基于遗传算法的新安江模型日模拟参数优选研究

陈坰烽[1,3],张万昌[2]

(1.南京大学国际地球系统科学研究所,江苏 南京 210093;
2.中科院大气物理所全球变化东亚区域研究中心,北京 100029;
3.南京大学地理与海洋科学学院,江苏 南京 210093)

摘 要:在概念性水文模型的参数率定中,目前还没有一个传统优化方法能够提供保证足够高效和稳定性的算法。为了克服传统优化方法中局部收敛性的缺点,近年来利用遗传算法通过计算机准确稳定地进行概念性水文模型的参数优选的尝试得到越来越多的重视和发展。目前优选水文模型待定参数,大多是从次洪模型的方面去讨论,有关日模拟模型的遗传算法参数优选讨论的较少。本文系统分析了基于遗传算法的新安江模型日模拟参数的自动优选,同时针对遗传算法在模型参数众多的情况下时间效率低下问题,通过利用新安江模型参数分层原理与模型参数敏感性分析对优选结果影响,提出一套简化的日模型参数遗传算法优选方案。经过流域模拟检验,该优选方案可行,运行效率高,可以作为类似模型遗传算法参数率定快速、有效的方案。

关键词:新安江模型;遗传算法;参数优选;日模型

中图分类号:P338+.9 **文献标识码**:A **文章编号**:1000-0852(2006)04-0032-07

1 引言

新安江模型的日尺度模拟(日模型)是洪水模拟(次洪模型)的基础,新安江模型次洪模拟的土壤湿度参数须由日模拟的结果中得到。因此,进行较准确的日模型模拟对洪水模拟有一定影响[1]。

如何根据实测水文资料来确定模型参数,是新安江模型日模型应用的主要问题。模型参数率定工作一般是用手工或者是计算机自动完成。要做好手工调试,工作人员必须有丰富经验,而且很耗时,参数的优选不仅困难而且往往带有主观性。因此,自动参数率定受到广泛关注[2]。在概念性水文模型的参数率定中,传统的优化方法由于局部收敛性问题,往往无法提供全局最优解,因此不具有足够高效和稳定性。这些传统的自动率定方法也无法真正地在实际应用中推广[3]。近年来,非经典数学方法,特别是遗传算法(Genetic Algorithms)由于其很好的全局搜索特性、思想简单、易于实现等优点为各领域所广泛接受,并在模式识别、人工智能、管理决策等领域得到应用。实验证明它优于很多的传统优化算法[4-5]。同样的这些先进的数学方法也被引入到概念性水文模型参数优选领域。谭炳卿[6]以新安江模型为例,分别使用罗森布郎克(Rosenbrock)、改进的单纯形(Simplex)和遗传算法(Genetic)优选模型参数,并对优化方法的收敛速度,初值对结果的影响进行比较分析,得出了使用遗传算法进行优选效果最好的结论,并指出利用遗传算法进行参数率定缺点是收敛速度慢。杨晓华等[7]针对传统遗传算法收敛速度慢的问题,在实编码遗传算法中加入了单纯形搜索算子和加速搜索算子,提出了混合加速遗传算法。并在新安江模型参数率定中应用。陆桂华等[8]将单纯形法、模式搜索法与二进制加速遗传算法、实编码加速遗传算法、实编码单纯形混合加速遗传算法进行了比较,发现在新安江模型参数率定中,遗传算法全局优化性能较好,而其中混合加速遗传在精度和稳定性方面有很好的优选结果。武新宇等[9]使用并行算法技术和遗传算法

收稿日期:2005-10-08

基金项目:国家重点基础研究发展计划973项目(2001CB309404和2006CB400502)和中科院"百人计划"项目(8-047401)资助。

作者简介:陈坰烽(1982-),男,福建龙岩人,硕士,主要从事水文学理论及遥感水文学应用方面工作。

结合,利用 JAVA 编程在微机集群环境下,快速稳定的获得新安江模型参数,使算法优选时间缩短。

上述文章主要是从算法本身或者算法实现的角度研究模型参数自动优选中收敛速度慢的问题,取得了一定的效果。主要研究对象是次洪模型。由于参数优选是将模型所有参数一并代入算法优选计算,而较少考虑参数的物理意义和参数分层特性,致使有些方法的实现需要有大量设备及复杂软件的投入。鉴于此,本文从模型构造自身的规律性入手,提出了一套运用于日模型的较简洁的遗传算法参数自动优选方案。该方案以模型刻画的水文过程为出发点,把参数分为几个相互独立的层次,分目标函数加入遗传算法进行快速参数率定,使优选参数方法客观实际,具有应用价值;并利用遗传算法优选模型主要敏感参数,忽略不敏感参数,从而达到提高模型优选效率的目的。

