的方法在本文后面的部分予以叙述。

## 1 随机 GP 法优化概念性水文模型参数原理

### 1.1 水文模型参数优化函数关系的描述

运用基于导数的方法优化需要将模型输入输出变量转变为用公式表达的函数关系, 根据水文模型模拟过程可建立以下了目标函数表达式:

$$F_i(\theta) = f(Q_{obs}^i, Q_{sim}^i(\theta)) \quad (1)$$

式中  $F_i(\theta)$  为评价第  $i$  次水文过程模拟效果优劣的目标函数值;  $f(\cdot)$  为评价水文过程模拟效果的函数;  $Q_{obs}^i$  为第  $i$  次水文过程的实际输出流量;  $Q_{sim}^i(\theta)$  为第  $i$  次水文过程的模拟输出流量;  $\theta$  为水文模型参数集。

评价水文模型模拟效果的函数  $f(\cdot)$  采用 Nash 与 Sutcliffe 在 1970 年提出了 Nash-Sutcliffe 系数 (模型确定性系数) 来评价模型模拟结果的精度, 确定性系数是归一化表现形式, 更直观的体现了实测与模拟流量过程的拟合程度, 对于单次水文过程, 确定性系数的公式如下:

$$R^2 = \left( 1 - \frac{\sum_{i=1}^N w_i^2 [Q_{obs}^i - Q_{sim}^i(\theta)]^2}{\sum_{i=1}^N w_i^2 [Q_{obs}^i - \bar{Q}_{obs}^i]^2} \right) \times 100\% \quad (2)$$

式中,  $R^2$  为确定性系数,  $w_i$  为第  $i$  个时段的比重,  $N$  为水文过程的时段数。

$$Q_{sim}^i(\theta) = H(Iset^i, \theta, M_{org}) \quad (3)$$

式中,  $H(\cdot)$  为概念性水文模型在输入集与输出集之间建立的映射;  $Iset^i$  为第  $i$  次水文过程降雨、蒸发等水文气候资料的输入集合;  $M_{org}$  为水文模型表示模型模拟水文过程开始时的各含水层的初始值。

$$F(\theta) = \sum_{i=1}^N F_i(\theta) \quad (4)$$

式中,  $F(\theta)$  为评价水文模型模拟效果的总函数值;  $N$  为模型模拟水文过程的总次数。

### 1.2 近似梯度的计算方法

基于导数的方法有 Newton 法, 由于评价函数的不连续, 且目标函数与模型参数是高度非线性的, 因而用 Newton 法求局部最优解时, 会使求得的解不在合理的参数空间内, 即得到的参数超出了其所代表的物理意义的范围, 因而, 本文采用近似梯度的方法确定搜索方向。

计算梯度的公式如下:

$$g(\theta) = \left[ \frac{\partial F}{\partial \theta_1}, \frac{\partial F}{\partial \theta_2}, \dots, \frac{\partial F}{\partial \theta_n} \right]^T \quad (5)$$

式中  $g$  为目标函数  $F$  对模型参数  $\theta$  的梯度;  $n$  为模型参数  $\theta$  包括的参数个数。

由 (1) - (4) 式可知, 目标函数  $F$  中包括概念性水文模型及评价函数等高度非线性函数映射关系, 因此, 无法得到解析的梯度表达式。在已发表的文献中, Gupta 于 1985 年提出了对水文模型不同的产流模式求解表达式, 后期文献大多采用

了这种求导数的方法。但有些模型如新安江模型,其产汇流过程无法求出精确的解析表达式,因此,本文建议采用求极限的方法求梯度的近似值,具体求解公式如下:

$$\frac{\partial F}{\partial \theta_i} = \lim_{\Delta \theta_i \rightarrow 0} \frac{F(\theta_i + \Delta \theta_i) - F(\theta_i)}{\Delta \theta_i} \quad (6)$$

