

基于遗传算法的四维变分资料同化技术的研究

胡娅敏^{1, 2} 丁一汇² 沈桐立¹

1 南京信息工程大学大气科学系, 南京 210044

2 中国气象局国家气候中心, 北京 100081

摘 要 将遗传算法应用于四维变分资料同化问题, 提供一种新的较为有效的数值预报初始场优化方法, 给出了相应的理论依据和详细算法, 并结合变分问题本身的特点, 设计合理的遗传编码、遗传操作和遗传参数。同时, 以二维的浅水方程模式为例建立基于遗传算法的变分同化模型, 并从多方面与伴随模式变分同化方案进行分析比较。数值试验结果表明, 基于遗传算法的变分同化方案取得了比较满意的同化效果。

关键词 遗传算法 变分资料同化 交叉 变异

文章编号 1006-9895(2006)02-0248-09

中图分类号 P456

文献标识码 A

A Research of Four-Dimensional Variational Data Assimilation Based on Genetic Algorithm

HU Ya-Min^{1, 2}, DING Yi-Hui², and SHEN Tong-Li¹

1 *Department of Atmospheric Science, Nanjing University of Information Science and Technology, Nanjing 210044*

2 *National Climate Center, China Meteorological Administration, Beijing 100081*

Abstract Four-dimensional (4D) variational data assimilation (VDA) has been recognized as one of the effective methods to improve numerical weather predication (NWP) initial field. But, the method still has some shortcomings, for example, 1) the optimization method used in VDA should satisfy continual presumption, while the discontinuity of a forecast model, such as the parameterization physical process, will make the presupposition available no more; 2) the cost function always presents multimodal distribution owing to its non-linear dynamic constraint, whereas the present descent algorithms of VDA only aim at the local optimization; 3) furthermore, the adjoint VDA requires higher computing resource in operation. Therefore, it is necessary to seek for a new algorithm with weaker demand to mathematical properties of cost function and higher skill in searching out the global optimum or proximity, together with less computer time to match with the operational demand.

In this context, the genetic algorithm (GA) is applied to 4DVDA. The rational genetic coding, manipulation and parameters, together with related theoretical basis and detailed procedure are designed based upon the properties of the variational technique. The algorithm will be described as follows in detail: 1) real encoding of the parameters; 2) population initialization with model solutions trajectory and the reciprocal of cost function as fitness equivalent; 3) selection of the intermediate population with roulette wheel method; 4) adoption of self-adaptive and mandatory crossover probability to produce better individuals towards greater fitness; 5) application of the quasi-elitist strategy to directly reproduce parent individuals of higher fitness to next generation for better individual reserved; 6) introduction of the steepest descent method, in order to exert its superiority of calculation velocity and precision, to make the mutation manipulation along with the descent direction; 7) selection of genetic control parameters also playing an

收稿日期 2004-10-28, 2005-04-20 收修定稿

资助项目 国家自然科学基金项目 40075023

作者简介 胡娅敏, 女, 1978 年出生, 博士研究生, 目前主要从事区域气候数值模拟和变分同化的研究。E-mail: huyamin@gov.cma.cn

important role in GA, whose rationality depends on the solution convergence and quality after running times without number. Thus, it can be seen that in the process of assuring the population diversity, all of the operations make the population evolve towards greater fitness and accelerate the convergence velocity of GA global optimization.

As an example, the output of a GA-based VDA model constructed under the constraint of 2-dimensional shallow water equations has been compared with that of a scheme with an adjoint VDA model. The results suggest that the convergence precision of GA-based scheme is in common with that of the adjoint equivalent to great degree. When the assimilation window is 6 hours, the convergence velocity of GA-based method is nearly in agreement with that of the adjoint scheme. If the window increases to 12 hours, the former is superior to the latter. Consequently, the GA-based VDA scheme shows a better performance than the adjoint counterpart with the assimilation time window increasing. The most important is that the GA-based VDA model improves greatly mathematical properties of cost function, only utilizing adaptability information independent of higher-order demands, such as the gradient of cost function. In conclusion, the new method suggests a better performance than the adjoint counterpart based on the present experimental case. The GA-based VDA scheme is of superiority to some degree, which enriches the VDA contents and provides more extensive application prospect for future work.

Key words genetic algorithm (GA), variational data assimilation (VDA), crossover, mutation

1 引言

变分同化技术被公认为是提高数值模式初始场质量的有效方法之一,但预报模式(参数化物理过程)的不连续会导致变分法中使用的优化方法需满足连续条件这一前提不再成立^[1]。另一方面,以非线性动力模式系统作为约束条件的目标函数通常是一个多峰函数^[2],而目前的变分同化技术中所使用的下降算法常常是局部优化方法^[3]。同时,当前的伴随同化技术在业务中要求有较高的环境才能实现^[4]。因此,寻找一种对目标函数数学性态要求较弱的,并能够找到或接近全局最优解的算法,而且计算量相对较小、能适合实际业务需要的算法就显得十分必要。

遗传算法(Genetic Algorithm,简称 GA)^[3, 5~11]是以适应度函数作为寻优判据,与模型的具体表达方式无关,在不可微甚至不连续的函数优化问题中,GA 都能求得全局最优解。同时,GA 采用概率化的寻优方法,能自适应地调整搜索方向,数值计算结果表明,它可以越过局部极值,使群体向全局最优值迁移。另外,GA 具有隐含的并行性能^[12],可以利用大型工作站同步分担优化问题各个部分的计算量,从而在其他条件不变的情况下求解速度几十倍上百倍地提高,这一优点正好与现在发展迅速的高性能大容量并行计算机技术相适应。因此,将 GA 与变分法结合在一起,就有可能更好的解决数值预报初始场的资料同化问题。

