

基于并行遗传算法的新安江模型参数优化率定方法

武新宇, 程春田, 赵鸣雁

(大连理工大学 土木水利学院, 辽宁 大连 116024)

摘要: 本文结合新安江模型参数的特点, 以洪峰流量、峰现时间和洪水总量的合格率为评价目标, 定义了评价洪水性能目标的模糊合格率, 提出了新安江模型参数率定的并行遗传算法, 并在微机集群环境下, 利用JAVA语言进行了算法编程。串行和并行遗传算法计算结果的比较表明, 本文提出的并行遗传算法可以大大缩短优化过程的时间, 得到较为稳定的模型参数。

关键词: 并行计算; 遗传算法; 参数率定; 新安江模型; 集群

中图分类号: P334 **文献标识码:** A

概念性降雨径流模型参数率定一直是水文预报的重要内容, 国内外研究得较多的参数率定优化方法主要有Rosenbrock^[1]、SCE-UA^[2,3]、遗传算法(GAs)^[4]等。由于前两种方法有可能陷于局部最优, 因此自文献[4]首次将GAs方法应用于流域水文预报模型参数率定以来, 遗传算法作为模拟生物在自然环境中的遗传和进化过程而形成的一种自适应全局优化概率搜索算法, 因其简单通用、鲁棒性强、可以处理非解析式的目标函数和约束等特点, 已成为水文模型参数率定方面一个活跃研究方向^[4~7]。文献[7]结合遗传算法和模糊优选模型提出了新安江模型参数率定新方法, 有效的提高了参数率定的效率。但将产汇流分开率定的方式不利于实际应用, 需要在进行汇流参数率定之前, 对率定好的产流结果进行必要调整, 以消除对汇流参数率定的影响; 若将产汇流作为整体进行率定, 由于参数变量增加而导致了解空间急剧增加, 增加了利用遗传算法寻优的难度, 这将不可避免的带来过长的计算时间和获得理想解的难度, 因此有必要探索更为有效的参数率定方法。

自20世纪70年代计算机技术在流域水文模型参数建模领域应用以来, 计算机现已成为流域运行管理最基本的工具。现在计算机性能价格比已经得到极大提高, 利用集群计算机技术提高流域水文建模的效率已经成为可能。考虑到遗传算法天然并行性和集群计算的高速并行性, 本文结合新安江模型参数率定特点, 提出了新安江模型参数率定的并行遗传算法^[8~10], 以提高算法的计算速度和求解质量。为了区别解的非唯一性, 针对洪水预报规范要求的评价洪水性能3个目标: 洪水总量、洪峰、峰现时间合格率, 提出和定义了二种模糊合格率。论文工作采用Java语言进行编程, 基于SQLServer 2000数据库, 利用JPVM消息传递接口^[11]在Windows 2000微机集群上实现基于粗粒度并行遗传算法的新安江模型参数的自动率定, 计算结果和获得的并行信息表明, 本文提出的方法是可行的, 显著提高了求解速度和质量。

1 新安江模型参数优化率定的并行遗传算法

1.1 新安江模型 新安江模型是赵人俊教授于1973年提出的分散性模型^[12], 主要用于湿润与半湿润地区, 它把全流域分成许多个单元流域, 对每个单元流域作产汇流计算, 得出单元流域的出口流量过程。再

收稿日期: 2003-10-27

基金项目: 国家自然科学基金资助(50479055)

作者简介: 武新宇(1979-), 男, 吉林长春人, 博士生, 主要研究方向为并行计算与网格计算。

进行出口以下的河道洪水演算，得到该单元出流过程，所有单元流域的出流过程相加，得出流域出口的总出流过程。本文采用其三水源模型，采用蓄满产流计算产流量；径流分为地面、地下和壤中流3种；采用3层蒸散发计算模型，3种水源均按线性水库计算河网总入流，河网汇流采用分段马斯京干连续演算法。模型共有16个参数，包括7个产流参数：上层张力水容量 U_m ，下层张力水容量 L_m ，深层张力水容量 D_m ，张力水蓄水容量曲线方次 B ，不透水面积比例 I_m ，蒸发能力折算系数 K ，深层蒸散发系数 C ；9个汇流参数：表土自由水蓄水容量 S_m ，表土自由水蓄水容量曲线方次 E_x ，自由水蓄水水库对地下水的出流系数 K_g ，自由水蓄水水库对壤中流的出流系数 K_i ，壤中流的消退系数 C_i ，地下水水库的消退系数 C_g ，河网蓄水量的消退系数 C_s ，马斯京干法的单元河段的两个参数 K_e 、 X_e 。