式中,  $\Delta \theta_i$  表示  $\theta_i$  的微增量,其他符号含义与前式相同。

运用近似梯度的方法逐步逼近局部最优点,可以使优化过程始终在使水文模型参数有意义的合理空间内寻优,从而避免因步长过大而使优化解陷入“空洞”的参数空间中。

### 1.3 局部空间的 GP 搜索方法

本文建议的搜索方法是对局部连续的函数响应表面寻优的方法,该方法是基于近似梯度的最速下降法 (the approximate gradient-based steepest descent algorithm) 与模式搜索法 (the pattern search algorithm) 混合优化方法,简称为 GP 搜索法。

其具体方法如下:

1.3.1 基于近似梯度的最速下降 (上升) 法 对于任意给定具有物理意义的模型参数空间的点,基于近似梯度的最速下降法可以较快的找到局部最优解。由于本文采用确定性系数为目标函数,优化的方向是使目标函数增大至最优效率,因此,模型参数的优化表达为:

$$\theta_{\text{next}} = \theta_{\text{now}} + k \frac{g(\theta)}{\|g(\theta)\|} \quad (7)$$

式中,  $\theta_{\text{next}}$  为更新后的模型参数集;  $\theta_{\text{now}}$  更新前的模型参数集;  $k$  为更新步长。

当采用最速下降 (上升) 法时,若步长固定,则实际步长  $k$  影响搜索结果,当  $k$  值较大时,可以保证搜索过程有效接近最小点,但在局部最小附近发生振荡,不能精确收敛;当实际步长  $k$  较小时,搜索效率低,收敛速度慢,因此,本文采取动态调整步长  $k$  的自适应策略,采用以下规则更新  $k$ : 如果目标函数连续上升了  $m$  次 (目标函数的最大值即是最优解),则将  $k$  增大  $p\%$ 。

1.3.2 模式搜索法 上述优化方法中,当目标函数响应表面不连续时,会偏离原来所在的分段连续性的表面而发生跳跃,这可能使一些较大的优化解在优化过程中“遗失”,即漏掉一些分段连续表面的局部最优解,为了避免上述现象发生,有必要对最速下降 (上升) 法予以改进,本文建议结合模式搜索法来完善局部寻优问题。

模式搜索法主要由交替进行的“探测搜索”

和“模式移动”组成。探测搜索的出发点称为参考点,探测搜索的目的是在参考点的周围寻找比它更好的点,从而确定一个有利的前进方向。对于目标函数极小化问题,如果能够找到这样的点,那末称为基点。如果有基点的函数值小于参考点的函数值,自然想到,从基点出发,沿从参考点到基点的方向,目标函数有可能继续下降。这样的向量称为“模式”。下一步就进行模式移动,模式移动的起点是基点,它的终点是新的参考点。于是探测搜索与模式移动就可以交替进行下去。迭代开始时,基点和参考点相重合,并都在初始点外。经过探测搜索得到新的基点,然后经过模式移动得到新的参考点。再“探测”,再“移动”,迭代点将逐渐向极小点靠近。模式搜索法适用于参数比较少的无约束问题。

1.3.3 混合的 GP 搜索法 本文建议改进的方法如下:当以近似最速下降 (上升) 法优化遇到不利于优化方向的目标函数值时返回到上一个优化解并以其为基点,采用模式搜索法,以近似梯度为基点的方向,缩小步长搜索,直至找到更优化的解,若由于计算误差而使得在近似梯度方向得不到更优解时,则沿着近似梯度的反方向变步长搜索寻找更优解,若未找到更优解则搜索终止。在整个优化过程中目标函数始终朝着更优的方向搜索,以获取局部最优解。

### 1.4 随机优化策略

概念性水文模型参数集与目标函数存在高度非线性的复杂函数关系,本文以目标函数的响应表面进行大量数据分析后有以下结论:

(1) 目标函数的响应表面是分段连续的粗糙的多峰表面。

(2) 受到模型参数间的相互关系及水文模型复杂性的影响,在模型参数空间中存在着没有对应的目标函数值的区域。即所谓的数据“空洞”。

(3) 当以 Nash 确定性系统为目标函数时,各个离散的分段表面片段的局部最优解存在较大差别。

以上实验结果在文章后面部分予以说明。根据上述数据实验的结果,本文提出了一种筛选模型参数空间的优化方法,即通过对目标函数局部优化的结果确定目标函数较优参数的分布空间,通过反复不断的筛选,使参数空间逐步收敛到最优解。优化具体步骤见图 1。

随机优化程序中  $Limit$  值  $Nsum$  值对于优化效果有明显的影 响,其中  $Limit$  初值宜选取较小值,

以便能以较大的概率选中高效率存在的区域并运用 GP 法进行局部优化。*Limit* 更新算法影响函数的收敛速度，采用如下公式计算：

$$Limit = mean\{F(\theta_{opt,i})\} \quad i = 1, 2, \dots, m \quad (8)$$

*Nsum* 是每次优化时的取样密度，对优化结果有影响，取样密度过大，会占用过多的计算资源及时间，从而降低优化效率，若取样密度不足，则优化的结果会“遗失”一些局部优化的解集，使结果不能足够接近全局最优，因此，*Nsum* 的采用下式进行计算：

$$Nsum \geq d_{\Theta} / \min\{d_{opt,i}\} \quad i = 1, 2, \dots, m \quad (9)$$

式中  $d_{\Theta}$  为模型参数空间  $\Theta$  的欧氏距离； $d_{opt,i}$  为第  $i$  个样本初始值与 GP 优化后的局部最优解的欧氏距离。

上述程序中的数值选取方法将在下面的例子中予以说明。

在传统方法基础上，随机 GP 法做了以下改进：将近似梯度法与模式搜索法相结合解决了在分段连续的表面上局部最优化问题。随机空间筛选策略的引入可以使优化过程逐步逼近全局最优解。

## 2 应用实例

### 2.1 研究区概况

本文选取杨楼水文站控制的单元流域作为研究区域，简称为杨楼单元流域，该流域属淮河流域内一子流域，地处南北气候过渡带，气象条件复杂。

### 2.2 水文模型及参数说明

新安江模型是一个半分散性概念性水文模型，在国内有着广泛的应用。应用新安江模型需按下垫面和地形等情况对流域进行分块（称为分单元或子流域），然后将每个子流域的出口流量演算至整个流域的出口断面。这样单元流域的新安江模型有 14 个参数，新安江模型的结构及原理参考相关文献<sup>[8]</sup>。

本文选取流域进行模拟计算时，未考虑流域内河道汇流。模型应用 14 个参数进行模拟，各参数的物理意义如下： $K$  为流域蒸散发系数； $IM$  为流域不透水面积比； $B$  为蓄水容量曲线系数； $WM$  为流域缺水量； $WUM$  为上层最大含水量； $WLM$  为下层最大含水量； $C$  为深层蒸散发系数； $EX$  为自由水蓄水容量曲线系数； $SM$  为自由水蓄水容量； $KI$  为壤中流出系数； $KG$  为地下水出流系数； $CI$  为壤中流消退系数； $CG$  为地下径流消退系数； $CS$  为河网消退系数。各参数的取值范围见表 1。

其中各参数间有以下关系： $K > 0$ ； $WM > WUM > 0$ ； $SM > 0$ ； $KI > 0$ ； $(1 - KI) \geq KG > 0$ ； $CS > 0$ ； $CI > CS$ ； $CI > CG$ 。

