

新安江模型参数率定的多核并行遗传算法

王伟*, 金新峰, 刘洋, 余豪

中国电建集团中南勘测设计研究院有限公司, 湖南 长沙

收稿日期: 2021年12月3日; 录用日期: 2022年2月2日; 发布日期: 2022年2月28日

摘要

新安江模型参数众多, 率定计算工作量大、传统率定参数精度难以满足实际生产需要的特点。本文提出了一种参数率定的多核并行遗传算法, 以洪峰流量、峰现时间和洪水总量为评价目标, 建立多准则参数率定模型, 并采用模糊集思想把多准则问题转化为单一目标优选问题。其次针对遗传算法个体适应度计算时间较长, 采用主从式并行策略实现遗传算法的并行化, 提高系统计算效率。最后以我国南方地区西水流域凤滩水库为例, 该方法不仅可获得高质量的模型参数, 保证模型的预报准确性, 同时可解决新安江模型参数率定耗时长等问题, 有效提高模型参数率定效率, 为模型参数校核提供参考。

关键词

新安江模型, 并行遗传算法, 参数率定

Multi-Core Parallel Genetic Algorithm for Parameter Calibration of Xin'anjiang Model

Wei Wang*, Xinfeng Jin, Yang Liu, Hao Yu

Power China Zhongnan Engineering Corporation Limited, Changsha Hunan

Received: Dec. 3rd, 2021; accepted: Feb. 2nd, 2022; published: Feb. 28th, 2022

Abstract

Aiming at the characteristics of many parameters of Xin'anjiang model, heavy workload of calibration calculation and difficulty of traditional calibration parameter accuracy to meet the needs of actual production, a multi-core parallel genetic algorithm for parameter calibration is proposed. Taking the peak discharge, peak time and total flood as the evaluation objectives, a multi criteria parameter calibration model is established, and the multi criteria problem is transformed into a single objective optimization

作者简介: 王伟, 出生于 1984 年 7 月, 硕士研究生, 高级工程师, 研究方向为水文水资源、水利信息化等。

*通讯作者 Email: zhouhh@ms.xjb.ac.cn

文章引用: 王伟, 金新峰, 刘洋, 余豪. 新安江模型参数率定的多核并行遗传算法[J]. 水资源研究, 2022, 11(1): 77-83.

DOI: 10.12677/jwrr.2022.111008

problem by using the idea of fuzzy set. Secondly, in view of the long computing time of individual fitness of genetic algorithm, the master-slave parallel strategy is adopted to realize the parallelization of genetic algorithm and improve the computing efficiency of the system. Finally, taking Fengtan Reservoir in You-shui Basin in southern China as an example, this method can not only obtain high-quality model parameters and ensure the prediction accuracy of the model, but also solve the problems of long time-consuming in Xin'anjiang model parameter calibration, effectively improve the efficiency of model parameter calibration and provide reference for model parameter verification.

Keywords

Xin'anjiang Model, Parallel Genetic Algorithm, Parameter Calibration

Copyright © 2022 by author(s) and Wuhan University.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

概念性降雨径流模型是常用的洪水预报模型，新安江模型作为一种概念性降雨径流模型广泛应用于湿润地区与半湿润地区[1]，然而新安江模型含有 16 个参数，并且这些参数是相互依赖的，这使得参数率定十分困难，其推广与应用受到严峻挑战[2] [3]。

目前常用的模型参数率定方法包括人工试错法与智能算法，其中人工试错法是在计算机欠发达时常用的方法，然而由于其缺少判断参数优劣判断标准，缺少客观性，并且对个人经验要求较高，当个人经验不足时很难得到理想的结果[4]。随着计算机技术的发展，人工智能算法广泛应用于新安江模型参数率定，目前常用的方法主要有 Rosenbrock [4]、SCE-UA [5] [6]、遗传算法[7]。然而 Rosenbrock 与 SCE-UA 法在搜索过程中很容易陷入局部最优。遗传算法由于其简单通用、鲁棒性强、全局寻优等特点，常用于非解析式的目标函数与约束等问题，成为水文工作者参数率定的重要手段[8] [9]。近些年多核计算技术的普及与发展，多核并行技术具有计算成本低廉、系统运行环境稳定、并行环境容易实现等优点，已在水资源领域广泛应用[6] [10]。本文结合遗传算法的天然可并行性，提出一种多核并行遗传算法，用于求解多准则新安江模型参数率定问题，以我国南方地区西水流域凤滩水库为例，表明本文所提方法能显著提高参数率定速度。