2　优选方案的确定

2.1　模型参数的分层优选

模型参数众多,相互关系复杂,参数率定计算开销大,算法运行效率是个瓶颈。解决这一瓶颈的希望之一是通过对某一个或某几个参数进行独立优选,获取不同的目标函数,进行分层调试以达到全局最优的目的。

2.1.1　理论基础

分系统分层次进行参数优选是以大系统的分解理论为依据的。其基本思想是把系统分解为几个独立部分,如果分解的部分具有独立性(自身优选过程不影响其他部分),则整体优选和分解优选的结果完全相同。如函数:

$$F(x_1,x_2)=(x_1)^2+2(x_2-4)^4+3x_1+x_2+231$$

可以分解为:

$$F_1(x_1)=(x_1)^2+3x_1+231$$

$$F_2(x_2)=2(x_2-4)^4+x_2$$

显然,使用 $F(x_1,x_2)$ 直接优选 x_1,x_2 和 $F_1(x_1)$、$F_2(x_2)$ 分别优选是没有差别的。因此,对于可以分解为独立子系统的目标函数的优化,就可以采用分解优选的方法。这是新安江模型分层优选的理论依据。

2.1.2　水文系统的分层理论

流域水文现象按其性质可分为几个层次,对于湿润地区,一般可以分为蒸散发、产流、分水源、汇流四个层次。第一个层次蒸散发主要由气候因素决定,最为稳定,能决定长时段内的产流总量,但对于产汇流过程作用很小。第二层、第三层决定产流量在水源上与时间上的分配,与降雨过程及流域条件的关系很密切,变化比蒸散发敏感得多。第四层把流域面上各种水源的产流过程汇集成为流域出口的出流过程,流量变化非常敏感[10]。

上述层次的顺序也就是模型的计算步骤,上一层次的计算结果可作为下一层次的计算输入,交叉不多。每一层次中都有一些公式与参数,以实现模拟计算。这种计算方法的前提是下一层次的参数值对上一层次的计算结果影响很小。例如:汇流参数很少影响分水源计算,分水源参数不影响产流总量的计算等。这种层次之间的相对独立性,在水文现象上是有根据的。

2.1.3　新安江模型参数的分层和优选目标函数

基于以上水文特性分析,结合分析新安江模型的模型架构及其各部分表征的水文物理意义,模型参数可以分为蒸散发、产流、分水源、汇流等四大类(如图1),各类之间基本上是独立的。

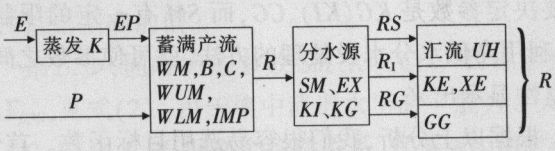


图1　新安江模型分层结构图

(1) 蒸散发和产流。蒸散发和产流层共有7个参数,分别是 K、WM、WUM、WLM、IMP、B、C。由上面的分层理论容易得知,蒸散发参数决定产流总量,其效果是使多年的降雨、蒸发与径流之间水量得到平衡。当然其他参数 WM、WUM 等也影响产流总量,只不过影响较小,模型简化结构图如图2。

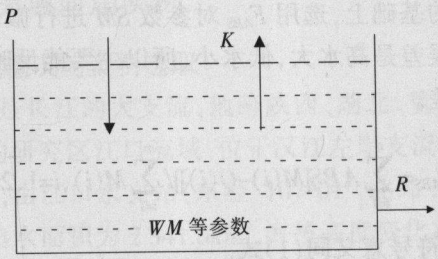


图2　模型结构简化图

图2中,将模型进行抽象得到模型的输入为 P(降水量),输出为 R(径流量)与 K 值控制的蒸发量。可以列出水量平衡方程 $P-F(WM\text{等})-R-E(K)=0$;在 P 已知的情况