1.2 并行遗传算法设计

1.2.1 编码和解码 将并行遗传算法应用于新安江模型参数率定，首要工作是要将问题的解表示成符合遗传算法结构形成的编码方式。遗传算法通常用染色体表示问题的解，一个染色体通常由多个基因组成。新安江模型16个参数组成的参数集相当于一个染色体，每一个参数变量相当于一个基因。可以用二进制、整数和实数变量表示问题的编码，本文采用十进制整数。

用 Z_t ($t=1, 2, \dots, 16$)表示新安江模型有待率定的16个参数， $Z_{t, \min}$ 、 $Z_{t, \max}$ 分别为参数 t 的最大值和最小值，欲使十进制编码精度为 λ ，则应把变量取值区间划分为 $1/\lambda$ 等份，得到的 i 个分点对应编码 i ，由编码组成的整数列进行各种遗传操作。

1.2.2 约束处理 应用马斯京干法进行河道演进时，参数 K_e 、 X_e 必须满足下式的条件，否则会导致马斯京干法河道演进计算公式的失效。

$$2K_e X_e \leq t \leq 2K_e - 2K_e X_e \quad (1)$$

式中： t 为计算时段步长

在产生初始种群和交叉，变异计算中出现违反此约束的情况时，一律重新进行该项操作，直到符合约束条件为止。

1.2.3 适应值函数 适应值计算因实际问题而异，也是遗传算法具有广泛适应性的主要原因。根据洪水预报规范，洪水总量、洪峰、峰现时间合格率是评价参数率定好坏的重要标准，上述3个目标很难用解析式表示成目标函数。当采用上述目标进行洪水率定性能评价时，由于参数之间相互依赖性，往往出现不同参数具有相同合格率，即产生了参数率定过程中优化解非唯一性问题，这时使用常规的方法难以对个体适应度进行评价，为此需要引入新的处理方法。

一般来说，理想的参数模拟结果应满足两个基本条件：具有较高的合格率和对于不合格场次的洪水误差累计和较小。为此，引入评价目标的“模糊合格率”概念，其隶属函数图形如图1。对于峰量和洪水总量合格的隶属度函数值 y_i 定义为

$$y_i = \begin{cases} 1.0 & \delta_i \leq 0.2 \\ 1.25\lambda(1.0 - \delta_i) & 0.2 < \delta_i \leq 1.0 \\ 0 & \delta_i > 1.0 \end{cases} \quad (2)$$

δ_i 为峰量或洪水总量的预报相对误差绝对值，注意到仅仅当 δ_i 不大于20%表示此项合格，因此上述定义反映了洪水预报规范的要求。

是相对于某一目标不合格洪水总的模糊合格率的控制参数，它的取值应使所有不合格洪水的模糊合格率总和小于1，即定义(2)应使实际的洪水合格率大小不变，而又能根据误差大小区别开相同合格率标准的洪水性能。值由下式确定

$$=4.0/N \quad (3)$$

式中：N为参与率定洪水总数。

式(3)隐含在满足最低乙级方案标准(总的洪水合格率 75%)下，不合格洪水的模糊合格率总和将不大于1。

类似地，峰现时间模糊合格的隶属度定义为

$$y_i = \begin{cases} 1.0 & \tau_i \leq 1 \\ \lambda(10 - \tau_i)/9 & 1 < \tau_i \leq 10 \\ 0 & \tau_i > 10 \end{cases} \quad (4)$$