把 GA 用于变分资料的同化研究,最早是由国外学者 Ahrens 提出^[2],他用非标准化的 GA 对弱约束形式的 Lorenz 模型进行全局最小搜索,并与 Evensen 和 Fario^[13]的局部优化方法进行比较,取得了较好的结果,但弱约束形式中算法要求有大量的控制参数,因而算法的应用只针对简单系统。国内最早由王顺凤等^[14]将 GA 引入到气象资料变分同化中,为资料同化的研究提出新的思路,证明了 GA 解决气象资料变分同化问题的可行性。

本文的目的在于,研究相同条件下,分析比较基于 GA 的四维变分资料同化这一新的方案和伴随模式变分同化方案的同化效果,以此来讨论基于 GA 的变分同化方案解决气象资料变分同化问题的有效性及优越性。为了排除其他因素的影响,也为了节省计算时间,研究采用一个二维的浅水方程模式。第 2 节将讨论 GA 在四维变分资料同化中是如何实现的,第 3 节叙述数值试验的设计和结果,最后给出结论和讨论。

2 遗传算法在四维变分资料同化中的实现

2.1 遗传算法简介

GA 是一种宏观意义下的仿生算法,它通过模拟达尔文“优胜劣汰,适者生存”的原理鼓励产生好的结构,并通过模仿孟德尔遗传变异理论在迭代过程中保持已有的结构,同时寻找更好的结构的一种自适应全局优化的搜索算法。GA 以编码空间代替

问题的参数空间,以适应度函数为评价依据,以编码群体为进化基础,以对群体中个体染色体的遗传操作实现选择和遗传机制,来建立一个迭代过程,在这一过程中,通过随机重组编码染色体中重要的基因,使新一代的染色体集合优于老一代的染色体集合。GA 在整个进化过程中的遗传操作是随机性的,但它所呈现出的特性并不是完全随机搜索,它能有效地利用历史信息来推测下一代期望性能有所提高的寻优点集,这样一代代地不断进行,最后收敛到一个最适应环境的个体上,求得问题的最优解。参数编码、初始群体的设定、适应度函数及遗传操作设计和控制参数设定五个要素共同组成了 GA 的核心内容。标准 GA 的基本流程和结构如图 1 所示。

从图 1 可以看出,GA 的运行过程为一个典型的迭代过程,其基本步骤如下:

(1) 选择编码策略,把参数集合转换为染色体

- 结构空间;
- (2) 定义适应度函数;
 - (3) 确定遗传参数,包括群体大小、最大世代数、交叉概率、变异概率等,并确定选择、交叉、变异的方法;
 - (4) 随机初始化生成群体 P ;
 - (5) 计算群体中个体染色体解码后的适应值;
 - (6) 按照遗传策略,运用选择、交叉和变异算子作用于群体,形成下一代群体;
 - (7) 判断群体性能是否满足某一指标,或者已完成预定迭代次数,不满足则返回步骤(6),或者修改遗传策略再返回步骤(6)。