下,只要以水量平衡方程为目标,优选其他的待定参数(K,WM等)即可。同时可以发现由于WM是表示流域平均蓄水容量,其值不会随时间增加而增大,多年的降水量平衡就能够使WM的影响降到最低。同时其他层次参数的取值并不影响总量平衡。所以最后优选K及其他产流层参数值(如WM、WUM等)使用的目标函数需使多年产流总量的误差最小,即$\triangle R$最小,公式如下:

$$\triangle R=\sum_{i=1}^{n}ABS[M(i)-Q(i)],i=1,2,\cdots,n \quad (1)$$

式中,$M(i)$为实测流量;$Q(i)$为计算流量。

(2)分水源和汇流。在分水源计算结束后,求得的是河网总入流。汇流计算只处理河网汇流问题,与水源划分无关。因此第三层参数和第四层参数之间性质上是完全独立的。但是在优选参数时,都只能根据流量过程线,于是考虑如何在目标函数中尽可能地将参数独立出来[11]。由于地面径流控制高水部分,主要的决定参数是SM、EX、$KG(KI)$、UH。地下径流控制低水部分,主要决定参数是$KG(KI)$、CG,而SM有一定的限制作用。利用高低水分水段处理的方法,就可使参数之间的独立性显示出来。

根据以上分析,我们很容易选用目标函数。首先,选用目标函数F_{LOG}对参数SM、KG、CG进行优选。由于EX与SM值有较大相关性,而且其值稳定在1~1.5,不参加优选[11]。F_{LOG}的取值是低水大,高水小,主要能反映低水部分的误差。

$$F_{LOG}=\sum_{i=1}^{n}ABS\{LOG[M(i)/Q(i)]\}/\sum_{i=1}^{n}LOG[M(i)],i=1,2,\cdots,n \quad (2)$$

式中,符号意义同(1)式。

其次,鉴于SM对高水部分的控制,于是在上一步优选的基础上,选用F_{ABS}对参数SM进行微调。F_{ABS}绝对值误差是高水大,低水小,所以主要能反映高水部分的误差。

$$F_{ABS}=\sum_{i=1}^{n}ABS[M(i)-Q(i)]\}/\sum_{i=1}^{n}M(i),i=1,2,\cdots,n \quad (3)$$

式中,符号意义同(1)式。

以上就构成了新安江模型参数分层优选的算法基础,各类参数在模型刻画的水文过程中有其各自的作用与目标。

2.2 蒸散发和产流层优选敏感参数

在众多水文模型参数中,有的参数很敏感,数量稍有变化对输出的影响就很大。而有的参数则反应迟钝,对输出影响不大。对敏感参数仔细优选,非敏感参数则可以据一般经验固定下来,是传统人工优选模型参数常用的方法,以往众多实验证明这种方法可以满足应用的要求。

由于蒸散发和产流层需要优选的参数最多,共有7个,是遗传算法参数率定最耗时的部分,而模型敏感性分析表明参数中只有1个较为敏感。因此比较了使用遗传算法同时优选7个参数和单独优选最为敏感的参数K对模拟精度的影响和各自的运行时间,定量讨论使用遗传算法单独优选敏感参数的可行性。