2. 新安江及参数率定标准

模型参数的率定是一个迭代过程，不断调整参数以找到一组参数以实现模拟洪水流量与观测值之间的最佳拟合。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。为此引入了模糊集的概念。

当评价目标为洪峰流量和洪水总量时，各方案对应评价目标的模糊合格率为：

$$x_{ij} = \begin{cases} 1.0 & \zeta_{ij} \leq 0.2 \\ 1.25\lambda(1.0 - \zeta_{ij}) & 0.2 < \zeta_{ij} \leq 1 \\ 0 & \zeta_{ij} > 1 \end{cases} \quad (1)$$

式中：当 $j=1$ 时， x_{i1} 为第 i 个方案对应洪峰流量的模糊合格率， ζ_{i1} 为第 i 个方案的洪峰流量实际值与预报值相对误差的绝对值；当 $j=2$ 时， x_{i2} 为第 i 个方案对应洪水总量的模糊合格率， ζ_{i2} 为第 i 个方案的洪水总量实际值与预报值相对误差的绝对值； λ 为修正系数。

当评价目标为峰现时间时，各方案对应评价目标的模糊合格率为：

$$x_{ij} = \begin{cases} 1.0 & \tau_{ij} \leq 3 \\ \lambda(10 - \tau_{ij})/9 & 3 < \tau_{ij} \leq 10 \\ 0 & \tau_{ij} > 10 \end{cases} \quad (2)$$

式中：当 $j=3$ 时， x_{i3} 为第 i 个方案对应峰现时间的模糊合格率， τ_{i3} 为第 i 个方案的峰现时间，所述峰现时间是指预报洪峰时间与实际洪峰时间之差。

每个方案对应不同评价目标的模糊合格率 x_{ij} 进行归一化处理，由归一化处理后的模糊合格率 r_{ij} 构成归一化矩阵 R_{ml} ，其中 $R_{ml} = (r_{ij})_{ml}$ ， $r_{ij} = x_{ij} / \sqrt{\sum_{j=1}^l x_{ij}^2}$ ， $i=1, 2, \dots, m$ ， $j=1, 2, \dots, l$ 。

根据权重决策矩阵 V_{ml} 计算正理想解 A^+ 和负理想解 A^- ，具体计算为：

$$A^+ = [a_1^+, a_2^+, \dots, a_l^+] = \max(v_{ij} | i=1, 2, \dots, m) | j=1, 2, \dots, l; \quad (3)$$

$$A^- = [a_1^-, a_2^-, \dots, a_l^-] = \min(v_{ij} | i=1, 2, \dots, m) | j=1, 2, \dots, l; \quad (4)$$

根据正理想解 A^+ 和负理想解 A^- ，计算距正理想解的距离 d_i^+ 和距负理想解的距离 d_i^- ，具体计算为：

$$d_i^+ = \sqrt{\sum_{j=1}^l (a_j^+ - v_{ij})^2}, d_i^- = \sqrt{\sum_{j=1}^l (a_j^- - v_{ij})^2} \quad (5)$$