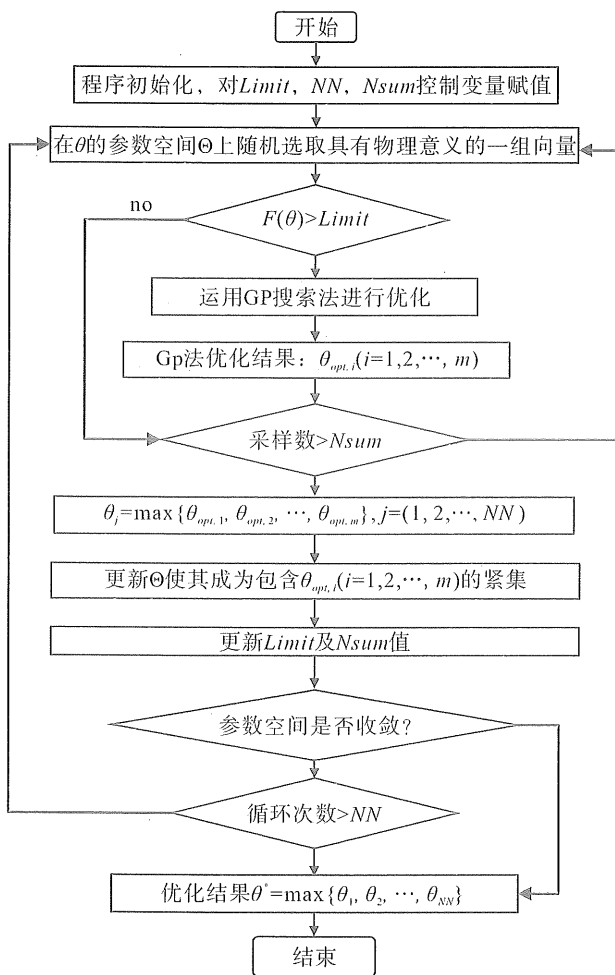


图 1 随机 GP 优化法程序图

Fig. 1 Optimization program of stochastic GP algorithms

表 1 水文模型参数取值范围

Table 1 Value of parameter

参数	$K$	$IM$	$B$	$WM$	$WUM$	$WLM$	$C$	$EX$	$SM$	$KI$	$KG$	$CI$	$CG$	$CS$
最大值	0.5	0	0.1	50	0.5	5	0.05	0	1	0.01	0	0.1	0.8	0.01
最小值	1.0	0.1	2.0	250	50	125	0.25	2.0	120	0.51	1	0.8	1.0	0.91

### 2.3 参数的率定及验证

杨楼流域径流年际变化悬殊, 年内分配不均, 区域内土质主要为沙质壤土。土壤渗透力强, 蒸发量大, 产流只发生在雨季。本文采用杨楼单元流域

出口断面有实测流量的日资料进行模型的率定和验证, 其中, *Limit* 和 *Nsum* 的初值分别为  $-500$ ,  $10^6$ 。参数优选结果见表 2, 资料的选用及率定结果见表 3。

表 2 水文模型参数优化结果

Table 2 Results of parameters value optimized by stochastic algorithms

<i>K</i>	<i>IM</i>	<i>B</i>	<i>WM</i>	<i>WUM</i>	<i>WLM</i>	<i>C</i>	<i>EX</i>	<i>SM</i>	<i>KI</i>	<i>KG</i>	<i>CI</i>	<i>CG</i>	<i>CS</i>
1.0	0	0.1	216.6	38.5	117.9	0.14	0.89	118.6	0.2	0.8	0.63	1	0.77

表 3 模型参数率定及验证表

Table 3 Performance of the calibrated and validated parameters

率定期 年份	实际洪峰 $/(m^3 \cdot s^{-1})$	模拟洪峰 $/(m^3 \cdot s^{-1})$	洪峰误差 /%	峰现时间 误差/d	Nash 效率 系数/%
1986	0.28	0.19	-31.6	+1	74.8
1987	0.43	0.39	-8.9	-1	95.7
1989	0.54	0.42	-21	0	92.2
1990	2.1	2.0	-4.2	-1	73.2
2005	8.93	7.54	-15.5	0	95.5
验证期 年份					
2006	5.16	4.72	-8.6	0	88.1
2007	6.58	5.52	-16.1	0	85.5