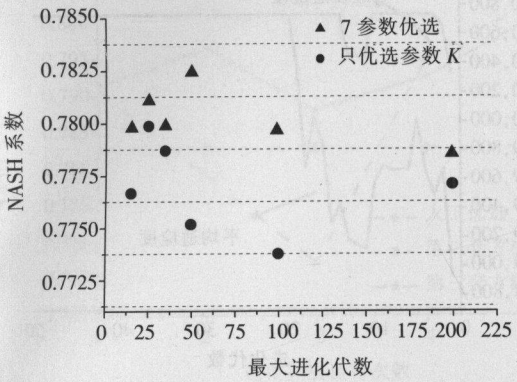
2.2.1 可行性分析

首先,选择汉江流域江口水文站控制的h流域内1981~1983年逐日水文观测资料为试验时段,利用新安江模型日模型进行模拟,将方案第一步提出的只优选敏感参数K的方法和同时优选7个参数方法(其他层次的模型参数取相同值)得到的结果进行对比,两者得到的NASH系数十分相近,如图3(左),第2、3和第6个点之间NASH系数差别十分微小,约为0.0013;而最大的差别在于第4个点,为0.0075,平均差别小于0.005。如图3(右),在多年径流量绝对误差总和方面,同时优选7个参数得到的结果要明显好于单独优选方法,这是由于遗传算法优选使用的目标函数即多年径流量绝对误差总和。可以发现,两种方法得到的结果平均相差约1 250m³/s,但是由于率定期时长是3年,而分到每天误差不到1m³/s,这样的误差也是在允许范围之内。同时通过循环6次的遗传算法计算发现,得到的结果很稳定,变化幅度不大,如图3。单独优选的NASH系数平均值为0.776,而绝对误差总量平均值为19 250m³/s。因此认为使用遗传算法第一步中只单独优选K参数是可行的,对模拟的效果影响不大。

2.2.2 时间效率分析

由于多参数同时优选导致遗传算法初始种群参数呈指数上升趋势,因此7参数优选的运行时间比只优选参数K耗费得多。两者的运行时间比较如表1。

因此本文中,在蒸发层和产流层中使用遗传算法优选敏感参数K,而其他6个不敏感参数则依照流域特征和经验进行定值。这样既结合模型自身的水文特性,同时充分发挥了遗传算法优越性,提高了优选运行效率。以最小的代价换取遗传算法参数优选效率的提高,达到事半功倍效果。



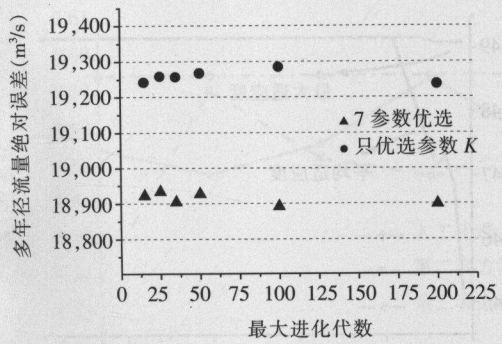


图 3　只优选每感参数 K 与 7 参数同时优选模拟结果比较
(左图以 NASH 系数为标准,右图以多年径流量绝对误差为标准)

表 1 只优选敏感参数 K 与 7 参数同时优选运行时间比较

遗传算法最大进化代数	15	25	35	50	100	200
只优选参数 K ('分"秒)	3'50"	6'31"	9'15"	13'1"	26'23"	53'0"
7 参数同时优选	102'49"	175'30"	249'45"	351'27"	715'42"	1 431'55"

2.3　微调 *SM* 的作用

加入目标函数 F_{ABS}(式 3)后模拟结果全部好于不加入第三步微调的模拟结果，因此有必要加入微调参数 *SM* 这一步骤。

微调 *SM* 对模拟结果的影响见图 4。

2.4　遗传算法优选方案

通过以上分析,最终确定日模型参数的遗传算法自动优选方案。根据水文模型参数分层特性,一共分三步分别优选模型不同层次的参数。具体方案如下:

第一步,优选敏感参数 *K*,目标函数使用 $\triangle R$,见式(1)。其他参数 *B*、*C*、*WUM*、*WLM*、*WM*、*IMP*、*EX* 根据