计算评价方案与对应负理想解的相对贴近度，具体计算为：

$$c_i = d_i^- / (d_i^- + d_i^+) \quad (6)$$

式中： c_i 为第 i 个方案与该方案对应的负理想解的相对贴近度；其中 c_i 值越大，对应的方案越优。

3. 并行计算

Fork/Join 框架是 Java7 提供的一个充分挖掘多核 CPU 计算能力的并行计算框架，通过问题分解、并行计算与结果合并三个主要步骤实现并行求解。其基本思想是采用“分治策略”将计算规模较大的问题，分解为若干个规模较小、相对独立的问题进行求解，然后求得各个子问题的解，合并子问题的解得到原问题的解，其计算执行过程如图 1 所示。该框架通过阈值控制子问题的规模大小，任务划分时，首先判断任务规模是否大于设定的阈值，如果任务规模小于设定的阈值，则直接求解即可；如果任务规模大于设定的阈值，则对任务规模进行划分，直到子问题的任务规模小于设定的阈值。在并行过程中如果阈值过大，则子问题任务数目较小，无法充分发挥多核优势；如果阈值太小，则子问题数目较多，任务管理开销加大。为充分利用资源、提高计算效率，按下式确定阈值：

$$\beta = \langle q/p \rangle \quad (7)$$

式中： $\langle \rangle$ 表示取最小正整数， β 为阈值， p 为多核 CPU 的核数， q 为原问题的计算规模。

4. 具体流程

1) 染色体的确定：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。本文采用十进制整数进行编码。

2) 初始化种群：假设种群规模为 A_{mm} ，记 $A_{mm} = (a_{ij})_{mm}$ ，其中 m 为种群个数， n 为新安江模型参数数量， a_{ij} 为第 i 个染色体对应的第 j 个参数的取值。