2.2 遗传算法在四维变分资料同化中的实现

本文提出一种能以较大概率在四维变分资料同化问题中成功求得全局最优解的 GA,该算法在四维变分资料同化过程中是通过以下工作流程来实现

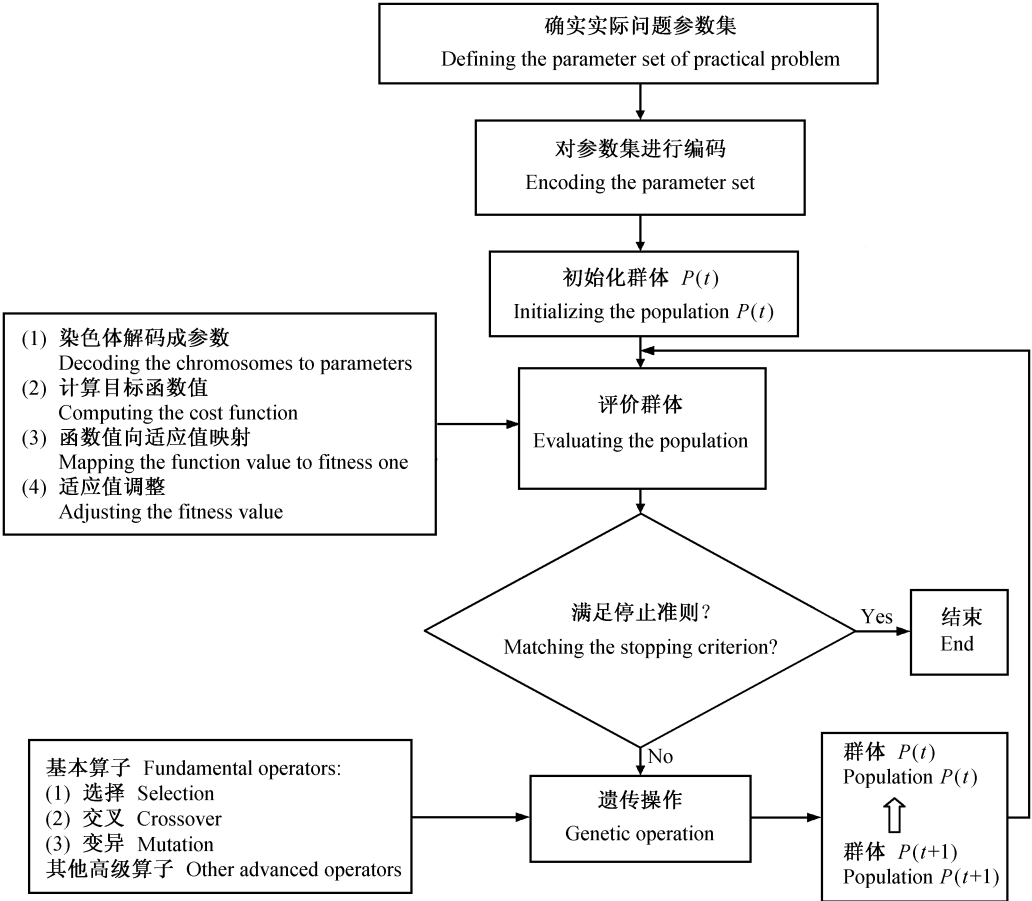


图 1 标准 GA 基本流程框图
Fig. 1 The basic flow diagram of standard GA

的。

2.2.1 参数编码

考虑到需要同化的物理变量的复杂性, 采用实值编码, 它是连续参数优化问题直接的自然描述, 不存在编码和解码过程, 在初始种群中容易定义并实现一个可行解, 并针对实值编码的特点进行尺度因子^[15]的选取, 以增加进化速度。

2.2.2 初始化种群

GA 与传统随机搜索算法的最大区别之一在于它的整个算法是在解的群体上进行的, 正是这一特点使 GA 具有了搜索过程的并行性、全局性和鲁棒性。可见, 群体的设定对整个 GA 的运行性能具有基础性的决定作用。选取理想场向量为 $\mathbf{Y}(t_0)$, 使用伪随机数来产生随机场中的误差, 使初估场尽可能体现其分布, 则初估场向量 $\mathbf{Y}_r(t_0)$ 为

$$\mathbf{Y}_r(t_0) = \mathbf{Y}(t_0)(1 + \alpha\mu), \quad r = 1, 2, \dots, N \quad (1)$$

其中, α 为服从某种分布的随机数, μ 为随机误差系数, N 为种群数。文中选用模式计算得出的解的轨迹来构造初始种群, 这样可使 GA 更快地达到收敛解。

2.2.3 适应性评价

GA 将问题空间表示为染色体空间, 为了执行适者生存的原则, 必须对个体的适应性进行评价, 因此适应度函数就构成了个体的生存环境。适应值是群体中个体生存机会选择的唯一确定性指标, 根据个体的适应值, 就可决定它在此环境下的生存能力。一般来说, 好的染色体结构具有比较高的适应函数值, 即可获得较高的评价, 具有较强的生存能力。因此, 适应度函数的形式直接决定着群体的进化行为, 并影响到 GA 的收敛速度及其能否找到最优解。对于给定的优化问题, 必须建立适应度函数与目标函数的映射关系, 保证映射后的适应值是非负的, 而且目标函数的优化方向对应于适应值增大的方向。因此, 适应度函数取为

$$f[\mathbf{Y}_r(t_0)] = \frac{1}{1 + D[\mathbf{Y}_r(t_0)]}, \quad r = 1, 2, \dots, N \quad (2)$$

其中, $D[\mathbf{Y}_r(t_0)]$ 是变分同化问题中的目标函数。

2.2.4 遗传操作

GA 的操作算子一般包括选择、交叉和变异三种基本形式, 它们构成了 GA 具备强大搜索能力的

核心, 是模拟自然选择以及遗传操作过程中发生的繁殖、杂交和变异现象的主要载体。GA 利用遗传算子产生新一代群体来实现群体进化, 算子的设计是遗传策略的主要组成部分, 也是调整和控制进化过程的基本工具。

(1) 选择 选择实现了达尔文的适者生存原则。选择的目的是为了从当前群体中选出优良的个体, 使它们有机会作为父代为下一代繁殖子孙。进行选择的原则是适应性强的个体为下一代贡献一个或多个后代的概率大。选择算子从群体中按某一概率成对选择个体, 某个体被选择的概率与其适应值成正比。

(2) 交叉 交叉操作是 GA 中最主要的遗传操作, 通过交叉操作可以得到新一代个体, 新个体组合了其父辈个体的特性, 体现了信息交换的思想。交叉算子将被选中的两个个体按概率进行交叉, 生成两个新的个体。

(3) 变异 变异本身是一种局部随机搜索, 它为新个体的产生提供了机会, 变异与选择、交叉算子结合在一起, 在保证 GA 的有效性和局部随机搜索能力的同时使 GA 保持种群的多样性, 以防止出现未成熟过早收敛。