验证期的模拟效果见图 2、图 3 所示。

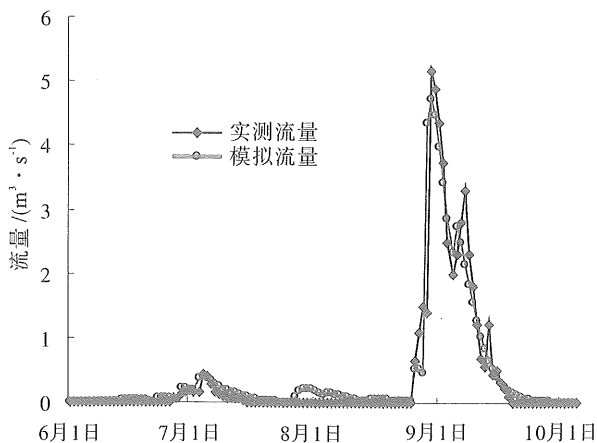


图 2 2006 年模拟与实测流量图

Fig. 2 The simulated and observed hydrographs for 2006

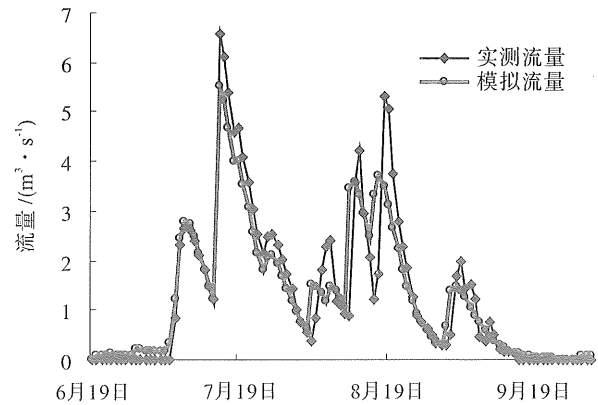


图 3 2007 年模拟与实测流量图

Fig. 3 The simulated and observed hydrographs for 2007

## 3 结 论

本文选取了 Nash-Sutcliffe 系数作为评价概念性水文模型模拟的水文过程的目标函数, 给出了目标函数对水文模型参数梯度的近似计算方法, 提出了基于近似梯度的最速下降法与模式搜索法相结合的模型参数空间局部优化方法 (GP 搜索法), 运用对模型参数空间的筛选策略从而得到模型参数全局最优化。杨楼流域实验结果表明, 该方法可以有效的找到模型参数的近似全局最优解集。

### 参考文献:

- [1] GUPTA V K, SOROOSHIAN S. The relationship between data and the precision of parameter estimates of hydrologic models [J]. Journal of Hydrology, 1985, 81: 57-77.
  - [2] 谭炳卿. 水文模型参数自动化优选方法的比较分析 [J]. 水文, 1996 (5): 8214.
- TAN B Q. Comparison of automatic optimization algorithms for hydrologic model parameters [J]. Journal of Hydrology, 1996 (5): 8214.

(下转第 26 页)

且方程 (1) 化为:

$$\Delta_h(r(t)(\Delta_h[x(t) + p(t)x(t - \tau)])) + \sum_{i=1}^m Q_i(t)f_i(x(t - \sigma_i)) = 0, t \in h\mathbb{N}_0 \quad (5)$$

其中  $m \geq 1$  为整数,  $\tau, \sigma_i \in h\mathbb{N}_0$  为常数;  $r, p, Q_i \in C_{rd}(h\mathbb{N}_0, \mathbb{R})$ ,  $r(t) > 0$ ,  $f_i \in C(\mathbb{R}, \mathbb{R})$ ,  $i = 1, \dots, m$ 。

**推论 2** 假设存在非负常数  $c_1$  和  $c_2$ , 使得  $c_1 + c_2 < 1$ ,  $-c_2 \leq p(t) \leq c_1$ , 且

$$\sum_{j=\frac{t_0}{h}}^{\infty} \sum_{k=\frac{t_0}{h}}^{j-1} \frac{|Q_i(jh)|}{r(kh)} < \infty, i = 1, 2, \dots, m$$