3) 采用 Fork/Join 并行框架创建并行遗传算法，并将 m 个染色体分配到不同的线程池中。

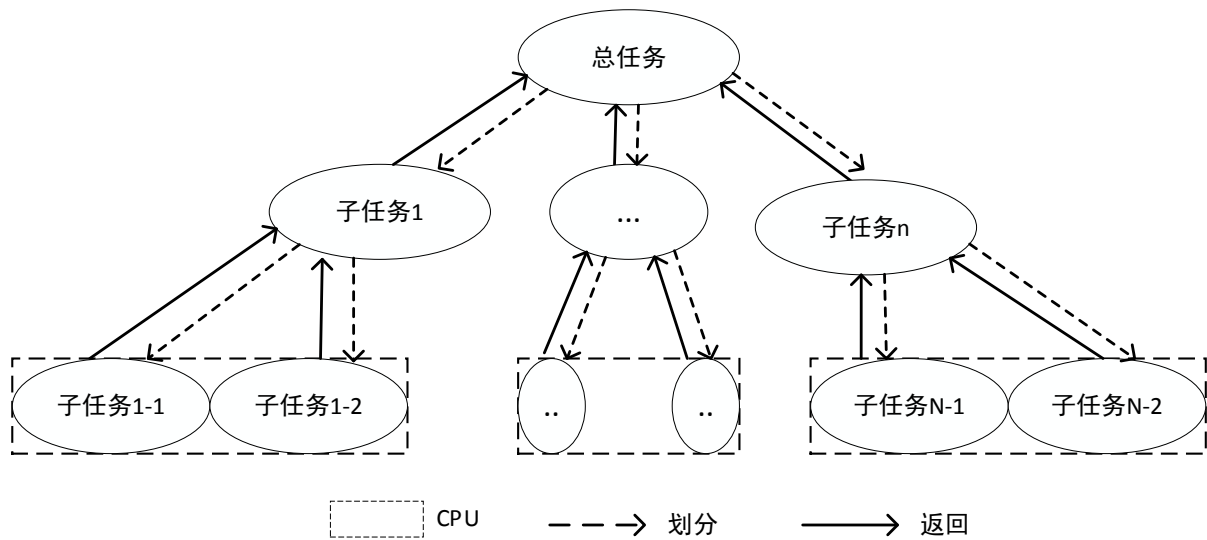


Figure 1. Flow chart of parallel computing
图 1. 并行计算流程图

- 4) 计算各线程池中评价目标的模糊合格率。
- 5) 根据所述步骤 4 计算的模糊合格率，创建由 m 个方案和 l 个评价目标构成的多目标评价矩阵 X_{ml} ，记 $X_{ml} = (x_{ij})_{ml}$ ，其中 x_{ij} 为第 i 个方案对应第 j 个评价目标的模糊合格率， X_i 为第 i 个方案对应的目标向量，记为 $X_i = [x_{i1}, x_{i2}, \dots, x_{ij}, \dots, x_{il}]$ 。
- 6) 根据所述多目标评价矩阵 X_{ml} 计算并确定各线程池中的最优方案，将各线程池对应的最优方案返回给主线程。
- 7) 所述主线程对各线程池对应的最优方案进行比对，确定当前最优计算结果。
- 8) 根据所述步骤 7 的当前最优计算结果确定该当前最优计算结果对应的方案，从而确定模型的当前参数组合。
- 9) 对于所述步骤 8 的当前参数组合，计算该当前参数组合中各参数在不同水平下相对贴进度 c_i 的平均值 \bar{T}_i ，选取平均值 \bar{T}_i 中的最大值 \bar{T}_{\max} 所对应的参数的水平取值，得到最优参数组合。
- 10) 判断所述步骤 9 的最优参数组合中各参数的水平取值是否满足精度需求，即 $|a_{i+1j} - a_{ij}| \leq \varepsilon$ (a_{ij} 为染色体 A_{mm} 中第 $i + 1$ 个方案对应的第 j 个参数的取值， ε 为精度)，如果满足，则停止计算，所述步骤 9 得到的最优参数组合作为最终计算结果；否则重复步骤 2~9，直到满足精度需求。

5. 具体算例

以我国南方地区西水流域凤滩水库为例，验证所提方法的有效性。该流域为亚热带季风气候，流域面积为 1.7 万 km^2 ，平均温度 18 摄氏度。3~7 月的降雨占全年的 65% 以上，雨量充沛，降雨集中，中长期径流预报难度较大。流域内的预报降雨与蒸发数据根据流域内的 27 个气象站，采用泰森多边形计算得到。本次计算选取 2000 年 1 月~2018 年 12 月共计 18 年的 103 场降雨数据进行参数率定，以 2019 年 1 月到 2020 年 12 月的 19 场洪水进行校核。

采用本文所提方法对新安江三水源模型进行参数校核，算例中遗传算法的初始种群规模为 200，交叉概率为 0.65，变异概率为 0.05，进化代数为 500 代，计算终止条件 $\varepsilon = 0.01$ ，所有仿真程序均用 java 语言编写，运行环境为 4 核 CUP、主频 3.2 GHZ、内存为 16 GB、硬盘 500 GB、Windows 系统的 ThinkPad 计算机。为了保证计算结果的合理性，计算 5 次，计算结果如表 1 所示：

Table 1. The calculation results of the proposed method in this paper
表 1. 本文所提方法计算结果

模型	次数	计算时间(s)	率定合格率(%)			检验合格率(%)		
			洪量	洪峰	峰现时间	洪量	洪峰	峰现时间
遗传算法	1	5013	92.23	79.18	86.41	84.21	73.68	89.47
	2	5162	91.26	79.18	85.54	89.47	78.64	89.47
	3	4999	91.26	80.58	83.49	89.47	73.68	89.47
	4	5065	93.20	84.46	86.41	89.47	78.64	94.74
	5	4991	93.20	79.18	85.54	84.21	73.68	94.74

取最优计算结果, 模型各个参数如表 2 所示:

Table 2. Parameter calibration results of Xin'anjiang Model
表 2. 新安江模型参数率定结果表

参数	KC	B	WM	UM	LM	C	Im	Sm	Ex	Kg	Ki	Ci	Cg	Ke	Xe
值	0.79	0.53	145	28	81.3	0.15	0.02	13.4	1.6	0.19	0.56	0.2	0.58	2.5	0.15

根据表 1 可以看出, 遗传算法 5 次计算结果不一致, 原因是遗传算法由于交叉与变异概率的随机性。取最优计算结果, 其洪峰流量的合格率为 84.46%, 洪量的合格率为 93.20%, 峰现时间的合格率为 86.41%。对参与校核的 19 场洪水进行分析, 其洪峰流量的合格率为 78.64%, 洪量的合格率为 89.47%, 峰现时间的合格率为 94.74%, 本文所提方法洪峰、洪量、峰现时间合格率均大于 70% 满足乙级标准。