上述各种算子的实现是多种多样的, 而且许多新的算子正在不断地提出, 以改进 GA 的某些性能。文中采用轮盘赌选择方法随机选择一对个体, 并结合变分同化的特点, 在交叉运算中采用自适应交叉概率^[16]和强制交叉概率^[17]相结合的算法向适应值大的方向产生新个体, 同时采用类似精英个体保留策略, 把适应值较大的父个体直接复制到下一代, 这样使得更好的个体得以保留, 从而保证算法的整体收敛性。然后, 在变异操作中引入最陡下降算子, 对交叉后的个体按自适应变异概率^[16]沿下降方向进行变异, 这种变异操作发挥了最陡下降法在计算速度与计算精度上的优势。可见, 这种交叉和变异操作在保证群体多样性的同时使群体一直向适应度大的方向进化, 加速 GA 在全局优化问题上的收敛性。本文自适应交叉概率 P_c 和自适应变异概率 P_m 采取如下表达式计算:

$$P_c = \begin{cases} P_{cl} - \frac{(P_{cl} - P_{c2})(f' - f_{avg})}{f_{max} - f_{avg}}, & f' \geq f_{avg} \\ P_{cl}, & f' < f_{avg} \end{cases} \quad (3)$$

$$P_m = \begin{cases} P_{m1} - \frac{(P_{m1} - P_{m2})(f_{\max} - f)}{f_{\max} - f_{\text{avg}}}, & f \geq f_{\text{avg}} \\ P_{m1}, & f < f_{\text{avg}} \end{cases} \quad (4)$$

其中, $P_{c1} = 1.0$, $P_{c2} = 0.9$, $P_{m1} = 0.7$, $P_{m2} = 0.001$, $f_{\max}$ 为群体中最大的适应值, f_{avg} 表示群体的平均适应值, f' 为交叉的两个个体中较大的适应值, f 表示变异个体的适应值。

2.2.5 控制参数

在运行 GA 程序时,需要对种群数 N (大群体含有较多模式,为 GA 提供足够的模式采样容积,可以改进 GA 搜索的质量,防止成熟前收敛。但大群体增加了个体适应性评价的计算量,从而使收敛速度降低)、交叉概率 P_c (交叉概率控制着交叉算子的应用频率,在每一代新的群体中,需要对 $P_c \times N$ 个个体的染色体结构进行交叉操作。交叉概率越高,群体中新结构的引入越快,已获得的优良基因结构的丢失速度也相应升高;而交叉概率太低则可能导致搜索阻滞)和变异概率 P_m (变异操作是保持群体多样性的有效手段。变异概率太小,可能使某些基因位过早丢失的信息无法恢复;而变异概率过高,则遗传搜索变成随机搜索)等遗传控制参数做事先选择,这些参数对 GA 的性能都有很重要的影响。对于具体问题而言,衡量参数设置恰当与否,要依据多次运行的收敛情况和解的质量来判断。

3 基于遗传算法的四维变分资料同化系统的数值试验

为了比较 GA 和伴随方法的同化效果,并希望完全排除其他因素的影响,在数值研究中采用了一个简单的二维浅水方程模式,需同化的物理变量包括风速分量 u 、 v 和位势高度场 z , 定义目标函数如下:

$$D = \frac{1}{2} \sum_{ijn} [W_u(u_{i,j,n} - u_{o,i,j,n})^2 + W_v(v_{i,j,n} - v_{o,i,j,n})^2 + W_z(z_{i,j,n} - z_{o,i,j,n})^2], \quad (5)$$

式中, W_u 、 W_v 和 W_z 是权重函数, u 、 v 及 z 为模式预报值, u_o 、 v_o 及 z_o 为观测值, 下标 i 、 j 表示格点位置, n 表示时间步数。采用二次守恒平流格式编制模式, 空间步长和时间步长分别为: $\Delta x = \Delta y = 300 \text{ km}$, $\Delta t = 600 \text{ s}$, 模式网格为 20×16 点, 边界

条件取固定边界条件。

由于大气模式及实际观测资料的误差都是不可知的, 最容易的办法是假设模式正确, “观测资料”是由模式产生的理想资料, 这样容易知道问题的准确解^[18]。同化方案如下: 以 1973 年 4 月 29 日 08 时(北京时, 下同)我国东北、华北地区 500 hPa 等压面作为理想初始高度场, 初始风场由地转关系得到, 其向前积分得到的结果作为同化期间的观测资料。就实际气象资料的观测误差而言, 可认为风场和气压场的观测误差是服从正态分布的^[19], 因此在理想初始场(图 2)上叠加一正态分布的扰动场作为初估场(图 3), 其向前积分得到的结果作为同化期间的模式预报资料, 假设同化过程每小时加一次观测值, 并在所有网格点上有效, 同化窗口分别为 3、6、12 小时。