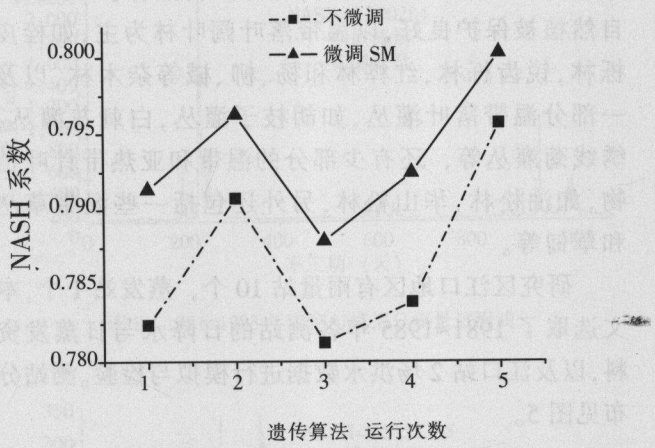


图 4　微调 SM 与不微调的模拟结果的比较

流域特征大致估值[11] [12],他们对日模型的模拟效率影响很小，前面已经对其做了分析。其他待优选参数如 *SM* 等可以取在取值范围[11]内的任意值。它们的取值对参数 *K* 的优选结果产生很小的影响,在后面的步骤中将优选出正确的值。

第二步,同时优选参数 *SM*、*KG*、*CG*。目标函数使用 F_{LOG},见式(2)。由于壤中流退水历时一般为 3 天,因此 *KG*+*KI*=0.7。根据这个隐性关系,优选 *KG* 同时得到 *KI*。*K* 值取第一步优选得到的参数值。其他参数蒸散发层和产流层的参数如 *B* 等仍然使用第一步中固定的值。

第三步,微调参数 *SM*。目标函数 F_{ABS},见式(3)。

需要注意的是,汇流参数 *Ke*,*Xe* 根据流域的水力学特性求得[13],*UH* 由单独的一次或者多次降雨产生的日尺度的洪水过程流量数据求得[13]。经过以上三步,就可确定新安江模型日模型中的所有参数。

3　江口流域应用实例

3.1　流域概况与资料选取

汉江是长江的大支流,地跨陕西、湖北、重庆等省市。本文的研究区江口流域,位于汉江左岸支流褒河流域的上游,由江口水文站(东经 107°01′,北纬 33°42′)控制,其集水面积为 2 341.6km²,海拔高度变化在 950~3 400m,年降水量变化在 600~900mm 之间,主要集中在夏季的 6~9 月。

研究区内土壤类型以黄棕壤和棕壤为主，主要有石灰岩黄棕壤,坡积洪积黄棕壤,花岗片麻岩黄褐土,坡积洪积黄褐土,以及石灰岩黄棕壤等。

该流域植被覆盖度高，与许多自然保护区毗邻，自然植被保护良好，以温带落叶阔叶林为主，如栓皮栎林，锐齿栎林，红桦林和杨、柳、槭等杂木林，以及一部分温带落叶灌丛，如胡枝子灌丛，白刺花灌丛，绣线菊灌丛等，还有少部分的温带和亚热带针叶植物，如油松林，华山松林，另外还包括一些温带草丛和草甸等。

研究区江口地区有雨量站10个，蒸发站1个，本文选取了1981~1985年各测站的日降水与日蒸发资料，以及江口站2场洪水数据进行模拟与检验。测站分布见图5。

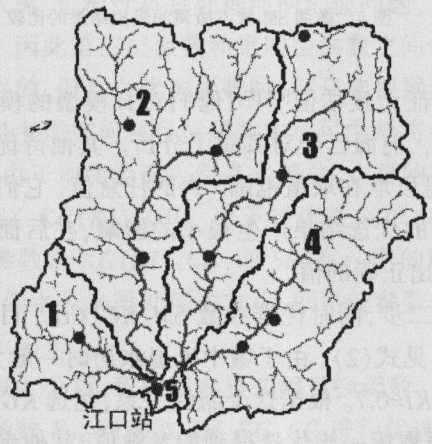


图5 江口流域单元流域的划分及测站分布

3.2 采用的遗传算法介绍

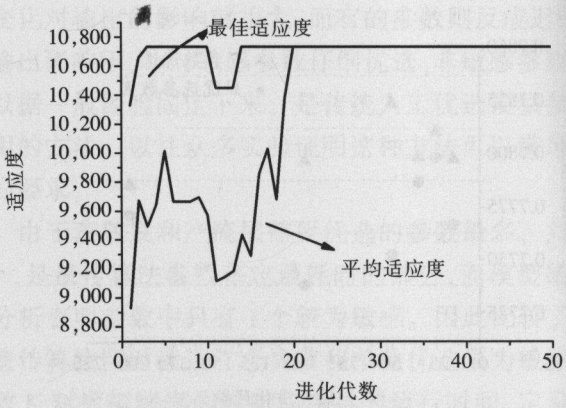
由于只采用单独目标函数，所以本文采用二进制编码的基本遗传算法[14]对参数进行优选。根据De Jong的研究结论[15]，使用保留最佳个体模型或者期望值模型的遗传算法会比基本遗传算法的性能有明显的改进。因此本文在基本遗传算法的选择算子中加入保留最佳个体模型(Elitist Model)。