为了进一步验证并行算法的并行计算性能, 将并行遗传算法与串行遗传算法进行对比分析。测试基于并行核数与终止条件 ε 进行分析, 并采用衡量并行算法的两个重要指标来评价加速比 S_p 与效率 E_p 来测试并行遗传算法的性能, 其中 S_p 与 E_p 分别采用下式计算:

$$S_p = \frac{t_s}{t_p}, E_p = \frac{S_p}{C} \quad (8)$$

式中: t_s 为串行遗传算法计算一次所需要的时间, t_p 为并行遗传算法计算所需时间, C 为并行遗传算法中的并行核数。

由表 3 可知, 串行遗传算法非常耗时, 且随着离散精度的不断减少所需时间不断增加, 当离散精度为 $\varepsilon = 0.001$ 时, 串行遗传算法所需时间为 156 分钟(9356 s), 无法满足时效性的要求, 相同离散精度下, 采用并行遗传算法可大大缩短计算时间, 在所实验的范围内, 参数率定所需要的时间随着并行核数的增加而减少。当离散精度为 $\varepsilon = 0.001$, 2 核环境下所需的时间为 4873 s, 大约是串行时间的一半; 在 16 核的环境下, 仅需 824 s 即可完成计算。因此并行遗传算法可有效提高参数率定速率, 是改善大规模复杂问题的一个切实可行的方法。在两核环境下, 离散精度从 $\varepsilon = 0.05$ 依次到 $\varepsilon = 0.001$, 加速比从 1.82 到 1.92, 效率由 0.91 增加到 0.96; 并且随着并行核数的增多, 加速比与效率随着离散精度的提高而更加明显。

Table 3. Calculation results under different parallel environments
表 3. 不同并行环境下的计算结果

方法	核数	结果	终止条件 ε					
			0.001	0.002	0.005	0.01	0.02	0.05
串行	--	t_s	9356	8204	6651	4991	4365	3740

Continued

		t_s	4873	4295	3519	2683	2359	2055
2		S_p	1.92	1.91	1.89	1.86	1.85	1.82
		E_p	0.96	0.96	0.95	0.93	0.93	0.91
		t_s	2488	2182	1837	1414	1262	1116
4		S_p	3.76	3.76	3.62	3.53	3.46	3.35
		E_p	0.94	0.93	0.91	0.88	0.87	0.84
		t_s	1789	1621	1360	1048	937	831
6		S_p	5.23	5.06	4.89	4.76	4.66	4.50
		E_p	0.87	0.84	0.82	0.79	0.78	0.75
		t_s	1455	1300	1080	833	739	640
8		S_p	6.43	6.31	6.16	5.99	5.91	5.84
		E_p	0.80	0.79	0.77	0.75	0.74	0.73
		t_s	1026	906	778	621	565	519
12		S_p	9.12	9.05	8.55	8.04	7.72	7.2
		E_p	0.76	0.75	0.71	0.67	0.64	0.60
		t_s	824	747	636	503	474	441
16		S_p	11.36	10.98	10.45	9.93	9.21	8.48
		E_p	0.71	0.69	0.65	0.62	0.58	0.53

6. 结论

本文探讨了基于并行遗传算法的新安江模型参数率定方法,充分利用了遗传算法天然可并行的优势,采用并行计算,多核计算机的计算特性,极大地提高了计算速度,参数率定准确率达到乙级标准。但本文未解决遗传算法容易陷入局部最优的问题,且算法的稳定性没有确切的数据说明,这在以后的参数率定中需要进一步完善。