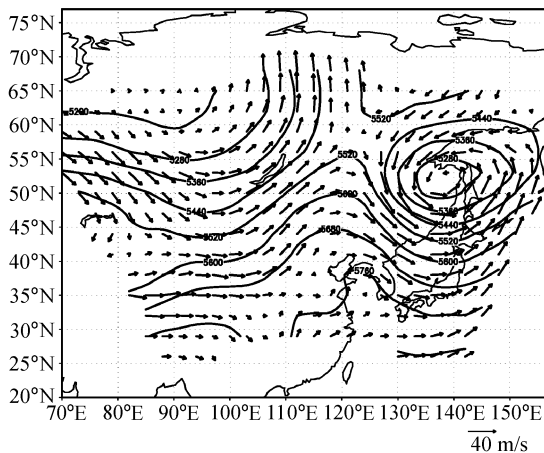


图 2 理想初始高度场及风场

Fig. 2 The ideal initial height and wind fields

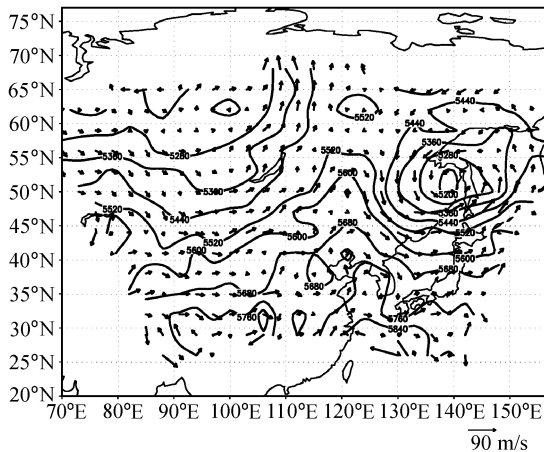


图 3 同化前的初估高度场和风场

Fig. 3 The initial height and wind fields before assimilation

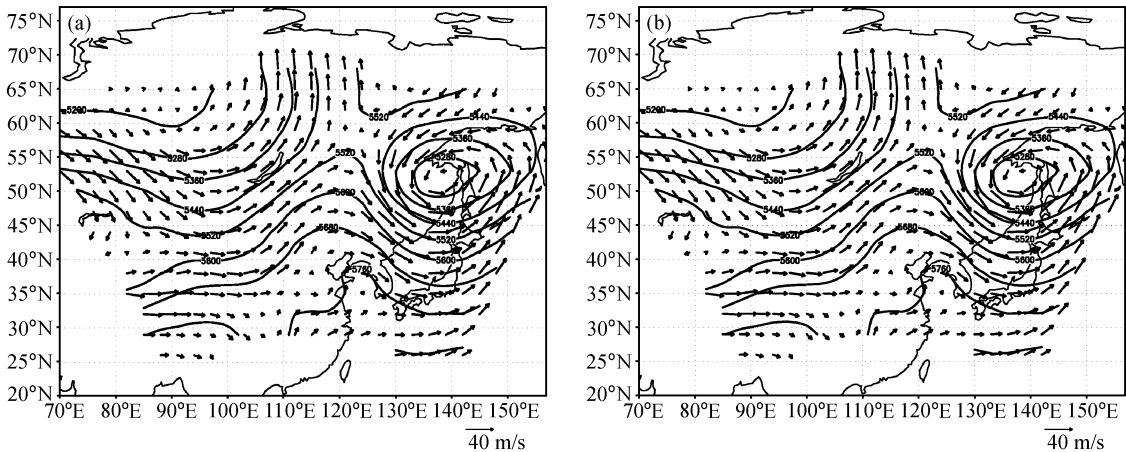


图 4 同化后的初始高度场和风场：(a) GA 同化；(b) 伴随同化
Fig. 4 The initial height and wind fields after (a) the GA-based and (b) adjoint VDA schemes

用基于 GA 的变分同化方案和伴随模式进行数值试验，随机误差系数 $\mu=0.1$ ，正态分布误差的期望 $U=0.0$ ，均方差 $G=0.1$ 。GA 中取权重系数 $W_u=W_v=10^{-2}$ ， $W_z=10^{-4}$ ，尺度因子 $S_u=S_v=10^4$ ， $S_z=10^8$ ，种群数 $N=120$ ，交叉概率 $P_{c1}=1.0$ ， $P_{c2}=0.9$ ，变异概率 $P_{m1}=0.7$ ， $P_{m2}=0.001$ ，最大世代数 $\text{Maxgen}=150$ 。伴随模式中取权重系数 $W_u=W_v=10^{-2}$ ， $W_z=10^{-4}$ ，尺度因子 $S_u=S_v=1$ ， $S_z=10^2$ 。下面考察在不同同化窗口下，从不同方面分析比较基于 GA 的变分同化方案与伴随模式变分同化方案对有限区域初始高度和风场的同化效果，以此来讨论基于 GA 的变分同化方案的有效性及其优越性。

图 4 给出同化窗口为 6 小时，采用基于 GA 的变分同化方案和伴随模式变分同化方案同化后所对应的初始场分布。从图 4a、b 可以看出，两种变分同化方案对加入正态分布误差后产生的初估场都具有很好的修正能力，同化后的初始场与理想初始场十分接近，同化前初始场的虚假高低中心和小波动在同化后得到了较大的纠正，同化效果明显。图 5 则直观地给出两种方案的最佳适应值随迭代步数的变化情况，从图中可以看出，基于 GA 的变分同化方案在迭代 20 步时最佳适应值已收敛，而伴随方法要迭代到 110 步才收敛，可见在合适的遗传控制参数的选取下，GA 随迭代步数的变化较伴随方法有较快的收敛。

表 1、2 定量给出基于 GA 的变分同化方案和伴随模式变分同化方案在不同的同化窗口下，同化

前后初始高度场和风场的均方根误差、最佳适应值、目标函数值及平均运行时间。从表 1 可以看出，基于 GA 的变分同化方案在不同的同化窗口下都具有较好的同化能力，同化前后高度场和风场的均方根误差的修正比都高达 99.999% 以上，而且它们的最佳适应值都趋近于 1，目标函数值都趋近于零。

同样，从表 2 可以看出，伴随模式变分同化方案在不同的同化窗口下也都具有较好的同化能力，其同化前后高度场和风场的均方根误差的修正比都高达 99.998% 以上，而且它们的目标函数值也都趋近于零，最佳适应值都趋近于 1。