τ_i 为峰现时间差，其隶属函数图形如图2。

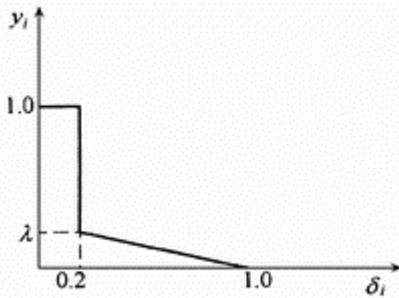


图1 峰量和洪水总量的模糊合格率定义

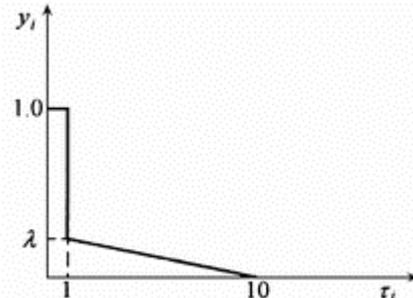


图2 峰现时间的模糊合格率定义

根据式(2)和式(4)，对于第j组染色体，其模糊合格率为

$$x_{ij} = \sum_{i=1}^N y_i^j / N \quad (i=1, 2, 3; j=1, 2, \dots, P_{size}) \quad (5)$$

注意到GAs的每一循环的个体适应度评价、最佳染色体选择、交叉、变异等过程其实是一有限方案多目标评价问题，为了快速方便得到较好的染色体适应度评价，引入多目标模糊优选模型^[7]。

记经过交叉、变异产生的染色体个数为n，多目标评价目标数为m。某一染色体对应着一组模型参数，据此可求得参与率定所有洪水场次的洪峰流量、峰现时间和场次洪水总量，根据水文水情预报规范，相应可得到该组模型参数所对应的洪峰流量合格率、峰现时间合格率及场次洪水总量合格率的大小。若第j个染色体对应的评价目标值用 x_{ij} 表示($i=1, 2, \dots, m; j=1, 2, \dots, n$)，则可得到有限方案的多目标评价矩阵 $X=(x_{ij})_{m \times n}$ ，上述问题可用多目标模糊优选模型进行求解。为了增加目标可比性，需要对目标作规格化，对效益性和成本性目标，分别用式(6)和式(7)。

$$r_{ij} = x_{ij} / x_{i \max} \quad (6)$$

$$r_{ij} = 1 - x_{ij} / x_{i \max} \quad (7)$$

式(6)和式(7)中 $x_{i \max} = \bigvee_{j=1}^n x_{ij}$ ，规格化后转化成相对隶属度矩阵 $R = (r_{ij})_{m \times n}$ 。方案的隶属度为

$$u_j = \frac{1}{1 + \frac{\sum_{i=1}^m [w_i (r_{ij} - 1)]^2}{\sum_{i=1}^m (w_i r_{ij})^2}} \quad (8)$$

式中： w_i 为目标的权重， $i=1, 2, \dots, m$ ；染色体的适应度大小可用隶属度 u_j 来表示。

通常权重的选取根据决策者的经验及规范要求，本文3个目标按等权考虑，即峰值、峰现时间和洪水总量的权重均为1/3。

1.2.4 交叉运算 本文采用部分离散交叉。设 P_c 为交叉概率，对经过选择后得到的 P_c ，母体进行两两随机配对。然后，产生随机数，若小于 P_c 则随机设置某一基因之后的位置为交叉点，交换这一点以后的所有基因以形成新的个体。

1.2.5 变异运算 本文采用非均匀变异，将变异算子与进化代数联系起来，使得在进化的初期，变异的范围相对较大，进行均匀随机搜索；而随着代数的增加，变异的范围越来越小，越来越接近于局部搜索。

1.2.6 ($\mu+$)选择策略 ($\mu+$)选择策略根据种群内 μ 个染色体个体产生一个个体(根据交叉和变异操作)，然后从 $\mu+$ 个个体中再选择 μ 个最优个体。采用扩大采样空间的选择模式，使双亲和经过交叉变异产生的后代均有同样的生存竞争机会，依照双亲和后代组成的新群体的各染色体的适应度大小，在保持原有种群规模大小的前提下淘汰适应度小的个体。