3.3 遗传算法进化过程及最终参数优选结果

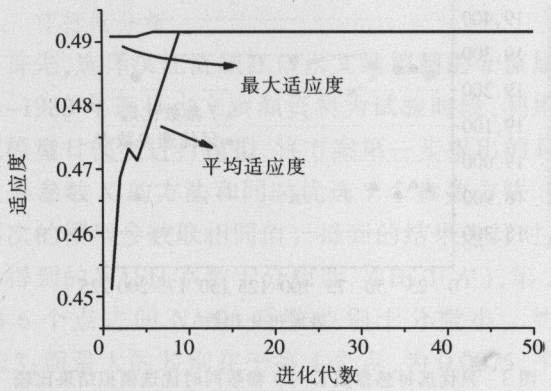
遗传算法进化过程及最终参数优选结果见图6、表2。

表2 日模型参数率定成果表

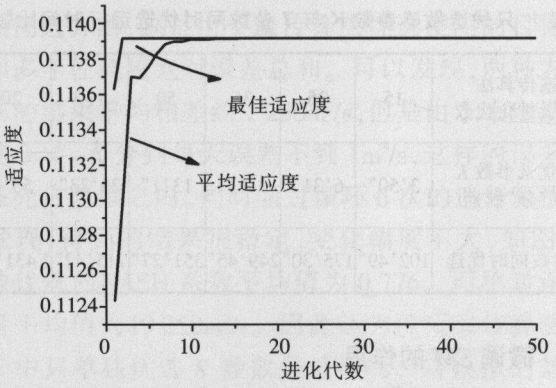
K	WM	WUM	WLM	IMP	B	C	KG
0.887488	120	20	80	0.01	0.3	0.15	0.4042
KI	KKG(CG)	KKSS(CI)	EX	SM	KE	XE	UH
0.2958	0.996135	0.7	1.5	15.82	24	0.2	3.1;0.5;0.2



(a)遗传算法优先参数第一步进化过程



(b)遗传算法优先参数第一步进化过程



(c)遗传算法优先参数第一步进化过程

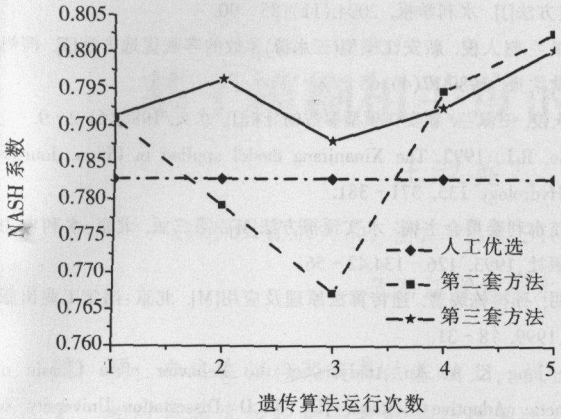
图6 遗传算法参数优选进化过程

((a)图是优选K,(b)图优选SM,KG,CG,(c)图微调SM)

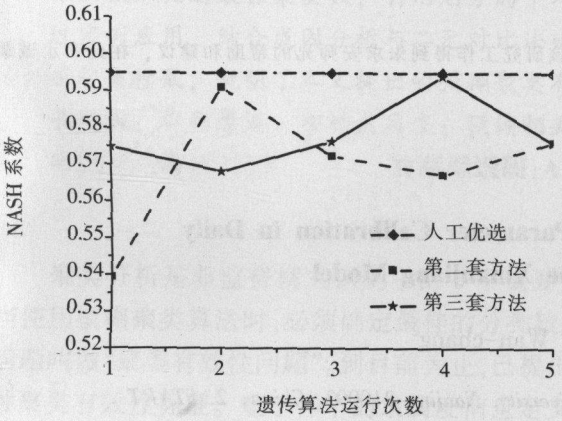
3.4 方案的检验

以率定期确定性系数和校验期确定性系数为评价标准，比较本文提出的第三套方案与传统手工方法、传统计算机自动率定(第二套方案)的优选结果见图7。

分析结果：利用第三套方案优选得到的NASH系数总体高于第二套方案和人工优选方法，而且优选的结果比较稳定，说明优选方法是可靠的。而在检验期



(a)率定期三套方案模拟结果比较(NASH 参数)