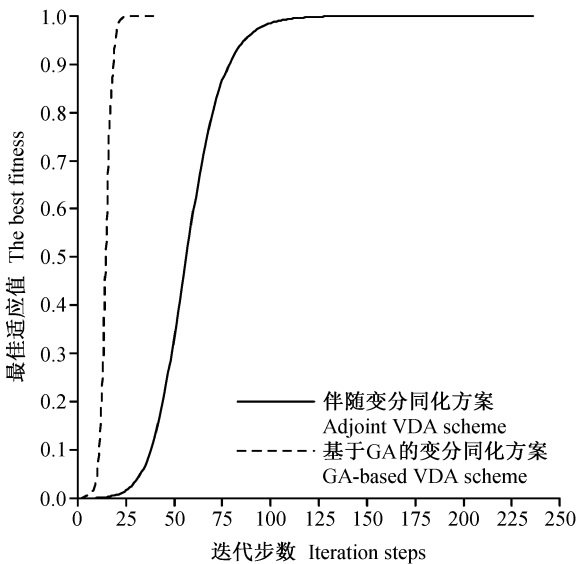


图 5 GA 和伴随方法的收敛速度
Fig. 5 The convergence velocity of the GA-based and adjoint VDA schemes

表 1 基于 GA 的变分同化系统的运行性能

Table 1 The performance of the GA based VDA scheme

	同化窗口 Assimilation window	均方根误差 RMSE			最佳适应值 Best fitness	目标函数值 Cost function value	平均运行时间 Average calculation time/s
		高度场 Height field	东西风场 WE wind field	南北风场 NS wind field			
同化前 Before	3 h	50. 7378942	19. 4423780	12. 0845659	0. 99999994	$7. 9850304 \times 10^{-8}$	150
同化后 After		$1. 4388775 \times 10^{-4}$	$8. 9051574 \times 10^{-5}$	$6. 9688675 \times 10^{-5}$			
修正比 Ratio		99. 9997164%	99. 9995420%	99. 9994233%			
同化前 Before	6 h	50. 8166784	19. 3980093	12. 0945393	0. 99999993	$6. 0985925 \times 10^{-8}$	220
同化后 After		$9. 6952785 \times 10^{-5}$	$1. 2674686 \times 10^{-4}$	$5. 0608128 \times 10^{-5}$			
修正比 Ratio		99. 9998092%	99. 9993466%	99. 9995816%			
同化前 Before	12 h	50. 8296068	19. 3832135	12. 0968061	0. 99999993	$7. 3844960 \times 10^{-8}$	390
同化后 After		$9. 1779120 \times 10^{-5}$	$1. 5685661 \times 10^{-4}$	$1. 2982066 \times 10^{-4}$			
修正比 Ratio		99. 9998194%	99. 9991908%	99. 9989268%			

表 2 伴随模式变分同化系统的运行性能

Table 2 The performance of the adjoint VDA scheme

	同化窗口 Assimilation window	均方根误差 RMSE			目标函数值 Cost function value	最佳适应度 Best fitness	运行时间 Calculation time/s
		高度场 Height field	东西风场 WE wind field	南北风场 NS wind field			
同化前 Before	3 h	50. 5311871	19. 5479842	12. 0385700	$8. 3585748 \times 10^{-8}$	0. 99999992	35
同化后 After		7. 7486683	4. 7821633	8. 8502584			
修正比 Ratio		99. 9984666%	99. 9997554%	99. 9992648%			
同化前 Before	6 h	50. 5311871	19. 5479842	12. 0385700	$9. 6749013 \times 10^{-8}$	0. 99999990	250
同化后 After		$9. 6958872 \times 10^{-4}$	$6. 7487564 \times 10^{-5}$	$1. 1520427 \times 10^{-4}$			
修正比 Ratio		99. 9980812%	99. 9996548%	99. 9990430%			
同化前 Before	12 h	50. 5311871	19. 5479842	12. 0385700	$9. 8047424 \times 10^{-8}$	0. 99999990	1280
同化后 After		$1. 0207927 \times 10^{-3}$	$7. 0355943 \times 10^{-5}$	$1. 2355713 \times 10^{-4}$			
修正比 Ratio		99. 9979799%	99. 9996401%	99. 9989737%			

比较表 1 和表 2 可以发现, 基于 GA 的变分同化方案和伴随模式变分同化方案都具有较好的同化效果。以二维浅水方程为例的理想数值试验方案及个例来看, 基于 GA 的变分同化方案在对初估场的同化效果和目标函数的收敛精度上与伴随模式变分同化方案相当。在运行时间上, 当同化窗口为 3 h, 基于 GA 的变分同化是伴随同化时间的 4 倍; 当同化窗口为 6 h, 基于 GA 的变分同化与伴随同化时间相当; 而当同化窗口为 12 h, 基于 GA 的变分同化是伴随同化时间的 1/3。造成这种现象的原因是因为 GA 在整个解空间上进行的一种随机化搜索过程, 无论是同化几小时, 首先它都必须对初始种群中的每个个体进行适应性评价, 在进行每一次的交叉、变异操作过程后, 它都必须对遗传操作过的个体进行向前积分, 然后再进行适应性搜索, 直到找到当前代群体中的最佳个体, 才进行下一次遗传进

化过程, 所以对不同的同化窗口时要消耗相当的计算机机时, 其同化窗口为 3 h、6 h、12 h 的运行时间比值为 1 : 1. 5 : 2. 6; 而伴随同化方法则是对每一次下降算法的迭代, 先把预报模式向前积分一次, 然后把伴随方程向后积分到初始时刻, 得到需要的梯度, 根据下降算法的要求通过该梯度对初值进行修正后再重复前面的过程, 直到目标函数达到极小值后, 得到的模式解就是所需的分析场。因此, 在同化时间不长时, 它的优越性较明显, 但随着同化时间的增加, 伴随模式则需要较大的存贮量与计算量, 其同化窗口为 3 h、6 h、12 h 的运行时间比值为 1 : 7 : 36. 6。