1.2.7 终止条件 终止条件主要有三种，一是预设最大迭代代数 T_{\max} ，当达到最大代数时终止；二是在连续若干代以后染色体的适应度没有明显的改进时，亦可终止；三是染色体的适应度达到预定的要求时，即可终止。

1.2.8 粗粒度并行遗传算法实现模型 并行遗传算法的实现模型主要分为主从式、粗粒度和细粒度模型。其中粗粒度模型又称为分布式或孤岛式模型，是适应性最广的遗传算法并行化模型。整个种群分成多个子种群，各子种群在不同的处理器上相对独立的并发执行进行操作，种群间通过迁移进行联系，即各子种群交换若干个体以引入其它子种群的优秀基因。粗粒度并行遗传算法不但加速了遗传算法的搜索过程，而且由于种群规模扩大和各子种群相对隔离，使子群多样性得以丰富和保持，减少了未成熟收敛的可能性，从而提高了求解质量和计算速度。同时，粗粒度模型通信开销较小，适于运行在通信带宽较低的集群系统上。

(1)子种群规模。在总的种群规模 $m_s \times pop_s$ (m_s 为子种群个数， pop_s 为子种群规模)一定的情况下， m_s 的适当增加有利于提高优化的质量。本文的问题中，取 $m_s=8$ ， $pop_s=50$ 可得到较稳定结果。

(2)子种群间连接拓补。采用双向环拓补方式，得到的优化结果较好，即每个种群与其两侧相邻的子种群交换个体，如图3。

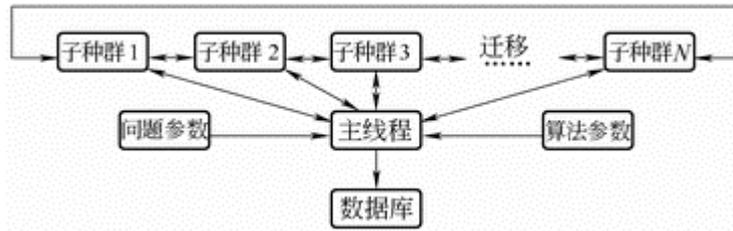


图3 并行遗传算法双向环拓扑结构

(3)子种群间通信方式。子种群间通信分同步和异步方式，本文采用较常用的同步方式，即在一定的进化代，进行迁移操作。

(4)迁移策略。每个子种群每隔预定的代数(epoch)选择最优的 n_m 个个体向外“移民”，并从其它种群接收优秀个体以替换自身较差的个体。本文的算法中，迁移间隔epoch过小，不利于得到高质量的解，取epoch=20, $n_m=2$ 。

2 计算方法

2.1 计算过程 (1)由一个主线程数据库读取数据，包括流域分单元信息；各雨量站信息；模型参数的取值范围；率定期逐日逐站降雨量，流域逐日蒸发量；各场洪水逐时段逐站降雨量和实际入库流量过程。(2)主线程分配若干子线程，每个实现一个子种群，并将(1)中读取的数据传递给各子种群。(3)各个子种群独自进行进化计算，将每一代的最优个体传递给主线程；主线程比较各子种群传来的个体，得到各代的最优个体。(4)在预设的子种群迁移代，各子种群将最优的一部分个体传递给其它各子种群；并从其它各子种群接收的个体，用以替代该子种群内较差的一部分个体。(5)达到终止条件后，主线程和所有子种群的线程结束。对计算过程中得到的最优解进行校核，得出最终结果。

计算流程如图4。

3 应用和结果

对双牌水库1984~1995年共34场洪水进行率定，并以1999~2000年11场洪水对结果进行校核。采用的参数为：编码精度=0.001，交叉概率0.65，变异概率0.05，选择操作中 $\mu=$ ，进化代数100。并行遗传算法参数：子种群数8，子种群规模50，迁移间隔20，迁移规模2。