参考文献

- [1] 何斌,刘宝库,卓闻涛,等.考虑玉石水库影响的碧流河水库洪水预报方式[J/OL].水力发电:1-5[2021-12-16].
<http://kns.cnki.net/kcms/detail/11.1845.TV.20211209.0156.004.html>
HE Bin, LIU Baoku, ZHUO Wentao, *et al.* Flood forecasting method of Biliuhe reservoir considering the influence of Yushi reservoir. *Hydropower*: 1-5[2021-12-16]. <http://kns.cnki.net/kcms/detail/11.1845.TV.20211209.0156.004.html> (in Chinese)
- [2] 邱文怡,朱永军,湛忠宇.新安江产流和TANK汇流耦合模型在南方湿润地区的应用[J].人民珠江,2021,42(2):19-24.
QIU Wenyi, ZHU Yongjun and ZHAN Zhongyu. Application of Xin'anjiang runoff generation and tank confluence coupling model in humid areas of South China. *People's Pearl River*, 2021, 42(2): 19-24. (in Chinese)
- [3] 孙娜,周建中,张海荣,等.新安江模型与水箱模型在柘溪流域适用性研究[J].水文,2018,38(3):37-42.
SUN Na, ZHOU Jianzhong, ZHANG Hairong, *et al.* Study on applicability of Xin'anjiang model and water tank model in Zhexi Basin. *Hydrology*, 2018, 38 (3): 37-42. (in Chinese)
- [4] LIAO, S. L., SUN, Q. Y., CHEN, C. T., *et al.* Multicore parallel genetic algorithm with tabu strategy for rainfall-runoff model calibration. *Journal of Hydrologic Engineering*, 2017, 22(8): 1136-1149.
[https://doi.org/10.1061/\(ASCE\)HE.1943-5584.0001542](https://doi.org/10.1061/(ASCE)HE.1943-5584.0001542)
- [5] 张超,姜景山.新安江模型参数自动优化率定的SCE-UA算法应用研究[J].三峡大学学报,2020,42(6):18-23.
ZHANG Chao, JIANG Jingshan. Application of SCE-UA algorithm for automatic optimization and calibration of Xin'anjiang model parameters. *Journal of China Three Gorges University*, 2020, 42(6): 18-23. (in Chinese)

- [6] 阚光远, 洪阳, 梁珂, 等. 基于 GPU 加速的水文模型参数率定[J]. 人民长江, 2019, 50(5): 65-69+75.
KAN Guangyuan, HONG Yang, LIANG Ke, *et al.* Parameter calibration of hydrological model based on GPU acceleration. Yangtze River, 2019, 50(5): 65-69+75. (in Chinese)
- [7] 刘佩瑶, 郝振纯, 王国庆, 等. 新安江模型和改进 BP 神经网络模型在闽江水文预报中的应用[J]. 水资源与水工程学报, 2017, 28(1): 40-44.
LIU Peiyao, HAO Zhenchun, WANG Guoqing, *et al.* Application of Xin'anjiang model and improved BP neural network model in hydrological prediction of Minjiang River. Journal of Water Resources and Water Engineering, 2017, 28(1): 40-44. (in Chinese)
- [8] 陆旻皎. 新安江模型研究的回顾和展望[J]. 水利学报, 2021, 52(4): 432-441.
LU Minjiao. Review and prospect of Xin'anjiang model research. Journal of Hydraulic Engineering, 2021, 52(4): 432-441. (in Chinese)
- [9] 刘志萍, 杨华, 周雨, 等. 基于遗传算法的江西七一水库来水流量新安江预报模型参数优化[J]. 气象与减灾研究, 2020, 43(2): 149-154.
LIU Zhiping, YANG Hua, ZHOU Yu, *et al.* Parameter optimization of Xin'anjiang prediction model for inflow of Qiyi reservoir in Jiangxi Province based on genetic algorithm. Meteorology and Disaster Reduction Research, 2020, 43(2): 149-154. (in Chinese)
- [10] 刘永和, 冯锦明, 徐文鹏. 分布式水文模型的 GPU 并行化及快速模拟技术[J]. 水文, 2015, 35(4): 20-26.
LIU Yonghe, FENG Jinming and XU Wenpeng. GPU parallelization and fast simulation technology of distributed hydrological model. Hydrology, 2015, 35(4): 20-26. (in Chinese)