图 6 比较了不同的同化窗口下, 基于 GA 的变分同化方案和伴随模式变分同化方案的收敛速度, 从图中可以看出对不同的同化窗口, GA (图 6a) 迭代到近 25 步就收敛到最佳适应值, 而伴随模式

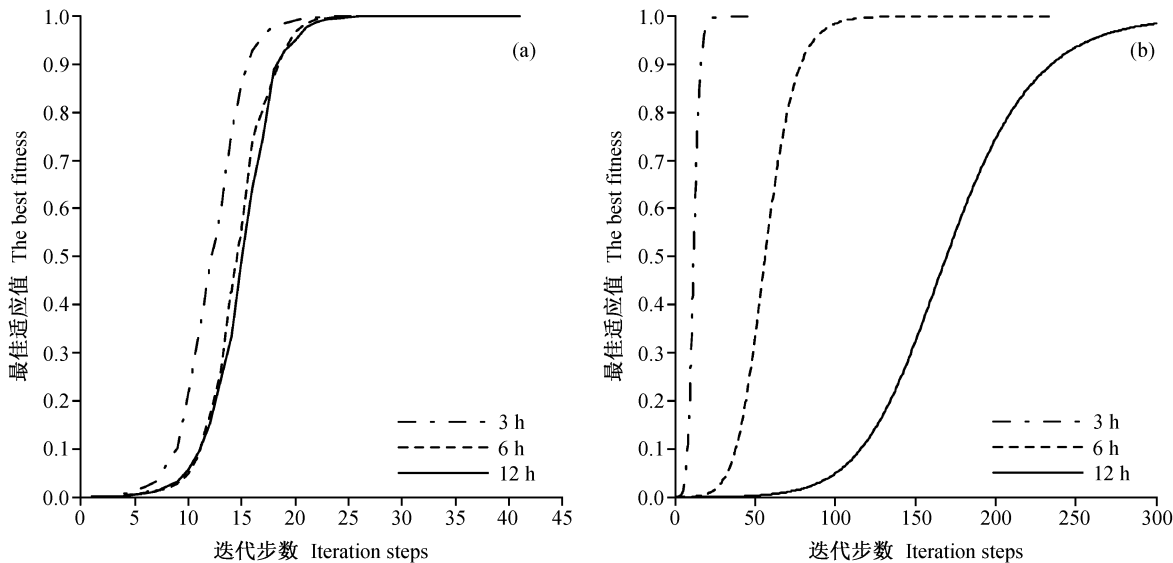


图 6 不同变分同化方案的收敛速度：(a) GA 方法；(b) 伴随方法
Fig. 6 The convergence velocity of different variational assimilation schemes; (a) The GA-based; (b) the adjoint

(图 6b) 在同化时间不长时其收敛较快，但随着同化时间的增加，其迭代步数也增加。

一般来讲，理论上 GA 可以在合理的计算时间内找到满足条件的最优解，但是随着研究问题维数的不断增加及复杂程度的不断提高，尤其是针对气象这一复杂的有约束的高维非线性规划问题，还有许多工作有待于进一步深入研究，比如，关于约束项的更有效处理，复杂物理过程的处理，遗传参数、遗传算子以及收敛准则的确定等方面都还存在很多理论和技术上的难点，需要做大量的理论研究和数值试验。

4 结论和讨论

本文将 GA 应用于四维变分资料同化问题中，提供一种新的较为有效的 NWP 初始场优化方法，给出相应的理论依据和详细算法，并结合问题本身的特点设计合理的遗传编码、遗传操作和遗传参数，同时以二维浅水方程为例建立了基于 GA 的变分同化方案，并与伴随模式变分同化方案从多方面进行分析比较，数值试验分析结果表明：

(1) 以二维浅水方程模式为约束条件，构造基于 GA 的变分同化系统，取得了比较满意的同化效果。其试验结果表明，将 GA 应用于四维变分资料同化这一工作丰富了变分同化的内容，对解决变分同化问题是一次有益的尝试，使变分同化技术作为

提高数值预报初始场质量的这一有效技术得到了进一步的应用。

(2) 基于 GA 的变分同化方案和伴随模式变分同化方案都具有较好的同化效果。从目前的试验方案及个例分析结果来看，基于 GA 的变分同化方案在目标函数的收敛精度上与伴随模式变分同化方案相当，当同化窗口为 6 h 时其收敛速度与伴随模式相当，但同化窗为 12 h 的 GA 变分同化方案优于伴随模式方案。

(3) 与伴随模式的变分同化系统相比，基于 GA 的变分同化系统对目标函数的数学性态要求上有了极大改进，它只利用适应性信息，不需要梯度等阶信息，而且随着复杂模式系统及复杂物理过程的加入，目标函数的结构变得越来越复杂。这时，伴随模式同化方法由于梯度信息的约束，可能使目标函数陷入局部最优值而造成迭代不收敛，而基于 GA 的变分同化系统则可以较好地避免这一情况的发生，这使得基于 GA 的变分同化系统有了更广泛的适用范围和应用前景。

总之，从目前试验方案及个例来看，基于 GA 的变分同化方案具有较好的同化效果。数值试验结果表明，基于 GA 的变分同化方案是具有一定优越性的，它为以后的研究提供了广阔的应用前景。