为了比较串行和并行遗传算法计算效率和解的质量，测试了不同种群规模对计算结果的影响；为消除遗传算法因随机计算对结果带来的影响，对于每一组参数，分别连续计算5次，计算结果如表1。从表1可以看出，对于串行遗传算法，当种群规模小于800时，参数率定和检验结果均不是很稳定；并且随着种群规模的减少，解的质量可能会变得更差；这说明，尽管较小的种群规模也可能得到较好的率定结果，但为了减少寻优过程时间，减少种群规模将会损害解的质量；另一方面也说明，为了保证获得较可靠的率定结

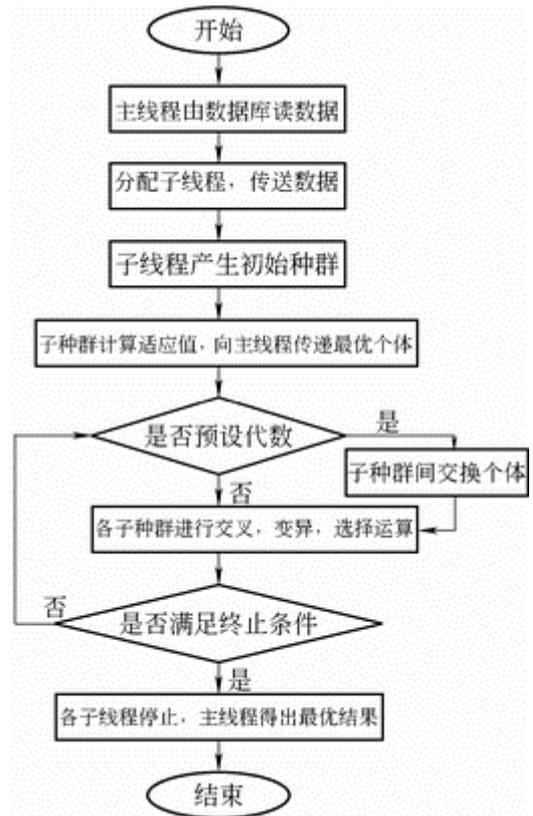


图4 基于并行遗传算法的新安江模型计算流程

果, 需要保证一定种群规模。对于PGA, 较小的种群规模也可获得较为稳定的结果, 种群为 50×8 时, 结果介于SGA种群规模600与800之间, PGA的种群缩小为SGA的 $2/3 \sim 1/2$, 主要原因在于并行遗传算法种群相对隔离保持了种群多样性, 算法的迁移增强了种群适应能力; 同时, 由于计算机集群技术的采用, 计算时间大大缩短(从SGA种群规模800的95min、种群规模600的73min缩短到现在的7min)。

表1 SGA和PGA计算结果比较

模型	种群规模	进化代数	机器数	次数	时间/min	率定合格率(%)			检验合格率(%)			
						水量	洪峰	峰时	水量	洪峰	峰时	
SGA	200	100	1	1	24	97	85	88	90	81	100	
				2		94	79	88	72	63	100	
				3		94	85	85	100	90	100	
				4		97	79	88	72	63	100	
				5		94	79	88	90	72	100	
	400				1	49	94	85	85	100	100	100
					2		97	82	88	63	72	100
					3		94	85	85	100	100	100
					4		94	85	85	100	100	100
					5		97	82	88	90	81	100
	600				1	73	94	85	88	90	81	100
					2		97	82	88	90	72	100
					3		94	85	85	90	100	100
					4		94	85	88	90	90	100
					5		94	85	88	100	81	100
	800				1	95	94	85	88	100	90	100
					2		97	85	88	90	90	100
					3		97	85	88	90	90	100
					4		94	88	88	90	100	100
					5		94	85	88	90	90	100
1000				1	118	94	88	88	100	100	100	
				2		97	85	88	90	90	100	
				3		97	85	88	90	90	100	
				4		94	85	88	90	90	100	
				5		94	88	88	90	90	100	
PGA	50	100	8	1	7	94	85	88	90	90	100	
				2		94	85	88	100	100	100	
				3		94	85	88	100	100	100	
				4		94	85	85	100	90	100	
				5		94	85	85	90	81	100	