则 (5) 存在有界非振动解。

#### 参考文献:

- [1] HILGER S. Analysis on measure chains—a unified approach to continuous and discrete calculus [J]. Results Math, 1990, 18(1–2): 18–56.
- [2] AGARWAL R, BOHNER M, O'REGAN D, PETERSON A. Dynamic equations on time scales: A survey [J]. J Comput Appl Math, 2002, 141(1–2): 1–26.
- [3] BOHNER M, PETERSON A. Dynamic Equations on Time Scales [M]. Boston: Birkhäuser, 2001.
- [4] BOHNER M, PETERSON A. Advances in Dynamic Equations on Time Scales [M]. Boston: Birkhäuser, 2003.
- [5] MATHSEN R M, WANG Q R, WU H W. Oscillation for neutral dynamic functional equations on time scales [J]. J Difference Equ Appl, 2004, 10(7): 651–659.
- [6] ZHU Z Q, WANG Q R. Frequency measures on time scales with applications [J]. J Math Anal Appl, 2006, 319(2): 398–409.
- [7] ZHU Z Q, WANG Q R. Existence of nonoscillatory solutions to neutral dynamic equations on time scales [J]. J Math Anal Appl, 2007, 335(2): 751–762.
- [8] HUANG H, WANG Q R. Oscillation of second-order nonlinear dynamic equations on time scales [J]. Dynam Systems Appl, 2008, 17(3–4): 551–570.
- [9] YU Z H, WANG Q R. Asymptotic behavior of solutions of third-order nonlinear dynamic equations on time scales [J]. J Comput Appl Math, 2009, 225(2): 531–540.
- [10] ZHU S M, LI Y H, RONG Z X. Region qualitative analysis of predator-prey systems on time scales [J]. Ann. Differential Equations, 2008, 24(1): 121–126.
- [11] KULENOVIC M R S, HADZIOMERSPAHIC S. Existence of nonoscillatory solution of second order linear neutral delay equation [J]. J Math Anal Appl, 1998, 228(2): 436–448.
- [12] ZHOU Y. Existence for nonoscillatory solutions of second-order nonlinear differential equations [J]. J Math Anal Appl, 2007, 331(1): 91–96.

(上接第 22 页)

- [3] 陈炯烽, 张万昌. 概念性水文模型遗传算法多目标参数优选研究 [J]. 水利水电技术, 2007, 38(6): 5–11.  
CHEN J F, ZHANG W C. Study of multi-objective optimization for hydrologic model using generic algorithms [J]. Water Resources and Hydropower Engineering, 2007, 38(6): 5–11.
- [4] YANG J, REICHERT P, ABBASPOUR K C. Hydrological modeling of the Chaohe in China: Statistical model formulation and Bayesian inference [J]. Journal of Hydrology, 2007, 340: 167–182.
- [5] DUAN Q, SOROOSHIAN S, GUPTA V K. Effective and efficient global optimization for conceptual rainfall-runoff models [J]. Water Resources Research, 1992, 28(4): 1015–1031.
- [6] DUAN Q, SOROOSHIAN S, GUPTA V K. Optimal use of SCE-UA global optimization method for calibrating watershed models [J]. Journal of Hydrology, 1994, 158: 265–284.
- [7] 杨晓华, 杨志峰, 郦建强, 等. 水文模型参数识别算法研究及展望 [J]. 自然科学进展, 2006, 16(6): 657–661.  
YANG X H, YANG Z F, LI J Q, et al. Development of parameters identification for hydrologic model [J]. Progress in Natural Science, 2006, 16(6): 657–661.
- [8] 赵人俊, 王佩兰. 新安江模型参数的分析 [J]. 水文, 1988(6): 228.  
ZHAO R J, WANG P L. Analysis of Xin-an-jiang model [J]. Journal of Hydrology, 1988(6): 228.