(b)校验期三套方案模拟结果比较(NASH 参数)

图 7　三套日模型优选方案拟结果比较

((a)图为率定期 1981~1983,(b)图为校验期 1984~1985 年)

中,由于流量平均值与率定期差别较大的影响,校验期确定性系数与率定期相比有较大的下降幅度。但是这并不影响做三个方案的横向对比。第三套方案优选得到的校验期的确定性系数仍然稳定,平均值要好于第二套优选方案,而且第四次优选的结果接近最优值。

3.5　运行时间

遗传算法参数优选运行时间见表 3。

表 3　遗传算法参数优选运行时间

	优选 *K* 参数	优选 *SM*、*KG*、*CG*	微调 *SM*	总和
种群大小 PopSize	10	40	10	
进化代数 MaxGeneration	25	25	50	
运行时间	6'28"	26'10"	13'23"	46'1"

3.6　流量过程线

1981~1983 年、1984~1985 年实测与模拟日流量过程线见图 8、图 9。

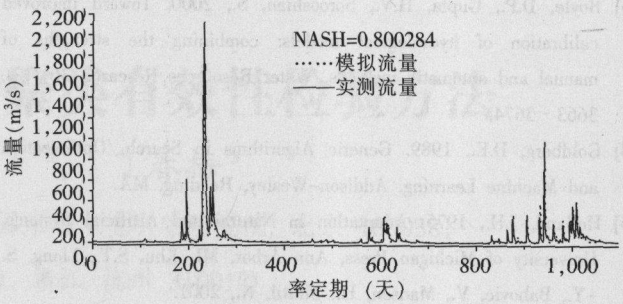


图 8　1981~1983 年实测与模拟日流量过程线

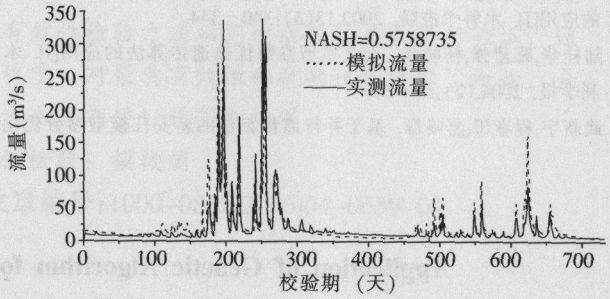


图 9　1984~1985 年实测与模拟日流量过程线

4　结论

(1)遗传算法是一类具有较强鲁棒性的优化算法,特别是对一些大型、复杂非线性系统,表现出了比其他传统优化方法更加独特和优越的特性。全局搜索特性和隐含并行性是它的两大特性,因此适合在概念性水文模型参数的自动优选中应用。

(2)通过计算发现,遗传算法的最大进化代数对模型参数优选影响不大,在较早的时候到达收敛,因此具体计算时最大进化代数的值可以不必设置很大。

(3)本文提出与模型自身规律结合的遗传算法的模型参数自动识别方法,在保证模拟精度的同时,提高了参数优选的运行效率。利用参数敏感性减少优选参数个数,同时以 NASH 系数为标准,定量分析了只优选敏感参数对模拟精度的影响。

(4)本文虽局限于讨论新安江模型的参数自动优选方案,但该方法也可为其他概念性水文模型参数优选提供借鉴。

参考文献:

[1] 赵人俊. 流域水文模拟-新安江模型与陕北模型[M]. 北京:水利电力出版社, 1984. 116-118.

[2] Gupta, V.K., Sorooshian, S., 1985. The automatic calibration of conceptual catchment models using derivative-based optimization algorithms. Water Resources Research 21 (4), 473-485.