4 结论

遗传算法已成为概念洪水预报模型参数率定的有效方法, 然而随着模型参数增加, 传统的遗传算法求解效率将随之下降, 因此需要寻求更为有效的计算方法。近些年来, 随着计算机性能价格比不断提高, 构

建简单的集群微机系统对于普通用户已成为可能, 由此为快速进行模型参数率定提供了可能。

本文计算结果表明, 为了获得稳定的理想解, 串行遗传算法需要保证一定种群规模, 否则会损害解的质量; 基于集群的并行遗传算法由于将集群计算的并行性与遗传算法的天然并行性相结合, 大大提高了算法的计算速度和求解质量。应用本文方法, 使用者只需要根据流域的自然地理特性和降雨径流资料的实际情况, 选定好模型参数大致范围, 剩下参数率定工作可以全部交由程序自动完成, 大大方便和简化了模型参数率定过程。由于计算效率提高, 用户现在可以在较短时间完成一次率定过程, 这样可以在有限时间对不同模型参数进行敏感性测试, 从而可以发现更为深刻的模型规律, 有利于提高洪水预报精度。

参考文献:

- [1] Rosenbrock H H. An automatic method for finding the greatest or least value of function [J]. Computer Journal, 1960, 3: 175-183.
- [2] Qingyun Duan, Soroosh Sorooshian, Vijai Gupta. Effective and efficient global optimization for conceptual rainfall-runoff models [J]. Water Resources Research, 1992, 28(4): 1015-1031.
- [3] Hapuarachchi H A P, 李致家, 王寿辉. 湖泊科学 [J]. 2001, 13(4): 304-314.
- [4] Wang Q J. The genetic algorithm and its application to calibrating conceptual rainfall-runoff models [J]. Water Resources Research, 1991, 27(9): 2467-2471.
- [5] Franchini M. Using a genetic algorithm combined with a local search method for the automatic calibration of conceptual rainfall-runoff models [J]. Hydrological Sciences Journal, 1996, 41(1): 21-40.
- [6] Franchini M, Galeati G. Comparing several genetic algorithm schemes for the calibration of conceptual rainfall-runoff models [J]. Hydrological Sciences Journal, 1997, 42(3): 357-379.
- [7] Cheng C T, Ou C P, Chau K W. Combining a fuzzy optimal model with a genetic algorithm to solve multi-objective rainfall-runoff model calibration [J]. Journal of Hydrology, 2002, 268: 72-86.
- [8] 王小平, 曹立明. 遗传算法——理论、应用与软件实现 [M]. 西安: 西安交通大学出版社, 2002.
- [9] 郭彤城, 暮春棣. 并行遗传在无线通讯网基站优化选址中的应用 [J]. 系统工程理论与实践, 2003, (2): 57-61
- [10] 张宝琳, 等. 数值并行计算原理与方法 [M]. 北京: 国防工业出版社, 1999.
- [11] Adam J. Ferrari JPVM: Network Parallel available from: <http://www.cs.virginia.edu/~ajf2j/jpvm.html>.
- [12] 赵人俊. 流域水文模拟 [M]. 北京: 水利电力出版社, 1984.

Parameter calibration of Xinanjiang rainfall-runoff model by using parallel genetic algorithm

WU Xin-yu, CHENG Chun-tian, ZHAO Ming-yan
(Dalian University of Technology, Dalian 116024, China)

Abstract: A parallel genetic algorithm (PGA) for calculation of Xinanjiang rainfall-runoff model is proposed.

The algorithm evaluates the fitness function based on the definition of fuzzy qualified ratios of floods. The PGA is written in Java language and executed in a cluster of PCs. The comparison of calculation results shows that the proposed method is remarkably better than serial genetic algorithm. The time consumption for optimization is greatly reduced and the more stable parameters can be attained.

Key words: parallel computation; genetic algorithm; calibration; parameter; Xinanjiang model; cluster