致谢 在撰写过程中曾受到南京大学数学系[孙麟平]教授的悉心指导，在此表示深深的谢意。

参考文献 (References)

- [1] 罗乔林. 数据处理之进展——见于气象四维同化问题. 系统工程理论与实践, 1997, **17**(9): 98~103
Luo Qiaolin. Data processing in meteorological data assimilation. *Systems Engineering-Theory and Practice* (in Chinese), 1997, **17**(9): 98~103
- [2] Ahrens B. Variational data assimilation for a Lorenz model using a non-standard genetic algorithm. *Meteorology and Atmospheric Physics*, 1999, **70**: 227~238
- [3] 王小平, 曹立明. 遗传算法——理论、应用与软件实现. 西安: 西安交通大学出版社, 2002. 76pp
Wang Xiaoping, Cao Liming. *Genetic Algorithm—Theory, Application and Software Implementation* (in Chinese). Xi'an: Xi'an Jiaotong University Press, 2002. 76pp
- [4] 沈桐立, 田永祥, 葛孝贞, 等. 数值天气预报. 北京: 气象出版社, 2003. 241pp
Shen Tongli, Tian Yongxiang, Ge Xiaozhen, et al. *Numerical Weather Forecast* (in Chinese). Beijing: China Meteorological Press, 2003. 241pp
- [5] Holland J H. Genetic algorithms. *Scientific American*, 1992, **7**: 44~50
- [6] Barth N H. Oceanographic experiment design II: Genetic algorithms. *J. Atmos. Oceanic Technol.*, 1992, **9** (4): 434~443
- [7] Hernandez F, Le Traon P-Y, Barth N H. Optimizing a drift-cast strategy with a genetic algorithm. *J. Atmos. Oceanic Technol.*, 1995, **12** (2): 330~345
- [8] Lakshmanan V. Using a genetic algorithm to tune a bounded weak echo region detection algorithm. *J. Appl. Meteor.*, 2000, **39** (2): 222~230
- [9] 李敏强, 寇纪淞, 林丹. 遗传算法的基本理论与应用. 北京: 科学出版社, 2002. 5
Li Minqiang, Kou Jisong, Lin Dan. *Genetic Algorithm—Basic Theory and Application* (in Chinese). Beijing: Science Press, 2002. 5
- [10] 玄光男, 程润伟. 遗传算法和工程优化. 北京: 清华大学出版社, 2004. 21
Xuan Guangnan, Cheng Runwei. *Genetic Algorithm and Engineering Optimization* (in Chinese), Beijing: Tsinghua University Press, 2004. 21
- [11] Mwale D, Gan T. Wavelet analysis of variability, teleconnectivity, and predictability of the September–November East African rainfall. *J. Appl. Meteor.*, 2005, **44**(2): 256~269
- [12] 刘峰, 胡非. 用遗传算法优化工业污染源布局. 城市环境与城市生态, 2003, **16**(6): 158~160
Liu Feng, Hu Fei. Genetic algorithm to optimize planning of industrial pollutant sources. *Urban Environment and Urban Ecology* (in Chinese), 2003, **16** (6): 158~160
- [13] Evensen G, Fario N. Solving for the generalized inverse of the Lorenz model. *J. Meteor. Soc. Japan*, 1997, **75**: 229~243
- [14] 王顺凤, 沈桐立, 张建伟, 等. 基于进化方向遗传算法的四维变分资料同化方法. 南京气象学院学报, 2002, **25**(6): 740~746
Wang Shunfeng, Shen Tongli, Zhang Jianwei, et al. A method of four-dimensional variational data assimilation based on the evolution direction genetic algorithm. *Journal of Nanjing Institute of Meteorology* (in Chinese), 2002, **25** (6): 740~746
- [15] 马刚, 王强, 邱崇践. 用变分法修正黑河试验数据的初步研究. 大气科学, 2002, **26**(2): 271~278
Ma Gang, Wang Qiang, Qiu Chongjian. A research to validate the data from Heihe River field experiment by variation. *Chinese Journal of Atmospheric Sciences* (in Chinese), 2002, **26**(2): 271~278
- [16] Srinivas M, Patnaik L M. Adaptive probabilities of crossover and mutation in genetic algorithms. *IEEE Transactions on Systems, Man and Cybernetics*, 1994, **24**(4): 656~667
- [17] 郑双泉, 章兢. 改进的遗传算法及其在优化 IC 中的应用. 湖南教育学院学报, 1997, **15**(2): 44~48
Zheng Shuangquan, Zhang Jing. A modified genetic algorithm and its application in the optimization of intelligent controller. *Journal of Hunan Educational Institute* (in Chinese), 1997, **15** (2): 44~48
- [18] 邱崇践. 变分同化中使用背景场时尺度匹配的数值研究. 大气科学, 2001, **25**(1): 103~109
Qiu Chongjian. Matching scales of the retrieved variables with background scale in variational assimilation: a numerical study. *Chinese Journal of Atmospheric Sciences* (in Chinese), 2001, **25**(1): 103~110
- [19] 王必正, 曾庆存, 穆穆. 伴随方程在水汽资料四维同化中的应用 I. 理论. 气候与环境研究, 2000, **5**(3): 273~278
Wang Bizheng, Zeng Qingcun, Mu Mu. The four-dimension assimilation problem of water vapor by means of the adjoint equation Part I. Theory. *Climatic and Environmental Research* (in Chinese), 2000, **5**(3): 273~278