[3] Boyle, D.P., Gupta, H.V., Sorooshian, S., 2000. Toward improved calibration of hydrological models: combining the strengths of manual and automatic methods. Water Resources Research 36 (12), 3663 - 3674.

[4] Goldberg, D.E., 1989. Genetic Algorithms in Search, Optimization and Machine Learning, Addison-Wesley, Reading, MA.

[5] Holland, J.H., 1975. Adaptation in Natural and Artificial Systems, University of Michigan Press, Ann Arbor, MI. Khu, S.T., Liong, S.-Y., Babovic, V., Madsen, H., Muttil, N., 2001.

[6] 谭炳卿. 水文模型参数自动优选方法的比较分析 [J]. 水文, 1996, (5):8 - 13.

[7] 杨晓华,陆桂华,郦建强. 混合加速遗传算法在流域模型参数优化中的应用[J]. 水科学进展, 2002,13(3):340 - 344.

[8] 陆桂华,郦建强,杨晓华. 水文模型参数优选遗传算法的应用[J]. 水利学报, 2004,(2):50 - 56.

[9] 武新宇,程春田,赵鸣雁. 基于并行遗传算法的新安江模型参数优化率定方法[J]. 水利学报, 2004,(11):85 - 90.

[10] 王佩兰,赵人俊. 新安江模型(三水源)参数的客观优选方法[J]. 河海大学学报, 1989,17(4):65 - 68.

[11] 赵人俊, 王佩兰. 新安江模型参数的分析[J]. 水文, 1988,(6):2 - 9.

[12] Zhao, R.J., 1992. The Xinanjiang model applied in China. Journal of Hydrology 135, 371 - 381.

[13] 长江水利委员会主编. 水文预报方法[M]. 第二版, 北京:水利电力出版社,1993. 126 - 134,42 - 56.

[14] 周明, 孙树栋编著. 遗传算法原理及应用[M]. 北京:国防工业出版社,1999. 18 - 31.

[15] De Jong K A. An Analysis of the Behavior of a Classic of Genetic Adaptive Systems [D]. Ph.D Dissertation,University of Michigan,No.76-9381,1975.

致谢：该研究工作得到朱求安师兄的帮助和建议，在此表示诚挚的感谢。

Application of Genetic Algorithm for Model Parameter Calibration in Daily Rainfall-Runoff Simulations with the Xinanjiang Model

CHEN Jiong-feng[1,3], ZHANG Wan-chang[2]

(1.*International Institute for Earth System Science, Nanjing University, Nanjing* 210093, *China*; 2. *START Regional Center for Temperate East Asia, Institute of Atmospheric Physics, CAS, Beijing* 100029, *China*; 3. *College of Geography and Ocean Science, Nanjing University, Nanjing* 210093, *China*)

Abstract: By now, there are still no traditional optimization methods being capable of providing enough steady and efficient algorithm to determine the prior-calibrated modeling parameters in conceptual model for rainfall - runoff simulations. The most knotty problem for this issue mainly attributes to the local convergence of the traditional optimization methods, the genetic algorithm capable of yielding accurate and stable optimization of the modeling parameters in conceptual models, therefore, has attract much attention and has rapidly developed with the computing advances. However, most of the efforts on genetic algorithm applications for modeling parameter calibration are concentrated on flooding simulations, few attempt has been made to modeling parameter optimizations with genetic algorithm in daily stream-flow simulations. Bearing this in mind, this study presents an approach to systematically automate-optimize the calibrated parameters in daily stream-flow simulations by using the Xinanjiang hydrological model. For resolving the frequently occurred problem of rather low computation efficiency in genetic algorithm applications, a set of simplified method in parameter optimization based on hierarchy principle of the model parameters and model sensitivity analyses is proposed. An experimental study on model parameter calibration by using the Xinanjiang hydrological model is conducted to a 2 341.6 km2 watershed located on the upper stream of the Hanjiang River Basin. The preliminary study shows that the proposed approach is feasible and of high computation efficiency, and can be transferred to model parameter calibrations for conceptual hydrological models in the similar categories.

Key words: Xinanjiang model; genetic algorithm, model calibration, daily rainfall - runoff model